Structure of the Picornavirus Replication Platform: A Potential Drug Target for Inhibiting Virus Replication

Meghan Suzanne Warden
Old Dominion University, mbrown_34@hotmail.com

Follow this and additional works at: https://digitalcommons.odu.edu/chemistry_etds

Part of the Biochemistry Commons, Chemistry Commons, Epidemiology Commons, and the Physiology Commons

Recommended Citation
Warden, Meghan S.. "Structure of the Picornavirus Replication Platform: A Potential Drug Target for Inhibiting Virus Replication" (2018). Doctor of Philosophy (PhD), Dissertation, Chemistry & Biochemistry, Old Dominion University, DOI: 10.25777/wyvk-8b21
https://digitalcommons.odu.edu/chemistry_etds/22

This Dissertation is brought to you for free and open access by the Chemistry & Biochemistry at ODU Digital Commons. It has been accepted for inclusion in Chemistry & Biochemistry Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
STRUCTURE OF THE PICORNAVIRUS REPLICATION PLATFORM:
A POTENTIAL DRUG TARGET FOR INHIBITING VIRUS REPLICATION

by
Meghan Suzanne Warden
B.S. May 2011, Lambuth University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY
CHEMISTRY

OLD DOMINION UNIVERSITY
December 2018

Approved by:
Steven M. Pascal (Director)
Lesley H. Greene (Member)
Hameeda Sultana (Member)
James W. Lee (Member)
John B. Cooper (Member)
ABSTRACT

STRUCTURE OF THE PICORNAVIRUS REPLICATION PLATFORM:
A POTENTIAL DRUG TARGET FOR INHIBITING VIRUS REPLICATION

Meghan Suzanne Warden
Old Dominion University, 2018
Director: Dr. Steven M. Pascal

Picornaviruses are small, positive-stranded RNA viruses, divided into twelve different genera. Members of the Picornaviridae family cause a wide range of human and animal diseases including the common cold, poliomyelitis, foot and mouth disease, and dilated cardiomyopathy. The picornavirus genome is replicated via a highly conserved mechanism involving a presumed cloverleaf structure located at the 5’ noncoding region of the virus genome. The 5’ cloverleaf consists of three stem loops (B, C, and D) and one stem (A), which interact with a variety of virus and host cell proteins during replication. In this dissertation, human rhinovirus serotype 14 (HRV-14) SLB and the 5’cloverleaf (5’CL) solution structures were determined using nuclear magnetic resonance (NMR) and small-angle X-ray scattering (SAXS). HRV-14 SLB adopts a predominantly A-form helical structure. The stem contains five Watson–Crick base pairs and one wobble base pair and is capped by an eight-nucleotide loop. The wobble base pair introduces perturbations in the helical parameters, but does not appear to introduce flexibility. The helix major groove appears to be more accessible than in typical standard A-form RNA. Flexibility is seen throughout the loop and in the terminal nucleotides. The pyrimidine-rich region of the loop, the apparent recognition site for the poly(C) binding protein, is the most disordered region of the structure.
The solution structure of HRV-14 5'CL was determined in the absence and presence of magnesium. In the absence of magnesium, the structure adopts an open, somewhat extended conformation. In the presence of magnesium, the structure compacts, bringing SLB and SLD into close contact, a geometry that creates an extensive accessible major groove surface, and permits interaction between the proteins that target each stem loop. A deeper understanding of these structures will offer invaluable information regarding the picornavirus replication mechanism. The results from these studies have the potential to elucidate unique drug targets with broad spectrum efficacy against a range of picornaviruses.
Copyright, 2018, by Meghan Suzanne Warden, All Rights Reserved.
This dissertation is dedicated to my family for their infinite love and support and also for pretending to understand my research.
ACKNOWLEDGMENTS

There are many people who I would like to thank for their contributions to the success of this dissertation. I would first like to thank my advisor, Dr. Steven Pascal, for his expert guidance, support and patience during my time in his lab. Without his encouragement and commitment to my success this dissertation would not have been possible. Second, I would like to thank Dr. Komala Ponniah for her continuous training and expert advice in the lab. I would like to acknowledge the previous and current Pascal Lab members for the entertaining conversations, companionship, and lending a helping hand whenever possible. To my committee members, Dr. Greene, Dr. Sultana, Dr. Lee, and Dr. Cooper for endless support and guidance to the success of this dissertation.

I would like to thank our collaborators at the National Magnetic Resonance Facility at Madison (NMRFAM). Dr. Gabriel Cornilesceu and Dr. Marco Tonelli for performing the NMR experiments for both projects presented in this dissertation. Dr. Kai Cai, for performing the SAXS data and analysis, and Dr. Samuel Butcher for expert advice on RNA preparation and structure calculations. To Isaiah Ruhl and Dr. Ravi Garimella at COSMIC for guidance in the set-up and troubleshooting of various NMR experiments. Lastly, I would like to thank my loving husband Andy for his never-ending support, love, and encouragement. None of my success would be possible without you by my side.
NOMENCLATURE

3D  RNA-dependent RNA polymerase
RV  Rhinovirus
NCR  Non-coding region
IRES  Internal ribosome entry site
ICAM  Intracellular adhesion molecule
LDLR  Low density lipoprotein receptor
VPg  Viral protein genome linked
PCBP  Poly(C)-binding protein
SA  Stem A
SLB  Stem loop B
SLC  Stem loop C
SLD  Stem loop D
PABP  Poly(A)-binding protein
HRV-14  Human rhinovirus serotype
CVB3  Coxsackievirus B3
NMR  Nuclear magnetic resonance
SAXS  Small-angle X-ray scattering
CL  Cloverleaf
TBE  Tris-borate-EDTA
EDTA  Ethylenediaminetetraacetic acid
kDa  Kilodalton
NOESY  Nuclear Overhauser effect spectroscopy
TOCSY  Total coorelation spectroscopy
HSCQ  Heteronuclear single quantum coherence
RDC  Residual dipolar coupling
ARTSY  Amide RDCs by TROSY spectroscopy
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>NMPs</td>
<td>Nucleotide monophosphates</td>
</tr>
<tr>
<td>WC</td>
<td>Watson-Crick</td>
</tr>
<tr>
<td>KH</td>
<td>K-homology</td>
</tr>
<tr>
<td>P(r)</td>
<td>Pairwise distance distribution function</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>U</td>
<td>Uracil</td>
</tr>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>Rg</td>
<td>Radius of gyration</td>
</tr>
<tr>
<td>ddH₂O</td>
<td>Double distilled water</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. STRUCTURE OF RNA STEM LOOP B FROM THE PICORNAVIRUS REPLICATION PLATFORM</td>
<td>16</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>16</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>18</td>
</tr>
<tr>
<td>RESULTS</td>
<td>28</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>39</td>
</tr>
<tr>
<td>III. CONFORMATIONAL FLEXIBILITY IN THE PICORNAVIRUS REPLICATION PLATFORM</td>
<td>46</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>46</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>49</td>
</tr>
<tr>
<td>RESULTS</td>
<td>54</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>66</td>
</tr>
<tr>
<td>IV. CONCLUSIONS AND FUTURE WORK</td>
<td></td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>76</td>
</tr>
<tr>
<td>FUTURE WORK</td>
<td>76</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>78</td>
</tr>
<tr>
<td>APPENDICES</td>
<td></td>
</tr>
<tr>
<td>A. CHEMICAL SHIFTS OF SLB AND 5'CL</td>
<td>89</td>
</tr>
<tr>
<td>B. PLASMID PREPARATION OF THE 5'CL</td>
<td>91</td>
</tr>
<tr>
<td>C. MC-SYM PROTOCOL FOR STARTING STRUCTURES</td>
<td>93</td>
</tr>
<tr>
<td>VITA</td>
<td>96</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1. Picornavirus Classification</td>
<td>2</td>
</tr>
<tr>
<td>2. Summary of Structural Constraints and Refinement Statistics</td>
<td>39</td>
</tr>
<tr>
<td>3. Structural Statistics for HRV-14 5’CL</td>
<td>69</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Schematic representation of the <em>Enterovirus</em> genome</td>
<td>5</td>
</tr>
<tr>
<td>2.</td>
<td>Circularization of the virus genome</td>
<td>6</td>
</tr>
<tr>
<td>3.</td>
<td>Polyprotein processing during enterovirus translation</td>
<td>7</td>
</tr>
<tr>
<td>4.</td>
<td>Secondary structure predictions of SLD from (A) human rhinovirus B14, (B) coxsackievirus B3 and (C). an enterovirus consensus structure</td>
<td>11</td>
</tr>
<tr>
<td>5.</td>
<td>NMR solution structures of SLD from (A) HRV-14, (B) CVB3 and (C) enterovirus consensus</td>
<td>12</td>
</tr>
<tr>
<td>6.</td>
<td>NMR structure of HRV-14 SLD and 3C</td>
<td>14</td>
</tr>
<tr>
<td>7.</td>
<td>Secondary structure prediction of HRV-14 5’CL</td>
<td>23</td>
</tr>
<tr>
<td>8.</td>
<td>Assigned imino proton portion of the 250 ms $^1$H,$^1$H-NOESY spectrum at 278 K of HRV-14 SLB in H$_2$O with the 1D $^1$H NMR spectrum shown above</td>
<td>34</td>
</tr>
<tr>
<td>9.</td>
<td>Base to H1’/H5 region of a 250 ms $^1$H,$^1$H-2D NOESY spectrum of SLB in D$_2$O</td>
<td>35</td>
</tr>
<tr>
<td>10.</td>
<td>Superposition of the final ten structures of HRV-14 SLB determined in this study</td>
<td>40</td>
</tr>
<tr>
<td>11.</td>
<td>(A) Comparison between SLB $^{13}$C chemical shifts and chemical shift of nucleotide monophosphates (NMPs)</td>
<td>43</td>
</tr>
<tr>
<td>12.</td>
<td>A space-filling view of the average calculated structure of HRV-14 SLB</td>
<td>47</td>
</tr>
<tr>
<td>13.</td>
<td>Secondary structure prediction of HRV-14 5’CL</td>
<td>53</td>
</tr>
<tr>
<td>14.</td>
<td>Imino NOESY spectra in the absence (blue) and presence (red) of magnesium.</td>
<td>60</td>
</tr>
<tr>
<td>15.</td>
<td>$^1$H chemical shift changes within SLB (magenta) and within SLD (green) induced (A) by incorporation into the 5’CL (without magnesium present) and (B) by the addition of magnesium (in the context of the 5’CL)</td>
<td>63</td>
</tr>
<tr>
<td>16.</td>
<td>SAXS data plots for HRV-14 5’CL</td>
<td>65</td>
</tr>
</tbody>
</table>
17. Structure of HRV-14 5´CL .................................................................68
18. Superposition of 5´CL and SAXS \textit{ab initio} model ..........................70
19. Changes in orientation induced by magnesium..............................75
20. Conformational subgroups in the absence of magnesium..............76
21. Major groove width of HRV-14 5´CL ............................................79
22. RNA-protein interaction modeling..............................................80
23. RNA-protein interaction modeling..............................................81
CHAPTER I
INTRODUCTION

Picornaviruses

Picornaviruses, meaning literally small RNA viruses, contain a single-stranded, positive sense RNA genome, of the same sense as mRNA. The highly conserved replication mechanism of the genome does not involve a DNA intermediate\textsuperscript{1,2}. Instead, the RNA is directly replicated via a virus encoded RNA replicase (3D). The \textit{Picornaviridae} family consists of more than 28 different species grouped into 12 genera; \textit{Enterovirus}, \textit{Cardiovirus}, \textit{Aphthovirus}, \textit{Hepatovirus}, \textit{Parechovirus}, \textit{Erbovirus}, \textit{Kobuvirus}, \textit{Teschovirus}, \textit{ Sapelovirus}, \textit{Senecavirus}, and \textit{Tremovirus}\textsuperscript{3}. The classification, species, number of types and the associated diseases are shown in Table 1.

Picornaviruses cause afflictions in vertebrates, primarily mammals, who serve as the natural hosts for the \textit{Picornaviridae} family. Poliomyelitis, a human disease caused by the poliovirus, has proof of existence dating back to 1300 BC where ancient Egyptian paintings depicted children with walking sticks and deformed limbs\textsuperscript{4}. Picornaviruses also include several other debilitating human and animal diseases such as hepatitis A, cardiomyopathy, foot-and-mouth disease and the common cold.
### TABLE 1: Picornavirus Classification

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species</th>
<th>No. of types</th>
<th>Associated disease(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enterovirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Human enterovirus A</em></td>
<td>21</td>
<td>Herpangina, paralytic disease, aseptic meningitis, encephalitis, ataxia</td>
</tr>
<tr>
<td></td>
<td><em>Human enterovirus B</em></td>
<td>59</td>
<td>Epidemic pleurodynia, aseptic meningitis, paralytic disease, encephalitis</td>
</tr>
<tr>
<td></td>
<td><em>Human enterovirus C</em></td>
<td>19</td>
<td>Acute anterior poliomyelitis, respiratory disease, aseptic meningitis</td>
</tr>
<tr>
<td></td>
<td><em>Human enterovirus D</em></td>
<td>3</td>
<td>Hand-foot-and-mouth disease, acute hemorrhagic conjunctivitis</td>
</tr>
<tr>
<td></td>
<td><em>Simian enterovirus A</em></td>
<td>1</td>
<td>None known</td>
</tr>
<tr>
<td></td>
<td><em>Bovine enterovirus</em></td>
<td>2</td>
<td>Abortion, stillbirth, infertility, neonatal death, enteritis, respiratory disease</td>
</tr>
<tr>
<td></td>
<td><em>Porcine enterovirus B</em></td>
<td>2</td>
<td>Skin lesions</td>
</tr>
<tr>
<td></td>
<td><em>Human rhinovirus A</em></td>
<td>75</td>
<td>Respiratory disease</td>
</tr>
<tr>
<td></td>
<td><em>Human rhinovirus B</em></td>
<td>25</td>
<td>Respiratory disease</td>
</tr>
<tr>
<td></td>
<td><em>Human rhinovirus C</em></td>
<td>10</td>
<td>Respiratory disease</td>
</tr>
<tr>
<td><strong>Cardiovirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Encephalomyocarditis virus</em></td>
<td>1</td>
<td>Myocarditis, encephalitis, reproductive failure</td>
</tr>
<tr>
<td></td>
<td><em>Theilovirus</em></td>
<td>12</td>
<td>Polioencephalomyelitis, neurological disease</td>
</tr>
<tr>
<td><strong>Aphthovirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Foot-and-mouth disease virus</em></td>
<td>7</td>
<td>Vesicular disease, myocarditis</td>
</tr>
<tr>
<td></td>
<td><em>Equine rhinitis A virus</em></td>
<td>1</td>
<td>Respiratory disease, abortions</td>
</tr>
<tr>
<td></td>
<td><em>Bovine rhinitis B virus</em></td>
<td>1</td>
<td>Mild respiratory disease</td>
</tr>
<tr>
<td><strong>Heptavirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Hepatitis A virus</em></td>
<td>1</td>
<td>Hepatitis</td>
</tr>
<tr>
<td><strong>Parechovirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Human parechovirus</em></td>
<td>14</td>
<td>Encephalitis, meningitis, neonatal sepsis, enteritis, respiratory disease</td>
</tr>
<tr>
<td></td>
<td><em>Ljungan virus</em></td>
<td>4</td>
<td>Diabetes</td>
</tr>
<tr>
<td><strong>Erbovirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Equine rhinitis B virus</em></td>
<td>3</td>
<td>Respiratory disease</td>
</tr>
<tr>
<td><strong>Kobuvirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Aichi virus</em></td>
<td>1</td>
<td>Gastroenteritis</td>
</tr>
<tr>
<td></td>
<td><em>Bovine kobuvirus</em></td>
<td>1</td>
<td>None known</td>
</tr>
<tr>
<td><strong>Teschovirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Porcine teschovirus</em></td>
<td>11</td>
<td>Polioencephalomyelitis, reproductive failure, enteritis</td>
</tr>
<tr>
<td><strong>Sapelovirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Porcine sapelovirus</em></td>
<td>1</td>
<td>Enteritis</td>
</tr>
<tr>
<td></td>
<td><em>Simian sapelovirus</em></td>
<td>3</td>
<td>None known</td>
</tr>
<tr>
<td></td>
<td><em>Avian sapelovirus</em></td>
<td>1</td>
<td>None known</td>
</tr>
<tr>
<td><strong>Senecavirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Seneca Valley virus</em></td>
<td>1</td>
<td>None known</td>
</tr>
<tr>
<td><strong>Tremovirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Avian encephalomyelitis virus</em></td>
<td>1</td>
<td>Encephalomyelitis</td>
</tr>
<tr>
<td><strong>Avihepatovirus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Duck hepatitis A virus</em></td>
<td>3</td>
<td>Hepatitis</td>
</tr>
</tbody>
</table>

Table adapted from Ehrenfeld et al. 2010®
Picornavirus infections cause major public and economic burdens worldwide. Although vaccinations have reduced the incidence of polio by 99%, recent outbreaks have occurred in Nigeria, Pakistan, and Afghanistan. Hepatitis A is an infectious disease of the liver and is preventable by vaccination. However, the virus is endemic in parts of the world and outbreaks have recently increased in California due to the rising homeless population and fecal-contaminated environments. The coxsackievirus B3 is the leading cause of myocarditis, an inflammation of the heart muscle typically affecting young, otherwise healthy individuals. The condition is initially asymptomatic and normally resolves without treatment. Yet, in some cases the condition progresses to dilated cardiomyopathy ultimately resulting in congestive heart failure. Foot-and-mouth disease, caused by the foot-and-mouth disease virus, is a highly contagious condition affecting the health and weight of livestock. If the disease is identified in an area, millions of infected and susceptible animals in the area are typically slaughtered, to prevent further spread. The prevention itself causes devastating economical affects.

