RNA STRUCTURE-BASED ESTABLISHMENT, MAINTENANCE, AND ALTERATION OF
TRANSLATION READING FRAMES:
LESSONS FROM A VIRAL IRES AND A CELLULAR FRAMESHIFTING SIGNAL

By
MARISA DEMAREE CARLSON RUEHLE
B.S., University of Denver, 2007

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Molecular Biology Program
2015
This thesis for the Doctor of Philosophy degree by

Marisa Demaree Carlson Ruehle

has been approved for the

Molecular Biology Program

by

Thomas C. Evans, Chair

Richard E. Davis

Rui Zhao

Aaron M. Johnson

Ruben L. Gonzalez, Jr.

Jeffrey S. Kieft, Advisor

Date: 12/18/2015
Structured RNAs play diverse roles in the cell, and there is an increasing appreciation that structured, *cis*-acting elements found on messenger RNAs can act as regulatory molecules that influence their own translation. Two types of these *cis*-acting signals are internal ribosome entry sites (IRESs) and -1 programmed ribosomal frameshifting (PRF) signals. This thesis integrates detailed studies of each of these types of signals to understand how a translation reading frame can be established, maintained, and reprogrammed using specific RNA structures. Studies of the viral IGR IRESs show that, while the compact architecture of the RNA is important for ribosome recruitment, a conformationally dynamic loop in a pseudoknot domain helps the IRES to drive two essential, non-canonical translocation events within the ribosome’s decoding groove. These data suggest that there is an important balance between stable RNA structures and conformational flexibility during IGR IRES-driven ribosome manipulation. Investigation of the CCR5 -1 PRF element found in mammalian cells shows that its structure and function can be influenced by a microRNA. But, how the microRNA recognizes the -1 PRF structure or promotes formation of an active frameshifting conformation remains unclear. From my preliminary data, I propose that microRNA
binding helps to establish a key pseudoknot interaction that is necessary to elicit a frameshift when encountered by the ribosome. From a detailed dissection of a well-known RNA-driven mechanism to the first forays into an unstudied molecular interaction, my work shows that conformational dynamics are essential to the function of these RNAs, and additionally, supports the hypothesis that structured RNAs can alter the thermodynamic landscape of ribosome function in defined ways to provoke desired responses.

The form and content of this abstract are approved. I recommend its publication.

Approved: Jeffrey S. Kieft
Learn from yesterday, live for today, hope for tomorrow. The important thing is to not stop questioning.

–Albert Einstein
ACKNOWLEDGEMENTS

I first would like to acknowledge the Molecular Biology Program and the Department of Biochemistry and Molecular Genetics for proving me with a phenomenal training environment during my graduate school career. Additionally, I am thankful to have been supported by the Molecular Biology Program’s NIH T32 training grant, and by an American Heart Association Predoctoral Fellowship.

My family has supported me in many fundamental ways from childhood on, and I am blessed with sisters, parents, aunts, uncles, cousins, and grandparents who exemplify the joy and friendship of family bonds. I am grateful for their support of my interests in science and anything else I felt I needed to fully invest my time and physical or emotional energy in, but especially for just being good people who I love spending time with. I thank my parents in particular for financing my college education and otherwise providing me with anything they could—materially and emotionally—to help me succeed. My mom’s fascination with nature and with the rigor and excitement of working in a lab planted the first seeds of my own passion for science. My dad tirelessly strives to improve himself and with undaunted work ethic sets an example of what it means to give your all everyday working towards worthwhile goals. My grandfather is my hero, and his good-natured personality is admirable and infectious. These examples have set a bar in my work and my life that is both challenging and fun to pursue. My husband’s family has been equally significant; Nancy is the mother-in-law everyone wishes they could have, kind and genuinely supportive of anything Alex and I chase. My father-in-law, Ray, lived a life of passion for many diverse (often eccentric) hobbies, and
admirably the line between work and play was a blurred one for him. I am deeply grateful for all of my family’s influence in helping me find my own expression of a life well lived.

The Kieft Lab has been a home-away-from-home for the past six years, and I value the brother-sister relationships that have emerged from it. Each person challenged me in different ways and I learned new things everyday in the lab; it is truly a unique environment and I am humbled to have been a part of this constantly changing but always intelligent and fun group of people. A few people played especially important roles. Megan has a bubbly enthusiasm for life and science that is infectious, and on more than one occasion she gave me the boost I needed to push through especially difficult challenges. I also cherish many memories with Terra and Yumeng. Brian never ceases to make me laugh, and he continues to be a helpful sounding board in both experimental design and life. I cannot thank each of these people—and everyone in the Kieft lab—enough for their friendship.

Of course, the lab would not be what it is without Jeff. Jeff is a scientist to the core— he has a passion for discovery that extends far beyond RNA structure, but in many ways is disguised by a laid-back, affable persona. I think this allows him to be patient as he lets his trainees find their own love for what they are doing, something that is essential for success. Yet, it is evident that Jeff is deeply invested in the success of his people and he takes the responsibility of training seriously. I need to thank him for being patient with me as I learned how to do biochemistry and be a productive part of the lab, not to mention talking me down from the occasional freak-out, and never
letting me off the hook of my own goals. For these things, and the friendship that developed along the way, I am sincerely thankful.

Finally, I extend my deepest gratitude to my husband, Alex. In high school I had this idea that I wanted to be a scientist, but I had no idea of how to pursue it and it seemed out of my league and unattainable. Alex believed I could actually achieve this goal and that I could excel at it, and then he helped me to believe I had it in me, too. It inspires me to see him pursue his own goals and passions so tenaciously. I respect him deeply for it. Traveling this road together makes the challenges and sacrifices worthwhile; I feel very fortunate to have found that kind of friendship at such a young age.
# TABLE OF CONTENTS

## CHAPTER

### I. INTRODUCTION: RNA Structure-Based Ribosome Manipulation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>1</td>
</tr>
<tr>
<td>General Tenets of RNA Structure and Pseudoknots</td>
<td>2</td>
</tr>
<tr>
<td>A Brief Introduction to the Ribosome</td>
<td>5</td>
</tr>
<tr>
<td>RNA Structure-Driven Establishment and Maintenance of a Reading Frame</td>
<td>11</td>
</tr>
<tr>
<td>Canonical, Cap-Dependent Translation Initiation</td>
<td>11</td>
</tr>
<tr>
<td>Translation Elongation</td>
<td>13</td>
</tr>
<tr>
<td>Alternate Initiation Mechanisms and IRES Biology</td>
<td>18</td>
</tr>
<tr>
<td><em>Dicistroviridae</em> IGR IRESs Merge Initiation and Elongation in a Streamlined Mechanism</td>
<td>21</td>
</tr>
<tr>
<td>RNA Structure-Driven Reprogramming of a Reading Frame</td>
<td>25</td>
</tr>
<tr>
<td>An Overview of -1 PRF</td>
<td>26</td>
</tr>
<tr>
<td>How the Translating Ribosome Resolves mRNA Structure</td>
<td>30</td>
</tr>
<tr>
<td>How the Translating Ribosome Maintains the Reading Frame</td>
<td>31</td>
</tr>
<tr>
<td>Mechanistic Models of -1 PRF</td>
<td>36</td>
</tr>
<tr>
<td>The Search for New Eukaryotic -1 PRF Signals Genome-Wide</td>
<td>37</td>
</tr>
<tr>
<td>Regulation of -1 PRF</td>
<td>39</td>
</tr>
<tr>
<td>CCR5 and miR-1224</td>
<td>40</td>
</tr>
</tbody>
</table>

### II. MATERIALS AND METHODS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods for Chapter III</td>
<td>42</td>
</tr>
<tr>
<td>Plasmid Construction and Cloning</td>
<td>42</td>
</tr>
</tbody>
</table>
Radiolabeling RNA, primers, and microRNAs ........................................... 62
Electrophoretic Mobility Shift Assay ........................................................... 62
EDTA Native Gel Analysis ........................................................................... 63
Native RNase T1 Probing .............................................................................. 63
Fluorescence Anisotropy-Stoichiometry of Binding ....................................... 65
Site-Specific UV Crosslinking ...................................................................... 65
Crosslink Mapping ....................................................................................... 66
1D DMS and SHAPE Probing ...................................................................... 67
2D Mutate and Map SHAPE Probing .......................................................... 70

III. AN UNSTRUCTURED REGION OF THE IGR IRES PLAYS MULTIPLE ROLES IN TRANSLATION INITIATION ......................................................... 72

Introduction .................................................................................................. 72

Results .......................................................................................................... 76

Is Loop 3 Important for Translation? .......................................................... 76
Can Loop 3 Mutants Bind to Ribosomes? .................................................... 80
Does Loop 3 Promote Translocation? .......................................................... 83
Is the Correct Reading Frame Maintained? .................................................. 88
What Is the Relative Rate of Tetrapeptide Synthesis in the Sequence Mutants? ................................................................................................. 88
Is tRNA Association with IRES-Ribosome Complexes Altered When Loop 3 Is Mutated? ....................................................................................... 90
Do the Loop 3 Length Mutants Fail Because They Cannot Execute the First Pseudotranslocation or Because tRNA Cannot Be Delivered After the First Pseudotranslocation? ......................................................... 96
Comparison of Results in Lysate and Reconstituted Systems ................. 100
IV. IGR IRES STRUCTURAL ANALYSIS: THE INFLUENCE OF LOOP 3 MUTATIONS AND RIBOSOME BINDING ................................................................. 111

Introduction ........................................................................................................ 111

Results ................................................................................................................ 113

Does Mutation to Loop 3 Alter Its Flexibility and/or the Structure of Domain III? ................................................................. 113

How Does the Δ3 Mutation Alter the Ribosome-Bound Structure of the IRES? ................................................................. 117

Does the IGR IRES Change Structure Upon Binding in Different tRNA Binding Sites? How Does the Translation Machinery Interact with the IRES and What Are the Relevant Conformational Changes? .......... 120

Discussion ........................................................................................................... 121

V. STRUCTURAL ANALYSIS OF THE CCR5-1 PROGRAMMED RIBOSOMAL FRAMESHIFTING SIGNAL AND ITS INTERACTION WITH MICRO RNA-1224 .......... 125

Introduction ........................................................................................................ 125

Results ................................................................................................................ 131

Does miR-1224 Stabilize a Preexisting Conformation of the CCR5 -1 PRF RNA or Induce a New Structure? ......................................................... 131

What is the Stoichiometry of miR-1224 Binding to the CCR5 -1 PRF Element? .......................................................................................................... 135

What Are the Important Regions of CCR5 for miR-1224 Association? ... 137

Where Does miR-1224 Bind to CCR5? ............................................................. 144

What Is the Secondary Structure of the CCR5 -1 PRF Element? Are There Alternate Models That Fit the New Data? ............................................. 146

Discussion ........................................................................................................... 155

Future Directions ................................................................................................. 157
LIST OF TABLES

TABLE

3.1 Loop 3 composition and length in diverse IGR IRESs ..............................................79

3.2 Activity of IGR IRESs in RRL and mutations tested .....................................................80
LIST OF FIGURES

FIGURE

1.1 RNA Structure and Pseudoknots........................................................................................................3
1.2 The Ribosome ........................................................................................................................................8
1.3 The Translation Elongation Cycle .......................................................................................................14
1.4 IGR IRES Domains and Structure ..................................................................................................22
1.5 Frameshift Signals ............................................................................................................................28
1.6 Structural Elements Involved in Reading Frame Maintenance ..................................................33
3.1 IGR IRES Mechanism and Loop 3 .................................................................................................73
3.2 IGR IRES Genomic Location, Conservation, and Structure .....................................................75
3.3 Function of diverse WT and loop 3 mutant IGR IRESs in RRL .................................................77
3.4 Loop 3 Mutant Ribosome Binding Assays .....................................................................................82
3.5 Ribosome Docking, Translocation, and Reading Frame Maintenance ........................................84
3.6 Toeprinting with Hygromycin B .....................................................................................................87
3.7 Schematic Overviews of Experiments Performed in the Reconstituted System .....................89
3.8 Codon- and Factor-Dependent tRNA Binding to IRES-80S Complexes ....................................91
3.9 Identification of Specific Translocation Steps Affected by Loop 3 .............................................93
3.10 Single-Molecule Colocalization Experiment ...............................................................................97
3.11 Single-Molecule Colocalization of IRES and tRNA Within 80S Ribosomes .........................98
4.1 Structure Analysis of Unbound CrPV IGR IRES Loop 3 Mutants .............................................114
4.2 Cocrystallization of domain III Δ3-70S complexes .....................................................................119
4.3 Electron Microscopy of WT CrPV IGR IRES-80S complexes ....................................................122
5.1 Programmed Ribosomal Frameshifting in CCR5 ................................................... 126
5.2 Models of CCR5 and Proposed Interaction with miR-1224................................. 128
5.3 miR-1224 Binds Specifically to the WT CCR5 -1 PRF RNA ................................. 132
5.4 miR-1224 Binding to CCR5 Induces Structural Changes................................. 133
5.5 Stoichiometry of miR-1224 Binding to CCR5.................................................. 136
5.6 Mutants of CCR5 ............................................................................................... 138
5.7 miR-1224 Binding to CCR5 Mutants................................................................. 140
5.8 RNase T1 Probing of CCR5 Mutants................................................................ 142
5.9 Site-Specific UV Crosslink Mapping of 4SU-Containing miR-1224 to WT CCR5 -1 PRF RNA ........................................................... 145
5.10 Chemical Probing of CCR5............................................................................... 148

6.1 A Speculative Model for miR-1224 Binding to CCR5....................................... 172
LIST OF ABBREVIATIONS

(d)dNTP- (di)deoxynucleotide triphosphate
4SU- 4-thio-uridine
Ac-tRNA- aminoacyl-tRNA
CCR5- c-c chemokine receptor 5
cDNA- complimentary DNA
CHX- cycloheximide
cpm- counts per minute
CrPV- Cricket Paralysis Virus
cryoEM- cryogenic electron microscopy
DMS- dimethyl sulfate
DMSO- dimethylsulfoxide
DTT- dithiolthreitol
eEF- eukaryotic elongation factor
EF- elongation factor
eIF- eukaryotic initiation factor
EMSA- electrophoretic mobility shift assay
eRF- eukaryotic release factor
EtBr- ethidium bromide
EtOH- ethanol
fLuc- firefly luciferase
FRET- Forster (fluorescence) resonance energy transfer
GAC- GTPase activating center
GTP- guanosine triphosphate
HCV- Hepatitis C Virus
HDV- Hepatitis Delta Virus ribozyme
HH- Hammerhead ribozyme
HiTRACE- High-throughput robust analysis for capillary electrophoresis
HIV- Human Immunodeficiency Virus
HygroB- hygromycin B
IF- initiation factor
IGR- intergenic region
IRES- internal ribosome entry site
KO- knockout
M²- Mutate and Map
miRNA- microRNA
mRNA- messenger RNA
MS- mass spectrometry
MWCO- molecular weight cut-off
ncRNA- noncoding RNA
NMD- nonsense mediated decay
NMIA-N-methylisatoic anhydride
NMR- nuclear magnetic resonance spectroscopy
OH- hydroxyl
ORF- open reading frame
PAGE- polyacrylamide gel electrophoresis
PCR- polymerase chain reaction
PIC- preinitiation complex
PK- pseudoknot
PRF- programmed ribosomal frameshifting
PSIV- *Plautia Stali* Intestine Virus
PTC- peptidyl transferase center
RBD- ribosome binding domain
RLU- relative light units
rLuc- *Renilla* luciferase
RNA- ribonucleic acid
RNase- ribonuclease
RPM- revolutions per minute
RRL- rabbit reticulocyte lysate
rRNA- ribosomal RNA
RT- reverse transcriptase/transcription
Rz- ribozyme
SHAPE- selective 2’ hydroxyl acylation analyzed by primer extension
sm- single molecule
SRL- sarcin-ricin loop
TAE- Tris Acetic Acid EDTA
TBE- Tris Borate EDTA
TC- ternary complex
tRNA- transfer RNA
TSV- Taura Syndrome Virus
UTR- untranslated region
UV- ultraviolet
V- volts
VpG- viral protein G
W- watts
WT- wildtype
CHAPTER I

INTRODUCTION:

RNA STRUCTURE-BASED RIBOSOME MANIPULATION

Preface

The translation of messenger RNA (mRNA) into proteins is a high fidelity process that is executed by the ribosome. The ribosome is a large macromolecular complex consisting of RNA and proteins that, through an intricate fold and the structure of the subunits, uses dynamic conformational changes to decode mRNAs. A central aspect of translation is the establishment and maintenance of the reading frame throughout this process—functions that are ascribed to the initiation and elongation phases of translation, respectively. Throughout this thesis, I will explore the role that RNA structure and dynamics play in manipulating the ribosome to either establish and maintain an open reading frame (in the case of the IGR IRESs), or purposely alter it (in the case of the CCR5 -1 programmed ribosomal frameshifting PRF signal). Recurring themes in my work are the concepts of structural dynamics of both the ribosome and the RNAs that bind to it, as well as the role that RNA pseudoknots play in these processes.
General Tenets of RNA Structure and Pseudoknots

RNA structure is classified into three different levels, or hierarchies. RNA primary structure is the sequence of RNA nucleotides—adenosine, guanosine, cytidine, and uridine—in an RNA molecule. The sequence is reported in the 5’ to 3’ direction. RNA secondary structure is defined as the double- or single-stranded regions that are the result of direct base pairing interactions, which generally occur in an antiparallel orientation. Tertiary structure consists of long-range interactions between these secondary structure elements. All of these interactions are guided by hydrogen bonding of functional groups on different edges of the nucleotides (Fig. 1.1A) (extensively reviewed in (Leontis and Westhof 2001)) as well as by base-stacking via pi-pi interactions and van der Waals forces between aromatic rings of the nitrogenous bases. Combinations of these generic rules, sugar pucker conformations, and several other parameters lead to an incredible diversity of structures in RNA (reviewed in (Butcher and Pyle 2011)).

A common RNA tertiary structure is the hairpin-type pseudoknot. H-type pseudoknots occur when base pairing between helical strands is not nested; in other words, when two halves of a base pairing region are separated by one half of a different base pairing region (Fig. 1.1B) (Peselis and Serganov 2014; Staple and Butcher 2005). Another way of envisioning this is as a hairpin/stem-loop structure where the nucleotides in the loop base pair with an element outside of the hairpin. This results in a structure with two different helical segments, which are usually coaxially stacked, and—
Figure 1.1 – RNA Structure and Pseudoknots
Figure 1.1 — RNA Structure and Pseudoknots

A. A canonical, Watson-Crick A-U base pair. Numbering in the ribose is indicated on the A nucleotide (left), and wedge/dash bond notation is used to illustrate antiparallel orientation of the base pair. Different interacting edges are indicated by dashed lines: sugar edges are green, Watson-Crick are red, the purine Hoogsteen edge is cyan, and the pyrimidine CH edge is purple. The N-glycosidic bonds are indicated with asterisks (*). The major and minor grooves are indicated, as well as the identity of the nucleotide, A=adenine, U=uridine.

B. Linear structure of an H-type RNA pseudoknot. Below, two different representations of the helical arrangements. Loop 3 interactions with the minor groove of stem 1 are indicated with a dashed arrow.

C. Example of an A-minor motif with a Watson-Crick G-C base pair and adenine (pink) making minor groove interactions along the sugar edges.
theoretically—three different loop structures (Fig. 1.1B). Practically, loop 2 sometimes does not exist due to the quasi-continuous stacking of the two helices. Often when this happens, the loop that connects stem 1b to stem 2b, which should be called loop 3, is named loop 2 instead. This has led to some ambiguity in the literature. Regardless, loop 3 is usually the longest of the three loop structures (Xin et al. 2008; Aalberts and Hodas 2005), and most atomic-resolution structures of pseudoknots show that it docks neatly into the minor groove of stem 1. An important exception to this observation is loop 3 from Pseudoknot I of the IGR IRESs which instead shows a high degree of conformational flexibility (Jan and Sarnow 2002; Pfingsten et al. 2010; 2007). This example will be discussed in depth later. In most RNA pseudoknots, minor groove contacts are mediated by adenosine residues which are capable of making interactions with Watson-Crick pairs, forming hydrogen bonds usually with at least one of the 2’ hydroxyl groups as well as other areas on the sugar edges of the nucleotides (Fig. 1.1C) (Nissen et al. 2001). These interactions are called A-minor interactions, but are not exclusively observed with adenosine. Each of the other three bases have been observed to make similar types of interactions in an A-form RNA minor groove (Xin et al. 2008). These interactions add to the stability of the RNA fold.

A Brief Introduction to the Ribosome

Small, rounded, electron-dense particles were first identified under an electron microscope associating with the endoplasmic reticulum by George Emil Palade in 1954 (PALADE 1955). These “granules from the microsomal fraction” had previously been
studied in isolation by other biochemists of the day (PETERMANN and HAMILTON 1952; PETERMANN et al. 1953; 1954; SLAUTTERBACK 1953; BARNUM and HUSEBY 1948) and were given a variety of names including “microsomes”, “ultramicrosomes”, and “macromolecules”. With the discovery that they contained a large quantity of ribonucleic acid, they were renamed “ribosomes” at the second annual meeting of the Biophysical Society in 1958. Mounting evidence supported an essential role of ribosomes in protein synthesis. Understanding ribosome structure has come a long way from the original images unveiled by Dr. Palade. As David Slautterback noted, “...it was felt that structure might give some clue to function...” (SLAUTTERBACK 1953). Advances in X-ray crystallography and cryo-electron microscopy have provided near-atomic to atomic resolution views of the intricate structure of this massive molecular machine. Here, I will describe several general characteristics and relevant structural features of the ribosome although this is by no means a comprehensive discussion. Inextricably tied to ribosome structure and function are its dynamics; a few of the pertinent dynamic motions necessary during the process of translation are also mentioned here and described in subsequent sections.

The ribosome consists of two subunits: “small” (30S in bacteria, 40S in eukaryote cytoplasm) and “large” (50S, 60S). The small subunit is formed by the 16/18S

---

* This is in contrast to mitochondrial or chloroplastic ribosomes which are of bacterial evolutionary origin (though quite different from bacterial ribosomes as well). When I refer to eukaryotic ribosomes, I mean ribosomes found in the cytoplasm of eukaryotes.  
† Bacterial/eukaryotic
ribosomal RNA (rRNA), and the large subunit is formed primarily by the 23/25/28S\(^\ddagger\) rRNA as well as the 5S and 5.8S rRNAs (5.8S in eukaryotes only). Proteins decorate these RNA components. Proteins associated with the small ribosomal subunit are given the prefix “S”, whereas large subunit proteins are given the prefix “L”. Although eukaryotic ribosomes are much larger in size, the “core” is conserved throughout all kingdoms of life, with kingdom-specific differences manifesting primarily on the outer, solvent exposed surfaces (Melnikov et al. 2012). This conserved core is the location of transfer RNA (tRNA) and mRNA binding and peptidyl transferase activity. Much of what we know about the function of ribosomes has come from studies of bacterial systems, but because of the conservation of function in the conserved core, these lessons are considered to be applicable to eukaryotes as well. Nevertheless, understanding eukaryotic ribosomes is an ongoing focus of many labs worldwide.

The two subunits come together during translation via the formation of intersubunit bridges (Chapman and Noller 1977; Herr and Noller 1979; Herr et al. 1979). tRNAs and the mRNA bind in the intersubunit space. During the course of translation, the intersubunit bridges are reversibly remodeled to facilitate important conformational changes while still allowing the subunits to remain associated with one another (Zhang et al. 2009).

Each subunit possesses structural features worth discussing. The small subunit has been described generically as having a structure similar to a duck, with gross features such as the head, beak, shoulder, body, and feet (Fig. 1.2A)

\(^\ddagger\) bacterial/yeast/higher eukaryotes
Figure 1.2 — The Ribosome
Figure 1.2 — The Ribosome

A. Intersubunit space views of the 30S small subunit (yellow, left, PDB ID: 1J5E) and 50S large subunit (blue, right, PDB ID: 4Z3R). Important features are identified in the figure. The path of mRNA through the decoding groove of the small subunit is shown with a green dashed line. Transfer RNA binding sites are indicated: A= aminoacyl, P= peptidyl, E= exit. Yellow star shows the general location of the peptidyl transferase center. 

B. Side and Front views of the 70S ribosome (PDB ID: 4XEJ). The channel visible from the side view is the location of tRNA binding. The side view is shown such that the A site is closest to the reader. 

C. Cartoon diagram of a ribosome during translation from the front view, which will be used frequently in this thesis to diagram the ribosome and the RNAs bound to it. Colored circles represent different amino acids in the nascent peptide chain.
(Verschoor et al. 1996; Boulik and Hellmann 1978). The groove between the head and the body (the neck) of the subunit defines the decoding groove, which is where mRNA lays down and tRNA anticodon loops interact by base pairing to the mRNA codons. mRNA enters the decoding groove between the beak and the shoulder, and exits between the platform and the back of the head, with an arch-like trajectory (Yusupov et al. 2001; Yusupova and Yusupov 2014; Wilson and Doudna Cate 2012). There are three tRNA binding sites: the aminoacyl or acceptor (A) site, the peptidyl (P) site, and the exit (E) site (Moazed and Noller 1989; 1990). The A site is closest to the mRNA entry channel and the E site is closest to the mRNA exit channel.

The large subunit has two stalk regions, the L7/L12 stalk (called the P stalk in eukaryotes) and the L1 stalk, on opposing sides of the subunit (Ban et al. 2000; Ben-Shem et al. 2011; Boulik and Hellmann 1978), both of which exhibit a high degree of structural dynamics (Bocharov et al. 2004; Fei et al. 2008; Cornish et al. 2009). The L7/L12 stalk is made entirely of ribosomal proteins. It is closest to the A site and is responsible for interacting with many protein factors involved in elongation and termination of protein synthesis (Wahl and Möller 2002). The L1 stalk consists of helices 76-78 of the 23S rRNA and ribosomal protein L1, and is closest to the E site. The L1 stalk interacts with tRNAs to facilitate their movement through and out of the ribosome. The large subunit also contains a “central protuberance” which interacts with the head of the small subunit and harbors the 5S rRNA. The space between the central protuberance and the rest of the large subunit is the location of binding for tRNA aminoacyl acceptor stem ends. The rRNA surrounding the CCA-aminoacyl 3’ ends of
tRNAs in the P and the A sites is responsible for peptidyl transferase catalytic activity (Nissen et al. 2000; Moore and Steitz 2003; Monro 1967; Sonenberg et al. 1975), although the exact mechanism by which this catalysis occurs remains enigmatic.

**RNA Structure-Driven Establishment and Maintenance of a Reading Frame**

Translation is divided into three general phases: initiation, elongation, and termination/recycling. I will focus on the first two of these phases in this thesis. Translation initiation is the process by which ribosomes are recruited to an mRNA and find the proper initiation codon that establishes the reading frame for the protein to be made. Once the initiation step is complete, the ribosome transitions to the elongation phase of translation, which allows amino acids to be linked together in the sequence dictated by the mRNA instructions. The way these processes occur is orchestrated in large part by protein factors, and much of what we know surrounds these protein-based mechanisms. To understand how RNA structure might drive these steps via alternate mechanisms, it is first important to understand the canonical processes.

**Canonical, Cap-Dependent Translation Initiation**

In eukaryotic cells, mRNAs are capped at the 5’ end by a 7-methyl-guanosine nucleotide via a 5’ to 5’ linkage (Adams and Cory 1975). The structure of the cap at the 5’ end recruits the eukaryotic initiation factor (eIF) 4F multi-protein complex to the mRNA (Shatkin 1976). eIF4F consists of the cap-binding protein eIF4E, the scaffolding protein eIF4G, and a helicase protein eIF4A. Binding of eIF4F to the capped 5’ end of
mRNAs allows association of eIF4B and eIF4H. Together, these proteins promote unwinding of RNA structure within the 5’ untranslated region (UTR) (Merrick 1992), which allows the 43S preinitiation complex (PIC) to associate with the mRNA via eIF3-eIF4G interactions (Lamphear et al. 1995). The 43S PIC contains the 40S ribosomal subunit, eIF3, eIF2- Met-tRNA$^{\text{Met}}$-GTP ternary complex, eIF1, eIF1A, and eIF5. 43S binding occurs upstream of the initiation codon that defines the open reading frame, and upon mRNA binding is referred to as the 48S complex. The 48S complex uses eIF1 and 1A to scan down the 5’ UTR and locate the AUG initiation codon (Pestova and Kolupaeva 2002; Passmore et al. 2007; Hinnebusch 2011). Initiation codon identification by base pairing to the anticodon of the initiator tRNA in the 40S P site pauses scanning, and is impacted by the so-called “Kozak” consensus sequence immediately upstream of the AUG (Kozak 1986). Verification of the codon-anticodon interaction in the P site by eIF1A through specific RNA-protein contacts alters ribosome conformation. Specifically, the 40S changes from an “open” (scanning) state to a “closed” state by formation of a latch between the beak and the shoulder of the ribosome (Hussain et al. 2014). These conformational changes are thought to impact downstream eIF-driven steps, and this highlights a single example of the importance of ribosome structural changes in the process of reading frame establishment. Dissociation of all eIFs except 1A is concomitant with eIF5B-GTP and 60S subunit joining (Unbehaun et al. 2004). GTP hydrolysis by eIF5B allows for its own dissociation, followed by that of eIF1A (Fringer et al. 2007). The end result of this process is an 80S ribosome that contains mRNA in its decoding groove with a methionyl-tRNA in the P site decoding the first codon of the open reading frame.
(Hershey et al. 2012). This 80S complex is now ready to begin the elongation phase of translation.

**Translation Elongation**

Unlike translation initiation where millions of years of evolution have lead to divergent strategies between prokaryotes and eukaryotes, the process of translation elongation is remarkably well conserved throughout all kingdoms of life. This conservation refers both to the steps and protein factors involved, as well as the ribosome and tRNA conformational changes that occur. Because most of what we have learned about translation elongation was elucidated from studies of bacterial ribosomes, I will describe this process using the prokaryotic terminology, and point out the eukaryotic nomenclature and any proposed kingdom-specific differences where appropriate.

Translation elongation (Fig 1.3) begins with the delivery of aminoacyl-tRNA to the A site of the ribosome by elongation factor (EF) – thermogenous (Tu) (Lucas-Lenard and Lipmann 1966) (homolog of eukaryotic eEF-1A). EF-Tu binds to the GTPase Activating Center (GAC) of the 50S subunit which consists in part of the sarcin-ricin loop (SRL) of the 23S rRNA, while maintaining interactions with the acceptor stem of the tRNA (Miller 1972; Fischer et al. 2015). The anticodon stem can then query the mRNA codon sitting in the A site of the 30S subunit. Perfect base pairing between the anticodon of the incoming tRNA and the codon leads to the “flipping out” of 16S rRNA residues A1492 and A1493 from their normal orientation in towards the decoding
Figure 1.3—The Translation Elongation Cycle
Figure 1.3—The Translation Elongation Cycle
1) The elongation cycle begins with a ribosome harboring a peptidyl-tRNA in the P site and an empty A site.

2) a. tRNA is delivered by EF-Tu (green oval). This initial binding is referred to as the A/T state, since the anticodon of the tRNA is in the A site of the ribosome. In the process of decoding and proofreading, two adenosine bases make minor groove interactions to verify the codon-anticodon interaction is correct. b. If it is correct, the latch between the beak and shoulder closes, leading to changes in the sarcin-ricin loop in the large subunit (blue loop on the large subunit), GTP hydrolysis, and EF-Tu conformational change that releases the tRNA. c. Torsional strain in the tRNA from the conformational changes described above allows the tRNA to relax and travel into the A site on the large subunit. This is called the A/A classical state.

3) a. Immediately upon entering the A/A state, a peptide bond is formed. b. This allows the tRNAs to sample P/E and A/P hybrid states, the subunits to rotate relative to one another, the L1 stalk to move inwards, in addition to many other important conformational changes within the complex that collectively are referred to as the “Ratcheted State”. Evidence suggests that ribosome complexes at this stage of elongation fluctuate freely between the ratcheted and non-ratcheted states, indicated by the bidirectional arrow.

