The structure and function of the HIV-1 frameshift site stem-loop

By

Kathryn D. Mouzakis

A dissertation submitted in partial fulfillment

Of the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 04/05/2013

The dissertation is approved by the following members of the Final Oral Committee:
John L. Markley, Professor, Biochemistry
Ann C. Palmenberg, Professor, Biochemistry
Douglas B. Weibel, Assistant Professor, Biochemistry
Karen M. Wassarman, Professor, Bacteriology
Samuel E. Butcher, Professor Biochemistry
Abstract

Human immunodeficiency virus type I (HIV-I) requires a programmed -1 ribosomal frameshift for translation of nearly one-third of its genome. This frameshifting event is driven by two elements in the HIV-1 messenger RNA: a heptanucleotide slippery sequence (UUUUUA) followed by a downstream 11 base-pair stem-loop. Together, these elements induce frameshifting during approximately 5% of translation events of the structural open reading frame. This frequency ultimately dictates the ratio of structural (0 frame) to enzymatic (-1 frame) proteins produced. The goal of my work is to determine the relationship between the stem-loop structure and its function in translational reprogramming.

First, I studied the functional role of the 11 base-pair stem-loop during the -1 frameshift event. A strong correlation between frameshift efficiency and the local thermodynamic stability of the first 3-4 base-pairs in the stem-loop base was observed. My data demonstrated that frameshifting is driven by a “thermodynamic block” to translocation: the stronger the thermodynamic block, the greater the amount of frameshifting. This thermodynamic block is determined by the local stability of stem-loop base-pairs proximal to the ribosomal mRNA entrance tunnel during frameshifting.

In HIV-1, the frameshift site stem-loop folds into an extended structure, with the 11 base-pair stem-loop separated by a three-purine bulge from an 8 base-pair helix. I used this RNA as a model to investigate the dynamic properties of RNAs with purine-rich bulges, a common structural motif. Using NMR spectroscopy, the interhelical dynamics within this extended structure were examined as a function of potassium and magnesium concentration. The stem-loop was found to undergo large-scale interhelical motions under low-ionic strength conditions that are quenched by physiological concentrations of magnesium. Comparison of my results to
previous studies suggested that the impact of monovalent and divalent ions on RNA domain dynamics is a general phenomenon and independent of bulge sequence.

In summary, my work demonstrated that the local stability of the HIV-1 frameshift site 11 base-pair stem-loop is critical to frameshift stimulation. My findings also show the effect of solution conditions on the inherent dynamics of the helical elements within the extended structure. Continued structural and biochemical investigations of the HIV-1 frameshift site, based on my findings, will clarify the relationship between -1 frameshift efficiency and viral replication, and the importance of structural characteristics, such as mechanical stability, in -1 frameshift stimulation.
Acknowledgements

Over the past five and a half years I had the absolute pleasure of doing my graduate work in the Department of Biochemistry at the University of Wisconsin – Madison. I would first and foremost like to thank my advisor, Sam Butcher, for providing the opportunity to work on HIV-1 frameshifting in his lab. Sam pushed me to be a better scientist, and for that I am thankful. I am especially grateful for his willingness to let me pursue my outside interests in teaching, running, and coaching, while maintaining a high-level of research productivity. The time he gave me to develop as a teacher and mentor was instrumental for my career.

My committee members, Dr. John Markley, Dr. Doug Weibel, Dr. Karen Wassarman, and Dr. Ann Palmenberg contributed invaluable feedback and direction to my graduate work. My work required a significant amount of help and support from the staff of NMRFAM, including Dr. Larry Clos II, Dr. Marco Tonelli, Dr. Milo Westler, and Dr. Mark Anderson. I would additionally like to thank the Marchel Hill and other members of the Palmenberg lab for my continued use of their luminometer and many enjoyable conversations.

My HHMI Teaching Fellow mentors, Dr. Katrina Forest and Sarah Miller, were an inspiration to me as an instructor. I am grateful for their support, friendship, and mentoring. The time I spent with my HHMI Teaching Fellow colleagues, Mindy Wesley, Mellissa Cordes, Dr. Brent Berger, Dr. Gilbert Jose, and Dr. Julie Keating, is unforgettable.

The past and present members of the Butcher Lab made coming to lab every day a pleasure. Their insights, encouragement, and occasional diversions were helpful and refreshing. I would especially like to thank my lab mate, Dr. Jordan Burke, for her constant support as a friend and colleague. I am thankful to Dr. Ryan Marcheschi and Dr. Larry Clos II for their scientific expertise and assistance with data collection and experimental design. I am forever grateful to the undergraduate students I had the opportunity to work with. Andrew Lang and Preston
Easterday contributed directly to my work, and my experience working with them gave me a foundation as a mentor.

I would like to thank my immediate and extended family for their constant support, encouragement, and unconditional love. They gave me perspective when I needed it most and pushed me to follow and realize my dreams. Specifically, my father, John Mouzakis, step-mother, Christina Mouzakis, twin-sister, Liz Elhaj, brother, Dennis Mouzakis, sister-in-law, Courtney Mouzakis, and brother-in-law, Rick Elhaj, all experienced the ups and downs of my graduate career with me and I am truly grateful we were able to share the experience. I am also thankful for my extended family in Madison, who truly made me feel like one of their own.

When I look back on my time in Madison it will always be with fond memories of the friends I made and the adventures I had. Although there are too many people to mention, I specifically recognize Dr. Amanda Hickman, Dr. Marielle Gruenig, Dr. Khanh Ngo, Dr. Jordan Burke, and Jonathan Neumann. These people were especially amazing, loyal, and supportive friends over many years and I am very grateful to have each of them in my life.
# Table of Contents

Abstract ............................................................................................................................................. i

Acknowledgements .............................................................................................................................. iii

Table of Contents ............................................................................................................................... v

List of Tables ...................................................................................................................................... xii

List of Figures ................................................................................................................................... xiii

Chapter 1: Introduction ...................................................................................................................... 1

1.1 Ribosome fidelity .......................................................................................................................... 5

1.1.1 Errors in decoding .................................................................................................................... 11

1.1.1 Errors in processivity ................................................................................................................. 11

1.2 Canonical vs. non-canonical initiation ......................................................................................... 13

1.2.2 Cap-independent non-canonical initiation ................................................................................. 19

1.2.1 Cap-dependent non-canonical initiation ..................................................................................... 20

1.3 Canonical vs. non-canonical elongation ....................................................................................... 24

1.3.1 +1 PRF ...................................................................................................................................... 25

1.3.2 -1 PRF ..................................................................................................................................... 26

1.3.3 -1 PRF mechanism ..................................................................................................................... 31

1.3.4 Modulators of -1 frameshift efficiency ...................................................................................... 35

1.4 Canonical vs. non-canonical termination ..................................................................................... 37

1.4.1 Readthrough ............................................................................................................................. 37
Chapter 2: HIV-1 frameshift efficiency is primarily determined by the stability of base-pairs positioned at the mRNA entrance channel of the ribosome. 

2.1 Abstract

2.2 Introduction

2.3 Materials and methods

2.3.1 Plasmid construction

2.3.2 RNA synthesis and purification

2.3.3 Frameshift assay

2.3.4 ΔG\text{Global} and ΔG\text{Local} measurements

2.3.5 Correlation between frameshift efficiency and RNA stability

2.3.6 Modeling the HIV-1 frameshift site onto the eukaryotic ribosome

2.4 Results

2.4.1 Frameshift efficiency correlates with local stability of the HIV-1 frameshift site stem-loop

2.4.2 Influence of spacer length in frameshifting

2.4.3 The HIV-1 3HJ modulates frameshift efficiency

2.5 Discussion
Chapter 3: Investigating RNAs involved in translational control by NMR and SAXS

3.1 Introduction

3.1.1 From RNA secondary structure prediction to biophysical studies

3.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

3.2.1 NMR analysis of base-pairing interactions and secondary structure

3.2.2 Towards solving NMR structures

3.2.3 Solution structures of large RNAs

3.2.4 RNA dynamics

3.3 RNA structure analysis by small angle X-ray scattering (SAXS)

3.3.1 Sample requirements for SAXS

3.3.2 Interpreting SAXS Data

3.3.3 Ab initio modeling of molecular shapes from SAXS data

3.3.4 All-atom molecular modeling

3.3.5 Combining SAXS and NMR

3.4 Using NMR and SAXS to study RNAs involved in translational control

3.4.1 Programmed ribosomal frameshift sites

3.4.2 Untranslated regions
Chapter 4: Domain motions in the HIV-1 frameshift site RNA

4.1 Abstract

4.2 Introduction

4.3 Materials and Methods

4.3.1 RNA synthesis and purification

4.3.2 NMR spectroscopy

4.3.3 Order tensor analysis

4.3.4 Fluorescence-monitored nucleotide stacking

4.4 Results

4.4.1 The FS-SL is dynamic in variable concentrations of potassium

4.4.2 Magnesium promotes coaxial alignment and rigidification of the FS-SL structure.

4.5 Discussion

4.5 Acknowledgements

Chapter 5: Conclusions and future directions

5.1 Conclusions

5.1.1 Overview

5.1.2 Investigation of HIV-1 -1 PRF

5.1.3 Characterization of HIV-1 Frameshift Site Stem-loop Domain Motions
5.2 Future directions ........................................................................................................................................... 166

5.2.1 Examining HIV-1-1 PRF in vivo .................................................................................................................. 166

5.2.2 Analysis of the relationship between stem-loop mechanical stability and -1 PRF ....................................... 168

5.2.3 Investigation of the HIV-1 three-helix junction .......................................................................................... 169

5.2.4 HIV-1 stem-loop dynamics in near physiological solution conditions ....................................................... 170

Appendix I: Structure of the IAPV IGR IRES PK1 domain ........................................................................... 171

A1.1 Overview ......................................................................................................................................................... 172

A1.2 Materials and methods ................................................................................................................................. 175

A1.2.1 RNA sample preparation .......................................................................................................................... 175

A1.2.2 NMR data collection ................................................................................................................................. 175

A1.2.3 SAXS data collection ................................................................................................................................. 176

A1.2.4 Ab initio structure calculation .................................................................................................................. 176

A1.2.5 Molecular modeling and filtration of models by SAXS and RDCs ......................................................... 177

A1.2.6 Molecular modeling and filtration of PKIΔ56-70 models by SAXS and RDCs ........................................ 177

A1.3 Results ............................................................................................................................................................... 179

A1.3.1 Global structure of the IAPV IGR IRES PKI domain ............................................................................. 179

A1.3.2 NMR spectroscopy of PKI ......................................................................................................................... 184

A1.3.3 Modeling the structure of PKI in solution ................................................................................................. 190

A1.4 Conclusions and future directions ............................................................................................................. 194
A1.4.1 Conclusions .................................................................................................................. 194
A1.4.2 Future directions .......................................................................................................... 194

Appendix II: HIV-1 frameshift site stem-loop domain motions in the presence of cobalt-
hexamine and doxorubicin ................................................................................................. 196

A2.1 Overview ....................................................................................................................... 197
A2.2 Materials and methods ................................................................................................. 198
A2.2.1 DNA template design and synthesis ...................................................................... 198
A2.2.2 RNA synthesis and purification .............................................................................. 198
A2.2.3 NMR spectroscopy ................................................................................................. 199
A2.2.4 NMR analysis ......................................................................................................... 199
A2.2.5 Fluorescence-monitored titrations ....................................................................... 204
A2.3 Results .......................................................................................................................... 205

Appendix III: Structure of the HIV-1 frameshift site three-helix junction ......................... 207
A3.1 Overview ....................................................................................................................... 208
A3.2 Materials and Methods ............................................................................................... 209
A3.2.1 Plasmid Construction ............................................................................................. 209
A3.2.2 RNA Synthesis and purification ............................................................................ 209
A3.2.3 NMR Spectroscopy ............................................................................................... 210
A3.3 Results .......................................................................................................................... 211
A3.3.1 The 3HJ does not stably form in solution ............................................................... 211
References

216
List of Tables

Table 2-1. Sequences inserted into the p2luc vectors for plasmid construction.................................52
Table 2-2. Overall and local thermodynamic stabilities (kcal/mol) at 37 °C.................................58
Table 2-3. In vitro frameshift efficiencies.........................................................................................65
Table 2-4 Correlations between ΔG_{Local (N bp)} and frameshift efficiency..............................................67
Table 4-1 HIV-1 FS-SL RDCs measured in 20 mM potassium cationic conditions .......................139
Table 4-2 HIV-1 FS-SL RDCs measured in 150 mM potassium cationic conditions ....................140
Table 4-3 HIV-1 FS-SL RDCs measured in 20 mM potassium and 2 mM magnesium cationic conditions .................................................................................................................................141
Table 4-4 Order tensor analysis of RDCs measured in the HIV-1 FS-SL RNA ..........................142
Table A1- 1 Filtering of structural models of PKI........................................................................192
Table A2 - 1 HIV-1 FS-SL RDCs measured in 20 mM potassium and 0.5 mM cobalt-hexamine. ........................................................................................................................................202
Table A2 - 2 HIV-1 FS-SL RDCs measured in 20 mM potassium and 1.3 mM doxorubicin ....203
Table A2 - 3 Order tensor analysis of RDCs in the HIV-1 FS-SL RNA..............................................206
Table A3- 1 DNA template sequences used for run off transcription............................................210
List of Figures

Figure 1-1 Features of the 80S ribosome complex ................................................................. 4
Figure 1-2 Translation elongation .............................................................................................. 8
Figure 1-3 Errors in translation. .................................................................................................. 10
Figure 1-4 Model of the canonical pathway of eukaryotic translation initiation .................... 16
Figure 1-5 Examples of non-canonical translation .................................................................. 18
Figure 1-6 Canonical -1 PRF ................................................................................................... 30
Figure 1-7 Four models of the state of the ribosome just prior to a -1 frameshift .................... 34
Figure 1-8 The HIV-1 genome and site of -1 PRF .................................................................. 42
Figure 2-1 The HIV-1 frameshift site. ....................................................................................... 50
Figure 2-2 Determination of ΔG ............................................................................................. 59
Figure 2-3 Measured ΔG_{Global} values vs. predicted values. .................................................. 63
Figure 2-4. Frameshift efficiency is plotted as a function of RNA stability. ............................ 66
Figure 2-5. The role of the spacer in frameshift stimulation. ..................................................... 71
Figure 2-6 Defining the relationship between stem-loop local stability and frameshift efficiency. .................................................................................................................................................... 72
Figure 2-7 HIV-1 frameshift efficiencies in the context of a 7 nt spacer. .................................. 73
Figure 2-8. The HIV-1 frameshift site stem-loop modeled onto the eukaryotic ribosome. ......... 75
Figure 2-9. Effect of the 3HJ on frameshift efficiency. ............................................................ 78
Figure 3-1. ^1H 1D spectrum of a 32 nt RNA. ............................................................................ 91
Figure 3-2. ^1H-^1H 2D NOESY experiment. ........................................................................... 93
Figure 3-3. The HNN-COSY experiment directly detects base-pairing by correlating the acceptor and donor nitrogen atoms involved in the hydrogen bond................................. 94
Figure 3-4. RDCs (D_{ij}) between spins i and j provide long-range distance restraints for the average orientation (θ) of the internuclear vector relative to the magnetic field (B_0). .............. 95
Figure 3-5. NMR measurements used to study picosecond (ps) to second (s) RNA dynamics. 101
Figure 3-6. Small angle X-ray scattering (SAXS) data. ................................................................. 107
Figure 3-7. The Kratky plot, q^2I vs. q, indicates the extent of folding within an RNA molecule. ................................................................. 111
Figure 3-8. The pair distance distribution function (PDDF) or p(r) reflects the shape of the molecule. ............................................................................................................................. 112
Figure 3-9. Homology modeling and refinement of tRNA^Val using SAXS and RDC measurements. ............................................................................................................................. 116
Figure 3-10. Modeling the structure of an RNA with MC-Sym. ............................................... 117
Figure 3-11. NMR and SAXS refinement of an RNA:RNA complex ........................................ 119
Figure 3-12. Solution structure of the TCV RBSE [273] .......................................................... 124
Figure 4-1 RNA constructs utilized for investigation of HIV-1 FS-SL dynamics. ................. 134
Figure 4-2 Secondary structure of the HIV-1 FS-SL and measured RDC data ...................... 148
Figure 4-3 Order tensor analysis and the derived average orientations of the FS-SL in variable potassium and magnesium concentrations. ............................................................. 150
Figure 4-4 Bulge nucleotide stacking. ..................................................................................... 154
Figure 4-5 RNA dynamics revealed by resonance intensities. ............................................. 156
Figure 4-6 Magnesium exchange broadening in non-helical regions of the SL ................. 159
Figure A1-1 Proposed secondary structure of the IAPV IGR IRES PKI domain. .............. 174
Figure A1-2 SAXS of IAPV PKI and PK1Δ56-70. ................................................................. 181
Figure A1-3 Ab initio structures of PKI and PK1Δ56-70....................................................... 182
Figure A1-4 Identification of helices within the PKIΔ56-70 envelope................................. 183
Figure A1-5 Secondary structure of PKI as determined by NMR ....................................... 186
Figure A1-6 ^1H and ^15N imino chemical shift assignments for PKI .................................. 187
Figure A1-7 Secondary structure of the 55-nt PK1Δ56-70 RNA as determined by NMR .... 188
Figure A1-8 $^1$H and $^{15}$N imino chemical shift assignments for PK1Δ56-70.................................189
Figure A1-9 Filtering structural models by SAXS and NMR data.........................................................191
Figure A1-10 Preliminary model of the PKI structure filtered using SAXS and RDC measurements.........................................................................................................................193
Figure A3-1 Three-helix junction (3HJ) constructs used for NMR..........................................................212
Figure A3-2 Overlay of the 3HJ* and 3HJ P2 $^1$H-$^1$H 2D NOESY and $^1$H 1D experiments........214
Figure A3-3 Base-pairing inferred from $^1$H-$^1$H NOESY experiments................................................215
Chapter 1: Introduction

When the central dogma of molecular biology was developed, RNA was considered an intermediate that mainly served as a message for protein production. However, the discoveries of reverse transcriptase, catalytic RNA, metabolite-sensing RNA (e.g. riboswitches), regulatory RNA (e.g. miRNA), and cis-acting RNA elements (e.g. internal ribosome entry and programmed ribosomal frameshift sites) revealed other diverse functions of RNA. Cis-acting RNA elements in viral messenger RNA (mRNA) have important roles in the translational regulation required for successful viral replication. Structural and functional studies of RNA elements involved in these processes are critical for understanding how unique RNA sequences adopt highly specific functions. Further, such studies can lead to development of effective strategies to inhibit viral replication.

The ribosome is an RNA and protein (ribonucleoprotein) complex (Figure 1-1) that is responsible for protein synthesis. The ribosome core structure is universally conserved across prokaryotes and eukaryotes and functions to decode mRNA information and catalyze peptide bond formation [1]. Highly structured ribosomal RNAs (rRNAs) are responsible for this catalysis [2, 3]. During translation, the ribosome decodes mRNA in tri-nucleotide steps [4], with each consecutive set of three nucleotides making up a single codon. Each codon in turn defines the specific amino acid to be incorporated into the growing polypeptide chain. The translational reading frame is defined by the location of the 5’ start codon (usually AUG) in the mRNA and dictates the triplets by which the mRNA code is deciphered. Translation of all mRNAs requires three different phases: initiation, elongation, and termination.

Translational reprogramming is defined as one or more a directed changes to the process of translation. Such events can occur during all three phases of translation and are often directed
by cis-acting RNA elements and occasionally by trans-acting factors [5, 6]. Whereas changes to canonical translation processes can be detrimental to protein synthesis, multiple strategies for controlling post-transcriptional gene expression also can be advantageous for an organism. As compared to transcriptional regulation, translation control at the mRNA level can allow for faster changes in the concentrations of encoded proteins, and consequently, can regulate the maintenance of cellular homeostasis in times of stress [7].

In this chapter, several examples of non-canonical translation processes are discussed. Central to each mechanism, is the structure and function of the cis-acting RNA elements involved in the programmed translational event. For perspective the function of these elements, during canonical translation and ribosome fidelity is also discussed. While many of these non-canonical examples relate specifically to viral translation, cis-acting RNA elements are also important to the translation of cellular mRNAs [6, 8, 9]. Although these functions may vary depending on the mRNA context [8, 10], translation regulation via cis-acting RNA elements is essential to all organisms, whether they be humans or yeast [6].
Figure 1-1 Features of the 80S ribosome complex

(A) The 80S ribosome complex. Ribosomes are composed of two subunits. The 60S and 40S from eukaryotes are shown. At the subunit interface, three cavities serve as binding sites for transfer RNAs (tRNAs). In the ribosomal aminoacyl site (A-site), a tRNA is used to decode an mRNA codon. In the ribosomal peptidyl-transfer site (P-site), the ribosome orients the peptide end of the tRNA proximal to the peptidyl transfer center, then maintains base-pairing between the tRNA anticodon and mRNA codon. Within the peptidyl transfer center, which overlaps the A and P-sites, the ribosome catalyzes peptide bond formation between the peptidyl-tRNA nascent polypeptide chain C-terminus, and the aminoacyl group of the amino acid bound to the A-site tRNA. The 60S subunit has an exit tunnel for the growing polypeptide chain. Following catalysis, the ribosome expels the P-site tRNA after deacylation, and subsequent translocation to the exit-site (E-site). The 40S subunit has channels for mRNA entry and exit. During translation initiation, mRNA is wrapped around the 40S neck then threaded through a tunnel between the 40S neck and shoulder [11-15]. The mRNA travels through the A- and P-sites on the 40S surface and it exits near the solvent side of the 40S platform [11-15].

(B) Structures of the 80S Saccharomyces cerevisiae ribosome (PDB ID 3U5B, 3U5C, 3U5D, and 3U5E), combined with tRNA and mRNA from the 30S Thermus thermophilus structure (PDB ID 3I8G). In this depiction, rRNA and ribosomal proteins are shown in dark and light grey, for the large and small subunits, respectively. The tRNAs of 30S Thermus thermophilus, PDB ID 3I8G, occupy the A, P, and E-sites are shown in blue, light green, and orange, respectively. The mRNA is shown in light yellow.

(C) The 60S and (D) 40S subunits are shown with the tRNAs and mRNA. The orientation of these subunits is identical to (B).
1.1 Ribosome fidelity

Successful protein synthesis is dependent upon accurate decoding of the mRNA. This requires that ribosomes faithfully and efficiently convert genetic information, in the form of RNA, into protein sequence. Because increased rates of translation results in decreased ribosomal accuracy [16], a balance between these opposing objectives leads to the observation that ribosomes make approximately $10^{-4}$ errors per codon [17]. These errors result in the rare incorporation of incorrect amino acids into the growing polypeptide chain.

Errors in decoding and processivity typically occur during the elongation phase of translation (Figure 1-2). During elongation the ribosome decodes mRNA codons by binding appropriately aminoacylated tRNAs (aa-tRNAs). Each tRNA has a 3 nucleotide anticodon loop at one end that is complementary to specific mRNA codons. A covalently attached amino acid, specific to that codon, is at the other end of the tRNA (Figure 1-3A cognate aa-tRNA).

The genetic code has 64 codons; 61 code for 20 amino acids and three code for stop signals (reviewed in [18]). Most amino acids are encoded by more than one codon, making the genetic code degenerate. 40 distinct tRNAs are required for translation. The difference in number of tRNAs and codons is resolved by tRNAs that pair with more than one codon sequence. Increased decoding capacity is achieved by way of post-transcriptional modifications to the first position within and the first nucleotide 3’ of a tRNA anticodon loop (reviewed in [18, 19]). In one example, modification of the first nucleotide in this loop, which forms a base-pair with the third nucleotide in the mRNA codon, to an inosine nucleoside allows wobble base-pair formation between the inosine and an A, U, or C nucleotide within the codon [19].
Figure 1-2 Translation elongation

Eukaryotic translation elongation is depicted in eight steps. The large and small ribosomal subunits are illustrated in grey, with the A, P, and E-sites indicated. The mRNA (black), nascent polypeptide (white spheres), and next amino acid to be added (yellow sphere) are shown. tRNAs are depicted as inverted “T”s (black), and have a single sphere attached to their termini when they are aminoacylated. The process of elongation is as follows: (1) the ribosomal A-site is bound by the aa-tRNA-eEF1A(GTP) complex. (2) Accurate base-pairing between the A-site codon and aa-tRNA anticodon triggers eEF1A(GTP) hydrolysis. This is coupled to a conformational change in the ribosome that (3) results in eE1FA(GDP) dissociation and aa-tRNA reorientation. (4) The ribosome catalyzes peptide bond formation between the α-amino group of the A-site tRNA amino acid and the ester carbon of the P-site tRNA polypeptide chain. (4) tRNAs adopt hybrid P/A and E/P conformations, while (5) EF-2(GTP) binding and (6) EF-2(GTP) hydrolysis promotes ribosome translocation. EF-2(GDP) dissociation (7) followed by deacylated tRNA exit [20] (8), resets the ribosome for further rounds of elongation.
Figure 1-3 Errors in translation.
The three most commonly occurring errors in translation are shown. Component coloring is with the same as Figure 1-2. Wobble and Watson-Crick base-pairs are shown as red dots or lines, respectively. (A) Selection of a near-cognate tRNA during accommodation, in place of cognate tRNA, can result in the incorporation of an incorrect amino acid (yellow sphere) into the growing peptide chain. (B) Premature peptidyl-tRNA dissociation results in the synthesis of a truncated protein if the mRNA has not been completely translated. (C) A spontaneous change in reading frame by the ribosome is rare, but would result in synthesis of a C-terminally incorrect polypeptide. When translation continues in the alternative reading frame incorrect amino acids (the first one is shown in red) are incorporated into the growing peptide chain until the ribosome encounters a stop codon and translation is terminated.
1.1.1 Errors in decoding

The most frequent translational error is a decoding error where the ribosome incorrectly selects an aa-tRNA during the aa-tRNA accommodation steps in elongation (Figure 1-2 steps 1-3) [21-24]. Inaccurate tRNA selection can result in incorporation of an incorrect amino acid (Figure 1-3A). When the ribosome initially binds the aa-tRNA in its A-site, it employs kinetic proofreading (Figure 1-2 step 1) to differentiate between cognate tRNAs (no mismatches), near-cognate tRNAs (one mismatch), and noncognate tRNAs (two or more mismatches) (reviewed in [25]). Such tRNAs are distinguished by their rates of codon-binding and eEF1A-GTPase activation. Cognate tRNAs have fast on and slow off rates for binding [26, 27], while near-cognate tRNAs have over two orders of magnitude faster off-rates for binding and slower GTPase activation rates. The difference in these rates for cognate and near-cognate tRNAs improves accuracy in cognate aa-tRNA selection (reviewed in [25]).

To increase translational accuracy, the ribosome employs an additional proofreading step after tRNA accommodation and before the peptidyl transfer reaction occurs (Figure 1-2 step 4) [21, 24, 25, 28-31]. This step is necessary because conformational changes during ribosome accommodation of tRNAs do not distinguish between cognate and near-cognate tRNAs [24]. Structural data suggest that the conformational change in the small subunit constrains the geometry of the decoding center such that the first two base-pairs in the codon-anticodon complex are limited to Watson-Crick geometries. If wobble pairs or mismatches in the first two base pairs are forced into this position, distortions in the anticodon-codon mini-helix promote the efficient rejection of the near-cognate tRNA from the ribosome [24, 27].

1.1.1 Errors in processivity

Faithful maintenance of base pairing between mRNA codons and tRNA anticodons is critical for correct mRNA decoding. Processivity errors can arise from premature P-site tRNA
dissociation (Figure 1-3B) or changes in reading frame (also known as translational frameshifting) (Figure 1-3C). Both types of errors produce truncated and typically non-functional proteins [32]. After frameshifting, the ribosome will continue elongation in the alternative frame and incorporate incorrect amino acids into the growing peptide chain until it encounters a stop codon and translation is terminated. Alternative reading frames frequently contain stop codons (1 in 20 codons in E. coli [32]) ensuring the quick termination of inaccurate mRNA translation.