Human rhinoviruses are responsible for over 50% of common cold infections. Although the common cold was thought to be a nonthreatening upper respiratory infection, it is now evident that human rhinoviruses are linked to many other secondary infections, including exacerbations of chronic pulmonary disease, asthma development, bronchiolitis and pneumonia. Currently, there are no vaccines available for the common cold and treatment remains predominantly over-the-counter, with medication aimed to help relieve symptoms. Considering the prevalence and the global impact of human rhinoviruses, finding a treatment and preventative agent would have significant positive effects on public health.
Human Rhinovirus

This project focuses on the human rhinovirus (HRV) as a model system for picornavirus replication. As such, additional background on HRV is warranted. Rhinoviruses, from the Greek word rhinos meaning nose, are members of the Enterovirus genus, first discovered by Christopher Howard Andrewes in 1950, with over 200 different serotypes now confirmed. Currently, there are three classifications of rhinoviruses (RV), RV-A, RV-B, and RV-C. These species were once assigned to their own genus, Human Rhinovirus, because the associated disease (common cold) was very much different from other diseases caused by enteroviruses (such as polio and myocarditis).

In 1985, the first crystal structure of a human rhinovirus (B14 serotype) was determined in conjunction with the first full sequence of the whole genome. It became evident that, the global organization of the rhinovirus genome was identical to poliovirus. Therefore, at that time, human rhinoviruses were reclassified as species within the Enterovirus genus.

HRV Genome

Genetic information is carried as a small, single-stranded RNA molecule (~7–8.5 kb) that is replicated via a highly conserved mechanism. The rhinovirus genomic RNA consists of three distinct regions: the 5´ and 3´ noncoding regions (NCRs) flanking a gene encoding a single ~250 kDa polyprotein (Figure 1). Every picornavirus genome has complex secondary structures located in the 5´NCR. The first domain provides the binding site for host and virus proteins involved in RNA replication. In enteroviruses,
FIGURE 1: Schematic representation of the Enterovirus genome.

This first domain folds into a presumed cloverleaf-shaped structure (~83-86 bases) that when bound to proteins, facilitates the circularization of the virus genome, an essential step in virus replication (Figure 2)\textsuperscript{19}. The second domain in the 5'NCR contains the internal ribosome entry site (IRES), a series of complicated stem loop structures, which function in translation. Immediately following the 5'NCR is a large gene, encoding a single polyprotein, which is processed by virus encoded proteases. The processing of the large polyprotein yields mature virus proteins necessary for different stages in the replication cycle. Finally, the 3'NCR contains a short poly-(A) tail that helps stabilize the virus genome and participates in genome circularization\textsuperscript{20}. 
FIGURE 2: Circularization of the virus genome. Ternary complex is formed between PABP, PCBP, 3CD, and the 5’cloverleaf. The red arrow depicts the beginning of negative strand synthesis. For simplicity the cloverleaf is shown as a grey square.

Translation

Enteroviruses enter the cell using a variety of cell surface receptors including, intracellular adhesion molecule type1 (ICAM-1) and low density lipoprotein receptor (LDLR)\(^2\). Upon entry into the cell, the virus capsid dissociates, releasing the positive sense virus RNA. The virus RNA then undergoes cap-independent translation, using the IRES as the translation initiation site. The open reading frame is translated into a single polyprotein that undergoes a series of proteolytic cleavage events (Figure 3).

The overall organization of the polyprotein is as follows; the P1 region includes the virus capsid proteins (1A-1D) and the P2 and P3 regions contain the replication
proteins (2A-2C and 3A-3D) (Figure 3). In enteroviruses, the first cleavage event is performed by 2A protease, which separates the P2 region (nonstructural proteins) from the P1 region (structural proteins). The second cleavage event is performed by 3C protease, cleaving between 2C and 3A, separating the P2 and the P3 regions. The remaining cleavage events are also performed by 3C protease, or the fusion protein 3CD, producing intermediate and mature proteins, as summarized in Figure 3.

**FIGURE 3**: Polyprotein processing during enterovirus translation. The protease cleavage sites for 2A (red) and 3C (green) are shown as arrows.
P3 Region Proteins

The nonstructural proteins in the P3 region of the polyprotein are directly involved in virus replication. The cleavage of the P3 region results in two stable precursor proteins (3AB and 3CD), later cleaved into four mature proteins; 3A, 3B, 3C and 3D. 3AB is a nonspecific RNA binding protein and can interact with the presumed cloverleaf located at the 5'NCR. This interaction may be necessary to anchor the replication complex to the cytoplasmic membrane for efficient replication. 3B, also known as VPg (viral protein genome linked), attaches to the 5' and 3' end of the virus genome and effectively acts as a primer for RNA synthesis.

The precursor protein, 3CD, contains the sequences of both 3C protease and RNA-dependent RNA polymerase (3D). 3CD is responsible for the proteolytic cleavage of the P1 virus capsid proteins. Also, 3CD binds to the 5'cloverleaf, an essential interaction for efficient virus replication. Mammalian cells do not contain the machinery needed to replicate RNA from an RNA template. Therefore picornaviruses encode the required polymerase, RNA-dependent RNA polymerase (3D), for virus replication. In addition to functioning in the polyprotein cleavage events, 3C protease also cleaves host protein poly(C)-binding protein (PCBP), changing the RNA-binding properties of the protein, which promotes a switch from translation to replication. Briefly, PCBP is an RNA binding protein that recognizes and binds to poly(C) rich regions and contains three K-homologous domains; KH1, KH2, and KH3. To facilitate translation, full-length PCBP interacts with C-rich regions in stem IV of the IRES. However, when full-length PCBP is cleaved by 3CD to produce a truncated version of PCBP (contains only KH1 and KH2 domains), the truncated protein is unable to function
in translation. Instead, the truncated protein binds to the 5' cloverleaf, effectively promoting a switch from translation to virus replication.

**Replication**

In enteroviruses, the 5' cloverleaf is the protein binding site for initiation of virus replication. Mutational analysis confirmed stem loop D (SLD) and stem loop B (SLD), located in the 5' cloverleaf, are the specific binding sites for 3CD and PCBP\textsuperscript{18, 26, 27} and the above interactions (3CD-SLD and PCBP-SLB) are required for replication\textsuperscript{19}. During replication, the positive strand is used as a template to synthesize a negative strand intermediate. Since synthesis of the negative strand occurs at the 3' end of the virus genome, the involvement of the 5' end during replication was unclear. The poly(A)-binding protein (PABP) interacts with the 3'NCR by binding to the poly-(A) tail and also interacts with the 5'NCR binding to PCBP and 3CD\textsuperscript{19}. The interaction between the 3' and 5'NCR circularizes the virus genome, repositioning 3D polymerase to the 3'NCR to begin negative strand synthesis. 3D polymerase catalyzes the uridylylation of the VPg primer using the poly-(A) tail as a template\textsuperscript{23, 28, 29}. The VPg-pUpU is elongated to produce the full-length, positive sense RNA molecule\textsuperscript{23} which can be used for translation or be packaged into the virus capsid proteins and exit the cell via cell lysis.
Stem Loop D

There have been three published solution structures of stem loop D; from human rhinovirus-14, coxsackievirus B3, and a enterovirus consensus structure. Although the sequence of stem loop D in enteroviruses are not conserved (Figure 4), the three-dimensional structures are relatively conserved (Figure 5).

Human rhinovirus-14 (HRV-14) SLD contains twenty-seven nucleotides and is distinct from other enteroviruses by containing a triloop instead of the more common tetraloop. HRV-14 SLD is predicted to form two regions of Watson-Crick base pairs (stem I and stem II) separated by a pyrimidine mismatch region that should form a bulge (Figure 4). Unusual imino resonances, appearing outside of the Watson-Crick base pair chemical shift range, suggested the appearance of stable pyrimidine-pyrimidine base pairing for U5-U23 and U7-U25. No imino resonances were observed for U24, however residual dipolar couplings and NOEs suggested the appearance of stable base pairing between C6-U24. Similar base pairing patterns were also observed in the consensus enterovirus and the coxsackievirus SLD structure.

A-form RNA contains a narrow and deep major groove (~3.0 Å), restricting access for protein interactions to the minor groove side. However, all three of the SLD structures display widened major groove widths (~13 Å) (Figure 5) atypical of A-form RNA. The accessible major groove, when it does exist, has often been shown to play a role in protein recognition; permitting interaction with protein alpha helices. This feature in SLD could be significant as SLD is known to bind to the virus protein 3CD during enterovirus replication.
FIGURE 4: Secondary structure predictions of SLD from (A) human rhinovirus B14, (B) coxsackievirus B3 and (C) an enterovirus consensus structure. The structures are predicted to contain two stems separate by a pyrimidine mismatch region (U:yellow, C:orange). Structures were predicted using MC-Sym$^{35}$. 
FIGURE 5: NMR solution structures of SLD from (A) HRV-14\textsuperscript{30}, (B) CVB3\textsuperscript{31} and (C) enterovirus consensus\textsuperscript{32}. The mismatch region is colored as in Figure 4 and the loop nucleotides are colored as A (blue), C (orange), U (yellow), and G (green).

HRV-14 3C protease is 182 amino acids in length and contains two anti-parallel $\beta$ barrel each with six $\beta$ strands. The two $\beta$ barrels are connected by a flexible linker that binds the SLD triloop. Sequences of the 3C protease from enteroviruses are not highly conserved, however the secondary and tertiary structure is well conserved. HRV-14
3C protease contains a catalytic triad (Glu71, His40, and Cys146) positioned at the opposite side of the RNA binding region.

Interestingly, HRV-14 3C protease can bind to the more common tetraloop in other enteroviruses SLDs, however CVB3 and poliovirus 3C protease cannot bind to HRV-14 SLD. The tetraloop sequence found in many enteroviruses structurally is actually a biloop as the first and fourth bases pair. Therefore, it is possible that binding of enterovirus 3C protease to the triloop in HRV-14 SLD is abrogated by inability to fit the extra loop base in the triloop.

The interaction between HRV-14 SLD and HRV-14 3C protease has been characterized using NMR and SAXS\textsuperscript{36}. The protein:RNA complex shows HRV-14 SLD extended away from the protease, creating an oblong structure for the complex (Figure 6). Many proteins with multiple RNA binding domains contain flexible linker regions between the RNA binding domains to allow contact with multiple flexible regions of RNA. The 3C protease has a flexible linker region between two β barrels that is key to the interaction with HRV-14 SLD\textsuperscript{36}. There are several key residues of HRV-14 3C protease that make direct contact with SLD; L10 and R12-M16 from the N-terminal α helix and the subsequent strand, G27-R33 and C35 from β strands two and three, V49 and Q52 from the loop region connecting strands four and five, and finally residues L77, R79-R84, and I86 that are located in the linker region between the two β barrels\textsuperscript{36} (Figure 6). The nucleotides of HRV-14 involved in protein binding are U13-U17; the triloop and two additional nucleotides located on the 3’ strand (Figure 6). The non-Watson Crick base pairs (U5-U23, C6-U22 and U7-U21) are not involved in the interaction with 3C protease, however the presence of non-canonical base pairs in a
FIGURE 6: NMR structure of HRV-14 SLD and 3C$^{36}$. The amino acids (listed in the text) involved in RNA binding are colored red. The RNA base coloring is as Figure 5.
helix can introduce an atypical widened major groove. The widened major groove, shown in HRV-14 SLD\textsuperscript{30}, provides access for protein interactions as well as provides specific recognition of bases through the major groove.

The HRV-14 SLD-3C complex provides insight into the mechanisms involved in picornavirus replication. This complex is a crucial first step into developing a high-resolution structure of the 5’cloverleaf replication complex essential for picornavirus replication.

**NMR Spectroscopy**

RNA has the ability to adopt multiple conformations including stem-loops, pseudoknots, and bulges, giving rise to various biological functions\textsuperscript{37}. Currently, 37\% of the RNA structures deposited into the RCSB Protein Databank have been determined using nuclear magnetic resonance (NMR) spectroscopy\textsuperscript{38}. NMR provides atomic-level structural information about biological macromolecules and also allows researchers to study the dynamics of a molecule in solution. Both structural and dynamic information are essential for determining how a macromolecule behaves under physiological conditions. In order to characterize a structure in solution, several conformational constraints must be collected; hydrogen-hydrogen distances, \(^3\)J-coupling values, and residual dipolar couplings.

Hydrogen-hydrogen distances are derived from nuclear Overhauser effect spectroscopy (NOESY) spectra. Protons within 5-6 Å give rise to a nuclear Overhauser effect (NOE), identifying hydrogen atoms that are close in space. These distance constraints are crucial for defining the overall fold of RNA in solution. An NOE arises
when the disturbance of one spin population has been transmitted to nearby spins, causing a change in their populations. This change can be directly correlated to the chemical shift of the original disturbed spin, thus detecting protons close in space to one another. Typically, NOEs are divided into three groups; intranucleotide, sequential and long-range NOEs. Intranucleotide NOEs are between protons within the same nucleotide, sequential NOEs are between sequential nucleotides (H_i-H_{i+1}), and long-range NOEs are between protons typically separated by at least two nucleotide positions. Even though long-range NOEs account for a very small portion of observable NOEs, they are the most important for defining helical structure and unique tertiary interactions.

In NMR, the magnetic interactions of nuclei separated by three bonds, also known as $^3$J-coupling, provides information about bond geometry. J-coupling arises when the state of one spin affects the surrounding electrons which relay this affect to nearby spins typically within three bonds. J-coupling values in RNA are used to define backbone dihedral angles and sugar pucker conformations. A sugar pucker refers to the conformation of the ribose ring in RNA; C2’ or C3’-endo. Sugar pucker conformations can be estimated using a $^1$H,$^1$H-TOCSY experiment. TOCSY, total correlation spectroscopy, correlates protons within a spin system (a group of protons separated by three bonds) via $^1$H-$^1$H $^3$J-coupling. Weak $J_{H1'-H2'}$ presented in a TOCSY spectrum correspond to C3’endo conformation whereas strong $J_{H1'-H2'}$ corresponds to C2’endo conformation. The estimation of sugar puckers helps define whether the RNA structure resembles typical or atypical A-form RNA. Typically, A-form RNA adopts a C3’-endo sugar pucker conformation which creates a short phosphate-phosphate distance (~5.9
A) compared to B-form DNA\(^40\). The shorter phosphate-phosphate distance results in a very narrow and deep major groove width typical of A-form RNA. The exact opposite is true for C2’endo sugar pucker conformation. This conformation creates a longer phosphate-phosphate distance (~7.0 Å) atypical of A-form RNA.

Residual dipolar couplings (RDC) measures the angle between the \(^1\text{H}-^{15}\text{N}\) or \(^1\text{H}-^{13}\text{C}\) bond and the external magnetic field to define the relative orientations of parts of a biomolecule in solution\(^41\). RDC values can be used to refine solution structures previously determined using NOEs and J-coupling values. Both NOEs and J-coupling values provide important information on local conformation. NOEs provide distance information between protons, and J-coupling values define sugar pucker conformation and dihedral angles. However, neither NOE nor J-coupling values are able to distinguish the orientation of a structure or parts of the structure in solution; for example, whether the molecule is straight or slightly curved. In solution, the rapid tumbling of a molecule averages the net effect of the dipolar coupling to zero. Therefore, the molecule must be partially aligned relative to the magnetic field in order to measure RDCs. RNA is measured in isotropic media to obtain the J-coupling values \(^1\text{J}_{\text{H-X}}\) values (X referring to \(^{15}\text{N}\) or \(^{13}\text{C}\)) and then also measured in alignment media to obtain RDC + \(^1\text{J}_{\text{H-X}}\) values. The difference between the aligned and isotropic values results in RDCs for the RNA molecule.