4) a. The translocase EF-G (red) binds the complexes in 3 and stabilizes the ratcheted state. b. Conformational changes at the sarcin-ricin loop activate GTP hydrolysis on EF-G, promoting movement of the codon-anticodons into the adjacent sites. This is mediated in part by the L1 stalk “pulling” the P/E hybrid tRNA into the E/E classical state.

5) EF-G dissociates from the complex, and the deacylated tRNA is released. At this point, the complex is again considered to be in the non-ratcheted state, and the cycle can recommence. See text for more description and references.
groove to assess the first two positions in the minor groove of the codon-anticodon
minihelix (Carter et al. 2000; Ogle et al. 2001). When the anticodon does not match the
codon, these A-minor interactions cannot be stably formed. Stabilization of the A-minor
contacts promotes a closed conformation wherein the beak and shoulder make close
contacts, triggering conformational changes that distort the tRNA and which translate
allosterically up to the GAC, where EF-Tu binds (Schmeing et al. 2009). Changes in
conformation at the SRL orients the catalytic histidine residue to promote GTP
hydrolysis on EF-Tu (Voorhees et al. 2010). This leads to conformational change in EF-Tu
which releases the acceptor stem of the tRNA, allowing it to migrate into the A site of
the 50S subunit by relaxing torsional strain, and promoting dissociation of EF-Tu. Upon
A-site tRNA accommodation, the tRNAs are in classical binding states (they occupy the
same sites on both the small and large subunits, for example the A site in the small
subunit and the A site in the large subunit, referred to as A/A), ribosomal subunits are in
the non-rotated state, and the head of the small subunit is not swiveled; collectively,
these conformations have been referred to as the “Non-Ratcheted” state.

Once the acceptor stem of the A-site tRNA is oriented properly in the peptidyl
transferase center (PTC) a peptide bond is made between the peptidyl-tRNA in the P site
and the aminoacyl-tRNA in the A site (Trobro and Aqvist 2005). This reaction is catalyzed
by rRNA in the PTC, proceeds spontaneously and irreversibly, and results in the nascent
peptide chain being transferred to the A-site tRNA (Samaha et al. 1995; Rodnina and
Wintermeyer 2003). Immediately upon peptide bond formation, the ribosomal subunits
rotate relative to one another and the tRNAs sample “hybrid states” wherein the
aminoacyl acceptor stem ends that associate with the 50S move into the adjacent tRNA binding site, while the anticodon stem ends remain bound in their original sites on the small subunit (Dorner et al. 2006; Agirrezabala et al. 2008). In this state, the L1 stalk moves inward to contact the elbow of the P/E hybrid tRNA (Fei et al. 2008). Additionally, the head of the small subunit swivels, resulting in a slight reorientation of the small subunit tRNA binding sites (Ratje et al. 2010). These conformational changes appear to be reversible until EF-G associates with the complex (Chen et al. 2011a), and are referred to as the “Ratcheted” state.

Sampling of these conformational changes are thought to promote EF-G (eukaryotic eEF-2 homolog) binding to the ribosome (Frank and Agrawal 2001; Chen et al. 2013a). EF-G and EF-Tu share structural similarity through the majority of the protein where they interact with the GAC (Nissen et al. 1995). EF-G, however, also contains a long extended domain 4, which is thought to mimic the anticodon stem loop of a tRNA (Agrawal et al. 1998). Association of EF-G permits domain 4 to interact with the codon-anticodon minihelix of the A/P hybrid tRNA (Brilot et al. 2013; Chen et al. 2013b). GTP hydrolysis causes domain 4 to translocate the tRNAs, moving the codon-anticodon interactions in the A and the P sites into the P and the E sites, respectively (Zhou et al. 2013; Ramrath et al. 2013). Movement of the L1 stalk outward helps to pull the P/E tRNA fully into the E site (Fei et al. 2008; Trabuco et al. 2010). This resolves the hybrid states of tRNAs, moving them back into classical states of tRNA binding. These movements are thought to be concomitant with un-swiveling of the small subunit head and reversal of rotation of the ribosomal subunits, both of which facilitate the tRNA
translocation process. The end result of this step is a 70S ribosome with deacylated tRNA bound in the E site, peptidyl tRNA in the P site, and an empty A site with the next codon waiting to be decoded. It is worth mentioning as a foundation for future discussions that EF-G can efficiently translocate an anticodon stem loop (in the absence of the rest of the tRNA structure) from the A site to the P site as long as a full tRNA exists in the P site (Joseph and Noller 1998).

There are different ideas about when deacylated tRNAs separate from their mRNA codons and dissociate from the E site of the ribosome (Yusupov et al. 2001). Some evidence suggests that early rounds of translocation keep E-site tRNA until the correct A-site tRNA binds, but at later rounds the E-site tRNA is ejected prior to A-site tRNA delivery (Chen et al. 2011b). Regardless, at some point tRNA in the E site leaves, guided by the L1 stalk, and this permits another round of tRNA delivery and translocation to follow.

**Alternate Initiation Mechanisms and IRES Biology**

As mentioned above, translation elongation is well conserved across kingdoms of life, but initiation is not. The canonical, cap-dependent initiation mechanism is complicated; it requires many different proteins, a specific mRNA modification, and temporal execution of steps to be effective. Not surprisingly, viruses as well as cells have evolved variations on the theme of translation initiation (reviewed in (Plank and Kieft 2012)). Generally, this provides a means by which specific messages can circumvent global translational regulation mechanisms executed within the cell, permitting
continued or enhanced translation during times of otherwise suppressed protein synthesis. Initially, these alternate mechanisms were identified in picornaviruses and were shown to be both cap- and 5’ end- independent, leading to the name “Internal Ribosome Entry Site” (Pelletier and Sonenberg 1988; Jang et al. 1988). Subsequently, some cellular messages were also characterized to contain IRESs (Komar and Hatzoglou 2011). However, more recently the field has begun to appreciate that not all translation initiation mechanisms fall cleanly into the cap-dependent or IRES-dependent classes. Rather, it appears that there may be a broad spectrum of mechanisms that exhibit varying initiation factor dependencies, trans-acting factors, proximity to 5’ ends, and even RNA modifications. An intriguing example of this is the translation initiator of short 5’ UTRs (TISU) element, which directs the cap-dependent, but scanning-independent translation of mRNAs harboring very short 5’ UTR lengths (average of 12 nucleotides) (Elfakess and Dikstein 2008; Elfakess et al. 2011). The existence of these “mix-and-match” set of mechanisms provides evidence that the common denominator in this process, the ribosome, is the key to unlocking translation.

Viral IRESs provide a unique opportunity to observe how ribosomes can be “unlocked” in highly streamlined ways. Although related viruses use similar IRES elements, there is surprising diversity in the RNA sequences and structures that can elicit viral IRES-driven translation initiation (Plank and Kieft 2012). Reflecting this diversity, viral IRESs can be classified into four groups based on the protein factors they use to initiate translation (Filbin and Kieft 2009). Group 4 IRESs most closely reflect canonical, cap-dependent initiation, using many eIFs, so-called “IRES trans-acting
factors” (ITAFs), and scanning to locate the proper start codon. Examples from this
group are Poliovirus and Hepatitis A. Group 3 IRESs also use many eIFs and ITAFs, but
are placed directly at the initiation codon, so scanning does not occur. This is the case in
foot-and-mouth disease virus and encephalomyocarditis virus. The Hepatitis C (HCV)
and HCV-like viruses characterize group 2 IRESs. Group 2 IRESs interact directly with the
ribosome but only require a very limited set of initiation factors to deliver tRNA and
perform other functions (although unpublished investigations from our lab suggest
these IRES may use a wider set of canonical factors in the context of infection, (Jaafar et
al. unpublished)). Finally, group 1 IRESs bind directly to the ribosome and require no
initiation factors at all. The only known examples of group 1 IRESs are the Intergenic
Region (IGR) IRESs from the *Dicistroviridae* family of viruses. It is intriguing to note that
the decreasing dependence on initiation factors moving from group 4 to group 1 IRESs
correlates with an increased amount of structured RNA forming the IRES itself. Further,
an increasing body of evidence suggests that despite the differences between group 1
and group 2 IRES structures, they both contact the ribosome in the same important
places which allows for the process of translation to begin (Filbin et al. 2013; Quade et
al. 2015; Fernández et al. 2014). By studying these examples whereby structured RNAs
replace the role of multiple proteins, we reveal the basic important steps or structures
in “unlocking the ribosome” to begin protein synthesis. The first part of this thesis will
focus entirely on the RNA structure-dependent process that the IGR IRESs use to initiate
translation by establishing the reading frame in the A site and maintaining that reading
frame through non-canonical translocation events on the ribosome.
Dicistroviridae IGR IRESs Merge Initiation and Elongation in a Streamlined Mechanism

The *Dicistroviridae* family of viruses use single-stranded, positive-sense RNA genomes to infect a variety of arthropod hosts. The species they infect impose significant ecological and economical implications, contributing to phenomena such as colony collapse disorder of honeybees, decaying shrimp populations, and other problems (Cox-Foster et al. 2007; Granberg et al. 2013; Overstreet et al. 1997; Audelo-del-Valle et al. 2003). The name “Dicistroviridae” is derived from the genomic organization of the virus which contains two open reading frames, one encoding the non-structural viral proteins and the other encoding the structural viral proteins (Isawa et al. 1998). Both of these open reading frames are translated by ribosomes that are initiated using IRESs (Wilson et al. 2000b; Sasaki and Nakashima 1999), but the downstream, intergenic region (IGR) IRES has garnered great interest from the translation field due to its ability to initiate translation independently of any initiation factors, as described above.

A wealth of biochemical and structural information has revealed how this structured RNA can bind directly to ribosomes and promote translation initiation. Structurally, the IRES is characterized by three domains, each consisting of an RNA pseudoknot (Fig. 1.4) (Kanamori and Nakashima 2001; Jan and Sarnow 2002). Domains I and II, containing pseudoknots II and III, fold into a compact element (Costantino and Kieft 2005) from which two stem-loop (SL) structures emanate (Fig. 1.4B), SL IV and V (Pfingsten et al. 2006). These two loops contain sequences that are required for 40S
Figure 1.4—IGR IRES Domains and Structure

A. Secondary structure of the *Plautia Stali* Intestine Virus (PSIV) IGR IRES. Domains I and II (dI and dII) form the ribosome binding domain (RBD). Stem loops required for specific interaction with the 40S ribosomal subunit are shown with blue arches. The green box indicates the non-AUG start codon immediately downstream of Pseudoknot (PK) I which is contained within Domain III (dIII).

B. X-ray crystal structure of the PSIV IGR IRES RBD (PDB ID: 2IL9).

C. X-ray crystal structure of domain III from the Cricket Paralysis Virus IGR IRES bound in the P site of a 70S ribosome (PDB ID: 3PYU). Inset shows the interface view of the 30S subunit from that structure with domain III bound.
subunit binding via ribosomal proteins eS25 and uS7 (Jan and Sarnow 2002; Spahn et al. 2004b; Costantino and Kieft 2005; Nishiyama et al. 2007; Landry et al. 2009). Additionally, this element uses a bulge between two stems (L1.1) to effect recruitment of the 60S subunit (Pfingsten et al. 2010), predominately via the L1-stalk (Spahn et al. 2004b; Schüler et al. 2006; Fernández et al. 2014), suggesting this region may mimic portions of tRNAs that L1 usually binds. The complex architecture of this often-called Ribosome Binding Domain (RBD) is thought to provide the structural framework to orient the stem-loops and the bulge in the proper positions such that they can interact with these specific parts of the ribosome (Pfingsten and Kieft 2008). Apparent equilibrium dissociation constant ($K_d$) measurements of the IGR IRESs for the 40S subunit show approximately 2nM binding affinity (Costantino and Kieft 2005).

Domain III of the IGR IRES (Fig. 1.4C) is entirely comprised of pseudoknot I, and sometimes these names are used interchangeably. Binding studies show that this domain is dispensable for ribosome association (Jan and Sarnow 2002; Costantino and Kieft 2005), however, mutations that prevent stem 2 of the pseudoknot from forming are deleterious to IRES function (Wilson et al. 2000a; Jan and Sarnow 2002). Thus, ribosome binding is necessary but not sufficient for IRES activity. Further studies showed that stem 2 of this pseudoknot mimics the interaction of a tRNA anticodon loop base paired to an mRNA codon (Costantino et al. 2008), and that this pseudoknot binds in the decoding groove of the ribosome (Zhu et al. 2011). The first amino acid of the protein to be made is dictated by the triplet codon immediately downstream of this pseudoknot (Fig. 1.4A, green box) (Wilson et al. 2000b). Thus, pseudoknot I formation establishes
the reading frame by presenting the ribosome with an artificial codon-anticodon interaction. This led to the hypothesis that the IRES operates by putting the ribosome into an elongation-competent state, wherein tRNAs are already decoding the mRNA (mimicry of domain III), and this might allow other tRNAs to bind to this “elongation competent” complex.

Early IGR IRES literature reports that upon ribosome binding, domain III is positioned in the P site (Wilson et al. 2000a; Pestova et al. 2004). However, recent literature and reinterpretation of those original studies have led the field to agree that domain III first binds in the A site (Zhu et al. 2011; Koh et al. 2014; Fernández et al. 2014; Muhs et al. 2015). This is important from a mechanistic point of view because it means the first step of IGR IRES initiation is translocation (without a peptide bond being formed, “pseudotranslocation”) of domain III into the P site before the first codon of the open reading frame is placed in the A site. Presumably this pseudotranslocation is executed by eEF2, although direct experimental evidence for this is lacking. Once the first codon is in the A site, eEF1A delivers the cognate tRNA to the IRES-ribosome complex (Yamamoto et al. 2007); eEF2 greatly enhances this step. In the absence of a peptide bond, the ribosome is recognized by eEF2 and translocated again (Pestova et al. 2004; Pestova and Hellen 2003). At this point, domain III is in the E site of the ribosome, there is an aminoacyl-tRNA in the P site, and the second codon of the protein to be made is in the A site. Subsequent delivery of tRNA to this complex is now reminiscent of a canonical elongation cycle since a peptide bond is formed between the P and A site amino acids, and the normal elongation cycle is thought to ensue.
The above mechanism of IGR IRES initiation reveals a remarkable use of RNA structure encompassing less than 200 nucleotides to functionally replace over 12 multi-protein complexes in the process of translation initiation. The IRES uses specific structured domains to first bind the ribosome and then co-opt elongation factors to act on the complex, bringing in the tRNAs needed to make the structural proteins for the virus. This is accomplished in large part through various levels of tRNA mimicry: 1) The IGR IRES binds in the decoding groove mostly in the E site, where essential contacts with large subunit rRNA for translocation are known to be needed. 2) Domain III structurally mimics an anticodon stem loop, thus presenting a suitable substrate for eEF2 recognition. However, although we understand the structure of the RNA and how it binds the ribosome, ribosome binding is not sufficient to initiate translation. To understand how this RNA structure-driven process works, we need to know how the RNA promotes movement through each of the tRNA binding sites, and the corresponding ribosomal structural changes. Clearly, domain III and the tRNA mimicry it exhibits play an important role in this process. Chapters III and IV of this thesis address the question of how a specific part of domain III controls pseudotranslocation of the IRES within the decoding groove, influencing the ribosome to use and maintain the viral open reading frame.

**RNA Structure-Driven Reprogramming of a Reading Frame**

Clearly, establishing a translation reading frame is an important and highly regulated process on cellular messages and viral RNAs alike. Once established, the
intricate Brownian motions of the ribosome, influenced by interactions with other components of the translational machinery, ensure that the reading frame is maintained throughout the elongation cycle. These multiple levels of control ensure that the ribosome steps down a message by exactly 3 nucleotides during each elongation cycle, maintaining codon-anticodon interactions as it goes, and ultimately producing a single protein of a specified amino acid sequence from that mRNA. But, what happens if this process is disrupted? Presumably, loss of reading frame maintenance or other forms of translational recoding in general could be deleterious. However, a wealth of evidence supports that such recoding events can not only be advantageous but also that some messages program the ability to alter their reading frames at specific locations into the sequence of the mRNA itself. This provides a means by which to regulate protein synthesis during the elongation phase rather than during initiation, which is more common. Although there are several forms of translational recoding known, including stop codon suppression, hopping, +1 frameshifting, -2 frameshifting, selenocysteine incorporation, and pryolysine incorporation, I will focus on -1 Programmed Ribosomal Frameshifting (-1 PRF) in this thesis.

**An Overview of -1 PRF**

A *cis*-acting signal that stimulated -1 frameshifting was first discovered in Rous Sarcoma Virus at the viral RNA’s *gag-pol* junction (Jacks et al. 1988). This set the stage for subsequent widespread discovery of -1 PRF signals in related retroviruses and other viruses as well. As a result, much of our knowledge comes from studying these viral
examples. In these cases, -1 PRF regulates the production of a protein with an alternate C-terminal amino acid sequence (Fig. 1.5B). These C-terminal chimeric proteins have functional significance for the virus. Frameshifting from these -1 PRF elements displays varied efficiencies, and the efficiency defines the molar ratios of each protein being made. Alteration of frameshifting efficiency (and therefore ratios of each protein) downward was shown to inhibit virus propagation and infectivity to the same degree in a study of HIV-1 (Dulude et al. 2006), making frameshifting signals an attractive target for antiviral therapeutics (Dinman 2012).

In eukaryotes and eukaryotic viruses, -1 PRF is thought to be elicited by a bipartite signal in the mRNA open reading frame consisting of a slippery heptanucleotide sequence (generically, X XXY YYZ§) followed by an RNA pseudoknot (Fig. 1.5A) (Jacks et al. 1988; Dam et al. 1990). These elements are separated by a short linker sequence less than 12 nucleotides in length. In bacteria, a stable hairpin can replace the RNA pseudoknot, but these signals also usually contain an internal Shine-Dalgarno-like sequence (Larsen et al. 1994). Through a mechanism that remains mysterious, these signals promote ribosome pausing and tRNA “slippage”, that is, breaking of codon-anticodon interactions within the decoding groove and reestablishing them by sliding backwards (5’) by one nucleotide. Once reestablished, the new codon-anticodon interactions allow the ribosome to resume translation in the -1 frame.

§ Where X is any nucleotide, Y is all Us or As, and Z is A, U, or C. Spacing indicates the 0 reading frame.
Figure 1.5 — Frameshift Signals
Figure 1.5 — Frameshift Signals

A. General architecture of a frameshift signal. B. When the ribosome encounters a viral -1 PRF signal it can either continue to make the normal protein, or slip back by one nucleotide and make a fusion protein. C. When the ribosome encounters a eukaryotic -1 PRF signal it can either continue to make the normal protein, or slip back by one nucleotide which directs the ribosome to a premature termination codon and triggers mRNA decay through the Nonsense Mediated Decay (NMD) pathway.
How the Translating Ribosome Resolves mRNA Structure

To understand how RNA structure might drive a programmed frameshifting event, it is important to understand how structural elements in mRNAs are normally encountered and unwound by the ribosome. In order for codons to be decoded by tRNAs, they must enter the decoding groove in a single-stranded state. The ribosome possesses intrinsic helicase activity to deal with secondary structure in the mRNA (Takyar et al. 2005). This is notable given that many other known helicase-dependent cellular processes use independent proteins to unwind nucleic acid hybrids. Instead, ribosomal proteins uS3, uS4, and uS5, which collectively form a ring and line the interior of the mRNA entry channel, are thought to be responsible for this critical activity on the small subunit of the ribosome (Kurkuoglu et al. 2008). The internal diameter of the entry channel is less than that of an RNA A-form helix, necessitating structure unwinding before entering the channel. Rather than being an ATP-dependent process as is common with dedicated helicases, uS3, uS4, and uS5 are thought to bias the thermal fluctuations of helical segments towards the open/unwound state as well as couple the energetics of conformational changes associated with translocation to the mechanical separation of mRNA secondary structure (Qu et al. 2011; Kim et al. 2014). Together, these mechanisms allow a basal level of protein synthesis in the cell.

Instances where the ribosome pauses due to encountering structured elements, as during -1 PRF, are therefore thought to result from the structure possessing extraordinary physical characteristics. However, thermal and mechanical stability of RNA structures, and even the duration of ribosome pausing, are poor indicators of
frameshifting efficiency (Tu et al. 1992; Kontos et al. 2001; Ritchie et al. 2012). A systematic analysis of NMR-derived structures showed that a bent conformation of the pseudoknot positively correlates to frameshifting function (Kang and Tinoco 1997; Chen et al. 1996; Shen and Tinoco 1995), whereas another study shows correlation to the ability of the pseudoknot to sample alternate structures (Ritchie et al. 2012).

Furthermore, several groups have shown an important role for RNA triplex structures within the pseudoknot (Dinman et al. 2002; Nixon and Giedroc 2000; Su et al. 1999; Chen et al. 2009). All of these data implicate ribosome-specific interactions in eliciting frameshifting from -1 PRF signals. How these characteristics work together within the structure, how these structures interact at the mRNA entry channel, and the mechanism by which this leads to tRNA repositioning over a slippery site are largely unanswered questions.

**How the Translating Ribosome Maintains the Reading Frame**

Accurate movement of intact codon-anticodon helices through the decoding groove at exactly 3 nucleotide intervals is credited to the establishment of defined steric boundaries between each tRNA binding site on the small subunit and their remodeling during translocation. The beginning and the end of the decoding groove are bounded by the mRNA entry channel and mRNA exit channel. The mRNA entry channel transitions between open and closed conformations upon tRNA delivery to the A site, as discussed previously. Similarly, the mRNA exit channel expands and contracts during different states of tRNA binding and movement (Frank and Agrawal 2000). Both the entry and exit
channel use small subunit rRNA nucleotides to intercalate between mRNA bases up and
downstream of mRNA-tRNA interactions, acting as anchor points along the mRNA
trajectory (Fig. 1.6A) (Zhou et al. 2013).

To define the boundaries between the A and the P site and between the P and
the E site, more elaborate mechanisms are used. After peptide bond formation but
before translocation, the head of the ribosome undergoes a counter clockwise (from the
front view) swiveling movement about the neck of the small subunit. This motion opens
a 16S rRNA “gate” between the P and the E sites formed by nucleotides G1338-U1341**
on the head and A790 on the platform (Fig. 1.6B) (Schuwirth et al. 2005). The
nucleotides on the head make minor groove interactions with the P-site tRNA anticodon
stem. In the non-swiveled state, the distance between the gate elements is about 12Å—
too narrow for the A-form anticodon helix of P-site tRNA to pass through. Upon head
swiveling, that distance increases to ~23Å (Yamamoto et al. 2014), and the motion is
thought to “pull” tRNAs via the minor groove interactions into an intermediate position
on the small subunit. Subsequently, during translocation, tRNAs complete their
movements into the adjacent binding sites, somehow breaking the minor groove
contacts with G1338-U1341, and the head rapidly returns to its non-swiveled position
which reestablishes the P-E gate (Frank et al. 2007; Spahn et al. 2004a; Taylor et al.
2007). The ribosome has often been referred to as a “Brownian ratchet”; several groups
propose that the P-E gate acts as the “pawl”, coupling subunit rotation and head swivel

** E. coli numbering
Figure 1.6 — Structural Elements Involved in Reading Frame Maintenance
Figure 1.6 — Structural Elements Involved in Reading Frame Maintenance
A. mRNA bound to the 70S ribosome in an intermediate stage of translocation. rRNA residues (yellow) intercalate bases in the mRNA (magenta) within the entry and exit channels. PDB ID: 4V9L. The numbering of the mRNA is shown where “+1” refers to the first nucleotide in the P site. B. The P-E gate forms between rRNA nucleotides on the head and body of the small subunit. In the non-ratcheted state, the gate is closed and the distance is too small for a tRNA to pass through (top panel, PDB ID: 4V5D). In the ratcheted state when the head is swiveled, the gate opens (bottom panel, PBD ID: 4V9L). C. Ribosomal protein uS13 (orange spheres) extends a tail between the A- and the P-site tRNAs (blue). PDB ID: 4W29.
with tRNA movement and preventing the tRNA “teeth” from moving backwards after translocation is complete (Dunkle and Cate 2010).

While the P and the E sites are separated by the rRNA gate described above, the A and the P sites appear to be defined by a different means. Specifically, the mRNA path as it threads from A to P exhibits a distinct kink between the codons which is induced by rRNA underneath the mRNA (Yusupova et al. 2001). Additionally, ribosomal protein uS13 on the head has a flexible C-terminal tail that projects into the space between A- and P- site tRNAs (Fig. 1.6C). Specific interactions of amino acids in the tail are thought to be important to establish A and P site tRNA binding as well as play an important role during translocation (Cukras et al. 2003). Given uS13’s additional role as a structural element in intersubunit bridge 1a, it also has the ability to directly couple the subunit rotation motions during translocation to tRNA movement (Cukras and Green 2005). More mechanistic information about how this impacts reading frame maintenance is needed to better understand this process.

From the discussion above, it should be clear that the ribosome has evolved several “built-in” features to maintain the proper reading frame during elongation, safeguarding against what could otherwise be a highly error-prone process. How -1 PRF elements manipulate these features to allow reading frame alteration is still unknown. Interesting recent work on other types of translational recoding point to these same ribosomal features, supporting the idea that these varied recoding strategies might mechanistically converge in some way.
**Mechanistic Models of -1 PRF**

Several groups have worked to identify the mechanism by which -1 PRF signals interact with the ribosome to induce tRNA slipping and repositioning in the -1 frame. Currently, five different models attempt to explain this phenomenon, each describing tRNA slippage at a different point during the elongation cycle. In the “9Å model” (Plant et al. 2003), frameshifting is thought to occur upon delivery of tRNA to the ribosome when the third codon of the slippery site is in the A site. In the “Simultaneous Slippage model” (Jacks et al. 1988), tRNA slippage occurs after tRNA decoding of the third codon of the slippery sequence, but notably, before a peptide bond has been made. The “Dynamic model” (Weiss et al. 1989), on the other hand, suggests tRNA delivery and peptide bond formation occur normally, but frameshifting occurs during subunit rotation as the complex samples the hybrid or ratcheted state. The “Mechanical model” (Namy et al. 2006; Moran et al. 2008) posits that EF-G/eEF2 is required to reposition tRNAs during formation of the post-translocation state. Finally, an as-yet unnamed model proposes that frameshifting occurs during tRNA delivery to decode the third codon of the slippery site, but in a state wherein tRNAs in the E and P sites sample non-canonical states of binding (Qin et al. 2014). Although the experiments performed to develop each of these models provides a fascinating foray into ribosome biochemistry, a thorough description of the data that support these models is outside the scope of what I hope to portray in this thesis, which is simply that mechanistic dissection of this phenomenon is still an ongoing and debated effort. I will note that the most detailed mechanistic information in the field has been gathered from illuminating studies of the
Infectious Bronchitis Virus (IBV) -1 PRF signal that regulates expression of its 1a/1b gene (Namy et al. 2006), as well as of the bacterial dnaX -1 PRF signal that promotes expression of the \( \gamma \) subunit of DNA polymerase III (Kim et al. 2014; Qin et al. 2014; Chen et al. 2014).

The recent application of single molecule force and fluorescence experiments to the question of programmed frameshifting has rapidly moved the field towards detailed mechanistic answers (Kim et al. 2014; Qin et al. 2014; Chen et al. 2014). Still, ambiguities exist, and the apparent conflict of some studies underscores the importance of using multiple approaches and being cautious with interpretation. Nevertheless, it is conceivable that there are multiple mechanistic pathways to achieve -1 PRF (Dinman 2012). In the parlance of energy landscapes, different PRF signals could present different energetic obstacles along the canonical path taken by ribosomes to accept and move tRNAs from one site to the next; different -1 PRF signals could conceivably re-route the ribosome to circumvent these barriers in different ways. Despite these multiple possibilities, specific ribosome features will likely play central roles in each case, particularly the small subunit head given its importance in movement of codon-anticodon interactions.

**The Search for New Eukaryotic -1 PRF Signals Genome-Wide**

Like many molecular mechanisms originally discovered in viruses, it was hypothesized that frameshifting might also be a phenomenon used in cellular life as well. With advances in sequencing technologies and computational power, searching
existing databases for these signals became a reasonable possibility beginning in the
1990s and continuing after the turn of the century. Mining the data was accomplished in
two ways. First, genomes were searched for overlapping open reading frames (Moszer
et al. 1995; Moszer 1998; Médigue et al. 1993). This approach makes the assumption
that the purpose of frameshifting in cells is the same as in viruses—to condense as much
coding sequence as possible into a limited amount of genetic material. However, cells
have evolved to accommodate much larger genomes compared to their viral invaders,
therefore obviating the need for strong selective pressures to constrain coding
sequences and maintain overlapping reading frames. Consistent with this idea, these
searches were not fruitful (Baranov et al. 2002).

The second approach eliminated this assumption and was based on the
sequence motifs defined by viral -1 PRF signals (Hammell et al. 1999; Jacobs et al. 2007).
Specifically, searches were performed in several eukaryotic species’ genomes from yeast
to humans for the motif: heptanucleotide slippery site- short spacer- predicted RNA
pseudoknot. The results of these genome-wide searches identified a significant number
of motif hits over random in each species, found -1 PRF signals in homologous genes
from multiple species, and correlated known disease alleles with predicted pseudoknot
disruption. Together, this suggests that frameshifting is a widespread and important
phenomenon in eukaryotic life. Intriguingly, most of the motif hits from this screen
showed that the frameshift event directs the ribosome to a premature stop codon, and
it was later verified that these messages are targeted for degradation by the nonsense
mediated decay pathway (Fig. 1.5C) (Plant et al. 2004; Belew et al. 2014). Thus, the
function of -1 PRF in eukaryotes—unlike in viruses—appears to be to regulate mRNA levels within the cell (Dinman 2006).

**Regulation of -1 PRF**

In viruses, the efficiency of -1 PRF dictates precise molar ratios of proteins needed by the virus during its infection and life cycle; in general, this does not require additional layers of regulatory control. In cells, however, -1 PRF stimulates mRNA decay providing a means to fine tune gene expression co-translationally. It makes sense that this would be subject to regulation depending on changing cellular conditions. Furthermore, because -1 PRF signals were identified in many genes of varied functions, globally up- or down-regulating frameshifting would not provide the sensitivity needed to elicit gene-specific responses. The idea that specific regulatory molecules may exist to control specific -1 PRF signals arose from these observations. Proteins are obvious candidates for such a molecule. The literature is rife with examples of sequence- and structure- specific RNA binding proteins. However, only one instance of a trans-acting protein required for frameshifting has been reported, and this is a viral protein whose binding functionally replaces RNA structure as the energetic barrier to the ribosome (Li et al. 2014). Noncoding (nc) RNAs are an attractive alternate. ncRNAs have been shown to play regulatory roles in a wide variety of cellular processes (Cech and Steitz 2014), and they inherently have sequence specificity through the ability to hybridize to their target RNAs. Increasing evidence supports that some ncRNAs have defined secondary and possibly tertiary structures, but how that structure relates to their function is
unclear. Furthermore, apart from the well-documented case of microRNAs which interact with their targets via direct base pairing, our understanding of how ncRNAs recognize the structure of their specific binding partners is limited. These will likely be important avenues of investigation as our knowledge of the functions of ncRNAs in the cell expands.

**CCR5 and miR-1224**

One of the validated motif hits from the genome-wide screen described above is the chemokine receptor, CCR5. CCR5 is part of the C-C family of chemokine receptors which are 7-transmembrane G-protein coupled receptors (Alkhatib 2009). It has several known chemokine ligands that are normally used to modulate immune responses in an organism (Wu et al. 1997; Samson et al. 1996), but overexpression of CCR5 protein has been linked to a number of types of cancer and as such is the target of several anti-cancer therapeutics currently in clinical trials (González-Martín et al. 2012; Mañes et al. 2003). Furthermore, CCR5 is a major co-receptor used by the HIV-1 virus during cell entry (Wu et al. 1996; Westby and van der Ryst 2010). Clearly, understanding CCR5 expression and regulation will shed light on an important regulator in human health and disease.