Despite the fact that the ribosome undergoes large-scale inter-subunit movements during translocation [33-36], the ribosome has evolved to maintain the integrity of the anticodon-codon complex such that the possibility of frameshifting is minimized, with an error frequency of less than $10^{-5}$ per codon, or less than 10% of all ribosomal errors [32]. In a prokaryotic ribosome, extensive interactions between the ribosome and A- and P-site tRNAs stabilize the anticodon interactions with the mRNA codons [29, 37]. Additional contacts between rRNA and ribosomal proteins also take place downstream and upstream of the decoding center (Figure 1-1A) [38]. Immediately downstream of the A site, the 16S rRNA (18S in eukaryotes) and ribosomal protein S3 (also S3 in eukaryotes [1]) make extensive contacts with the mRNA through hydrogen bonding and base stacking interactions [29]. These contacts are preceded by the sequence-independent hydrogen bonding of ribosomal protein basic side chains to the mRNA phosphate backbone [38]. Upstream of the P-site, the first base in the E-site codon (the nucleotide in the -3 position) base pairs with the tRNA anticodon (Figure 1-2) while the 16S rRNA hydrogen bonds with the phosphate group of the -4 mRNA nucleotide [38]. Finally, numerous naturally occurring post-transcriptional modifications to the tRNA anticodon and nucleotides 3’ of the anticodon are important for maintenance of the translational reading frame [39-42]. These complex layers of interaction between the ribosome, tRNAs, and mRNAs ensure reading frame maintenance.
1.2 Canonical vs. non-canonical initiation

Canonical eukaryotic translation initiation takes place on mRNAs with a 7-methylguanosine cap (m7G) and 3’ poly(A) tail. Four major successive events are required to initiate translation reviewed in [43]). First, the mRNA is activated for translation. The 5’ cap is recognized by eukaryotic initiation factor (eIF) 4E, which then recruits several other initiation factors (eIF4G, eIF4A, and eIF4B) and the 3’ end of the mRNA via polyA binding protein. Together this complex recruits a pre-formed 43S preinitiation complex into the second stage of initiation (Figure 1-4). The 43S preinitiation complex is composed of the 40S small ribosomal subunit, eIF2-Met-tRNAi-GTP (ternary complex), and four additional initiation factors (eIF3, eIF1, eIF1A, and eIF5). The mRNA-bound factors and 43S preinitiation complex consolidate together on the mRNA to form a 48S initiation complex, which then scans the mRNA, 5’ to 3’, until it encounters the first AUG codon [43, 44]. During the third phase of initiation, the 48S complex localizes the AUG codon into its P-site, promoting eIF2(GTP) hydrolysis, and locking the ribosome in a closed conformation, with the methionine initiator tRNA (Met-tRNAi) in the P-site. Finally, the 60S ribosomal subunit joins the 40S, in a process that displaces several initiation factors (eIF1, eIF2-GDP, and eIF5), eIF5B GTP hydrolysis mediates the release of eIF1A and eIF5B(GDP) [43]). The newly formed 80S ribosome bound to mRNA, at the correct AUG, is now competent to begin translation elongation.

The majority of translation regulation occurs during initiation (Figure 1-4) [7]. For example, in times of cellular stress, such as viral infection or apoptosis, cap-dependent translation is inhibited by phosphorylation, proteolysis, and sequestration of initiation factors (reviewed in [6, 45]). While advantageous for the cell defenses, global inhibition of translation limits the synthesis of proteins required for the stress response or for cellular recovery. Thus, the occasional need for selective mRNA translation necessitates alternative mechanisms of translation initiation. Growing evidence shows that in these situations, alternative mechanisms
of initiation maintain the translation of specific mRNAs [6]. What better way is there to ensure that non-canonical initiation occurs only on intended mRNAs than to mark them with a signal within the transcript? RNA elements critical to non-canonical initiation events are often located in the 5′ untranslated region (UTR) (Figure 1-5). These include structural elements that can interact with trans-acting factors, internal ribosome entry sites (IRESes), and regulatory upstream open reading frames.
Figure 1-4

mRNA activation

43S preinitiation complex

43S attachment to mRNA
48S 5' to 3' scanning

Initiation codon recognition
eIF2-GTP hydrolysis

Subunit joining
Factor displacement
eIF5B GTP hydrolysis
Model of the canonical pathway of eukaryotic translation initiation.

(Adapted from [43]) The pathway of translation initiation is shown as four stages, indicated by red text. These stages ultimately result in formation of an 80S ribosomal initiation complex. This complex has Met-tRNA\textsubscript{Met} base paired with the AUG start codon in the P-site and is competent to being translation elongation. The stages are: mRNA activation, during which the mRNA cap (m7G) is bound by eIF4-E, eIF4-G, eIF4-A, and eIF4-B; 43S preinitiation complex attachment to the mRNA, consolidation into the 48S complex, and 48S of scanning of the 5’ UTR in a 5’ to 3’ direction; start codon recognition switches the scanning complex to a ‘closed’ conformation and promotes eIF2 GTP hydrolysis and inorganic phosphate release; joining of the 60S subunit to the ‘closed’ initiation complex and concomitant displacement of eIF2(GDP) and other initiation factors (eIF1, eIF3, eIF4B, and eIF4F) is mediated by eIF5B GTP hydrolysis. Release of eIF5B(GDP) results in assembled 80S ribosomes with mRNA competent for elongation.
Figure 1-5

Canonical translation

Non-canonical initiation

A) Internal ribosome entry

B) Ribosome shunting

C) Leaky scanning

D) Reinitiation

E) Non-AUG initiation

Non-canonical elongation

F) Frameshifting

Non-canonical termination

G) Readthrough

H) Stop-carry on
Figure 1-5 Examples of non-canonical translation
(adapted from [5]) Translation is initiated (grey arrow) on an AUG codon within an mRNA (black) that bears a 5’ 7-methylguanosine cap (m7G) (brown) and 3’ poly(A) tail (Aₙ). Elongation continues through the open reading frame (ORF) (boxed) until translation is terminated at a stop codon (red stop sign). Non-canonical initiation can occur on (A) uncapped mRNAs by way of an internal ribosome entry site (IRES). IRESes directly recruit the ribosome to positions within mRNA for translation initiation. Non-canonical initiation on m7G capped RNAs can occur through (B) ribosome shunting: ribosome scanning-independent bypass (green) of a complex leader sequence; (C) leaky scanning: a subset of ribosomes will bypass the first start site and initiate translation on a downstream AUG codon (large grey arrow); (D) reinitiation, the 40S subunit remains attached after translating a very short ORF (blue), resumes scanning (green), and reinitiates translation (small grey arrow) on a downstream start codon; (E) non-AUG initiation: a fraction of ribosomes initiate translation on a non-AUG codons (small grey arrow) with strong primary sequence context. Non-canonical elongation occurs by way of (F) programmed ribosomal frameshifting. During elongation a fraction of ribosomes will pause on a frameshift site, slip in a defined direction, and continue translation (green arrow) in the alternative reading frame until a stop codon (red stop sign) is encountered. Non-canonical termination occurs through (G) stop-codon readthrough: ribosomal bypass of a stop codon (small grey arrow), and (H) stop-carry on: a premature release of the nascent polypeptide with subsequent continuation of elongation in the original reading frame (small grey arrow).
1.2.2 Cap-independent non-canonical initiation

As discussed above, multiple means of non-canonical initiation can be advantageous for an organism in times of viral infection [6]. Many viruses actually capitalize on the cellular stress-response by including similar RNA elements in their 5’ UTRs, facilitating selective translation of viral transcripts. Additionally, even if a virus can abate the programmed cellular response to infection, many viruses forgo the machinery to add a 5’ cap or poly-A tail to their transcripts [5]. Some viruses, such as human rhinoviruses, inhibit cap-dependent translation by viral protein mediated cleavage of specific initiation factors [46, 47]. Other viruses have evolved to include internal ribosome entry sites (IRESes) for ribosome recruitment and initiation [48, 49].

There is no singular IRES sequence or structure [50]. IRESes are defined as RNA units that can recruit ribosomes to positions within mRNA for translation initiation, thus facilitating cap-independent translation (Figure 1-5A) [51]. Some viruses have IRESes in the 3’ region of their 5’ UTRs. Initiation without scanning from the very 5’ end of the mRNA allows protein binding of other 5’ UTR elements, like packaging and replication signals, to be unperturbed during translation [5, 52]. The dependence of viral IRES structures on ribosomal initiation factors and IRES trans-activating factors for initiation varies, according to sequence [49]. Data suggest that the more stably folded an IRES, the fewer cellular factors are required for non-canonical initiation. For example, the IRES from the intergenic region of dicistroviruses folds into a highly compact tertiary structure [49, 53-55], enabling direct recruitment of the 40S ribosome, formation of an 80S complex, and initiation, in the absence of trans-acting factors [53, 55]. In contrast, the IRESes of most picornaviruses fold into extended and largely flexible structures and thus require numerous initiation factors (eIF4G, eIF4A, rIF3, eiF5B) and additional trans-acting proteins not typically involved in initiation for 40S binding [49], such as LA autoantigen [56] and poly(rC) binding protein 2 [57].
Although it has been suggested that as many as 10-15% of cellular genes may contain IRESes [6, 58], many of these genes were identified in experiments biased for IRES-mediated translation [51]. Cellular IRESes are found in several mRNAs encoding proteins needed for cellular stress management or for entering into programmed cell death [6, 58], but careful controls are needed before many of the these putative IRESes are functionally verified [51]. Interestingly, what we do know about viral IRESes structure and function has not been directly applicable to cellular IRESes [49]. For example, stable mRNA secondary structures in many yeast IRESes inhibit efficient initiation [59]. One explanation for the negative correlation between IRES activity and structural stability is that the IRES activity may depend on the cellular levels and localization of trans-activating IRES factors, which can change in response to different cellular conditions [6, 49]. In times of cap-dependent translation inhibition, these factors may decrease the IRES structural stability, which would allow efficient initiation and translation of these messages. Regulation of IRES mediated initiation limits translation of such mRNAs to defined cellular environments.

1.2.1 Cap-dependent non-canonical initiation

Retrovirus genomes are integrated in host DNA, and as a result they must rely on that host for mRNA synthesis (transcription) and processing (5’ capping, splicing, and 3’ polyadenylation). While translation of such mRNAs is considered to be cap-dependent, RNA elements in the 5’ UTR often inhibit efficient ribosome scanning and initiation [52]. For example, the HIV-1 5’ UTR contains the primer binding site, packaging signal, genome dimerization signal, and Tat trans-activation response element [60]. Because many of these elements are upstream of the 5’ splice site, they are present in the 5’ UTRs of approximately 30 alternatively spliced mRNA forms [61, 62]. The inhibitory elements located in the HIV-1 5’ UTR are found in many retroviruses (reviewed in [52]). As such, retroviruses encode cis-acting RNA
elements in their 5' UTRs that overcome inherently inefficient scanning and initiation from the 5' cap.

Highly structured RNAs promote ribosome shunting (Figure 1-5B) [5]. Shunting allows ribosomes to access downstream open reading frames (ORFs) by scanning-independent bypass of a complex leader sequence [5, 63]. In the genomes encoded by the Caulimoviridae family the long ORF encoding most viral proteins is inaccessible without prior translation of a short upstream ORF followed by a stable RNA structure [64]. The ribosome is able to bypass the structure and resume mRNA scanning downstream in a process called shunting. The stop codon in the upstream ORF and proximal RNA structure act in cis to promote discontinuous scanning. Ribosomal shunting in other cases is initiated by ribosomal pausing near the leader sequence independently of translation of the upstream ORF [63].

Non-canonical initiation can increase mRNA coding capacity through the selection of alternative start sites. The efficiency of canonical initiation at an AUG codon is modulated by its sequence context [65, 66]. Optimal sequence context in mammals is defined by a 5' GCCRCCCAUGG 3' sequence, where R is a purine. Changes in this sequence context will reduce the initiation efficiency and ribosomes will bypass the first start site to initiate translation at downstream AUG codon with optimal sequence context [5]. This process is termed leaky scanning (Figure 1-5C) and is utilized by many RNA viruses for translational control of gene expression. For example, the HIV-1 vpu AUG has a weak initiation sequence context, allowing leaky scanning and initiation on the downstream env AUG initiation codon [52]. Here, Env translation is dependent upon decreased selection of the upstream AUG start site in the vpu ORF.

Leaky scanning is promoted through two other mechanisms. First, when the primary AUG is close to the 5' end of the transcript, within 30 nucleotides, it is often recognized inefficiently
[67], allowing a fraction of ribosomes to continue scanning and initiate at a downstream start site. Second, leaky scanning is increased by occasional ribosome recognition of a downstream AUG within close proximity to the primary during the process of mRNA scanning [68]. Data suggest that initiation on a downstream AUG results from forward and backwards thrusts by ribosomes in the process of scanning [5].

While infrequent, ribosomes can also be directed to reinitiate (Figure 1-5D) on a downstream AUG after translation termination, by cis-acting RNA elements [5]. Following canonical termination, the 40S and 60S subunits dissociate from the mRNA. However, the 40S subunit can occasionally remain attached if it has translated a very short ORF (e.g. less than 30 codons), then resume scanning, and reinitiate translation on a secondary downstream start codon (reviewed in [69]). In this instance of translational reprogramming, the short ORF acts in cis to facilitate non-canonical termination and cap-independent reinitiation. Reinitiation after translation of an ORF encoding a functional product is very rare, and requires additional cis-acting RNA elements or trans-acting protein factors [5, 70]. For caliciviruses, translation of a subset of viral structural proteins is dependent upon reinitiation [70-72]. Here, 40S ribosome retention after termination on a functional ORF is enabled by a cis-acting mRNA sequence 40-90 nucleotides upstream of the reinitiation site [73, 74].

Some viruses use non-AUG initiation (Figure 1-5E) as part of a leaky scanning mechanism to translate multiple proteins from alternative reading frames or translate multiple N-terminal isoforms of the same protein (reviewed in [5]). Similar to leaky scanning, non-AUG initiation is modulated by the primary sequence surrounding the codon, which is enhanced by optimal 5' and 3' initiation sequence context (5' GCCRCCNNNG 3' [65, 66], with the non-AUG codon shown as NNN) and AUG codon similarity. Non-AUG initiation can be highly efficient if it is specifically directed by IRES elements. For example, the dicistrovirus intergenic IRES only
initiates translation on a non-AUG codon. This is achieved by mRNA structural mimicry of the P-site codon-anticodon complex [53, 75].

Although the most efficient protein synthesis in mammalian mRNAs is usually initiated at AUG codons, initiation on GUG and CUG codons is common [76]. Interestingly, translation of a significant number of cellular mRNAs that initiate with non-AUG codons, drives production of truncated or extended protein isoforms [76-78]. In a well characterized example, translation initiation of the human fibroblast growth factor (FGF) mRNA occurs at any of four different CUG codons or a downstream in-frame AUG. In addition to being dependent on sequence context and AUG similarity, the non-AUG initiations are modulated by five cis-acting RNA structures within the FGF 5’ UTR and the alternatively translated region [79]. The differential localization of these various N-terminal protein isoforms within a cell is important to FGF function (reviewed in [78]).
1.3 Canonical vs. non-canonical elongation

During elongation (Figure 1-2), the ribosome decodes consecutive mRNA codons in a series of translocation-pause-translocation events [4]. During the pause phase, aa-tRNA accommodation and the peptidyl transfer reaction occur. After pausing, translocation advances the ribosome on the mRNA by a single codon step. Each phase is briefly discussed below.

A typical elongation cycle is initiated by binding into the ribosomal A-site of an aa-tRNA-eEF1A(GTP) complex (Figure 1-2 step 1) (reviewed in [25]). Accurate base pairing between the mRNA codon and aa-tRNA anticodon triggers eEF1A(GTP) hydrolysis [25, 80] (Figure 1-2 step 2). This is coupled to a conformational change in the ribosome that results in eE1FA(GDP) dissociation and aa-tRNA reorientation [25, 80], such that the aminoacyl end is within the peptidyl transfer center [81] (Figure 1-2 step 3). Next, the ribosome catalyzes peptide bond formation between the α-amino group of the A-site tRNA and the ester carbon of the P-site tRNA-bound polypeptide chain (Figure 1-2 step 4). Following catalysis, the acceptor ends of the P and A-site tRNAs then have a greater affinity for the E- and P-sites and will alternate between their original states, or “classic states”, and a P/E and A/P “hybrid states” conformation [34, 82].

Ribosome translocation is triggered by binding of EF-2(GTP) near the A-site (Figure 1-2 step 5). This event creates a sizeable 40S subunit movement, or ratcheting, relative to the large subunit, triggering EF-2 mediated GTP hydrolysis [34, 83, 84] (Figure 1-2 step 6). In this state EF-2(GDP) docks into the A-site decoding center, forcing the translocation of the A/P and P/E site tRNA:mRNA complexes into the P- and E-sites [34]. EF-2(GDP) dissociation (Figure 1-2 step 7) followed by deacylated tRNA exit [20] (Figure 1-2 step 8) resets the ribosome for further rounds of elongation.
Failure of the ribosome to maintain the reading frame can result in incorrect protein synthesis and premature termination (discussed in section 1.1). However, programmed changes in reading frame can result in the translation of functional proteins, if additional genes are encoded in alternative reading frames generated by a +1, +2, -1, or -2 nucleotide ribosomal shift (Figure 1-5F) [85]. Since viruses have selective pressure to maximize their genomic coding capacity, they frequently utilize this type of frameshifting to express multiple proteins from alternative reading frames. Translational reprogramming through frameshifting provides a way to produce multiple protein products from a single translation start site. This eliminates the need for redundant genomic material within the virus particle and can dictate the molar ratio of proteins produced from the original and alternate reading frames.

Programmed changes in reading frame may be stimulated by cis-acting RNA elements and occur during ribosomal elongation (Figure 1-5F). Such events are facilitated by ribosomal pausing on a “slippery” sequence when it is positioned in the P- and A-sites of the ribosome. Next the ribosome slips by a specific number of nucleotides in a defined direction. The tRNAs translocate into the new reading frame and the ribosome continues translation. Programmed ribosomal frameshift (PRF) events have multiple mechanisms: a 1 nucleotide, or 2 nucleotide shift in the 5’ direction (-1 or -2 PRF), or 3’ direction (+1 or +2 PRF) [86]. -2 and +2 PRF will not be discussed here because they have only been documented in a few limited cases [85, 87-89].

1.3.1 +1 PRF

+1 PRF sites were identified in Escherichia coli and multiple eukaryotic mRNAs (reviewed in [8, 10, 90]). Although the components and mechanism of +1 translational reprogramming seem to be case specific, it appears always to be driven by cis-acting RNA elements. In most instances of +1 PRF, P-site tRNA slippage on a heptanucleotide slippery sequence is required for a productive frameshift [10]. Definition of a canonical +1 PRF site is confounded by several
factors: 1) most slippery sequences, but not all, have a rare codon positioned in the A-site [91-95], 2) 5’ and 3’ RNA elements can contribute to frameshifting [94, 96], but are not always present, and 3) productive slippery sequences can preclude P-site tRNA repairing in the +1 frame [97]. The outcome of +1 PRF in the E. coli prfB [92], eukaryotic ornithine decarbozylase antizyme [93, 94], and the Saccharomyces cerevisiae retrotransposable Ty1 [91] genes is synthesis of functional proteins.

1.3.2 -1 PRF

The typical -1 frameshift site contains three RNA elements (Figure 1-6A): a heptanucleotide slippery sequence, a single-stranded spacer (5-8 nucleotides), and a downstream RNA structure, typically a pseudoknot or stem-loop. The cis-acting RNA elements involved in this translational reprogramming event are coupled to yield highly specific frameshift efficiencies [98-101], which predetermine the ratio of 0 frame to -1 frame proteins produced. -1 PRF efficiencies vary widely and can range from values as low as 5% (HIV-1 [99]) to values as high as 50% (E. coli dnaX [102]).

The slippery sequence has the consensus X XXY YYZ (shown in the 0 reading frame), where X can be any nucleotide type, Y can be A or U, and Z is not G in eukaryotes [103, 104]. The P- and A-site tRNAs in the 0-frame (XXY YYZ) slip one nucleotide in the 5’ direction and repair in the -1 frame (XXX YYY). Repairing in this alternative reading frame is energetically allowed because the third (wobble) position in the anticodon can tolerate multiple pairings. While the spacer region has been suggested to be anywhere from 5 to 8 nucleotides in length, in the context of a ribosome positioned to frameshift, the spacer length should be ≥7 nucleotides (discussed in Chapter 2). Structural stimulators of frameshifting can be pseudoknots (Figure 1-6A) (reviewed in [101]), stem-loops (Figure 1-6A) [102, 105-109], or double stranded RNA junctions [110, 111].
-1 PRF sites are found in many viruses, including coronaviruses [107, 112, 113], astroviruses [114, 115], giardiviruses [116, 117], plant viruses [118, 119], and many retroviruses [52, 99, 120-125]. These translational reprogramming sites are utilized for the translation of C-terminally extended polyproteins. In most retroviruses a -1 PRF site lies between the gag and pol ORFs (Figure 1-6B), with pol in the -1 reading frame relative to gag. Some retroviruses, such as human T-cell leukemia virus (HTLV) and Mason-Pfizer monkey virus, separate pol into two different reading frames: pro and pol, and utilize a downstream -1 PRF site within pro to translate pol [124, 126]. The gag ORF encodes the viral structural proteins, while the pro and pol ORFs encode the enzymatic proteins (Pro- protease and Pol- reverse transcriptase and integrase). During translation of viral mRNA, the majority of ribosomes terminate at a stop codon at the end of the gag, producing the Gag polyprotein (Figure 1-6B & C). However, a subset of ribosomes will pause on the slippery site and shift into the -1 reading and continue translation to produce Gag-Pol (Figure 1-6B) or Gag-Pro (Figure 1-6C) polyproteins [98, 99, 121, 127, 128]. The full-length Gag-Pro polyprotein is only synthesized in the absence of frameshifting on the downstream -1 PRF site (Figure 1-6C). A subset of ribosomes translating pro will pause, frameshift, and continue translation in the -2 reading frame relative to gag, producing Gag-Pro-Pol polyproteins (Figure 1-6C). The frameshift efficiencies at each -1 PRF site predetermine the ratio of enzymatic to structural proteins produced. Maintaining strict molar ratios of viral proteins is essential for retroviral replication and infectivity [103, 127, 129-136]. For example, a three-fold decrease in -1 frameshift efficiency inhibits viral replication in Rous sarcoma virus [132].

Numerous -1 PRF sites in archae, prokaryotic, and eukaryotic genomes have recently been identified (reviewed in [8, 10]). While the cis-acting RNA components directing the programmed event are the same as viral examples, the functional outcome of frameshifting is quite different. Experimentally verified productive -1 frameshift sites in cellular mRNAs encode premature termination codons (PTCs) in their -1 reading frames. PTCs are
differentiated from termination codons by their mRNA context (reviewed in [137]). It is suggested that when a translating ribosome encounters a PTC in its A-site, translation is arrested and terminated, and the mRNA is degraded through nonsense mediated mRNA decay (NMD) pathway. Several studies have shown that inclusion of a -1 PRF signal in the coding region of an mRNA greatly decreases the mRNA half-life in an NMD dependent fashion [138, 139]. Additionally, sufficiently long ribosomal pausing on the frameshift site can lead to ribosome disengagement and mRNA decay via the no-go decay pathway [138, 140]. Therefore, inclusion of -1 PRF sites in cellular mRNAs could represent an efficient manner of post-transcriptional regulation, whereby mRNA abundance levels are indirectly controlled by the cis-acting RNA elements within the frameshift site. If the frameshift efficiencies in any of these sites are modulated by trans-acting factors, mRNA levels could be easily altered in different cellular conditions.
Figure 1-6

A) Diagram of a stem-loop and pseudoknot structure with a slippery sequence.

B) Diagram showing reading frames and translation products for gag and pol genes.

<table>
<thead>
<tr>
<th>Reading frame</th>
<th>Translation products</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Gag</td>
</tr>
<tr>
<td>0, -1</td>
<td>Gag-Pol</td>
</tr>
</tbody>
</table>

C) Diagram showing reading frames and translation products for gag, pro, and pol genes.

<table>
<thead>
<tr>
<th>Reading frame</th>
<th>Translation products</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Gag</td>
</tr>
<tr>
<td>0, -1</td>
<td>Gag-Pro</td>
</tr>
<tr>
<td>0, -1, -2</td>
<td>Gag-Pro-Pol</td>
</tr>
</tbody>
</table>
Figure 1-6 Canonical -1 PRF

(A) A -1 PRF site is defined by three elements: a heptanucleotide slippery sequence, a single stranded spacer, and a downstream structure. The downstream structure is typically an RNA stem-loop or pseudoknot. (B) In most retroviruses, the frameshift site is located in the 3’ end of the gag ORF, which codes for viral structural proteins. The pol ORF is in the -1 reading frame relative to gag and is accessed via a -1 PRF event. The location of the frameshift site in gag is indicated by a star. If the ribosome does not change reading frames during translation of gag it will produce the Gag polyprotein; a fusion protein including all of the viral structural proteins. Occasionally, the ribosome will shift one nucleotide in the 5’ direction when paused on the frameshift site, and continue translation in the -1 reading frame, thus producing the Gag-Pol polyprotein. This fusion protein includes the viral structural and enzymatic proteins. (C) In some retroviruses the ORF coding for enzymatic proteins is split into two, one coding for the viral protease and the other coding for the viral reverse transcriptase and integrase proteins. The pol and pro ORFs are both accessed through independent -1 PRF events. The frameshift sites in gag and pro are indicated by stars.
1.3.3 -1 PRF mechanism

Multiple models have been proposed to explain the -1 frameshift mechanism [34, 86, 99, 113, 128, 141-144, 145]. The following steps are fundamental to each: 1) during elongation, the ribosome pauses when the slippery sequence (XXY YYZ in the 0 frame) is positioned in the ribosomal P- and A-sites [104, 146-148]. 2) While paused, the ribosome slips one nucleotide in the 5’ direction at a frequency specific to the frameshift site and continues elongation in the -1 reading frame. Each model is distinguished by the point in translation when the -1 frameshift occurs.

Two groups have suggested -1 frameshifting occurs during aa-tRNA binding to the A-site before E-site tRNA exit (Figure 1-7 Model 1) [128, 143, 149]. This model is consistent with observations that mutations within rRNA near the E-site and the codon 5’ of the slippery sequence in the HIV-1 affect its frameshift efficiency [143]. However, a recent study observed that deacylated tRNA exit is only coupled to aa-tRNA binding in very early stages of elongation [20]. Thus, the E-site tRNA likely exits the ribosome spontaneously, before aa-tRNA binding. If this is the case, the deacylated tRNA upstream of the slippery sequence should not influence frameshifting. The identity of the E-site codon could still indirectly influence frameshifting by changing the rate of translation (reviewed in [150]).

Alternatively, the -1 frameshift could occur after tRNA accommodation, but before peptidyl transfer (Figure 1-7 Model 2). In the 9 Å model [141], aa-tRNA accommodation coupled mRNA movement is strained by blockage of the mRNA entry channel by the downstream RNA structure. The structure’s resistance to unwinding could generate tension that is sensed in the codon-anticodon interaction. This tension could be relieved by the coincident slippage of the P- and A-site tRNAs by 1 nucleotide in the 5’ direction. This model is consistent with the simultaneous slippage model proposed for the HIV-1 -1 PRF site by Jacks and Varmus [99]. The
and simultaneous slippage [99] models are generally consistent with the amino acid sequence of the HIV-1 Gag-Pol polyprotein, in which the slippery sequence is decoded in the 0 (Phe-Leu) and -1 (Phe-Phe) reading frames with 70% and 30% frequency, respectively [99].

Namy et al. suggested that frameshifting occurred during EF-2 catalyzed translocation (Figure 1-7 Model 3) [142]. This model is based on a cryo-EM structure of an 80S ribosome stalled by a downstream -1 PRF pseudoknot. In this structure the P-site tRNA was unusually compressed and EF-2 occupied the A-site. It was hypothesized that when the ribosome is paused by the pseudoknot, EF-2 catalyzed translocation is stalled after the A-site tRNA is moved into the P-site. Stalling creates tension in the mRNA, which causes the P-site tRNA to bend like spring. This unnatural level of tRNA compression [34, 37, 142] could promote breakage of hydrogen bonds in the anticodon-codon complex [34], allowing the P-site tRNA anticodon to repair in the -1 frame.

Recent crystallographic data suggest that P-site tRNA deformation occurs normally during translocation after aa-tRNA accommodation, peptidyl transfer, and EF-2 GTP hydrolysis [34]. If the model proposed by Namy et al. [142] was updated to include the A-site tRNA, it would be consistent with a model proposed by Weiss et al. and Harger et al. (Figure 1-7 Model 4) [151, 152]. In this model, frameshifting occurs during translocation when the P- and A-site tRNAs are in hybrid conformations. The large changes in tRNA conformation [153] associated with its hybrid state may weaken the codon-anticodon interactions, facilitating frameshifting in this position. Alternatively, the -1 frameshift event could result from an incomplete translocation [145]. However, data from several studies suggests that post-peptidyl transfer ribosomes do not slip (reviewed in [141]).

Finally, the ‘many pathways model’ of -1 PRF does not require a distinct point in translation for the frameshift event to occur [144]. Instead, it proposes that frameshift efficiency
is the sum of frameshift events occurring along different points in elongation (Figure 1-7). This model reconciles the differences in slippery sequence decoding observed during HIV-1 -1 frameshifting [99].
Figure 1-7 Four models of the state of the ribosome just prior to a -1 frameshift.