The refinement of HRV-14 SLD structure using RDC values proved to be advantageous\(^30\). Inclusion of RDCs gave more precise (RMSD of 0.59 Å) and compact structures (axial length 47 Å) compared to those structures determined without RDCs (RMSD 1.00 Å and axial length of 52 Å). RDCs were critical for maintaining hydrogen
bond pairing for nucleotide U22 in the pyrimidine mismatch region (Figure 4A)\textsuperscript{30}. This unique base pair region contributes to the widened major groove width of SLD, a feature potentially important for protein interactions.

NMR has become a principle technique for determining structures of biological macromolecules. However, one of the major disadvantages of NMR is its size limitation. The larger the molecule, the faster the relaxation rates and the slower the tumbling, causing decreased sensitivity in experiments. To overcome this, researchers have developed a method to divide large RNA structures into individual duplexes\textsuperscript{42-44}. The structure and orientation of the individual duplexes are determined using experimental NMR constraints; NOEs, RDCs, J-coupling values. The global RNA structure is then determined using RDCs, SAXS and the experimental information from the individual duplexes. In this approach, it is assumed (and later verified) that chemical shifts of the individual duplexes are conserved when incorporated into the global structure, therefore the previously determined conformation of the individual duplexes remains intact.

**Small-angle X-ray Scattering**

Small-angle X-ray scattering (SAXS) provides information regarding the global shape of a biological macromolecule or complex in solution. SAXS is a medium to low resolution technique providing precise information about size and shape, requiring low concentrations of sample, no size limitations, and can be performed in biologically relevant solvents with varying ion concentrations (for recent review see\textsuperscript{45}). It is a simple technique involving an X-ray beam of a single wavelength (<3º) which illuminates the
sample containing the biomolecule. The scattered radiation is recorded on a detector. The scattering pattern is described by the intensity ($I$) as a function of the momentum transfer ($q$). Structural characteristics such as, radius of gyration, maximum particle diameter and molecular weight can be determined using SAXS. With no prior knowledge of structural information, SAXS provides a basic understanding into the behavior of a biomolecule in solution; for example, flexibility, compactness, and oligomerization. The plot of intensity vs $q$ does not directly provide information on the structure: the plot must be Fourier transformed to obtain a pairwise-distance distribution function ($P(r)$)\textsuperscript{46}. $P(r)$ is used to describe the distances between a pair of atoms within a structure and is useful in investigating conformational changes within a molecule e.g. due to changes in ionic environment.

SAXS can be used as a method for structure validation. For example, the X-ray crystal structure of release factor 1 (RF1), a protein that directs translation termination in \textit{E.coli}, presented the protein in a compact conformation spanning a width of 23 Å\textsuperscript{47}. However, this compact conformation would not allow the protein to span the appropriate distance needed to shutdown translation\textsuperscript{47}. Therefore, the structure was investigated with SAXS, and an equilibrium was detected between an open and closed conformation in solution with the equilibrium favoring the open conformation. The more compact conformation observed in the crystal structure is more than likely an artifact of crystallization\textsuperscript{47}. As such, SAXS has the ability to provide new structural and conformational information in solution as well as being used for structure validation for previously determined crystal structures.
NMR and SAXS are complementary techniques used in combination to determine a high-resolution structure of RNA. SAXS provides the overall shape of the structure while NMR provides local structural information based on specific atom and nucleotide constraints, together providing comprehensive information about biological macromolecules in solution.

Research Aims

The aim of this dissertation project was to determine the solution structures of HRV-14 SLB and the 5’cloverleaf (5’CL) using NMR spectroscopy and SAXS. This work provides the first high resolution structures of SLB and 5’CL in picornaviruses. These structures are necessary to understand protein:RNA interactions involved in picornavirus replication and can potentially provide further insight into the development of therapeutic agents for inhibiting replication.
CHAPTER II

STRUCTURE OF RNA STEM LOOP B
FROM THE PICORNAVIRUS REPLICATION PLATFORM

PREFACE

The content of this chapter was published in Biochemistry in May 2017.


INTRODUCTION

The aim of the research described in the following chapter was to produce milligram quantities, and high quality, solution NMR samples of HRV-14 SLB and determine the first high resolution structure of HRV-14 SLB. As mentioned in Chapter 1, the rhinovirus genomic RNA consists of three distinct regions: the 5´ and 3´ untranslated regions (UTRs) flanking a gene encoding a single ~250 kilodalton polyprotein. After translation, the polyprotein is cleaved into several structural and nonstructural proteins by virus proteases that are part of the polyprotein. Most of these cleavages are performed by the 3C protease, which also performs a second task: some 3C protease remains covalently joined to the virus-encoded RNA-dependent RNA polymerase (3D) to form the 3CD fusion protein. The 3CD protein directly interacts with the 5´UTR as an essential step in replication\textsuperscript{2, 29, 34, 48, 49}. The 5´UTR itself contains two
regions: the large (typically 400-500 bases) internal ribosome entry site (IRES) that directs translation, and a smaller (<85 bases) region, predicted to form a cloverleaf (5´CL, see Figure 7) in most picornaviruses, that serves as a platform for RNA replication\textsuperscript{1, 50, 51}. Stem loop D (SLD) from the 5´CL attracts the 3CD fusion protein\textsuperscript{34}, effectively delivering the 3D RNA polymerase to the required site to begin replication. The C-rich region of stem loop B (SLB) attracts the host polyC-binding protein (PCBP) that helps to circularize the genomic RNA\textsuperscript{52, 53}.

The above interactions of 5´CL with 3CD and with PCBP are not independent. PCBP facilitates both translation and replication, by binding C-rich loops in both the IRES and the 5´CL\textsuperscript{33, 34, 54-56}. However, PCBP can be cleaved by 3C, and the product subsequently functions only in replication, effecting a switch from translation to replication\textsuperscript{57, 58}. Furthermore, interaction of PCBP with the 5´CL SLB facilitates binding of 3CD to SLD. A better understanding of the structural basis for these interactions and their interplay could lead to ways to inhibit virus replication.

In this study, the NMR solution structure of a 24-nucleotide RNA hairpin representing the 5´CL SLB (Figure 7) HRV-14 was determined. The structure contains five stable Watson-Crick base pairs and one wobble base pair with an 8-nucleotide C-rich hairpin loop. The loop is highly flexible, with the C-rich region exposed, while the stem region presents an accessible major groove that may be important for protein interactions. This study represents an important step in the process of building a high resolution structural understanding of picornavirus replication.
FIGURE 7: Secondary structure prediction of HRV-14 5’CL. SLB, the subject of this investigation, is circled. SLB is predicted to contain six Watson-Crick base pairs with one wobble base pair shown in color and an eight nucleotide hairpin loop. Nucleotides implicated in PCBP-binding are located in the loop and are shown in color. The structure was predicted using MC-Sym\textsuperscript{35}.

MATERIALS AND METHODS

Expression and Purification of Bacteriophage T7 RNA Polymerase

The pUC19 plasmid of His\textsubscript{6}-tagged T7 RNA polymerase was a gift from the Butcher lab and was provided on Whatman filter paper. The plasmid was cut from the filter paper with sterile scissors and placed into a 1.5 mL Eppendorf tube. 100 mL of double distilled water was added to the Eppendorf tube and vortexed gently for one
minute. 2 µL of the His6-tagged T7 RNA polymerase plasmid was added to 50 µL of BL21 (DE3) E.coli electrocompetent cells and placed into a cold 0.1 cm Gene Pulser electroporation cuvette (Biorad). The cells were electroporated with a 1.8 kV pulse and then pipetted in 350 µL of Luria Broth (LB). The cells were transferred to a clean 1.5 mL Eppendorf tube and incubated for 45 minutes at 37°C. 100 µL of the cell culture was plated onto LB (containing 100 µg·µL⁻¹ of ampicillin) agar plate and incubated overnight at 37°C.

One colony from the BL21 (DE3) E.coli transformed with His6-tagged T7 RNA polymerase from a plate of LB agar (containing 100 µg·µL⁻¹) was inoculated with 5 mL of LB containing 100 µg·µL⁻¹ ampicillin and grown overnight at 37°C. Two 800 mL aliquots of the overnight culture were pipetted into two separate Eppendorf tubes and 200 µL of sterile 75% glycerol was added to each. The glycerol stock tubes were vortexed thoroughly to mix and stored at -80°C.

10 mL of LB was added to three sterile 50 mL falcon conical centrifuge tubes (total of 30 mL of LB). A sterile loop was used to scrap cells from the glycerol stock of BL21 (DE3) E.coli cells transformed with His6-tagged T7 RNA polymerase and added to each falcon conical centrifuge tube with 100 µg·µL⁻¹ of ampicillin. The tubes were incubated overnight at 37°C. Six 2.5 L flasks containing 500 mL of LB with 100 µg·µL⁻¹ of ampicillin was inoculated with the 30 mL overnight culture (5 mL of overnight culture per flask) and incubated at 37°C to an OD₆₀₀ of 0.7. The cells were then induced by the addition of 500 µL of 1 M IPTG and incubated for 3 hours at 37°C. The cells were harvested by centrifugation at 5,000 rpm for 25 minutes at 4°C. At this point, the harvested cells can be stored at -80°C until ready for purification.
The harvested cells were dissolved on ice with 20 mL of buffer A (50 mM NaPi (pH 8.0), 300 mM NaCl, 5% glycerol, 20 mM imidazole, and 0.5 mM DTT) and 2 mL of 1X protease inhibitor solution (Thermo Scientific Pierce Protease inhibitor tablets). The cells were added to a sterile 50 mL falcon conical centrifuge tube with 30 mg of lysozyme. The solution was sonicated at 40% amplitude for 10 seconds on, 50 seconds off for a total of 2 minutes. The sonicated protocol was repeated twice to ensure the lysis of cells. The lysate was spun down at 16,000 rpm for 30 minutes at 4°C and the supernatant was filtered first through a 0.45 um filter and then through a 0.22 um filter. The supernatant containing His6-tagged T7 RNA polymerase was purified using a 5 mL precharged nickel immobilized metal ion affinity chromatography column (HisTrap HP; GE Healthcare). The column was operated on an AKTA liquid chromatography system. The column was equilibrated with buffer A at 2 mL·min⁻¹ for 6 column volumns. The supernatant was applied to a 50 mL superloop prefilled with buffer A and injected into the column at 1 mL·min⁻¹. The column flowthrough was collected in a 50 mL falcon conical centrifuge tube and stored at 4°C. The superloop was flushed with 5 mL of buffer A to ensure all the supernatant was injected into the column. The column was washed with buffer A for 10 column volumes at 1 mL·min⁻¹ and the column flowthrough was collected in a 50 mL falcon conical centrifuge tube and stored at 4°C. The His6-tagged T7 RNA polymerase was eluted from the column using a 3 step gradient of 5 column volumes each (14.3%, 37% and 100%) of buffer B (50 mM NaPi (pH 8.0), 300 mM NaCl, 5% glycerol, 500 mM imidazole, and 0.5 mM DTT). 2 mL column fractions were collected and stored at 4°C. Finally, the column was equilibrated with 5 column volumes of buffer A at 2 mL·min⁻¹.
The column flowthrough from the injection of the sample, wash, and eluted fractions were run on a 10% SDS PAGE gel. The fractions containing the His\textsubscript{6}-tagged T7 RNA polymerase were collected and added to regenerated cellulose dialysis tubing (6,000-8,000 molecular weight cutoff; Fisher Scientific). The dialysis tubing was placed in dialysis buffer (20 mM NaPi (pH 7.7), 100 mM NaCl, 1 mM EDTA, 50% glycerol, and 1 mM DTT) and dialyzed overnight at 4\(^\circ\)C. The sample was dialyzed for an additional 4 hours in fresh dialysis buffer and the sample was aliquoted and stored at -20\(^\circ\)C to be used for in-vitro transcription.

**In-vitro Transcription of HRV-14 SLB**

In-vitro transcription using T7 RNA polymerase requires a double stranded DNA template with a T7 polymerase promoter site. The following forward sequence, \(5\text{'TTCTAATAAGACTCAGTATAAGCGGATGGGTATCCCACCATTCA3'}\) and its corresponding reverse complement sequence were purchased from Integrated DNA Technologies and contained a 17 nucleotide T7 polymerase promoter site (shown in bold) followed by 24 nucleotides which correspond to HRV-14 SLB sequence. The forward and reverse oligonucleotides were resuspended in Ultra-Pure, DNase and RNase free water (Invitrogen) to create a 1000 \(\mu\text{M}\) stock concentration. The oligonucleotides were diluted with sterile annealing buffer (10 mM Tris-HCl (pH 8), 1 mM EDTA, and 100 mM NaCl) to a 100 \(\mu\text{M}\) stock. 100 \(\mu\text{L}\) of the forward sequence was added to a clean 1.5 mL Eppendorf tube. 100 \(\mu\text{L}\) of the reverse sequence was added to the forward sequence and diluted with 800 \(\mu\text{L}\) of sterile annealing buffer to make a 10 \(\mu\text{M}\) DNA concentration. The DNA solution was mixed thoroughly and incubated on a
heat block at 95°C for 5 minutes and then slowly cooled for 1 hour. The DNA was spun down and stored at -20°C.

Several 10 µL small scale transcription reactions were performed in order to optimize the components in the transcription reaction to give the greatest yield of RNA product; transcription buffer, DNA template, ribonucleotide triphosphates, T7 polymerase, and pyrophosphatase. It is well known that T7 RNA polymerase prefers to initiate transcription with a guanosine as the first base in the RNA transcript. Therefore, in order to increase overall RNA yield in the transcription reaction, it was decided to extend the sequence of HRV-14 SLB to contain a G at the 5´end of the sequence. To create a symmetrical stem loop, the sequence was extended to contain an A at the 3´end. To test whether RNA yield could be further increased, an additional G was added to the 5´end and several small scale transcription reactions were performed to determine whether the 1G or 2G template yielded a higher concentration of RNA. It was determined the yield of the 1G and 2G template were identical, therefore transcription reactions were performed with the 1G template and not the 2G.

A natural abundance RNA sample, 5´GCGGAUGGGUAUCCCACCAUUCGA3´, corresponding to HRV-14 SLB, was produced by in vitro transcription using T7 RNA polymerase (expressed and purified in house) with RNAPoly reaction buffer [New England Biolabs; 40 mM Tris-HCl, 6 mM MgCl₂, 10 mM DTT, and 2 mM spermidine (pH 7.9)], each rNTP at 5 mM (Sigma), 1 µM double-stranded DNA (as a template), and 0.001 unit·mL⁻¹ pyrophosphatase (Fisher). A 20 mL transcription reaction mixture was incubated at 37°C for 3 h; then 2.5 volumes of 100% ethanol and 0.1 volume of 3 M
sodium acetate were added to the reaction mixture, and the mixture was incubated overnight at -20°C. The solution was spun down at 9,200xg and 4°C.

**Purification of HRV-14 SLB**

The resulting pellet was washed in 70% ethanol and spun down for an additional 10 min at 9,200xg. The pellet was dissolved in 8 M urea loading buffer [8 M urea, 25 mM EDTA, 500 mM Tris-HCl, 4.6 mM xylene cyanol, and 3.7 mM bromophenol blue (pH 7.9)] and heated for 3 min at 98 °C prior to gel purification. The RNA was resolved on a 15% denaturing polyacrylamide gel containing 8 M urea and visualized via ultraviolet (UV) shadowing at 254 nm, and the corresponding band was excised from the gel. The RNA was recovered from the gel pieces using an Elutrap device (Whatman) in 1X Tris-borate-EDTA (TBE) buffer (pH 7.9). The concentration of the eluted RNA was determined using a Biodrop UV-visible spectrophotometer. The RNA was concentrated and buffer exchanged using a Vivaspin spin concentrator (3 kDa cutoff) to a concentration of 0.5 mM in NMR buffer [10 mM sodium phosphate and 0.1 mM EDTA (pH 7.0)]. A uniformly 13C- and 15N-labeled RNA sample of SLB was similarly produced using each 13C- and 15N-labeled rNTP at 2.5 mM (Cambridge Isotope Laboratories) and 0.00075 unit/mL pyrophosphatase.