The -1 PRF signal in the CCR5 mRNA sits at nucleotide 407, contains the slippery site sequence U UUA AAA, and is predicted to have an unusually large pseudoknot structure 96 nucleotides long. Validation of -1 PRF from this signal was performed by mass spectrometry of the alternate, truncated form of CCR5, mining existing ribosome
profiling data, as well as conducting luciferase-based reporter assays (Belew et al. 2014). To investigate the possibility that the CCR5 -1 PRF element could be regulated by a ncRNA, microRNA (miRNA, miR) target prediction software was used and identified three different miRNAs with base pairing potential to the 103 nucleotide CCR5 -1 PRF element: miR-711, miR-141, and miR-1224. The effect of these miRNAs was determined using a dual luciferase translation-based frameshifting assay in hamster, monkey, and human cell lines. Only miR-1224 stimulated CCR5 -1 PRF activity in all three cell types. Affinity-based binding assays showed that miR-1224 associates specifically with the CCR5 -1 PRF element in HeLa cells. These studies provided the first compelling evidence that miRNAs (a class of ncRNAs) could regulate frameshifting in eukaryotic cells, and opened the possibility that this may be a widespread means of regulation. Yet, how miR-1224 recognizes and interacts with the CCR5 mRNA to elicit -1 PRF remains mysterious. Chapter V of this thesis is directed at unraveling details of the interaction between these two molecules, providing further insight into the way structured RNAs can manipulate the translation machinery.
CHAPTER II
MATERIALS AND METHODS

Methods for Chapter III

Plasmid Construction and Cloning

The constructs used in the experiments presented in Chapter III were either dual luciferase mRNA constructs, shorter RNAs containing the IRES only, or the IRES with some extensions on the 5’ or 3’ end. The dual luciferase vector pCrPV1-1 uses the T7 promoter to drive transcription of the *Renilla* luciferase (rLuc) gene followed by a mutant EMCV IRES sequence (to prevent ribosome read-through from the upstream ORF), the WT CrPV IGR IRES sequence, some viral coding sequence, and finally the firefly luciferase (fLuc) gene. The shorter IRES constructs were made either by PCR generation of transcription templates from the pCrPV1-1 vector (such as the toeprint RNAs) or from the CrPV4 pUC19 vector which harbors the full IRES sequence and the WT GCU start codon with an upstream Hammerhead ribozyme and a downstream hepatitis delta virus (HDV) ribozyme. See Appendix 2.1-3 for all constructs and relevant primers. All transcription was driven by the T7 RNA Polymerase promoter sequence, and all vectors contained the ampicillin resistance gene for selection purposes. The pCrPV1-1 dual-luciferase vector was a kind gift from Dr. Eric Jan. Reporter vectors containing wild-type

---

1 Portions of this chapter are taken with permission from: Ruehle, M. D. et al. “A dynamic RNA loop in an IRES affects multiple steps of elongation factor-mediated translation initiation.” eLife, in revision.
IAPV, HoCV, KBV, HiPV, TSV, SInV and ABPV IGR IRES sequences were generated by PCR amplification of the IRES sequence (plasmids were gifts from Dr. Eric Jan and Dr. Sunnie Thompson) and subsequent ligation into a dual luciferase vector (pDBS, derived from pBluescript, a gift from Dr. Les Krushel). Mutagenesis was employed using the QuickChange (Agilent) method. DNA sequences encoding the RNA for assembly assays (“CrPV4”: full IRES RNA sequence including GCU start codon) and RNase T1 probing (“CrPV11”: domain III only, no start codon) were cloned into pUC19-derived vectors with a T7 promoter and a 5’ Hammerhead ribozyme and 3’ HDV ribozyme flanking the IRES sequence. Constructs for reconstituted functional analysis (“FVKM RNAs”) were built by PCR from the CrPV1-1 vector using primers that contained the appropriate mutations and flanked with restriction sites for cloning into pUC19 (without ribozymes). All cloned sequences including the luciferase open reading frames were verified by standard sequencing methods using appropriate primers.

**RNA Preparation**

RNAs for translation assays were *in vitro* transcribed from XbaI-linearized vectors using the MEGAscript Kit (Life Technologies). RNA purification was performed by extraction with TriReagent (Sigma) followed by chloroform extraction and column purification using the RNeasy Kit (Qiagen) (Plank et al. 2013). RNAs for all other assays were made by *in vitro* transcription using T7 RNA polymerase and PCR-generated DNA templates, as described previously (Pfingsten et al. 2007). These RNAs were purified on 10% polyacrylamide-urea denaturing slab gels, visualized by UV-shadowing, passively eluted
at 4°C into nuclease-free water, then concentrated and buffer-exchanged using appropriate MWCO centrifugal ultrafiltration devices (Millipore). All RNAs were assessed for quality using denaturing PAGE.

Radiolabeling RNA and Primers

RNAs not made with ribozymes were treated with rAPid Alkaline Phosphatase (Roche) to remove the 5’ triphosphate, whereas no treatment was needed for RNAs made with a 5’ ribozyme or for synthetic primers (IDT), which have a 5’ hydroxyl. RNA was 5’ end-labeled using T4 polynucleotide kinase (New England Biolabs) and $^{32}$P-gamma-ATP (PerkinElmer), then purified by denaturing gel electrophoresis, eluted, and precipitated as described previously (Kieft et al. 1999).

General RNA Folding Method (all RNAs)

RNA folding was typically performed as follows. No more than 1µg of RNA was added to 1µL of 300mM HEPES-KOH pH 7.5 buffer in 9µL of nuclease-free water. This 9µL reaction was heated at 85°C for 30 seconds to 1 minute to denature the RNA. Immediately upon removal from heat, 1µL of MgCl$_2$ ranging in concentrations from 20-100mM depending on the downstream experiment was added to the reaction, and cooled to room temperature on the benchtop for at least 5 minutes, then transferred to ice. In some cases, after adding magnesium the sample was “snap-cooled” by immediately placing on ice. This 10µL reaction was often scaled up and then aliquotted to ensure that samples with the same RNA were treated exactly the same way.
In Vitro Translation Assays

Pure dual-luciferase reporter RNAs (0.5µg) were incubated in 35µL of rabbit reticulocyte lysate (RRL; Promega) supplemented with 150 mM potassium acetate (final concentration) and amino acids (final reaction volume of 50µL) for 90 min at 30°C, as previously described (Pfingsten et al. 2010). Luciferase production was measured using the Dual Luciferase Reporter Assay System (Promega) and the GloMax Multi Detection plate reader. Data shown are from 5 independent experiments.

Assembly Assays

1000 cpm of 5'-32P end-labeled CrPV RNAs were incubated in 30µL of RRL and incubated at 30°C for 5 min. Sometimes these lysates were supplemented with 150mM KOAc as in the translation assays, but this did not have a significant impact on the results. Samples were diluted in 500µl ribosome association dilution buffer (RADB, 50mM Tris pH 7.5, 50mM NaCl, 5mM MgCl₂, 1mM DTT) and separated by 15-30% sucrose gradient density fractionation in an SW41 rotor for 3 hours at 36,000 rpm, 4°C. Fractions were collected on a BioComp gradient maker and fractionation system. The amount of 32P in each fraction was determined by filter binding (described in depth below) and exposure to a phosphorscreen.
Filter Binding Assays

Approximate on-rate: IRES RNAs and a negative control RNA (Murray Valley Encephalitis Virus xrRNA) were 5’ end-radiolabeled. The RNAs were diluted to 100cpm/µl in RNase-free water, which resulted in RNA concentrations in the attomolar range. 100cpm of RNA was used per 50µL reaction. RNAs were heated at 85°C for 1 min in 30mM HEPES-KOH pH 7.5 and removed from heat. MgCl₂ to 10mM final concentration was added and the RNAs were allowed to cool on the benchtop for 5 min. Pure shrimp ribosomes were added to the RNA at room temperature to a final concentration of 30nM, and then 50µL aliquots were removed from the reaction at defined time points out to 12 min and immediately pipetted through a membrane sandwich of nitrocellulose (BioRad) (on the top), Hybond nylon membrane (GE Healthcare) (middle), and Whatman filter paper (VWR) (bottom), on a dot-blot vacuum manifold. Membranes were air-dried then exposed to a phosphorscreen. The screens were imaged on a Typhoon phosphorimager scanner. The data were analyzed by drawing equal sized circles around each dot using ImageQuant software and obtaining a volume/intensity report for each circle. Fraction bound was then calculated from the intensity signals as follows: \( \frac{\text{Nitrocellulose}}{\text{Nitrocellulose} + \text{Nylon}} \).

Approximate off-rate: 100cpm of RNA per 50µL reaction was folded as described above in 30mM HEPES-KOH pH 7.5 and 10mM MgCl₂. 15nM purified yeast 40S and 60S ribosomal subunits were added to the folded RNA and incubated at 37°C for 15 min. 5µg of unlabeled RNA was added to each reaction (WT RNA added to the WT reactions, and G-rich RNA added to the G-rich reactions, ~240nM), and 50µL aliquots were removed at
defined time points out to 30 min and immediately applied to the membrane sandwich as described above. Data were analyzed as described above.

**Toeprinting Assay**

For unbound IRES RNAs, 0.5µg of toeprint RNA was mixed with 1.5µl of 10X Toeprint Buffer A (1X: 20mM Tris pH 7.5, 100mM KOAc, 2.5mM MgOAc$_2$, 2mM DTT, 1mM ATP, 0.25mM spermidine), 0.5µl of RNasin Plus (40 U/µl, Promega), and nuclease-free water to a final volume of 15µl. For ribosome-bound RNAs (purified yeast 40S and 60S subunits or purified rabbit 40S), reactions were set up in the same way as above but included 8pmol of each purified subunit. For rabbit reticulocyte lysate (RRL)-incubated RNAs, 11µl of RRL was pre-incubated with 1µl of 45mg/mL cycloheximide or 1µl nuclease-free water for 5 min at 37°C, and added to RNA and 10X buffer A as above. All reactions were incubated at 30°C for 5 min to allow for folding and binding. 1µl of 40,000 cpm/µl toeprint primer (internal photinus) and 24µl of 1X Buffer A were added and incubated at 30°C for 5 min for primer annealing. Reverse transcription was performed by addition of 4µl dNTPs (1.25mM each), 1µl 320 mM MgOAc$_2$, and 0.5µl avian myoblastosis virus reverse transcriptase (25 U/µl, Promega) to each reaction. Primer extension was performed at 30°C for 45 min, and was quenched with 4µl of 4M NaOH and heated at 85°C for 5 min to hydrolyze RNA. 100µl of nuclease-free water was added to each reaction before extraction with phenol:chloroform:isoamyl alcohol (PCIAA, 24:24:1, ThermoFisher), followed by CIAA (24:1) (ThermoFisher) extraction, and ethanol precipitation with 3 volumes of 100% ethanol and 1/10 volume of 3M NaOAc.
pH 5.3. Pellets were washed with 70% cold ethanol. Precipitated RNA pellets were dried and resuspended to equal counts/µl in 1X TBE + 9M urea loading buffer, then equal volumes (typically 10 µl) were loaded on a 10% polyacrylamide sequencing gel (1 mm gel thickness) with a sequencing ladder of the WT RNA (made by dideoxy-NTP incorporation as previously described (Filbin et al. 2013)) and electrophoresed at 65W for approximately 2 hours. Gels were dried and exposed to a phosphorscreen overnight. Gels were imaged on a Storm scanner (GE Healthcare) and analyzed in ImageQuant.

“Percent translocated” toeprints were calculated for each RNA in RRL with cycloheximide treatment by quantifying the intensity of the +14/15 toeprint and the +20/21 toeprint in equal sized boxes in ImageQuant, and using these values in the equation: (+20/21)/(+14/15 + +20/21). Toeprinting assays using concentrated hygromycin B were performed essentially as described above, however 1µl of 30mg/mL hygromycin B (Roche) was added to the RRL and pre-incubated for 5 min at 37°C. For toeprinting assays in the presence of dilute hygromycin B, 0.5µg of each RNA was incubated for 1 min in RRL / Buffer A / RNasin mix (as above) at 30°C before adding 1µl of 0.05mg/mL hygromycin B (“+”) or nuclease free water (“-”). Reactions were incubated at 30°C for 5 min before adding radiolabeled primer and buffer as above. Reverse transcription and gel analysis were performed as described above.

**Ribosome and Elongation Factor Purification**

Both yeast (*S. cerevisiae*) and shrimp (*A. salina*) eggs were used as sources of 40S and 60S ribosomal subunits. Yeast subunits were purified from strain YAS2488 as described
(Acker et al. 2007). Briefly, cells were lysed using a liquid nitrogen mill, and clarified lysates were spun through 250mM sucrose cushions under high-salt conditions to obtain clean 80S ribosomes. Subunits were separated by treatment with puromycin and resolved on 5-20% sucrose gradients. Crude shrimp egg 80S ribosomes were prepared from dried, frozen cysts as previously described (Iwasaki and Kaziro 1979; Thiele et al. 1985) with some modifications. After the shrimp cysts were ground open, debris was removed by centrifugation at 30,000xg for 15 min and crude 80S ribosomes were precipitated from the supernatant by addition of 4.5% (w/v) PEG 20K according to previous methods (Ben-Shem et al. 2011). Subunits were resolved on 10-30% sucrose gradients after puromycin treatment. eEF1A was purified from yeast according to published methods (Thiele et al. 1985). His<sub>6</sub>-eEF2 was isolated from an overexpressing yeast strain (TKY675), obtained from Dr. Terri Kinzy, and purified as described (Jørgensen et al. 2002).

**Tetrapeptide Kinetics Assay**

Preinitiation complexes (Pre-ICs) were formed by incubation of shrimp egg 40S and 60S subunits with FVKM IRES RNA constructs at 37°C for 5 min in buffer 4 (40 mM Tris-HCl pH 7.5, 80 mM NH₄Cl, 5 mM MgOAc₂, 100 mM KOAc, 3 mM β-mercaptoethanol). tRNAs were charged with appropriate amino acids as described (Pan et al. 2009). Phenylalanine, valine, lysine, and <sup>35</sup>S-methionine ternary complexes (TCs) with purified yeast eEF1A were formed as separate complexes by incubating the relevant charged tRNA (1.6 µM, based on amino acid stoichiometry) with eEF1A (8 µM) in buffer 4.
supplemented with 1 mM GTP and 1 mM ATP at 37°C for 5 min. Tripeptide complexes were made by mixing Pre-ICs with 1 µM eEF2 and F, V, and K TCs at 37°C for 15 min. Using a quench-flow instrument, tetrapeptide complexes were made by mixing the tripeptide complexes with 35S-Met TC for defined time points on the millisecond scale. Reactions were quenched with 0.8 M KOH and peptide was released from tRNA by further incubation at 37°C for 3 h. Samples were neutralized with acetic acid, lyophilized and suspended in water. Following centrifugation to remove particulates (which contained no 35S), the supernatant was analyzed by thin layer electrophoresis as previously described (Youngman et al. 2004). The identities of the tri- and tetrapeptides were confirmed by their comigrations with authentic samples obtained from GenScript (Piscataway, NJ). A further demonstration of tetrapeptide identity was provided by matrix-assisted laser desorption/ionization (MALDI) mass spectrometric analysis (Ultraflex III TOF/TOF, Bruker).

A-site tRNA Binding: Anisotropy

Phe-tRNA^{Phe}(prf) was prepared as previously described (Wintermeyer and Zachau 1974; Betteridge et al. 2007). Ternary complex (0.1 µM, 250 µL) was incubated with 80S or 80S-IRES complex (0.1 µM, 250 µL) in Buffer 4 for 15 min at 37°C and then kept on ice until anisotropy measurement, which was performed at 23°C. Steady-state fluorescence anisotropy was determined using a Photon Technology International (PTI) QuantaMaster fluorometer with polarizer in L-format, with excitation at 462 ± 2 nm and fluorescence emission collected at 490 ± 2 nm. Instrument-integrated monochromators
were used as filters for the fluorescence emission and the excitation light. The g-factor and anisotropy value were calculated using the instrument software. The instrument was calibrated by using suspended nonfat dry milk aqueous solution as scatter. Experimental data were processed and analyzed by Felix software (from PTI).

**P-site tRNA Binding: Sucrose Cushion Cosedimentation**

80S-IRES complexes containing Phe-tRNA$_{\text{Phe}}$ in the P site were formed by incubation of pre-IC (16 pmol) and Phe-TC (32 pmol) at 37°C for 15 min in the presence of 1 μM eEF2, in a total volume of 40 μL. The 80S-IRES complexes were isolated by ultracentrifugation at 4°C (540,000xg) for 40 min through a 1.1 M sucrose cushion, with 600 pmol of pure 30S bacterial ribosome subunits added as carrier to enhance pelleting and allow facile calculation of complex recovery. The pellets were gently washed twice with buffer 4 and dissolved in 100 μL of buffer 4 for $A_{260\text{nm}}$ determination. Recoveries typically varied between 60 and 80%. $^3$H counts from the pellet were measured to determine the amount of $[^3\text{H}]$-Phe-tRNA$_{\text{Phe}}$ bound to the complex.

**Translocation Efficiency Analysis**

The percent A-site (Fig. 4.9C) and P-site (Fig. 4.9B) tRNA binding levels were each divided by the percent of A site binding for the WT, Δ1, and Δ2 mutants, then multiplied by 100%. This permits analysis of the percentage of A-site tRNA that was moved to the P site for each of these RNAs.
Single Molecule Colocalization Assays

WT and Δ3 IRES RNAs for single-molecule analysis were generated with a 5’ extension of sequence (5’)-CA AAU CAA CCU AAA ACU UAC ACA-(3’) such that a complementary, 3’-biotinylated DNA oligo ((5’)-TGT GTA AGT TTT AGG TTG TG/3Biotin/-(3’)) could be hybridized to the IRES constructs. The biotin at the 3’ end of the DNA oligo that had been hybridized to the IRES RNAs could then be used to tether the 80S-IRES ribosome complexes to the polyethylene glycol-, biotin-polyethylene glycol-, and streptavidin-derivatized quartz surface of a microfluidic observation flowcell (Fei et al. 2008; Blanchard et al. 2004; Ha et al. 2002). The 3’ end of the IRES RNAs contained one codon for Phe (UUC), followed by the hepatitis delta ribozyme to generate a clean 3’ end. 2’-3’ cyclic phosphates were removed as previously described (Kieft et al. 1999). IRES RNAs were labeled using Cy3-maleimide (GE Healthcare) and the 3’ DNA End-Tag Kit (Vector Labs), which added one additional dG residue harboring the Cy3 label to the 3’ end of the IRES construct using the terminal transferase enzyme. IRES(Cy3) RNAs were purified from free dye by multiple phenol extractions and ethanol precipitation, or centrifugal filtration with a 10,000 Da MWCO (Millipore). Labeling efficiencies determined by A\textsubscript{260nm} and A\textsubscript{550nm} readings were typically low, ranging from 3%-20%. A diagram of the RNA constructs is shown in Fig. 4.10. Stocks of IRES(Cy3) RNAs that had been hybridized to the biotinylated DNA oligo were prepared by incubating a 10-fold excess (50 nM) of the 3’-biotinylated DNA oligo with either 5 nM WT IRES(Cy3) or 5 nM Δ3 IRES(Cy3) RNA (in a reaction volume of 100 µL) at 95°C for 2 min, slowly cooling the hybridization reactions to room temperature, transferring the hybridization reactions to ice, aliquoting, flash-
freezing in liquid nitrogen, and storing the stocks at -80°C. These stocks, therefore, had 5 nM of either WT IRES(Cy3) or Δ3 IRES(Cy3) RNA.

Purified *E.coli* tRNA\textsuperscript{Phe} (Sigma) was fluorescently labeled with Cy5-NHS ester (GE Healthcare) at the primary aliphatic amino group of its naturally modified acp\textsuperscript{3}U47 residue, according to previously published protocols (Fei et al. 2010). The labeling reaction was quenched with 0.3 M NaOAc (pH 5.2), phenol-chloroform extracted, ethanol precipitated, and the Cy5-labeled tRNA\textsuperscript{Phe} (tRNA(Cy5)) was separated from unlabeled tRNA\textsuperscript{Phe} by hydrophobic interaction chromatography (HIC) using a TSK gel Phenyl-5PW column (Tosoh Biosciences) attached to a ÄKTA FPLC system (GE Healthcare) as previously described (Fei et al. 2010). The HIC-purified tRNA\textsuperscript{Phe}(Cy5) was charged with phenylalanine (Sigma) as described using *E.coli* Phe-tRNA synthetase that was overexpressed and purified as previously described (Fei et al. 2010). The charging reaction was quenched with 0.3M NaOAc (pH 5.2), phenol-chloroform extracted, ethanol precipitated, resuspended in 10 mM ice-cold KOAc (pH 5), passed through a Micro Bio-Spin Gel Filtration spin-column (Bio-Rad), aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C. Charging efficiency was estimated by running an aliquot through a Phenyl-5PW column to detect the charged Cy5-Phe-tRNA\textsuperscript{Phe} and uncharged Cy5-tRNA\textsuperscript{Phe}, separated by HIC. The typical charging efficiency in these reactions was >90%.

For each colocalization experiment, IRES-80S ribosome complexes were initially assembled using 1.25 nM oligo-hybridized-Cy3-IRES RNA and 100 nM each of yeast 40S and 60S subunits in 1X Eukaryotic Polymix Buffer (EPB: 50 mM Tris-acetate at pH 7 at
25°C, 100 mM KOAc, 10 mM MgOAc₂, 0.5 mM spermidine, and 10 mM β-mercaptoethanol). In a separate reaction tube, a ternary complex was prepared using 500 nM Phe-tRNA^{Phe}(Cy5), 5 µM eEF1A, and 2 mM GTP in 1X EPB. Each of these two reaction tubes were incubated at 37°C for 10 min. 1 µM eEF2 and 2 mM GTP were then added to the IRES-80S ribosome complex to initiate the first pseudo-translocation reaction and the reaction was allowed to proceed for an additional 10 min at 37°C (during which the reaction tube containing the ternary complex was kept on ice). Subsequently, the ternary complex was added to the IRES-80S ribosome complex (containing eEF2 and GTP) and the entire reaction incubated for another 10 min at 37°C. Finally, the entire reaction was diluted 5-fold in 1X EPB and the diluted reaction was delivered into the polyethylene glycol-, biotin-polyethylene glycol-, and streptavidin-derivatized quartz microfluidic observation flowcell (Blanchard et al. 2004). The 80S-IRES ribosome complex was incubated in the flowcell for 5 min and components that remained untethered to the surface of the microfluidic flowcell at the conclusion of the 5 min were washed out of the flowcell using an imaging buffer composed of 1X EPB and a protocatechuic acid/protocatechuic acid-3,4-dioxygenase based oxygen scavenging system (Aitken et al. 2008). Cyclooctatetraene (COT, Sigma), and 0.012% v/v 3-Nitrobenzyl alcohol (NBA, Sigma) were included as triplet state quenchers in these experiments.

Surface-tethered, Phe-tRNA^{Phe}(Cy5)-bound 80S-IRES ribosome complexes were imaged using a custom-built, prism-based total internal reflection fluorescence microscope. Cy3 and Cy5 fluorophores were excited with a 532 nm laser and a 640 nm
laser (CrystaLaser®), respectively, with their powers attenuated such that the laser beams measured ~8 mW when they hit the prism. Emission data were directed to the image sensor of an electron-multiplying charge-coupled device (EMCCD) camera that records the fluorescence emission as a ~2-min movie with a frame rate of 100 msec. Prior to striking the image sensor of the EMCCD camera, the fluorescence emission from Cy3 and Cy5 are wavelength-separated using dichroic beamsplitters such that they could be directed onto the two separate halves of the image sensor. Colocalization data were analyzed from the imaged frames, using the standard software MetaMorph®, as follows: the 256 pixel X 256 pixel imaged frames were split into the green and red halves, each half being 128 pixel X 256 pixel. Spots were picked from the red frame, using automated features in MetaMorph® and designated as ‘Areas’. The red frames were then stacked on the green frames and the ‘areas’ were transferred from the red to the green frames. Spots were then manually counted on the green frame that overlapped with the areas from the red frame and the value reported as percentage of co-localized spots.

For the experiments designed to test the effect that the absence of eEF2, prior to addition of the ternary complex, had on the colocalization, the first 10-min incubation step of the IRES-80S complex with eEF2-GTP was omitted. For these experiments, after imaging the IRES-80S complexes with Phe-tRNA_{Phe}(Cy5) delivered by eEF1A, the same channel was washed three times with 1X EPB to remove all unbound components and a fresh mix of pre-incubated eEF2-eEF1A-GTP-Phe-tRNA_{Phe}(Cy5) was delivered to the flowcell prior to a second round of imaging aimed at monitoring the rescue of co-
localization by addition of eEF2. Similarly, in experiments targeted to detect the effect of eEF1A on co-localization, eEF1A was not added to the initial reaction tube in which the ternary complex was set up. In this case, after imaging the IRES-80S complexes with Cy5-Phe-tRNA\textsuperscript{Phe}, the channel was washed with 1X EPB and a fresh mix of pre-incubated ternary complex containing eEF2-eEF1A-GTP-Phe-tRNA\textsuperscript{Phe}(Cy5) was delivered to the flowcell to detect restoration of co-localization. All experiments were performed at least in duplicate and data from at least 5 movies for each experiment were averaged to calculate the co-localization percentage under a given set of conditions.

Methods for Chapter IV

RNA Constructs

Preparation of the RNA constructs used in this chapter is described in the Methods for Chapter III section above. Please refer to Appendix 2.1-3 for relevant cloning and design information.

Native Gel Analysis

CrPV-11 WT, Δ3, and G-rich RNAs (1μg) were added to 1x TH-Magnesium loading dye (1x: 66mM Tris-HCl, 34mM HEPES pH 7.4, 40% glycerol, 10mM MgCl\textsubscript{2}, trace xylene cyanol and bromophenol blue), heated at 85°C for 1 minute and cooled on the benchtop for 5 minutes. Magnesium-containing gels (66mM Tris-HCl, 34mM HEPES pH 7.4, 10mM MgCl\textsubscript{2}, 12% polyacrylamide) were pre-run in buffer of the same composition (without acrylamide) for roughly 30 minutes at 8W in a 4°C cold room. Gels were loaded while current was applied, and ran for 5.5 hours at 8W. When complete, gels were taken
down, stained with 0.5µg/mL ethidium bromide, and imaged with a UV transilluminator. Samples for EDTA native gel analysis were prepared the same way as described above except the loading buffer, gels, and running buffer contained 10mM EDTA and no magnesium.

**Native RNase T1 Probing**

40,000 cpm of 5’ 32P labeled CrPV-11 (domain III only) WT, G-rich, and Δ3 RNAs were folded by heat-cooling in 30mM HEPES-KOH pH 7.5, 10 mM MgCl₂, in the presence of 1µg carrier tRNA. RNase T1 (Roche) digestion was performed by adding 0.1 U of enzyme and incubating at 37°C for 2 min. RNAs were ethanol precipitated overnight and resuspended to equal counts per µl in 1X TBE + 9M urea loading buffer. RNase T1 (G) ladders for each RNA and a hydrolysis ladder of the WT CrPV-11 RNA were generated as previously described (Kieft et al. 1999). Samples were loaded on a 12% polyacrylamide-urea sequencing gel (1 mm gel thickness) and run for 2 hours at 65W. For analysis, data were normalized to total amount of radiation in each lane before subtracting the appropriate non-native T1 cleavage signal (G ladders) from the native T1 cleavage signal.

**IGR IRES SHAPE Probing**

SHAPE probing of WT, Δ3, and G-rich CrPV mutants was performed with N-methylisatoic anhydride (NMIA) as previously described (Filbin and Kieft 2011). Briefly, 8pmol of each RNA (toeprint RNA constructs, see Chapter III Methods) was folded in HEPES buffer and 10mM MgCl₂, as above. RNAs were probed for 35 minutes at 35°C with 75mM NMIA in
DMSO, or using neat DMSO as a no-modification control. Samples were desalted using G-25 Sephadex spin columns. Reverse transcription was performed using the YH-SHAPE thermocycling protocol (defined below) with the exception that the annealing temperature was 45°C, and the reverse transcriptase (RT) was SuperScript III. 3µM 6-FAM internal photinus primer was used. RT reactions were desalted as above and dried to ~20µL. 20µL of Hi-Di formamide and 1µL of Rox 500 size standard (ABI) were added to each sample, then transferred to a 96-well ABI conical plate and sequenced on an ABI 3400 sequencer. Data were analyzed in an early version (Jan 2012) of HiTRACE using freely available MATLAB scripts. Raw reactivates were used to calculate differences from WT at each position and were plotted in Microsoft Excel.

**X-ray Crystallography**

**Sample Preparation:** CrPV-11 Δ3 RNA was added to 1x High Mg** Buffer G (5mM HEPES-KOH, 20mM Mg(OAc)₂, 10mM NH₄Cl, 50mM KOAc) at room temperature. 70S ribosomes from *Thermus thermophilus* were previously prepared by Brian Wimberly, as described (Selmer et al. 2006). 70S ribosomes were thawed on ice and gently vortexed, then diluted in the buffer above. RNA and ribosomes were then combined, and deoxy big chap detergent was added to a final concentration of 1.5mM, then gently vortexed. The sample was incubated at 55°C for 30 minutes. Condensation was spun down, and the sample was mixed by pipetting up and down. This sample was used to set up 24-well crystal trays with 500µL of well solution in the well and 3µL of sample plus 4µL of well solution in the drop. Drops were pipetted up and down five times. Several sample
conditions ranging from 8-10 fold excess domain III over ribosomes were tested. Final concentration of ribosomes in the sample used to set up the drops was 4.5µM.

**Crystallization Conditions:** Crystal trays contained well solutions of 0.175M Arginine-HCl, 11% MPD (2-methyl-2,4-pentanediol), and increasing concentrations of Polyethelene glycol 20,000 from 2.3% to 2.6% v/v. Trays were covered with clear tape and stored at 18°C. Crystals grew in all of these conditions.

**Cryoprotection and Harvesting:** Crystals were harvested after growth for one week. Cryoprotection solutions were made in 1x Buffer G supplemented with 25%, 30%, and 35% final concentrations of MPD, and 5-10µM of folded CrPV-11 Δ3 RNA. Slits were cut in the tape over wells harboring crystals to be harvested. 5µL of 25%, then 30%, then 35% cryoprotection solutions were added to the drops with 15 minute wait intervals between each addition. Finally, 10µL were carefully removed from the drop, followed by addition of 10µL of 35% cyroprotection solution. This step was repeated after 15 minutes. Then, 15µL were carefully removed from the drop, followed by addition of 20µL of 35% cryoprotection solution. This step was repeated after 15 minutes. Tape covering wells to be harvested was cut away to expose the drop. Crystals were scooped into nylon loops and immediately plunged into liquid propane in 2mL cryo tubes in a liquid nitrogen bath. Cryo tubes containing crystals in propane “popsicles” were stored in liquid nitrogen until data collection.

**Data Collection and Processing:** X-ray diffraction data were collected at the Brookhaven National Laboratory National Synchotron Light Source on the X25 beamline. Data were indexed using HKL2000 software, integrated and merged in the XDS package, and scaled
using xscale. Structure determination was executed in Phenix using a rigid body alignment to an empty 70S ribosome structure of the same crystal form.

**Electron Microscopy**

**Sample Preparation:** Sample preparation and microscopy was performed in the laboratory of Tamir Gonen at Janelia Farm Research Campus. Samples were made by performing assembly assays essentially as described above, with the following exceptions. 1) The IRES RNA was the WT CrPV FVKM construct described in Chapter III methods and was not radiolabeled. 2) Rabbit reticulocyte lysates were preincubated with either 0.7mg/mL hygromycin B, 0.1mg/mL cycloheximide, or 10nM bouvardin (final concentrations) for 5 minutes at 37°C. 3) RRL and IRES were incubated together for 15 minutes before loading on the sucrose gradients. Sucrose gradients for complex separation were prepared with the appropriate antibiotic/inhibitor in the solution. Fractions corresponding to the 80S ribosome peak were buffer exchanged by dialysis to remove sucrose.

**Negative Staining:** Samples still in sucrose solution were used for negative staining with 3% uranyl formate on carbon-coated grids. 3μL of sample was applied to the grid and incubated for 1 minute at room temperature. Excess sample was wicked away on filter paper, then the grid was touched to two 50μL drops of 3% uranyl formate with excess liquid wicked away by touching filter paper between steps. The sample was incubated in stain for another minute at room temperature before excess was removed by wicking. Grids were then washed with nuclease-free water 5 times by similar 50μL drop-touching
techniques. In the final step, excess liquid was wicked away and then a gentle suction was applied to dry the grid.