Component coloring is consistent with Figure 1-2. Here, the HIV-1 frameshift site is used as a model to demonstrate when a -1 PRF could occur during translation. In model 1, frameshifting occurs during aa-tRNA accommodation before peptidyl transfer and E-site tRNA exit. In model 2, the frameshift happens after tRNA accommodation, but before peptidyl transfer. In model 3, the frameshift event takes place during translocation when the ribosome is stalled and the P-site tRNA is unnaturally compressed. In model 4, frameshifting occurs after peptidyl transfer during translocation when the tRNAs are in the P/E and A/P conformations.
1.3.4 Modulators of -1 frameshift efficiency

Multiple factors may modulate -1 frameshift efficiency, such as slippery site sequence [98, 104, 126, 154], tension generated during translocation transmitted through the spacer [88], RNA stability [100, 102, 136, 155-160], conformational plasticity [161], specific structural interactions with the ribosome [101], trans-acting factors [162-164], and translation initiation rates [164, 165]. The composition of the slippery site sequence is the greatest modulator of frameshift efficiency, as differences in sequence can change -1 frameshift efficiency dramatically [103, 104, 126, 154]. Tension generated during translocation is fundamental to several of the proposed -1 PRF models (Figure 1-7), and is likely influenced by the downstream structure’s resistance to unwinding and its distance from the A-site codon. It was recently hypothesized that shorter spacers can increase the tension in mRNA [88]. However, the validity of this hypothesis depends on the plasticity in the distance between the ribosomal A-site and mRNA entry channel entrance. Data presented in Chapter 2 suggests that the channel length is maintained and changes in spacer length simply extend the ribosomal footprint into the RNA structure. Therefore, a constant spacer length is maintained when the ribosome is positioned on the slippery sequence [100]. Changes in tension may still impact frameshifting, but these changes appear to result from changes in the downstream structural stability rather than changes in spacer length.

RNA structural stimulation of frameshifting may depend on multiple structural characteristics. General trends between stem-loop overall stability and frameshift efficiency have been observed [102, 136, 155, 156]. However, it appears that local structural stability is a primary determinant of stem-loop stimulated frameshifting [100] and may also modulate pseudoknot -1 PRF systems [159]. Alternatively, mechanical stability may control structurally stimulated frameshifting [157-160]. However a recent study contradicts this relationship and
instead observed a correlation between frameshifting and the ability to form alternative structures [161].

Site-specific interactions between the RNA structure and ribosome could prime the ribosome for frameshifting [101, 142, 166, 167]. Binding of the coronavirus infectious bronchitis virus -1 PRF pseudoknot to the surface of the 80S ribosome was directly visualized by cryo-EM [142]. Interactions between the HIV-1 frameshift site stem-loop and a prokaryotic ribosome stalled on the frameshift site have also been detected using toe-printing experiments [167]. Additionally, trans-acting factors could impact frameshifting indirectly through interactions with the ribosome [162-164]. A study by Cassan et al. argues against existence of a specific viral or cellular trans-acting modulator of -1 frameshift efficiency in HIV-1 [168].

The rate of translation initiation controls the loading rate of ribosomes and affects -1 PRF in HIV-1 [164]. Increased initiation rates leads to increased poly-ribosome (aka polysome) density and decreased spacing between individual ribosomes. If two ribosomes stack end to end at the frameshift site, the first ribosome would encounter the folded RNA structure, but the trailing ribosome would not because the ribosomal footprint (approximately ~28 nucleotides [169]) will preclude folding of the structure in between ribosome encounters. As a result, the measured frameshift efficiency should decrease with increased ribosome stacking. Studies examining the secondary structure of the HIV-1 genomic RNA within capsids predict that the frameshift site is part of a conserved three-helix junction secondary structure [170, 171]. As described in Chapter 2, this additional secondary structure indirectly decreases the in vitro HIV-1 frameshift efficiency [100]. Formation of additional secondary structures within long mRNAs is probable and likely impacts ribosome density and frameshifting in vivo in all translation contexts.
1.4 Canonical vs. non-canonical termination

Ribosomes continue translation along an mRNA until a stop codon (UAA, UAG, or UGA) is positioned in the A-site. Stop codons are recognized directly [172] by release factors eRF1 and by eRF3 (in eukaryotes), and by RF1 and RF2 (in bacteria) that resemble aa-tRNA-eEF1A-GTP complexes and signal for translation termination [173]. In eukaryotes, eRF1 and eRF3 act in concert [174, 175]. Direct contacts between eRF3 and the stop codon promote a small subunit conformational change, which reorients eRF1 so it can mediate peptide release. A universally conserved glutamate in eRF1 facilitates hydrolysis by positioning a water molecule for nucleophilic attack of the ester linkage between the nascent polypeptide and peptidyl-tRNA [175-178]. The hydrolysis reaction releases the nascent polypeptide from the ribosome. Termination is completed when the mRNA, release factors, and deacetylated tRNAs leave the ribosome, which then dissociates into 40S and 60S components with assistance from the ABCE1 ribosome-recycling factor [179] and EF-G [180, 181].

1.4.1 Readthrough

Although typically translation termination is a highly accurate process, the ribosome can bypass stop codons through a process called “readthrough” (Figure 1-5G). During readthrough, ribosomes decode the stop codon using a suppressor or near-cognate tRNA and continue translation (reviewed in [5, 182]). The frequency of readthrough is influenced by the sequence and structure downstream of the stop codon, making some sequences especially ‘leaky’. Interestingly, in viruses that exhibit readthrough, a downstream RNA structure is commonly positioned eight nucleotides away from the stop-codon [5], mimicking the position of RNA structures blocking the mRNA entry channel during -1 PRF [100]. While the exact mechanism of readthrough is unclear, the position of the downstream structure suggests that the mechanism may be similar to structure stimulated -1 PRF.
1.4.2 Stop-carry on

Stop-carry on is defined by premature release of the nascent polypeptide and continuation of elongation in the original reading frame (Figure 1-5H) [5, 183]. This atypical termination is dependent on the amino acid sequence, or “2A” motif, in the C-terminus of the nascent polypeptide (D(V/I)EXNPGP) and is not dependent on stop-codon recognition. Additionally, there is usually a string of rare codons downstream of the terminal proline codon that slows the passing ribosome. Therefore, a single ORF can be utilized to synthesize two distinct polypeptides if the mRNA sequence codes for this motif. A genera of picornaviruses, including Cardioviruses and aphthoviruses, utilize the 2A motif to promote co-translational processing of their polyproteins [5, 184]. The 2A peptide interacts with the ribosome exit tunnel and peptidyl transfer center (Figure 1-5A) to prevent formation of a peptide bond between glycine and the terminal proline [185]. eRF1 and eRF3 interactions mediate the release of the primary polypeptide despite the absence of a stop codon. Elongation continues with nearly 100% efficiency, and the rest of the ORF is translated as an independent polypeptide with an N-terminal proline [5].
1.5 -1 PRF in HIV-1

HIV-1 is a complex retrovirus that is the causative agent of acquired immunodeficiency syndrome (AIDS). In addition to encoding the standard structural, enzymatic, and envelope genes (e.g. gag, pol, and env) found in simple retroviruses, HIV-1 encodes species-specific accessory proteins. The 9.4 kb genome has 9 ORFS that are translated using a primary transcript and several alternatively-spliced transcripts [61, 62], to produce a total of 15 unique proteins (Figure 1-8A). A -1 PRF site is located in the 3′ end of gag and is required for Pol protein synthesis, which in turn, is decoded in the -1 reading frame relative to gag (Figure 1-8B).

The HIV-1 PRF site is composed of a heptanucleotide slippery sequence (U UUU UUA in the 0 frame) followed by a downstream RNA stem-loop (Figure 1-8A). The frameshift site UUU (phenylalanine) and UUA (leucine) codons were evolutionarily selected from two and six possible codon options, respectively. Only this codon sequence allows near-cognate and cognate re-pairing of the A and P site tRNA anticodons, respectively, in the -1 reading frame. The slippery sequence of HIV-1 is especially “slippery”, and, in the absence of the stem-loop, increases the basal level of ribosomal frameshifting from approximately 0.0001% per codon to 0.1% [98, 126]. Frameshifting is further stimulated to the levels required for viral replication by the RNA stem-loop [100, 108, 186-188].

The RNA structure within the PRF site of HIV-1 group M is a stem-loop 11 base pairs long, originally identified by Jacks et al. [99]. A combination of thermodynamic measurements, structure determinations [108, 187, 189], mutational and functional analyses [127, 128, 190], and sequence co-variation [190], indicates that the RNA folds into an extended stem-loop (Figure 1-8A). This structure is characterized by an extremely stable 11 base pair, ΔG = #, upper helix separated from a moderately stable 8 base-pair, ΔG = #, lower helix by a GGA bulge [108]. At the onset of my work, the role of the stem-loop in frameshift stimulation was unclear.
Alteration of HIV-1 frameshifting efficiency can affect viral replication and infectivity [103, 127, 131, 134-136]. Specifically, relatively small decreases in frameshift efficiency greatly inhibit viral replication [103, 127, 136]. Therefore, targeted manipulation of frameshift efficiency has the potential for development of novel antiretroviral therapies [103, 145, 191].
Figure 1-8 The HIV-1 genome and site of -1 PRF
This figure was modified from [192].  (A) Organization of the coding regions of the HIV-1 genome.  The genome encodes nine open reading frames (gag, pol, vif, vpu, rev, tat, env, and nef), which are translated into fifteen individual proteins.  There are also many cis-acting RNA regulatory elements.  Here, the gag and pol coding regions are indicated in blue and beige, respectively, with the location of the frameshift indicated by a star.  The frameshift site consists of a UUUUUA slippery sequence, immediately followed by an extended stem-loop composed of an eight base-pair lower helix (blue), an eleven base-pair upper helix (red), and a three purine GGA bulge (purple).  (B) During translation of HIV-1 mRNA, the majority of ribosomes terminate at the stop codon at the end of the gag ORF, producing the Gag polyprotein, a fusion of the structural proteins, matrix (MA), capsid (CA), nucleocapsid (NC), and p6.  However, approximately 5% of ribosomes pause at the frameshift site, indicated by a star, and shift into the -1 reading frame, thus producing the Gag-Pol polyprotein, which is consists of the MA, CA, NC, p6*, protease (PR), reverse transcriptase (RT), and integrase (IN) proteins, fused as a single polyprotein.
1.6 Dissertation overview

My work aimed to generate an improved understanding of the structure and function of the HIV-1 frameshift site RNA stem-loop. It was known that the stem-loop acted in cis with an upstream heptanucleotide slippery sequence to stimulate -1 PRF [100, 107]. Although this translational reprogramming event is critical to viral replication [103, 127, 131, 134-136], the frameshift mechanism was unclear.

Chapter 2 describes my experiments on the role of the stem-loop structure in frameshift modulation. The findings now provide an enhanced understanding of the importance of base-pairs within the stem-loop for frameshift stimulation. This study lead to the development of a quantitative model that predicts HIV-1 -1 frameshift efficiency with good statistical confidence. This model can be extended to other frameshifting systems to explain the actions of multiple cis-acting RNA structures in -1 PRF.

Chapter 3 describes how I used nuclear magnetic resonance (NMR) spectroscopy and small-angle X-ray scattering (SAXS) to study the structure of the HIV-1 frameshift site stem-loop. NMR methods presented in Chapter 3 were then expanded to investigate domain motions of the frameshift site stem-loop, as described in Chapter 4. Chapter 5 summarizes my thesis work and gives potential future experimental directions based on my novel results.

Three appendices are included. Appendix I describes a preliminary structure of an RNA pseudoknot important to IRES function in the Israeli acute paralysis virus. This work is ongoing and employs a combination of NMR and SAXS for structure determination. Appendix II describes the effect of two compounds, cobalt-hexamine and doxorubicin, on HIV-1 frameshift site stem-loop domain motions. Finally, Appendix III documents a preliminary NMR analysis of base-pair formation of the HIV-1 three-helix junction, which encompasses the frameshift site and ~90 nucleotides of its surrounding sequence.
Chapter 2: HIV-1 frameshift efficiency is primarily determined by the stability of base-pairs positioned at the mRNA entrance channel of the ribosome

The majority of work presented in this chapter has been published as:


A.L.L and P.D.E. collected UV thermal melt and in vitro frameshift data for a subset of the MS and 3HJ constructs, respectively.

K.A.V provided technical assistance for determination of ΔG.

Ari J. Salinger cloned the p2luc 3HJ M1-M3 plasmids, and contributed them to this study.

The remainder of the work was by K.D.M.
2.1 Abstract

The human immunodeficiency virus (HIV) requires a programmed -1 ribosomal frameshift for Pol gene expression. The HIV frameshift site consists of a heptanucleotide slippery sequence (UUUUUUA) followed by spacer region and a downstream RNA stem-loop structure. Here we investigate the role of the RNA structure in promoting the -1 frameshift. The stem-loop was systematically altered to decouple the contributions of local and overall thermodynamic stability towards frameshift efficiency. No correlation between overall stability and frameshift efficiency is observed. In contrast, there is a strong correlation between frameshift efficiency and the local thermodynamic stability of the first 3-4 base-pairs in the stem-loop, which are predicted to reside at the opening of the mRNA entrance channel when the ribosome is paused at the slippery site. Insertion or deletions in the spacer region appear to correspondingly change the identity of the base-pairs encountered 8 nucleotides downstream of the slippery site. Finally, the role of the surrounding genomic secondary structure was investigated and found to have a modest impact on frameshift efficiency, consistent with the hypothesis that the genomic secondary structure attenuates frameshifting by affecting the overall rate of translation.
2.2 Introduction

Translation is a high fidelity process in all organisms. Failure to maintain reading frame typically results in incorrect protein synthesis and/or early termination. However, a programmed change in reading frame can result in the translation of new proteins, thereby maximizing genomic coding capacity. Many retroviruses, including human immunodeficiency virus type 1 (HIV-1) [99], and some coronaviruses, such as severe acute respiratory syndrome [107] and infectious bronchitis virus (IBV) [112], use a programmed -1 ribosomal frameshift (-1 PRF) to control translation levels of their enzymatic proteins [86, 101, 148, 193]. In the retroviruses, the -1 PRF site lies between the gag and pol open reading frames (ORFs), with pol in the -1 reading frame relative to gag. The gag ORF encodes the viral structural proteins, while the pol ORF encodes the enzymatic proteins. During translation of HIV-1 mRNA, the majority of ribosomes terminate at a stop codon at the end of the gag ORF, producing the Gag polyprotein [107, 194]. However, the HIV -1 PRF induces approximately 5% of ribosomes to shift into the -1 reading frame, thus producing the Gag-Pol polyprotein [98, 99, 127, 128]. The 5% frameshift efficiency determines the ratio of viral proteins produced and is important for viral replication and infectivity [103, 127, 129-131, 134-136]. A decrease in frameshift efficiency can inhibit viral replication [103, 127, 134-136].

The HIV-1 frameshift site is composed of a heptanucleotide slippery sequence (UUUUUUA) followed by a downstream RNA stem-loop (Figure 2-1). The slippery sequence follows a general XXXYYYYZ consensus sequence, where X can be any nucleotide (nt) type, Y can be A or U, and Z is not G in eukaryotes [103, 104]. This sequence allows near-cognate and cognate re-pairing of the A and P site tRNA anticodons, respectively, in the -1 reading frame. HIV-1’s slippery sequence is especially “slippery,” and in the absence of a downstream structure increases the basal level of ribosomal frameshifting from approximately 0.0001% per codon to 0.1% [98, 126]. However, in order to further stimulate frameshifting to the levels required for
viral replication, the slippery site must be followed by a stable RNA structure [98, 99, 108, 146, 156, 167, 168, 186-190] (Figure 2-1A). Thus, frameshifting is achieved by the *cis* coupling of the slippery site and downstream structure [98, 99, 127, 128, 168].

Multiple models have been proposed to explain the frameshift mechanism [34, 86, 99, 113, 141-145]. Common amongst them are the following steps: 1) during translation, the ribosome pauses when the slippery sequence (UUU UUA in the 0 frame) is engaged in the ribosomal A- and P-sites [104, 146-148]. The pause is triggered by the downstream structure’s resistance to unwinding. 2) While paused, approximately 5% of ribosomes slip 1 nt in the 5’ direction and continue elongation in the -1 reading frame. The proposed models are differentiated by the exact step at which the frameshift occurs: during aminoacylated-tRNA accommodation [141], after accommodation, but before peptidyl transfer [99], after large subunit translocation [101], or after peptidyl transfer due to an incomplete translocation [145]. Alternatively, the ‘many pathways model’ of -1 PRF suggests that frameshift efficiency is the sum of frameshift events occurring, each of which could occur at these different points in elongation [144].

An important role of the downstream structure is to induce ribosomal pausing on the slippery sequence, which is necessary but not sufficient to promote efficient levels of frameshifting [195, 196]. Interestingly, the pause length does not appear to correlate with frameshift efficiency [195]. Interactions with the translational machinery have also been hypothesized to contribute to frameshifting [101, 142, 166, 167]. Previous studies have observed general trends between HIV-1 stem-loop thermodynamic stability and frameshift efficiency [136, 155, 156]. However, a quantitative correlation between thermodynamic stability and frameshift efficiency has not been described, and the role of individual base-pairs has not been systematically investigated.
For frameshift sites with a downstream pseudoknot structure, mechanical stability has been proposed to be a determinant of frameshift efficiency [157-160]. It has been hypothesized that mechanical tension lowers the energy barrier for frameshifting [101], where the amount of tension sensed by the ribosome is proportional to the mechanical stability of the translocation barrier [157-159, 197, 198]. However, a recent study found no correlation between pseudoknot mechanical stability and frameshift efficiency, but instead observed a correlation between frameshifting and the ability to form alternative structures [161].

Other factors can modulate the frameshift efficiency, such as translation initiation rates [164, 165]. Increased translation initiation rates lead to increased polysome density, which can cause ribosomes to stack at the frameshift site. This in turn affects the rate of mRNA refolding during translation and leads to a decrease in overall frameshift efficiency [145, 164]. Ribosome stacking can be promoted by RNA structure that precedes the frameshift site. Studies examining the secondary structure of the HIV-1 genomic RNA within capsids have revealed that the frameshift site is part of a conserved three-helix junction (3HJ) [170, 171]. It has been hypothesized that the role of this secondary structure is to decrease the rate of translation [170], which may affect frameshifting by facilitating pausing and inducing ribosome stacking.

Here, we investigate the role of the HIV-1 RNA structure in frameshifting, focusing on elucidating the relationships between frameshift efficiency and 1) the downstream RNA stem-loop thermodynamic stability, 2) spacer length, and 3) surrounding genomic secondary structure. By systematically altering the base-pair composition of the stem-loop, we dissect the contributions of global and local thermodynamic stability on frameshifting. These data reveal that the thermodynamic stability of the first 3-4 base-pairs in the stem-loop is a primary determinant of frameshift efficiency. Our data further indicate that the base-pairs important for frameshifting are located at a distance of 8 nt from the slippery site, which corresponds to the
length of the spacer and is consistent with a structural model of the ribosome paused at the frameshift site. Finally, we find that the conserved genomic RNA secondary structure serves to attenuate the frameshift efficiency, likely by affecting the overall rate of translation. Importantly, our study describes the first quantitative and predictive model for frameshift inducing stem-loops, which can be generally applied to many -1 PRF viral systems.
Figure 2-1

A) Two cis-acting RNA elements are separated by a single stranded spacer. B) Twelve mutant stem-loop (MS) RNA constructs were designed to discern the relative contributions of local and global RNA stability on HIV-1 frameshift efficiency. Sequence changes are indicated in bold and italicized. Predicted overall, ΔG\text{Global}, and local, ΔG\text{Local}, thermodynamic stabilities are in units of kcal/mol.

---

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>MS1</th>
<th>MS2</th>
<th>MS3</th>
<th>MS4</th>
<th>MS5</th>
<th>MS6</th>
<th>MS7</th>
<th>MS8</th>
<th>MS9</th>
<th>MS10</th>
<th>MS11</th>
<th>MS12</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>U</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>U</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>A</td>
<td>G</td>
<td>G</td>
<td>U</td>
<td>A</td>
<td>U</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>U</td>
</tr>
<tr>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>U</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>U</td>
<td>A</td>
<td>U</td>
<td>A</td>
<td>U</td>
<td>A</td>
<td>U</td>
<td>A</td>
<td>U</td>
<td>A</td>
<td>U</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>U</td>
<td>G</td>
<td>U</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>U</td>
</tr>
<tr>
<td>U</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>C</td>
<td>G</td>
<td>U</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>U</td>
</tr>
<tr>
<td>ΔG\text{Global}</td>
<td>-21.8</td>
<td>-24.1</td>
<td>-22.0</td>
<td>-19.1</td>
<td>-13.7</td>
<td>-18.6</td>
<td>-19.6</td>
<td>-17.3</td>
<td>-6.8</td>
<td>-24.5</td>
<td>-16.3</td>
<td>-20.8</td>
<td>-11.8</td>
</tr>
<tr>
<td>ΔG\text{Local} (3bp)</td>
<td>-7.5</td>
<td>-9.8</td>
<td>-9.8</td>
<td>-9.0</td>
<td>-9.0</td>
<td>-7.9</td>
<td>-7.5</td>
<td>-7.5</td>
<td>-6.4</td>
<td>-2.0</td>
<td>-2.0</td>
<td>-2.0</td>
<td>-2.0</td>
</tr>
</tbody>
</table>
2.3 Materials and methods

2.3.1 Plasmid construction

DNA templates used for the dual-luciferase frameshift assay were cloned into a p2luc vector between the *rluc* and *fluc* reporter genes. Briefly, complementary synthetic oligonucleotides (Integrated DNA Technologies (IDT), Inc.) with BamH I and Sac I compatible ends were cloned into the p2luc vector using the BamH I and Sac I sites between the *rluc* and *fluc* reporter genes. Oligonucleotides comprising the template sequences (Table 2-1) and their complements were phosphorylated, annealed, and ligated into the p2luc vector to produce the experimental constructs. This places the *fluc* gene in the -1 reading frame relative to *rluc*; analogous to the orientation of the *gag* and *pol* genes in the HIV-1 genome. For the spacer mutation constructs (MS13-17), a compensatory number of nucleotides were added or removed downstream of the frameshift site to maintain the appropriate reading frame of the downstream reporter gene. The “wild-type” (WT) sequence utilized here corresponds to the most frequently occurring sequence found in HIV-1 group M subtype B NL4-3 laboratory strain (56). Positive control sequences and their complements were also cloned into the p2luc vector and have two thymidine residues (Table 2-1 bold) in the slippery sequence (Table 2-1 underlined) replaced with cytidines, and an additional nt inserted immediately before the Sac I complementary sequence (GAGCT), which places the *rluc* and *fluc* genes in-frame. In all constructs, a Pml I restriction site was included at the end of the template to allow for run-off transcription after digestion with the Pml I enzyme (NEB). Resultant products were transformed into *E. coli* competent cells (DH5α). Plasmid DNA was purified from cell cultures (Qiagen) and the sequences of all constructs were verified (University of Wisconsin Biotechnology Center).
Table 2-1. Sequences inserted into the p2luc vectors for plasmid construction

<table>
<thead>
<tr>
<th>Construct</th>
<th>Sequence (5’- 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS1</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS2</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS3</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS4</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS5</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS6</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS7</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS8</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS9</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS10</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS11</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS12</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS13</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS14</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS15</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS16</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>MS17</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>3HJ WT</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>3HJ Mut</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>3HJ MS1</td>
<td>GATCCGGCAAGTGGAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCCTTCAGAGCCAGAGGCAAGCCGACCAGAGCT</td>
</tr>
<tr>
<td>3HJ M1</td>
<td>GATCCCCGACAAACCCTTGAGACAGGGCTGTTGGAAATGTGGAAAGGACACCAAATGAAAGATTGTA CTGAGAGACAGGCTAAATTTTTAGGGAAGATCTGCTGCCCTCCCAAAGGGAAGG CCAGGGAATTTCAGAGCGAGACAGACCAGAGCCACAGCCCCACCCAGAGCT</td>
</tr>
<tr>
<td>3HJ M2</td>
<td>GATCCGGGCTGTTGGAAATGTGGAAAGGAGTTCTTCTTCACCCAAATGAAAGATTGTA CTGAGAGACAGGCTAAATTTTTAGGGAAGATCTGCTGCCCTCCCAAAGGGAAGGCC AGGGAATTTCAGAGCGAGACAGACCAGAGCCACAGCCCCACCCAGAGCT</td>
</tr>
<tr>
<td>3HJ M3</td>
<td>GATCCGGGCTGTTGGAAATGTGGAAAGGAGTTCTTCTTCAGAGCGAGACAGACCAGAGCCACAGCCCCACCCAGAGCT</td>
</tr>
</tbody>
</table>

φ The wild-type HIV-1 sequence represents a naturally occurring HIV-1 M subtype B sequence ([http://www.hiv.lanl.gov](http://www.hiv.lanl.gov), accession number AB078005).
2.3.2 RNA synthesis and purification

Microgram quantities of RNA for the frameshift assay were transcribed in vitro using linearized p2luc plasmid DNA, purified His6-tagged T7 RNA polymerase (10x), 11.25 mM NTPs, and 2 units of RNasin Plus RNase Inhibitor (Promega), in 200 µL for 90 min at 37 °C. Pyrophosphate was pelleted by centrifugation (10 min, 13,200 rpm, room temperature), and RNA was phenol/chloroform extracted. Unincorporated NTPs and salt were separated from the RNA using size-exclusion chromatography (2 Econo-Pac P6 cartridges (Bio-Rad) in series). Monomeric RNA folding was achieved by denaturation at 95 °C for 5 min followed by incubation on ice for 30 min. RNAs were lyophilized to dryness and resuspended in water to a concentration of 1 µg/µL and stored in aliquots (~25 µL) at -80 °C. RNA integrity and purity were checked with 1% agarose gel electrophoresis. Finally, RNAs used for UV spectroscopy were purchased from IDT.

2.3.3 Frameshift assay

In vitro frameshift assays were completed with each RNA reporter (experimental and positive control) using a Rabbit Reticulocyte Lysate (RRL) System (Promega, nuclease treated, L416A). Differences from our previously described protocol [199] include the following: translation reactions contained 1.25 µg RNA, 10 units of RNasin Plus RNase Inhibitor (Promega, N2615), and 8.75 µL of RRL in 12.5 µL. Following a 90 min incubation at 37 °C, reactions were quenched with the addition of 0.5 µL 0.156 M EDTA pH 8.0 (6 mM final), as described previously [190]. For each reporter, a minimum of three independent frameshift assays were completed. Each independent assay included six replicate reactions.

Luminescence was measured using the Dual-luciferase Reporter Assay (Promega) as previously described [109]. Readings were taken with a Veritas microplate luminometer equipped with dual-injectors (Turner Bio-systems) for 10 seconds after 25 µL of the respective
substrate was injected into the reaction mixture (2 second lag time prior to measurement). Ratios of firefly/Renilla luminescence were calculated for each of the experimental and control translation reactions. The frameshift efficiency was calculated by taking the ratio of the experimental/control luminescence (firefly/Renilla). Frameshift efficiencies were averaged and their standard deviations were propagated through to yield a standard error of the mean (SEM).