**NMR Spectroscopy of HRV-14 SLB**

NMR spectra of the unlabeled SLB sample were a one-dimensional (1D) 1H spectrum, 1H,1H-NOESY in H2O and D2O, and 1H,1H-TOCSY spectra. The spectra obtained with the labeled sample were 1H,13C and 1H,15N-HSQC, 1H,13C-ARTSY, 1H,15N
and \(^1\text{H}, ^1\text{H}, ^{13}\text{C}\)-edited three-dimensional NOESY-HSQC spectra. Two copies of each spectrum measuring \(^{13}\text{C}\) shifts were obtained, with the \(^{13}\text{C}\) transmitter centered on the C6/C8/C2 region or the C1′/C5 region. A 0.5 mM unlabeled RNA sample in 10 mM sodium phosphate and 0.1 mM EDTA (pH 7.0) containing a 90% H\(_2\)O/10% D\(_2\)O mixture was used for the analysis of exchangeable protons. For the analysis of nonexchangeable protons, the 0.5 mM sample was vacuum concentrated using a Thermo Scientific SpeedVac concentrator and resuspended in 99.9% D\(_2\)O. NMR experiments were performed at 750, 800, and 900 MHz (Varian) at the National Magnetic Resonance Facility at Madison (NMRFAM) equipped with a cryoprobe. For the assignment of nonexchangeable base proton resonances, \(^1\text{H}, ^1\text{H}\)-NOESY spectra in D\(_2\)O with mixing times of 125 and 250 ms were recorded at 26.5°C with 256 increments, 4096 data points, 32 scans per fid, and a 3 second recycle time with presaturation for suppression of the water signal. A \(^1\text{H}, ^1\text{H}\)-NOESY spectrum in H\(_2\)O with a mixing time of 250 ms was recorded at 5°C using a 1−1 echo NOESY pulse sequence with 256 increments, 4096 data points, and 96 scans for the assignment of exchangeable imino proton resonances. Two \(^1\text{H}, ^1\text{H}\)-TOCSY spectra with mixing times of 50 and 200 ms with a 5 kHz spin-lock field were recorded at 26.5°C to obtain sugar pucker conformations. Two \(^1\text{H}, ^{13}\text{C}\) HSQC spectra were recorded for C6/C8/C2 and C1′/C5 with spectral widths of 13,020 and 14,534 Hz for \(^1\text{H}\) and 4,000 and 5,200 Hz for \(^{13}\text{C}\), respectively. Residual dipolar coupling (RDC) experiments were performed on a 750 MHz Bruker spectrometer. The samples contained 18 mg/mL Pf1 filamentous bacteriophage, and one-bond \(^1\text{D}_{\text{NH}}\) and \(^1\text{D}_{\text{CH}}\) couplings were measured from \(^1\text{H}, ^{13}\text{C}\) and \(^1\text{H}, ^{15}\text{N}\)-ARTSY
spectra. Data for all spectra were processed using NMRPipe and visualized and analyzed using NMRViewJ.

**Structural Constraints**

NOE distance constraints were calculated using $^1$H,$^1$H-NOESY and three dimensional $^{13}$C-edited NOESY-HSQC spectra. The interproton distances calculated for samples in D$_2$O were calibrated according to the intranucleotide H5–H6 intensities (distance of 2.4 Å) using the equation $I = kr^6$, where $r$ is the distance between two protons. The NOE distances corresponding to assignments in the $^1$H,$^1$H-NOESY spectrum in H$_2$O were calibrated using the intranucleotide H5-H3 distance of uracil (4.1 Å). A 33% upper limit was added to all NOE distances up to a maximal upper limit of 6 Å. Because of the high degree of flexibility associated with the loop and terminal regions, NOE distances corresponding to loop nucleotides (G9-A16) and terminal nucleotides (G1, C2, G23, and A24) were treated more conservatively with a further 1 Å added to the upper limit. The lower limit of all NOE distances was set to 1.8 Å. Sugar puckers were determined using $^1$H,$^1$H-TOCSY spectra. Nucleotides for which strong H1′-H2′ cross peaks were observed were constrained to the C2′-endo sugar pucker conformation with $\nu_0$-$\nu_4$ set to $-4.2 \pm 15.0^\circ$, $24.9 \pm 15.0^\circ$, $-34.9 \pm 15.0^\circ$, $33.0 \pm 15.0^\circ$, and $-18.3 \pm 15.0^\circ$, respectively. Nucleotides without strong H1′–H2′ cross peaks were constrained to C3′-endo sugar conformations with $\nu_0$-$\nu_4$ set to $5.8 \pm 15.0^\circ$, $-26.2 \pm 15.0^\circ$, $36.5 \pm 15.0^\circ$, $-33.0 \pm 15.0^\circ$, and $16.8 \pm 15.0^\circ$, respectively. Nucleotides with intermediate H1′-H2′ cross peaks were left unconstrained. On the basis of preliminary structure calculations showing the stem region is consistent with A-form geometry,
backbone dihedral angle constraints corresponding to the A-form were applied to nucleotides G3-G8 and C17–U22 with α, β, γ, δ, ε, and ζ set to 292 ± 15.0°, 178 ± 15.0°, 54 ± 15.0°, 82 ± 15.0°, 207 ± 15.0°, and 289 ± 15.0°, respectively. CURVES analysis of the initial calculated structures revealed that G4 of the wobble base pair (G4-U21) exhibited atypical α and γ backbone dihedral angles that were correlated with dihedral angle violations in the XPLOR calculations. Typically, α and γ form a gauche-/gauche+ conformation in A-form geometry; however, the α and γ backbone dihedral angles of G4 formed a trans/trans conformation. This conformation is typical of wobble base pairs embedded in the helix of RNA. Therefore, α and γ backbone dihedral angles of G4 were left unconstrained in the final structure calculations, and U21 backbone dihedral angles were given an additional tolerance of ± 30.0° to alleviate any possible strain experienced due to unusual dihedral angles of G4.

Because of the absence of imino resonances for nucleotide G23 in the 1H,1H-NOESY spectra, no hydrogen bond constraints for C2-G23 were included in the calculation. The properties of C2 and G23 resemble those observed for the loop nucleotides: C2′-endo sugar conformation assigned to C2 and partial C2′-endo character for G23 with no stable hydrogen bonding. They display a degree of flexibility similar to those of nucleotides G1, G9-A16, and A24 (see Results). Therefore, for NOE constraints involving C2 and G23, an additional 1 Å was added to the upper distance limit, and no backbone dihedral angle constraints were included in the structural calculation for C2 and G23.

Residual dipolar couplings were measured from partial alignment of the labeled SLB sample via 18 mg/mL Pf1 filamentous bacteriophage. One-bond 1D_{NH} and 1D_{CH}
couplings were obtained from two dimensional ARTSY spectra. RDC errors were estimated via comparison between isotropic peak heights, aligned sample peak heights, and the average spectral noise. RDC values for nucleotides G3-G7 and C17-U22 were entered as constraints in the simulated annealing protocol. The RDCs for the loop and terminal regions display values significantly smaller than those of the stem, most likely as a result of flexibility associated with the loop and terminal nucleotides. Therefore, loop and terminal RDC values were not used as constraints, to prevent incorrect alignment of the associated bonds in these disordered regions.

Structural Calculation and Refinement

A total of 400 structures were calculated using a simulated annealing protocol in XPLOR-NIH. The initial annealing temperature was set to 3000 K with a final temperature of 25 K. The force constants for NOE and hydrogen bond constraints were ramped from 2 to 50 kcal mol$^{-1}$ Å$^{-2}$ and from 0.02 to 5 kcal mol$^{-1}$rad$^{-2}$ for RDC constraints. The structural calculations contained 257 NOE distance constraints, 115 sugar puckers, 73 backbone dihedral angles, 30 hydrogen bond constraints, 12 planarity constraints, and 37 $^1$H-$^{15}$N and $^1$H-$^{13}$C residual dipolar couplings. The 10 lowest-energy structures containing no NOE violations of >0.5 Å and no dihedral angle violations of >5° were selected to represent the structural ensemble. The structures were visualized with PyMol$^{59}$ and analyzed using CURVES$^{60}$. 
RESULTS

Assignment of Imino Protons

The number of observable imino resonances in a 1D \(^1\)H NMR spectrum of RNA is correlated to the number of base pairs. Resonances of imino protons in standard Watson-Crick base pairs typically appear in the spectral range of \(\sim 13-15\) ppm (uracil) and \(\sim 11-13.5\) ppm (guanine). The absence of an imino resonance suggests that the nucleotide is not base-paired, allowing the imino proton to rapidly exchange with the solvent. Imino resonances outside of the spectral ranges mentioned above indicate the presence of non-Watson-Crick base pairs, such as G-U wobble pairs (10–12 ppm). Cross peaks in the NOESY spectrum should be seen between imino protons from adjacent base pairs, providing a means to walk through the helix and obtain sequential imino assignments. HRV-14 SLB contains 12 imino protons, seven of which were observed in the 1D \(^1\)H spectrum (top of Figure 8) and in the \(^1\)H,\(^1\)H-NOESY spectrum in H\(_2\)O (Figure 8). The strongest five imino resonances were sequentially assigned to G3, G4, U6, U20, and U21 on the basis of the imino–imino NOE cross peaks labeled in Figure 8. The two weaker resonances were assigned to G7 and G8. These two assignments were made via cross-strand NOEs to C18 and C17 amino protons, respectively. As a result, the moderately weak imino peak (\(~11.7\) ppm) is assigned to G7, and the weakest observed imino is assigned to G8, consistent with the fact that the G8-C17 base pair is the closing base pair. In fact, the G8 imino resonance produced no observable diagonal peak. We considered the possibility that the broad resonance
assigned to the G8 imino proton was instead the G9 imino proton, and that the G8 base is not base-paired to an extent allowing detection of its imino proton.

**FIGURE 8**: Assigned imino proton portion of the 250 ms $^1$H,$^1$H-NOESY spectrum at 278 K of HRV-14 SLB in H$_2$O with the 1D $^1$H NMR spectrum shown above. A diagram of the helix is shown on the right.

If G8 is extruded from the helix, this could allow the formation of a G9-C17 base pair and an additional base pair could then potentially form between U10 and A16 to extend
the helix. However, the base-sugar NOE walk (Figure 9) is continuous through the sequence as currently assigned. Such continuity would not be expected if the G8 base

![Diagram](image)

**FIGURE 9.** Base to H1’/H5 region of a 250 ms 1H,1H-2D NOESY spectrum of SLB in D2O. Peaks representing intranucleotide H1’-H6/H8 NOEs are labeled with nucleotide number. The black lines connect the labeled peaks through sequential H1’-H6/H8 NOE cross peaks. The H2 resonances not involved in the walk are underlined.

were extruded. Furthermore, the U10 and A16 peak characteristics (see Figure 11) are consistent with disorder, not with stable base pair formation. Also, the U10 imino peak is
not observed. For these reasons, we maintained the assignment of the broadest observed imino proton to G8. Therefore, the structural calculations included hydrogen bond constraints for the following base pairs: G3-C22, G4-U21, A5-U20, G7-C18, and G8-C17. To account for broadening of the G8 imino proton, the hydrogen bond parameters for this pair were weakened as described by Baba et al. The absence of the imino resonances for G1, G9, U10, U12, and G23 indicates that these nucleotides are not involved in stable hydrogen bonding.

**Assignment of Sugar and Nonexchangeable Base Protons**

Intraneucleotide and sequential H1′–H6/H8 NOEs provide a means to “walk” down a polynucleotide chain to obtain sequential chemical shift assignments. The walk region of the SLB 250 ms mixing time 1H-1H NOESY spectrum in D2O (Figure 9) was used to assign the H1′, H2, H5, H6, and H8 resonances of nucleotides C2–A24. The corresponding 13C resonances were assigned using 1H-13C HSQC spectra. NOESY peak overlap in the U12–C15 region was resolved, and all assignments were confirmed, using 13C-edited three dimensional NOESY-HSQC spectra. The H1′, H8, C1′, and C8 resonances of G1 were assigned via NOEs to A24, because no G1–C2 NOEs were detected. In all, every nonexchangeable base resonance was assigned along with all sugar H1′ and H2′ resonances. Of the remaining sugar resonances, 19 H3′ and 13 H4′ resonances were assigned.
**Sugar Conformations**

Standard A-form RNA contains predominantly C3′-endo sugar conformations that help to create a short sequential phosphate-phosphate distance (~5.8 Å). As a result, A-form RNA exhibits a helical structure more compact than that of the B-form helix that is typical of DNA. The SLB 1H,1H-TOCSY spectra revealed strong intra-nucleotide H1′-H2′ cross peaks for nucleotides G1, C2, G9-A16 and A24, consistent with the C2′-endo sugar conformation at these sites. The absence of intranucleotide H1′-H2′ cross peaks for nucleotides G3-G8 and C17-C22 is consistent with the C3′-endo sugar conformation expected in an A-form helix. Nucleotide G23 exhibits weak but detectable H1′-H2′ cross peaks typical of an intermediate sugar pucker or equilibrium between C3′-endo and C2′-endo conformations. Therefore, the sugars of nucleotides G3–G8 and C17–C22 were constrained to C3′-endo. Nucleotides G1, C2, G9–A16, and A24 were constrained to C2′-endo. The sugars of G23 were left unconstrained.

**Structural Calculations**

Structures were calculated using a total of 524 constraints (Table 2). The calculation included 257 NOE-derived distances along with 30 hydrogen bond-derived constraints. The latter represent 15 hydrogen bonds in the 6 bp identified above via imino proton analysis. Twelve planarity constraints, 37 1H-15N and 1H-13C residual dipolar coupling- based constraints, and 188 dihedral angle constraints were applied. The majority of the dihedral constraints specify sugar pUCKERS, but backbone dihedral angles of Watson-Crick base- paired nucleotides determined to have C3′-endo sugar puckers (G3, A5-G8, C17-U20, and C22), along with U21 from the wobble pair, were
constrained to typical A-form helix ranges.23 Backbone dihedral angle constraints were not added for G23 because of the observed intermediate H1′-H2′ J-couplings. Initially, nucleotide G4 from the wobble base pair was constrained to A-form backbone dihedral angles, but NOE violations and dihedral angle violations were observed. Therefore, the α and γ G4 backbone dihedral angles were left unconstrained in the final structure calculations. The structures were calculated using a standard simulated annealing protocol in XPLOR-NIH.62 A total of 400 structures were calculated. The 10 lowest-energy structures containing no NOE violations of >0.5 Å, no RDC violations of >3 Hz, and no dihedral angle violations of >5° were selected to represent the structural ensemble (Figure 10).
<table>
<thead>
<tr>
<th>TABLE 2: Summary of Structural Constraints and Refinement Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NOE constraints</strong></td>
</tr>
<tr>
<td>Intranucleotide NOE</td>
</tr>
<tr>
<td>Internucleotide NOE</td>
</tr>
<tr>
<td>Long-range NOE</td>
</tr>
<tr>
<td><strong>Dihedral angle constraints</strong></td>
</tr>
<tr>
<td>Sugar pucker</td>
</tr>
<tr>
<td>Backbone</td>
</tr>
<tr>
<td><strong>Hydrogen bonds</strong></td>
</tr>
<tr>
<td><strong>Planarity</strong></td>
</tr>
<tr>
<td><strong>Residual dipolar couplings</strong></td>
</tr>
<tr>
<td><strong>Number of violations</strong></td>
</tr>
<tr>
<td>NOE violations (&lt;0.5 Å)</td>
</tr>
<tr>
<td>Dihedral violations (&lt;5°)</td>
</tr>
<tr>
<td>RDC violations (&lt;3 Hz)</td>
</tr>
<tr>
<td><strong>Mean RMSD from ideal geometry</strong></td>
</tr>
<tr>
<td>Bond length (Å)</td>
</tr>
<tr>
<td>Bond Angle (°)</td>
</tr>
<tr>
<td>Improper bond angle (°)</td>
</tr>
<tr>
<td><strong>Pairwise RMSD</strong></td>
</tr>
<tr>
<td>Stem atoms (Å)</td>
</tr>
<tr>
<td>All atoms (Å)</td>
</tr>
</tbody>
</table>
FIGURE 10: Superposition of the final ten structures of HRV-14 SLB determined in this study. The loop and terminal regions are shown as grey ribbons. The bases of hydrogen-bonded base-pairs are shown in the following colors: cytosine (orange), uracil (yellow), guanine (green), and adenine (blue).
Solution Structure of SLB

HRV-14 SLB forms a six base pair helical stem (including G3-G8 and C17-C22) that is largely consistent with A-form geometry, and an 8-nucleotide hairpin loop (G9-A16). The helical region is well-defined, with a pairwise RMSD of 0.415 Å, and contains a wide and shallow minor groove (~10.5 Å wide) typical of A-form RNA. However, its major groove (~8.7 Å wide) shares some characteristics with the major groove of B-form DNA, as will be discussed further below. The overall SLB RMSD of 2.249 Å reflects the high degree of conformational variability in the regions outside of the helix.