Cryo grid preparation: Samples freshly out of dialysis were used to make cyro-grids on a Vitrobot (FEI) with the following conditions: 3.5µL of sample, 120s wait time, 4.5s blot time, 0s drain time, 100% humidity, 15°C temperature, 0 blot force, and 1 blot. Sample images were taken on an FEI T12 cryo-electron microscope.

Methods for Chapter V

Plasmid Construction and RNA Preparation

A gBlock (Integrated DNA Technologies) containing wild type CCR5 -1 PRF RNA sequence (nucleotides 1-103) and harboring the T7 promoter and Hammerhead ribozyme sequences upstream of the CCR5 sequence and the HDV ribozyme downstream was cloned into the pUC19 cloning vector using EcoRI and BamHI restriction site. This construct is referred to as CCR5-1 (Appendix 2.3).

Constructs CCR5-10, -12, -14, -15, -16, and -17 were cloned into pUC19 vectors in essentially the same way, except the inserts were built by PCR using the CCR5-1 vector as a template and using overlapping forward primers to incorporate the necessary changes in the CCR5 sequence and hammerhead ribozyme. The reverse primer for these PCR reactions was M13 Reverse -27.

Constructs CCR5-1.89, -1.S2D, and -1.polyU12miR1224 were not cloned, but templates for in vitro transcription were generated by PCR using the CCR5-1 vector as the template and M13 Forward primer and the reverse primers containing the appropriate sequences to generate the desired 3’ ends (Appendices 2.1 and 2.3). These
constructs did not use 3’ ribozymes. RNAs were *in vitro* transcribed as described in the methods section for Chapter III. microRNAs (mir-1224 and 6144U) were synthesized by Dharmacon or IDT, deprotected, and HPLC purified by the company.

The methods for template construction and *in vitro* transcription of RNAs for 1D and 2D Probing will be described in those sections below.

**Radiolabeling RNA, primers, and microRNAs**

Radiolabeling was achieved exactly as described in the Chapter III section. Synthesized microRNAs have 5’ OH chemistry and needed no enzymatic modification before labeling.

**Electrophoretic Mobility Shift Assay**

10,000 cpm of 5’ ^32^P labeled CCR5 RNAs were added to freshly made 1x TH-Magnesium loading dye (1x: 66mM Tris-HCl, 34mM HEPES pH 7.4, 40% glycerol, 10mM MgCl$_2$, trace xylene cyanol and bromophenol blue), heated at 85°C for 30 seconds and cooled on the benchtop for 5 minutes. 1µL of 100µM miR-1224 or 6144U control miRNA or nuclease-free water were added to the appropriate samples, then incubated for 10 minutes at 37°C. For samples where radiolabeled miR-1224 and unlabeled CCR5-1 RNA were used, the sample was prepared in essentially the same way, but with 10,000 cpm of 5’ ^32^P labeled miR-1224 and 1µg of unlabeled CCR5-1 RNA.

Magnesium-containing gels at room temperature (66mM Tris-HCl, 34mM HEPES pH 7.4, 10mM MgCl$_2$, 10% 29:1 bis:tris acrylamide) were pre-run in buffer of the same
composition (without acrylamide) for roughly 20 minutes at 16W, then turned down to 8W for sample loading. Gels were loaded while current was applied, and ran for 6 hours at 16W. When complete, gels were taken down, wrapped in plastic, and exposed to a phosphorscreen for roughly one hour before visualization on a Typhoon imager.

**EDTA Native Gel Analysis**

Samples for EDTA native gel analysis were prepared the same way as described for EMSA experiments except no miRNA was added and the loading buffer contained 5mM EDTA and no magnesium. Gels with the same buffer conditions were pre-run at 8W for 30 min at 4°C, samples were loaded while current was applied, and continued running at 8W for 12 hours.

**Native RNase T1 Probing**

1μL of 40,000 cpm/μL 5’-32P labeled CCR5 RNAs were added to folding reactions containing 30mM HEPES-KOH pH 7.5, 1μg “carrier” tRNA, and 10mM MgCl₂ in a final volume of 8μL with heating at 85°C before magnesium was added. 1μL of 10μM unlabeled miR-1224 or 6144U control miR, or nuclease-free water were added to appropriate tubes, then incubated at 37°C for 15 minutes. 1μL of 1:1000 freshly diluted RNase T1 (0.1 Units) was added to each tube and incubated at 37°C for 2 minutes. Samples were ethanol precipitated as described in Chapter III methods using a dry ice/ethanol bath for 20 min before pelleting and washing with 70% ethanol. Samples were dried using a speed vac and resuspended to equal counts per μl in urea loading
dye. Alkaline hydrolysis and denaturing RNase T1 ladders were made as described in Chapter IV methods. All samples were electrophoresed on a 10% acrylamide-urea sequencing gel at 65W for 1 hour 15 minutes to 2 hours (to better resolve bands at the 3’ end of the RNA). Gels were transferred to Watman paper, dried, exposed to a phosphorscreen for at least 1 hour, and then imaged with a Typhoon imager.

Quantitative analysis of RNase T1 probing was performed by using ImageQuant software to calculate the amount of total radiation in each lane by drawing equal-length lines through each lane. The CCR5-1 Only lane was used as the reference signal to which all other lanes were normalized. The normalization factors to set the total radiation levels equal to one another were then applied to “volume” values of boxes that were drawn around each G band in each lane. This process was repeated for gels with longer run times to obtain better data for Gs at the 3’ end of the RNA sequence. The volume values of each G in the CCR5-1 RNA alone lane were each set equal to 1 by dividing the value by itself. For example, (volume of G8 box in CCR5 Only sample) / (volume of G8 box in CCR5 Only sample) = 1. Then, each G for each of the other samples was divided by the CCR5-1 RNA Only value for a specific box to obtain the fold change upon miR-1224 binding. For example, fold change = (volume of normalized G8 box in CCR5+miR sample) / (volume of G8 box CCR5 Only) = X. This analysis was only performed for the “5x miR/6144U” samples shown in Fig 5.4.
Fluorescence Anisotropy-Stoichiometry of Binding

50nM TAMRA-miR-1224 (Dharmacon) in 1mL of 30mM HEPES-KOH pH 7.5, 100mM KOAc, and 10mM MgCl₂ was added to a tall, 1mL quartz cuvette and adapter. Fluorescence anisotropy of the TAMRA “probe” alone was calculated at room temperature on a FluoroMax-3 fluorometer in L-format with polarizers in and excitation at 550nm and emission at 580nm. Dark value corrections were enabled. G-factors were calculated automatically by the instrument. 1µL of 10µM CCR5-1 folded RNA in the same buffer conditions was added to the cuvette (leading to 10nM final CCR5 concentration), then inverted 5 times before collecting 3 anisotropy measurements as described above over the course of 8 minutes. Similar 1µL additions with increasing concentrations of CCR5-1 RNA were performed to obtain data points at 10nM increments up to 100nM, then 150nM, 200nM and 300nM CCR5-1 RNA. Increases in total volume were considered negligible to the concentration of the TAMRA probe. The anisotropy values collected were averaged and plotted as a function of CCR5-1 concentration using KaleidaGraph software. Two separate linear regressions were calculated for points up to 50nM and for points after 50nM of added CCR5-1 RNA.

Site-Specific UV Crosslinking

4-thio-uridine at position 17 in miR-1224 (4SU(17)-miR-1224) was synthesized by Dharmacon. CrPV IGR IRES RNA and CCR5 RNAs were folded as follows: 2.5µg of RNA were added to 30mM HEPES-KOH pH 7.5 and 100mM KOAc, in a final volume of 8.5µL. These samples were heated at 85°C for 30 seconds, 1µL of 100mM MgCl₂ was added,
and then allowed to cool at room temperature for 5 min. 0.5μl of 40,000cpm/μl 4SU(17)-miR-1224 was added and samples were incubated at 37°C for 15 minutes.

Samples were crosslinked on ice by opening the 1.7mL eppendorf tubes they were in and setting a hand-held UV lamp set to long wavelength (365nm) over the samples in a dark room for 10-60 minutes (for time course experiment, Fig 5.7) or for 30-40 minutes for all other experiments. Samples were analyzed by electrophoresis on a 10% acrylamide-urea gel and exposure to a phosphorscreen.

**Crosslink Mapping**

Crosslinked samples were generated as described above except with 660pmol CCR5-1 RNA and 1.5μmol of unlabeled 4SU(17)-miR-1224, scaled up to 200μL folding reactions. After crosslinking, samples were hard-spun to pellet debris, and supernatants were brought up to 3mL in nuclease-free water.

Samples were loaded on a PLRP-S HPLC column (8μm inner diameter), with a 2ml/min flow rate from 2% B to 60% B elution gradient, (A=100mM Triethyl Ammonium Acetate (TEAA), 0% Acetonitrile; B= 100mM TEAA, 50% Acetonitrile). A_{260} was monitored. Fractions corresponding to UV-exposed CCR5 only or the UV-induced CCR5-miR-1224 crosslink were each pooled, concentrated, and buffer exchanged into nuclease-free water using 10,000 Da MWCO spin filters (Millipore), and then ethanol precipitated with 1μL of 20mg/mL glycogen to enhance pelleting. Pellets were washed with 70% ethanol and dried, then resuspended in 10μL of nuclease-free water.
RNA concentrations were determined by NanoDrop absorbance reading at 260nM. 10pmol of each RNA in a final volume of 9µL were transferred to PCR tubes and 1µL of 40,000cpm/µL $^{32}$P 5’ end-labeled RT primers (G41 or G80, see Appendix 2.1) were added. Reverse transcription was performed using the “YH SHAPE” program:

- 65°C for 5 min (melting)
- 42°C for 20 min (primer annealing)
- 42°C for 1 min (Add dNTPs and extension buffer)
- 42°C for 10 min (Add 80U GoScript Reverse Transcriptase)
- 95°C for 10 min (Add 4M NaOH, RNA hydrolysis)
- 95°C for 5 min (Add Acid Stop Mix, neutralize base)

Buffer composition is as described for SHAPE RT analysis above. DideoxyNTP sequencing ladders were made using this program and as previously described (Filbin and Kieft 2011). Samples were immediately electrophoresed on 10% acrylamide-urea sequencing gels. Gels were dried, exposed to a phosphorscreen, and analyzed using ImageQuant software.

**1D DMS and SHAPE Probing**

**DNA template preparation by primer assembly:** Six primers (three forward and three reverse, named 1-6 from 5’ to 3’) were designed to contain the WT CCR5-1 RNA sequence with two reference hairpins up and downstream of the CCR5 sequence. Even numbered primers were forward primers and odd numbered primers were reverse primers. A T7 promoter sequence is included upstream of the 5’ hairpin, and a 3’ “tail” sequence is included downstream of the 3’ hairpin (schematic, Fig 5.10B; actual sequences, Appendices 2.1 and 2.3) This construct is named “CCR5.2hp_tail2”. Double
stranded templates were generated by PCR-based extension of these primers (100µM primer 1 and 6, 1µM primers 2-5) in a single reaction with Phusion DNA polymerase and with the following PCR cycling protocol: 98°C x 30s, 35 cycles of 98°C x 10s, 55°C x 30s, 72°C x 30s, final extension 72°C x 10min, hold 4°C indefinitely. DNA was purified using AMPure magnetic beads (Agencourt), and products were verified by 4% agarose gel electrophoresis in 1X TBE containing 0.5µg/mL of ethidium bromide. Concentration of purified product was determined using a NanoDrop spectrophotometer.

**RNA synthesis and purification:** In vitro transcription was performed in 80µL volumes with 16pmol of DNA template, 40U T7 RNA polymerase (NEB), 1mM each NTP, 40mM Tris-HCl pH 8.1, 25mM MgCl₂, 2µM spermidine, 4% PEG-1200, and 0.01% Triton-X-100. Transcriptions were carried out at 37°C for 4 hours, and RNAs were verified on a 4% agarose-formaldehyde gel with SYBR Green II (Invitrogen) staining. RNAs were purified using AMPure beads as described above, and concentrations were determined using a NanoDrop spectrophotometer.

**RNA folding and chemical modification:** RNA at 0.06µM final concentration was added to 100mM Na-HEPES pH 8.0, heated at 90°C for 3 minutes, then cooled at room temperature for 12 minutes. Either 1M NaCl or 10mM MgCl₂ (final concentrations) were added and then incubated for another 30 minutes at 50°C and cooled on the benchtop for 12 minutes. Reaction size at this point was 10µL. 5µL of modifying reagent—either 0.5% DMS in ethanol/water mixture or 1.25mg/mL 1M7 in DMSO—or nuclease-free water (“No modification” control) was added and incubated for 15 minutes at room temperature. Modification reactions were quenched at room temperature for 10
minutes with 250mM Na-MES pH 6.0 (1M7 reactions) or 14.3M β-mercaptoethanol (DMS reactions), 1.5M NaCl, 6.25nM FAM-A20 Tail2 Primer, and 15% by volume hydrated and washed Poly-dT magnetic beads. The A20 Tail2 primer contains a poly A stretch, allowing it to hybridize to the modified RNA and purify it by binding the Poly-dT beads. Beads were separated by application of magnetic force for 7 min, and the solution was discarded. Beads were washed 3 times with 100μL of 70% ethanol. Samples were resuspended in 2.5μL of nuclease-free water.

Reverse transcription and capillary electrophoresis: An enzyme mix containing 2x “First strand buffer” (Life Technologies), 0.01M DTT, 1.6mM each dNTP, and 20U of SuperScript III Reverse Transcriptase was prepared and 2.5 μL of this mix were added to each 2.5μL modified sample. Samples were incubated at 42°C, 48°C, or 55°C for 30 minutes. 5μL of 0.4M NaOH was added to quench the reaction, followed by incubation at 90°C for 3 minutes, then incubation on ice for 5 minutes. 5μL of Acid Quench was added to neutralize the sample, then beads were separated by application of magnetic force, and the supernatant was discarded. Beads were washed three times with 70% ethanol then air-dried. Samples were resusupended in 10μL of Hi-Di-Formamide and incubated for 20 minutes, then removed from magnetic stand. Dilutions of the final sample were made into 8μL of Rox 350 (Life Technologies) size standard, transferred to a conical 96-well plate and sequenced on an ABI 3100 sequencer.

Data analysis: The sequencing data files were processed using the HiTRACE (Yoon et al. 2011) MATLAB scripts which are freely available as a github repository at https://github.com/hitrace/hitrace. The analysis procedure has been described in depth
previously (Kladwang et al. 2011b). Briefly, signals from multiple capillaries were aligned using the Rox 350 size standard. Then, sequence assignment and band annotation was performed. Quantification of bands is determined automatically by the program by automated peak-fitting of the capillary traces to Gaussian distributions. The Fold and SHAPEKnots executables in the RNAStructure package were used to infer secondary structures from the reactivities. Secondary structure images were prepared in VARNA.

**2D Mutate and Map SHAPE Probing**

Mutate-and-map primer design: An automated MATLAB script was used to design primers containing single mutations at each position in the CCR5 sequence of the CCR5.2hp_tail2 construct. The T7 promoter, reference hairpins, and tail sequences were unaltered. These “mutate primers” were ordered in 96-well format, with mutate primers for each primer number in a different plate (for example, primer 3 which is 39 bases long had 39 number of mutate primers in a single plate, which was different from the 55 number of mutate primers in different wells of a different 96-well plate for primer 4). The position of the mutate primer within the plate corresponds with the position of the mutation in the sequence. Mutations were made using “library 1”, which changes A to U, U to A, C to G, and G to C.

DNA template preparation by primer assembly: DNA templates were prepared essentially in the same way as the 1D experiment described above, with the exception that all reactions (here and throughout the remaining steps) were performed in 96-well
format, and incorporated the appropriate mutate primers with the appropriate WT primers 1-6.

**RNA synthesis and purification:** RNA was prepared and purified as described for the 1D experiment, maintaining the 96-well format of the experiment.

**RNA folding and chemical modification:** SHAPE modification of the RNA was performed exactly as described above, except without a “No Modification” control or ddNTP sequencing ladder, and the experiment was only performed in 10mM MgCl₂.

**Reverse transcription and capillary electrophoresis:** These steps were performed exactly as described above, except only the 48°C extension temperature was used. It was very important to remove residual ethanol during the wash steps carefully with a pipet, being cautious not to disrupt the beads or cross-contaminate the samples.

**Data analysis:** Data analysis was performed essentially as described above, however Z-scores were then extracted from the band intensities as previously described (Kladwang et al. 2011b). The Z-score is a difference in intensity signal from the mean across all mutants for a single nucleotide. Z-scores are then used for structure inference in Fold and SHAPEknots as described above with bootstrapping over 100 iterations of the analysis. Secondary structure images were prepared in VARNA.
CHAPTER III

AN UNSTRUCTURED REGION OF THE IGR IRES PLAYS MULTIPLE ROLES IN
TRANSLATION INITIATION

Introduction

The IGR IRESs contain three domains which are each defined by the presence of an RNA pseudoknot (Nishiyama et al. 2003). While domains I and II fold together and independently from domain III to create a ribosome binding element (Nishiyama et al. 2003; Costantino and Kieft 2005; Pfingsten et al. 2006; Jan et al. 2003), the pseudoknot that forms domain III of the IGR IRES structurally mimics a tRNA anticodon stem loop interacting with an mRNA codon (Fig. 3.1B) (Costantino et al. 2008; Zhu et al. 2011). This interaction effectively establishes the reading frame of the protein to be made, as the first codon of the ORF sits immediately downstream of the last nucleotide of the pseudoknot. The IRES therefore initiates translation by co-opting the elongation cycle of the ribosome and driving multiple rounds of translocation and tRNA delivery (Fig. 3.1A). Mutations to domain III have centered on abrogating the codon-anticodon mimicking interaction, which causes a failure to initiate translation even though ribosome

---

2 This chapter is taken with permission from: Ruehle, M. D. et al. “A dynamic RNA loop in an IRES affects multiple steps of elongation factor-mediated translation initiation.” eLife, in revision.
Figure 3.1 — IGR IRES Mechanism and Loop 3
A. Schematic of the IGR IRES initiation factor-independent translation initiation mechanism. The IGR IRESs occupy the same binding sites as tRNAs in the ribosome. Elongation factor-catalyzed steps are shown in red type and arrows, proposed reverse reactions are shown with gray arrows.

B. Secondary structure cartoon of an IGR IRES with domain III boxed and loop 3 in red. PKI denotes the pseudoknot base pairs that mimic the codon-anticodon interaction.

C. Cryo-EM reconstruction of the TSV IGR IRES bound to *S. cerevisiae* 80S ribosomes (Koh et al. 2014). The TSV IRES RNA model is shown in yellow, with loop 3 in red. Density within 8 angstroms of the IRES model is shown, at a threshold of 2.5. To the right is a close-up view of loop 3.

D. Same as panel C, but of a CrPV IGR IRES bound to *K. lactis* 80S ribosomes (Fernández et al. 2014). Density within 4 angstroms of the IRES model is shown, at a threshold of 2.5.

E. Same as panel C, but of a CrPV IGR IRES bound to *O. cuniculus* 80S ribosomes with eRF1 bound (Muhs et al. 2015). Density within 5 angstroms of the IRES model is shown, at a threshold of 3.0.

F. Diagram of the dual luciferase reporter RNA used in all *in vitro* translation assays. IRES activity is determined as a ratio of Firefly luciferase activity to Renilla luciferase activity.
recruitment is unaffected. Thus, domain III of the IGR IRES is essential for post-80S IRES function.

Although the structural similarity between domain III and an authentic codon-anticodon interaction is striking, an important difference between these is that domain III uses an intramolecular interaction (via the pseudoknot) to accomplish the mimicry of an intermolecular interaction. In order for this to work, domain III uses a single stranded loop to connect the tRNA-like part to the mRNA-like part of the domain. I refer to this loop as “loop 3” according to classical H-type pseudoknot nomenclature, although it has also been referred to as the variable loop region, or VLR, in other instances in the literature.

Previous chemical probing experiments have shown that loop 3 is highly dynamic (Jan and Sarnow 2002; Pfingsten et al. 2010; 2007). This is consistent with the crystal structure that was solved of domain III bound in the P site of the 70S ribosome (Zhu et al. 2011), which contained no discernable electron density for any of the nine CrPV or PSIV loop 3 residues. On one hand, this is intuitive; loop 3 is the only part of domain III that deviates from being a perfect codon-anticodon mimic and by remaining unstructured it is in a sense “invisible” to the ribosome. On the other hand, it is highly unusual that a loop in a pseudoknot is unstructured. Typically, loop 3 of pseudoknots form base triples in the minor groove of stem 1 resulting in a stable, compact fold. Furthermore, the ribosome’s decoding groove is a crowded place; various parts of the ribosomal RNA and proteins extend into the tRNA binding sites to control proper movement of tRNAs during the elongation cycle. It makes sense that loop 3 might
Figure 3.2 — IGR IRES Genomic Location, Conservation, and Structure

A. Diagram of the *Dicistroviridae* RNA genome. The IGR IRESs initiate translation of the second open reading frame. B. Alignment of domain III sequences from 14 *Dicistroviridae* family members (class I and II). The location of loop 3 is indicated in red. Conserved sequence is in bold. Abbreviations: Cricket Paralysis Virus (CrPV), Aphid Lethal Paralysis Virus (ALPV), Black Queen Cell Virus (BQCV), Drosophila C Virus (DCV), Himetobi P Virus (HiPV), *Homalodisca coagulata* Virus (HoCV), *Plautia stali* Intestinal Virus (PSIV), *Rhopalosiphum padi* Virus (RhPV), Triatoma Virus (TrV), Acute Bee Paralysis Virus (ABPV), Israeli Acute Paralysis Virus (IAPV), Kashmir Bee Virus (KBV), *Solenopsis invicta* Virus-1 (SInvV), Taura Syndrome Virus (TSV). C. A model of the CrPV IGR IRES from cryo-electron microscopy (magenta) bound to an 80S ribosome (PDB ID: 4CUX) (Fernández et al. 2014) overlaid with A-, P-, and E-site tRNAs (green) bound in a 70S ribosome in the presence of paromomycin (PDB ID 2WDK) (Voorhees et al. 2009). Domain 3 of the IRES is boxed in red and loop 3 is indicated with an arrow.
interact with these elements as it transits these same spaces, and that this may contribute to the manipulation of the ribosome that the IGR IRESs exert.

It is apparent that the IGR IRES initiation strategy involves unique, initiation factor-free ribosome recruitment, however there are several other steps downstream of ribosome binding—including three noncanonical translocation events—that must occur before an IRES-initiated, translating ribosome assumes a “normal” elongation cycle with only tRNA ligands. Little is known about the elements of the IRES that direct these post-80S steps. In the work described in this chapter, I asked: Is dynamic loop 3 important for IRES-driven translation, and if so, does it regulate post-80S steps in the initiation mechanism?

Results

Is Loop 3 Important for Translation?

In order to address whether loop 3 plays an important role in translation from the IGR IRESs, I made several types of mutations to loop 3 within the context of the CrPV IGR IRES and tested the effects of these mutations in a cell-free translation assay from rabbit reticulocytes. One set of mutations to loop 3 was made to elucidate the importance of the length of loop 3. I reasoned that loop 3 may exhibit flexible character in part because the number of residues in the loop provides more length than the minimum distance it needs to cover to connect the tRNA-like part of the structure to the codon-like part, akin to having slack in a rope. By deleting residues from loop 3, I might decrease the amount of slack in the loop, thus decreasing its flexibility and dynamics. In
Figure 3.3 — Function of diverse WT and loop 3 mutant IGR IRESs in RRL

A. Activity of different WT IGR IRESs. Mutant CrPV-K/O has pseudoknots III and I disrupted and is the negative control (Jan et al. 2003; Costantino and Kieft 2005).

B and C. Function of WT IRESs (black bars) and loop 3 mutants (gray bars). WT levels are normalized to 1 for each IRES.

D. Diagrams of CrPV IGR IRES domain III mutants. Mutations are boxed and X indicates deletion of a nucleotide.

E. Activity of CrPV loop 3 mutants in RRL. Error bars represent standard error of the mean over at least three biological replicates.
this set of mutations, I deleted one, two, or three residues (Δ1, Δ2, Δ3 mutants) with eight, seven, and six residues, respectively, remaining in the loop (Fig. 3.3D). In another set of mutations, I tested whether the sequence of the loop is important. To do this, I introduced point mutations to several positions in loop 3 without altering the length. The GGC mutant was made because I noticed from a sequence alignment of the nine Class I IGR IRESs that A6205, A6207, and U6211 were nearly perfectly conserved. I therefore mutated these positions to a G, G, and C respectively, maintaining the original purine or pyrimidine class to assess the identity of the residue itself, and not simply the amount of space it might occupy. I also observed that loop 3 has high adenosine content, and addressed this characteristic in the G-rich mutant. The high adenosine content is interesting because adenosine residues are commonly observed to participate in base triples through the minor groove of another helix. If loop 3 makes a minor groove interaction as seen with most H-type pseudoknots its high adenosine content could mediate these contacts. Therefore, I mutated the non-conserved residues A6204, A6208, and A6209 to Gs (Fig. 3.3D).

Bicistronic mRNAs were generated containing the WT or loop 3 mutant CrPV IGR IRES between an upstream *Renilla* luciferase gene and a downstream firefly luciferase gene (Fig. 3.1F). Translation of these mRNAs revealed that deleting a single nucleotide from loop 3 has a modest defect in translation (roughly 80% of WT), but deletion of two or three nucleotides abolishes IRES activity (Fig. 3.3E). Similarly, the G rich and GGC mutants are severely decreased in their IRES activity, although the GGC
mutant still operates at about 20% of WT. From these data I concluded that loop 3 is important for CrPV IGR IRES-driven translation.

To determine whether this was a widespread and conserved role of loop 3 in all IGR IRESs, I mentored a rotation student, Ryan Sheridan, who first assessed the relative translation efficiencies of several different WT IGR IRESes from both the *Cripa*- and *Apara*- virus subfamilies (representing the Class I and Class II IGR IRESs) (Fig 3.3A, Table 3.2) and then made similar types of mutations to loop 3 (Fig. 3.3 B and C, Table 3.2). Strikingly, despite differences in the sequence of loop 3 and overall secondary structure of domain III, shortening of loop 3 and point mutations all negatively impact translation from the IGR IRESs. Thus, the functional importance of loop 3 is conserved throughout the *Discistroviridae* family IGR IRESs.

**Table 3.1 Loop 3 composition and length in diverse IGR IRESs**

<table>
<thead>
<tr>
<th>Virus</th>
<th>Loop 3 length (# of bases)</th>
<th>Loop 3 A content</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrPV</td>
<td>9</td>
<td>56%</td>
</tr>
<tr>
<td>ALPV</td>
<td>6</td>
<td>50%</td>
</tr>
<tr>
<td>BQCV</td>
<td>8</td>
<td>50%</td>
</tr>
<tr>
<td>DCV</td>
<td>9</td>
<td>67%</td>
</tr>
<tr>
<td>HiPV</td>
<td>9</td>
<td>56%</td>
</tr>
<tr>
<td>HoCV</td>
<td>10</td>
<td>40%</td>
</tr>
<tr>
<td>PSIV</td>
<td>9</td>
<td>56%</td>
</tr>
<tr>
<td>RhPV</td>
<td>7</td>
<td>43%</td>
</tr>
<tr>
<td>TrV</td>
<td>9</td>
<td>22%</td>
</tr>
<tr>
<td>ABPV</td>
<td>9</td>
<td>56%</td>
</tr>
<tr>
<td>IAPV</td>
<td>9</td>
<td>56%</td>
</tr>
<tr>
<td>KBV</td>
<td>9</td>
<td>56%</td>
</tr>
<tr>
<td>SinV</td>
<td>10</td>
<td>50%</td>
</tr>
<tr>
<td>TSV</td>
<td>8</td>
<td>40%</td>
</tr>
</tbody>
</table>
Table 3.2. Activity of IGR IRESs in RRL and mutations tested

<table>
<thead>
<tr>
<th>Virus</th>
<th>WT activity</th>
<th>Loop 3 mutants activity*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>G-rich</td>
</tr>
<tr>
<td>Class I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrPV</td>
<td>++++</td>
<td></td>
</tr>
<tr>
<td>HiPV</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>HoCV</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>PSIV</td>
<td>+</td>
<td>UUAGGGGCCG</td>
</tr>
<tr>
<td>Class II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABPV</td>
<td>++++</td>
<td></td>
</tr>
<tr>
<td>IAPV</td>
<td>++++</td>
<td>GAUGGGCCA</td>
</tr>
<tr>
<td>KBV</td>
<td>++</td>
<td>GAAUGGCCG</td>
</tr>
<tr>
<td>SinV</td>
<td>++++</td>
<td></td>
</tr>
<tr>
<td>TSV</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

*Site of mutation is shown in bold italics and underlined. Site of deletion is shown as a dash. Abbreviations: Cricket Paralysis Virus (CrPV), Himetobi P Virus (HiPV), Homalodisca coagulata Virus (HoCV), Plautia stali Intestinal Virus (PSIV), Acute Bee Paralysis Virus (ABPV), Israeli Acute Paralysis Virus (IAPV), Kashmir Bee Virus (KBV), Solenopsis invicta Virus-1 (SInV), Taura Syndrome Virus (TSV).

Can Loop 3 Mutants Bind to the Ribosome?

Domain III has previously been shown to be dispensable for IRES binding to ribosomes (Costantino and Kieft 2005), however it is possible that loop 3 mutations introduce an unforeseen effect that could alter ribosome recruitment, leading to the decreased protein synthesis observed in the translation assays above. To verify that the loop 3 mutant IRESs can bind 80S ribosomes, I performed assembly assays in which I radiolabeled IRES RNAs, incubated them in rabbit reticulocyte lysate (RRL) to allow for ribosome assembly on the IRES, and then separated the lysate over a sucrose gradient via ultracentrifugation. Dense complexes like ribosomes travel into the gradient, carrying the labeled IRES RNA if it is bound, whereas unbound RNAs and other lighter components of the lysate remain at the top of the gradient where the solution is less dense.
Performing this experiment with the WT and loop 3 mutants reveals that loop 3 mutation has no effect on 80S ribosome assembly. The only exception to this is that the G-rich mutant reproducibly exhibited a decreased 80S peak (Fig. 3.4A).

To address this question using a complimentary experiment, I also performed ribosome binding assays over a short time course with purified ribosomal subunits from yeast and shrimp and measured association with the IRES using a filter binding method. This experiment revealed that there is no difference in the approximate rate of ribosome binding between WT and any of the mutants tested (Fig. 3.4B). In addition, a competition experiment in which radiolabeled WT or G rich IRESs were allowed to reach binding equilibrium with purified ribosomes, followed by addition of excess, cold IRES to compete off the labeled RNA showed that there is no difference in the approximate off-rate between these two IRES constructs (Fig. 3.4C).

One caveat of the filter binding experiments is that it is impossible to distinguish 40S-bound IRES from 80S-bound IRES species. It is possible that this explains the discrepancy between these results and those of the assembly assays for the G-rich mutant. However, later experiments also argue against the G-rich mutant failing to initiate translation as a result of defective ribosome binding. Taken together, the assembly assays and the filter binding assays show that the loop 3 mutants are equally capable of ribosome recruitment, pointing to loop 3 playing an important role in a post-80S step of the IGR IRES initiation mechanism.
Figure 3.4 — Loop 3 Mutant Ribosome Binding Assays

A. Assembly of 80S ribosomes on CrPV IGR IRES loop 3 mutants in rabbit reticulocyte lysate. Radiolabeled IRES RNAs were incubated in RRL for 5 min before separation of initiation complexes on a 15-30% sucrose gradient. Free, 40S-, and 80S-bound IRES complexes are indicated. These experiments did not contain cycloheximide.

B. Approximate on- and off-rates of IRES-ribosome binding measured by filter binding. The on-rate experiment measures the association of IRES with ribosomes or ribosomal subunits as a function of time. Pure shrimp ribosomes were used for the on-rate experiment. The off-rate experiment used unlabeled competitor IRES RNA to detect dissociation of IRES from ribosomes as a function of time. Purified yeast subunits were used for the off-rate experiment.
Does Loop 3 Promote Translocation?