2.3.4 $\Delta G_{\text{Global}}$ and $\Delta G_{\text{Local}}$ measurements

For the WT and a subset of the mutant stem-loop (MS) RNAs (Table 2-2), RNA overall thermodynamic stability, $\Delta G_{\text{Global}}$, was measured using UV absorbance at 260 nm as a function of temperature with a Cary Model 400 Bio UV-visible spectrophotometer equipped with a Peltier heating accessory and temperature probe. All samples contained 10 mM potassium phosphate buffer, pH 7.0, 2 µM RNA, in a volume of 1 mL. For RNAs that were too stable to measure $\Delta G_{\text{Global}}$ under these conditions, urea was added to 4, 6 and 8 M, and the $\Delta G_{\text{Global}}$ was deduced by extrapolating to 0 M urea as described below. Prior to data collection, samples were heated from 20 °C to 95 °C, at 10 °C/min, held at 95 °C for five min, and cooled from 95 °C back to 20 °C at the same rate to ensure homogenous folding. Samples were heated at 1 °C degree per min from 20 °C to 95 °C. Identical traces were obtained by cooling, indicating a lack of hysteresis. $A_{260}$ data was collected in 0.5 min intervals and raw data were baseline corrected by subtraction of $A_{320}$ values at each temperature. The average hyperchromicity (Equation 1) and temperature were calculated from four curves and the standard error of the mean was determined for each average.

$$Y = \frac{\sum_{k=1}^{4} A_{260}^{X \degree C} - A_{260}^{20 \degree C}}{A_{260}^{X \degree C}}$$  \[1\]
For RNA with a single melting transition, the average hyperchormicity can be fit by Equation 2 to measure ΔH and Tm.

\[ Y = A_f + \frac{K}{1 + K} * A_f - A_u \]  \[ 2 \]

Here, \( A_f \) and \( A_u \) are the temperature dependent \( A_{260} \) of the folded and unfolded forms of the RNA, determined to be linear functions of the temperature. K is given by Equation 3, where \( T \) is the desired temperature (Kelvin, 310K for our calculations) for the \( \Delta G \) calculation, and \( R \) is the gas constant in units of kcal/(mol*K).

\[ K = e^{-\frac{\Delta H}{RT} \times \left(1 - \frac{T}{T_m}\right) + \frac{\Delta C_P}{R} \times \left(\frac{T_m}{T} - 1 + \ln\left(\frac{T}{T_m}\right)\right)} \]  \[ 3 \]

With an average \( T_m \) and \( \Delta H \) extrapolated using Equation 2, assuming a \( \Delta C_P \) of zero, the \( \Delta G_{\text{Global}} \) can be calculated using Equation 4.

\[ \Delta G = \ln K * -RT \]  \[ 4 \]

Error in \( \Delta G_{\text{Global}} \) is calculated using standard propagation of error, where:

\[ \Delta G = A - B, \text{where } A = \Delta H \text{ and } B = T_m \]

If \( \Delta G = A - ATB^{-1} \), then the error in \( \Delta G \) is the following:

\[ \sigma_{\Delta G}^2 = \sigma_A^2 \left( \frac{d \Delta G}{dA} \right)^2 + \sigma_B^2 \left( \frac{d \Delta G}{dB} \right)^2 \]

where,

\[ \sigma_A^2 = (\sigma \Delta H)^2, \sigma_B^2 = (\sigma T_m)^2, \frac{d \Delta G}{dA} = 1 - \frac{T}{B} = 1 - \frac{T}{T_m}, \text{ and } \frac{d \Delta G}{dB} = AB^{-1} = \frac{\Delta H * T}{T_m^2}. \]

Average hyperchormicity data for each RNA were fit using Equation 2 and overall thermodynamic stabilities were calculated using Equation 4 (Prism 4.3, GraphPad). For RNAs with melting temperatures approaching or greater than 95 °C, a linear extrapolation of \( \Delta G_{\text{Global}} \)
vs. urea concentration was applied to determine the $\Delta G_{\text{Global}}$ at 0 M urea (Figure 2-2) [200]. For all other RNAs, determination of $\Delta G_{\text{Global}}$ at standard buffer conditions was sufficient to produce minimal error in $\Delta G_{\text{Global}}$. All data used to calculate the reported $\Delta G_{\text{Global}}$ values were established using a minimum of three independently prepared samples at each buffer condition. A small and non-cooperative transition was observed in the 40-50 °C range for RNAs with less stable terminal or loop closing base-pairs (WT, MS2, MS6, MS7, MS10, and MS12). This transition was not present for RNAs with very stable terminal base-pairs (e.g., MS1 and MS5) and can be attributed to helical fraying. For all RNAs, only the major cooperative unfolding transitions were used in data fitting. Starting values were determined by examining the first derivative plots for each set of averages. Local stabilities, $\Delta G_{\text{Local}}$, for base-pairs were calculated using nearest-neighbor parameters at 1 M NaCl, 37 °C [201-203] (Table 2-2).
Table 2-2. Overall and local thermodynamic stabilities (kcal/mol) at 37 °C.

<table>
<thead>
<tr>
<th>RNA</th>
<th>10 mM Potassium Phosphate</th>
<th>1 M NaCl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measured ( \Delta G_{\text{Global}} )</td>
<td>Extrapolated ( \Delta G_{\text{Global}} )</td>
</tr>
<tr>
<td>WT</td>
<td>-14.7 (0.3)</td>
<td>-14.9 (0.2)</td>
</tr>
<tr>
<td>MS1</td>
<td>-16.6 (1.7)</td>
<td>-16.8 (0.5)</td>
</tr>
<tr>
<td>MS2</td>
<td>-14.7 (0.5)</td>
<td>-14.3 (0.02)</td>
</tr>
<tr>
<td>MS3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MS4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MS5</td>
<td>-10.9 (0.1)</td>
<td>-</td>
</tr>
<tr>
<td>MS6</td>
<td>-12.5 (0.3)</td>
<td>-12.4 (0.4)</td>
</tr>
<tr>
<td>MS7</td>
<td>-10.4 (0.1)</td>
<td>-</td>
</tr>
<tr>
<td>MS8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MS9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MS10</td>
<td>-11.1 (0.4)</td>
<td>-11.2 (0.1)</td>
</tr>
<tr>
<td>MS11</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MS12</td>
<td>-6.8 (0.1)</td>
<td>-</td>
</tr>
<tr>
<td>MS13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MS14</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MS15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MS16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MS17</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Standard errors shown in parenthesis.
- Not determined.
Extrapolated = determined using a urea titration (Materials & Methods).
\( \Delta G_{\text{Local \( (3bp) \)}} \) is shown for three base-pairs.
Figure 2-2 Determination of $\Delta G$

A) Representative thermal denaturation experiments in 10 mM potassium phosphate, pH 7.0, with various concentrations of urea. Red lines are indicative of the fit. The calculated $\Delta G_{\text{Global}}$ is shown for each. B) $\Delta G_{\text{Global}}$ at 0 M urea was extrapolated using the linear regression fit of $\Delta G_{\text{Global}}$ at 4, 6, and 8 M urea.
2.3.5 Correlation between frameshift efficiency and RNA stability

Frameshift efficiency was plotted as a function of overall and local thermodynamic stability. Overall and local thermodynamic stabilities were predicted at 1 M NaCl. Data were fit to a one-phase exponential decay function (Equation 5) (Prism 4.3, GraphPad). Here, the Amplitude, K, and Plateau are variables and $X_o$ is used to offset the exponential fit. $X_o$ was set to the most negative X value in the data set. Errors were determined using a 95% confidence interval.

$$y = Amplitude \cdot e^{-K(x-X_o)} + Plateau \ [5]$$

2.3.6 Modeling the HIV-1 frameshift site onto the eukaryotic ribosome

The 3.1 Å crystal structure of the Thermus thermophilus 30S ribosomal subunit in complex with mRNA and tRNAs [38], PDB ID 3I8G, was aligned to the 3.0 Å crystal structure of the Saccharomyces cerevisiae 40S ribosomal subunit [13], PDB IDs 3U5B and 3U5C, over 20 conserved nts in the rRNA decoding center [1, 11, 13, 14] (residues 780-800 in 3I8G and 991-1011 in 3U5B). The 5′ end of the HIV-1 frameshift site stem-loop, PDB ID 1Z2J [108] (residues 7-35), was connected to mRNA extending from the A site, PDB ID 3I8G [38]. The HIV-1 stem-loop was oriented to prevent steric clash using PyMOL (The PyMOL Molecular Graphics System, Version 1.5, Schrödinger, LLC.).
2.4 Results

2.4.1 Frameshift efficiency correlates with local stability of the HIV-1 frameshift site stem-loop

We utilized a well-established dual luciferase frameshift assay [109, 204] to quantitatively measure frameshift efficiency in rabbit reticulocyte lysate (RRL). The sequence of the frameshift site stem-loop was varied to dissect the relative contributions of local and overall RNA stability on HIV-1 frameshift efficiency. We hypothesized that once the ribosome is paused on the slippery sequence, the thermodynamic stability of the base-pairs encountered at the base of the stem-loop should be a critical determinant for frameshifting. After the ribosome has completed one translocation step, it moves away from the slippery site and the reading frame is established. Therefore, we further hypothesized that downstream base-pairs in the stem-loop should have a much lower impact on frameshifting. To test this hypothesis, 12 mutant stem-loop (MS) constructs (Figure 2-1B) were created using nearest-neighbor parameters [201-203, 205] to systematically alter the stability of different regions of the stem-loop. We define local stability ($\Delta G_{\text{Local}}$) as the thermodynamic stability of base-pairs directly downstream of the spacer (Figure 2-1A), as determined by their nearest neighbor interactions [201-203]. Global stability ($\Delta G_{\text{Global}}$) is defined as the overall thermodynamic stability of the stem-loop.

The thermodynamic stabilities ($\Delta G_{\text{Global}}$) were experimentally determined for a subset of RNAs (WT, MS1-2, MS5-7, MS10, and MS12) using UV monitored thermal denaturation (Table 2-2, Figure 2-3). Owing to the extreme stabilities of the structures [108], thermal denaturation curves were measured at low ionic strength (10 mM potassium phosphate buffer) in the presence of varying concentrations of urea, and extrapolated back to 0 M urea (Figure 2-2) [200]. Results followed the same trend as those predicted from nearest-neighbor parameters [201-203, 205] ($R^2 = 0.95$) (Figure 2-3). As expected, the measured stabilities were lower than the
predicted values at 1 M NaCl [206]. Upon correction of the experimental values to 1 M monovalent ionic strength [207], we observe an excellent agreement ($R^2 = 0.95$) between experimental and predicted free energies (Figure 2-3). Indeed, free energy prediction is robust for small, stable RNAs with no competing suboptimal folds [206].
Figure 2-3 Measured $\Delta G_{\text{Global}}$ values vs. predicted values.

Black symbols, data predicted with RNAstructure [205] using revised UG parameters [203] ($r^2 = 0.95$). Grey symbols, experimental data corrected to 1 M monovalent ionic strength ($r^2 = 0.95$). A linear regression function was used to fit both data sets.
Frameshift efficiencies for the different RNA constructs were measured (Figure 2-4, Table 2-3). Increases in the local stability of the first 3 base-pairs resulted in significant increases in frameshift efficiency (Figure 2-4A, MS1-5). In contrast, sequence changes that significantly lowered the local stability of the first 3 base-pairs resulted in decreased frameshift efficiencies (Figure 2-4A, MS10-12, Table 2-3). No correlation between frameshift efficiency and overall thermodynamic stability is observed (Figure 2-4B). Instead, we observe a strong correlation ($R^2 = 0.88$) between frameshift efficiency and local stability of the first 3 base-pairs at the base of the stem-loop using a one-phase exponential decay function (Figure 2-4C, Table 2-4). The frameshift efficiency for each variant frameshift site can be predicted using the parameters derived from the correlation, and each predicted frameshift efficiency falls within one standard deviation of its measured value (data not shown). These results support the hypothesis that the stability of the base-pairs at the base of the stem-loop is a primary determinant of frameshift efficiency.

Extremely stable RNA structures can promote long-term ribosomal stalling or “roadblocking,” [113, 208]. In the dual luciferase assay, roadblocking would result in decreased translation levels of the downstream firefly luciferase reporter gene product. However, the dual luciferase assay controls for this, as frameshift efficiencies are normalized relative to in-frame control constructs [109, 204]. Nevertheless, we asked if differential degrees of roadblocking might occur for our various constructs. The luminescence data reveal a consistent ratio of firefly/Renilla activity (data not shown) for all constructs (Table 2-3). These values were calculated using the luminescence data from the positive control dual-luciferase constructs, where the Renilla and firefly genes are in frame and the slippery site is mutated such that the ribosome cannot frameshift. The consistency in the relative expression levels of the reporter genes indicates that roadblocking, if occurring, is uniform for all constructs.
Table 2-3. In vitro frameshift efficiencies

<table>
<thead>
<tr>
<th>RNA</th>
<th>Frameshift Efficiency (%) (± SEM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>4.6 (0.5)</td>
</tr>
<tr>
<td>MS1</td>
<td>18.6 (2.2)</td>
</tr>
<tr>
<td>MS2</td>
<td>23.2 (2.1)</td>
</tr>
<tr>
<td>MS3</td>
<td>17.2 (1.9)</td>
</tr>
<tr>
<td>MS4</td>
<td>8.1 (0.9)</td>
</tr>
<tr>
<td>MS5</td>
<td>7.7 (1.4)</td>
</tr>
<tr>
<td>MS6</td>
<td>3.3 (0.2)</td>
</tr>
<tr>
<td>MS7</td>
<td>5.1 (1.2)</td>
</tr>
<tr>
<td>MS8</td>
<td>2.7 (0.4)</td>
</tr>
<tr>
<td>MS9</td>
<td>4.8 (0.5)</td>
</tr>
<tr>
<td>MS10</td>
<td>3.4 (0.7)</td>
</tr>
<tr>
<td>MS11</td>
<td>3.2 (0.5)</td>
</tr>
<tr>
<td>MS12</td>
<td>3.9 (0.3)</td>
</tr>
<tr>
<td>MS13</td>
<td>5.4 (0.6)</td>
</tr>
<tr>
<td>MS14</td>
<td>7.8 (0.7)</td>
</tr>
<tr>
<td>MS15</td>
<td>13.8 (1.4)</td>
</tr>
<tr>
<td>MS16</td>
<td>5.6 (0.4)</td>
</tr>
<tr>
<td>MS17</td>
<td>5.2 (0.5)</td>
</tr>
<tr>
<td>3HJ WT</td>
<td>2.5 (0.4)</td>
</tr>
<tr>
<td>3HJ Mut</td>
<td>4.7 (0.5)</td>
</tr>
<tr>
<td>3HJ MS1</td>
<td>12.2 (0.2)</td>
</tr>
<tr>
<td>3HJ M1</td>
<td>4.9 (0.8)</td>
</tr>
<tr>
<td>3HJ M2</td>
<td>4.5 (0.5)</td>
</tr>
<tr>
<td>3HJ M3</td>
<td>5.7 (0.7)</td>
</tr>
</tbody>
</table>

SEM = standard error of the mean
Figure 2-4. Frameshift efficiency is plotted as a function of RNA stability.

Data represent an average of at least three replicates, with the SEM indicated. A) *In vitro* frameshift assay results for WT and MS1-12 RNAs. Error bars representing the standard error of the mean are shown. The dashed line is indicative of the WT mean. B) Overall thermodynamic stability (at 1 M NaCl) vs. frameshift efficiency. C) Local stability (at 1 M NaCl) vs. frameshift efficiency. A single exponential decay function was fit to the data ($y = 18(\pm2)e^{-0.9(\pm0.3)(X+9.8)} + 3(\pm1), R^2 = 0.88$).
Table 2-4 Correlations between $\Delta G_{\text{Local}}$ (N bp) and frameshift efficiency.

Correlations were determined using the WT and MS1-12 RNAs

<table>
<thead>
<tr>
<th>N (base-pairs)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.61</td>
</tr>
<tr>
<td>2</td>
<td>0.70</td>
</tr>
<tr>
<td>3</td>
<td>0.88</td>
</tr>
<tr>
<td>4</td>
<td>0.86</td>
</tr>
<tr>
<td>5</td>
<td>0.87</td>
</tr>
<tr>
<td>6</td>
<td>0.87</td>
</tr>
<tr>
<td>7</td>
<td>0.84</td>
</tr>
<tr>
<td>8</td>
<td>0.55</td>
</tr>
<tr>
<td>9</td>
<td>0.44</td>
</tr>
<tr>
<td>10</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Correlations examined using a one-phase exponential decay equation.
2.4.2 Influence of spacer length in frameshifting

It has been hypothesized that during frameshifting, the mechanical force of translocation causes a build-up of tension that is transmitted through the spacer region (Figure 2-1A) and sensed at the anticodon-codon level [101]. We therefore investigated the influence of nt deletions and insertion in the spacer region (Figure 2-5). The WT construct was compared to a version with an adenosine insertion that increases the spacer length by 1 nt (Figure 2-5 MS13). Additionally, we created spacers with a single nt deletion (Figure 2-5 MS14) and 2 nt deletions (Figure 2-5 MS15-17). The resulting frameshift efficiencies were measured (Figure 2-5D).

Interestingly, MS15 shows a large increase in frameshift efficiency. This cannot be due to the 2 nt deletion, since MS16 and MS17 have the same spacers yet display WT-levels of frameshifting. We hypothesized that the 2 nt deletion in the spacer of MS15 increased frameshift efficiency by altering the base-pairs in the stem-loop encountered by the ribosome during frameshifting. In other words, by deleting 2 nt in the spacer, a ribosome footprint may extend 2 nt further into the stem. In support of this hypothesis, a very stable set of base-pairs are located 2 nt from the base of the stem (5’-GGC-3’/5’-GCC-3’). In MS16 and MS17, we replaced these base-pairs with the less stable base-pairs normally found at the base of the stem (5’-CUG-3’/5’-CAG-3’) (Figure 2-5C). Indeed, when these changes are made, the frameshifting efficiency is indistinguishable from WT, despite the apparent 2 nt spacer deletion (Figure 2-5D). Interestingly, the overall stability of MS17 is increased relative to MS16 (Figure 2-5C), yet the frameshift efficiency is unaltered (D). These data indicate that changes in the spacer region correspondingly alter the base-pairs encountered by the ribosome when it is engaged with the slippery site.

When plotting the data from all 18 RNA constructs studied (including the MS13-17 RNAs) as a function of overall RNA stability, no correlation is observed (Figure 2-5E). However, we observe a strong correlation between frameshifting and the thermodynamic stability of the first 3-4 base-pairs 8 nt downstream of the slippery site (Figure 2-5F and G). Conversely, the
correlations grow considerably weaker as more base-pairs are considered in the analysis (Figure 2-6). Likewise, no correlation is observed between local stability of base-pairs at the top of the stem-loop and frameshift efficiency (Figure 2-6L). The observed correlations are exponential functions with baselines of 2-4% frameshifting, which correspond to the lowest observed frameshift efficiencies in the presence of a stem-loop secondary structure downstream of the slippery site.

The stem-loop is flanked by a 5’ U and 3’ G (Figure 2-1A), that could potentially form a U-G wobble at the base of the stem. Inclusion of this wobble pair in the local stability term produces consistently weaker correlations between frameshift efficiency and local stability (Figure 2-7). To further address whether or not this U-G wobble pair can form during frameshifting, we modelled the frameshift site stem-loop and spacer onto the eukaryotic ribosome (Figure 2-8). The spacer was connected to the terminal nucleotide in the A-site, to recapitulate the position of the ribosome when it is engaged on the slippery sequence in the 0 reading frame. The model indicates that the minimal spacer distance between the slippery site and the stem-loop is 7 nt; however formation of a U-G wobble at the base of the stem is blocked by steric clash with the ribosomal S3 protein (Figure 2-8B and D). Therefore, experimental data and structural modelling support an HIV-1 frameshift site spacer length of 8 nts.
Figure 2-5. The role of the spacer in frameshift stimulation.

A) Changes in spacer and stem-loop sequence are shown for MS13-17. B) Spacer sequences for WT and MS13-17, with nt insertions shown in bold and underlined. C) Changes in stem-loop sequence for MS16-17 are bold and italicized. D) In vitro frameshift assay results for the WT and MS13-17 RNAs. The dashed line is indicative of the WT mean. Data represent an average of at least three replicates, with the SEM indicated. E) Overall thermodynamic stability (at 1 M NaCl) vs. frameshift efficiency. F) Local stability (3bp, at 1 M NaCl) vs. frameshift efficiency. A single exponential decay function was fit to the data ($y = 16(\pm2)e^{-0.8(\pm0.2)(X+9.9)} + 3(\pm1)$, $R^2 = 0.80$). G) Local stability (4bp, at 1 M NaCl) vs. frameshift efficiency. A single exponential decay function was fit to the data ($y = 9(\pm2)e^{-1.0(\pm0.3)(X+12.5)} + 4(\pm1)$, $R^2 = 0.83$).
Figure 2-6 Defining the relationship between stem-loop local stability and frameshift efficiency.

(A-J) Using a one-phase exponential decay equation, correlations between local stability and frameshift efficiency for WT and MS1-17 were examined. Local stability was calculated using nearest-neighbor parameters for the indicated number of base-pairs. K) Global stability vs. frameshift efficiency for WT and MS1-17. L) Local stability of the top of the stem-loop does not correlate to frameshift efficiency. A representative plot is shown.
Figure 2-7 HIV-1 frameshift efficiencies in the context of a 7 nt spacer.

A) MS RNA secondary structures redrawn in the context of a 7 nt spacer. Local stabilities are shown for 3 to 6 base-pairs. B) Local stability vs. frameshift efficiency. A single exponential decay function was fit to the data. The number of base-pairs used to calculate local stability is indicated.
Figure 2-8
Figure 2-8. The HIV-1 frameshift site stem-loop modeled onto the eukaryotic ribosome.
The frameshift site stem-loop, PDB ID 1Z2J, was modeled onto the 40S *Saccharomyces cerevisiae*
ribosomal subunit, PDB IDs 3U5B and 3U5C, using tRNA and mRNA coordinates from the 30S
*Thermus thermophilus* structure, PDB ID 3I8G. A) Ribosomal RNA and proteins are shown in
grey, with ribosomal protein S3 surface in light green and the HIV-1 stem-loop in red. Part of
the A site tRNA is visible in dark blue. B) Close-up view of the HIV-1 stem-loop on the surface
of the ribosome, tilted and rotated approximately 180° relative to A. The potential U-G wobble
pair would occur between the last U in the spacer (light blue) and the G (red) at the base of the
stem. C) “Side” view. tRNAs from the 30S *Thermus thermophilus*, PDB ID 3I8G, are blue, light
green, and orange for the A, P and E sites, respectively. D) Cut-away view showing the path of
the mRNA. The mRNA corresponding to the slippery site is shown in yellow, and the 8 spacer
nts are shown in light blue.
2.4.3 The HIV-1 3HJ modulates frameshift efficiency

Within viral capsids, the HIV-1 frameshift site RNA is part of a conserved 3HJ secondary structure (Figure 2-9A) [170, 171]. It has been hypothesized that the role of this secondary structure is to slow down the rate of translation [170], which in turn may modulate frameshift efficiency. We therefore compared the 3HJ secondary structure (3HJ W[186]T) to a similar construct with mutations designed to disrupt secondary structure formation in the P1 and P2 helices (Figure 2-9B, 3HJ Mut). We observe a significant decrease in frameshift efficiency, from 4.6(±0.5) to 2.5(±0.4)%, when the 3HJ secondary structure is present (Figure 2-9; Table 2-3, compare 3HJ WT to 3HJ Mut). As expected, there is no significant difference between the observed frameshifting efficiencies of the 3HJ Mut and the WT construct used above (4.7(±0.5)% and 4.6(±0.5)%, respectively). The observed frameshift efficiencies for our 3HJ WT and WT reporter constructs in RRL both fall within the range of previously measured frameshifting efficiencies for HIV-1 in vivo, which range from 2 - 5% [98, 103, 168, 186].

We hypothesized that the 3HJ P1 helix (Figure 2-9A) was primarily responsible for 3HJ frameshift modulation. Two 3HJ mutants (3HJ M1-2) were designed to test the functional role of individual helices within P1. A third mutant, 3HJ M3, was created as a control, as it is not predicted to impact the 3HJ (Figure 2-9A) or WT (Figure 2-1A) secondary structures. Disruption of base-pairing in two different regions of P1 returned frameshifting efficiency to WT levels (Table 2-3, 3HJ M1- 5.9(±0.8)% and 3HJ M2- 4.5(±0.5)%). Surprisingly, the 3HJ M3 construct increased frameshifting to 5.7(±0.7)% (Figure 2-9C). Mutations in this region may have disrupted secondary structure in an alternative 3HJ structure in conformational exchange with the SHAPE detected 3HJ secondary structure [170, 171] (personal communication with K. Weeks).
Next, we tested the local stability hypothesis in the context of the 3HJ secondary structure by increasing the local stability of two base-pairs in the P3 helix (Figure 2-9A). The two base-pair change (3HJ MS1) results in a large, 5-fold increase in frameshift efficiency (Figure 2-9C, Table 2-3). This increase is similar to the 4-fold difference between the MS1 and WT RNAs (Figure 2-4A, Table 2-3). We conclude that the 3HJ secondary structure indirectly modulates frameshifting, likely by altering the kinetics of translation, as previously hypothesized [170]. This effect must happen prior to frameshifting, as the ribosome disrupts the 3HJ secondary structure as it encounters the slippery site. Once the ribosome is engaged with the slippery site, local stability is the primary determinant of frameshifting efficiency, as illustrated by comparison of 3HJ WT to 3HJ MS1 (Figure 2-9C).
Figure 2-9. Effect of the 3HJ on frameshift efficiency.

A) The secondary structure of the wild-type 3HJ (3HJ WT) is shown. Sequence changes in 3HJ MS1, 3HJ M1, 3HJ M2, and 3HJ M3 are boxed. B) Sequence changes in the 3HJ mutant (3HJ Mut) are shown in bold and italicized. C) *In vitro* frameshift assay results for the 3HJ RNAs. Data represent an average of at least three replicates, with the SEM indicated.
2.5 Discussion

In this work, we report a strong correlation between the thermodynamic stability of the first three to four base-pairs at the base of the stem-loop and frameshift efficiency in HIV-1. We therefore hypothesize that the frameshift mechanism involves a thermodynamic block to translocation, determined by the local stability of base-pairs positioned directly at the mRNA entrance channel. This is in agreement with previous studies investigating antisense-induced frameshifting using either mixed Locked nucleic acid/DNA [111] or morpholino/RNA oligonucleotides [110]. When these oligonucleotides were used to direct antisense induced frameshifting, the local stability of the duplex was also critical to frameshift stimulation [110, 111].

In light of the local stability hypothesis, we can re-examine data from prior studies that investigated trends between thermodynamic stability of the HIV-1 stem-loop and frameshift efficiency [136, 155, 156]. Indeed, we find that these results are generally consistent with local stability being the primary determinant in frameshift efficiency. For example, Bidou et al. [156] investigated the frameshift efficiencies of five mutant stem-loops in both yeast and mouse NIH3T3 cells, and observed decreased frameshift efficiency when local stability was decreased. Telenti et al. [136] used a yeast frameshift reporter assay to test naturally occurring stem-loop variants, and observed the greatest decrease in frameshift efficiency for a point mutation that disrupts the C-G base-pair at the base of the stem, while little to no effect was observed for mutations in the upper regions of the stem-loop. Hill et al. [155] investigated mutant HIV-1 virus replicating in 293T cells. Of the four mutant stem-loops investigated in this study, two eliminated base-pairing at the base of the stem and these showed large reductions in frameshifting; the other two mutants had similar local stabilities to WT and similar levels of frameshifting.
Mutations in the frameshift site that arise in response to cytotoxic T-cell escape [209] and protease inhibitor resistance [210, 211] are also consistent with our results. Prado et al. [209] investigated the frameshift efficiency of four HIV-1 strains with mutations in the frameshift site. In this study, the only mutation that produced a significant change (decrease) in frameshift efficiency was one with a decreased local stability due to disruption of the first base-pair in the stem-loop. Nijhuis et al. [210] examined the effects of three point mutations in the spacer region, all of which do not impact frameshift efficiency. Knops et al. [211] investigated mutations in the frameshift site that either mutate the ACAA tetraloop to CCAA, or change a C-G base-pair in the upper stem-loop to a U-G wobble (C2108U). Consistent with the local stability hypothesis, none of these mutations alter the frameshift efficiency.

If the local stability of three to four base-pairs is sufficient to determine frameshift efficiency, why does the HIV-1 frameshift site stem-loop contain eleven Watson-Crick base-pairs? We can see two possible explanations for this. First, the additional base-pairs ensure that the stem-loop has a high probability of folding and cannot be out-competed by suboptimal folds, which can severely impact HIV-1 replication [212]. In the >9,400 nt genome, there are a total of only 11 helices that are equal or larger in size [170]. Secondly, the additional base-pairs serve to cooperatively stabilize the base-pairs at the base of the stem. These effects may explain why severely truncated constructs produce lower frameshift efficiencies compared to stem-loops with identical local stability (MS8 vs. WT, MS3 vs. MS4). Cooperative stabilization of local stability may also explain why severe truncations of a hairpin downstream of the Simian retrovirus type-1 (SRV-1) slippery site results in lower frameshift efficiency [105].

Results here demonstrate a lack of sequence dependence on frameshift efficiency. For example, an RNA stem-loop with an entirely different sequence, decreased overall thermodynamic stability, and increased local stability (MS5) increased frameshifting above WT
levels. Therefore, sequence-specific interactions with the translational machinery are not required, in agreement with conclusions drawn for the SRV-1 mechanism [105].

The observed frameshift efficiencies for the 3HJ WT and WT reporter constructs in RRL both fall within the range of previously measured frameshifting efficiencies for HIV-1 in vivo, which range from 2 - 5% (9,15,21,24). The wide range of observed frameshifting efficiencies in vivo is likely influenced by viral and cellular factors, for example, modulation of translation initiation by the HIV-1 TAR RNA structure [165], and polysome density [164]. We find that the conserved 3HJ secondary structure in the HIV-1 genomic RNA [170] causes a significant decrease in frameshift efficiency (Figure 2-9). Disruption of base-pairing in the PI helix restores frameshifting to WT levels (3HJ M1-2: Figure 2-9 and Table 2-3). A decreased level of frameshifting in 3HJ WT is consistent with the previous hypothesis that the 3HJ secondary structure induces ribosomal pausing [170]. Pausing at the upstream secondary structure may promote stacking of consecutive ribosomes [169], promoting a net decrease in frameshift efficiency, because the mRNA would have less time to refold between ribosomes. Our data support this model, but also indicate that local stability has a far greater impact on frameshift efficiency (Figure 2-9).