Each of the loop and terminal nucleotides adopts a C2´-endo sugar conformation, except G23, which is C2´-endo in eight structures and C3´-endo in two structures. This result accurately reflects the observed intermediate G23 H1´-H2´ J-coupling and the observed strong couplings for the other loop and terminal nucleotides. The helix nucleotides adopt exclusively C3´-endo sugar pucker conformations. This is consistent with the weak H1´-H2´- couplings detected for each helix nucleotide, G3-G8 and C17-C22.

Evidence of Disorder in HRV-14 SLB Loop and Termini

Flexible regions within an RNA structure can be identified via analysis of chemical shift, resonance intensity, and residual dipolar couplings. Chemical shifts from highly flexible regions tend toward the chemical shifts of NMPs. The difference between the Carbon (13C) chemical shifts of SLB and NMP values are shown in Figure 11A. As expected, the loop region displays chemical shifts more similar to those of the NMPs. The four terminal residues (G1, C2, G23, and A24) were not constrained to A-form geometry.
due to the absence of observable imino peaks and the presence of strong H1´-H2´ couplings. They do however have chemical shifts more closely resembling the helical region. This suggests that the terminal residues may have some helical or other partially ordered character.

Flexible regions are also expected to produce narrow, intense resonances. Figure 11B displays intensities from a ¹H,¹³C-HSQC spectrum. Resonance intensities for the loop region and terminal region are noticeably higher than those of the helix. There is a gradual decrease in intensity proceeding down the helix toward the loop, suggesting increase rigidity near the loop. Within the loop, U11-A16 are the most intense, suggesting that this region is the most dynamic on the ns-ps time scale. However, the C8 resonance of G9 was not observed. This suggests ms-µs time scale dynamics causing exchange broadening at the 5´-end of the loop.

Disordered residues typically have lower RDC values compared with ordered regions. In Figure 11C the RDC values for C5/C6/C8/C2 carbons are presented. The low RDC values in the loop and termini are again consistent with mobility in these regions. Some relatively low RDC values within the helix, such as at G3 and C17, are also consistent with mobility. However, rigid regions may be oriented in such a way that the RDC value is small. That appears to be the case within the helix, since inclusion of these RDCs as orientational constraints resulted in a well-ordered ensemble that is largely consistent with A-form geometry.

In summary, chemical shift, resonance intensity and RDC values are each consistent with the calculated disorder in the loop and termini. There may be a slight trend, particularly in the 5´-strand, showing more order in the end of the helix nearer the
loop (see Figure 11). However, all regions of the helix appear to be far more ordered than the loop and termini are.

**FIGURE 11:** (A) Comparison between SLB $^{13}$C chemical shifts and chemical shift of nucleotide monophosphates (NMPs). The chemical shift difference $\delta$(NMP)-$\delta$(SLB) is shown in ppm. (B) Normalized intensity of C6/C8/C2 resonances in $^1$H,$^{13}$C-HSQC spectrum. (C) SLB $^1$H,$^{13}$C RDC values extracted from $^1$H,$^{13}$C ARSTY spectra. In each section, the regions comprising the helix are shaded grey, and the color/shape coding of the bars is defined to the right of the figure.
DISCUSSION

Helical Parameters

Rigid body parameters, including shear, rise and twist can be used to describe the geometry of a base pair and the sequential base pair steps. Typical A-form RNA parameters include the following parameters: base pair shear ~0.0 Å and a helical twist ~31°. Deviation from these parameters is often observed when a non-Watson-Crick base pair is incorporated into a helix. The most prevalent non-Watson-Crick base pair is the G-U wobble base pair.

The rigid body parameters of the final ten structures of SLB were analyzed using the program CURVES. In each of the ten structures, the shear, twist and rise parameters fall close to standard A-form values, with the following exceptions: (i) the shear of G4-U21 and A5-U20 base pairs are ~ -1.9 Å and ~ +1.5 Å, respectively and (ii) the twist of inter-base pair steps G4/A5 and A5/U6 are ~41.9° and ~17.0°, respectively. Shear is an intra-base pair parameter that is directly related to the hydrogen bonds and indicates the sliding of one base with respect to the other. Therefore, the shear value is correlated with the twist value, the latter of which is measured via the relative position sequential C1’ atoms. The average observed shear value for the G4-U21 wobble base pair (~ -1.9 Å) is a feature consistent with wobble base pairs presented in the literature, as are unusual twist parameters near a G-U pair. The observed twists can be characterized as an over-twisting of the helix at the G4/A5 step and an under-twisting at the next step, A5/U6. Ananth et al. report that over-twisting (~47°) is common at Watson-Crick(WC)/GU base pair steps while under-twisting is often observed in the
next base pair step (GU/WC $\sim 19^\circ$). However, the over-twisting and under-twisting near a G-U pair is often reported at different positions. For instance, Joli et al. reported the WC/GU base pair step experiences an under-twisting while the GU/WC step is slightly over-twisted. Here, SLB contains a similar pattern to that reported by Ananth et al., an over-twist followed by an under-twist. However, this pattern appears one base pair further down the chain relative to the wobble pair.

To ensure the over-twisting/under-twisting of the helix was not influenced by artificially restricted backbone dihedral angles, an additional 400 structures were calculated excluding backbone dihedral angle constraints for nucleotides G3, G4, and A5. In the ten lowest energy structures, the helical parameters, including the position of the over-twisting and under-twisting, were similar to those discussed above, and no other noticeable changes were observed.

**Major and Minor Groove Width**

Typical A-form RNA contains a narrow (~3.0 Å) but deep major groove and a wide (~11 Å) but shallow minor groove. Therefore, protein-RNA base interactions are largely restricted to the wider minor groove side. In contrast, B-form DNA has a wide (~11 Å) major groove that often accommodates protein alpha helices. The groove width is defined via the shortest cross-strand P-P distances on the relevant side of the double-strand. For A-form RNA, these shortest distances are from $P_i$-$P_{i+6}$ (major groove) and from $P_i$-$P_{i+3}$ (minor groove).

In HRV-14 SLB, the minor groove width, as measured from cross-strand $P_i$-$P_{i+3}$ distances, averages 10.7 Å, which is close to the typical A-form range. However, the
shortest cross-strand phosphorous distances in the SLB major groove are from $P_i-\overline{P}_{i-5}$ and yield an average major groove width of 8.7 Å (Figure 12), atypically large for A-form RNA. This is despite the fact that other helical parameters are consistent with A-form geometry.

The reason for this discrepancy is the length of the helix: typical A-form RNA contains eleven base pairs per turn of the helix, with seven base pairs being the minimum required to produce a $P_i-\overline{P}_{i-6}$ contact to close off the major groove. SLB forms only six base pairs, and so the stem, though it follows the pattern of standard A-form, is not long enough to close off access to the major groove. Thus, the SLB major groove, as calculated, is more accessible than is the major groove in a longer A-form stem. This accessibility could play a role in interactions with PCBP or other unidentified proteins and molecules.

Interestingly, the major groove of SLD of the HRV-14 5´CL has previously been shown to contain a similarly accessible major groove $^{30, 83}$, also with the potential to enhance interactions with host or virus proteins. It remains to be seen whether these two major grooves remain as accessible in the context of the intact 5´CL RNA. Also to be determined is their relative orientation, which could influence the juxtaposition of proteins that bind to each of these stem loops, including potentially to their major grooves.
FIGURE 12: A space-filling view of the average calculated structure of HRV-14 SLB. The helix region is shown in grey and the loop (top) and termini (bottom) are blue. The red dashed line shows the shortest distance between phosphorus atoms P\textsubscript{i}-P\textsubscript{i-5} across the major groove (G3-C17).
Role of SLB and PCBP

Picornaviruses are positive-sense single-stranded RNA viruses. The known role for SLB from the picornavirus 5´CL in virus replication is to attract the host PCBP. The PCBP then coordinates with host poly-(A) binding protein (PABP) that binds to the A-rich 3´UTR of the virus RNA. This interaction effectively circularizes the virus RNA and assists negative strand RNA synthesis, using the positive sense genomic RNA strand as a template. The nascent negative strand is then used as a template to create multiple positive strands, effecting virus replication. Thus, interaction of SLB with PCBP is a critical step in the replication process, and detailed structural knowledge of this interaction and its constituents may lead to ways to block replication.

Human PCBP contains three K-homology (KH) domains, the first and third of which are capable of binding C-rich single-stranded RNA while the second KH domain is important for dimerization of the protein. The first KH domain (KH1) is the major determinant of 5´CL binding. The presumptive site on SLB for PCBP is the UCCCA sequence (nucleotides 12-16) located in the loop (G9-A16) (Figure 7).

Two crystal structures of the KH1 domain from PCBP, in complex with a single stranded C-rich DNA strand representing a human telomere, have been determined, along with a third structure in which the DNA is replaced with its RNA equivalent. In each case, the RNA-binding cleft in PCBP is narrow, therefore direct contact is limited to a single-stranded nucleic acid sequence of four nucleotides in length. The crystal structures, though similar in many ways, show an interesting difference in the register of the interacting polynucleotide: KH1 contacts the sequence ACCC, CCCT, and CCCU in the three structures, respectively. Each crystal structure shows
essentially identical interactions between the KH1 domain and the two central (cytosine) nucleotides from the interacting tetrad. These two cytosine nucleotides hydrogen bond to two guanidinium groups (from KH1, R57 and R40, respectively) via the cytosine hydrogen bond acceptors O2 and N3. Specific interactions were not observed between KH1 and the first nucleotide (A or C) of the tetra-nucleotide sequence. The last nucleotide hydrogen bonds to KH1 via side chains of E51 (if the nucleotide is C) or via the carbonyl group of I49 (if the nucleotide is T or U). The reason for this apparent difference in register of the CCC sequence is not completely clear at this time.

Mutational analysis has been done to help identify the site on SLB for PCBP recognition. PCBP binding is inhibited when the CCC nucleotide sequence (C13-C15) is mutated to CAC, identifying C14 as important. Also, deletion of CCCA sequence (C13-A16) significantly decreases binding of PCBP to SLB. These studies do not however specify whether PCBP contacts the UCCC (nucleotides 12-15) or CCCA (nucleotides 13-16) tetrad in SLB. Some insight can be gained by analysis of the SLB structure. The four consecutive pyrimidine bases located in the loop region (nucleotides 12-15) are solvent exposed in most of the final calculated structures. The increased flexibility observed within this region of the loop (Figure 11) suggests all or some of these nucleotides would be immediately accessible for protein interactions. Therefore, UCCC may be the tetra-recognition site for PCBP binding. This would place C13 and C14 central in the recognition tetrad, with conserved contacts to the arginine as discussed above for the crystal structures. If this is correct, then C15 would likely interact with the PCBP glutamic acid group, as in the first structure discussed above.
with perhaps non-specific interactions involving U12. However, structural analysis of
SLB with PCBP is needed in order to confirm the tetra-recognition sequence.
CHAPTER III

CONFORMATIONAL FLEXIBILITY IN THE PICORNAVIRUS
RNA REPLICATION PLATFORM

PREFACE

The content of this chapter submitted to RNA in November, 2018 and is currently under review. The formatting has been altered to include the supplemental material in the manuscript. The IBC number for this project is 12-007.

INTRODUCTION

The aim of the research described in the following chapter was to determine the first high resolution structure of HRV-14 5’CL and study the structural effects in the absence and presence of magnesium. As mentioned in Chapter 1, Picornaviruses are responsible for a number of diseases and conditions ranging from poliomyelitis to the common cold. Their genetic information is carried as a small single-stranded RNA molecule (~7.5-8.3 kb) that is replicated through a highly conserved mechanism. The genome consists of three distinct sections: an extensive 5´ non-coding region (NCR) and a much smaller 3´-NCR, that together flank a large single open reading frame. This open reading frame encodes a single large polyprotein (~250 kDa).

The 5´-NCR can be further subdivided into two regions. A large (~400-500 bases) internal ribosome entry site (IRES) is immediately 5´ to the ORF, where it directs translation of the polyprotein without the need for a capped mRNA structure. At the extreme 5´-end of the genomic RNA, a smaller (~83 base) region, predicted to form a
cloverleaf structure (5´CL) in most picornaviruses, serves as the main replication platform for virus replication \(^1,^{50-52}\). Sequence analysis of the 5´CL predicts the presence of one stem (stem A or SA) and three stem loops (SLB, SLC and SLD) (see Figure 13). SLD attracts the fusion protein 3CD, delivering the 3D polymerase to the required site to begin synthesis of a negative strand RNA\(^{34}\). This negative strand then serves as a template to produce multiple copies of the positive strand, which can be used either as additional mRNA or as genetic material for nascent viral particles that are released from the cell and can go on to infect other cells.

The loop region of SLB contains a C-rich sequence that attracts the host poly(C)-binding protein (PCBP). PCBP functions both in translation, via binding to C-rich regions in the IRES, and in replication, via binding to the 5´CL SLB\(^{52,53}\). However, when PCBP is attached to the IRES, PCBP can be cleaved by the virus-encoded 3C protease. The cleaved PCBP can then only function in replication. Thus, cleavage of IRES-bound PCBP serves effectively as a switching mechanism from translation to replication \(^{33,34,54-56}\). Furthermore, the binding of PCBP to SLB facilitates the binding of 3CD to SLD and triggers circularization of the RNA genome, two events that are essential for the synthesis of the negative strand\(^{25}\). A better understanding of the structural basis of these interactions and their interdependence could lead to ways to inhibit virus replication, leading to effective treatment or immunization against the large number of diseases and conditions associated with picornavirus infection.
In this study, a combined nuclear magnetic resonance (NMR) and small angle X-ray scattering (SAXS) approach has been used to determine the solution structure of the 5’CL RNA from isotype 14 of the human rhinovirus (HRV-14). Results show that the cloverleaf conformation, in particular its compactness, conformational stability and relative orientation of the two largest stem loops (SLB and SLD), is highly dependent upon the absence vs. presence of magnesium. This study represents an important step in the process of building a high-resolution structural understanding of picornavirus replication.

**FIGURE 13:** Secondary structure prediction of HRV-14 5’CL. The structure is predicted to form a four junction containing stem A (SA; blue), stem-loop B (SLB; magenta), stem-loop C (SLC; red) and stem-loop D (SLD; green). Protein binding site in SLB (magenta stars) and SLD (green stars) are shown, along with colored numbering for key nucleotides discussed in the text. Structure was predicted using MC-Sym\textsuperscript{35}. 

\[
\text{\footnotesize 5' 1' U A A-80 U C C 3'}
\]

\[
\begin{array}{c}
\text{SA} \\
\text{SLB} \\
\text{SLC} \\
\text{SLD}
\end{array}
\]
MATERIALS AND METHODS

RNA Preparation

A DNA transcription template was designed to contain a T7 RNA polymerase promoter followed by a 5’ cis-acting hammerhead ribozyme sequence and 83 nucleotides corresponding to HRV-14 5´CL. The double stranded DNA plasmid was cloned into the pUC19 vector using the EcoR1 and BamH1 restriction sites. For in-vitro transcription, the DNA plasmid was linearized using Bsa1. A natural abundance RNA sample of HRV-14 5´CL was produced by in vitro transcription using T7 RNA polymerase (expressed and purified in house) with RNAPoly reaction buffer [New England Biolabs; 40 mM Tris- HCl, 6 mM MgCl2, 10 mM DTT, and 2 mM spermidine, pH 7.9], each rNTP at 5 mM (Sigma), 0.02 mg/mL double-stranded DNA template, and 0.001 unit/mL pyrophosphatase (Fisher). A 10 mL transcription reaction mixture was incubated at 37 °C for 3 h and the RNA was ethanol precipitated with 2.5 volumes of 100% ethanol and 0.1 volumes of 3 M sodium acetate and the mixture was incubated overnight at -20 °C. The solution was spun down at 9,200xg and 4 °C. The resulting pellet was washed in 70% ethanol and spun down for an additional 10 minutes at 9,200xg. The pellet was dissolved in 8 M urea loading buffer [8 M urea, 25 mM EDTA, 500 mM TrisHCl, 4.6 mM xylene cyanol, and 3.7 mM bromophenol blue, pH 7.9] and heated for 3 min at 98 °C prior to gel purification. The RNA was resolved on a 6% denaturing polyacrylamide gel containing 8 M urea and visualized via ultraviolet (UV) shadowing at 254 nm, and the corresponding band was excised from the gel. The RNA was recovered from the gel pieces using an Elutrap device (Whatman) in 1×Tris-borate-
EDTA (TBE) buffer, pH 7.9. The concentration of the eluted RNA was determined using a Biodrop UV-visible spectrophotometer. The RNA was concentrated and buffer exchanged using a Vivaspin spin concentrator (3 kDa cutoff) to a final concentration of 6 mg/mL in NMR buffer [10 mM sodium phosphate and 0.1 mM EDTA, pH 7.0]. The sample was purified with Superdex 200 Increase 10/300 GL gel filtration column and further diluted to create 3 mg/mL and 1.5 mg/mL samples (SAXS samples). All three samples (6, 3, and 1.5 mg/mL) were dialyzed thoroughly with degassed NMR buffer. The dialysis buffer was retained for SAXS solvent blank measurements. An additional three samples (6, 3, and 1.5 mg/mL) were purified and dialyzed in NMR buffer containing 5 mM MgCl₂. A labeled sample of the 5´CL was similarly produced using 2 mM each of ¹³C and ¹⁵N-labeled rGTP and rUTP (Cambridge Isotope Laboratories) with 8 mM each of natural abundance rATP and rCTP (Sigma). The sample was divided to produce one labeled sample with 25 mM MgCl₂ and one sample containing no MgCl₂.