After the IGR IRES assembles an 80S ribosome, domain III is placed in the A site where new tRNAs are delivered to decode the next codon in the mRNA (data to support this statement is relatively new in the field; previously domain III was thought to bind in the P site. For more discussion on this subject, see Chapter I and the section below). In order for the first codon of the ORF to enter the A site for decoding, domain III must be translocated from the A site to the P site. Following intervening tRNA delivery events, the IRES must go through two more rounds of translocation (from P to E and from E to no longer associating with the ribosome) before the ribosome resembles a canonical elongation complex, with only tRNA ligands bound (Fig. 3.1A).

If loop 3 regulates a post-80S step in the initiation mechanism, investigating whether loop 3 mutants can perform these noncanonical translocation events could illuminate which step is being controlled. Therefore, to test translocation in these mutants, I performed a toeprinting experiment to map where the edge of the ribosome is on the IRES mRNA. This assay also has the benefit of showing whether 80S ribosomes are properly positioned on the IRES. WT CrPV IGR IRES with pure yeast 40S and 60S ribosomal subunits yields a “pretranslocated” toeprint band at the +14/15 position (where the +1 position is defined as the 5’ most nucleotide of the codon in the P site) (Fig. 3.5A lanes 2 & 18). In RRL supplemented with the elongation inhibitor cycloheximide (CHX) the WT IRES translocates twice (+20/21 toeprint, Fig. 3.5A lanes 3 & 18) as previously observed (Jan et al. 2003). Without CHX no strong toeprints are seen,
Figure 3.5 — Ribosome Docking, Translocation, and Reading Frame Maintenance

A. Toeprinting analysis of CrPV WT IRES and loop 3 mutants in the free (f) and yeast 80S ribosome-bound (80S) forms, and in RRL with or without 3 mg/ml cycloheximide (+/− CHX). The +14/15 toeprint indicates the position of the edge of the pretranslocation ribosome, and the +20/21 toeprint shows the position of the edge of the 2x translocated ribosome. Gels are representative of at least six independent experiments. B. Quantification of translocated toeprint bands (+20/21/((+14/15)+(+20/21))) in RRL+CHX (n = 6−9), error bars represent standard error of the mean. C. In vitro translation assay of dual luciferase reporters with +0 (normal), +1, or +2 reading frames. Error bars represent standard error of the mean of three independent experiments.
indicating that the antibiotic traps IRES-ribosome complexes that can be observed in this assay.

Like WT, all length mutants (Δ1, Δ2, Δ3) have a pretranslocated toeprint at +14/15 when bound to pure ribosomal subunits, indicating these IRESs are correctly positioned within the decoding groove of 80S complexes (Fig. 3.5A lanes 6, 10, 14). However, in RRL the loop 3 length mutants retain the +14/15 toeprint both with and without CHX, showing that pseudotranslocation is inhibited (lanes 7, 8, 11, 12, 14, 15). A mutation that abrogates codon-anticodon base pairing in PKI does not generate a pretranslocation toeprint at all (Jan et al. 2003), indicating the tested mutations likely do not disrupt pseudoknot formation. Furthermore, the +20/21 toeprint is decreased in the Δ2 mutant, and is completely missing in the Δ3 mutant. Quantification of the change in toeprint band intensities over many replicates shows that the loss of translation activity by shortening loop 3 is due to inhibiting pseudotranslocation (Fig. 3.5B). The G-rich and GGC sequence mutants also form 80S complexes that are properly positioned at the +14/15 location (Fig. 3.5A lanes 22 and 26). The G-rich pretranslocation toeprint is reproducibly less intense (lane 22) than WT and all other mutants, likely indicating a decreased binding to ribosomes (consistent with the assembly data in Fig. 3.4A). However, in RRL the sequence mutants quantitatively match WT’s ability to generate a strong +20/21 band (lanes 23 and 27), suggesting they can translocate (Fig. 3.5B). To verify the results with CHX, I performed toeprinting with a different translocation inhibitor, hygromycin B, which binds the ribosome in a different location and has a different mechanism of action than CHX (Fig. 3.6A). The WT, G-Rich, and GGC mutants
pseudotranslocate once (+17/18 toeprint), but the length mutants do not execute the first pseudotranslocation event. To our knowledge this is the first demonstration of an antibiotic trapping the IGR IRES after one round of pseudotranslocation, making it a useful tool to study this event and will be discussed in Chapter V.

To identify the step at which the G-rich and GGC mutants are inhibited, I adapted the toeprinting assay to look at rounds of translocation after the two allowed by CHX. Dilute hygromycin B was added to RRL after IRES RNA (in previous experiments, RRL was pretreated with high concentrations of hygromycin B or CHX). WT IRES toeprinting shows 4-5 rounds of translocation (Fig. 3.6B, lane 2). As expected, Δ1 behaved similarly to WT while the Δ2 and Δ3 mutants did not proceed past the initial binding location (lanes 6 and 8). Surprisingly, the sequence mutants displayed toeprinting patterns similar to WT (lanes 10 and 12) although the bands generated from the first few rounds of translocation are less intense (Fig. 3.6B, right). Thus, the G-rich and GGC mutants can translocate at least 4-5 times in RRL, and the source of their reduced translation initiation activity must be more subtle than a complete failure to translocate. Although all of the mutants showed defects in translation initiation (Fig. 3.3), the toeprinting data indicate that the reasons differ between the length and sequence mutants. The G-rich and GGC mutants do not block translocation while the length mutants do, indicating loop 3 has two independent roles in IGR IRES driven translation initiation.
Figure 3.6 — Toeprinting with Hygromycin B
A. Toeprinting analysis in RRL without or with 0.66 mg/mL hygromycin B (-/+). (B) Toeprinting analysis in RRL without or with 3.33 µg/mL hygromycin B (-/+ added after 1 min of incubation of the IRES in lysate. Normalized traces of the WT, Δ3, G-rich, and GGC IRES RNAs in RRL+ hygromycin B are shown at right. Image is from a single gel, asterisk indicates where two irrelevant lanes were removed.
Is the Correct Reading Frame Maintained?

The ability of the G-rich and GGC mutants to translocate in the toeprinting assays suggests they disrupt a different process than do the length mutants. Domain III is essential for establishing the proper reading frame, so the mutations might induce the ribosome to initiate translation out-of-frame. To test this, I measured translation using constructs with one or two additional nucleotides inserted immediately before the AUG of the firefly luciferase open reading frame (+1 and +2 frames), which could rescue out-of-frame initiation (Fig. 3.5C). Neither alternate frame rescues IRES activity in the G-Rich or GGC loop 3 mutants, indicating the G-rich and GGC mutants do not induce out-of-frame initiation.

What Is the Relative Rate of Tetrapeptide Synthesis in the Sequence Mutants?

If the G-rich and GGC mutants initiate in-frame and can translocate at least 4 times as indicated by the toeprinting assay, why is their translation activity decreased? It is unlikely that loop 3 acts after the IRES no longer interacts with the ribosome, the presumed situation after 4 translocation events. Alternatively, decreased toeprint band intensity in these mutants (Fig. 3.6B lanes 10 and 12) suggested there could be subtle changes in kinetics of the translocation events. Because toeprinting is not an ideal assay to examine this, we collaborated with Dr. Barry Cooperman’s group at the University of Pennsylvania to directly explore differences in the rate of peptide synthesis between the WT and the sequence mutants in an in vitro reconstituted translation system (shrimp ribosomes and yeast elongation factors) by quench-flow (diagrammed in Fig. 3.7).
Figure 3.7 — Schematic Overviews of Experiments Performed in the Reconstituted System

Data from these experiments is presented in Figure 5. Top: Tetrapeptide formation assay. Initiation complexes with the IGR IRES and A. salina (shrimp) ribosomes were first assembled before adding ternary complexes (TCs) consisting of F-, V-, and K-charged tRNAs with eIF1A-GTP (the tRNA delivery factor) and eEF2-GTP (the translocase) to form tri-peptides. These complexes were then combined with $^{35}$S-Met-tRNA$^{Met}$ TC, quenched with strong base, and resultant peptides were analyzed by thin layer electrophoresis. Bottom left: P site tRNA binding by cosedimentation. Bottom right: A site tRNA binding by anisotropy. Details of the assays can be found in the methods section.
Because toeprinting suggested at least 4 rounds of translocation on the G-rich and GGC mutants, we first assayed the rate of conversion of tripeptide to tetrapeptide with the coding sequence for the peptide FVKM placed downstream of the IRES. Compared to WT, both the G-rich and GGC mutants displayed substantially decreased abilities to convert tripeptide to tetrapeptide, at levels that reflected their relative translation activities (Fig. 3.9A). These data suggest that the loss of translation activity in the loop 3 sequence mutants is imparted by at least one defective elongation step at or preceding tetrapeptide formation.

**Is tRNA Association With IRES-Ribosome Complexes Altered When Loop 3 Is Mutated?**

The decreased peptide synthesis described above could result from inhibition of any step preceding tetrapeptide formation, including binding of the first ac-tRNA to the IRES-80S ribosome complex. To measure the efficiency of this step, we delivered $[^3]$H]Phe-tRNA$^{_{Phe}}$ to WT and mutant 80S-IRES (coding for FVKM) ribosome complexes in the presence of eEF1A-GTP (which forms a ternary complex, TC, with ac-tRNA) and eEF2-GTP and collected these complexes by ultracentrifugation through a sucrose cushion (diagrammed in Fig. 3.7). As expected, ac-tRNA delivered by eEF1A and translocated to the P site by eEF2 bound stably enough to survive this purification, whereas A-site associated ac-tRNA did not (Fig. 3.8) (Yamamoto et al. 2007). Furthermore, ac-tRNA delivery and binding to the P site depended on a cognate codon-tRNA anticodon interaction (Fig. 3.8). Therefore, this assay measures the efficiency of completion of all three eEF-dependent steps (Fig. 3.1A). As expected, stable $[^3]$H]Phe-
Figure 3.8 — Codon- and Factor-Dependent tRNA Binding to IRES-80S Complexes

The graph depicts the amount of Phe-tRNA$^{3}$H or Arg-tRNA$^{3}$H recovered after cosedimentation with shrimp 80S ribosome complexes through a sucrose cushion. The IRES RNA construct is from the WT CrPV IGR IRES, but initiates with a UUC (Phe) codon. The identity and source of the isolated tRNA is indicated beneath the bars. The presence or absence of factor(s) and IRES RNA is indicated beneath the graph. Arg-tRNA$^{\text{Arg}}$ was used as a control for association of non-cognate tRNA. Phe-tRNA$^{\text{Phe}}$ from yeast or E. coli stably associates with the ribosomes in an eEF1A- and eEF2-dependent manner, while non-cognate Arg-tRNA$^{\text{Arg}}$ did not. Omitting either elongation factor greatly decreased stable binding. Error bars represent one standard deviation from the mean.
tRNA\textsuperscript{Phe} binding was observed with WT IRES with eEF2 (Fig. 3.9B), consistent with previous reports (Yamamoto et al. 2007). When mutants Δ1, Δ2, and Δ3 were assayed, they showed a progressive decrease in bound \[^{3}\text{H}]\text{Phe-tRNA}^{\text{Phe}}. Interestingly, the G-rich and GGC mutants also showed decreased P-site ac-tRNA association with IRES-80S ribosome complexes at levels that mirror their relative translation activities. Therefore, mutations to loop 3 length and base composition cause decreased association of the first ac-tRNA in the P site.

Because eEF2-GTP was included in the above experiment, we could not distinguish whether decreased ac-tRNA association in the P site resulted from reduced eEF2-driven pseudotranslocation of domain III from the A site to the P site, subsequent ac-tRNA delivery to the A site, or the second pseudotranslocation that moves ac-tRNA from the A site to the P site. To help discriminate between these possibilities, we employed a fluorescence anisotropy experiment in which proflavin-labeled \text{Phe-tRNA}^{\text{Phe}}[\text{Phe-tRNA}^{\text{Phe}}(prf)] TC was delivered to WT and mutant IGR IRES-80S ribosome complexes in the absence of eEF2 (diagrammed in Fig. 3.7). The measured anisotropy of unbound Phe-tRNA\textsuperscript{Phe}(prf) was 0.205 +/- 0.002. As expected, addition of eEF1A-GTP to the ac-tRNA resulted in an increase in measured anisotropy to 0.210 +/- 0.003, consistent with formation of the eEF1A+GTP+Phe-tRNA\textsuperscript{Phe}(prf) ternary complex (TC). Addition of empty 80S ribosomes (lacking an mRNA or IRES, indicated as “no IRES”) resulted in only a slight increase in the change in anisotropy relative to the TC alone (Fig. 3.9C). However, when a complex of CrPV IGR IRES bound to 80S ribosomes was added
Figure 3.9 — Identification of Specific Translocation Steps Affected by Loop 3
Characterization of early steps in IGR IRES initiation in a reconstituted translation system, using purified shrimp ribosomes and yeast elongation factors. A. Time course of tetrapeptide formation from tripeptide. Data are representative of two independent experiments. B. \(^{[3}H\)Phe-tRNA\(^{Phe}\) binding to the P site in the presence of eEF2. Triplicate reads were averaged and normalized to set WT equal to 1. C. Anisotropy measurements of Phe-tRNA\(^{Phe}\) (prf) binding to IRES-80S ribosome complexes. For each set of experiments performed, a determination was made of the anisotropy difference (\(\Delta\)) between free TC and TC added to the WT IRES-80S complex, and differences between TC added to other complexes and free TC were normalized to this value. Error bars represent one standard error from the mean of 2-4 replicates. D. Translocation efficiency of ac-tRNA from the A to the P site in the \(\Delta1\) and \(\Delta2\) mutants. Data were normalized to set the anisotropy-based A site binding levels (data from C) to 1, and those factors were applied to the cosedimentation-based P site binding levels (data from B).
to the TC, we observed a much larger increase in anisotropy, to 0.272 +/- 0.006. This change in anisotropy between TC alone and in the presence of 80S ribosomes+IRES (0.061 +/- 0.003) is consistent with delivery of ac-tRNA to the A site of the IRES-80S ribosome complex by the TC.

To verify that IRES-dependent delivery of tRNA was specific for the first codon following the IRES, we delivered ac-tRNA to an IRES-80S ribosome complex in which the UUC codon for tRNA$^{\text{Phe}}$ was replaced by the non-cognate GCU codon (“non-cognate”, Fig. 3.9C). This resulted in a smaller increase in anisotropy compared to the IRES with a cognate Phe codon, but larger than the “no IRES” control. Importantly, the observation that eEF2-independent ac-tRNA binding to the ribosome requires a cognate codon is consistent with the idea that the first codon is entering the A site and is queried by the ac-tRNA anticodon. This supports the idea that domain III can spontaneously move to the P site to some degree, perhaps akin to the observed ability of tRNAs to undergo slow spontaneous translocation on bacterial ribosomes (Gavrilova et al. 1976; Gavrilova and Spirin 1971; Pestka 1969; Southworth et al. 2002; Fredrick and Noller 2003; Moore 2012; Robertson and Wintermeyer 1987; Semenkov et al. 1992). The nature of the ac-tRNA’s association with the ribosome likely differs depending on whether an IRES RNA with a non-cognate or cognate codon is present; the former probably represents transient TC interaction with the tRNA in a A/T state during a decoding step, the latter likely represents full and longer-lived accommodation of the tRNA into the A/A state.

The results outlined above validate the use of this assay to explore the effect of loop 3 mutations on ac-tRNA association with the IRES-ribosome complex independent
of eEF2 activity. Mutants Δ1, Δ2, and Δ3 showed a progressive decrease in anisotropy (Fig. 3.9C), following the trend established by the translation initiation data. These data indicate that these mutants have a defect in initial ac-tRNA binding; in the case of Δ3, this defect is more severe than the effect of a non-cognate codon. This may be because the movement of the first codon into the A site has been compromised. ac-tRNA delivery to IRES-80S ribosome complexes with the Δ1 and Δ2 mutants was less than to WT, but equal to or greater than to the IRES with a non-cognate codon. To approximate the percentage of these A-site ac-tRNAs that successfully translocated to the P site, I normalized their P site binding levels to the A site interaction levels (Fig. 3.9D). For Δ1, the percentage is ~80% while for Δ2 it is ~25%. When we consider these data in light of the proposed mechanism of IGR IRES-driven initiation (Fig. 3.1A), they suggest that these mutants have defects in both pseudotranslocation events and these defects become progressively worse as loop 3 is shortened. In contrast, the G-rich and GGC mutants display ac-tRNA binding similar to the WT IRES (Fig. 3.9C). Thus, the defect in these sequence mutants is restricted to the second pseudotranslocation event which moves ac-tRNA from the A site to the P site, and domain III from the P site to the E site. Taken together, the data from all mutants suggest that loop 3 has two independent functions to facilitate two elongation factor-driven steps, which depend on loop 3 length and base composition.
Do the Loop 3 Length Mutants Fail Because They Cannot Execute the First Pseudotranslocation or Because tRNA Cannot Be Delivered After the First Pseudotranslocation?

The anisotropy data show that loop 3 is important for initial ac-tRNA association with the ribosome, but do not directly address eEF2’s role in this process. The decreased ac-tRNA association in mutant IRES-80S ribosome complexes observed in the anisotropy experiment could result from a decrease in domain III spontaneous vacating of the A site, or from decreased TC association even if the A site is available after domain III moves to the P site. To address this, I collaborated with Dr. Ruben Gonzalez’s group at Columbia University to use single-molecule total internal reflection fluorescence microscopy to directly visualize the colocalization of Cy5 fluorophore-labeled Phe-tRNA$^{\text{Phe}}$ with Cy3 fluorophore-labeled IRES-80S ribosome complexes that had been tethered (via the IRES RNA) to the surface of a microfluidic observation flowcell (Fig. 3.10). This colocalization data reports on the ac-tRNA occupancy of the 80S-IRES ribosome complexes. We chose WT and Δ3 IRESs to study as they exhibited the most differing behaviors in the previous experiments. As expected, addition of just Phe-tRNA$^{\text{Phe(Cy5)}}$+GTP (without eEFs) to 80S-IRES ribosome complexes, followed by incubation and subsequent flushing of the flowcell to remove unbound ac-tRNA, revealed very low ac-tRNA occupancies for both WT and Δ3 IRESs (Fig. 3.11). When GTP+eEF2 was included with the Phe-tRNA$^{\text{Phe(Cy5)}}$ (but no eEF1A) the ac-tRNA occupancy of the IRES-80S ribosome complexes formed with WT IRES increased to 9.7 ± 2.5%, consistent with a low, but enhanced level of eEF1A-independent ac-tRNA binding.
Figure 3.10 — Single-Molecule Colocalization Experiment

A. 80S complexes, assembled on IRES (Cy3) molecules (hybridized to a biotinylated DNA), were tethered to microscope slide surfaces coated with PEG-Biotin, via a streptavidin bridge. In the cartoon shown here, the IRES has undergone the first pseudotranslocation event so that domain 3 sits at the P site and the Phe-\text{tRNA}_{\text{Phe}}^{\text{Cy5}} is delivered to the A site by eEF1A, thereby generating a surface-tethered complex with spatially colocalized Cy3 and Cy5 spots. Sample green and red frames, corresponding to the imaged Cy3 (false colored as green) and Cy5 (false colored as red) spots, are depicted in B and C, respectively, while D represents the superposition of the two frames; regions with colocalized green and red spots are false colored as yellow for visual clarity. The panels below B-D show a representative region from the corresponding frames, magnified 6X, to demonstrate the well-resolved distribution of spots and the precision of colocalization.
Figure 3.11 — Single-Molecule Colocalization of IRES and tRNA Within 80S Ribosomes

Effect of eEF2 on colocalization of Phe-tRNA$^{phe}$ (Cy5) with individual 80S ribosome-IRES complexes formed with either WT(Cy3) IRES or Δ3(Cy3) IRES. Addition of elongation factors and Phe-tRNA$^{phe}$ (Cy5) (tRNA(Cy5)) to 80S ribosome-IRES complexes formed with either (left) WT(Cy3) IRES (black bars) or (right) Δ3(Cy3) IRES (gray bars) are depicted as percent Cy3-Cy5 colocalized spots. The presence or absence of factor(s) is indicated beneath the graphs and error bars represent one standard deviation from the mean. Elongation factors and ribosomes are from yeast.
When this experiment was repeated with the Δ3 IRES, we observed a lower ac-tRNA occupancy (1.5 ± 1.1%) compared to the WT IRES. Higher eEF1A-independent, but eEF2-dependent, ac-tRNA occupancy on WT IRES complexes compared to Δ3 IRES complexes suggests that the difference between these two IRESs in the anisotropy experiment (Fig. 3.9C) is not due to altering eEF1A function. Rather, those data may indicate a decrease in clearing of the A site by the Δ3 mutant, suggesting the Δ3 mutant’s main defect is in the first pseudotranslocation and not in the A-site ac-tRNA binding event itself.

To examine eEF1A-dependent ac-tRNA delivery, we assembled TC with Phe-tRNA^{Phe}(Cy5)+eEF1A+GTP and delivered this to the immobilized IRES-80S complexes without eEF2. Compared to the reactions lacking eEF1A, both IRESs show increased and similar ac-tRNA occupancies (WT: 17.9 ± 4.8%, Δ3: 20.8 ± 5.4%). These data initially seem at odds with the anisotropy data in which eEF2-independent ac-tRNA association with 80S-WT IRES ribosome complexes is much greater than complexes with Δ3. This apparent discrepancy is likely due to the fact that anisotropy data are obtained under equilibrium conditions where transient interactions are observed, whereas the single-molecule fluorescence data are collected after the flowcell is flushed and thus only show stable long-lived association. Combining the data from both experiments reveals that eEF2-independent ac-tRNA association to WT IRES-80S ribosomes is transient and is inhibited by the Δ3 mutation.

Finally, when eEF2+GTP+TC was delivered to the tethered 80S-IRES ribosome complexes, we observed a dramatic increase in the ac-tRNA occupancy on complexes formed with the WT IRES (82.8 ± 15.7%), but not with the Δ3 IRES (26.6 ± 10.9%).
demonstrates that the Δ3 mutation inhibits the IRES-ribosome complex from using eEF2 to facilitate stable ac-tRNA delivery. Overall, our data suggest that loop 3 is important for eEF2’s ability to catalyze both pseudotranslocations, the first of which moves domain III to clear the A site for ac-tRNA binding and the second which moves the first ac-tRNA to the P site.

**Comparison of Results in Lysate and Reconstituted Systems**

The toeprinting experiments performed in RRL and experiments conducted with reconstituted systems show some differences. Specifically, toeprinting with the G-rich and GGC mutants in RRL+CHX shows at least two rounds of translocation (Fig. 3.5A) and at least four in RRL+ hygromycin B at low concentrations and post-treatment (Fig. 3.6B). However, in the reconstituted assays these mutants fail before two rounds of pseudotranslocation (Fig. 3.9B). I consider it unlikely that this discrepancy is due to differences in the species of ribosomes used (purified subunits were made from yeast and shrimp sources, versus rabbit subunits in RRL) because IGR IRESs function in diverse systems and contact highly conserved ribosome features. A more likely possibility is that the presence or effective concentrations of various components (ribosomes, ac-tRNAs, GTP, or unidentified factors) is different in the lysate as compared to the reconstituted system, which may alter the kinetics of the translocation reactions. In addition, the presence of antibiotics such as CHX or hygromycin B (which I only used in RRL-based experiments) may suppress the effects of sequence mutation to loop 3 by altering ribosome conformational dynamics (Wilson 2014). Despite this uncertainty, taken
together my data clearly identify loop 3 as important in more than one round of pseudotranslocation and also illustrate the importance of employing multiple experimental approaches.

**Discussion**

To function, IGR IRESs must have affinity for the ribosome, promote subunit joining, manipulate elongation factor action, and move through all of the tRNA binding sites. In this body of work I have shown that conformationally dynamic loop 3 in the tRNA-mimicking domain controls two independent, non-canonical translocation events, demonstrating how a viral RNA can carry out intricate ribosome manipulation using dynamic RNA structure. This strengthens the previously postulated idea that structured regions are important for overall IGR IRES architecture and ribosome positioning, whereas conformationally dynamic regions help drive the IRES through the ribosome in elongation factor-dependent steps to initiate translation (Pfingsten et al. 2010). The strategy of using a combination of conformationally flexible elements with stably structured domains is likely a strategy used by many RNAs that control dynamic cellular machines.

My data show that the length and sequence of loop 3 are both important for function. A previous study also examined the effect of loop 3 length and sequence on IGR IRES translation efficiency (Au and Jan 2012) using similar dual luciferase translation assays and a slightly different application of the toeprinting experiment. That work showed that increases in loop 3 length were tolerated up to four additional nucleotides,
but further increases were detrimental to IRES activity in the translation assay. I did not explore the role of loop 3 length increases in my work. Au and Jan also made mutations to shorten the length of loop 3 by deleting 1 or 2 nucleotides (but not 3) and showed similar translation defects as I show here, but also showed that the location of the 1 nucleotide deletion within the loop had an effect on the severity of the defect. In toeprinting assays, the deletion mutations that were tested were capable of generating a pretranslocation toeprint at the same position as the WT RNA using the antibiotic edeine as an inhibitor in the lysate, but the intensity varied from 61-126% of WT. This suggests that toeprint intensity does not correlate with translation activity, and this is consistent with our data. This highlights a central problem with the toeprinting experiment, which is that some reports use the toeprint as an indicator of ribosome binding and even is extended as a semi-quantitative measure of ribosome binding efficiency. In other instances, the toeprint is used as an indicator of ribosome positioning of the RNA at nucleotide-level resolution within the decoding groove. We feel the latter is the only appropriate interpretation, and that extension of the assay as a semi-quantitative measure of ribosome binding is unjustified due to the many variables associated with the experiment and processing of the samples (different extension and cDNA extraction efficiencies, for example). Illustrating this point, in Fig 3.5A two gels are shown that each have the same set of conditions for the WT IRES RNA. These experiments were performed on the same day and the WT samples were generated as true replicates. However, there is clearly variability in the intensity of the toeprints between replicates, even though the position of the toeprints are the same. These
observations underscore that there is a degree of variability in the intensity of the bands in the toeprint experiments that should be expected, and extracting quantitative results should be done only after normalization to the total signal in the lane (at least), and with averaging over multiple replicates (as we do in Fig. 3.5B). Taking this into consideration, it is unclear in the experiments presented by Au and Jan whether the mutations they made altered 80S binding efficiency since they did not use an appropriate assay to measure ribosome binding directly. Au and Jan conclude that “ribosome positioning” is affected by these mutations, but the position of the toeprint does not change in their experiments, only the intensity.

Au and Jan also investigated the role of the identity of nucleotides in the loop via translation and toeprinting assays. They made single point mutations to 4 different conserved positions within the loop (A6205, A6207, A6208, and U6211), systematically altering them to each of the other three nucleotides, and then testing the effect of these mutations in the dual luciferase functional assay. For reference, the G-rich mutation I made altered A6204, A6208, and A6209 and the GGC mutation altered A6205, A6207, and U6211. The mutations from Au and Jan had variable functional affects, whereas both the G-rich and GGC mutants had deleterious effects on translation. Mutation of A6207 to any other nucleotide had no effect on IRES activity. Mutation of A6205 to any other nucleotide led to roughly 50% loss of activity. Mutation of A6208 to G decreased IRES activity to 70%, although C or U mutations had little impact. Similarly, U6211 mutation to G decreased IRES activity to 30%, whereas mutation to C permitted 80% activity and mutation to A stimulated translation to nearly
140% of WT. Toeprinting experiments were only performed on the A6205C/G/U and A6207C/G/U mutants and shows that the pretranslocation (with edeine) and 2x translocated (with cycloheximide) toeprints were decreased in intensity in each of the A6205C/G/U mutants, but not in the A6207C/G/U mutants, which also correlates with translation levels. However, Au and Jan conclude that A6205 affects ribosome positioning, even though the position of the toeprint (albeit lighter) is the same as WT. Again, binding of these mutants to ribosomes was not tested directly, and no further work was done to explore these mutants in more depth.

In contrast, my data shows that a different set of mutations to loop 3 length and sequence, which all have pronounced translation defects, do not alter the position of the pretranslocation toeprint. We note that in repeated experiments, the intensity of the pretranslocation toeprint of the G-rich mutant is usually lighter than WT, and that this correlates with the ribosome assembly in RRL shown in Fig. 3.4A. However, when performing direct ribosome binding assays and examining approximate on- and off-rates, the G-rich mutant does not show any difference compared to WT. It is possible that the G-rich mutant alters the conformation of the ribosome in such a way that the reverse transcriptase is not as strongly inhibited by the leading edge of the ribosome, such as stabilizing the open conformation of the latch between the beak and the shoulder of the 40S subunit and permitting the RT to proceed past its usual stopping point. This could result from domain III not being properly “docked” or locked into the A site (perhaps this is what Au and Jan mean by “positioning”). Regardless of the reason for this difference, subsequent experiments show that translocated toeprints can arise
and that A-site tRNA is successfully delivered to G-rich IRES-80S complexes. These data suggest that the reason for the difference in pretranslocation toeprint band intensity in the G-rich mutant does not prevent downstream events in the initiation mechanism.

In summary, while the data presented here are not at odds with what Au and Jan reported, the conclusions we reach are slightly different. This mainly centers around the nuanced interpretation of the toeprinting data and the relevance of toeprint band intensities. Because I discovered novel mutants with more pronounced translation defects, I could dissect the specific mechanistic role of loop 3 in more depth using a wider set of biochemical assays designed to interrogate each step in the initiation pathway. My data indicate that domain III’s loop 3 is involved in the two non-canonical pseudotranslocation events following initial IGR IRES recruitment of the 80S ribosome.

Although domain III was originally proposed to first bind in the P site, the most recent structural and mechanistic models, based on both additional structural information and reexamination of earlier published biochemical data, places domain III in the A site (Fig. 3.1A) (Fernández et al. 2014; Koh et al. 2014; Muhs et al. 2015; Zhu et al. 2011). In this mechanistic model, initial pseudotranslocation by eEF2 is needed to clear the A site before ac-tRNA can bind the ribosome. Consistent with this, my data and other studies show that stable association of ac-tRNA with the IRES-ribosome complex depends on eEF2 (Yamamoto et al. 2007). Additionally, eukaryotic release factor 1 (eRF1) only binds in the A site of IRES-80S ribosome complexes (and induces a change in the toeprint) in the presence of eEF2 (Muhs et al. 2015; Jan et al. 2003). However, no pseudotranslocation is observed with pure WT IGR IRES-80S ribosome complexes.
treated with eEF2 only (assayed by toeprinting) (Pestova and Hellen 2003). A mechanistic model that reconciles this observation posits that eEF2 first moves domain III from the A site to the P site, but this is a transient state and without immediate ac-tRNA delivery domain III spontaneously reverse-translocates to the A site, first proposed in Fernandez et al. 2014. This is validated by the toeprinting experiment demonstrating one round of translocation in RRL pre-treated with high concentrations of hygromycin B (Fig. 3.6A), an antibiotic which has been shown to potently inhibit reverse translocation (Borovinskaya et al. 2008; Szaflarski et al. 2008). If this explanation is true, the transient position of domain III in the P site would preclude detection of this state by traditional biochemical approaches; possibly, the toeprinting assay itself may facilitate reverse-translocation. This mechanistic model is supported by my data and agrees with all previously published data.

Assuming domain III begins in the A site, shortening loop 3 appears to inhibit movement of domain III to the P site before any ac-tRNA is bound. Given that domain III and loop 3 are positioned to interact with components of the 40S subunit head known to be involved in translocation (ribosomal protein uS13 when domain III is in the A site, for example (Cukras et al. 2003)), my data favor a mechanistic model where the loop 3 length mutants fail to efficiently execute the first pseudotranslocation event and this blocks access of ac-tRNA to the A site. This is supported by the single molecule colocalization data (Fig. 3.11) and the anisotropy data with the non-cognate RNA (Fig. 3.9C) which show an increase above background levels established by the no-IRES control. This likely indicates the transient binding of the ac-tRNA TC to the A site and
subsequent rejection due to incorrect codon-anticodon base pairing. In comparison, the fact that the Δ3 mutant yields even lower anisotropy levels than the non-cognate RNA suggests that the TC can never bind the Δ3 IRES-ribosome complex even transiently. This is consistent with the idea that the initial movement of domain III does not occur with this mutant, either spontaneously or with eEF2, and domain III remains in the A site. Given that our sequence mutants (G-rich and GGC) inhibit the second pseudotranslocation, this interpretation makes loop 3, despite being a short and apparently conformationally dynamic element, a key player in non-canonical translocation events that move the IGR IRES through all three tRNA binding sites.