Prokaryotic ribosomes use two active mechanisms during translation to unwind RNA [197]. In the first mechanism, the ribosomal helicase activity raises the free energy of an encountered base-pair by +0.9 kcal/mol [197]. This destabilizes the base-pair, which can then be opened by the mechanical force generated by translocation. If the base-pair is resistant to this force, tension may be created which is sensed at the codon-anticodon base-pairs [141, 142, 166, 197, 198]. Because GC pairs require more force for unwinding [197], the tension sensed in the decoding center would be proportional to the local RNA stability [198]. The mechanical tension may either cause the tRNAs to slip 1 nt in the 5’ direction [101, 141, 142, 144, 197], or
alternatively, cause the ribosome to translocate incompletely by 2 instead of 3 nt [145], which would also result in a -1 frameshift.

When the HIV-1 frameshift site is modelled onto the eukaryotic ribosome, base-pairs critical for frameshifting are positioned at the entrance to the mRNA entry channel (Figure 2-8B and D), in agreement with chemical probing and toeprinting results with a prokaryotic ribosome stalled on the HIV-1 frameshift site [167]. Interestingly, the decoding center and the mRNA channel are highly conserved between prokaryotes and eukaryotes [1] and a bacterial translation system produces similar levels and changes in frameshift efficiency in response to changes in the HIV-1 stem-loop sequence [128].

Our data support an 8 nt spacer length between the slippery site and the stem-loop, as the effect of deletions in the spacer correlate with corresponding changes in stem-loop local stability 8 nt downstream of the slippery site (Figure 2-5). Consistent with this idea, deletion of 1 nt in the spacer region of the beet western yellow virus (BWYV) -1 PRF site promotes the melting of the first base-pair in the downstream structure [213]. If the mRNA channel length is maintained, it follows that spacer lengths in all -1 PRF sites should be ≥ 7 nts in length. Yet, some frameshift sites have been drawn with 5-6 nt spacers (reviewed in [101, 148]), including that of BWYV [213]. Our data suggest that these frameshift site structures may be partially unwound at the time of frameshifting, in order to accommodate the requisite spacer length and positioning of the slippery sequence in the ribosomal decoding center. Unfortunately, there are currently no high-resolution structural views of ribosomes engaged with frameshift site structures. In conjunction with functional studies such as the one presented here, high-resolution structural views will ultimately be required to define the frameshifting mechanism. Prior studies have observed relationships between spacer length and -1 PRF efficiency in various systems [88, 122, 146, 147] and are consistent with a spacer length of 7-8 nts and local
stability being the primary determinant in frameshift efficiency. Spacer lengths of 6-8 nts produced the highest level of -1 PRF for the antisense oligonucleotides, stem-loop and pseudoknot stimulatory structures [88]. In agreement with our conclusions, these spacers position base-pairs with strong local stabilities at the entrance to the mRNA channel.

HIV-1 group M subtype B is the dominant form of HIV-1 in North and South America, Europe, Japan, Thailand, and Australia. The less common non-B subtypes (A, C, D, E, F, G, H, J, and K) have decreased local stabilities; for example, a frequent C to U mutation in the first base-pair of the stem-loop in these subtypes results in formation of a U-G wobble pair in place of a C-G [190, 214]. Interestingly, the exponential relationship we observe predicts that such a change would have little effect on frameshift efficiency. For instance, mutants MS10-12 all incorporate U-G wobble pairs in the stem-loop base, which significantly destabilize the local stability by +5.5 kcal/mol relative to WT (Figure 2-1). Nevertheless, these mutations result in near WT frameshift efficiencies (Figure 2-4A), owing to the exponential relationship between local stability and frameshift efficiency (Figure 2-4B and Figure 2-5F and G). These results are consistent with the observed frameshift efficiencies of the less common sub-types [190]. Finally, a randomized trial of HIV patients receiving protease inhibitor therapy examined mutations in the gag-pol frameshift site and found no relationship between overall stability of the stem-loop and virological response [215]. This observation is consistent with our results that show no correlation between overall stability and frameshifting (Figure 2-4B and Figure 2-5E).
2.7 Acknowledgements

We thank Jordan Burke, Ashley Richie, and Lauren Michael for helpful discussions. We also thank Raymond Gesteland (University of Utah) for the generous gift of the p2luc plasmid DNA and Prof. A.C. Palmenberg and her laboratory (University of Wisconsin-Madison) for equipment use. Finally, UV data were obtained at the University of Wisconsin - Madison Biophysics Instrumentation Facility, which was established with support from the University of Wisconsin - Madison and grants BIR-9512577 (NSF) and S10 RR13790 (NIH).

2.8 Funding

The University of Wisconsin - Madison; NSF [BIR-9512577]; National Institutes of Health [S10 RR13790] to obtain UV data at the University of Wisconsin - Madison Biophysics Instrumentation Facility. National Institutes of Health [GM072447 to S.E.B.] Funding for open access charge: NIH GM072447
Chapter 3: Investigating RNAs involved in translational control by NMR and SAXS

The work presented in this chapter was published as:


J.E.B wrote and generated figures for section 3.3, and also created figure 3-2 and figure 3-12.

The remainder of the work was by K.D.M.
3.1 Introduction

Translational control, broadly defined, is the post-transcriptional regulation of gene expression within a cell. The majority of translational control is thought to rely on the interaction of regulatory proteins with messenger RNA (mRNA) [216]. However, a growing body of evidence indicates that mRNA structures play an important role in translational control. For example, certain classes of riboswitches form dynamic RNA structures that influence translation [217]. Additionally, RNA-RNA interactions can broadly control translation, as in the case of microRNA [218]. Therefore, it is important to consider the role of RNA structure and intrinsic dynamic motions that can modulate RNA function during translation [219, 220].

Historically, the most informative structural information related to RNA structure and its role in translation has been derived from X-ray crystallography, with the most striking example of this being the structure of the ribosome [221-227]. However, obtaining diffraction quality crystals of RNA can be tremendously challenging. Alternatives to X-ray crystallography are NMR and small angle X-ray scattering (SAXS), which are solution-based methods. In addition to providing structural information, NMR can be used to investigate dynamic motions. In this chapter, we review methods for investigating RNA structure and dynamics in solution. In sections 3.2 and 3.3, we focus predominately on NMR spectroscopy and SAXS as complementary techniques for analyzing RNA structure and dynamics in solution. In section 3.4, specific examples of RNAs involved in translational control are discussed.

3.1.1 From RNA secondary structure prediction to biophysical studies

The first step to understanding RNA structure using any solution technique is identification of the RNA sequence of interest and determination of its secondary structure. Prediction of RNA secondary structure is routinely accomplished with Mfold [228]. Mfold predicts
secondary structure of RNA using free energy minimization based on thermodynamic nearest-neighbor parameters. Mfold is approximately 73% accurate for RNAs less than 700 nt in length [202], and its predictive power greatly improves for smaller RNAs. While the Mfold algorithm alone cannot predict base-pairs involved in pseudoknot formation, several alternative programs have been developed that can [229-231]. The accuracy of secondary structure prediction by Mfold can be improved by incorporating experimental data from sources such as chemical probing [232], NMR [231] and/or microarray [233] experiments. Of these, NMR provides the most direct and rigorous experimental verification of secondary structure, since it is the only method that can directly detect the hydrogen-bonded protons involved in base-pair formation. For larger RNA sequences, accurate prediction of secondary structure is more challenging due to increased base-pairing possibilities leading to alternative secondary structures with similar overall free energies. Prediction of these large secondary structures can be assisted by the use of phylogenetic analysis [234, 235] and chemical probing [232, 236, 237]. If the sequence of interest is derived from a larger RNA, it is critical to verify that the secondary structure of the RNA was not perturbed due to the truncation. This can be demonstrated with NMR, by confirming that the chemical shifts for each subdomain are consistent with those observed in the context of the larger RNA. This approach was recently used to examine the secondary structure of a 356 nt HIV-1 genome packaging 5’ leader RNA [238].

In vitro transcription with T7 RNA polymerase is standard for production of the milligram quantities of RNA required for NMR [239-241]. The majority of the DNA template for transcription can be single-stranded, but the 20 base-pair T7 RNA polymerase promoter region must be double stranded, as described [239]. Chemical synthesis of the DNA template strand with phosphoramidite chemistry is customary for synthesis of oligonucleotides 90 nts in length or smaller. For DNA constructs longer than 90 nts, standard cloning techniques [242] are used to insert the DNA duplex into a DNA plasmid. Once milligram quantities of plasmid DNA are
produced, the plasmid is linearized with a restriction enzyme that cuts immediately following the end of the DNA template sequence [231]. This linearized product is then used for in vitro transcription. Additionally, T7 RNA polymerase requires at least one but preferably two guanosine residues at the 5′ end of the RNA construct for initiation of transcription [239].

Purification techniques for milligram quantities of RNA have historically relied on a denaturing purification process followed by a refolding step [231]. These techniques have been explained in great detail elsewhere [231, 243]. Generally NMR data acquisition requires low monovalent salt concentrations, 100 mM or less, because higher salt concentrations decrease signal to noise [244]. Excess salt from the purification process can be removed with size-exclusion chromatography, buffer exchange using dialysis, or centrifugal filtration [245]. Once the RNA sample is in low ionic strength buffer, it can be subjected to a refolding procedure by diluting to ~10 µM, heating to 90 – 95°C for 1 to 2 min., followed by rapid cooling on ice. This step promotes homogenous folding of the RNA and disrupts inter-molecular associations (non-native base-pair formation between molecules) that can occur at higher concentrations. The chemical integrity of the RNA and conformational homogeneity should be analyzed using denaturing and non-denaturing PAGE [246], respectively.

A handful of approaches for native purification of RNA have also been published [245, 247-249]. These studies employ size exclusion [245] and affinity chromatography [247-249]. Native purification is advantageous because it eliminates 1) RNA aggregates that can form during the ethanol precipitation step after transcription [245], and 2) RNA misfolding [247-249]. RNA aggregation and misfolding can be particularly problematic for RNAs as small as 50 nucleotides. Furthermore, for some RNAs co-transcriptional folding is critical for their homogeneity [250] and catalytic activity [249]. As will be discussed in section 1.3, sample homogeneity is required for high quality SAXS data acquisition.
3.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectroscopy is a powerful tool for investigating the solution structures of small molecules, proteins, and nucleic acids. Its power stems from direct detection of atoms (typically $^1$H, $^{13}$C or $^{15}$N) that report structural and dynamic information in solution. RNA secondary, tertiary and quaternary structure, as well as dynamics, can be probed using NMR. There are numerous examples of RNA solution structures solved using NMR; 489 NMR structures with RNA have been deposited in the Protein Data Bank (www.pdb.org).

NMR has several limitations. The first limiting factor is molecular size. The signal-to-noise ratio in an NMR experiment is dependent upon spectrometer sensitivity, the sample concentration and the NMR relaxation rate. As molecular weight increases, the NMR relaxation rate also increases, which results in line broadening and rapid loss of signal. Transverse relaxation optimized spectroscopy (TROSY) has greatly extended the molecular weight limit for molecules studied with NMR [251, 252]. TROSY partially counters the dipole-dipole relaxation between neighboring nuclei using constructive interference from chemical shift anisotropy [240, 253]. Currently, the practical size limitation for RNA structure determination by NMR is approximately 40 kDa, or around 100 nucleotides. However, in some favorable cases, NMR can be used to determine secondary structures of much larger RNAs [238]. Approaches for resolving chemical shift overlap for large RNAs are discussed in greater detail in section 3.2.3.

The second potentially limiting factor is sample concentration. NMR is an inherently insensitive technique, so the method generally requires high concentrations of sample. For biological macromolecules, this is typically between 0.1-2 mM. Even with 1 mM sample concentrations, signal averaging of multiple experiments is typically used to increase the signal-to-noise.
Resolution is the third major challenge in NMR, and is particularly problematic for RNA. Unlike proteins, which are composed of 20 distinct amino acids, nucleic acids are made up of only four chemically similar nucleotides, which results in a small degree of chemical shift dispersion (Figure 3-1). Additionally, the proton density within a nucleotide is concentrated primarily in the ribose ring, and the majority of protons in the ribose ring resonate in a narrow spectral window of ~1 ppm. This high level of chemical shift overlap can make it challenging to obtain complete resonance assignments. Therefore, isotopic $^{13}$C and $^{15}$N labeling [254, 255] is required to resolve $^1$H chemical shifts into additional dimensions by correlating $^1$H shifts to their bonded carbon and/or nitrogens. Such multidimensional heteronuclear NMR experiments are instrumental in resolving overlapped $^1$H chemical shifts [240].
Figure 3-1. $^1$H 1D spectrum of a 32 nt RNA.

(PDB ID 1XHP) Characteristic $^1$H chemical shift ranges are indicated for nitrogen-bonded protons (white bars) and carbon-bonded aromatic and ribose protons (black bars). Imino protons involved in Watson-Crick (WC) base-pairing have a chemical shift range of 10 – 15 ppm. DSS is 4,4-dimethyl-4-silapentane-1-sulfonic acid, a chemical shift calibration standard.
3.2.1 NMR analysis of base-pairing interactions and secondary structure.

The first step in investigating RNA structure using NMR is measuring a 1D $^1$H spectrum. The observation of imino proton resonances from G and U nucleotides (note there are no imino protons on A and C), in the 10-15 ppm range is indicative of hydrogen bond formation and base-pairing (Figure 3-1). The imino resonances can be assigned to their respective nucleotides using $^1$H-$^1$H 2D nuclear Overhauser effect spectroscopy (NOESY) experiments. The NOESY experiment provides information about the number and type of base-pairs, as well as their sequential neighbors [256]. In a $^1$H-$^1$H 2D NOESY experiment, cross-peaks (NOEs) are observed between protons that are less than 6 Å apart. In a standard A-form helix, an NOE cross-peak is observed between adjacent base-pairs [240] (Figure 3-2B). Thus, sequential neighbors for base-pairs are identified by “walking-through” these cross-peaks (Figure 3-2C).

The HNN-COSY experiment [257] is extremely useful for investigating base-pairing in $^{15}$N-labeled RNA, because it directly detects hydrogen bonding. The simplest application of this experiment correlates the donor and acceptor $^{15}$N atoms across N-H•••N hydrogen bonds in Watson-Crick base-pairs [251, 257]. During the experiment, magnetization is transferred in two steps (Figure 3-3): first, from the imino proton to its directly bonded donor nitrogen and second, from the donor nitrogen through the hydrogen bond to the acceptor nitrogen. The magnetization is transferred back to the imino proton through the reverse of the first two steps. These experiments have been extended to detect NH•••O=C, OH•••N-, OH•••O=P and NH$_2$•••O=P hydrogen bonds [258].
Figure 3-2. \( ^1H-^1H \) 2D NOESY experiment.

This experiment was utilized for assignment of imino protons in the SIV frameshift stem-loop RNA structure [109]. (A) RNA secondary structure. (B) Base-pairs stacked in A-form helical regions are oriented such that the G and U imino protons are within 5 Å and will exhibit a cross peak in the \( ^1H-^1H \) 2D NOESY. (C) Zoomed-in region of 2D NOESY showing imino crosspeaks. A sequential walk is shown. Colored lines correspond to (A).
Figure 3-3. The HNN-COSY experiment directly detects base-pairing by correlating the acceptor and donor nitrogen atoms involved in the hydrogen bond. Magnetization is transferred from the imino proton to the donor nitrogen (1), and then from the donor nitrogen through the hydrogen bond to the acceptor nitrogen. (2) After labeling the magnetization, it is transferred back to the imino proton through the reverse process. Adapted from [251].
Figure 3-4. RDCs (D$_{ij}$) between spins i and j provide long-range distance restraints for the average orientation ($\theta$) of the internuclear vector relative to the magnetic field ($B_o$).

Here, $\gamma_i$ and $\gamma_j$ are the gyromagnetic ratios of spins i and j, respectively, $r_{ij}$ is their internuclear distance, $\hbar$ is Plank’s constant, $\mu_o$ is the magnetic permittivity of vacuum, and the angular term is shown in brackets. RDCs are not observed in isotropic solution conditions because the angular term averages to nearly zero. Adapted from [259].
3.2.2 Towards solving NMR structures

The quality of the ensemble of NMR structures is dependent upon the quality of restraints, which includes the overall number of distance restraints (NOEs) measured [260]. RNA structure determination by NMR has historically relied on the measurement of many short range distance restraints. This approach requires unambiguous chemical shift assignment for as many of the protons in the RNA as possible. For RNAs larger than 30-40 nucleotides, uniform $^{13}$C, $^{15}$N labeling may not be sufficient to resolve proton spectral overlap. To circumvent this problem, RNAs can be labeled using nucleotide-type specific labeling strategies [240, 261]. Selective deuteration [262] and perdeuteration [261] can also be used to dramatically simplify NMR spectra and resolve chemical shift overlap (discussed in section 3.2.3).

Chemical shift assignment is accomplished in two basic ways. The first and most direct is via through-bond experiments that take advantage of scalar couplings between protons and neighboring $^1$H, $^{13}$C and $^{15}$N nuclei [231, 240, 251, 263]. Because chemical shifts are dominated by local electronic environment, the chemical shift of the proton and its directly bonded and neighboring nuclei can be used to help assign the identity of the proton. Thus, experiments that correlate a proton chemical shift to its directly bonded nucleus, such as 2D heteronuclear single quantum coherence (HSQC), are invaluable for proton resonance assignment [240]. Experiments that correlate multiple bonds can provide even more information. For example, triple resonance HCN-type heteronuclear NMR experiments can be used to correlate nucleobase aromatic protons to a particular ribose group by appropriately transferring magnetization through the glycosidic bond [264].

The second strategy for proton resonance assignment is based on through-space transfer of magnetization, via the NOE. The nucleotide structure contains a number of short (<6 Å) interproton distances that give rise to NOEs. Additionally, RNA structures are dominated by the A-form helical geometry, which gives rise to predictable NOE patterns that can be used to
obtain resonance assignments of neighboring nucleotides [240]. Because the intensity of the NOE is distance dependent \( r^{-6} \), NOEs obtained from these experiments are utilized as distance restraints for use in restrained molecular dynamics simulations to produce an ensemble of structures that satisfy the restraints [265].

Torsion angle restraints are also important for structure determination. The sugar-phosphate backbone and ribose ring are characterized by six backbone torsion angles and the ribose sugar pucker, respectively. Ribose sugar puckers in A-form RNA are C3'-endo, but in non-helical regions such as loops the ribose ring may adopt a C2'-endo conformation or undergo dynamic exchange between C3'- and C2'-endo conformations. The sugar pucker can be analyzed using 2D COSY or TOCSY experiments [240]. Backbone torsion angle measurements around the phosphodiester bond can be quantitatively measured [266]. However, experiments designed to quantitatively measure backbone torsion angles are often impractical for RNAs larger than 30 nucleotides, because they rely on the ability to resolve phosphorous chemical shifts, which are usually extensively overlapped.

Ideally, an NMR structure should be constrained by a large number of long-range NOEs between protons far apart in sequence [260]. However, the vast majority of NOEs occur between protons within a nucleotide (intraresidue) or between neighboring nucleotides. While these NOEs are important for establishing resonance assignments and defining local structure, they cannot sufficiently define the global RNA structure. RNA NMR structure quality can therefore be significantly improved by inclusion of long-range distance restraints. Such information can be obtained through the measurement of residual dipolar couplings (RDCs) [267, 268]. RDCs measure the orientation of a bond vector relative to the overall alignment of a molecule within the magnetic field. Therefore, RDCs provide angular restraints that improve both the local and global structural quality [251], and can also be used to characterize RNA dynamics (discussed in greater detail in section 3.2.4).
RDCs are not normally observed in solution under isotropic conditions because the overall tumbling of a molecule causes the angular term in the RDC equation (Figure 3-4) to average to a value that approaches zero [269]. However, by imparting a small degree of alignment to the sample, the dipolar coupling retains information about the angle of a bond vector relative to the alignment tensor. RDC values are determined by subtracting the measured dipolar coupling in isotropic conditions from the coupling in partially aligned conditions. For RNA, alignment is typically achieved by adding filamentous bacteriophage (Pf1) to the RNA [267, 270].

3.2.3 Solution structures of large RNAs

Solving solution structures of even moderately sized RNAs (50-100 nts) can be extremely challenging due to spectral overlap and fast relaxation rates. Only a handful of RNA structures greater than 75 nucleotides have been solved using NMR. These include the 77 nt HCV IRES domain II [271], an 86 nt tetraloop receptor complex [262], a 101 nt core encapsidation signal of Moloney murine leukemia virus (MoMuLV) [272] (see section 3.4), a 102 nt RBSE in TCV [273], and the 140 nt dimeric genome packing signal in MoMuLV [274]. Each of these studies employed a different approach to reduce spectral overlap and improve NMR relaxation rates associated with large macromolecules.

The most common approach for studying large RNAs has been the so-called “divide and conquer” approach (reviewed in [263]). Here, the RNA is separated into small, thermodynamically stable helical domains. By breaking the RNA into smaller domains, spectral overlap is reduced and assignment of the domain is often feasible. Unfortunately, the structure of the subdomain may change in the context of the larger, more biologically relevant RNA, if it is involved in tertiary interactions [275]. Therefore, it is critical to verify that the chemical shifts for each subdomain are consistent with those observed in the context of the larger RNA.
Several labeling strategies can be used to facilitate chemical shift assignment. Because $^2$H nuclei are not visible by $^1$H NMR, selective deuteration can greatly simplify highly crowded regions of the spectrum, like the ribose region [262]. Alternatively, incorporation of a protonated nucleotide-type into an otherwise perdeuterated RNA chain is another successful strategy for greatly simplifying NMR spectra of large RNAs [261, 272, 274]. Deuteration also improves relaxation properties for larger RNAs, leading to an improved signal-to-noise ratio [276]. Deuteration has been useful for several large RNA structures [262, 272, 274].

A second method facilitating chemical shift assignment uses selective nucleotide-type $^{13}$C, $^{15}$N labeling in conjunction with filter-edited NOESY experiments [272, 277]. Filter-edited NOESY experiments identify intermolecular NOEs between isotope labeled and unlabeled nucleotides. We found that 2D filter-edited experiments take approximately 24 hours and were sufficient for assignment of an 86 nt RNA complex [262].

### 3.2.4 RNA dynamics

It is increasingly clear that RNA can sample a wide range of structural conformations due to inherent dynamics [220]. A complete understanding of RNA structure and function requires knowledge of its dynamic motions. Many biophysical approaches have been developed to probe RNA dynamics: molecular dynamics simulations [278-280], single molecule fluorescence spectroscopy [281-283], time-resolved hydroxyl radical footprinting techniques [284, 285], single-molecule force measurements [286], and NMR methods [259, 287-289]. In this section we will highlight the different approaches for measuring RNA dynamics using NMR. NMR provides a unique way to investigate dynamics over time scales encompassing biologically relevant motions (Figure 3-5) [259, 287, 289]. Dynamic motions can occur at the atomic level or can correspond to entire domains [259]. RNA motions in the picosecond to nanosecond range reflect local motions [287]. These motions are characterized by measuring
the longitudinal ($T_1$) and transverse ($T_2$) relaxation rates for $^{13}$C or $^{15}$N nuclei, and the heteronuclear NOE between these nuclei and their bonded protons. The Lipari-Szabo model free approach [290] is the most popular method utilized to quantitatively describe these motional amplitudes and their rates. These types of experiments have proved very useful in characterizing local motions for several RNAs [291-296]. Slower motions in the microsecond to millisecond range are often reflective of more complex dynamic processes such as conformational changes, domain motions, folding and ligand binding. These motions can be investigated using relaxation dispersion experiments [291, 297, 298]. Several examples demonstrating the use of these techniques to study RNA are available [286, 292, 297, 299]. Slow processes on the millisecond to second timescale can be measured using ZZ-exchange NMR experiments [300]. Rates for processes that occur on even longer timescales can be measured by time-resolved NMR methods [288, 301, 302].
Figure 3-5. NMR measurements used to study picosecond (ps) to second (s) RNA dynamics. (Adapted from [287])
The study of RNA dynamics by NMR has been expanded by the measurement and analysis of RDCs. RDCs report changes in bond orientation relative to the magnetic field resulting from internal and overall molecular motions [269] (Figure 3-4). These couplings report on the weighted average of all conformations sampled by a macromolecule over the course of picoseconds to milliseconds. This time range encompasses both local motions as well as large-scale conformation changes associated with RNA domain motions (Figure 3-5).

A-form helices are the predominate building blocks of RNA structures. Because they are semi-rigid and highly regular, their overall geometries are independent of sequence as long as base-pairing is maintained [303]. Therefore, a large, complex RNA structure can be thought of as a combination of A-form helical domains, which can be directly linked or separated by single-stranded regions that form loops, turns, bulges, and linkers [304, 305]. These elements allow RNA to sample a wide range of conformations. NMR provides a unique way to quantitatively characterize domain motion and orientation using RDCs [259, 306]. Domain motion and orientation can be efficiently examined using RDCs by solving the order tensor solution [307], which describes the average orientation of each domain relative to the magnetic field and provides information on direction and amplitude of motion.

Investigation of RNA dynamics using order tensor analysis is efficient because solution structures and complete resonance assignments are not required. Order tensor analysis requires a minimum of 5 non-parallel RDCs for each helical domain (excluding RDCs from terminal base-pairs, which are dynamic due to fraying). Because helices are highly regular structures, they can be accurately modeled using RNA modeling programs [307]. Practically, a set of ≥ 11 RDCs is needed for a well-determined order tensor solution. Consider an RNA molecule with two helices separated by an asymmetrical bulge [307]. To determine the interhelical bend angle and dynamics between the two helical domains, order tensor solutions are determined for each helix independently. A $^{13}$C, $^{15}$N labeled RNA sample is used to
measure aromatic (CH), ribose (C1'H1', C2'H2', C3'H3', and C4'H4'), and imino (NH) dipolar couplings in isotropic and partially aligned conditions. The difference in dipolar coupling in the partially aligned and isotropic conditions yields the RDC measurement. The dipolar coupling is manifested as a measurable peak splitting (Hz), which should be measured in duplicate along the $^{13}\text{C}/^{15}\text{N}$ and $^1\text{H}$ dimensions to assess the RDC error. The order tensor solution for each helix can be computed using software such as RAMAH [308]. The Euler angles from the order tensor solution are used to rotate model A-form helices into their principle axis system (PAS). In their PAS, the helices are connected according to their covalent bonding and the interhelical bend angle can be calculated using the Euler-RNA software [307]. RNA dynamics are encoded by the general degree of order (GDO) and $\eta$ terms in the order tensor solution [259]. The GDO describes the structural rigidity and amplitude of helical motions, while $\eta$ describes the directionality of those motions.

The method described is limited by the $4^{n-1}$ degeneracy of the order tensor solution [309], where $n$ is the number of independent rigid helical domains that need to be oriented to each other. While this degeneracy can sometimes be partially overcome by covalent connectivity, steric clash or experimental restraints [307], it may be challenging to determine the correct orientation for RNAs with more than two domains [310]. Additionally, domain motions can occur on the same timescale as overall rotational motions [311], which convolutes dynamics analysis. One way to decouple domain motions from the overall reorientation of the molecule is through RNA helical extension [306] such that the domain motions occur on a different timescale from the overall rate of molecular tumbling.

The dynamics of the HIV-1 transactivation response (TAR) RNA [312-315] has been extensively studied using RDCs. Located in the 5' end of all HIV-1 pre-mRNA transcripts [316], the TAR hairpin regulates viral replication [317] through its interaction with the trans-activator protein (Tat) [318]. Al-Hashimi et al. demonstrated that in low salt conditions TAR has an
average inter-helical bend angle of 47° and moves with isotropic motions sampling all positions within a cone radius angle of 46° [312]. In the presence of magnesium chloride and increasing concentrations of sodium chloride, TAR’s dynamic motions are quenched and the RNA linearized [313, 315].

For the HIV-1 TAR RNA, using elongated-helical domains to study RNA dynamics yielded residue-specific measurements of internal motions in TAR that were not apparent using the smaller, non-elongated RNA [314]. Molecular dynamics and order tensor analysis were further combined to trace out the entire helical trajectory for TAR where the helices bend and twist in a correlated manner [319]. These studies of TAR demonstrate the utility of using NMR to study RNA dynamics.
3.3 RNA structure analysis by small angle X-ray scattering (SAXS)

Many RNA structures important for translational control of gene expression are large by NMR standards, making structure determination quite challenging and expensive. Fortunately, small angle X-ray scattering (SAXS) is a complementary technique for extracting global shape information about macromolecules in solution. SAXS also provides a method for monitoring the kinetics of macromolecular folding in real time, and can be used to assess and compare the solution state conformation of molecules to structures determined by X-ray crystallography and cryo-electron microscopy. Because SAXS alone provides low resolution structural information, the combination of SAXS with an NMR based approach is ideal for characterization of large RNA molecules in solution. SAXS experiments can be performed at much lower concentrations than NMR, requiring significantly less sample. Because data collection takes only milliseconds with synchrotron X-ray sources, SAXS is also a valuable method for characterization of folding kinetics and RNA dynamics. While characterization of large macromolecules by NMR is often difficult or intractable, SAXS effectively has no upper size limit. This section will address sample requirements, analysis of SAXS data, modeling of RNA structures based on SAXS and implementation of a combined NMR/SAXS approach.