**NMR Spectroscopy**

A 0.50 mM unlabeled RNA sample in 10 mM sodium phosphate and 0.1 mM EDTA (pH 7.15) containing 90% H₂O/10% D₂O mixture was used for the assignment of exchangeable protons. NMR experiments used for structure calculations were performed on a Bruker Avance III 750 MHz at the National Magnetic Resonance Facility at Madison (NMRFAM) equipped with a cryoprobe. For assignment of exchangeable proton resonances, ¹H,¹H-NOESY spectra in H₂O were recorded, in the absence and presence of 25 mM magnesium chloride, with mixing times of 150 ms at 5.1 °C with 256 increments, 2048 data points, 160 scans per fid using excitation sculpting with gradients
for water suppression. Samples with uniform $^{13}$C and $^{15}$N enrichment of G and U nucleotides, in the absence and presence of 25 mM magnesium chloride, were used for residual dipolar coupling (RDC) measurements. $^1$H,$^{15}$N-ARTSY spectra were recorded in isotropic media, and in the presence of 12 mg/mL Pf1 filamentous bacteriophage. The error of the RDC measurement was assessed via the ratio of the isotropic and attenuated peak intensities compared to the spectral noise. The data for all NMR spectra were processed using NMRPipe and visualized and analyzed using NMRDraw and NMRViewJ.

**Small-angle X-ray Scattering**

SAXS experiments were carried out on a Bruker Nanostar benchtop SAXS system (Bruker AXS) at NMRFAM equipped with a rotating anode (Cu) Turbo X-ray Source and a VÅNTEC-2000 (2048X2048 pixel) detector. Measurements were carried out in 10 mM NaPi, 0.1 mM EDTA, pH 7.15 in the absence and presence of 5 mM MgCl$_2$. The buffer used for the SAXS experiments contains 10 mM NaPi and 0.1 mM EDTA, pH 7.15. Samples of 5´CL and 5´CL with magnesium were clarified by passage through a 0.2 mm filter before loading into a glass capillary cell. SAXS data were collected at three concentrations 1.5, 3, and 6 mg/mL. No significant inter-particle interactions were observed for any of the concentrations. The 3 mg/mL data was used for model calculation. The sample-to-detector distance was set at ~1 m, allowing for the detection range: 0.012 > q > 0.25 Å$^{-1}$. Forty microliters of RNA and buffer samples were loaded separately into a capillary cell with 1 mm diameter, and scattering data were collected for 4 hours. The ATSAS software suite was used to process the SAXS data. The $R_g$ for each sample
was determined by using the Guinier approximation in the q range \((q_{\text{max}}R_g) < 1.3\). Pairwise distance distribution functions \((P(r))\) were obtained using the GNOM software.\(^9\) The output from GNOM was then used in conjunction with DAMMIF\(^6\) to generate 20 independent \textit{ab initio} dummy atom models to assess the molecular shape of each sample. No symmetries were enforced during the calculations. Most of the models exhibited excellent agreement with experimental data and had a normalized spatial discrepancy (NSD) < 1. CRY SOL software\(^7\) was used to compare the 5’CL structures with experimental SAXS data. Supcomb software\(^8\) was used to superimpose the 5’CL structures on to the SAXS \textit{ab initio} dummy atom models. \(V_c\) approach was used for the molecular weight calculation from the SAXS data.\(^9\)

**NMR/SAXS-based Structure Calculation and Refinement**

A total of 2000 preliminary structural models for HRV-14 5´CL were generated using MC-Sym\(^3\). The MC-Sym models were built using reference pdbs for each domain of the 5´CL, which consisted of the previously determined structures of HRV-14 SLB\(^1\) and SLD\(^3\),\(^6\),\(^8\) together with SA and SLC structures generated using RNAComposer\(^1\), the latter two based on secondary structure prediction and confirmed via assignment of imino proton peaks in the above \(1^H,1^H\)-NOESY spectra of the 5´CL. For the non-magnesium data, the 2000 models were filtered down to 500 models by comparing 20 \(1^H-15N\) experimental RDC vs calculated RDC values determined using PALES\(^1\). The 500 models were subject to \(\chi^2\) goodness of fit analysis between the experimental and the calculated SAXS profile, with the 10 structures having the lowest \(\chi^2\) values \((\chi^2 < 15.1)\) taken as preliminary structures to begin XPLOR-NIH\(^6\)
calculations. The procedure was repeated for the data in the presence of magnesium, and ten best structures with the lowest $\chi^2$ value ($\chi^2 < 16.5$) were chosen for further refinement with XPLOR-NIH$^{62}$.

In the case of the non-magnesium sample, for each of the 10 preliminary pdbs, 10 refined structural models were calculated using a simulated annealing and rigid body refinement protocol in XPLOR-NIH$^{62}$ that incorporated both the NMR and SAXS data with floating alignment tensor. In initial calculations in the absence of magnesium, the stem regions of the SA (A5:G9 and C75:U79), SLB (C10:G16 and C25:G31), SLC (A32:C34 and G40:U42) and SLD (U47:C57 and G61:A71) structures were held rigid and contained hydrogen bond constraints, while the nucleotides in the loop and junction regions were given torsional degrees of freedom along with two other nucleotides whose $^1$H imino resonances were not visible in the 2D $^1$H,$^1$H-NOESY (G46 and U80). For final calculations, rigidity was removed for nucleotides U52, U66, G12 and U29, however hydrogen bond constraints for these were retained, as described further in the Results section. In the presence of magnesium, based upon similar imino assignments and NOEs, the regions of the stem loops held rigid were the same as the initial calculations above for the absence of magnesium, with the following exceptions; nucleotides U52, U66, C10, and G31. The hydrogen bond constraints were retained for U52 and U66, but were removed for C10 and G31, as described further in the Results section.

In simulated annealing calculations, the initial annealing temperature was set to 3000 K with a final temperature of 25 K. The force constants for hydrogen bond constraints were ramped from 2 to 50 kcal mol$^{-1}$ Å$^{-2}$ and from 0.02 to 5 kcal mol$^{-1}$ rad$^{-2}$.
for RDC constraints. In the case of no magnesium, the structures were refined with 105 SAXS data points ($0.03 < q < 0.18 \, \text{Å}^{-1}$) using 20 one-bond $^1\text{H}^\text{15N}$ residual dipolar couplings, 50 planarity constraints, 124 hydrogen bond-derived constraints along with SAXS data and rigid body minimization in XPLOR-NIH$^{62}$. From the resulting 100 total refined structures, the eight lowest energy structures found with no RDC violations greater than 3 Hz and SAXS $\chi^2$ value of 1.44 or less were taken as the final ensemble, with a pair-wise RMSD of 7.735 Å (Figure 17a). Similarly, from the 100 refined structures in the presence of magnesium, eight structures were found with no RDC violations greater than 3 Hz and SAXS $\chi^2$ value of 0.924 or less. These eight structures were taken as the final structural ensemble in the presence of magnesium, with an average pair-wise RMSD of 1.812 Å. The constraint and geometry calculations for each condition are shown in Table 3.

**RESULTS**

**NMR Analysis of HRV-14 5´CL**

**Peak Assignments**

Imino protons, which are present only in guanine (G) and uracil (U) bases, are highly exchangeable with solvent protons. Therefore, the presence of the imino resonance implies protection, most notably via hydrogen bonding in a stable base-pair. The absence of an imino resonance conversely suggests the absence of a stable hydrogen bond and is often indicative of presence in a dynamic loop or bulge region. Resonances of imino protons in a Watson-Crick base pair range from ~11.5-13 ppm (G)
and ~13-15 ppm (U). Resonances observed outside of the above spectral range indicate the presence of a non-Watson-Crick base pair, such as a GU (10-12 ppm) or a UU base pair (10.4-11.3 ppm).

FIGURE 14: Imino NOESY spectra in the absence (blue) and presence (red) of magnesium. The doubling of nucleotide U29 (solid box) disappears in the presence of magnesium. The imino resonance of nucleotide U67 is only observed upon the addition of magnesium, with sequential imino-imino cross peaks to U67 shown in the dashed box.
The HRV-14 5´CL molecule contains a total of 45 imino protons, 26 of which were observed in the two dimensional ¹H,¹H-NOESY spectrum (Figure 14) in the absence of magnesium. Five strong resonances are located outside the spectral range of a Watson-Crick base pair and were assigned to nucleotides U50, U52, U66, U68 and G12 as described below. In general, imino proton resonances were assigned based on sequential imino-imino NOESY cross peaks (Figure 14), with the assistance of imino proton resonance assignments from previously published works with isolated HRV-14 SLB¹⁰⁰ and SLD³⁰. The observed pattern of sequential imino-imino NOE cross peaks confirmed the formation of SA, SLB, SLC and SLD helices, and the absence of additional imino resonances indicated nucleotides that are not involved in stable hydrogen bonding (for a full list of the hydrogen bond constraints see Materials and Methods).

**Chemical Shift Changes upon Incorporation into Cloverleaf**

The imino chemical shift difference between SLB and SLD in previously described isolated stem loops³⁰,¹⁰⁰ vs. when incorporated into the 5´CL (without magnesium) are shown in Figure 15a. The differences are small (RMSD 0.028 ppm), consistent with conservation of the structure, excluding the large difference observed for G11. Also, a new peak appears for the G31 imino resonance, indicating stabilization of the C10-G31 base pair only in the 5´CL, and not in the isolated SLB. This effectively extends SLB by one base pair, matching the predicted length of SLB (see Figure 13). This additional base pair evidently alters the chemical shift of the adjacent G11 resonance. In addition, the imino resonance of U29 appears to split into two or more
resonances in the cloverleaf (Figure 14), suggesting the presence of two or more slowly interconverting states, which was not observed in the isolated SLB structure. Since U29, G11 and G31 are from sequential base pairs at the junction end of SLB, it appears that alterations are localized to this region. To incorporate these data into 5’CL structure calculations in the absence of magnesium, hydrogen bonding constraints were added for C10-G31, and hydrogen bonding constraints were retained for G12-U29, but rigidity of the G12-U29 base pair was relaxed.

**Chemical Shift Changes upon Addition of Magnesium**

Upon addition of magnesium to the 5´CL, chemical shift changes indicate a return of SLB to its structure when isolated: the splitting of the U29 resonance disappears, the G31 imino resonance also disappears, and the G11 imino resonance returns to the chemical shift observed in the isolated SLB. Therefore, the above adjustments made to SLB for calculations in the absence of magnesium, were removed for calculations in the presence of magnesium.

Also, two significant changes are observed in SLD. First, a resonance appears for U67, indicating stabilization of the C51-U67 base pair from the pyrimidine mismatch region of SLD. Although the U67 imino resonance was not observed in the isolated SLD structure analysis\(^3\), other NOEs and RDCs clearly indicated the presence of C51-U67 hydrogen bonding in isolated SLD, and so no change was required in the calculations, other than incorporation of the U67 imino RDC.

The second significant change in SLD is alteration of the U66 imino chemical shift by approximately 0.3 ppm (Figure 15b). This alteration at the U52-U66 base pair is
FIGURE 15: $^1$H chemical shift changes within SLB (magenta) and within SLD (green) induced (A) by incorporation into the 5´CL (without magnesium present) and (B) by the addition of magnesium (in the context of the 5´CL).
likely related to the above-mentioned stabilization of the adjacent C51-U67 base pair. This change is also related to a required adjustment made in calculations: in initial calculations of the 5´CL structure, both in the presence and in the absence of magnesium, both the U52 and U66 imino RDCs were consistently violated. Non-Watson Crick base pairs containing two imino protons, such as U-U or G-U pairs, would be expected to have similar \(^1\text{H}-^{15}\text{N}\) RDC values if the base pair is nearly planar. However, the RDC values for the U52-U66 base pair were measured as 38.5 Hz for U52 and 16.3 Hz for U66 in the absence of magnesium, and 39.8 Hz for U52 and 32.8 Hz for U66 in the presence of magnesium. In the isolated structure of HRV-14 SLD, the imino bonds for U52 and U66 are nearly coplanar. Therefore, the conformation of the U52-U66 base pair in the 5´CL is not consistent with the isolated SLD structure. Therefore, for calculations in both the absence and presence of magnesium, hydrogen bonding constraints were retained for U52-U66, but rigidity of the base pair was relaxed, allowing the structure to adjust to the observed RDCs.

**SAXS: Global Structure of HRV-14 5´CL**

Small angle X-ray scattering (SAXS) probes the global shape of biological macromolecules in solution. SAXS is highly complementary to NMR: it is a mid-resolution technique that does not require crystallization or isotope labelling, has no size limitations and requires low sample quantities. Here, SAXS was used to investigate the global structure of HRV-14 5´CL in the absence and presence of magnesium. The measured radius of gyration \((R_g)\) and the maximum dimension \((D_{\text{max}})\) for the HRV-14 5´CL in the absence of magnesium were 26.8 and 85 Å, respectively. Upon the addition
FIGURE 16: SAXS data plots for HRV-14 5’CL. Plots are shown for data in the absence (blue) and presence (red) of magnesium. (A) Kratky plot (B) Pairwise distance distribution function (C) Overlay of the ab initio model of 5´CL in the absence (left) and presence (right) of magnesium with the helices colored as shown in Figure 13. (D) Overlay of experimental and calculated SAXS profiles.

of 5 mM MgCl₂, the Rg and Dmax decreased to 22.3 and 70 Å, indicating that magnesium triggers a conformational change to a more compact structure, consistent with the NMR results discussed above. A Kratky plot (q vs. q²•I) can be used to assess the overall shape and flexibility of the molecule. A bell-shaped Kratky curve indicates a well-folded globular and compact conformation. The Kratky curves in the presence and absence of magnesium are shown in Figure 16a. The HRV-14 5´CL in the absence of magnesium shows a bell-curved shape with a slight plateau at higher q values, whereas the 5´CL in the presence of magnesium displays a more evenly distributed plot. This confirms that upon the addition of magnesium the 5´CL forms a more well-folded compact conformation. This more compact shape with the presence of magnesium is consistent with the pair-wise distribution function (P(r)) plot in Figure 16b.

Full Structural Calculation: NMR and SAXS

In the absence of magnesium, 100 three-dimensional structures of HRV-14 5´CL were calculated as described in Materials and Methods, using simulated annealing and rigid body protocol as employed by XPLOR-NIH. The eight best final structures selected to represent the conformation have an average pair-wise RMSD of 7.735 Å
with the A-form helix regions having an RMSD of 6.981 Å (Table 3 and Figure 17a). A similar procedure was performed on HRV-14 5´CL in the presence of 25 mM MgCl₂ with the eight-lowest energy structures having an RMSD of 1.812 Å, and the A-form helical regions having an RMSD of 1.536 Å (Table 3 and Figure 17b). In the absence of magnesium, 5´CL adopts an open conformation with SLB and SLD positioned approximately 90° apart. In the presence of magnesium, a more compact structure forms with SLB and SLD in close contact and arranged in a parallel form. The positions of SA and SLC in both structures are less defined, at least in part due to the short length of the helices, which renders their position less sensitive to fitting of both the NMR and SAXS data.
FIGURE 17: Structure of HRV-14 5´CL. Superimposition of the 8 lowest energy structures of HRV-14 5´CL in (A) the absence of magnesium and (B) the presence of magnesium. The loop and terminal regions are colored in grey; the stem regions are colored as in Figure 13. Atoms from the bases of U29, U66 and U67, each of which show significant changes in NMR spectra upon addition of magnesium, are shown as white spheres.
Comparison of SAXS-only Models with SAXS/NMR-based Structures

Medium-resolution ab initio models of the 5´CL shape (Figure 16c), based only on the SAXS experimental profile, were calculated independently from the XPLOR-calculated structures using DAMMIF software. In the absence of magnesium, the ab initio model is somewhat Λ-shaped, but with one leg longer than the other. Overlay of this model the lowest energy SAXS/NMR-based structure from XPLOR-NIH calculations shows a good fit (χ² value of 1.44). The extended leg fits the longest stem loop (SLD), while the next longest stem loop, SLB, is fitted into the shorter leg. In this overlay, part of the SLB loop region extends outside of the ab initio model. However, this SLB loop has been shown to be highly dynamic, and so its position in this figure is somewhat misleading. As a dynamic loop, it can explore many conformations and positions, and so may not give rise to intense density in the ab initio model.