There is no obvious analogous structure to loop 3 in tRNA, raising the question of how this loop exerts its effects. One possibility is that loop 3 interacts directly with the ribosome in ways not yet clearly observed using structural methods. Recent cryoEM reconstructions of CrPV (Fernández et al. 2014) and TSV (Koh et al. 2014) IGR IRESs bound to 80S ribosomes in the pre-translocation (PRE) state (domain III in the A site) at resolutions of 3.8 and 6 Angstroms respectively and of CrPV-80S-eRF complexes in the post-translocation (POST) state (domain III in the P site) at 8.7 Angstroms (Muhs et al. 2015) provide structural models for loop 3. However, the local resolution for loop 3 is low in all structures, consistent with conformational dynamics (Fig. 3.1C-E). Interestingly, in the class I (CrPV) versus class II (TSV) IRESs, loop 3 spans somewhat different space when domain III is in the A site. In both structures, the 3’ ends of loop 3 terminate in the decoding center of the A site where they may interact with elements of the decoding groove. In contrast, the 5’ ends of loop 3 differ in these structural models.
In CrPV the 5’ nucleotides of loop 3 wrap around the 5’ terminal nucleotides of the PKI stem in the A site. In the TSV structural model, loop 3 interacts with the apical loop of rRNA helix 24, part of a constriction between the P and E sites (the P-E gate discussed in Chapter I). In bacterial ribosomes this constriction is essential for maintaining the P-site tRNA in its proper place to prevent slipping of the mRNA (Schuwirth et al. 2005), and must be remodeled by 30S subunit head swiveling for tRNA to translocate from the P to the E site (Fig. 1.6B) (Zhou et al. 2013; Ratje et al. 2010). If loop 3 contacts this constriction, it could affect a known structural regulator of translocation, affecting the conformation of the ribosome in a way that favors eEF2 function. In the POST structure with eRFs, loop 3 is modeled to interact with uS7, a key frame-maintenance and translocation regulator (Devaraj et al. 2009; Galkin et al. 2007; Robert and Brakier-Gingras 2003). Interestingly, the HCV IRES is also thought to communicate with uS7 (Filbin et al. 2013; Fukushi et al. 2001; Boehringer et al. 2005), pointing to this ribosomal protein as an important “gatekeeper” to ribosome function that is exploited by viral IRES RNAs. Precisely what loop 3 interacts with, how and when it makes these interactions, and how these interactions affect the conformation of the IRES-ribosome complex remains to be determined, as does the question of whether loop 3 functions differently in the two classes of IGR IRESs.

In addition to making contacts to the ribosome, loop 3 could also affect pseudotranslocation by altering the conformational landscape of domain III, which comprises an H-type pseudoknot. Many H-type pseudoknots use adenosines in loop 3 to make minor groove interactions with an adjacent helix. Although no minor groove
interactions have been identified in domain III, most IGR IRES loop 3s have adenosine content greater than 40% (Table 3.1); this may be an important feature of loop 3. Indeed, the G-rich and GGC mutations (22% and 33% adenosine, respectively), show substantially decreased translation activity. Transient or dynamic interactions between the loop and the rest of domain III may be important for altering the conformation of the pseudoknot as it moves through the ribosome. tRNAs are known to undergo substantial conformational changes as they transit through the ribosome (Dunkle et al. 2011; Fei et al. 2011); loop 3 could help domain III do the same. Alternatively, it may be important for loop 3 to remain unstructured. Indeed, structural probing of these mutants in the unbound form show decreases in loop 3 accessibility to single-stranded ribonuclease (Fig. 4.1). For a discussion of these data, see Chapter IV. The presence and importance of these changes within the ribosome are unknown, although it is tempting to speculate that a decrease in flexibility may drive the defects observed in this study.

There is growing evidence that molecular mimicry is a common tool viruses use to infect their host cells; indeed, several plant viruses display tRNA mimicry in their 3’ UTRs to enhance viral protein translation (Dreher 2010; Simon and Miller 2013). Yet, molecular mimicry is not limited to structural similarity; the binding partners of these mimics must also be fooled by conformational dynamics and overall molecular interactions. My work suggests that the flexible elements of the IGR IRES facilitate these additional aspects of mimicry that remain understudied. This discovery that IRES RNA flexibility rather than defined structure is important for function may be particularly important in the context of ribosome manipulation since the ribosome has been
suggested to act as a Brownian machine that fluctuates between conformational states (Frank and Gonzalez 2010), and thus it and other elements of the translation machinery are highly tuned to respond to and exploit the dynamics of their ligands.
CHAPTER IV
IGR IRES STRUCTURAL ANALYSIS: THE INFLUENCE OF LOOP 3 MUTATIONS AND RIBOSOME BINDING

Introduction
Cryo-electron microscopy derived structures of the IGR IRES bound to ribosomes were originally published at low to medium resolution, but revealed many important details that gave clues to the mechanism of IRES action (Schüler et al. 2006; Spahn et al. 2004b); most importantly, that the IRES was positioned in the decoding groove like a tRNA, although the precise position was not agreed upon until higher resolution structures were solved recently (Koh et al. 2014; Fernández et al. 2014). This discovery immediately made it apparent that the IRES functioned through tRNA mimicry and could be viewed by the ribosome as an authentic tRNA/mRNA ligand. Indeed, the requirement of elongation factors to initiate translation from the IRES verified that it is translocated through each tRNA binding site (Pestova et al. 2004; Pestova and Hellen 2003; Jan et al. 2003). Given the increasing knowledge of how the ribosome interacts with tRNAs to move them through the decoding groove in the process of peptide synthesis, we hypothesized that the IRES would tap into these same networks of interactions and conformational changes (specifically: subunit ratcheting, L1 stalk movement, and 40S subunit head swivel) perhaps with some IRES-specific solutions to drive these
conformations. However, in the absence of high resolution structures of the IRES in different states of ribosome binding this hypothesis remained untested.

In the previous chapter, I demonstrated that loop 3 of domain III of the IGR IRESs plays a crucial role in the function of the IRES, particularly in its capacity to perform non-canonical translocation events within the ribosome. However, it is unclear how a loop which has demonstrated flexible character can exert these effects. A wealth of bacterial ribosome structures and an increasing number of eukaryotic ribosome structures reveal that each tRNA binding site and the tRNAs themselves undergo subtle but significant local structural changes as the ribosome experiences many of the movements that drive ribosome function. It is conceivable that domain III in general— and loop 3 in particular— also experiences structural changes in each of these states of ribosome binding. Furthermore, such differences could drive pseudotranslocation of the IRES, which I have shown is a function of loop 3. Therefore, it is interesting to investigate how the structure of loop 3 changes upon length and sequence mutation, and how the structure of the IRES as a whole—and domain III specifically—changes in each tRNA binding site. In this chapter I investigated the structural ramifications of loop 3 mutation, whether domain III changes conformation when it is bound in each tRNA binding site of the ribosome, and how the ribosome’s structure changes in response to IRES binding and loop 3 mutation.
Results

Does Mutation to Loop 3 Alter Its Flexibility and/or the Structure of Domain III?

In order to determine whether altering the sequence and length of loop 3 (as in
the G-rich and Δ3 mutants) has an effect on the overall structure of domain III, I
performed native gel analysis of these mutants in the context of domain III alone
(nucleotides 6173-6216) compared to WT. Under non-denaturing conditions and in the
presence of EDTA, RNAs migrate through the gel according to their secondary structure.
The WT, G-rich, and Δ3 RNAs migrate similarly in the EDTA native gel (Fig. 4.1A, top
panel) indicating that the secondary structure of these RNAs is not radically different
upon loop 3 mutation. The slight difference in migration of the Δ3 mutant can be
explained by the loss of molecular weight since it is three residues shorter than the
other two RNAs. However, in non-denaturing conditions and in the presence of MgCl₂
RNAs fold into their native tertiary structures and migrate through the gel accordingly.
Both the Δ3 and G-rich RNAs exhibit different migration patterns compared to WT in the
MgCl₂ native gel analysis (Fig. 4.1A, bottom panel), suggesting that they adopt different
tertiary structures under these conditions or that the structures they form are less
stable, particularly in the G-rich mutant which displays a smear on the native gel rather
than a single, discrete band.

To examine the structural effects of these mutations at nucleotide-level
resolution, I performed RNase T1 probing. RNase T1 is a metal-ion-independent
endoribonuclease that cleaves RNA 3’ of single-stranded G bases. The pattern and
degree of RNase T1 cleavage can be used to determine if specific G bases are more or
Figure 4.1 — Structure Analysis of Unbound CrPV IGR IRES Loop 3 Mutants
Figure 4.1 — Structure Analysis of Unbound CrPV IGR IRES Loop 3 Mutants

A. Native gel analysis of CrPV-11 (domain III only) WT, Δ3, and G-rich constructs. Top image: 10mM EDTA, gel is smiling slightly (high on left and leveling out to the right). Bottom image: 10mM MgCl₂. Gels are stained with ethidium bromide. B. RNase T1 probing (single-stranded G bases) of unbound WT, Δ3, and G-rich CrPV IGR IRES RNAs containing only domain III (CrPV-11 constructs). Cleavage products in the denatured (no Mg²⁺) and native (+ Mg²⁺) states were resolved next to a hydrolysis ladder (OH) on a sequencing gel (left). Graph shows the difference in RNase T1 cleavage in the native state minus the denatured state after normalizing the total amount of radiation in each lane to the WT, no Mg²⁺ levels. The dashed box indicates residues that are in loop 3. G 6204, 6208, and 6209 are specific to the G-rich mutant. C. SHAPE probing of CrPV WT, Δ3, and G-rich toeprint construct RNAs with NMIA. Reactivity to the SHAPE probe was determined by analysis in the HiTRACE program (see methods), then WT reactivity at each nucleotide was subtracted from the mutant’s reactivity at each nucleotide to generate a difference map, which is shown on the graph. Bars above the horizontal center line show positions of greater SHAPE reactivity compared to WT, and bars below the center line show positions of weaker SHAPE reactivity compared to WT. Only domain III sequence is plotted, although the RNAs that were probed contained the whole CrPV IRES. The green box shows the location of loop 3. Error bars represent standard deviation across 3 replicates.
less single stranded under native conditions when compared to denaturing conditions. Probing the G-rich and Δ3 RNAs reveals a similar pattern of RNase T1 cleavage as compared to WT except through loop 3, where the G bases appear to have less single-stranded character in the mutants (Fig. 4.B). These data corroborate the native gel analysis and indicate that the overall secondary structure of the domain is unaltered by loop 3 mutation, however loop 3 itself becomes less single-stranded in nature. This may indicate a decrease in flexible character or the rise of new, more stable interactions with the rest of the domain (such as minor groove interactions with stem 1).

To further explore the flexibility of loop 3 in the G-rich and Δ3 mutants at nucleotide resolution, I performed Selective 2’ Hydroxyl Acylation Analyzed by Primer Extension (SHAPE) probing. SHAPE probing uses N-methylisatoic anhydride to acylate the 2’ OH groups of RNA sugars when they are flexible (Merino et al. 2005). These modifications subsequently block reverse transcriptase nucleotide addition at the position of acylation. This probing method is generally used for secondary structure determination since base paired nucleotides are less flexible than single-stranded nucleotides, but can also be used to assess the degree of flexibility in a single-stranded region. Performing this experiment in the context of the whole IRES (including domains I+II) revealed a decrease in flexibility in loop 3 in the G-rich and Δ3 mutants (Fig. 4.1C), consistent with the RNase T1 data. Increases in flexibility in stem 2a are also apparent particularly for the G-rich mutant, perhaps showing that the PKI interaction defined by stem 2 is less stable in the G-rich mutant, although equally corresponding increases in stem 2b modification are not observed.
Taken together, these three experiments all agree with the interpretation that mutation to flexible/unstructured loop 3 does not cause a radical change in the secondary structure of the domain, but does alter the tertiary structure in some way that is related to a decrease in flexibility or single-stranded character of loop 3. However, it remains unclear whether the changes observed in these experiments are also maintained in the context of the ribosome-bound IRES, in addition to whether these changes are relevant only to specific states of ribosome binding (when the IRES is in the A, P, or E sites, or in the rotated/non-rotated state, etc., for example). It is possible that the slight alterations in structure observed here could lead to nonproductive interactions in the decoding groove of the ribosome even if the RBD can properly recruit ribosomes, but I find this unlikely since it presumably would lead to a difference in ribosome positioning of domain III and we do not observe this type of defect in the toeprinting experiments presented in Chapter III.

How Does the Δ3 Mutation Alter the Ribosome-Bound Structure of the IRES?

I attempted several different types of chemical probing experiments to address how the structure of WT and mutant loop 3 changes as domain III occupies each tRNA binding site, however none of them provided clear, interpretable data due to the complexities of working with the lysate system††. As an alternative to the probing

†† These experiments were attempted before I had the components of the reconstituted system available to me. Although I could try these experiments again in the more controlled context of the purified system, recent high-resolution cryo-EM structures make this a less important direction to pursue.
experiments and to understand the effect of loop 3 shortening, I cocrystallized the Δ3 domain III only RNA (CrPV11 construct) in the P site of the *Thermus thermophilus* 70S ribosome (Fig. 4.2). Data collected from 6 crystals (28 wedges) were used to solve the structure to 3.35 Angstroms at I/σ of 1.0. From these data, clear electron density was visible for tRNA in the P site (Fig. 4.2C), indicating that the ribosome preparation retained contaminating tRNA, although likely at incomplete occupancy. Therefore, these data were unusable for Δ3 domain III structure determination. Efforts to eliminate tRNA contamination from the ribosome preparation are ongoing.

In complementary efforts, I collaborated with Dr. Joachim Frank’s group at Columbia University to solve the structure of the Δ3 IRES mutant (including domains I+II) bound to yeast 80S ribosomes by cryo-electron microscopy. Preliminary 2D classification of auto-picked particles into 200 classes revealed poorly defined structures, although they are clearly the expected size and shape of ribosomes. Improvements to the sample and grid preparation would likely remedy this problem. However, the advent of direct electron detection technology and subsequent publication of the high resolution structure of the WT CrPV IGR IRES bound to yeast 80S ribosomes (Fernández et al. 2014) led to the termination of this collaboration.

This important publication clearly revealed that domain III is positioned in the A site upon initial ribosome binding. However, given the essential role loop 3 plays in translocation of the IRES through the ribosome, I wanted to obtain structural information about the IRES when domain III binds the P site and the E site of the
Figure 4.2 — Cocry stallization of domain III Δ3-70S complexes

A. Pictures of representative rod-like crystals of domain III Δ3-70S complexes. B. Example diffraction data collected on similar crystals as shown in A. C. Electron density maps (purple mesh) calculated from diffraction data show residual tRNA in the P site. 70S ribosome is modeled in yellow and was used for rigid-body alignment. WT domain III is modeled in magenta (PDB ID: 3PYU). P-site tRNA is modeled in green (PDB ID: 2WDK).
eukaryotic ribosome and particularly under conditions where tRNAs associate with the IRES-ribosome complex. Furthermore, structures of elongation factor-bound IRES-80S complexes could reveal how the IRES manipulates them during its initiation mechanism, which has yet to be explored.

**Does the IGR IRES Change Structure Upon Binding in Different tRNA Binding Sites?**

**How Does the Translation Machinery Interact With the IRES and What Are the Relevant Conformational Changes?**

To solve the types of structures described above, I purified WT IRES-ribosome complexes from RRL containing bouvardin, hygromycin B, or cycloheximide. Bouvardin is a small molecule inhibitor of eEF2 that is expected to trap eEF2 on ribosomes. Under this treatment condition, I expected to generate IRES-80S complexes where domain III is in the A site (before translocation) or the P site (after translocation) with eEF2 bound and interacting with domain III. As described in Chapter III, hygromycin B halts ribosomes at the initial binding location as well as after a single round of translocation on the IGR IRES (domain III in the P site). IRES-ribosome complexes that have translocated in the presence of hygromycin B likely contain tRNA bound in the A site, given known interactions of the antibiotic with A site tRNA (Borovinskaya et al. 2008).

Cycloheximide allows for two rounds of translocation (domain III in the E site), and the P site is expected to harbor tRNA. Therefore, complexes built under each of these treatment conditions promise to reveal structures that are necessary to understand the mechanism of translation machinery manipulation by the IRES as a whole. These
structures would show how IRES elements change as it transits through the decoding groove, how tRNAs interact within these complexes and the structural changes elicited in the ribosome during this process. Together and at high resolution they could essentially capture the “movie” of IRES-driven initiation that the IGR IRES field has slowly gathered snapshots of using biochemical methods.

80S complexes generated under these conditions were crudely purified through a sucrose gradient and immediately used to make both negative stain (Fig. 4.3A) and cryo grids (Fig. 4.3B) for electron microscopy. While the negative stain images showed clear dispersion of 80S complexes, the cryo grid images showed high aggregation even after optimization, which prevents meaningful particle-picking and data analysis. While this remains a highly interesting avenue of research to pursue, lack of access to the needed facilities (these experiments were performed at Janelia Farm Research Campus) and the known competition from other successful cryo-EM groups led me to terminate further work on this project.

**Discussion**

The RNA structure-driven mechanism of the IGR IRES depends on its ability to manipulate the translation machinery. I have shown that even unstructured parts of the RNA are important for function. Therefore, the question of how the structure of the IRES RNA can interact with the myriad important structures sampled by the translation machinery during elongation is one of critical importance. It makes sense that the IRES
Figure 4.3—Electron Microscopy of WT CrPV IGR IRES-80S complexes
A. Uranyl formate negative stain of IRES-80S complexes purified from RRL with cycloheximide. B. Cryo-electron micrographs of the same complexes in A. Although ice thickness and clarity is good, samples are aggregated at the borders of the holey carbon, with a few single ribosomes visible (\*).
RNA might need to exhibit conformational plasticity (possibly mediated by flexible regions like loop 3) to successfully hijack this Brownian machine. In this chapter I have demonstrated that mutations to loop 3 alter the conformation of domain III and this may be the root cause of their defects in pseudotranslocation. Although it is possible that some of these mutations affect proper pseudoknot formation, which is a tertiary interaction previously shown to have deleterious effects on translation when abolished, I find this unlikely since each of these mutants is able to generate a toeprint at the appropriate position when bound to 80S ribosomes, unlike IRES mutants where pseudoknot formation is abrogated. Still, although there is evidence to suggest that loop 3 mutation decreases its flexibility, exactly what these IRES mutants look like, and the importance of loop 3 flexibility in the context of different positions of ribosome binding is unknown. Had the experiments I attempted here been successful, I would have expected to see differences in loop 3 structure between WT and the mutants, possibly identifying changes in the orientation of domain III in each binding site, or revealing key interactions with the rRNA or ribosomal proteins. As discussed in Chapter III, there are numerous components of the ribosome that appear to query the anticodon stem loops of tRNAs; it is possible that loop 3 interacts directly with these components to promote pseudotranslocation, or indirectly by modulating domain III conformation in general.

High-resolution structures of the IRES in each tRNA binding site are being pursued by several cryo-EM groups, and a structure (although artificially generated) of domain III in the P site was published early this year (Muhs et al. 2015). I look forward to seeing what these structures reveal about how the IRES uses its various domains to
navigate through the decoding groove. However, an important point of discussion remains. Although cryo-EM has reached a point where near-atomic resolution is possible, local resolution within a structure can vary greatly. This can be a result of a physical aspect of the sample, such as a particularly dynamic region, or a result of the method, such as having too few views of the complex in specific orientations to define that area well. Furthermore, structure modeling in cryo-EM can be performed at different contour levels or thresholds. Higher thresholds give you lower resolution information but higher confidence in the electron density information. Conversely, lower thresholds report on atomic-level features, but are subject to a high degree of noise, thereby lowering the confidence in the electron density. Standard practices to deal with these issues and perform de novo atomic-level structure modeling from cryo-EM datasets are in their infancy. The IGR IRES field could suffer from these issues given the known role of conformational dynamics in its mechanism. X-ray crystallography is another high-resolution structure determination method that holds promise; indeed, much of what we know about ribosome structure and interactions with its ligands has come from crystal structures of bacterial ribosomes. Yet, obtaining eukaryotic 80S ribosome crystals with relevant ligands in the decoding groove remains an extraordinary challenge. For these reasons cryo-EM is currently the best method available for determining each of these important IRES structures, and the ability to computationally separate non-homogenous complexes (such as ribosomes in different states of subunit rotation) are definite steps towards rectifying some of the method-associated problems mentioned above.
CHAPTER V

STRUCTURAL ANALYSIS OF THE CCR5 -1 PROGRAMMED RIBOSOMAL FRAMESHIFTING SIGNAL AND ITS INTERACTION WITH MICRO RNA-1224

Introduction

The -1 programmed ribosomal frameshifting (-1 PRF) signal in the chemokine receptor CCR5 was discovered from a genome-wide computational search for sequences that fit the criteria: heptanucleotide slippery sequence—six to nine nucleotide linker sequence—predicted RNA pseudoknot (PK) (Hammell et al. 1999) (Fig. 5.1A). Despite the known difficulties in de novo sequence-based PK prediction, this genome-wide analysis identified such a signal at position 407 in the CCR5 mRNA (among many other hits), and its frameshifting activity was validated by several cell culture and in vitro methods. Although frameshifting is an observed phenomenon throughout all kingdoms of life, the prevalence and function of frameshifting within human mRNAs is relatively unexplored, and this is particularly true for frameshifting elements without known viral ancestry (Dinman 2006). Thorough validation of other hits from the screen have not yet been reported in the literature, therefore the example from CCR5 provides a unique opportunity to explore how structured elements within coding regions of mRNAs can contribute to their own translational control. We now have a framework within which we can ask: What are the structures in eukaryotic messages that elicit frameshifting on
ribosomes? How are these structures regulated? What is the mechanism by which frameshifting occurs?

Figure 5.1 — Programmed Ribosomal Frameshifting in CCR5
A. Schematic of generic -1 PRF architecture. B. Diagram of the consequences of a ribosome encountering the -1 PRF signal in CCR5. The majority of the time, the ribosome stays in the normal reading frame (heavy arrow) to make full length, active protein. Less frequently (light arrow) the ribosome undergoes -1 frameshifting, which directs the ribosome to a premature stop codon and targets the mRNA for degradation. C. miR-1224 stimulates -1 PRF (heavy arrow) by binding directly to the CCR5 PRF element.

Based on the screen mentioned above, roughly 10% of genes contain a -1 PRF signal, and the majority of these signals target the ribosome to a premature stop codon, presumably leading to nonsense mediated decay (NMD) of the mRNA (Hammell et al. 1999; Dinman 2006). The relatively widespread incidence of these signals accompanied by the previous observation that global disruption of -1 PRF in yeast caused severe growth defects (Dinman and Wickner 1994) led to the idea that different PRF signals
may be differentially regulated from one another in the context of the living cell (Belew and Dinman 2015). Mechanisms to achieve this differential regulation could include non-coding RNA or protein targeting of the PK in a highly sequence-specific manner, and these interactions could conceivably enhance or suppress frameshifting. Intriguingly, microRNA (miRNA) target prediction programs based on base pairing potential identified two miRNAs that were also later validated to interact with the CCR5 -1 PRF element (CCR5) \textit{in vitro} and \textit{in vivo}. Interaction of one of these miRNAs, miR-1224, stimulated -1 PRF activity in multiple cell types (Fig. 5.1B and C) (Belew et al. 2014).

The structure of CCR5 was probed by a battery of chemical modification methods and several RNA folding programs guided the construction of the secondary structure model (Fig. 5.2A). However, a few components of the model are questionable. First, although the 5’ end of the signal is defined by the beginning of the slippery heptad (purple), there is no experimental data to define the 3’ end of the -1 PRF element. Presumably, the authors of the study used the 3’ end that the PK prediction software suggested, but these programs are known to be unreliable. Therefore, inclusion of probing data for nucleotides 3’ of the functional PK element could obscure the secondary structure model. More experiments need to be done to define the functional 3’ end of the RNA. Second, the secondary structure of the PK also includes the “linker” region in the sequence of the PK model (orange). This seems odd since, at least in terms of our knowledge of classical -1 PRF structures, the linker sequence should provide physical separation between the slippery heptad and the PK, rather than be a part of the
Figure 5.2 — Models of CCR5 and Proposed Interaction with miR-1224
Figure 5.2 — Models of CCR5 and Proposed Interaction with miR-1224

A. Secondary structure diagram of CCR5 published in (Belew et al. 2014). The slippery site is shown in purple, the linker in orange, and the rest in green, per the coloring scheme in Fig. 5.1A. B. Modified secondary structure diagram which is used throughout this chapter. Coloring scheme is as described in A. This model was drawn to make the slippery site and linker region single stranded, which I feel is more representative of what the RNA looks like when bound to the ribosome. C. Primary nucleotide sequence of the CCR5 -1 PRF element (black), and its predicted hybridization to miR-1224 (blue). Solid line indicated canonical Watson-Crick base pairs, colon indicates G-U wobble pairs, red X indicates a mismatch. D. Schematic of the published conceptual model for miR-1224 binding to CCR5. The conceptual model posits that the microRNA binds in the minor groove of stem 2 in a “triple” interaction.
structured PK (green). The linker, presumably, exists in a single stranded state in the mRNA entry channel of the ribosome, allowing the slippery site to sit in the A and the P sites—the location where tRNA slippage is thought to occur—once the PK has been encountered by the edge of the ribosome, inducing a pause. Furthermore, the base-pairing scheme within the linker region is tenuous; more work needs to be done to understand the role of the linker sequence within the structural model of the CCR5 -1 PRF RNA. Finally, the model for the PK lacks any residues for “loop 3” of this H-type PK. As discussed in previous chapters, loop 3 is usually the longest of the three loops in H-type PKs. This deviation from classical RNA folding patterns also calls the secondary structure model into question. Due to these issues, I chose to draw the secondary structure in a way that reflects these concerns (Fig 5.2B), although I did not have direct evidence to support this model.

Although miR-1224 was identified through base-pairing potential with CCR5 RNA, none of the three putative hybridization sites display particularly long stretches of complementarity (Fig. 5.2C). In order to understand how miR-1224 interacts with the CCR5 -1 PRF signal, SHAPE probing was performed in the presence and absence of miR-1224 (Belew et al. 2014). Surprisingly, no obvious changes between binding states were observed using this method. The authors concluded that miR-1224 stimulation of CCR5 -1 PRF occurred by direct binding and stabilization of a preexisting conformation of CCR5. This binding was proposed to occur through insertion of the miR in the minor groove of stem 2 of the PK by base-triple interactions (Fig 5.2D), although this idea was not
directly tested. How this binding mechanism might elicit sequence-specific regulation of CCR5 -1 PRF is unclear.

Despite the aforementioned problems with the structural model, the observation that a non-coding RNA can modulate translational recoding in trans is an exciting possibility. Clearly, more work needed to be done to understand this novel paradigm in translational control. In this chapter I will describe the work I have done to address the questions: What are the sequence and/or structural determinants of miR-1224 binding to the CCR5 -1 PRF element? Does the structure of the RNA change upon miR-1224 binding?

**Results**

**Does miR-1224 stabilize a preexisting conformation of the CCR5 -1 PRF RNA‡‡ or induce a new structure?**

Before I could begin examining the CCR5-miR-1224 interaction, it was first essential to verify that I could recapitulate miR-1224 binding to CCR5 in vitro. To do this, I performed an electrophoretic mobility shift assay (EMSA, Fig 5.3). Using ³²P-labeled WT CCR5 RNA (“CCR5-1”), I observed a distinct shift to a slower mobility species when I added unlabeled miR-1224 (lane 2), but not when I added a non-targeting, control miRNA (24 nucleotides, lane 3). Similarly, when I radiolabeled miR-1224 and added unlabeled WT CCR5, I observed a shift of miR-1224 to a slower mobility species (lane 4),

‡‡ Here and throughout this thesis, I will commonly refer to the CCR5 -1 PRF RNA element simply as “CCR5”. To eliminate confusion with the CCR5 protein, I will always specify if I am referring to the protein.
compared to miR-1224 alone (lane 5). In both of these experiments, I folded the CCR5 RNA in the presence of magnesium before adding miR-1224. These data show that I can recapitulate miR-1224-CCR5 binding in vitro and that this is a specific interaction that recognizes the native structure of CCR5.

Figure 5.3 — miR-1224 Binds Specifically to the WT CCR5-1 PRF RNA
Electrophoretic Mobility Shift Assay (EMSA) of $^{32}$P-labeled CCR5-1 RNA only (-), plus miR-1224 (+1224), or plus a control short RNA (+ctrl) (lanes 1, 2, and 3). Lanes 4 and 5 show $^{32}$P-labeled miR-1224 + unlabeled WT CCR5, or just $^{32}$P-labeled miR-1224 only. Red type indicates which component of the reaction is $^{32}$P-labeled.

A central component of the published model for miR-1224 recognition of CCR5 is that it stabilizes a native structure of the RNA, rather than inducing a new conformation. To investigate this hypothesis, I performed RNase T1 probing of CCR5 in the absence and presence of miR-1224 as well as the aforementioned control miRNA (Fig 5.4 A, E). As described previously, RNase T1 cleaves 3’ of single stranded G nucleotides and the intensity of cleavage can be compared between conditions (such as bound vs. unbound)
Figure 5.4 — miR-1224 Binding to CCR5 Induces Structural Changes
Figure 5.4 — miR-1224 Binding to CCR5 Induces Structural Changes

A. RNase T1 probing of $^{32}P$-labeled CCR5 RNA under native conditions with increasing amounts of miR-1224 or a control miRNA at defined stoichiometries (1/5x, 1x, or 5x). OH, hydrolysis ladder; G, RNase T1 digestion under denaturing conditions. B. RNase T1 probing of $^{32}P$-labeled miR-1224 RNA under native conditions with increasing amounts of WT CCR5. C. Normalized and quantitated data from A for the 5x condition, plotted as a fold change in amount of cleavage compared to the signal from each G in the CCR5 only condition. D. Changes in RNase T1 cleavage of CCR5 upon miR-1224 binding, plotted onto the modified secondary structure. Green indicates a specific increase in T1 cleavage, red indicates a specific decrease in T1 cleavage, and gray indicates no change or nonspecific changes in T1 cleavage. E. Same as in A except with miR-1224 in larger excess to CCR5, showing more drastic changes in the T1 cleavage pattern.
to identify changes in the single-stranded character of a given G base. Quantification of the RNase T1 probing data (Fig 5.4C) clearly indicate increases in T1 cleavage at Gs 29, 31, 34, 39, 41, 51, 52, and 53 and decreases at Gs 16, 17, 64, 68, and 80 in the presence of miR-1224, but not in the presence of a control miRNA. Thus, miR-1224 binding to CCR5 causes specific changes in the structure of the RNA.

To complement these data, I performed the same experiment but instead monitored changes in miR-1224 cleavage by RNase T1 directly (Fig. 5.4B). Qualitatively, there are decreases in T1 cleavage at the 3’ end and throughout the middle of the miRNA. The 3’ end of miR-1224 was not well-resolved in this experiment to allow for reliable quantitation. Nevertheless, these data indicate that the single-stranded nature of miR-1224 decreases upon interaction with CCR5. Whether these decreases are a result of direct base pairing, or general inaccessibility to the nuclease upon CCR5 binding (such as if it was participating in an RNA triplex), is unclear. Taken together with the above data, I concluded that miR-1224 interacts specifically to induce structural changes in the CCR5 -1 PRF signal.