3.3.1 Sample requirements for SAXS

Since all species in solution contribute to X-ray scattering, the sample must be pure. Typically, sample concentrations range from 0.1 to 3 mg/ml. RNA samples at concentrations of 3 mg/ml or higher may exhibit structure factors related to repulsion that can interfere with accurate measurement of molecular size; however, higher salt concentrations can be used to reduce these effects. Additionally, the buffer scattering must be subtracted and buffer matching is crucial and can be achieved by extensive dialysis.
SAXS measures the simultaneous scattering of all orientations of molecules in solution, which is represented by the scattering profile, $I(q)$ (Figure 3-6A). Species that are present even at low concentrations contribute to X-ray scattering, such that contamination from formation of higher order complexes or misfolded species can greatly interfere with obtaining high quality data. Misfolded species can be eliminated through the use of size exclusion purification immediately before data collection [320].

Two basic types of X-ray sources are available for collection of SAXS data: synchrotron radiation and bench-top sources. While synchrotron radiation provides a higher intensity of X-rays and thus reduces data collection time, it can also increase the risk of radiation damage to the sample. To minimize radiation damage, the sample can be flowed back and forth across the beam during data collection. Secondary radiation damage can also occur due to X-ray induced generation of hydroxyl radicals, which can promote RNA hydrolysis. To minimize this, tris buffer (~20 mM) can be included as an effective hydroxyl radical scavenger. Collection of SAXS data using a bench-top instrument requires a smaller sample volume (as little as 30 μl); however, the low intensity of the X-rays requires a much longer data collection time. While it is tempting to overcome this problem by using a higher concentration of RNA, this may result in interparticle repulsion as described above [321].
Figure 3-6. Small angle X-ray scattering (SAXS) data.

A. X-ray scattering intensity (I) vs. scattering angle (q). Arrows point to regions of plot that correspond to structural information. B. A Guinier plot, ln(I) vs. q², provides information about the size of the molecule and the quality of the sample. This plot should be linear at small values of q (qmax*Rg < 1.2). A sharp drop-off in the plot is indicative of repulsion while an upward curve is indicative of aggregation. Dashed line indicates data extrapolation, based on the Guinier equation, to q=0 (arrow).
3.3.2 Interpreting SAXS Data

The SAXS profile, I(q), results from observation of X-ray scattering from all orientations of the molecule in solution [322]. The resulting 2D scattering pattern can therefore be radially averaged and converted into a 1D scattering curve to maximize the amount of signal obtained from a given experiment. The scattering curve contains valuable information about the dimensions, volume and fold of an RNA molecule (Figure 3-6A). The scattering angle is expressed in reciprocal space as a function of q [323]:

\[ q = \frac{4\pi \sin \theta}{\lambda} \]  

(2)

where \( \theta \) is one-half the scattering angle from the incident beam and \( \lambda \) is the wavelength of the X-ray radiation. The maximum intensity of the scattering curve is dependent on the source and type of detector and is therefore frequently normalized to 1 to allow for comparison of different data sets. The smallest angles provide information about the size of the molecule (Figure 3-6A). For most molecules this falls in the range of \( q < 0.05 \text{ Å}^{-1} \). Scattering in the range of \( q < 0.3 \text{ Å}^{-1} \) contains information about the shape of the molecule. Peaks observed in the range of \( 0.3 < q < 1.0 \text{ Å}^{-1} \) arise from internal secondary structure within the molecule (Figure 3-6A). Unfortunately, due to the fact that the molecules in solution are randomly oriented, high resolution information cannot be extracted from this region for the purposes of structure analysis.

The radius of gyration (\( R_g \)) is the root mean square distance of electron density from the center of mass and provides an accurate measure of size and shape that is useful for comparison between different samples. \( R_g \) can be estimated using a Guinier transform which exploits an approximately linear relationship between \( \ln I(q) \) and \( q^2 \) at low values of q [323]:

\[ \ln I(q) = I(0) - \frac{1}{3}R_gq^2 \]  

(3)
where \( I(0) \) is the intensity of scattering at \( q=0 \). \( I(0) \) is related to the molecular weight of the molecule and should not change with sample concentration. The data range for this method should be limited such that \( q_{\text{max}}^*R_g < 1.2 \), where \( q_{\text{max}} \) is the maximum \( q \) value. The Guinier plot also provides information about sample quality (Figure 3-6B). Deviations from linearity indicate possible interparticle interactions such as aggregation or, more commonly with nucleic acids, repulsion (Figure 3-6B). Changes in \( R_g \) have been employed to measure both kinetics of binding and affinity in ribozyme systems [324]. Frequently, this type of analysis is adequate for studying global tertiary collapse during RNA folding [325] and has been used to characterize distinct stages during folding of RNase P [326].

Another transform of the scattering profile, the Kratky plot (Figure 3-7), is useful for inferring qualitative information about the level of structure in a molecule. Molecules that are completely unfolded exhibit a very different profile from molecules with both extensive secondary and tertiary structure (Figure 3-7) [327]. These profiles are distinct from a “random” fold, in which helices are randomly oriented in relationship to one another, or a non-globular or “extended” fold where the RNA helices are rigidly extended away from one another (Figure 3-7) [327]. Kratky analysis of time resolved SAXS experiments has been used to assess the kinetics of folding of RNase P [326] and the Tetrahymena ribozyme [328], as well as the contribution of cation concentration to the extent of collapse within the glycine riboswitch [329].

The pair distance distribution function (PDDF) or \( p(r) \) reflects structural features of the molecule (Figure 3-8) and is calculated via an indirect Fourier transform that can be performed using the GNOM program [330]. The PDDF represents the real space distribution of interatomic vectors within the RNA that can be thought of as the probability of finding two atoms a given distance \( (r) \) apart. \( R_g \) and \( I(0) \) are more accurately determined from the PDDF than from the Guinier transform [330]. Additionally, the PDDF provides an estimate of the
maximum dimension ($D_{\text{max}}$) of the molecule; however, the value of $D_{\text{max}}$ cannot be determined with the same level of accuracy as $R_g$, and is therefore not as widely employed.
Figure 3-7. The Kratky plot, $q^2I$ vs. $q$, indicates the extent of folding within an RNA molecule.

Molecules with an extensive tertiary fold (collapsed) result in a different profile than non-globular molecules (extended) or unfolded molecules. (Adapted from [327]).
The pair distance distribution function (PDDF) or $p(r)$ reflects the shape of the molecule.

The PDDF can be thought of as the probability of finding two atoms in the molecule a given distance ($r$) apart. (Adapted from [323]).
3.3.3 Ab initio modeling of molecular shapes from SAXS data

The information provided by the PDDF can be utilized to generate a low resolution envelope of an RNA using ab initio modeling techniques. Currently, DAMMIN [331] and its more recent implementation, DAMMIF [332], are the best software options for ab initio modeling of nucleic acid molecules. Both of these programs employ a set of dummy atoms or beads to simulate the general shape of the molecule. The predicted scattering amplitude of the dummy atom model is calculated and compared to the experimental scattering data. This process is repeated iteratively until a good fit is achieved. Typically only data in the range of $q < 0.33 \, \text{Å}^{-1}$ are used with this method for the reasons described in section 3.3.1.

Due to the inherently low resolution associated with ab initio modeling from SAXS data, the arrangement of domains in a molecule cannot always be unambiguously determined using this method. For RNA, helical extensions have been employed in a variety of studies to assign structural features within a low resolution structure [333]. Also, because these programs use the output of GNOM, it is essential that the parameters employed to generate the PDDF (e.g. $D_{\text{max}}$) are optimized before calculating an envelope. Typically, several envelope models are generated (>10) and then compared with each other to gauge the quality of the model. This can be achieved using the DAMAVER software package [334], which provides a measure of agreement between the molecules, the normalized spatial discrepancy (NSD). The NSD should fall between 0.7-0.9 for a unique, well-determined model [334].

3.3.4 All-atom molecular modeling

Because of the difficulty of inferring structural details about the position of molecular features from a low-resolution envelope, a preferable method for modeling molecular structure employs the use of all-atom models that have been selected for and/or refined against the scattering data [335]. Detailed structural information cannot be directly derived from the
observed scattering profile; therefore, selection of the best models is dependent upon accurate back-prediction of the scattering profile for each model.

If an existing structure is available for a related molecule, the best modeling method is homology modeling. This can be achieved using non-bonded non-crystallographic symmetry (NCS) restraints in XPLOR-NIH, as in the case of the structure of tRNA\textsuperscript{Val} [336] (Figure 3-9). A model of tRNA\textsuperscript{Val} was first built by constraining the model based on the crystal structure of tRNA\textsuperscript{Phe}, and then refined against SAXS (Figure 3-9A) and RDC measurements in XPLOR-NIH, resulting in a structural model that provided information about tRNA\textsuperscript{Val} (Figure 3-9B).

For RNA molecules, the most straightforward method in the absence of a homology model is the MC-Fold/MC-Sym pipeline [337]. Some models predicted using this software can deviate by as little as 2-4 Å from the experimentally determined structure [337]. Accuracy of models can be judged in part by observation of base-pairing, either by biochemical methods or NMR (see section 3.2.1). MC-Sym employs a library of structures from the protein data bank (PDB) to create mosaic models that include small fragments from matching sequences in the library. For small helical RNA structures, this modeling method can be surprisingly accurate (Figure 3-11) [338]. This method has been successfully employed with the TPP riboswitch to create all atom models that are consistent with SAXS measurements [333]. Another suite of software, the Simbios NAST/C2A package, allows for coarse grain modeling of RNA molecules and then conversion of the coarse grain models to all-atom models [339].

To determine which molecular models are consistent with the experimental data, the scattering profile for each model must be accurately predicted. Back-calculated scattering amplitudes for all-atom models can be determined using the Debye equation [340]:

$$ I(q) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(q) f_j(q) \frac{\sin(qd_{ij})}{qd_{ij}} $$

(4)
where \( d_{ij} \) is the distance between atoms i and j and \( N \) is the number of atoms in the molecule. Most available software uses some approximation to limit the computational time necessary. CRYSOL [341] and the FOXS web server [342] employ coarse grain approximations that create a bead for each residue in the molecule to obtain the \( \chi^2 \) goodness of fit between the predicted and experimental scattering amplitudes. CRYSOL provides not only the predicted scattering amplitudes but also the predicted PDDF and \( R_g \). FOXS is a higher throughput method and can calculate predicted scattering based on form factors for each atom type in the PDB file, making it a useful tool for nucleic acids. Fast-SAXS-RNA software [343] is fine-tuned specifically for use with nucleic acids and uses two form factor approximations per residue (one for the sugar moiety and one for the base). Model building can also be used in conjunction with the integrative model platform (IMP) [344], which refines structures using \( \chi^2 \) minimization. IMP employs rigid body treatment of user specified domains to continuously adjust the shape of the molecule until agreement with the experimental scattering curve is optimized.

Finally, an exciting new direction is available for analyzing inherently flexible RNA molecules and RNA-protein complexes in solution using SAXS. Two independently developed ensemble modeling methods, Ensemble Optimization Method (EOM) [345] and Minimal Ensemble Search (MES) [346], employ either a library of conformations or high temperature molecular dynamics simulations, respectively, to sample conformational space. From the generated conformations, a set that best represents the scattering profile is selected and can be subjected to energy minimization. These approaches help to define the conformational space sampled by flexible domains in a given molecule [346]. MES can also be used in conjunction with the FOXS server [342] if a large variety of structural models have already been generated.
Figure 3-9. Homology modeling and refinement of tRNA\textsuperscript{Val} using SAXS and RDC measurements.

(A) Predicted scattering curves of tRNA structures. (B) Comparison of the tRNA\textsuperscript{Phe} crystal structure (PDB ID 1TRA) with the refined tRNA\textsuperscript{Val} homology model (PDB ID 2K4C).
Figure 3-10. Modeling the structure of an RNA with MC-Sym.

Model of a small (43 nucleotide) tetraloop receptor RNA generated with the MC-Fold/MC-Sym pipeline (gray), which has a 4.1 Å RMSD from the NMR structure (red) [41].
3.3.5 Combining SAXS and NMR

Incorporation of both SAXS and NMR data can be very powerful for analyzing the structures of RNA molecules in solution. NMR structures are often underdetermined, in part because the data are often insufficient to accurately define the global molecular shape. SAXS provides low-resolution global information without any detailed information about interatomic distances. Thus, the combination of these two techniques is ideal for analyzing molecular structures in solution [347].

SAXS data can be used to further refine the overall size and shape of an already determined NMR structure [310, 348]. We have used this method to solve the structure of an 86 nt RNA/RNA complex in solution (Figure 3-11). Refinement with SAXS data improved the fit of the structure to both the measured \( R_g \) [310] and the calculated ab initio structure (Figure 3-11) (unpublished data).
Figure 3-11. NMR and SAXS refinement of an RNA:RNA complex.

Comparison of the *ab initio* structure (gray spheres) with the NMR structure of the complex refined in the absence (red) and presence (blue) of SAXS restraints. Inclusion of SAXS restraints results in better agreement with measured $R_g$ [310] and a better fit with the *ab initio* structure (unpublished data).
3.4 Using NMR and SAXS to study RNAs involved in translational control

RNA structural elements, such as internal ribosomal entry sites (IRESs) [349] and programmed ribosomal frameshift sites (PRFS) [86, 148], are widely use in biology to regulate translation. IRESs facilitate cap-independent translation [349, 350], range in size from 250 to 500 nucleotides, and are typically composed of several stable domains, including stem-loops and pseudoknots. PRFS, which increase genomic coding capacity by promoting a change in reading frame during translation, are composed of three essential elements: a heptanucleotide ‘slippery’ sequence, a linker region, and a downstream RNA structure, such as a pseudoknot or stem-loop [101]. While biochemical and genetic results highlight the functional importance of IRESs and PRFS, structural information is needed to understand the roles of structure and dynamics in the function of these RNAs. The following sections highlight several examples of RNAs involved in translation control that have been characterized using NMR and SAXS.

3.4.1 Programmed ribosomal frameshift sites

NMR has been employed to investigate the downstream RNA structures of numerous frameshift sites [108, 109, 187, 199, 351-354]. The most common type of frameshift is a -1 PRF, which promotes a one nucleotide shift of the translating ribosome in the 5’ direction. The downstream stimulatory structure for -1 PRF sites is most commonly an H-type pseudoknot [101], a stem-loop in which nucleotides from the 3’ strand fold back to base-pair with the loop, creating a second helical stem [355]. The first structure of a functional pseudoknot was determined with NMR [356]. A large bend and tertiary interactions in this structure suggested multiple possible mechanisms for frameshift enhancement.

NMR has been further employed to examine the differences in RNA structure and dynamics between frameshift stimulating (S) and non-stimulating (NS) H-type pseudoknots [351]. Identical stretches of sequence in S and NS pseudoknots have excellent 1H and 15N chemical
shift agreement, suggesting they have extremely similar structures. As a result, the characteristic bend observed for these pseudoknots cannot explain frameshift stimulation. Further, NMR relaxation rates for imino $^{15}$N nuclei within the helical regions for the S and NS pseudoknots did not reveal complex dynamics or striking differences between the two types.

While most frameshift sites contain a pseudoknot structure, structural studies using NMR showed that the structure in the human immunodeficiency virus type-1 (HIV-1) PRFS is an extended stem-loop [108, 187]. The HIV-1 PRFS RNA structure has an extremely stable upper stem-loop separated from a meta-stable lower helix by a dynamic three-purine bulge. While the lower helix must be unwound prior to frameshifting, the overall bend adopted by the RNA structure may be important for frameshift stimulation for reasons that are not yet entirely clear [108, 186]. In HIV-1, the frameshifting efficiency controls the ratio of structural to enzymatic proteins that is critical for virion infectivity [127]. Therefore, small molecules that target the HIV PRFS stem-loop and perturb its stability or structure may be able to attenuate viral replication.

The conserved three-purine bulge in the HIV-1 PRFS structure served as a target for a high-throughput screen for small molecules [199]. Compound binding was verified by NMR. Although the compound with the highest affinity only had a modest impact on frameshifting efficiency, this study demonstrates the utility of NMR to screen, validate and map RNA-small molecule interactions that have the potential to modulate -1 PRF during translation.

NMR revealed that the related simian immunodeficiency virus (SIV) PRFS also contained a stem-loop structure instead of a previously proposed pseudoknot structure [109]. Three RNA constructs of various lengths were examined by NMR. While two of the three constructs were designed to allow pseudoknot formation, the shortest construct contained only a stem-loop (Figure 3-2A). Chemical shift mapping indicated that all three constructs contained a stem-loop structure, and that a pseudoknot does not form. NMR structure determination revealed that the
SIV frameshift stem-loop forms two G-C base-pairs across the loop, which preclude pseudoknot formation [109].

3.4.2 Untranslated regions

The 5’ and 3’ untranslated regions (UTRs) of mRNA flank the protein coding region and are often highly structured. Many of the structures in these elements play a large role in translational regulation. Furthermore, the RNA structures found in UTRs are highly diverse, and can utilize very different strategies to control translation. This section will review the application of NMR and SAXS to study four different types of structures found in UTRs: an internal ribosome entry site (IRES), a cap-independent translation element, an “RNA thermometer”, and an adenine sensing riboswitch.

Internal Ribosome Entry Sites (IRESs)

Due to the inherently large size of IRESs, NMR investigations of their structure and function have focused on individual IRES domains. For example, NMR has been employed to investigate an IRES domain in hepatitis C viral (HCV) RNA [271, 357]. Translation of the HCV mRNA is initiated in the 5’ UTR on a highly structured IRES. While the 100 kDa IRES is made up of multiple domains, the 77 nt domain II is required for IRES activity [358] and contacts the 40S head region of the ribosome [358]. Lukavsky et al. demonstrated that the full length domain II folds independently in the context of the 100 kDa IRES using segmental $^{15}$N labeling and chemical shift mapping [271]. The NMR structure of domain II was solved using a divide and conquer approach, in which domain II of the HCV IRES was divided into two manageable pieces of 34 nt (domain IIb) and 55 nt (domain IIa). Local NOE distance restraints were combined with RDC data from the full-length construct to determine the solution structure of domain II. The structure adopts a bent L-shape stabilized by the internal stacking of 5 nts in the asymmetric bulge found in domain IIa.
The 90° bend in domain IIa of the HCV IRES appears to be critical for IRES function [357]. Several small molecules that bind to domain IIa with low micromolar affinity and inhibit HCV viral replication have been reported [359]. The compounds have binding sites that localized to the 5 nt bulge. The structure of domain IIa in complex with one such inhibitor, Isis-11, was solved by NMR [357]. Interestingly, compound binding trapped a nearly linear orientation of the RNA, eliminating the 90° bend observed in the free RNA. A structural understanding of this complex at the atomic level may facilitate structure-based drug design for future classes of improved antiviral inhibitors.

Ribosome-binding structural elements (RBSEs)

The global structure of a 102 nt ribosome-binding structural element (RBSE) found within the 3′ UTR of the Turnip Crinkle Virus (TCV) genomic RNA was characterized through a combination of NMR and SAXS [273] (Figure 3-12A). The RBSE assists in recruitment of the large ribosomal subunit and contains a large asymmetric loop that is essential for the coordination of translation of the viral proteins with genome replication. This study utilized the G2G software package that first models the individual A-form helices in the RNA, then aligns the helices by fitting them to the RDC data [347]. Linker sequences are then modeled into the aligned helices to create a complete model. The resulting model was iteratively refined against both SAXS and NMR data in XPLOR-NIH (Figure 3-12B). Interestingly, the RBSE RNA mimics a tRNA molecule in overall shape (Figure 3-12C), despite the fact that the RBSE secondary structure is quite distinct from the typical cloverleaf tRNA secondary structure (Figure 3-12A) [273]. Thus, the TCV RBSE likely binds to the ribosome via its ability to mimic a cellular tRNA.
Figure 3-12. Solution structure of the TCV RBSE [273].

(A) Secondary structure of the RBSE. Long-range base-pairing interactions are indicated by dashed lines. (B) The solution structure of the TCV RBSE (PDB ID 1KRL) as determined by a combination of SAXS and NMR reveals a twisted T-shape. (C) Despite the lack of similarity in secondary structure, the TCV RBSE (purple) resembles a tRNA (gray) in its 3-dimensional fold.
RNA thermometers

RNA thermometers regulate translation initiation through temperature-dependent RNA structure formation. Rinnenthal et al. have investigated a stem-loop structure found in the 5′ UTR of the Salmonella fourU RNA that base-pairs to the Shine-Dalgarno sequence at low temperatures, preventing ribosome recruitment [360]. At high temperatures, the structure unfolds, exposing the Shine-Dalgarno sequence and allowing translation to occur. To investigate the temperature dependent dynamics of this structure, NMR was used to measure the thermodynamic stability of individual base-pairs [360]. Thermodynamic stability of the base-pairs was quantified by measuring the imino $^1$H exchange rates along a 5°C to 50°C temperature gradient. Based on the measured imino $^1$H exchange rates, $\Delta G$, $\Delta H$, and $\Delta S$ were calculated for each base-pair. This information is consistent with a zipper-type of unfolding mechanism for this RNA thermometer [360]. Surprisingly, mutations were found to have long-range effects on base-pair stability over a distance of 6 base-pairs, an effect which may be transduced via perturbations of the RNA hydration shell [360].

Riboswitches

Riboswitches typically reside in the untranslated 5′ UTR regions of bacterial mRNA [361]. These regulatory RNA elements either increase or decrease expression of the downstream gene as a result of ligand binding [362]. Riboswitches are composed of two structural elements: an aptamer domain, which binds the ligand (usually a metabolite), and an expression platform, which is linked to the aptamer domain and undergoes a conformational change in response to ligand binding. Multiple crystal structures have been solved of various riboswitch aptamer domains bound to their ligands [363-365]. While the crystal structures reveal how ligands are bound, they do not reveal the dynamic interplay between ligand binding and the resulting conformational change in the expression platform. NMR has proved useful for characterization of purine riboswitch folding [366-368]. Recently, the folding of an adenine sensing riboswitch, which regulates synthesis of enzymes responsible for purine synthesis, was followed for three
minutes using time-resolved NMR [302]. Ultrafast acquisition of $^1$H-$^{15}$N 2D HSQC spectra revealed that after the addition of the adenine metabolite and magnesium to the free RNA solution, the core ligand-bound structure formed in the first 20 seconds. Within 30 seconds of core formation, tertiary interactions between two helical loops were stabilized; however stabilization of all anticipated base-pairs lasted up to 3 minutes. The sensitivity of SAXS to overall molecular shape also renders this technique appropriate for characterization of riboswitches [320, 329, 369-372], which may have multiple conformations in their unbound states.
3.5 Summary

Only 2% of the human genome is translated into protein, whereas more than 80% is transcribed into RNA [373]. Given the abundance of RNA in human biology, it is clear that RNA structural studies are going to be important for many years. As more RNAs are investigated, more functions will be discovered, including new strategies for regulating translation. Both NMR and SAXS methods are continuing to improve, and we are now witnessing the integration of these methods to provide a more detailed and comprehensive view of RNA structure, function and dynamics. In the future, we can expect that NMR and SAXS will play significant roles in advancing our understanding of translational control.
Chapter 4: Domain motions in the HIV-1 frameshift site RNA

The material presented in this chapter is a result of research performed by the author with contributions from Elizabeth A. Dethoff, Marco Tonelli, Prof. Hashim Al-Hashimi and Prof. Samuel E. Butcher.

E.A.D. contributed to RDC data analysis.

M.T. assisted with NMR data collection.

The remainder of the work was by K.D.M.
4.1 Abstract

The highly conserved human immunodeficiency virus type I (HIV-1) frameshift site RNA folds into a stable structure that is important for viral replication. We used NMR and order tensor analysis to investigate the interhelical motions of this RNA. The degree of base stacking was probed at each position in a conserved three-purine bulge by incorporating the fluorescent purine analogue 2-aminopurine. The RNA was studied under different ionic conditions (20 mM vs. 150 mM potassium, both with and without 2 mM magnesium). In potassium alone, the RNA adopts a ~43° interhelical bend and displays anisotropic interhelical motions. These motions are quenched by 2 mM magnesium, which promotes a near-coaxial conformation of the two helices. The coaxial helical conformation is accommodated by an extrahelical nucleotide in the bulge. Comparison of our results to previous studies suggests that the impact of monovalent and divalent ions on RNA interhelical dynamics is general and independent of sequence. In contrast, the extent of the coaxial alignment of helices induced by magnesium appears to be dependent upon the linker sequence.
4.2 Introduction

Complex RNA structures are composed of rigid A-form helical domains [303] connected by internal loops, bulges, single stranded regions and non-helical motifs [374]. It is becoming increasingly clear that RNA structures exist as an ensemble of conformations that are sampled through inter-domain motions [199, 220, 259, 314, 375, 376] on the nanosecond to millisecond time-scale [267, 269]. Individual nucleotides also experience local librations on the picosecond scale, which can be coupled to longer time-scale domain motions [314], adding another dimension to an already complex conformational trajectory.

RNA interhelical domain motions are controlled by the flexibility of the connecting motifs. These motions increase the functional potential of a single RNA sequence [259, 377, 378]. For example, large-scale interhelical reorientations facilitated by flexible junctions modulate the actions of many RNAs [362, 377]. Additionally, several RNAs undergo significant conformational changes in response to proteins or small molecule binding [379-382], which is the basis for riboswitch regulation of gene expression [302, 383, 384]. Finally, RNA domain motions may be important for protein recognition via conformational capture [385, 386].

Despite the enormous size and diversity of RNA transcriptomes [387-389], most of what is known about RNA dynamics has been derived from a relatively small collection of RNAs [298, 299, 386, 390-399]. Among these, the HIV-1 transactivation response (TAR) RNA element has been studied most extensively and has served as a model for investigating RNA motional amplitudes and dynamics with NMR [220, 286, 306, 311-315, 319, 375, 385, 391, 400-403]. TAR contains two helices separated by a flexible three-pyrimidine (UCU) bulge. In low-ionic strength, TAR has an interhelical bend angle of $47^\circ\pm5^\circ$ [400], is flexible [312], and samples a wide range of conformations in its topologically accessible ensemble [220, 319, 403] through highly correlated domain motions [319]. The addition of magnesium or high concentrations of
monovalent cations leads to quenching of TAR’s dynamics and promotes coaxial stacking of its helices [313, 315, 400]. Additionally, several ligand-bound conformations of TAR arrest global motions and trap the RNA in an orientation selected from its ensemble [312, 319, 385, 391].

It is not yet clear whether the observed large amplitude dynamic motions of TAR [312, 391] and their modulation by counter-ion and ligand binding [259, 313, 315, 400, 401] are general or are somehow specific to its UCU bulge. A previous study of RNA helices connected by A and U bulges of varying lengths revealed that magnesium ions affected the observed bend angle differentially for poly-U vs. poly-A bulges [404]. In contrast, magnesium was observed to have no impact on the average interhelical bend angle or dynamics of the RNase P P4 domain, which has a single uridine bulge [405].

RNA bulges are an extremely common motif [304, 374, 406, 407]. For dinucleotide and trinucleotide bulges in the PDB, the observed helical orientations are limited to less than 5% of predicted available conformational space [220]. Although this seems fairly restrictive, the number of conformations sampled within this range can be quite large. How much of this space is sampled is likely dependent on steric restraints encoded by the RNA [220, 375, 403, 408] and may be dependent on linker sequence and highly correlated bending and twisting motions [391]. The sequence of the bulge nucleotides, as well as the flanking base-pairs, may have a large effect on RNA dynamics [220, 375, 403].

The thermodynamic stability of bulges may impact RNA interhelical motions. While the free energies for RNA dinucleotide steps within duplex RNAs have been determined [202, 409, 410] and depends upon nearest neighbor interactions, the thermodynamic impact of single nucleotide bulges depend upon non-nearest neighbor interactions and are complex [411-414]. The thermodynamic contribution of single nucleotide stacking is always small for 5’ ends and depends on sequence for 3’ ends, but in general, 3’ purines are more stabilizing than 3’
pyrimidines [415, 416]. Therefore, purine bulges are expected to have higher stacking propensities than pyrimidine bulges, and these differences may significantly influence interhelical dynamics. A molecular dynamics study estimated that the unstacking of a single adenosine bulge nucleotide from within an RNA helix costs an additional ~1.5 kcal/mol of free energy relative to flipping a uridine nucleotide into solution [417]. To our knowledge, the thermodynamics of di- and trinucleotide bulges, and their impact on interhelical domain motions, have yet to be systematically investigated.