In the presence of magnesium, the ab initio SAXS structure is more symmetrical and compact. This matches well with the XPLOR-NIH calculated structure in the

<table>
<thead>
<tr>
<th>TABLE 3. Structural Statistics for HRV-14 5´CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS Constraint/Geometry Violations</td>
</tr>
<tr>
<td>-------------------------------------</td>
</tr>
<tr>
<td>PfR RDCs (Hz)</td>
</tr>
<tr>
<td>SAXS data fit (χ²)</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
</tr>
<tr>
<td>Bond angles (°)</td>
</tr>
<tr>
<td>Impropers (°)</td>
</tr>
<tr>
<td>Ensemble RMSD (Å)</td>
</tr>
<tr>
<td>Stems</td>
</tr>
<tr>
<td>SLB and SLD</td>
</tr>
<tr>
<td>All atoms</td>
</tr>
</tbody>
</table>
presence of magnesium ($\chi^2$ value of 0.924). The compactness arises when SLB and SLD align and come in to contact with each other. Together, these comparisons of *ab initio* SAXS-based models with the XPLOR models provide confirmation of structural veracity. Superimposition of the eight lowest energy structures and the SAXS *ab initio* model is shown in Figure 18.

**FIGURE 18**: Superposition of 5´CL and SAXS *ab initio* model. Eight lowest energy structures of 5´CL in the absence (A) and presence (B) of magnesium. In the absence of magnesium, the structures are divided into two subgroups; subgroup 1 (left) and subgroup 2 (right).
RNA Four-Way Junction: Helical Arrangement

RNA junctions are common structural elements found in, for instance, ribosomes, ribozymes, group II introns, and tRNAs. The topology of a junction is characterized by the pattern of coaxial stacking of helices; pairs of helices interacting end-to-end to form one continuous or nearly continuous helix. Based on coaxial stacking patterns, RNA four-way junctions such as 5´CL (with four helices from SA, SLB, SLC and SLD) can be classified into nine different families: families H, cH, and cL contain two pairs of coaxial helices; cK and π families contain one stacked helix pair each; cW, ψ, X and cX families contain no stacked helices\textsuperscript{104}. The HRV-14 5´CL structure in the absence of magnesium shows no clear evidence of coaxial stacking for three of the eight final structures whereas the remaining five present possible stacking between helices SA and SLB. Doubling present in $^1$H, $^1$H-NOESY spectrum, the longer axis length observed in the SAXS ab initio model and the $P(r)$ curve, and the decrease in intensity for the Kratky plot each support the notion that the structure in the absence of magnesium is somewhat extended (not maximally compact). Topologically, the structures fit either the cK or cX classification.

In the presence of magnesium, possible stacking of SA and SLB was also seen in two of the eight final structures, with no stacking evident in the remaining structures. Although somewhat similar coaxial stacking patterns are observed for both conditions, in the presence of magnesium the structure topology is very different, with more evidence of ordered compactness, and with the two longest stems, SLB and SLD, positioned parallel to each other and in close association. Topologically, most of the structures fit the cW or ψ classification.
DISCUSSION

Picornavirus replication is a highly conserved mechanism that is initiated by protein-RNA complexes forming upon the 5´CL RNA. A C-rich region of the SLB loop specifically binds to the host poly(C)-binding protein (PCBP), facilitating the binding of virus fusion protein 3CD with SLD\textsuperscript{25}. The PCBP then coordinates with the host poly(A)-binding proteins (PABP) that binds to the A-rich 3´UTR in the virus genome\textsuperscript{19, 54}. The PCBP-PABP interaction effectively circularizes the virus genomic RNA, allowing the RNA replicase (3D) to access the 3´UTR, where negative strand synthesis initiates. Thus, circularization requires protein-RNA interactions by PCBP, PABP and 3CD\textsuperscript{52, 53}. The resulting nascent negative strand is then used as a template to create a large number of new positive RNA strands that serve as mRNA and as genetic material for new virus particles. The determination of the three dimensional structure of HRV-14 5'CL, particularly the relative position and orientation of SLB and SLD, upon which PCBP and 3CD assemble, is a vital step toward understanding the replication mechanism of picornaviruses and toward developing therapeutic agents for inhibition of picornavirus replication.

Influence of Magnesium

The present study presents two distinct structures for HRV-14 5'CL, in the absence and presence of magnesium (Figure 17). Metal cations play a vital role in RNA folding and stabilization in many biological processes. The positive charge of metal cations can shield the repulsion of the phosphate ions present in the backbone of RNA, allowing RNA to fold into a more stable and compact tertiary conformation. Magnesium
is the most often identified cation in RNA structures and is considered the most important divalent cation for RNA stabilization \(^{105-107}\).

In the absence of magnesium, HRV-14 5'CL takes on an open conformation with SLB and SLD approximately perpendicular to one another (Figure 17a). As discussed in results, some structural dynamics is suggested by relatively broad NMR line shapes, doubling of some imino resonances in the \(^1H, ^1H\)-NOESY spectrum and a relatively low intensity in the Kratky plot and P(r) curve (Figure 16). In contrast, in the presence of magnesium, SLB and SLD move toward each other to form a more compact and more ordered structure (Figure 17b). Close association of the stems from SLB and SLD results in positioning of the two corresponding loop regions in close proximity (Figure 17b).

The magnesium-induced association of SLB and SLD suggests that magnesium binds somewhere along the SLB/SLD interface, helping to screen charge-charge repulsions between the phosphodiester backbones of the two stems. Consistent with this hypothesis, the largest NMR spectral changes in chemical shift upon addition of magnesium (Figure 15b) occur at nucleotides 11, 12, 29, 62 and 66, each of which are in SLB or SLD. In addition, as discussed previously, doubling of the U29 resonance disappears and a new peak arises for the U67 imino resonance.

The observed chemical shift changes seen at bases 11 and 12 upon magnesium addition are related to the destabilization of base pair C10-G31 as discussed in the results section. The remaining large changes occur in bases 29, 62, 66 and 67. Three of these bases, U29, U66 and U67, come into close contact with each other in the
magnesium structure (Figure 17b, depicted as spheres). This suggests stable site binding of magnesium very near these bases.

**Fold in Absence of Magnesium**

Closer examination of the structures in the absence of magnesium reveals the presence of two distinct structural subgroups, differing in the orientation of SA relative to SLB. The first subgroup contains five of the eight structures (Figure 19a) with the orientation of SA to SLB similar to the relative orientation in the presence of magnesium (Figure 19b). The remaining three structures effectively reverse the positions of SA and SLC, with SA appearing behind SLC in this view (Figure 20). Both subgroups fit the NMR and SAXS data, however, the first subgroup (Figure 19a) has somewhat lower calculated energy, and would require far less structural rearrangement to form the closed structure upon addition of magnesium. The first subgroup shows stacking of SA and SLB, while the second subgroup does not. Topologically, the junction classification of subgroup 1 is cK, while the classification of subgroup 2 is cX. The above evidence suggests that the cK structure is the correct fold in the absence of magnesium.

**The SAXS/NMR Approach**

The combined SAXS/NMR analysis of RNA structure is by now a relatively well-established technique (e.g. 42-44, for recent review see 108). Furthermore, when high resolution structures of the isolated stem loops are available (via conventional NMR or crystallographic analyses), and when the conservation of these separate structures is verified upon incorporation into the complex, the SAXS/NMR approach is capable of
FIGURE 19: Changes in orientation induced by magnesium. (A) Structure in the absence of no magnesium. The black arrow showing SLD’s change in position upon the addition of magnesium (B) Stem regions are colored as in Figure 13.

producing high resolution information. In the present case, prior high resolution structures are available for the largest two stem loops (SLB and SLD\(^{30, 100}\)); which together comprise 61% of the total cloverleaf. As discussed in results, chemical shift analysis, together with RDC and SAXS fitting, confirm that SLB and SLD structures are largely conserved upon incorporation into the cloverleaf, and pertinent measures were taken to adjust the individual structures to the few data that indicate localized rearrangement.

Thus, information regarding the positioning of SLB vs. SLD in the cloverleaf is unambiguous: SLB and SLD are oriented well apart from one another, nearly at a right
angle, in the absence of magnesium, while they come together to interact closely in the presence of magnesium.

**FIGURE 20**: Conformational subgroups in the absence of magnesium. (A) Eight lowest energy structures. (B) Five structures composing subgroup 1. (C) Three structures composing subgroup 2.

**SLB and SLD Major Grooves and Relative Helical Twist**

Typical A-form RNA contains a narrow (~2.4 Å width) but deep major groove and a wide but shallow minor groove. The narrow major groove restricts protein access for
base-specific interactions to the minor groove side. However, the isolated solution
structures of HRV-14 SLB and SLD each contain atypically wide and accessible major
grooves (9 Å and 13 Å width, respectively)\textsuperscript{30, 100}, a feature that is retained upon
incorporation into the cloverleaf (see below). In this context, it is interesting to note that
magnesium has been shown to help stabilize non-Watson Crick base pairs\textsuperscript{109}.

The observed accessibility of the SLB major groove is simply due to the brevity
of the SLB helix (6-7 base pairs): it contains too few base pairs to form a complete turn
(11 base pairs) in order to close off the major groove. The characteristics of SLB are not
otherwise unusual.

The SLD helix, on the other hand, contains 12 base pairs when the three central
U-U and U-C pairs are included. This length would be sufficient to close off the major
groove if the structure was standard A-form. However, the presence of the non-Watson
Crick pyrimidine mismatch regions in the center of SLD appears to disrupt standard A-
form geometry. Non-Watson Crick base pairs have been shown to induce widened
major grooves that serve as recognition sites for protein-RNA interactions\textsuperscript{110, 111} for
instance in the HIV RRE/Rev peptide interaction\textsuperscript{112}. This can be understood by noting
that base-pairing of two pyrimidines requires closer inter-helical approach than does a
Watson-Crick pyrimidine-purine base pair, a stress which must distort the helix\textsuperscript{110}.
Similar widening of the major groove was also reported for SLD from Coxsackievirus
B3\textsuperscript{31} and for a consensus enterovirus SLD sequence\textsuperscript{32}.

We find no evidence that this established characteristic of SLD is lost when
incorporated into the 5´CL. As an additional test, we computationally replaced the three
central pyrimidine-pyrimidine base pairs in SLD with Watson-Crick pairs, and created a
standard A-form SLD computational mutant. Using the same XPLOR-NIH routine, we were unable to find an orientation of the mutant SLD that fits the observed H-N RDCs (removing RDCs from the mismatch region). This, together with the contrasting ability to fit all data using the previously determined SLD structure, with modification only of one U-U base pair orientation as described in results, corroborates that the SLD structure is largely conserved in the cloverleaf.

Next, we analyzed the relative positions of the SLB and SLD accessible major grooves, in the presence of magnesium. The ability of the SAXS/NMR approach to determine the relative helical twist of stem loops has been previously reported (e.g. 43, 44). In the present case, the relative helical twist of SLB and SLD orients the two accessible major groove regions in alignment with each other, creating a single accessible groove surface (Figure 21). This surface is approximately 30 Å in length, and spans the width of the cloverleaf. Access to this groove surface could potentially stabilize interactions with host and virus proteins during replication, suggesting that it may prove to be an attractive therapeutic target.
FIGURE 21: Major groove width of HRV-14 5′CL. The 5′CL adopts a major groove width atypical of A-form RNA.

Protein-RNA Modeling

In order to begin to understand how the replication complex forms, interactions of the 5′CL with the virus 3C protein and the host PCBP KH1 domain were modeled (Figure 22). Modelling was based on the previously determined solution structure of isolated HRV-14 SLD complexed with 3C\textsuperscript{36}, and a previously determined structure of the KH1 domain of PCBP with an oligonucleotide of sequence AACCCTA, which is similar to the sequence in the 5′CL SLB C-rich loop\textsuperscript{87}. The model in Figure 23 is based on previously determined structure of the poliovirus 3CD fusion protein\textsuperscript{113}. In Figure 23, 3D polymerase is positioned in front of the accessible groove. The 3D protein,
FIGURE 22: RNA-protein interaction modeling. Modeled binding of PCBP KH1 domain (yellow) with SLB and 3C protease (orange) with SLD (A) in the absence and (B) in the presence of magnesium.

The ability of magnesium to alter the 5´CL structure provides evidence that the two structures presented here, the open and closed conformation, are of similar total energy. Thus, both of the conformations could be present during different steps in the replication cycle, under the appropriate conditions. Ionic environment, and also other factors, such as protein interactions, may tip this conformational balance. For instance, it is possible that the closed conformation binds PCBP and 3CD, but then circularization...
FIGURE 23: RNA-protein interaction modeling. Modeled binding of PCBP KH1 domain (yellow) with SLB and poliovirus 3CD protease (3C:orange, 3D:rose) with SLD.

via interaction of PCBP with PABP may cause or require a switch to the open conformation. Such a switch to the open conformation could in turn be required to correctly position 3D at the 3´NCR for initiation of negative strand synthesis. Further structural studies performed with the 5´CL and replication proteins will however be needed to establish the role, if any, of 5´CL conformational switching in the replication mechanism of picornaviruses.
The work in this dissertation provides structural insight into picornavirus replication. The presumptive RNA cloverleaf at the start of the 5’-NCR of the picornavirus genome is an essential element in replication. Stem loop B (SLB) of the cloverleaf is a recognition site for the host PCBP, which initiates a switch from translation to replication. In chapter 2, the solution structure of HRV-14 SLB was determined using NMR spectroscopy. SLB adopts a predominantly A-form helical structure with five Watson-Crick base pairs and one wobble base pair, and is capped by an eight nucleotide loop. The wobble base pair introduces perturbations in the helical parameters, but does not appear to introduce flexibility. Due to the short length of the helix, major groove appears to be accessible. The accessible major groove allows for the access of protein interactions, potentially of PCBP, a interaction essential in replication. Unlike the helix, flexibility is seen throughout the loop and in the terminal nucleotides, with the pyrimidine-rich region, the apparent recognition site for the PCBP, being the most disordered region of the structure.

In chapter 3, using previously determined information from isolate solution structures of HRV-14 SLB and SLD, the solution structure of HRV-14 5’CL was determined using a combination of NMR and SAXS. Magnesium is known to stabilize RNA tertiary folding by shielding the negatively charged phosphates of the sugar phosphate backbone. In the absence of magnesium, the structure adopts an open, somewhat extended conformation. In the presence of magnesium, the structure compacts, bringing SLB and SLD into close contact, a geometry that creates an extensive
accessible major groove surface, and permits interaction between the proteins that target each stem loop. This work provides the first solution structure of the 5’CL in picornaviruses, a discovery necessary for a thorough understanding of picornavirus replication.

In Figure 23, 5’CL is modeled with poliovirus 3CD\textsuperscript{113} and the KH1 domain from PCBP\textsuperscript{87}. Future studies would include docking 3CD and PCBP protein to the 5’CL using HADDOCK software\textsuperscript{114} to model the biomolecular complex. HADDOCK uses experimental data, NMR distance constraints and RDC values, to determine the relative orientation of the each component in a complex. The model in Figure 23 can be compared to the HADDOCK docking model for further validation.