**What Is the Stoichiometry of miR-1224 Binding to the CCR5 -1 PRF Element?**

An important aspect of the model of miR-1224 interaction with CCR5 is the number of miR-1224 molecules that recognize a single PK. The current model assumes a 1:1 stoichiometry of interaction. I worked with a graduate student in our lab, Erik Hartwick, to use a fluorescence anisotropy-based binding assay to determine the stoichiometry of the interaction of miR-1224 with CCR5. We used 50nM TAMRA-labeled
miR-1224 (well above the reported $K_d$ of 360pM) and titrated WT CCR5 from molar concentrations 5 fold below 1:1 (miR-1224 in excess) up to 6 fold over 1:1 (CCR5 in excess). Addition of CCR5 RNA over a molar ratio of 1:1 did not lead to an additional increase in anisotropy (Fig. 5.5). Thus, miR-1224 binding to the CCR5 -1 PRF element is 1:1.

**Figure 5.5 — Stoichiometry of miR-1224 Binding to CCR5**
Fluorescence anisotropy binding assay of TAMRA-labeled miR-1224 to titrated WT CCR5. TAMRA-miR-1224 was kept at 50nM and CCR5 was titrated from 10nM to 300nM. The data were fit with linear regressions for determination of apparent stoichiometry.
What are the Important Regions of CCR5 For miR-1224 Association?

The effect of mutations within CCR5 on miR-1224 binding is not well understood, but can provide insight to the location of miR association. Furthermore, the functional 3’ end of the CCR5 -1 PRF element has not been defined by experimental methods. To address the role of the 3’ end of the CCR5 structure, I made a construct that was truncated at the 3’ end to nucleotide 89 (CCR5-1.89)§§ (Fig. 5.6) Based on the current secondary structure, this would disrupt one end of stem 2, liberating nucleotides 24-35. Additionally, I made a construct in which the center of stem 2 is disrupted (CCR5-1.S2D Disrupt (82-89), or simply 1.S2D to abbreviate) by exchanging the sequence with the same sequence on the opposite side of the proposed helix. If the integrity of stem 2 is necessary for miR-1224 binding, or if miR-1224 directly hybridizes to these regions of mutation, I would expect to see a decrease in miR binding. Interestingly, I observed no difference in miR-1224 association with these CCR5 mutants by EMSA analysis (Fig 5.7A, lanes 10-13) or by site-specific crosslinking (Fig 5.7D, method described in more depth below) compared to the WT CCR5. In fact, the CCR5-1.S2D construct appears to possibly enhance miR-1224 binding, as evidenced by the loss of the lower “unbound” band in the EMSA gel. These data indicate that neither the 3’ end of CCR5 nor the integrity of stem 2 is necessary for miR-1224 binding, and disruption of this stem may actually promote miR-1224 association. Importantly, this is also consistent with the observation that

§§ A note about construct nomenclature: The number in front of the decimal point indicates the 5’ terminal residue and the number after the decimal indicates the 3’ terminal residue. For clarity, the “WT” RNA would be called CCR5-1.103 using this system since it starts at residue 1 (the first U in the slippery site) and ends at residue 103, the reported 3’ end of the PK.
<table>
<thead>
<tr>
<th>Mutant</th>
<th>Type of Mutation</th>
<th>Starts at Residue #</th>
<th>Ends at Residue #</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR5-1.89</td>
<td>truncation from 3’ end</td>
<td>1</td>
<td>89</td>
</tr>
<tr>
<td>CCR5-1.Stem 2 Disrupt</td>
<td>substitution of 10 bases</td>
<td>1</td>
<td>103</td>
</tr>
<tr>
<td>CCR5-10</td>
<td>truncation from 5’ end</td>
<td>10</td>
<td>103</td>
</tr>
<tr>
<td>CCR5-14</td>
<td>truncation from 5’ end</td>
<td>14</td>
<td>103</td>
</tr>
<tr>
<td>CCR5-15</td>
<td>truncation from 5’ end</td>
<td>15</td>
<td>103</td>
</tr>
<tr>
<td>CCR5-16</td>
<td>truncation from 5’ end</td>
<td>16</td>
<td>103</td>
</tr>
<tr>
<td>CCR5-17</td>
<td>truncation from 5’ end</td>
<td>17</td>
<td>103</td>
</tr>
<tr>
<td>CCR5-1.polyU12miR1224</td>
<td>WT +miR in cis</td>
<td>1</td>
<td>103*</td>
</tr>
</tbody>
</table>

Figure 5.6 — Mutants of CCR5
Table and secondary structure diagrams describing the mutants of CCR5 used in several subsequent experiments. Asterix indicates that the CCR5-1.polyU12miR1224 RNA contains WT CCR5 sequence to position 103, although the actual 3’ end terminates 31 nucleotides later (position 134).
nucleotides in stem 2a show enhanced RNase T1 cleavage upon miR binding suggesting this helix is disrupted when miR-1224 binds. These data directly contradict the published, “conceptual” model of the CCR5-miR-1224 interaction because it means that an intact stem 2 is not necessary for miR binding.

I also made a construct that contains the WT CCR5 sequence and miR-1224 sequence, separated by a 12 nucleotide polyuridine stretch. Essentially, this RNA provides the miR to CCR5 in cis, rather than in trans, and was originally designed for crystallography attempts. If the miR-1224 sequence provided in cis can properly interact with CCR5, then it might be expected that miR-1224 added in trans to this construct will not bind as well since its binding site would already be occupied. However, EMSA and crosslinking analysis show that miR-1224 added in trans to this construct can bind the RNA (Fig 5.7A, lanes 14 and 15). This either means that this construct as it is designed does not permit the cis-miR-1224 to bind properly to CCR5, thus the binding site is still available for miR binding in trans, or alternatively, that cis-miR-1224 binding is transient enough to allow for trans-miR-1224 association in these assays as well. Which of these possibilities is occurring is unknown. However, the observation that trans-miR-1224 can still associate with this RNA argues that CCR5 is not misfolded in this construct, and validates the possibility of using a cis-acting miR sequence as a tool in future studies like crystallography.

With the knowledge that the 3’ end is not necessary for miR-1224 binding, I next investigated the role of the 5’ end of CCR5 in miR association. To do this, I truncated the
Figure 5.7 — miR-1224 Binding to CCR5 Mutants

A. Electrophoretic Mobility Shift Assay (EMSA) of $^{32}$P-labeled CCR5-1 RNA only (-), plus miR-1224 (+1224), or plus a control short RNA (+ctrl) (lanes 1, 2, and 3). Lanes 4 and 5 show $^{32}$P-labeled miR-1224 + unlabeled WT CCR5, or just $^{32}$P-labeled miR-1224 only. Red type indicates which component of the reaction is $^{32}$P-labeled. (same as Fig. 5.3). Lanes 6-15, CCR5 mutants only (even lanes) or in the presence of cold miR-1224 (odd lanes). B. Non-denaturing EDTA gel of $^{32}$P-labeled CCR5 WT and mutants. The RNAs migrate similarly (changes are consistent with alterations in molecular weight), indicating there are no substantial changes in secondary structure. C. Time-course of crosslinking at 365nm of $^{32}$P-labeled 4SU(17)-miR-1224 to WT CCR5 or CrPV IRES RNA (negative control). D. Crosslinking for 40 min of 4SU(17)-miR-1224 to CCR5 mutants or CrPV IRES RNA (negative control).
RNA from the 5’ end, creating constructs CCR5-10, -14, -15, -16, and -17***. It should be noted that this deletes the slippery sequence (residues 1-7) and up to all of the proposed “linker” sequence. Surprisingly, deletion beyond 10 residues from the 5’ end is deleterious to miR-1224 binding as assayed by EMSA and site-specific crosslinking (Fig 5.7A and D). Thus, the 5’ sequence of CCR5 (which includes linker sequence) is important for association of miR-1224. It is unknown whether miR-1224 directly hybridizes to or interacts in some way with this 5’ sequence, or if the 5’ end is important to maintain a global structure that miR-1224 recognizes. Studies are ongoing to understand these possibilities.

As alluded to above, a caveat to these mutation experiments is that it is possible that the mutations disrupt the overall fold of CCR5, and therefore that loss of binding is actually a result of improper RNA folding. To determine whether mutation to CCR5 causes global structural changes within the RNA, I performed RNase T1 probing on each of the mutants. Additionally, I wanted to know whether the changes observed with WT CCR5 plus miR-1224 also manifested in the mutants. Therefore, the RNase T1 probing was performed in the absence and presence of miR-1224 and a control miRNA (Fig 5.8). Intriguingly, even in the absence of miR-1224, the 3’ mutants CCR5-1.89 and -1.S2D both show the characteristic changes that the WT displays in the presence of miR-1224 (discussed above). These changes are enhanced in the presence of miR-1224, but not by the control miRNA. This may suggest that these mutants favor the conformation of the RNA that is stabilized by miR-1224 binding.

*** All of these constructs end at nucleotide 103, and so would be properly named “CCR5-10.103” and so on, but I have left off “.103” for brevity.
Figure 5.8 — RNase T1 Probing of CCR5 Mutants
**Figure 5.8 — RNase T1 Probing of CCR5 Mutants**

RNase T1 probing of $^{32}$P-labeled CCR5 mutants in the absence (-) or presence of miR-1224 or a control miR. The probing patterns are generally maintained compared to WT (see Fig. 5.4), indicating a similar overall architecture of the RNAs. Several important regions are indicated by colored squares. Red squares indicate G39, a position that in the WT sequence is highly cleaved by T1 in the presence of miR-1224. Blue squares indicate Gs 51-53, nucleotides that are highly cleaved by T1 in the WT + miR-1224. Yellow squares show G70 and G68, nucleotides that decrease in cleavage in the WT + miR-1224. Additionally, the pink square indicates the 3 G residues that are introduced by the CCR5-1.S2D mutation. T1 cleavage after these G nucleotides appears to be enhanced in the presence of miR-1224. The purple square indicates at least one G nucleotide in the CCR5-1.polyU12miR1224 that is a part of the miR-1224 sequence.
Expectedly, the 5’ truncated CCR5 constructs do not display these characteristic changes, consistent with their defect in miR-1224 binding. However, it is important to note that deletion at the 5’ end leads to enhanced T1 cleavage at G62 and G64, consistent with the interaction of this part of the mRNA with the 5’ end that is proposed in the original secondary structure model. Because there are not radically different changes in the probing pattern in other parts of the RNA, these data support the idea that an entirely different global fold of the RNA is not caused by deletion of 5’ nucleotides.

**Where Does miR-1224 Bind to CCR5?**

Mutational analysis as performed above is useful to understand the general elements involved in miR-1224 binding, but it fails to provide information about specific locations of interaction. In order to pinpoint the location of miR-1224 binding on CCR5, 4-thio-uridine (4SU) was incorporated at the 17 position of miR-1224 (near the 3’ end, Fig. 5.9A) and then I site-specifically crosslinked this modified miR to WT CCR5 by exposure to long-wave UV light (Fig 5.7C for time course of reactivity with WT CCR5). I isolated and purified the crosslinked product by HPLC (Fig. 5.9B), and then used it as a template for reverse transcription (Fig 5.9C). Crosslinking of the 4SU(17) miR-1224 to CCR5 should prevent read-through by the RT and generate a truncated cDNA whose length correlates to the position of crosslinking. Using this method, I detected a site-specific, UV-induced band at G70 and possibly at G68 and U66 (Fig. 5.9D), although these are still preliminary investigations and need to be repeated. Fortuitously, there
Figure 5.9 — Site-Specific UV Crosslink Mapping of 4SU-Containing miR-1224 to WT CCR5-1 PRF RNA

A. Sequence of 4SU(17)-miR-1224. Purple indicates the position of the 4-thio-uridine residue.

B. Purification of crosslinked species. The figure shows HPLC chromatograms (A$_{260}$) of indicated samples. Only when the 4SU modification is present and the samples are exposed to UV light (365nm) is there a peak that elutes after CCR5 alone, indicating this is the crosslinked species. ** indicates miR-1224, * indicates CCR5, and *** indicates the mir-1224/CCR5 crosslinked species.

C. Reverse transcription of untreated CCR5 (RT ctrl), or the CCR5+UV (*) and crosslinked species (**), next to a dideoxyNTP sequencing ladder. Position of unextended primer and crosslink-induced stops are indicated on the gel.

D. Position of observed crosslink-induced stops are shown on the modified secondary structure of CCR5 in red type. The major stop at G70 is red and bolded, whereas less intense stops are simply red.
are two other uracil residues within miR-1224 at positions 2 and 9, providing relatively evenly spaced positions through which I can use this method to interrogate the location of the entire miR on CCR5 in the future.

**What Is the Secondary Structure of the CCR5 -1 PRF Element? Are There Alternate Models That Fit the New Data?**

The observation that the 3’ end, specifically residues 89-103, is unnecessary for miR-1224 binding suggests that the secondary structure as it is drawn may be incorrect. I revisited this question through a collaboration with Dr. Rhiju Das’s group at Stanford University. The Das group developed a method for secondary structure determination using a more quantitative means of chemical probing than has been performed before (Kladwang et al. 2014). To do this, stable hairpins are placed upstream and downstream of the RNA of interest. The modification through the loops and lack of modification through the helices of these hairpins are used as normalization tools throughout the experiment. Additionally, the RT reactions used to detect the locations of chemical modification are run on a capillary electrophoresis instrument (as opposed to traditional sequencing gels), which allows for better resolution of the cDNA products and the data are analyzed through the HiTRACE computational pipeline to extract quantitative reactivity information. Reactivities to each chemical are then provided along with the primary RNA sequence to various RNA folding algorithms (SHAPEknots, Fold, etc.) to predict a secondary structure that best fit the data.
First, we performed SHAPE and DMS probing on the WT CCR5 (residues 1-103) this time including the two up- and downstream normalization hairpins (Fig. 5.10B, C), essentially repeating the experiments already reported in the literature. Our data show a high degree of SHAPE modification in nucleotides 23-30, in direct contrast to the current model which shows residues 24-30 to be base paired with nucleotides at the 3’ end of the RNA. SHAPE reactivities in different salt conditions plotted on the published secondary structure reveal differences between the SHAPE data we collected here and the secondary structure model (Fig 5.10 D) proposed in the literature. All of the predictions of possible secondary structures based on these data are shown in Appendix 5.

To expand on this initial result and develop a high confidence model for the secondary structure, I performed “Mutate and Map” (M²) which is a variation on the chemical probing methods described above (Kladwang et al. 2011a; 2011b). In the M² experiment, single point mutations are made at each position throughout the RNA of interest and all of these mutants (103 in this example) are probed and analyzed using the HiTRACE program. If a specific nucleotide is base paired to another nucleotide in the WT sequence, abrogation of that interaction by mutation of the first nucleotide will lead not only to an increase in modification of that specific residue, but also will cause an increase in modification of the nucleotide to which it was paired in the original sequence. These compensatory changes in modification pattern by stepwise single point mutation of the RNA can be used to identify not only that a nucleotide is base paired, but also identifies specifically and without bias the partner in that base pairing.
Figure 5.10 A-C — Chemical Probing of CCR5
D.

SHAPE Modification of WT CCR5 PK
1M NaCl

SHAPE Modification of WT CCR5 PK
10mM MgCl₂

Figure 5.10 D — Chemical Probing of CCR5
Figure 5.10 E — Chemical Probing of CCR5
Figure 5.10 F — Chemical Probing of CCR5
Figure 5.10 — Chemical Probing of CCR5

A. Table showing several different chemical probes and their position of modification on RNA nucleotides. B. Diagram of the construct used for 1D and 2D chemical probing. C. 1D chemical probing data. Location of the unextended primer and the fully extended product are indicated. “No Mod.” = no modification of the RNA. Numbers below each lane indicate the extension temperature: 1 = 42°C, 2 = 48°C, 3 = 55°C. D. SHAPE probing results plotted on the published secondary structure of CCR5 in the presence of 1M NaCl (top) and 10mM MgCl₂ (bottom). SHAPE reactivity scale is indicated at bottom right. E. 2D “Mutate and Map” data set for WT CCR5. Yellow line shows the position of mutations, green lines show the two main helices, and the purple boxed lane shows an example of a point mutant (in this case G34C) that exhibits a completely different probing pattern compared to WT and the majority of the other point mutants. F. Secondary structure of CCR5 predicted from 1D and 2D probing data. SHAPE reactivity scale is shown bottom right and reactivity at each position is shown as a circle at each nucleotide. Thick orange lines indicate the base pairing scheme from the published secondary structure to illustrate the differences between that model and the model presented here. Bootstrap confidence intervals after 100 iterations of the SHAPEknots structure prediction algorithm are shown as percentages (green type) over each helical section.
interaction. This provides superior secondary structure prediction capability compared to traditional probing methods.

Performing this experiment with the 1M7 SHAPE probe (Fig. 5.10 E) revealed several notable observations. First, the data are fairly “clean” in that there are clear, consistent patterns of high and low modification throughout the library of mutants. This is consistent with the idea that the RNA mostly exists in a single conformation in the conditions of this experiment, and argues against the idea that the RNA exchanges between more than one structure, which would be characterized by moderate levels of modification (gray) throughout the sequence. Additionally, several mutants appear to dramatically change the probing pattern compared to the WT and the vast majority of the other single point mutants (see mutants G34C, purple box, for example). In these cases, it appears that the single point mutation leads to a completely different fold of the RNA. Remarkably, these alternative folds are consistent between the different mutants that appear to cause them. For example, many of these mutants show a drastic decrease in modification through residues 63-67, among other changes. It is intriguing that many of these differences in probing pattern are consistent with the characteristic changes I observe in RNase T1 experiments performed with the WT CCR5 + miR-1224. This led me to question whether these point mutants could be examples of RNAs that are “preformed” into the miR-bound conformation of CCR5 even in the absence of miR-1224. This idea will be explored more in Chapter VI.

We used the “1D” and “2D” (M²) data to predict the secondary structure of the WT CCR5 (Fig. 5.10 F). The data support a two-hairpin structure, which is different from
what was previously reported in the literature although the helix formed from nucleotides 38-60 and 68-89 contain the same base pairs. It is interesting to note that the predicted structure does not contain a PK as it is drawn. Most of the reactivities alone (base paired vs. single stranded) are largely conserved between our data and published data, with the exception of residues U23-U30 as mentioned previously, as well as residues G8-G16 and U90-A103 which are all shown to be single stranded in our experiments.

A major caveat to this experiment is that many of the secondary structure predictions obtained show interaction of the 3’ end of CCR5 with the “tail” sequence that is used as a primer binding site for reverse transcription (see Appendix 5). This interaction could lead to the formation of a competing structure that is irrelevant to the native fold of the CCR5 RNA, thus invalidating the probing data obtained. On the other hand, it is possible that if the 3’ end is irrelevant to the functional form of the CCR5 RNA (which is still just a hypothesis at this point since I have not tested function of these constructs) then perhaps by providing alternative base pairing partners for this region, we may have inadvertently but fortuitously allowed for the proper, functional structure to form. In this scenario, the probing data would still be useful. It is also possible that the 3’ end of the CCR5 RNA could base pair with the apical loop of the first stem of our new structure (residues 23-30), forming another PK, but that 3’ end sequestration by the tail prevents it from forming in this experiment. Base pairing of these nucleotides is a part of the original structural model, but our data suggest it could form in a different way, surrounded by a different architectural framework. Efforts to remedy and clarify
these issues are ongoing, as are performing these experiments in the presence of miR-1224.

Discussion

How the ribosome engages with the structure of frameshifting pseudoknots to promote recoding during the otherwise high-fidelity process of translation is a topic of ongoing research. A step towards understanding this process is acquiring knowledge about the three dimensional folds of RNAs that induce frameshifting, their stability, and how they are regulated. The surprising finding that microRNAs contribute to this regulation presents a new level of intrigue to this problem. MicroRNA targeting to mRNA open reading frames is not a widely observed phenomenon, although a few reports exist (Schnall-Levin et al. 2011). Even in these rare cases, microRNAs always find their mRNA targets through direct hybridization, not through structure-based recognition (Lau et al. 2001; Lagos-Quintana et al. 2001; Bartel 2009). This is due at least in part to the contribution of Argonaute (AGO) proteins that help microRNAs find their targets; the structure of the miR-1224-Ago complex itself would prevent an elaborate structure-based recognition strategy for miR-1224 binding to CCR5. These issues are discussed in more depth below. However, given the stimulatory effect of miR-1224 on CCR5 -1 PRF (Belew et al. 2014), and the known importance of RNA structure in this type of process (Hansen et al. 2007; Dinman 2012), the transitive conclusion is that miR-1224 must promote or stabilize the relevant frameshift-inducing RNA structure in CCR5. This represents a completely new way of thinking about microRNA function in translational
control. Understanding how a microRNA can interact with structured elements will greatly expand our knowledge of how noncoding RNAs exert regulatory effects within the cell.

Here, I have established experimental systems that can be used to understand the structure of -1 PRF signals and their interactions with putative RNA binding partners. Furthermore, I present data that serves as a foundation in understanding the -1 PRF signal in the CCR5 mRNA and its association with its regulator, miR-1224. Although this body of work is far from complete, my work disproves the published “conceptual” model for CCR5-miR-1224 interaction, which suggests that miR-1224 binds CCR5 to form an RNA triplex in stem 2 as it was originally drawn (Belew et al. 2014). The true mode of interaction remains elusive. However based on my data, I propose a new secondary structure model for the CCR5 -1 PRF signal and a speculative model for its binding with miR-1224 (Fig 6.1). Chapter VI discusses the evidence that supports this model from the data presented in this chapter. This model is supported by as much of the data as possible, although some observations are not accounted for; this simply is a place to start and provides for new hypotheses to be tested. The secondary structure is based on the two-hairpin prediction from the M² data, but also includes interaction of the 3’ end of the RNA with nucleotides 23-30 to form a pseudoknot. Most importantly, this model also shows a pseudoknot forming between G12-G16 and U67-U63, which I believe is promoted by miR-1224 binding.
Future Directions

Although this work provides a wealth of new information about how miR-1224 interacts with the CCR5-1 PRF element, we are still a long way from having an in-depth understanding of the precise nature of this interaction and how it promotes frameshifting in cells. The future directions for this project are directed towards enhancing our understanding of the molecular details of the interaction, as well as applying these findings to a functional frameshifting system.

A preliminary analysis of binding affinity between CCR5 and miR-1224 was performed using EMSA experiments and reported an apparent $K_d$ of 360 pM for the interaction (Belew et al. 2014). However, extracting $K_d$ from these types of experiments can lead to overestimation of the affinity due to caging effects in the gel which minimize the effect of off-rate (Fried and Liu 1994). It will be important to pursue a more accurate $K_d$ determination, as this will illuminate an important aspect of the nature of miR-1224 binding to CCR5 and inform future experimental design. Currently, I am working to adapt the fluorescence anisotropy binding assay that I used for stoichiometry experiments to assess the affinity of this interaction.

An obvious next experiment is to perform the 1D and 2D chemical probing experiments in the presence of miR-1224. Although no differences in SHAPE probing were observed in the original study when miR-1224 was added (Belew et al. 2014), the RNase T1 data presented here clearly indicate changes in structure. The expanded quantitative power of the 1D and 2D chemical probing methods might permit these changes to be observed. This experiment will also be greatly aided by the determination
of $K_d$ as discussed above because it will allow the experiment to be performed under conditions when all of the CCR5 RNA is bound by miR-1224. These data can then be used to predict the new structure associated with miR-1224 binding.

A limitation of the chemical probing experiments, however, is that they do not report on how specific nucleotides in miR-1224 interact with CCR5. To have a complete picture of the miR-1224-CCR5 interaction, we need to solve the high-resolution structure of the complex. Given the size of CCR5, there are technical limitations associated with using nuclear magnetic resonance (NMR) spectroscopy to solve this structure (Latham et al. 2005; Lu et al. 2010; Varani and Tinoco 1991), even though solving a solution structure could be extremely informative if there are regions of high flexibility that contribute to the conformational switching that I propose may occur when miR-1224 binds (see Chapter VI). Another method of high-resolution structure determination is cryo-EM (Kühlbrandt 2014). Cryo-EM is also capable of identifying multiple structures in solution, but practically, the CCR5-miR-1224 complex is too small to be able to identify these particles through the microscope (Glaeser and Hall 2011). Due to the problems associated with performing NMR and cryo-EM on this complex, I am attempting X-ray crystallography instead. Success in crystallography of RNA is linked to stability and lack of flexible regions within the molecule being studied (Reyes et al. 2009). For these reasons, the biochemical experiments shown in this chapter are extremely informative in designing and screening different constructs of the CCR5 RNA. For example, knowing that the 3’ end is single stranded (according to data from the $M^2$ experiment) and also is unnecessary for miR-1224 binding (according to EMSA and
crosslinking experiments), suggests that I can eliminate these residues from constructs that I try to crystallize. The M² experiments also identified single point mutants that I hypothesize are preformed or stabilized in the miR-bound conformation; it is possible that I can make these specific mutations to stabilize this conformation of the RNA and this may allow for crystallization of the miR-bound form specifically. This raises the issue that, if there is a conformational switch between two RNA structures, crystallography could be used to reveal the unbound or miR-bound structures. Either of these structures would be interesting to observe, however I am personally more interested in the miR-bound version because it promotes more efficient frameshifting and because it might reveal a unique miRNA target recognition strategy.

Finally, an expectation of establishing a detailed molecular understanding of CCR5 structure and its interaction with miR-1224 is that one could design the RNA or the interaction with a microRNA to enhance the frameshifting effect. To assess this possibility, I need to develop a functional assay that evaluates frameshifting efficiency. Additionally, the mutants that I designed in this chapter need to be tested for function; simply because miR-1224 binds the RNA does not mean that the construct is capable of inducing a frameshift. I have obtained dual luciferase vectors from the Dinman lab containing frameshifting signals of viral origin (Harger and Dinman 2003) to use as controls and I am working with a technician in our lab, Andrea MacFadden, to clone CCR5 RNA sequence into these vectors. These dual luciferase constructs have the upstream, *Renilla* luciferase stop codon knocked out, and the *Renilla* ORF is immediately followed by the frameshifting signal of interest. This allows ribosomes to continue to
translate the downstream firefly sequence only if the -1 frameshift occurs. The ratio of the amount of firefly to *Renilla* luciferase produced reports on the frameshifting efficiency of a given construct. I plan to perform these dual luciferase assays in a cell-free translation system such as rabbit reticulocyte lysate. This will allow for the enhanced ability to control the amounts of microRNAs added to the system and minimize the possibly confounding effect of active mRNA decay pathways found in cells. It will be especially imperative to test the constructs that I design for crystallography in this assay, as the structure solved can only be considered truly relevant if it is functional.

**Final Thoughts**

From a functional point of view, it is important to remember that CCR5 exhibits some level of frameshifting capacity on its own, even in the absence of miR-1224. Therefore, miR-1224 binding cannot be thought simply as an “on/off switch”, but rather as a “dimmer switch” instead. But, the observation that miR-1224 binding induces structural changes in CCR5 makes the situation more complicated than simply using microRNA binding to stabilize a preexisting, frameshifting structure. Instead, it raises the question of whether frameshifting from this signal can be elicited from two different structures of varying frameshifting efficiencies (unbound and weak, versus bound and strong) or if there is only one structure that elicits frameshifting that is sampled at a minor frequency in the absence of miR-1224 and at a greater frequency in the presence of miR-1224. In general, I favor the latter interpretation because it is mechanistically parsimonious, suggesting that there is a single active structure that engages the
ribosome to induce a frameshift. Arguing against this is the observation that the M² data do not show that the RNA alone exchanges between multiple structures within the resolution of the experiment. However, if my hypothesis is correct that the alternate structure that appears in the subset of M² point mutants is the “miR-bound” form of the RNA, then it implies that the RNA on its own is capable of forming the relevant frameshifting structure without miR-1224, although mutation is required to capture this state in that assay. If this is the case, it begs the question of what prompts the RNA to switch into the frameshifting structure when miR-1224 is absent? Are interactions with the ribosome important? Answering these questions lies at the heart of understanding how RNA structure-driven processes can be fine-tuned in general and how that tuning can be regulated specifically by microRNAs.

It is also important to consider the role that Argonaute and the RNA-Induced Silencing Complex (RISC) might play in this newly discovered microRNA-regulated process. MicroRNAs are not generally thought to be “free” in the cytoplasm, but rather are pre-bound to Argonaute before associating with their targets in the UTRs of mRNAs (Bartel 2009; Schirle and MacRae 2012; Schirle et al. 2014; Elkayam et al. 2012). Whether miR-1224 associates with Argonaute in the cell—let alone if Argonaute plays a role in directing miR-1224 to the CCR5 -1 PRF signal—is unknown. Knockdown of AGO1 and AGO2 does mildly suppress -1 PRF in CCR5 (Belew et al. 2014), but the reason for this is uncertain. One interpretation of this result is that Argonaute is required to deliver miR-1224 to CCR5. If this is the case, is AGO binding what actually stimulates frameshifting and not the structure of the RNA itself? This seems unlikely since CCR5 can
induce frameshifting at lower levels in the absence of miR-1224 and in AGO knockdown. Furthermore, it is known that part of the structural basis for AGO-miRNPs binding to their targets arises from the pre-configuration of the miRNA seed sequence (Lambert et al. 2011) within Argonaute. However, even though base pairing potential was the means by which miR-1224 was originally identified to possibly target CCR5, there is no obvious seed-matching between miR-1224 and the CCR5 -1 PRF signal. This argues against Argonaute interacting with the miR-1224-CCR5 complex. The contribution of AGO1 and 2 to PRF suppression could be explained by that at some frequency miR-1224 binds AGO (when not bound to CCR5) and this enhances its stability and half-life in the cell. Thus, knocking down AGO decreases the amount of miR-1224 available in the cell to stimulate PRF from CCR5. Of course, knockdown of any component of the RNA silencing machinery could have pleotropic effects on translation in general, and it is easy to speculate on many possible outcomes; more work needs to be done to understand this issue in depth.
CHAPTER VI
DISCUSSION

On the surface, it might not seem that IRESs and frameshifting signals have very much in common—they regulate very different functions, vary greatly in size, and can contact the ribosome in different locations. But, unifying these RNAs is the fact that they use complicated folds to contact conserved features of the ribosome, and it appears that their conformational dynamics are essential to exert their effects in manipulating this complex Brownian machine. These structured RNAs may be vestiges of the RNA World, giving us an opportunity to look back in time and learn how a single class of molecule with only four different building blocks can manifest the structural diversity that led to life as we know it today.

The work presented in this thesis was built upon decades of work that came before it, but there is still much left to learn and many unanswered questions remain. This chapter will outline those remaining questions for the IGR IRESs and the CCR5-1 PRF-miR-1224 projects, contextualize the findings presented, and speculate on new models that arise from the data. Finally, I will discuss the implications of these structured RNA elements and their influence on the free energy landscape of ribosome function.
IGR IRESs

The work I present in this thesis shows that the IGR IRESs cannot simply be thought of as a “ribosome landing pad” – a common moniker for IRESs in the literature. Rather, the IRES controls the execution of defined steps after ribosome recruitment using a conformationally dynamic loop. These steps include three non-canonical translocation events before the IRES no longer interacts in the decoding groove. An important element of this discovery is that differential alteration of this dynamic loop with regard to length and sequence leads to failure at different steps. Why is loop 3 length—but not sequence—important during the first pseudotranslocation? Similarly, why is sequence important in the second pseudotranslocation? These features of loop 3 may also be important at later stages in the IRES mechanism (for example, moving domain III out of the E site), but because they also play important roles earlier, these effects could not be interrogated.

Nevertheless, the observation that loop 3 controls multiple steps in IGR IRES-driven initiation clearly indicates that these transitions are mechanistically distinct processes, involving different ribosomal elements as described in Chapter I. I hypothesize that loop 3 flexibility is important during the first pseudotranslocation event, but specific interactions made through the loop are important in the second pseudotranslocation. It is possible that decreasing the flexibility of loop 3 by deleting nucleotides promotes minor groove interactions between loop 3 and stem 1. This could be problematic if the ability of the ribosome to make minor groove interactions with A-site tRNA is essential for translocation, similarly to the importance of P site anticodon
stem loop minor groove interactions during tRNA movement from the P to E sites.
Additionally, decreasing loop 3 flexibility might bend or twist the RNA in a manner that
decreases its overall structural likeness to an authentic tRNA. This could prevent the
ribosome or eEF2 from recognizing it as a legitimate substrate, and inhibit the first
pseudotranslocation. Furthermore, because the loop 3 sequence mutants share the
same length as the WT these point mutations may have a weaker impact on the
flexibility of loop 3, and therefore the first pseudotranslocation is unaffected in the
sequence mutants.