NMR is an ideal method for investigating RNA dynamics over a wide range of time scales [376, 418]. Dynamic motions occurring over the course of picoseconds to milliseconds can be studied by the measurement of residual dipolar couplings (RDCs) [259, 269, 309, 419]. These couplings report angular information on the weighted average of all conformations sampled by a macromolecule. The long-range information derived from RDCs can be used to efficiently quantify helical orientation and dynamics in a variety of solution conditions [376]. However, the coupling of internal and overall domain motions can complicate the interpretation of RDCs [314, 420, 421]. This situation becomes most problematic when internal motions alter the overall molecular alignment such that the ordering of both domains is indistinguishable, which interferes with the quantification of interhelical motions [307, 311]. RDC measurements on artificially elongated helices have proven exceptional at decoupling overall motions from internal motions because the overall alignment of an extended domain is less sensitive to internal motions [306, 311, 314, 391]. However, unnaturally long helical extensions may not always be necessary, as even moderate differences in helical lengths can be sufficient to uncouple these motions [311, 422].

Here, we investigate the dynamics of the HIV-1 frameshift site RNA [99, 100, 108, 187, 199]. This RNA has been shown to fold into an extremely stable upper helix separated from a less
stable lower helix by a GGA bulge [108, 186, 187] (Figure 4-1A). NMR and fluorescence spectroscopy were employed to examine conformational dynamics as a function of potassium and magnesium cation concentrations. We find that, in potassium alone, the RNA adopts an average interhelical bend of ~43° and experiences large amplitude domain motions. These dynamics are largely quenched upon addition of 2 mM magnesium. In comparison with previous studies [315, 400], our results suggest that the impact of potassium and magnesium on RNA dynamics is sequence independent. We show that the coaxial alignment of helices in 2 mM magnesium is facilitated by the unstacking of the central nucleotide in the purine bulge. Finally, the extent of coaxial stacking in magnesium appears to be linker sequence dependent.
Figure 4-1 RNA constructs utilized for investigation of HIV-1 FS-SL dynamics.
A) The HIV-1 FS-SL solution structure (PDB ID 1Z2J) and consensus sequence are shown for reference. B) Secondary structures of the modified FS-SL used for RDC collection (NMR construct). Sequence modifications are bold and italicized. C) The three 2-aminopurine (2AP) containing constructs used to study positional stacking within the bulge.
4.3 Materials and Methods

4.3.1 RNA synthesis and purification

The consensus sequence for the lower helix of the HIV-1 FS-SL (Figure 4-1A) was modified and truncated (Figure 4-1C) for optimal alignment of the longer, upper helix in Pf1 phage. The synthetic oligonucleotide (5’-TTCTAATACGACTCACTATAGGCGATCTGGCCTTCCCACAA GGGAAGGCCAGGGAATCGCC-3’) and its complement were purchased (Integrated DNA Technologies (IDT), Inc.) and utilized as a template for in vitro transcription.

RNA for NMR was transcribed in vitro using purified His6-tagged T7 RNA polymerase and synthetic DNA oligonucleotides (IDT), as previously described [108, 189, 199]. 13C/15N-labeled RNA samples were prepared using 13C/15N-labelled rNTPs (Cambridge Isotope Laboratories). RNA was purified using 12.5% denaturing polyacrylamide gel electrophoresis, identified by UV absorbance, and excised from the gel. RNA was recovered by diffusion into 0.3 M sodium acetate, precipitated with ethanol, purified on a High-Q anion exchange column (Bio-Rad), again precipitated with ethanol, and desalted on a sephadex G-15 (Sigma) gel filtration column. The purified RNA was lyophilized, resuspended in 20 mM KH2PO4 (pH 6.8) and exchanged into each solution by dialysis against 2L of buffer at 4° C. Partial alignment of RNA for RDC measurements was achieved by adding Pf1 filamentous bacteriophage (ASLA Ltd., Riga, Latvia) at a final concentration of ~10-15 mg/mL to 13C/15N-labeled samples. RNA used in fluorescence experiments was purchased from IDT.

4.3.2 NMR spectroscopy

All NMR spectra were obtained on a Varian 900 MHz or Bruker 700 MHz spectrometer at the National Magnetic Resonance Facility at Madison (NMRFAM). The spectrometers were equipped with z-axis pulsed field gradient cryogenically cooled probes. Chemical shifts were referenced to DSS by adding 2 μM sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) directly
to the samples. RDCs were measured in three different solution conditions: 20 mM potassium phosphate (the 20 mM potassium condition) supplemented with either 130 mM potassium chloride (the 150 mM potassium condition) or 2 mM excess magnesium chloride (the 20 mM potassium and 2 mM magnesium condition), all at pH 6.8. Magnesium titrations and a two-state interpretation of chemical shift (data not shown) indicate that the FS-SL (0.3 mM) is ~95% saturated in the magnesium “bound” form in the 20 mM potassium and 2 mM excess magnesium condition. 1H-13C-S3A-HSQC, 1H-13C-TROSY-HSQC, and 1H-15N-TROSY-HSQC experiments were used to collect scalar coupling values in non-aligned and partially aligned (~10-15 mg/ml buffer exchanged Pf1 filamentous bacteriophage (ASLA Ltd., Riga, Latvia)) samples. RDCs were calculated by determining the difference in scalar coupling values between non-aligned and partially aligned samples. RDC errors were calculated using the root-mean-square deviation (rmsd) for each bond type in both the direct and indirect dimensions and between duplicate experiments, as described [307].

A, U, C, and G spin-pair resonance intensities (peak heights) from 1H-13C-HSQC experiments were normalized to the intensity of the corresponding A25, U8, C16, and G27 spin-pair resonances. Such spin-pairs were chosen because they are expected to experience limited motions [376], given their location in an extremely stable upper helix [108], and their completeness in measurement across all conditions.

4.3.3 Order tensor analysis

RDCs measured in the Watson-Crick base pairs in the upper and lower helices (Tables 4-1, 4-2, and 4-3) were subject to order tensor analysis using idealized A-form helices as previously described [307] and validated [308, 312, 313, 423-425]. Upper and lower helices were constructed using Insight II (Accelrys Inc). A-form RNA parameters were checked using 3DNA [426]. Propeller twist was corrected using an in-house program as described [307].
RDCs were fit to the corrected A-form helices and order tensors were determined using RAMAH [308]. RDCs corresponding to terminal base pairs were excluded in the fits because of their departure from ideal A-form RNA characteristics [307]. To account for differences in alignment between 100% D$_2$O and 90% H$_2$O/10% D$_2$O NMR experiments, the $^1$H-$^{15}$N data collected in 90% H$_2$O/10% D$_2$O were uniformly scaled such that an optimal fit of the RDC data was achieved using RAMAH [308], as previously described [391]. Order tensor errors due to A-form structural noise and RDC error were calculated using Aform-RDC [425], with the error input being the average RDC error. Excellent fits to order tensors were achieved for RDCs measured under all conditions. In all instances, the rmsd compared favorably with the RDC measurement uncertainty (Table 4-4).

Order tensor solutions were used to rotate each A-form helix into its principal axis system (PAS) using EULER [307]. Order tensor solutions have a $4^{n-1}$ degeneracy, where $n$ is the number of helices, such that once the helix is in its PAS, rotation by 180° about each principal axis yields the same order tensor solution. To determine the correct helix orientation, two connectivity restraints were used. The first required connectivity between C7-P and U6-O3’ atoms and eliminated 2 of the 4 solutions. To satisfy this restraint, the lower helix was translated without rotation such that average A-form distances separated the C7-P and U6-O3’ atoms [427]. Specifically, the U6-O3’ atom was 1.58 Å from the C7-P atom, with a 102° O5’-P-O3’ bond angle and a 62° dihedral angle about the O5’-P bond. A final solution was selected by elimination of the model that violated the distant restraint between the G32-O3’ and A36-P atoms. This distance must be smaller than or equal to the theoretically allowed length of 21 Å [307]. Interhelical bend angles and dynamic parameters, GDO and $\eta$, were calculated in each condition as previously described [307]. To specifically compare the motion of one helix relative to the other, the general degree of order (GDO) for the dynamic helix is normalized by the GDO for the helix dominating the alignment, giving the internal GDO (GDOint).
The orientation of the upper helix Szz axis relative to the magnetic field z-axis was
determined to assess potential coupling of internal and overall motions. Here, the angular
difference was calculated by first orienting the upper helix along the z-axis of the magnetic
field. Next, RAMAH [308] was run using the upper helix RDCs and the z-axis oriented
coordinates. The resulting beta angle describes the difference between the magnetic field z-axis
and Szz, the principal component of the alignment tensor.
Table 4-1 HIV-1 FS-SL RDCs measured in 20 mM potassium cationic conditions

<table>
<thead>
<tr>
<th>Residue</th>
<th>RDC (Hz)</th>
<th>Error (Hz)</th>
<th>Residue</th>
<th>RDC (Hz)</th>
<th>Error (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>C1’H1’</td>
<td>11.31</td>
<td>0.85</td>
<td>C16</td>
<td>C5H5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.31</td>
<td>0.85</td>
<td>C16</td>
<td>27.43</td>
</tr>
<tr>
<td>G2</td>
<td>N1H1</td>
<td>-1.81</td>
<td>2.12</td>
<td>G23</td>
<td>C8H8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1.81</td>
<td>2.12</td>
<td>G23</td>
<td>31.05</td>
</tr>
<tr>
<td>C3</td>
<td>C1’H1’</td>
<td>19.07</td>
<td>0.85</td>
<td>G23</td>
<td>N1H1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19.07</td>
<td>0.85</td>
<td>G23</td>
<td>-14.69</td>
</tr>
<tr>
<td>C3</td>
<td>C6H6</td>
<td>16.65</td>
<td>3.18</td>
<td>G24</td>
<td>C8H8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16.65</td>
<td>3.18</td>
<td>G24</td>
<td>25.64</td>
</tr>
<tr>
<td>G4</td>
<td>C8H8</td>
<td>26.10</td>
<td>2.81</td>
<td>G24</td>
<td>N1H1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.10</td>
<td>2.81</td>
<td>G24</td>
<td>-12.36</td>
</tr>
<tr>
<td>G4</td>
<td>N1H1</td>
<td>-15.61</td>
<td>2.12</td>
<td>A25</td>
<td>C2H2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-15.61</td>
<td>2.12</td>
<td>A25</td>
<td>24.74</td>
</tr>
<tr>
<td>A5</td>
<td>C2H2</td>
<td>22.95</td>
<td>2.66</td>
<td>A25</td>
<td>C8H8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22.95</td>
<td>2.66</td>
<td>A25</td>
<td>22.49</td>
</tr>
<tr>
<td>A5</td>
<td>C8H8</td>
<td>28.35</td>
<td>2.81</td>
<td>A26</td>
<td>C2H2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>28.35</td>
<td>2.81</td>
<td>A26</td>
<td>15.75</td>
</tr>
<tr>
<td>A5</td>
<td>C1’H1’</td>
<td>15.56</td>
<td>0.85</td>
<td>A26</td>
<td>C8H8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.56</td>
<td>0.85</td>
<td>A26</td>
<td>15.29</td>
</tr>
<tr>
<td>U8</td>
<td>C5H5</td>
<td>37.23</td>
<td>2.19</td>
<td>G27</td>
<td>C8H8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>37.23</td>
<td>2.19</td>
<td>G27</td>
<td>22.95</td>
</tr>
<tr>
<td>G9</td>
<td>N1H1</td>
<td>-16.19</td>
<td>2.12</td>
<td>G27</td>
<td>N1H1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-16.19</td>
<td>2.12</td>
<td>G27</td>
<td>-14.49</td>
</tr>
<tr>
<td>G10</td>
<td>N1H1</td>
<td>-16.10</td>
<td>2.12</td>
<td>G28</td>
<td>C8H8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-16.10</td>
<td>2.12</td>
<td>G28</td>
<td>28.79</td>
</tr>
<tr>
<td>G10</td>
<td>C8H8</td>
<td>26.54</td>
<td>2.81</td>
<td>G28</td>
<td>N1H1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.54</td>
<td>2.81</td>
<td>G28</td>
<td>-17.61</td>
</tr>
<tr>
<td>U13</td>
<td>C5H5</td>
<td>35.19</td>
<td>2.19</td>
<td>U37</td>
<td>C5H5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35.19</td>
<td>2.19</td>
<td>U37</td>
<td>-21.90</td>
</tr>
<tr>
<td>U13</td>
<td>C1’H1’</td>
<td>-22.48</td>
<td>0.85</td>
<td>U37</td>
<td>N3H3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-22.48</td>
<td>0.85</td>
<td>U37</td>
<td>-9.60</td>
</tr>
<tr>
<td>U13</td>
<td>N3H3</td>
<td>-8.28</td>
<td>1.85</td>
<td>C38</td>
<td>C6H6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-8.28</td>
<td>1.85</td>
<td>C38</td>
<td>31.49</td>
</tr>
<tr>
<td>U14</td>
<td>C5H5</td>
<td>33.97</td>
<td>2.19</td>
<td>C38</td>
<td>C5H5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>33.97</td>
<td>2.19</td>
<td>C38</td>
<td>-8.41</td>
</tr>
<tr>
<td>U14</td>
<td>N3H3</td>
<td>-6.26</td>
<td>1.85</td>
<td>G39</td>
<td>N1H1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-6.26</td>
<td>1.85</td>
<td>G39</td>
<td>-5.58</td>
</tr>
<tr>
<td>C15</td>
<td>C5H5</td>
<td>29.69</td>
<td>2.19</td>
<td>G39</td>
<td>C1’H1’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29.69</td>
<td>2.19</td>
<td>G39</td>
<td>-17.26</td>
</tr>
<tr>
<td>C16</td>
<td>C1’H1’</td>
<td>-35.01</td>
<td>0.85</td>
<td>C40</td>
<td>C5H5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-35.01</td>
<td>0.85</td>
<td>C40</td>
<td>32.51</td>
</tr>
</tbody>
</table>
Table 4-2 HIV-1 FS-SL RDCs measured in 150 mM potassium cationic conditions

<table>
<thead>
<tr>
<th>Residue</th>
<th>RDC (Hz)</th>
<th>Error (Hz)</th>
<th>Residue</th>
<th>RDC (Hz)</th>
<th>Error (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>C8H8</td>
<td>0.68</td>
<td>G23</td>
<td>C8H8</td>
<td>30.67</td>
</tr>
<tr>
<td>G4</td>
<td>C8H8</td>
<td>23.68</td>
<td>G23</td>
<td>N1H1</td>
<td>-14.89</td>
</tr>
<tr>
<td>G4</td>
<td>N1H1</td>
<td>-16.66</td>
<td>G24</td>
<td>C8H8</td>
<td>27.30</td>
</tr>
<tr>
<td>A5</td>
<td>C2H2</td>
<td>27.31</td>
<td>A25</td>
<td>C2H2</td>
<td>26.17</td>
</tr>
<tr>
<td>A5</td>
<td>C8H8</td>
<td>30.69</td>
<td>A25</td>
<td>C8H8</td>
<td>21.43</td>
</tr>
<tr>
<td>A5</td>
<td>C1’H1’</td>
<td>16.69</td>
<td>A26</td>
<td>C2H2</td>
<td>19.40</td>
</tr>
<tr>
<td>U8</td>
<td>N3H3</td>
<td>-16.56</td>
<td>A26</td>
<td>C8H8</td>
<td>19.62</td>
</tr>
<tr>
<td>G9</td>
<td>N1H1</td>
<td>-19.07</td>
<td>A26</td>
<td>C8H8</td>
<td>19.41</td>
</tr>
<tr>
<td>G10</td>
<td>C1’H1’</td>
<td>-16.24</td>
<td>G27</td>
<td>C1’H1’</td>
<td>-46.47</td>
</tr>
<tr>
<td>U13</td>
<td>N3H3</td>
<td>-7.92</td>
<td>G27</td>
<td>C1’H1’</td>
<td>-46.47</td>
</tr>
<tr>
<td>U14</td>
<td>C6H6</td>
<td>10.60</td>
<td>U37</td>
<td>C5H5</td>
<td>-24.59</td>
</tr>
<tr>
<td>U14</td>
<td>N3H3</td>
<td>-7.60</td>
<td>U37</td>
<td>N3H3</td>
<td>-9.58</td>
</tr>
<tr>
<td>C15</td>
<td>C5H5</td>
<td>25.50</td>
<td>C38</td>
<td>C6H6</td>
<td>32.26</td>
</tr>
<tr>
<td>C16</td>
<td>C1’H1’</td>
<td>-41.97</td>
<td>C38</td>
<td>C5H5</td>
<td>-2.94</td>
</tr>
<tr>
<td>C16</td>
<td>C5H5</td>
<td>27.31</td>
<td>G39</td>
<td>C1’H1’</td>
<td>-19.86</td>
</tr>
<tr>
<td>G23</td>
<td>C1’H1’</td>
<td>-23.02</td>
<td>G39</td>
<td>N1H1</td>
<td>-6.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4-3 HIV-1 FS-SL RDCs measured in 20 mM potassium and 2 mM magnesium cationic conditions

<table>
<thead>
<tr>
<th>Residue</th>
<th>RDC (Hz)</th>
<th>Error (Hz)</th>
<th>Residue</th>
<th>RDC (Hz)</th>
<th>Error (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>-7.48</td>
<td>2.73</td>
<td>G24</td>
<td>-8.46</td>
<td>2.73</td>
</tr>
<tr>
<td>G4</td>
<td>18.44</td>
<td>3.34</td>
<td>A25</td>
<td>17.42</td>
<td>4.30</td>
</tr>
<tr>
<td>G4</td>
<td>-12.05</td>
<td>2.73</td>
<td>C1’H1’</td>
<td>-15.64</td>
<td>1.28</td>
</tr>
<tr>
<td>A5</td>
<td>2.92</td>
<td>1.28</td>
<td>A26</td>
<td>-16.24</td>
<td>1.28</td>
</tr>
<tr>
<td>U8</td>
<td>-11.85</td>
<td>1.27</td>
<td>A26</td>
<td>14.13</td>
<td>4.30</td>
</tr>
<tr>
<td>G9</td>
<td>18.82</td>
<td>2.41</td>
<td>C2H2</td>
<td>17.92</td>
<td>3.34</td>
</tr>
<tr>
<td>U13</td>
<td>-11.14</td>
<td>2.73</td>
<td>G27</td>
<td>-8.12</td>
<td>2.73</td>
</tr>
<tr>
<td>G10</td>
<td>-12.90</td>
<td>1.28</td>
<td>G28</td>
<td>16.18</td>
<td>3.34</td>
</tr>
<tr>
<td>U14</td>
<td>-12.05</td>
<td>1.27</td>
<td>G28</td>
<td>-10.32</td>
<td>2.73</td>
</tr>
<tr>
<td>C15</td>
<td>15.98</td>
<td>2.41</td>
<td>C38</td>
<td>16.55</td>
<td>4.3</td>
</tr>
<tr>
<td>C16</td>
<td>-17.57</td>
<td>1.28</td>
<td>C39</td>
<td>-23.35</td>
<td>1.28</td>
</tr>
<tr>
<td>G23</td>
<td>-12.43</td>
<td>2.73</td>
<td>C40</td>
<td>21.19</td>
<td>2.41</td>
</tr>
<tr>
<td>G24</td>
<td>-12.11</td>
<td>1.28</td>
<td>C5H5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4-4 Order tensor analysis of RDCs measured in the HIV-1 FS-SL RNA

<table>
<thead>
<tr>
<th>Cation condition</th>
<th>Helix</th>
<th>N</th>
<th>CN</th>
<th>rmsd (Hz)</th>
<th>$R^2$</th>
<th>$\eta$</th>
<th>GDO x $10^3$</th>
<th>GDOint</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 mM K$^+$</td>
<td>Upper</td>
<td>24</td>
<td>3.3</td>
<td>2.5</td>
<td>0.99</td>
<td>0.12 ± 0.04</td>
<td>1.67 ± 0.05</td>
<td>0.67 ± 0.03</td>
<td>43 ± 2</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>17</td>
<td>2.8</td>
<td>1.7</td>
<td>0.99</td>
<td>0.31 ± 0.03</td>
<td>1.12 ± 0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150 mM K$^+$</td>
<td>Upper</td>
<td>19</td>
<td>3.7</td>
<td>3.5</td>
<td>0.98</td>
<td>0.26 ± 0.04</td>
<td>2.0 ± 0.1</td>
<td>0.52 ± 0.05</td>
<td>44 ± 4</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>13</td>
<td>3.6</td>
<td>2.7</td>
<td>0.98</td>
<td>0.5 ± 0.1</td>
<td>1.04 ± 0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 mM K$^+$ 2 mM Mg$^{2+}$</td>
<td>Upper</td>
<td>25</td>
<td>3.0</td>
<td>2.0</td>
<td>0.98</td>
<td>0.01 ± 0.02</td>
<td>0.94 ± 0.07</td>
<td>0.94 ± 0.10</td>
<td>27 ± 7</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>11</td>
<td>4.8</td>
<td>3.2</td>
<td>0.96</td>
<td>0.50 ± 0.09</td>
<td>1.0 ± 0.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Shown for each helical domain is the number of RDCs (N), the condition number (CN), the root mean square deviation (rmsd), the correlation coefficient ($R^2$) between measured and back-calculated RDCs, the order tensor asymmetry ($\eta = |S_{yy} - S_{xx}|/S_{zz}$), the generalized degree of order (GDO = $\frac{2}{3}(S_{xx}^2 + S_{yy}^2 + S_{zz}^2)$, $S_{xx} \geq S_{yy} \geq S_{zz}$), the internal generalized degree of order (GDOint = GDOI/GDOi; GDOI < GDOi), and the interhelical bend angle ($\theta$). All values were generated using RAMAH software with errors estimated using the program AFORM-RDC.
4.3.4 Fluorescence-monitored nucleotide stacking

Fluorescence measurements were performed in triplicate similarly to those previously described [199]. Briefly, the 2-aminopurine substituted RNAs (HIV-1 G33-2AP, G34-2AP and A35-2AP) were excited at 309 nm and emission was measured at 360 nm with a Varian Cary Eclipse Fluorescence spectrometer. Magnesium chloride and potassium chloride were added to 0.002 M or 0.13 M, respectively, into 2 μM RNA and 20 mM potassium phosphate buffer (pH 6.8). Additionally, magnesium chloride was added to 0.002 M into a 2 μM RNA, 20 mM potassium phosphate buffer (pH 6.8), and 130 mM potassium chloride background. Using a 160 μL sample cell, fluorescence was measured for 2 seconds, five consecutive times, at 30 °C. Due to slight changes in lamp intensity, the data from each titration were uniformly scaled such that the RFU at 0 M titrant was 56 for G33-2AP, 200 for G34-2AP, and 230 for A35-2AP. The average of three replicates is reported and error was propagated through.
4.4 Results

The HIV-1 frameshift site stem-loop (FS-SL) structure is shown (Figure 4-1A, B) [108].

The lower helix was truncated to yield an overall alignment tensor dominated by the naturally longer 11 bp upper helix (Figure 4-1C). In all conditions tested, the upper helix was indeed found to dominate the overall alignment and was oriented on average nearly parallel to the magnetic field z-axis, with deviations of 7°(±2°), 11°(±3°) and 3°(±1°), in 20 mM potassium, 150 mM potassium, or 20 mM potassium with 2 mM magnesium, respectfully. The decreased magnitude of the lower helix RDCs relative to the upper helix RDCs is consistent with domain motions in the lower helix under variable potassium concentrations (Figure 4-2A & B). In 20 mM potassium and 2 mM magnesium, the degree of alignment for the upper and lower helices is identical within error (Table 4-4, GDO) and the magnitude of the RDCs is similar (Figure 4-2C). This can be attributed to either quenching of the interhelical motions upon addition of magnesium or a motionally-coupled state [311]. As the Szz is minimally changed in these solution conditions and motional quenching with magnesium has been observed for other RNAs [313, 315, 400], we hypothesize that the internal motions of the HIV FS-SL are effectively quenched by 2 mM magnesium. Such data suggest that the FS-SL is in a motionally-decoupled state in each condition tested. Importantly, order tensor analysis with RDCs measured from a motionally-coupled system would lead to an underestimation of domain motions and would not report a false read-out of domain motions [311].

RDCs measured in different ionic conditions were subjected to order tensor analysis [259, 307, 309, 428]. Order tensor solutions describe the average orientation of each helix relative to the magnetic field. The A-form helices were positioned in their principal axis system and oriented to satisfy connectivity restraint between U6 and C7 and the distant restraint between G32 and A36. RNA helical dynamics were assessed using two additional order tensor terms: the internal general degree of order (GDOint) and the asymmetry parameter (η) [259, 309, 428].
GDOint describes the amount of interhelical motion, while $\eta$ designates the directionality of that motion [428]. Both values range from 0 to 1, for maximal (GDOint=0) to no motions (GDOint=1) and no directional preference ($\eta=0$) to preference for only the x- or y-axis ($\eta=1$). Variations in $\eta$ values between helices are suggestive of anisotropic motions [269, 307, 391]. RNA dynamics defined by no directional preference are deemed isotropic and can be modeled assuming an isotropic cone motional model [290, 312, 429-431]. In contrast, anisotropic helical motions ($\eta_1 \neq \eta_2$) can have complex spatial trajectories [391]. We observe good agreement between the measured and back-predicted RDCs in all three conditions (Table 4-4: $R^2 = 0.96-0.99$, Figure 4-3), validating the use of order tensor solutions to describe interhelical dynamics.

4.4.1 The FS-SL is dynamic in variable concentrations of potassium

In 20 mM potassium, the RNA adopts an average interhelical bend angle of 43°(±2°) (Figure 4-3A) and experiences large amplitude motions (Table 4-4: GDOint = 0.67(±0.03)). The differences in helical $\eta$ values (Table 4-4, $\eta = 0.12(\pm0.04)$ and 0.31(±0.03) for the upper and lower helices, respectively) are indicative of asymmetric domain motions [391]. Prior studies of RNA two-way junctions suggest that the average interhelical bend angle for the FS-SL would be ~55° if it were uniformly sampling all topologically accessible conformations [403]. Therefore the FS-SL appears to unequally sample its accessible states, which is consistent with anisotropic domain motions.
Figure 4-2 Secondary structure of the HIV-1 FS-SL and measured RDC data.
A) The FS-SL analogue used in this study, where the lower helix has been truncated to facilitate decoupling of internal and overall motions. Circles (5’ strand) and triangles (3’ strand) are used to distinguish RDC values measured in different strands. Bars connecting residues denote formation of base-pairs as inferred using 1H-1H NOESY NMR experiments. Grey sections indicate RDCs not used in order tensor analysis. The horizontal dashed lines correspond to the average positive or negative RDC value in each helix. FS-SL RDCs (1DCH and 1DNH) measured between C8-H8 (black), C2-H2 (orange), C5-H5 (purple), C6-H6 (green), N1-H1 (dark blue), and N3-H3 (light blue) spin-pairs in base moieties; and the C1'-H1' (red) spin-pair in the ribose ring in B) 20 mM potassium, C) 150 mM potassium, or D) 20 mM potassium and 2mM magnesium chloride. RDCs are in alignment with the FS-SL secondary structure above.
Figure 4-3
Figure 4-3 Order tensor analysis and the derived average orientations of the FS-SL in variable potassium and magnesium concentrations.

Left panel, correlation plots between measured and back-calculated RDCs when order tensors are independently fit to an idealized A-form geometry for the upper (dark grey) and lower (light grey) helices. Error bars in the x-axis and y-axis are representative of the measured RDC measurement uncertainty (Supplementary Table 4-4) and calculated rmsd (Table 4-4), respectively. Right panel, the RDC derived average orientation of the FS-SL. Results determined in A) 20 mM potassium, B) 150 mM potassium, and C) 20 mM potassium and 2 mM magnesium cation concentrations.
We also examined changes in bulge nucleotide stacking in response to different ionic conditions. The purines in the bulge (5' GGA 3') were individually substituted with the fluorescent purine analogue 2-aminopurine (2AP). The incorporation of 2AP in these positions is justified by the observation that the GGA nucleotides do not participate in hydrogen bonding interactions [108] and therefore, replacement with 2AP should minimally alter the FS-SL structure [432]. 2AP is a sensitive reporter of local structure because it will fluoresce when solvent accessible, but its fluorescence is quenched upon stacking with other bases [432-442]. A small number of RNA helical junctions have been experimentally analyzed in this manner [433, 443].

In 20 mM potassium, the nucleotide at position 33 displays low fluorescence, consistent with a predominately stacked nucleotide conformation. This observation agrees with thermodynamic data that indicate a favorable stacking free energy for purine nucleotides adjacent to a helical 3' end [444] (nucleotide 33 is directly 3’ to the stable upper helix). This is due to the right-handed nature of the helix, which exhibits much more stacking for 3’ vs. 5’ terminal nucleotides [445]. In contrast, nucleotides 34 and 35 display significantly higher fluorescence relative to nucleotide 33 (Figure 4-4A), in line with partial unstacking of the bases at position 33 and 34. Such bulge nucleotide orientations are consistent with structures within the ensemble of the HIV-1 FS-SL (Figure 4-4B) [108, 187, 199].