Since isolated SLB retained its conformation when incorporated in the 5’CL, the complex of isolated SLB and KH1-KH2 domain of PCBP can be determined using solution NMR and SAXS. The SLB-PCBP complex could then be used, along with the previously solved complex of SLD-3CD, to model the entire picornavirus replication complex in solution (5’CL-PCBP-3CD).
REFERENCES


# APPENDIX A

## CHEMICAL SHIFTS OF SLB AND 5’CL

### TABLE A1: Chemical Shifts of SLB (ppm)

<table>
<thead>
<tr>
<th>Base</th>
<th>H1('C1)</th>
<th>H2('C2)</th>
<th>H3('C3)</th>
<th>H2(C2)</th>
<th>H5(C5)</th>
<th>H6(C6)</th>
<th>H8(C8)</th>
<th>H21(H23)</th>
<th>H41(H42)</th>
<th>H61(H62)</th>
<th>H1(N1)</th>
<th>H3(N3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>5.96(89.34)</td>
<td>4.84(4.72)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.12(140.0)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C2</td>
<td>5.71(93.18)</td>
<td>4.59(4.66)</td>
<td>-</td>
<td>5.91(97.82)</td>
<td>7.91(142.4)</td>
<td>-</td>
<td>7.03(8.52)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>G3</td>
<td>5.79(92.23)</td>
<td>4.59(4.69)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.75(137.0)</td>
<td>7.74</td>
<td>-</td>
<td>12.18(147.0)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>5.71(93.50)</td>
<td>4.63(4.40)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.16(137.0)</td>
<td>6.26(8.10)</td>
<td>-</td>
<td>10.49(142.6)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>5.54(93.17)</td>
<td>4.80(4.82)</td>
<td>7.83(154.0)</td>
<td>-</td>
<td>-</td>
<td>7.72(139.3)</td>
<td>-</td>
<td>-</td>
<td>6.94(7.82)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>U6</td>
<td>5.48(93.21)</td>
<td>4.82(4.45)</td>
<td>5.00(103.0)</td>
<td>7.51(140.9)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13.42(161.6)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G7</td>
<td>5.76(92.46)</td>
<td>4.49(4.57)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.61(136.3)</td>
<td>-</td>
<td>-</td>
<td>12.04(146.5)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G8</td>
<td>5.60(93.12)</td>
<td>4.36(4.46)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.16(136.3)</td>
<td>-</td>
<td>-</td>
<td>13.01(148.0)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G9</td>
<td>5.70(91.28)</td>
<td>4.52(4.46)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.52(140.5)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>U10</td>
<td>5.68(91.23)</td>
<td>4.30(4.57)</td>
<td>-</td>
<td>5.48(104.6)</td>
<td>7.61(142.7)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>A11</td>
<td>5.65(91.71)</td>
<td>4.75(4.68)</td>
<td>8.03(155.1)</td>
<td>-</td>
<td>-</td>
<td>8.24(140.9)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>U12</td>
<td>5.68(91.53)</td>
<td>4.37(4.53)</td>
<td>-</td>
<td>5.54(104.6)</td>
<td>7.65(143.0)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C13</td>
<td>5.65(92.84)</td>
<td>4.32(4.38)</td>
<td>-</td>
<td>5.74(98.50)</td>
<td>7.70(142.2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C14</td>
<td>5.69(92.66)</td>
<td>4.40(4.33)</td>
<td>-</td>
<td>5.68(98.52)</td>
<td>7.73(142.2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C15</td>
<td>5.68(92.88)</td>
<td>4.48(4.42)</td>
<td>-</td>
<td>5.71(98.32)</td>
<td>7.69(142.9)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>A16</td>
<td>5.68(91.59)</td>
<td>4.74(4.59)</td>
<td>8.15(155.3)</td>
<td>-</td>
<td>-</td>
<td>8.22(141.7)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C17</td>
<td>5.29(94.14)</td>
<td>4.36(4.37)</td>
<td>-</td>
<td>5.72(98.11)</td>
<td>7.75(141.4)</td>
<td>-</td>
<td>-</td>
<td>7.13(8.54)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C18</td>
<td>5.45(93.86)</td>
<td>4.48(4.55)</td>
<td>-</td>
<td>5.58(97.99)</td>
<td>7.82(141.1)</td>
<td>-</td>
<td>-</td>
<td>7.00(8.29)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>A19</td>
<td>5.00(93.04)</td>
<td>4.50(4.55)</td>
<td>7.29(153.0)</td>
<td>-</td>
<td>-</td>
<td>8.03(139.7)</td>
<td>-</td>
<td>-</td>
<td>6.44(7.85)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>U20</td>
<td>5.43(93.48)</td>
<td>4.48(4.32)</td>
<td>-</td>
<td>4.99(102.7)</td>
<td>7.49(141.1)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14.02(161.6)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>U21</td>
<td>5.48(94.06)</td>
<td>4.17(4.51)</td>
<td>-</td>
<td>5.70(104.8)</td>
<td>7.90(141.3)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11.84(158.2)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C22</td>
<td>5.49(93.43)</td>
<td>4.51(4.57)</td>
<td>-</td>
<td>5.61(97.82)</td>
<td>7.81(141.9)</td>
<td>-</td>
<td>-</td>
<td>6.73(8.29)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G23</td>
<td>5.61(92.49)</td>
<td>4.45(4.60)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.48(136.5)</td>
<td>6.71</td>
<td>-</td>
<td>12.11(146.7)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>A24</td>
<td>5.96(91.28)</td>
<td>4.19(4.32)</td>
<td>7.84(154.9)</td>
<td>-</td>
<td>-</td>
<td>7.92(140.5)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
## TABLE A2: Chemical Shifts of 5’CL (ppm)

<table>
<thead>
<tr>
<th>Base</th>
<th>Absence of Mg$^{2+}$</th>
<th>Presence Mg$^{2+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>G9</td>
<td>12.36(148.7)</td>
<td>12.40(148.2)</td>
</tr>
<tr>
<td>G11</td>
<td>11.67(148.1)</td>
<td>11.80(148.2)</td>
</tr>
<tr>
<td>G12</td>
<td>10.16(145.0)</td>
<td>10.07(144.9)</td>
</tr>
<tr>
<td>U14</td>
<td>13.09(160.4)</td>
<td>13.10(160.1)</td>
</tr>
<tr>
<td>G15</td>
<td>12.33(148.4)</td>
<td>12.50(149.7)</td>
</tr>
<tr>
<td>G16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>U28</td>
<td>13.67(159.8)</td>
<td>13.67(159.7)</td>
</tr>
<tr>
<td>U29</td>
<td>11.47(157.1)</td>
<td>11.54(157.1)</td>
</tr>
<tr>
<td>G31</td>
<td>12.43(148.8)</td>
<td>-</td>
</tr>
<tr>
<td>G40</td>
<td>11.78(148.1)</td>
<td>-</td>
</tr>
<tr>
<td>G41</td>
<td>12.21(148.7)</td>
<td>12.43(149.0)</td>
</tr>
<tr>
<td>U42</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>G46</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>U47</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>U50</td>
<td>10.95(156.6)</td>
<td>11.00(157.0)</td>
</tr>
<tr>
<td>U52</td>
<td>9.68(155.3)</td>
<td>9.66(155.3)</td>
</tr>
<tr>
<td>G53</td>
<td>12.39(148.8)</td>
<td>12.33(148.6)</td>
</tr>
<tr>
<td>G54</td>
<td>12.63(149.3)</td>
<td>12.64(149.3)</td>
</tr>
<tr>
<td>U55</td>
<td>13.38(160.3)</td>
<td>13.40(160.3)</td>
</tr>
<tr>
<td>G61</td>
<td>12.48(149.4)</td>
<td>12.50(149.7)</td>
</tr>
<tr>
<td>U66</td>
<td>10.92(156.3)</td>
<td>11.23(156.5)</td>
</tr>
<tr>
<td>U67</td>
<td>-</td>
<td>13.36(159.6)</td>
</tr>
<tr>
<td>U68</td>
<td>9.56(155.7)</td>
<td>9.49(156.3)</td>
</tr>
<tr>
<td>G69</td>
<td>12.84(149.4)</td>
<td>12.86(149.4)</td>
</tr>
<tr>
<td>U70</td>
<td>13.03(159.9)</td>
<td>13.08(159.7)</td>
</tr>
<tr>
<td>U76</td>
<td>13.71(160.4)</td>
<td>13.73(160.3)</td>
</tr>
<tr>
<td>G77</td>
<td>12.08(148.5)</td>
<td>12.18(148.5)</td>
</tr>
<tr>
<td>U78</td>
<td>13.89(160.4)</td>
<td>13.91(160.2)</td>
</tr>
<tr>
<td>U79</td>
<td>13.50(160.6)</td>
<td>13.51(160.5)</td>
</tr>
<tr>
<td>U80</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
APPENDIX B

PLASMID PREPARATION OF 5'CL

PCR was used to make the 5'CL DNA template for in vitro transcription. The PCR reaction was performed by Dr. Steven Pascal at University of Wisconsin-Madison. The PCR plasmid reaction was purified using a plasmid Giga Prep kit purchased from Qiagen. The purified DNA was stored in ddH₂O overnight at -20°C. The plasmid was linerazed using the following protocol. A 1.3 mL reaction was assembled with 1 mL of plasmid DNA (purified above), 130 µL of NEB buffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9), 13 µL of BSA, 77 µL of ddH₂O, and 80 µL of BsaI restriction enzyme. The reaction was incubated overnight at 37°C. The reaction was diluted to 1:25 and loaded onto a 1% agarose gel to ensure the plasmid was linerazed. To deactivate the enzyme, the reaction was heated for 20 minutes at 65°C. The DNA template was extracted by adding equal volumes of phenol-chloroform. The sample was vortexed and spun down at 4°C at 9,200xg for one minute. The top layer was recovered and put into a clean eppendorf tube. The bottom layer was washed with 100 µL of ddH₂O, vortexed and spun down at 4°C, 9,200xg for one minute. The top layer was recovered and put into a clean eppendorf tube. The DNA was ethanol precipitated with 0.1 volume of 3M sodium acetate (pH 5.2) and two volumes of 100% ethanol. The reaction was stored overnight at -20°C. The overnight reaction was spun down at 4°C, 9,200xg for 30 minutes. The resulting pellet was
washed with 70% ethanol and spun down for an additional 10 minutes. The pellet was resuspended in 500 µL of ddH₂O and stored at -20°C.
APPENDIX C
MC-SYM PROTOCOL FOR STARTING STRUCTURES

MC-Sym is a web-based server to predict RNA tertiary structures. Two thousand structures of the 5'CL were generated using MC-Sym (as described in Chapter 3). The protocol to generate the structures is below.

```plaintext
// ==================== -*-Mode: Mcsym-*- ====================
//
// MC-Sym 4.2 script generated by mcsymize.exe
// version: May 5 2011 11:05:10
// (c) Marc Parisien & Francois Major, University of Montreal
//
// web site FAQ: www.major.iric.ca/MC-Sym/faq.html
// advanced features: commands.html file of your working directory
// Status: Ok.
//
// ==================== Sequence ====================
sequence( r B1
UUAAAACAGCGAUUGGUUACCACAUUGGCUUGCUUGACUGCUACUGUACUGUACUGUACUGUACGCUUGUACGCU
//                       ...((((((((((...))))))))(((((....))))...((((...(((((...)))))...))))..))))))...
//                       AAAAAAAAABBBBBBBBBBBBBBBBBBBBBBCCCCCCCCCCC
DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD  AAAAAAAAA
//                       12345678901234567890123456789012345678901234567890123456789012345678901234567890123
//                       1         2         3         4         5         6         7         8
//
// ==================== Library ====================
//----- Fragment A -----   
// stem 1 -> 9, 75 -> 83
SA = library(
    pdb( "./ypkm9Xgc4c/SA.pdb" )
    #1:#18 <- B1:B9, B75:B83 )
//----- Fragment B -----   
// stem 10 -> 16, 25 -> 31
SLB = library(
    pdb( "./vQJ4pN1gHy/SLB.pdb" )
    #2:#23 <- B10:B31 )
//----- Fragment C -----   
// stem 32 -> 34, 40 -> 42
SLC = library(
    pdb( "./31B0l4ghcQ/SLC.pdb" )
    #1:#11 <- B32:B42 )
//----- Fragment D -----   
// stem 46 -> 57, 61 -> 72
SLD = library(
    pdb( "./UuKmDx5sW1/SLD.pdb" )
    #1:#27 <- B46:B72 )
lnk_01 = library(
    pdb( "MCSYM-DB/ss2/GC/*.pdb.gz" ) #1:#2 <- B9:B10 )
lnk_02 = library(
    pdb( "MCSYM-DB/ss2/CG/*.pdb.gz" ) #1:#2 <- B10:B11 )
lnk_03 = library(
```

pdb( "MCSYM-DB/ss2/GA/*.pdb.gz" ) #1:#2 <- B31:B32 )
lkB_04 = library(
    pdb( "MCSYM-DB/ss2/UG/*.pdb.gz" ) #1:#2 <- B42:B43 )
lkB_05 = library(
    pdb( "MCSYM-DB/ss2/GU/*.pdb.gz" ) #1:#2 <- B43:B44 )
lkB_06 = library(
    pdb( "MCSYM-DB/ss2/UA/*.pdb.gz" ) #1:#2 <- B44:B45 )
lkB_07 = library(
    pdb( "MCSYM-DB/ss2/AG/*.pdb.gz" ) #1:#2 <- B45:B46 )
lkB_08 = library(
    pdb( "MCSYM-DB/ss2/CG/*.pdb.gz" ) #1:#2 <- B72:B73 )
lkB_09 = library(
    pdb( "MCSYM-DB/ss2/GC/*.pdb.gz" ) #1:#2 <- B73:B74 )
// ===================== Backtrack =====================
// assemble the whole structure:
structure = backtrack
(
    SA
    merge( lnk_01 2.0)
    merge( lnk_02 2.0)
    merge( SLB 2.0)
    merge( lnk_03 2.0)
    merge( SLC 2.0)
    merge( lnk_04 2.0)
    merge( lnk_05 2.0)
    merge( lnk_06 2.0)
    merge( lnk_07 2.0)
    merge( SLD 2.0)
    merge( lnk_08 2.0)
    merge( lnk_09 2.0)
)
// ================ Backtrack Restraints ================
clash(
    structure
    1.5 !( pse || lp || hydrogen )
)
backtrack_rst
(
    structure
    width_limit = 25%,
    height_limit = 33%,
    method = probabilistic
)
// ================ Ribose Restraints ================
ribose_rst(
    structure
    method = ccm,
    pucker = C3p_endo,
    threshold = 2.0
)
// ================ Exploration Initialization =========
explore(
    structure
    option(
        model_limit = 2000,
        time_limit = 36h,
        seed = 3210
    )
    rmsd( 2.0 sidechain && !( pse || lp || hydrogen )
    )
    pdb("5pCL" zipped))
Since MC-Sym builds a backbone chain to connect linker regions in the structures (i.e. connects SA to SLB and SLB to SLC), there are small gaps in the resulting pdb files. The gaps in the structures need to be fixed before using XPLOR-NIH. To correct the gaps in the structures, the XPLOR protocol below was used. The input files are the MC-Sym pdb and a hydrogen bond table.

```python
# rebuild protons
protocol.initCoords(erase=True, selection="name H")

# fix nomenclature around Ps
protocol.initCoords(erase=True, selection="name O2P or name O1P")

protocol.addUnknownAtoms()

from repelPotTools import create_RepelPot
repel = create_RepelPot('repel')

from torsionDBPotTools import create_TorsionDBPot
tdb = create_TorsionDBPot('tdb', system='rna')

from xplorPot import XplorPot
hb = XplorPot('HBON')

from noePotTools import create_NOEPot
noeHB = create_NOEPot('noeHB', './hb.tbl')
noeHB.setShowAllRestraints(True)
noeHB.setScale(30)

# import regularize
try:
    # regularize.fixupCovalentGeomIVM(rigidRegions=[
    #     'resid 12:29',
    #     'resid 47:71',
    # ],
    # extraTerms=[repel, tdb, noeHB],
    # verbose=1,
    # )

# except regularize.CovalentViolation:

# pass

from simulationTools import minimizeRefine
minimizeRefine([repel, tdb, noeHB],
                rigidRegions=[
                    'resid 12:29',
                    'resid 47:71',
                ],
                )

from simulationTools import analyze
print analyze([XplorPot(term) for term in "BOND ANGL IMPR".split()] + [repel, noeHB])

protocol.writePDB('new.pdb')
```
VITA

MEGHAN SUZANNE WARDEN
mwarden@odu.edu
Department of Chemistry and Biochemistry
Old Dominion University
Norfolk, VA 23529

EDUCATION

Old Dominion University
PhD, Chemistry
Norfolk, VA
December 2018

Lambuth University
BS, magna cum laude, Chemistry
Jackson, TN
May 2011

PRESENTATIONS


M.S. Warden, M. Tonelli, G. Cornilescu, D. Lui, L. Hopersberger, K. Ponniah, S.M. Pascal "Structure of the RNA Stem Loop that Circularizes the Picornavirus Genome" Oral Presentation delivered at Graduate Achievement Day, Old Dominion University, Norfolk, VA, March, 2017.

PEER REVIEW PUBLICATIONS