Moving from the P to the E sites, on the other hand, may require specific
interactions of loop 3 with ribosomal elements, and therefore, this is the point at which
alteration of the sequence of loop 3 presents problems for the IRES. As mentioned in
the discussion section of Chapter III, recent structures of different IRESs bound in
different states of pseudotranslocation point to a few candidates for this interaction,
including the P-E gate. If this is true, this might suggest that these interactions are
mediated by conserved residues within loop 3. But, the G-rich mutant which does not
alter any highly conserved nucleotides also fails at this step, suggesting there is more to
this interaction than is immediately obvious. As an intriguing side note, the class II IGR
IRESs exhibit greater conservation throughout loop 3 than the class I IGR IRESs, and data
presented in Chapter III show that even a single point mutation in loop 3 of this class of
IGR IRESs leads to decreases in translation. With an additional stem loop added to
domain III in the class II IGR IRESs, there may be increased pressure to conserve these
hypothesized interactions of loop 3 with the ribosome, although the reason for this selective pressure is currently unclear.

Assessing IGR IRES and Ribosome Dynamics- Strategies and Challenges

Despite the intricate details of the interactions being made, both sets of mutants point to the idea that the movements associated with translocation are key points of IRES manipulation. Great advances in understanding translocation of tRNAs on the ribosome have recently come from X-ray and cryo-EM investigations of multiple pre- and post-translocation complexes at high resolution, and importantly, from single-molecule fluorescence-based experiments that report on the dynamics of moving parts in the system. As shown in Chapter IV of this thesis, I attempted to solve structures of various IRES-ribosome complexes to address the questions defined above. Not discussed in that chapter are the attempts I made to study these complexes using single molecule Förster Resonance Energy Transfer (smFRET). I tried multiple approaches to site-specifically, fluorescently label the PSIV and CrPV IGR IRESs and the L1 stalk to investigate their interactions with a labeled tRNA. One such labeling strategy is outlined in Fig. 3.10. This labeling strategy ultimately allowed us to perform single molecule co-localization experiments between the IRES and the first tRNA delivered to the complex (Fig. 3.11). However, we never observed FRET between the acceptor and donor dyes in these experiments, likely due to a photophysical problem related to constrained orientations of the dye molecules within the ribosome.
Beyond unfortunate photophysics, several key challenges exist with the single-molecule FRET approach that are both specific to the IGR IRES and general to the study of eukaryotic ribosome complexes, and this merits some discussion. With regard to the IRES, the finding that domain III is first positioned in the A site suggests it requires eEF2 to move into the P site before tRNA delivery, and other data support that domain III can reverse-translocate readily in the absence of A-site tRNA and eEF2 (Fernández et al. 2014). These traits make it very difficult to biochemically build and control IRES-ribosome complexes. For example, one of the main questions we wished to address with this approach is whether the first tRNA delivered can sample the A/P hybrid state upon tRNA delivery. Biochemically, this would require eEF2 to act on the IRES-ribosome complex once, but not twice; practically, this would entail extensive washing steps after adding eEF2 but before tRNA delivery. However, during this washing step the first translocation can readily reverse, preventing subsequent tRNA binding. Even if tRNA could be delivered, it is well documented in the literature that without eEF2 this is a labile complex. This makes it essentially impossible to address the question we set out to answer. Of course, all of these problems can only be appreciated retrospectively and upon gathering new information about the IRES. Regardless, this should underscore that the ability to add factors and control the composition of the ribosome complexes for smFRET experiments is essential, and this is quite difficult given these inherent features of IGR IRES dynamics. These problems are not dwarfed by the additional significant challenge of obtaining a FRET signal from two labeled components in the system within a reasonable distance of each other so that subtle conformational changes can be
detected. Furthermore, the labels need to be in locations that do not disrupt the
function of the complex. Labeling strategies are developed using structure modeling to
approximate distances between the location of FRET pairs, but ultimately whether these
strategies will be successful can only be determined empirically.

Another possibility when doing these types of experiments is to avoid labeling
the IRES itself, and instead label components of the ribosome or elongation factors as
well as tRNAs. Theoretically, this would allow previously described conformational
changes to be monitored as a result of IRES binding, such as subunit rotation. This
strategy has been employed extensively using bacterial ribosomes. However, although
site-specific labeling of the eukaryotic ribosome has been achieved (Petrov and Puglisi
2010), the positions of these labels have not been in a workable FRET range. Apart from
not having FRET pairs suitable for this approach, there are significant difficulties
associated with assembling canonically initiated eukaryotic complexes to use as a
proper comparison to IRES initiated ribosomes due to the complicated mechanism of
eukaryotic translation initiation. These are substantial hurdles, but they are not
insurmountable with continued effort.

Remaining Questions

Despite these difficulties, coupling the structural power of X-ray crystallography
and cryo-EM with the dynamic information from single molecule fluorescence
experiments will allow for a complete mechanistic dissection of how IGR IRESs
manipulate ribosomes. The following questions are, in my opinion, the most interesting to pursue using these methods:

**How does the IGR IRES recruit the 60S subunit and promote formation of intersubunit bridges?:** In prokaryotes and eukaryotes alike, large ribosomal subunit joining requires the conserved initiation factor IF2/eIF5B. Current models suggest that this factor helps to alleviate the electrostatic barrier to subunit joining by masking negatively charged interfaces (Marshall et al. 2009). Then, through GTP hydrolysis, these factors induce conformational changes that bias the establishment of non-rotated intersubunit bridges (Marshall et al. 2009; Acker et al. 2009). Putting the ribosome in a non-rotated state best prepares the ribosome to accept tRNA in the A site at the beginning of the elongation cycle. But, the IGR IRESs do not require tRNA delivery as the first step of their initiation mechanism, possibly making this a moot point for IRES-based initiation. How does the IRES (which is negatively charged itself) bring two massively negatively charged subunits together? Does IRES-driven subunit joining occur primarily through rotated or non-rotated intersubunit bridges?

**How is head swivel achieved and manipulated by the IGR IRES?:** IRES manipulation of the ribosome clearly involves promoting non-canonical translocation events. A key player in translocation is the small ribosomal subunit head. The RBD of the IGR IRES makes contacts to specific ribosomal proteins located at the back of the head (eS25 and uS7), and as discussed previously, domain III interacts with the head as well. These observations strongly suggest that ribosome head conformational dynamics might be a central target for IRES control. Does the IRES bias head dynamics in some way to
promote pseudotranslocation? The HCV IRES also makes extensive contacts with the 40S head on the solvent exposed side and extends a highly conserved domain into the decoding groove where it contacts uS7 as well. Is uS7 a “gatekeeper” to ribosome function, promoting certain head conformations over others?

**CCR5 -1 PRF and miR-1224**

The project surrounding the CCR5 -1 PRF RNA and its interaction with miR-1224 is still in an early stage. However, a key discovery from the work presented here is that the published conceptual model of miR-1224 binding is incorrect. Part of the problem with that model could be that the secondary structure of the RNA is incorrectly drawn. Indeed, data from the M² experiments I performed supports a different, two-hairpin structure, and other biochemical data I generated with mutant CCR5 constructs support this. Yet, it is surprising that the structure predicted from our M² experiments does not contain an RNA pseudoknot, since this is the most widely seen -1 PRF stimulatory structure, and since the CCR5 element was predicted to have an RNA pseudoknot. Because the two-hairpin secondary structure was derived from an experiment without the miR-1224 present, it is possible that miR-1224 binding promotes formation of a pseudoknotted structure. Based on the data from Chapter V, I propose a new, speculative model for the “active” form of the CCR5 -1 PRF element and its interaction with miR-1224 (Fig. 6.1). This model attempts to reconcile all of the data I have collected, as well as reflect ambiguities that remain.
A Speculative Model for the CCR5 -1 PRF Element and Its Binding with miR-1224

One of the characteristic changes in the RNase T1 experiment when miR-1224 is added is a decrease in single-stranded character through nucleotides G16, 17, 64, and 68. In the M² experiment (which was performed without miR-1224), most of the point mutants exhibit a high degree of SHAPE modification through these regions. However, point mutants G13C, G34C, G43C, G51C, G52C, G56C, G64C, and C76G-G81C (see purple boxed lane in Fig. 5.10E for an example of one of these point mutants) all appear to show decreased SHAPE modification in the single stranded region from G12-G17 and in the apical loop of the second stem that contains G64 and G68. Based on these data, I propose that a pseudoknot forms between these regions (indicated by connecting lines in Fig. 6.1). Additionally, in the RNase T1 experiment miR-1224 binding causes G31, 32, 34, 39, 41, 43, 51, 52, and 53 to exhibit more single-stranded character than in the unbound state. Similarly, the M² mutants mentioned above show increases in SHAPE modification at A37 and A38 in some cases, or at U42 and G43 in others. All of these mutants show increases in SHAPE modification at U50, G51, G52, G53, U54, C69, G70, G80, G81, A86, U87, U89, and U90. The RNase T1 experiment is limited to Gs, so it is difficult to compare these experiments directly, but overall the probing patterns for these M² mutants seem to be in line with the RNase T1 data when miR-1224 is bound to the WT CCR5. For these reasons, I propose that these M² mutants allow the RNA to
Figure 6.1 — A Speculative Model for miR-1224 Binding to CCR5
The new model for CCR5 folding and interaction with miR-1224 is based on the data presented in Chapter V. miR-1224 is shown in purple type and the model shows that miR-1224 promotes formation of a pseudoknot between G12-G16 and U67-U63 (solid lines). Gray dashed box shows uncertainty about the interaction of the 5’ end of miR-1224 with the 3’ region of CCR5. For an in depth discussion of this model, please refer to the “CCR5 ‐1 PRF and miR-1224” section of this chapter.
adopt a conformation that miR-1224 binding induces with the WT RNA, and these characteristic changes are incorporated into the new model. It is unclear, however, why mutation to these particular positions promotes this alternate conformation in some cases, particularly since some of the mutants are in regions that would abrogate a miR-induced base pairing interaction such as the G13C mutation (a recognized caveat with my proposed model). On the other hand, some of these point mutants make sense. For example the G51-53C mutants are in a region that becomes more single-stranded when miR-1224 binds according to RNase T1 data. It is possible that by mutating these residues, destabilization of the entire helix is promoted, and this appears to enhance microRNA binding as observed in the EMSA and crosslinking experiments with the 1.S2D mutant. Regardless, the position of these point mutations may identify these residues as essential for the architecture of the unbound form of the RNA, and can possibly be used in the future as tools to interrogate the idea that there exist at least two structures that the RNA can switch between.

Another major finding from this work is that the 5’ end of the CCR5 -1 PRF signal, which includes the linker sequence, is essential for miR-1224 association, but that the 3’ end is dispensable for binding. Destabilizing stem 2 from the original secondary structure does not prevent miR-1224 binding, but may actually promote it. The new model incorporates these observations by showing that the 3’ end of miR-1224 may base pair with G8-U10 and could possibly interact with the pseudoknot immediately following. The new model also shows base pairing between miR-1224 and U75-C78, which explains the increases in T1 cleavage/SHAPE modification through U50-G53 as
discussed above. This modeling also puts U17 of miR-1224 in close proximity to G70, G68, and U66, which were identified in the site-specific UV crosslink mapping experiment. G15 and G16 of miR-1224 do not show specific base pairing partners; it is possible that these nucleotides interact in the major or minor groove of the predicted pseudoknot helix to stabilize that interaction and contribute to the overall affinity of the microRNA for the structure of CCR5.

**Ambiguities in the Model**

There are many elements of this model that are uncertain. It is still unclear where the 5’ end of miR-1224 may interact; therefore, the new model shows the 5’ end of miR-1224 near residues G80-U90 of CCR5, but not necessarily interacting by direct base pairing since mutation of these nucleotides does not abrogate miR association. This ambiguity is reflected in the model as a gray dashed box around this region. It is possible that interaction of the 3’ end of miR-1224 with the 5’ end of CCR5 is sufficient for miR binding. The nature of these interactions and which regions of miR-1224 are necessary and sufficient for binding remain to be thoroughly explored. Nevertheless, the model as it is drawn incorporates the necessity for the 5’ end of CCR5 and dispensability of the 3’ end for miR-1224 association.

The model shows miR-1224 base pairing to sequences that are far apart in sequence in order to bring them close together in space. But, it is unclear how this could occur architecturally: Are RNA triplex interactions able to help bridge the pseudoknot helix? As discussed above, this could be the role of G15 and G16 of miR-1224. Are there
other non-canonical interactions that contribute to the binding? Before these questions can be answered, we need to first know more about the structure of the RNA on its own, and test the existence of the pseudoknot at the 5’ end.

**Additional Support and Precedence for the New Model**

Interestingly, a study of unfolding/disruption pathways of the CCR5 -1 PRF element by single molecule force measurements in an attempt to recapitulate ribosome-induced unwinding identified four different disruption pathways for the RNA (de Messieres et al. 2014). This study showed that the pathways corresponded to four different initial structures, rather than different unfolding pathways for a single starting structure. The dominant pathway depended on all nucleotides within the CCR5 RNA (supporting the published secondary structure), however alternate pathways existed which all depended on the first 20 nucleotides of the RNA, but not on the last 20 nucleotides. None of these experiments were performed in the presence of miR-1224. These studies reported the presence of a helix of roughly 18 base pairs at the 5’ end of the RNA in one of these alternate structures, and attempted to correlate them to M-Fold predicted RNA pseudoknot structures. Although the predicted structures are not precisely the same structure that I propose in the new model, the new model does show 3 helices totaling 18 base pairs that would be unwound before reaching a “break” corresponding to the other side of one of those helices. It is too premature to assess whether the convergence of these data are coincidental, or whether the single molecule studies in fact support the new model proposed here.
Another aspect of this new model is that two overlapping pseudoknots exist in
the structure. Although the classical view of frameshifting signals contain a single
pseudoknot, the -1 PRF signal at the HIV gag-pol junction was recently shown to contain
four pseudoknot structures, two of which are mutually exclusive, meaning they cannot
form at the same time (Huang et al. 2014). In this case, ribosome-mediated unwinding
of pseudoknots upstream of the frameshift-inducing structure permit formation of
intermediate PKs that continually present the ribosome with structural barriers as it
approaches the slippery site. These multiple PKs upstream of the frameshift signal are
thought to act as “speed bumps” that cause the ribosome to slow down before it
encounters the slippery site of the functional frameshifting signal (Huang et al. 2014;
Barry and Miller 2002). If this is the case, it makes sense that the strength of a
frameshifting signal might better be tuned by addition of multiple upstream elements
that slow the ribosome, rather than by strengthening the frameshifting pseudoknot
itself, which could present too difficult of a mechanical barrier for the ribosome to
handle. The precedent of the gag-pol junction not only supports the possibility that two
pseudoknots could exist in the CCR5 model, but also explains why elements in the linker
region might be involved in this “speed bump” structure as my new model indicates,
and why miR-1224’s role may be to promote formation of a “speed bump” rather than
inducing the active frameshifting structure. These are intriguing hypotheses, but they
can only be tested once more information about the structure of the CCR5 -1 PRF signal
is known.
Alternate Routes on the Free Energy Landscape of the Ribosome

The ribosome is a Brownian machine—a molecular complex that harnesses thermally available motions to drive large-scale conformational changes necessary for function. As such, it traverses a free energy landscape as it initiates translation and during each cycle of elongation. It uses initiation and elongation factors to help guide it through the landscape and lower certain energetic barriers. In light of this model, we can conceptualize the function of IRESs and -1 PRF signals as acting within this free energy landscape causing the ribosome to detour away from the “normal” trajectory that it follows in order to manipulate ribosome function. I will focus this discussion on the energy landscape specific to elongation since this is the setting for IGR IRES and CCR5 -1 PRF manipulation.

The IGR IRESs rely on tRNA molecular mimicry to “kick-start” protein synthesis from the middle of an elongation cycle. But, there are still many ways in which they are different from authentic tRNA substrates, and this may subject ribosomes bound to IRESs to a different “terrain” during the first two rounds of translocation. The IRES itself could use structural elements to guide (read: bias) the ribosome to take certain paths rather than others.

The -1 PRF bipartite signals, on the other hand, may influence the rise of a discrete energetic barrier within the otherwise “normal” landscape of ribosome function. This barrier is likely specifically related to the problem of pseudoknot structure unwinding in a single round of the elongation cycle. In this case, the ribosome is detoured off of its canonical path, navigating a new route that is based on the
surrounding energetic options. But, because the canonical path is also the one that is
the most energetically favorable, the ribosome having dealt with the barrier imposed by
the pseudoknot, finds it way back to its normal route through the landscape.

The “routes” and “trajectories” referred to above correspond to specific
conformational changes of the ribosome reflecting various degrees of freedom imparted
by the thermal fluctuations of each feature of the ribosome. Thinking of these RNAs in
this way allows their separation into two classes. One class, illustrated by the IGR IRES,
may completely alter the topography of the free energy landscape (akin to hiking in the
Appalachians versus the Rockies). A second class, illustrated by -1 PRF signals, set up
defined obstacles within the canonical terrain (similar to a tree falling over an
established hiking trail). It may be useful to consider these two classes when attempting
to understand the basis for ribosome manipulation by other ligands.

Concluding Remarks

The title of this thesis proposes that structured RNAs can manipulate the
ribosome at a variety of points throughout the process of translation, and if this is true,
some unifying principles should be able to be extracted about how this can be
accomplished (even if this thesis only presents an “n of 2”). One of these principles is
that pseudoknots are major players; they provide for a wide diversity of structures,
including intermolecular mimicry, as well as mechanical stability. A second principle is
that conformational dynamics are essential. This is not terribly profound: when the
machine in question depends on conformational changes itself, the RNAs that act on it
might also need to be able to respond to these changes, and we know that tRNAs undergo substantial structural changes when bound to the ribosome. Nevertheless, it bears repeating that we must think about the molecules that interact with the ribosome as dynamic entities and although studying them in isolation is very useful, mechanistic information will likely elude us if we do not study them in complex with ribosomes.

Nowhere else in the cell do RNA and protein converge in such an intricate and orchestrated manner to support a fundamental process as in the ribosome and translation (in my humble opinion). The ribosome is indeed an extraordinary machine, yet it is not an outlier and it has many lessons to teach. Understanding how the ribosome works promises to enlighten the study of complexes much smaller and much larger than it, imparting precedence to understand the link between molecular structure and function, and the navigation of changing thermodynamic landscapes.
REFERENCES


Schirle NT, MacRae IJ. 2012. The crystal structure of human Argonaute2. *Science* **336**: 1037–1040.


### APPENDIX A

#### 2.1- Primer Sequences

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13F -41</td>
<td>CGC CAG GGT TTT CCC AGT CAC GAC</td>
</tr>
<tr>
<td>M13R -27</td>
<td>CAG GAA ACA GCT ATG AC</td>
</tr>
<tr>
<td>Internal Photinus R/Toeprint</td>
<td>CGG ATA GAA TGG CGC CGG GCC</td>
</tr>
<tr>
<td>FVKM+BamHI Reverse</td>
<td>ATC GCG GAT CCG GCG GCG GCT TCT TTA TG</td>
</tr>
<tr>
<td>T7+CrpV1-1</td>
<td>TAA TAC GAC TCA CTAG TAG GGA GCA AAA ATG TGA TGC</td>
</tr>
<tr>
<td>PRL-CMV</td>
<td>GCA CCT GAT GAA ATC GGA</td>
</tr>
<tr>
<td>CCR5.2hp_tail2 RNA primer1</td>
<td>TTC TAA TAC GAC TCA CTAG TAG G</td>
</tr>
<tr>
<td>CCR5.2hp_tail2 RNA primer2</td>
<td>TCT ACT CGA GTC GTT TCC TATA GTG AGT CGT AT</td>
</tr>
<tr>
<td>CCR5.2hp_tail2 RNA primer3</td>
<td>ACG ACT CGA GTA GAG TCG AAA ATT TAA AAG CCA GGA CGG</td>
</tr>
<tr>
<td>CCR5.2hp_tail2 RNA primer4</td>
<td>GCC ACC ACCCAA GTG ATC ACA CTT GTC ACC ACCCAA AAG GTG ACC GTC CTG GCT T</td>
</tr>
<tr>
<td>CCR5.2hp_tail2 RNA primer5</td>
<td>TGG G TG GTG GCT GTG TTT GCG TCT CCA GGA ATC ATC TTT ACC AGA TCT CAA GTT</td>
</tr>
<tr>
<td>CCR5.2hp_tail2 RNA primer6</td>
<td>GTT GTT GTT GTT GTT TCT TTT GGA GTC TAC TCG ACT CCA ACT ATC GAT CTG GTA A</td>
</tr>
<tr>
<td>CCR5 G41 RT primer</td>
<td>GCC ACC ACCCAA GTG ATC AC</td>
</tr>
<tr>
<td>CCR5 G80 RT primer</td>
<td>GAT CTG GTA AAG ATG ATT CC</td>
</tr>
<tr>
<td>Internal Delta Reverse</td>
<td>CCA GCG AGG AGG CTG GGA CC</td>
</tr>
<tr>
<td>CCR5-0.89 Reverse</td>
<td>AGA TGA TTC CTG GGA GAG ACG C</td>
</tr>
<tr>
<td>CCR5-1 S2disrupt Reverse</td>
<td>TTG AGA TCT GTG AAT CAC ACT AGT CCT GGG AGA GAC GCA AAC AC</td>
</tr>
<tr>
<td>CCR5 polyU12 mir1224 Reverse</td>
<td>CCA CCT CCC GAG TCC TCA CAA AAA AAA AAA ATT GAG ATC TGG TAA AGA TGA TGA TCC C</td>
</tr>
</tbody>
</table>
## APPENDIX B

### 2.2- Important Sequences for Construct Design, Sequences of Loop 3 Mutants, and Anchor Sequence for Single-Molecule Experiments

<table>
<thead>
<tr>
<th>Other Important Sequences</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7</td>
<td>TAATACGACTCACTATAGGG</td>
</tr>
<tr>
<td>Hammerhead Ribozyme</td>
<td>AGAXXXAAXXAXXXCTGATGAGTCCGTGAGGACGAAACGGTACCCGCTACCGTC</td>
</tr>
<tr>
<td>HDV Ribozyme</td>
<td>GGGCGGCATGGTCCTCCTCGCTGGCGCCGCTGGGCAACATGCTTCGGCATGCGAATGGGACC</td>
</tr>
<tr>
<td>CrPV IGR IRES dIII WT</td>
<td>GGUUUUCAGAUUAGGUAGUCGAAAAAACCUAAGAAAUUACCU</td>
</tr>
<tr>
<td>CrPV IGR IRES dIII Δ1</td>
<td>GGUUUUCAGAUUAGGUAGUCGAAAAAACCUAAGAAAUUACCU</td>
</tr>
<tr>
<td>CrPV IGR IRES dIII Δ2</td>
<td>GGUUUUCAGAUUAGGUAGUCGAAAAAACCUAAGAAAUUACCU</td>
</tr>
<tr>
<td>CrPV IGR IRES dIII Δ3</td>
<td>GGUUUUCAGAUUAGGUAGUCGAAAAAACCUAAGAAAUUACCU</td>
</tr>
<tr>
<td>CrPV IGR IRES dIII G-rich</td>
<td>GGUUUUCAGAUUAGGUAGUCGAAAAAACCUGAGAGGUUACCU</td>
</tr>
<tr>
<td>CrPV IGR IRES dIII GGC</td>
<td>GGUUUUCAGAUUAGGUAGUCGAAAAAACCUGGGAACCUACCU</td>
</tr>
<tr>
<td>CrPV IGR IRES dIII ΔPKI/PKI KO</td>
<td>GGUUUUCAGAUUAGGUAGUCGAAAAAACCUAAGAAUUAAGGU</td>
</tr>
<tr>
<td>5' anchor smF (for tethering IRES to scope at the 5' end)</td>
<td>TGT GTA AGT TTT AGG TTG ATT TG/3Bio/</td>
</tr>
</tbody>
</table>
APPENDIX C

2.3- Summary of Main RNA Constructs: Cloning and Design

Construct Name: CrPV4
Cloned? Yes, pUC19
Ribozymes? Yes, 5’ HH, 3’ HDV
Primers associated with this construct: M13F and R
Sequence 5’ to 3’:
AGCAAAAATGTGATCTTGCTTGAATACAATTTTGTAGGGTTAATAAATTACAAAGTAGTGCTAT
TTTTGTATTTAGGTAGCTATTTAGCTTACGTCAGGATGCTAAGGCGCCACAATATCC
AGGAAGCCCTCTCTGGTTTTTCAGATTAGGATGTCGAAAAACCTAAGAAATTTACCTGCT

Construct Name: CrPV-11
Cloned? Yes, pUC19
Ribozymes? Yes, 5’ HH, 3’ HDV
Primers associated with this construct: M13F and R
Sequence 5’ to 3’: GGUUUUCAGAUUAGGUGUAGCCAAAACCUAAGAAAAUUUACC

Construct Name: CrPV Toeprint
Cloned? No, pCrPV1-1 was used as PCR Template
Ribozymes? No
Primers associated with this construct: T7+CrPV1-1 Forward and Internal Photinus/Toeprint Reverse
Sequence 5’ to 3’:
AGCAAAAATGTGATCTTGCTTGAATACAATTTTGTAGGGTTAATAAATTACAAAGTAGTGCTAT
TTTTGTATTTAGGTAGCTATTTAGCTTACGTCAGGATGCTAAGGCGCCACAATATCC
AGGAAGCCCTCTCTGGTTTTTCAGATTAGGATGTCGAAAAACCTAAGAAATTTACCTGCTAC
ATTTCAAGATACCATGGAAGACGCCAAAAACATAAAAGAAAAAGGCCGCCGCGCATTCTATCC

Construct Name: CrPV FVKM
Cloned? Yes, pUC19
Ribozymes? No
Primers associated with this construct: T7+CrPV1-1 forward, FVKM+BamHI reverse
Sequence 5’ to 3’:
GGGAGACCGGAATTCAAAGCAAAAATGTGATCTTGCTTGAATACAATTTTGTAGGGTTAATAA
ATTACAAAGTAGGTGCTATTTTGTATTTAGGTAGCTATTTAGCTTACGTCAGGATGCTAGT
GGCAGCCGCCCAATATCCAGGAAGCGCCCTCTCTCGGTGTTCCTCAGATTAGGATGTCGAAAAACCT
AAGAAATTTACC TTC GTG AAA ATG ACA TTT CAA GAT ACC ATG GAA GAC GCC AAA AAC
ATA AAG AAA GGC CCG GCC CC

Construct Name: CrPV smFRET 5’ tether
Cloned? Yes, pUC19
Ribozymes? 3’ HDV only
Primers associated with this construct: M13F and R
Sequence 5’ to 3’:
GGGGCAACCTAAAACTTACACAGGCCCCTAAGGAATAAAAAAAAAGCAAAAAATGTGATCTTGCTTG
TAATAACAATTTTGGAGGTTAATAAAAATTACAAAGTAGTGCTATTTTTGTATTTTAGTTAGCTATTT
AGCTTTACGTTCAGGATGCTAGTGCGAGGCCCCAATAATCCAGGAAGCCCTCTCCTGCGGTTTT
TCAGATTAGGTAGTCAAAAAACCTAAGAAATTACCT TTC

Construct Name: CCR5-1
Cloned? Yes, pUC19
Ribozymes? Yes, 5’ HH, 3’ HDV
Primers associated with this construct: M13F and R
Sequence 5’ to 3’:
TTTAAAAGCCAGGACGGTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTG
TTTGCGTCTCTCCAGGAATCATCTTTTACCAGATCTCAA

Construct Name: CCR5-10
Cloned? Yes, pUC19
Ribozymes? Yes, 5’ HH, 3’ HDV
Primers associated with this construct: M13F and R
Sequence 5’ to 3’:
CAGGACGGTCACCTTTGGGGTGGTGACAAGTGTGATCCTTTGGGGTGGTGCTGTGTTTGCGTC
TCTCCAGGAATCATCTTTTACCAGATCTCAA

Construct Name: CCR5-14
Cloned? Yes, pUC19
Ribozymes? Yes, 5’ HH, 3’ HDV
Primers associated with this construct: M13F and R
Sequence 5’ to 3’:
ACGGTCACCTTTGGGGTGGTGACAAGTGTGATCCTTTGGGGTGGTGCTGTGTTTGCGTC
TCTCCAGGAATCATCTTTTACCAGATCTCAA

Construct Name: CCR5-15
Cloned? Yes, pUC19
Ribozymes? Yes, 5’ HH, 3’ HDV
Primers associated with this construct: M13F and R
Sequence 5’ to 3’:
CGGTTCACCTTTGGGGTGGTGACAAGTGTGATCCTTTGGGGTGGTGCTGTGTTTGCGTC
TCTCCAGGAATCATCTTTTACCAGATCTCAA

Construct Name: CCR5-16
Cloned? Yes, pUC19
Ribozymes? Yes, 5’ HH, 3’ HDV
Primers associated with this construct: M13F and R
Sequence 5’ to 3’:
GGTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTGCGTCTCTCCCAGGAATCATCTTTACCAGATCTCTCAA

Construct Name: CCR5-17
Cloned? Yes, pUC19
Ribozymes? Yes, 5’ HH, 3’ HDV
Primers associated with this construct: M13F and R
Sequence 5’ to 3’:
GTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTGCGTCTCTCCCAGGAATCATCTTTACCAGATCTCTCAA

Construct Name: CCR5-1.89
Cloned? No
Ribozymes? 5’ HH only
Primers associated with this construct: M13F and CCR5-0.89 R
Sequence 5’ to 3’:
TTTTAAAGCCAGCAGGTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTGCGTCTCTCCCAGGAATCATCTCTCTCTCTCTCTACTAGTGTGA

Construct Name: CCR5-1.s2disrupt
Cloned? No
Ribozymes? 5’ HH only
Primers associated with this construct: M13F and CCR5-1 s2disrupt R
Sequence 5’ to 3’:
TTTTAAAGCCAGCAGGTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTGCGTCTCTCCCAGGAATCATCTCTCTCTCTCTCTACTAGTGTGA

Construct Name: CCR5-1.polyU12miR1224
Cloned? No
Ribozymes? 5’ HH only
Primers associated with this construct: M13F and CCR5- polyU12 mir1224 R
Sequence 5’ to 3’:
TTTTAAAGCCAGCAGGTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTGCGTCTCTCCCAGGAATCATCTCTCTCTCTCTCTACTAGTGTGA

Construct Name: CCR5-1 probing
Cloned? No
Ribozymes? 5’ HH only
Primers associated with this construct: M13F and internal delta R
Sequence 5’ to 3’:
TTTTAAAGCCAGCAGGTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTGCGTCTCTCTCTCTCTCTCTACTAGTGTGA
GTTTGCGTCTCTCCAGGAATCATCTTTACCAGATCTCAAGGGCGGCATGGTCCCAGCCTCCTCGCTGG

**Construct Name**: WT CCR5.2hp_tail2  
**Cloned**? No, made by primer assembly  
**Ribozymes**? no  
**Primers associated with this construct**: see Appendix 2.1  
**Sequence 5’ to 3’**: TTCTAATACGACTCACTATAGGAACGACTCGAGTAGAGTCGAAAATTTAAAAGCCAGGACGTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTGCGTCTCTCCAGGAATCATCTTTACCAGATCTCAAGTTGGAGTCGAGTAGACTCCAACAAAAGAAACAACAACAACACACG

**Construct Name**: miR-1224  
**Cloned**? No, synthesized  
**Ribozymes**? No  
**Primers associated with this construct**: N/A  
**Sequence 5’ to 3’**: GUGAGGACUCGGGAGGUGG

**Construct Name**: 6144U  
**Cloned**? No, synthesized  
**Ribozymes**? No  
**Primers associated with this construct**: N/A  
**Sequence 5’ to 3’**: AAUAUCUAGUGUACCCUCGUGCUC
APPENDIX D

5.1- Predicted Structures from 1D Chemical Probing Data