To qualitatively examine fast (low to sub-nanosecond) time-scale nucleotide dynamics, resonance intensities from $^1$H-$^{13}$C-HSQC experiments were measured (Figure 4-5). Substantial increases in intensities are indicative of motions on fast time-scales [314, 402]. In 20 mM potassium (Figure 4-5A), several nucleotides within the FS-SL tetraloop (A18, C20, A21, and A22) and bulge (G34 and A35) have higher than average resonance intensities, indicative of motions that are faster than the overall tumbling of the RNA. The largest normalized resonance
intensities arise in the FS-SL tetraloop (Figure 4-5A). Significant tetraloop nucleotide dynamics are consistent with previous observations [446, 447]. The average resonance intensity for A33 and the significantly higher intensities for G34 and A35 are in agreement with the 2AP data (Figure 4-4A) as well as the NMR structure (Figure 4-4B) [108]. Such fast time-scale motions of G34 and A35 likely allow unstacking of these nucleotides and provide a flexible hinge point to facilitate the observed interdomain motions.
Figure 4-4

A

- 20 mM K^+ (white bars)
- 150 mM K^+ (black bars)
- 20 mM K^+ 2 mM Mg^{2+} (gray bars)
- 150 mM K^+ 2 mM Mg^{2+} (hatched bars)

RFU

G33  G34  A35

2AP

B

- C7
- U6
- G32
- G33
- G34
- A36
- A35

C

- C7
- U6
- G32
- G33
- G34
- A36
- A35
Figure 4-4 Bulge nucleotide stacking.
A) Relative fluorescence levels for 2 μM G33-2AP, G34-2AP, and A35-2AP were measured with 20 mM potassium (white), 150 mM potassium (hashed), 20 mM potassium and 2 mM magnesium (grey), and 150 mM potassium and 2 mM magnesium (black) cation concentrations. Error bars representing the SEM are shown. B) Bulge nucleotide orientations in 20 mM and 150 mM potassium are consistent with the FS-SL solution structure (PDB ID 1Z2J, state 7, nucleotides 8-9 and 34-38 shown). Partially unstacked and dynamic nucleotides are shown in dark red and stacked nucleotides are shown in grey. C) In 20 mM potassium and 2 mM magnesium, G33 is stacked (grey), G34 adopts a predominantly unstacked conformation (bright red), and A35 is dynamic (dark red). These positions are also consistent with the FS-SL solution structure (PDB ID 1Z2J, state 3, nucleotides 8-9 and 34-38 shown).
Figure 4-5 RNA dynamics revealed by resonance intensities.

Normalized resonance intensities (peak heights) measured from non-constant-time 13C-1H-S3A HSQC spectra. Shown are the values for sugar C1’H1’ (diamonds) and base C2H2 (squares), C5H5 (triangle), C6H6 (X), and C8H8 (circles) for the FS-SL RNA. The intensity for each type of C-H spin was normalized by the intensity of the corresponding A25, U8, C16, and G27 spin-pairs. A) 20 mM potassium supplemented with B) 130 mM potassium chloride or C) 2 mM magnesium chloride (dialyzed).
Order tensor analysis revealed no significant change in the FS-SL average interhelical bend angle in 150 mM potassium (Figure 4-3B, Table 4-4) relative to the 20 mM potassium condition (Figure 4-3A). Additionally, no difference in stacking of the bulge nucleotides is observed (Figure 4-4A). The fast time-scale local motions in the RNA tetraloop and bulge are also similar in both potassium conditions (Figure 4-5A & B). Interestingly, the FS-SL interhelical domain dynamics and asymmetry modestly increase with increased potassium, as indicated by a reduction in GDOint and increase in η values (Table 4-4: compare GDOint and η values between 20 mM and 150 mM potassium).

4.4.2 Magnesium promotes coaxial alignment and rigidification of the FS-SL structure.

Addition of Mg²⁺ causes line broadening in the bulge and loop regions (Figure 4-6), which we attribute to unfavorable exchange dynamics induced by non-specific Mg²⁺ association. However, a sufficient number of helical RDCs were measured, which enabled calculation of accurate order tensor solutions (Figure 4-3C, Table 4-4). In 20 mM potassium and 2 mM excess magnesium, the FS-SL adopts a more coaxial orientation defined by a significant decrease in the interhelical bend angle (Figure 4-3C), from 43° (±2°) to 27° (±7°) (Table 4-4). Additionally, FS-SL interhelical domain motions were arrested under these conditions (Table 4-4: GDOint = 0.94(±0.1)). The fast time-scale bulge nucleotide dynamics observed in 20 mM and 150 mM potassium were also attenuated (Figure 4-5C). It is important to note that the decreased resonance intensities for the bulge nucleotides also could be reflective of intermediate exchange dynamics of RNA-associated magnesium ions, rather than a decrease in nucleotide dynamics per se.

The decreased bend angle upon addition of 2 mM magnesium chloride is accompanied by a small increase in the fluorescence of G33-2AP and A35-2AP and a large increase in the fluorescence for G34-2AP (Figure 4-4A). Thus, magnesium promotes unstacking of the central
base at position 34, but does not significantly alter the degree of stacking at positions 33 and 35, which are adjacent to helical termini. These nucleotide orientations also agree with structures within the ensemble of the HIV-1 FS-SL (Figure 4-4C) [108]. Unexpectedly, the tendency of G34-2AP to unstack in 20 mM potassium and 2 mM magnesium is abated in 150 mM potassium and 2 mM magnesium (Figure 4-4A). Thus, unstacking of G34 is competitively inhibited by 150 mM potassium. When G34 is unstacked (Figure 4-4C), the A35-P to A36-P distance is decreased by ~0.5 Å relative to a partially stacked position (Figure 4-4B). We hypothesize that magnesium’s relatively small ionic radius and +2 charge effectively stabilizes the increase in local negative charge density associated with an unstacked G34. Potassium, which has a +1 charge and ~1.4 fold larger ionic radius relative to magnesium, cannot counter the changes in charge density associated with an unstacked G34 and, at high concentrations, affectively shields the RNA from magnesium.
Figure 4-6 Magnesium exchange broadening in non-helical regions of the SL.
Non-constant-time $^1$H-$^{13}$C HSQC spectra are shown in 20 mM potassium (blue) and 20 mM potassium and 2 mM excess magnesium chloride (green). Resonances corresponding to bulge residues 33-35, loop residues 18-21, and nucleotides proximal to the bulge (residues 6, 36, 7, 32) are indicated. $^1$H-$^{13}$C-HSQC spectra optimized for the A) aromatic resonance region and the B) ribose resonance region.
4.5 Discussion

Overall, our results are remarkably similar to those that have been previously measured for TAR [220, 286, 306, 311-315, 319, 375, 385, 391, 400-403]. Both RNAs exhibit fast time-scale nucleotide dynamics in their trinucleotide bulge and tetraloop regions [286, 393, 402], have similar average interhelical bend angles (~43° (FS-SL) and ~47° (TAR)[312]), and display markedly anisotropic helical motions [311] that are quenched by physiological concentrations of magnesium [400, 448]. Interestingly, there are also differences in dynamics of these RNAs. TAR is more flexible than the FS-SL, with a GDOint of 0.45(±0.05) [311] compared to the FS-SL GDOint of 0.67(±0.03). The relative decrease in flexibility for the FS-SL may be the result of more favorable stacking interactions and possibly a greater level of steric constraints due to the purine bulge [375]. The base-pairs above and below the bulge are different for TAR and the FS-SL and can also influence RNA dynamics [403]. Together with previous studies [311, 312, 315, 391, 400], these results reveal general properties of RNA interhelical domain motions. Specifically, interhelical domain motions are sequence independent, anisotropic, and are quenched by physiological concentrations of magnesium in a low-monovalent background.

Both the TAR and FS-SL RNAs are coaxially aligned in physiological concentrations of magnesium [400, 403, 448]. However, the extent of alignment is not equivalent for the two RNAs under similar divalent solution conditions (~27° (FS-SL) vs. ~5° (TAR) [400]). These findings are corroborated by observations from Zacharias and Hagerman [404], which showed that helical domains connected by poly-U linkers more easily adopt coaxial conformations in the presence of divalent ions than their poly-A counterparts. The differences in extent of coaxial alignment can be attributed to the energetic penalty of purine nucleotide unstacking [417]. The propensity of purine nucleotides to stack [449, 450] may limit changes in interhelical motions in response to the stabilizing effect of divalent ions [451]. Therefore, the degree of flexibility in the
linker may depend on its sequence, with purine-rich sequences restricting domain motions to a greater degree than pyrimidine-rich sequences.

A small, but significant increase in FS-SL domain dynamics and motional asymmetry was observed in 150 mM potassium relative to 20 mM potassium. Increased RNA flexibility as a function of increasing ionic strength has been previously observed for a dynamic two-helix model RNA system using small angle X-ray scattering [408]. Thus, electrostatic repulsion can limit RNA conformational sampling [408].
4.5 Acknowledgements

The authors thank Dr. Jordan Burke, Dr. Lawrence Clos, and Allison Didychuk for helpful discussions. This study made use of the National Magnetic Resonance Facility at Madison, which is supported by NIH grants P41RR02301 (BRTP/NCRR) and P41GM66326 (NIGMS). Additional equipment was purchased with funds from the University of Wisconsin, the NIH (RR02781, RR08438), the NSF (DMB-8415048, OIA-9977486, BIR-9214394), and the USDA.
Chapter 5: Conclusions and future directions

5.1 Conclusions

5.1.1 Overview

Despite the necessity of -1 frameshifting in HIV-1 replication, the exact mechanisms that determine the recoding event still require further investigation. When I began my work, the role of the fundamental, stimulatory, structural element in the frameshift site was unclear. While many studies sought to explain the role of the downstream structure in -1 PRF through RNA overall thermodynamic stability \([155, 156]\), mechanical stability \([157-160]\), or conformational plasticity \([161]\), their conclusions were often inconsistent, generating considerable confusion in the field. The goal of my work was to define the relationship between the HIV-1 frameshift site stem-loop and its function in translational reprogramming.

My work has expanded our understanding of -1 PRF in HIV-1 through a multifaceted approach, including structural biology, biophysics, and biochemistry. The results have contributed to our understanding of the structure and function of the HIV-1 frameshift site stem-loop, where secondary structure in frameshift site plays a critical role in -1 frameshift stimulation (section 5.1.2). Furthermore, this frameshift site stem-loop was used as a model to investigate the dynamic properties of RNAs with purine-rich bulges, a common structural motif. The findings explain the effect of solution conditions on the dynamics of the helical elements within this structure (section 5.1.3).

5.1.2 Investigation of HIV-1 -1 PRF

Previous reports showed trends between HIV-1 frameshift efficiency and stem-loop overall thermodynamic stability, where stability enhances the -1 frameshift event \([136, 155, 156]\). The data presented here contradict this viewpoint (Chapter 2). Systematic mutagenesis was
employed to decouple the HIV-1 frameshift site stem-loop’s overall and local thermodynamic from frameshift efficiency. In contrast to previous studies, we found no correlation between overall thermodynamic stability and frameshift efficiency. Instead, a strong correlation ($R^2 = 0.83$) between HIV-1 stem-loop local thermodynamic stability and frameshift efficiency was observed. The results provide a new understanding of the importance of base-pairs in the base of the stem-loop for frameshift stimulation. We believe these base-pairs are positioned at the entrance to the mRNA entry tunnel and directly inhibit ribosome progression along the mRNA. Our data demonstrate that frameshifting is driven by a “thermodynamic block” to translocation: the stronger the thermodynamic block, the greater the amount of frameshifting. This thermodynamic block is determined by the local stability of stem-loop base-pairs positioned proximal to the mRNA entrance tunnel during frameshifting. Using our experimental data, we developed a quantitative model that predicts HIV-1 -1 frameshift efficiency. This model is applicable to other frameshifting systems to explain their mechanism of structural stimulation.

5.1.3 Characterization of HIV-1 Frameshift Site Stem-loop Domain Motions

In HIV-1 genomes, the frameshift site stem-loop folds into an extended structure, with the 11 base-pair stem-loop separated by a three-purine bulge from an 8 base-pair helix. This RNA was used as a model to investigate the dynamic properties of RNAs with purine-rich bulges, a common structural motif. Using NMR, the interhelical dynamics within this extended structure were examined as a function of potassium and magnesium concentrations. I found that under low- and high-monovalent solution conditions, the stem-loop undergoes large-scale interhelical motions. These motions are quenched by near physiological concentrations of magnesium in a low monovalent background. Comparison of my results to those of previous studies suggests
the impact of monovalent and divalent ions on RNA domain dynamics is general, and independent of sequence.
5.2 Future directions

5.2.1 Examining HIV-1 -1 PRF in vivo

The study presented in Chapter 2 used in vitro methods to develop a quantitative understanding of -1 frameshifting in HIV-1. This work can be expanded by employing in vivo methods, single-molecule techniques, and targeted changes to frameshift efficiency to broaden our understanding of the mechanism of -1 PRF in HIV-1.

We initiated a collaboration with Prof. Nathan M. Sherer (University of Wisconsin-Madison) to test the importance of stem-loop local thermodynamic stability in the HIV-1 frameshift mechanism in vivo. A preliminary investigation will use only a subset of our stem-loop mutants (Figure 3-1, MS1-3, 5-7, and 10 & 12). These mutants were chosen because of their limited range of sequence changes a large range of local or global stability, and large range of frameshift efficiencies. Using standard PCR-based methods, sequences coding for stem-loop mutants will be inserted into the NL4-3 ΔEnv, ΔVPR, ΔNef, +GFP provirus plasmid. 293T cells will be transiently transfected with the provirus plasmids and Gag and Gag-Pol protein products will be detected with standard western-blot analysis, and a capsid-specific antibody. Frameshift efficiency will be determined from the ratio of Gag:Gag-Pol. We expect that the quantitative relationship between local stability and the degree of frameshift efficiency observed in vitro, will be replicated in vivo.

Alteration of HIV-1 frameshifting efficiency can affect viral replication and infectivity [103, 127, 131, 134-136]. Three previous studies have examined the impact of increased Gag:Gag-Pol ratios on viral fitness [131, 134, 135]. In these studies the smallest change in the in vivo Gag:Gag-Pol ratio, a ~ 3.5 fold increase in Gag-Pol, decreased virion production by 40 % [131]. Greater changes in Gag-Pol levels resulted in an even greater decrease in viral fitness [131, 134, 135]. The impact of finer changes in frameshift efficiency on viral fitness has not been tested.
Additionally, relatively small decreases in frameshift efficiency greatly inhibit viral replication [103, 127, 136].

If the changes in frameshift efficiency associated with MS1-5 and MS7 (Table 2-3) are replicated in vivo, the impact of 1.1 - 5.0 fold increases in frameshift efficiency on viral fitness could be determined using mutant stem-loop provirus plasmids. Using a single-cell infectivity assay in HeLa cells, viral infectivity could be monitored utilizing the GFP reporter and fluorescence activated cell sorting. Virion assembly and budding could be monitored with time-lapse fluorescence microscopy if the mutant stem-loops are incorporated into a Gag-mCherry NL4-3 or similar provirus plasmid. However, a clear understanding of the relationship between frameshift efficiency and viral replication for many of these mutants may be confounded by changes in amino acid sequence in the gag and pol ORFs. The frameshift site sequence overlaps the p1-p6\(^{\text{p6*}}\) and the transframe protein p6\(^{*}\) coding regions in gag and pol, respectively (Figure 1-8). The p1-p6 polypeptide is important for virion budding (reviewed in [452]), whereas the majority of p6\(^{*}\) seems to be dispensable for viral replication [453, 454]. As many of the mutant stem-loops designed in Chapter 3 (Figure 3-1) have significant differences in frameshift site stem-loop sequence, utilization of a HIV-1 provirus plasmid that uncouples the gag and pol reading frames [453] would be especially useful in that it would allow easier design of mutations in the frameshift site that preserve the Pol amino acid sequence. Alternatively, mutant stem-loop sequences [100] with a limited number of changes in Gag and Gag-Pol sequence could be used in a typical NL4-3 provirus context.

Targeted disruption of the stem-loop base to eliminate its local stability and attenuate frameshifting may be a promising strategy of antiretroviral therapy. Several studies have targeted the HIV-1 frameshift site stem-loop using small-molecule and peptide ligands [129, 191, 199, 446, 455-460]. A subset of these ligands significantly decreased frameshift efficiency
[129, 199, 456], sometimes with the effect of impaired viral replication [129]. Given our current understanding of stem-loop stimulated -1 PRF, specific disruption of base-pairing within the stem-loop base may severely inhibit viral replication. Locked nucleic acid (2'-O, 4'-C-methylene linked bicyclic ribonucleotides, also known as LNA) oligonucleotides [461] complementary to the 3’ strand of the HIV-1 frameshift site stem-loop have the potential achieve similar inhibitory activity. LNAs are extremely stable RNA analogs, providing long term circulation in cellular systems [461]. The affinity of LNA for complementary sequences can be improved by including LNA in RNA or DNA oligonucleotides [111, 462]. This complementarity to the HIV-1 frameshift site should confer high target specificity [463]. Therefore, a variety of LNA/DNA/RNA oligonucleotides could be designed to target the RNA. Binding could be verified using NMR, while changes to frameshifting and viral fitness can be assayed using the methods described above.

5.2.2 Analysis of the relationship between stem-loop mechanical stability and -1 PRF

In -1 PRF sites with downstream pseudoknots, pseudoknot mechanical stability has been implicated as the major determinant of -1 frameshift efficiency [157-160]. However, a recent study suggested that mechanical stability does not correlate with frameshift efficiency, and instead, frameshifting is modulated by structural conformational plasticity [161]. Our stem-loop mutants have the potential to provide additional data to evaluate the importance of this structural characteristic to frameshift stimulation. Single molecule mechanical tweezers experiments could be used to determine the mechanical force required to pull each of the structures apart (reviewed in [464]). The relationship between the measured frameshift efficiencies, both in vitro and in vivo, can be subsequently examined as a function of mechanical stability.
Instead of contributing directly to frameshifting, mechanical stability of the downstream structure may directly control ribosomal pausing. Two studies suggested that ribosomal pausing on the slippery sequence is necessary, but not sufficient for -1 frameshifting [195, 196]. The duration of pausing promoted by the mutant stem-loops designed in Chapter 2 could be measured by using ribosomal heel print experiments [195]. If pausing is not directly related to frameshift efficiency, a direct correlation between mechanical stability and pausing may explain the disparity between several studies that have attempted to relate mechanical stability to frameshift efficiency [157-161].

5.2.3 Investigation of the HIV-1 three-helix junction

We hypothesize that 3HJ secondary structure modulates frameshifting by increasing the density of ribosomes stacked at the frameshift site (Chapter 2). This hypothesis could be examined directly by ribosomal profiling experiments [465]. Ribosomal profiling provides an output of ribosome density along mRNA. The experiments could be completed by using genomic RNA transcribed from the NL4-3 mutant plasmids in an in vitro translation extract or in vivo in 293T cells. To assess changes in ribosome stacking associated with additional secondary structure, results from the genomic RNAs could be compared to ribosome density on an RNA control that disrupts base-pairing in the P1 and P2 helices (Figure 2-9 3HJ mut). Because the RNA in the 5’ strands of the P1 and P2 helices is upstream of the frameshift site, it may be possible to disrupt base-pairing within these helices using only silent mutations.

Using SHAPE chemical probing, the HIV-1 three-helix (3HJ) secondary structure was detected in HIV-1 RNA extracted from viral capsids [170, 171]. The 3HJ includes the HIV-1 frameshift site and an additional ~90 nucleotides upstream and downstream of the frameshift site. However, this secondary structure may not be representative of the secondary structure when it is most relevant to frameshifting. Although our in vitro frameshift results suggested
that the ability to form the 3HJ secondary structure can indirectly modulate frameshifting, we have not directly verified formation of the structure in these contexts. A preliminary NMR investigation of the structure of the 3HJ construct yielded inconclusive results regarding formation of the 3HJ in solution (Appendix III). To directly investigate the genomic RNA structure in a translational context, HIV-1 genomic RNA could be purified from infected HeLa or 293T cells and probed using SHAPE, which provides an experimental measurement of nucleotide flexibility with single-nucleotide resolution [466]. Isolation of the genomic RNA from these cells would likely be technically challenging, but may be feasible with the assistance of experienced collaborators.

5.2.4 HIV-1 stem-loop dynamics in near physiological solution conditions

During our investigation of HIV-1 extended stem-loop domain motions, we determined that the impact of magnesium on bulge nucleotide stacking is abated in a high potassium background. Given that magnesium quenched stem-loop dynamics and promoted bulge nucleotide extrusion in low-ionic strength conditions, this result was particularly surprising. It will be interesting to see if RNA domain motions are present in cationic solution conditions with 2 mM magnesium and 150 mM potassium. This solution regime would be very relevant, as it approximates physiological magnesium and potassium concentrations [448]. Experiments mirroring those described in Chapter 4 could be used to quantitatively describe RNA dynamics in this condition using residual dipolar coupling measurements.
Appendix I: Structure of the IAPV IGR IRES PK1 domain

The material presented in this appendix is a result of research performed by the author with contributions from Jordan E. Burke, Prof. Samuel E. Butcher and Prof. Eric Jan (University of British Columbia).

K.D.M. conceived of the experiments, designed the RNA constructs, completed all RNA sample preparations, collected all SAXS data, executed the preliminary filtering of MC-SYM structural models by SAXS and NMR, and was responsible for all NMR data collection, analysis, and interpretation.

J.E.B assisted with SAXS data collection. Additionally, J.E.B completed the SAXS data processing, MC-Fold/MC-Sym structural modeling, ab initio modeling, structural docking, and analysis of the preliminary filtered structures.
A1.1 Overview

Israeli acute paralysis virus (IAPV) is a member of the *Dicistrovirdae* viral family [467, 468]. IAPV infection in honey bees is associated with honeybee colony collapse disorder [469]. The IAPV positive strand RNA genome is characterized by two non-overlapping reading frames (ORFs) [470]. Translation initiation of *Dicistrovirdae* non-structural genes is facilitated by a 5′ internal ribosome entry site (IRES) (see section 1.2.2) [471-474]. While, translation of the downstream structural genes is controlled by an intergenic (IGR) class II IRES, which facilitates cap-independent and factor-less translation initiation on a non-AUG codon in the ribosomal A-site [53, 75, 467, 475].

The IGR IRES is made up of two modular domains [467]. The first domain is composed of two different pseudoknot structures and recruits the ribosome to the viral RNA. The second domain is composed of a single pseudoknot structure (PKI) [467]. PKI binds within the ribosomal P-site and positions the ribosome for non-AUG initiation [53, 75]. It is thought that PKI structural mimicry of a canonical P-site codon-anticodon interaction allows the IRES to correctly position the ribosome for non-canonical initiation [53, 75, 476].

Although IGR IRESes are predicted to have a highly conserved secondary structure, two dicistrovirus IGR IRES classes are defined by variation in two structures [55, 477][Jan, 2006 #52]. IRES are identified by the size of a bulge region (L1.1) within domain one and the presence of an extra stem-loop in the PKI domain. The L1.1 region is responsible for 80S ribosome assembly on class I and II IRESes [54, 478]. The role of the additional stem-loop (Figure A1-1B, helix III) is unclear, but data suggest it is involved in ribosome binding and positioning [467, 479, 480]. The IGR class I PKI has been observed to form a tRNA-like structure [476, 479, 481, 482]. Although class II PKI structures are presumed to be similar, no structure of a class II PKI domain has been solved.
Structure determination of the IAPV class II IGR PKI domain would greatly enhance our understanding of structure mediated non-canonical translation initiation. Interestingly, PKI initiates translation \textit{in vitro} in the 0 and +1 reading frames with 80\% and 20\% frequency (Figure A1-1A) [483]. Data suggest that base-pair formation between U14 and G70 is critical for +1 initiation. Numerous point mutations within PKI can modulate the relative levels of 0/+1 initiation, however the impact on these mutations on the PKI structure is unclear ([483] and personal communication E. Jan).

Here we present a preliminary structural model of the IAPV IGR IRES PKI domain in solution as analyzed by NMR spectroscopy and SAXS. We used a recently developed method to sort a large number of fragment based models by SAXS and NMR residual dipolar coupling measurements [484, 485]. The resulting complex has a distinct overall fold that may be important during IRES initiation.
Figure A1-1 Proposed secondary structure of the IAPV IGR IRES PKI domain.

(A) The PKI domain of the IGR IRES can initiate translation on a non-AUG codon in the ribosomal A-site in the 0 or +1 frame. Numbering is consistent with [483]. (B) The 70-nt IAPV IGR PK1 domain was renumbered for this study. Structural features are Helix I (green), helix II (purple), helix III (blue), and helix IV (orange). (Black lines or circles) Experimentally determined base-pairs (see Figure A1-5) and single stranded nucleotides (grey). Asterisks (*) indicate nucleotides involved in non-Watson-Crick base-pairs that do not have identified pairing partners. (C) Proposed secondary structure of the 55 nucleotide PK1Δ56-70. Structural features are helix I (green), helix II (purple), and helix III (blue). Base-pairs are indicated as in (B).
A1.2 Materials and methods

A1.2.1 RNA sample preparation

RNA was transcribed in vitro using purified His$_6$-tagged T7 RNA polymerase. DNA templates were purchased from Integrated DNA Technologies (IDT). $^{13}$C-$^{15}$N labeled samples of PKI and PK1Δ56-70 were prepared using $^{13}$C-$^{15}$N labeled nucleotides (Cambridge Isotope Laboratories). RNA samples were purified using denaturing 7.5-12.5% PAGE with 8 M urea. Impurities were removed by DEAE anion exchange (Bio-rad) using a low-salt buffer (20 mM Tris-HCl at pH 7.6, 200 mM sodium chloride) to wash and a high-salt buffer (20 mM Tris-HCl at pH 7.6, 1.5 M sodium chloride) to elute the RNA. Samples were then diluted to between 1-10 μM in 20 mM potassium phosphate (pH 6.3), 200 mM KCl, 0.5 μM EDTA, denatured for 5 min in boiling water, and annealed on ice for 30 minutes. RNA samples were concentrated with 10K molecular weight cut off centrifugal filter units (Millipore) at 4 °C to 300 μL volumes. SAXS samples were subject to an additional size exclusion purification step using a HiLoad 16/500 Superdex 75 column (Amersham Biosciences) and dialyzed against 10 mM Tris (pH 6.3), 200 mM KCl, and 0.5 μM EDTA. All samples were assayed for folding homogeneity by 6% non-denaturing PAGE.

A1.2.2 NMR data collection

All spectra were obtained on Bruker Avance or Varian Inova spectrometers equipped with cryogenic single z-axis gradient HCN probes at the National Magnetic Resonance Facility at Madison. Resonances were assigned using $^1$H-$^1$H 2D NOESY with a mixing time of 100 msec and $^1$H-$^{15}$N 2D HMQC experiments at 10 °C. Partial alignment for RDC experiments was achieved by addition of 12.5 mg/mL Pf1 filamentous bacteriophage (ASLA) to a $^{13}$C, $^{15}$N U- and G-labeled sample. Pf1 phage concentration was confirmed by measuring $^2$H splitting at 700 MHz. Imino $^1$H RDC measurements were obtained using $^1$H-$^{15}$N 2D HMQC, $^1$H-$^{15}$N 2D TROSY HSQC, and $^1$H-$^{15}$N 2D Semi-TROSY HSQC experiments. RDC measurements were in the range
of -50 to 50 Hz. The uncharacteristically large imino RDCs are likely due to excessive sample alignment. Average RDC values were determined using the $^1$H half-splitting and $^1$H full splitting. RDC errors were determined by the difference in RDCs in the two different RDC determinations. RDCs with greater than 4.0 Hz errors were excluded from the data set.

**A1.2.3 SAXS data collection**

All SAXS data were obtained at Sector 12-ID-B and 5-ID-D of the Advanced Photon Source at Argonne National Laboratory. Measurements were carried out in 10 mM Tris (pH 6.3), 200 mM KCl, and 0.5 μM EDTA. RNA samples were loaded into a 1-mm capillary and flowed back and forth throughout the exposure. Twenty data collections of 0.5 sec each were averaged for each sample and buffer. The scattering intensity was obtained by subtracting the background scattering from the sample scattering. Subtraction of wide-angle scattering (WAXS) was adjusted until the contribution from buffer scattering was negligible. The scattering intensity at $q = 0 \text{ Å}^{-1}$ [$I(0)$], as determined by Guinier analysis, was compared between four different concentrations (0.5, 1.0, 1.5, and 2.0 mg/mL) to detect possible interparticle interference. WAXS and SAXS data were merged using the region between $q = 0.09 \text{ Å}^{-1}$ and 0.17 \text{ Å}^{-1} in PRIMUS [323]. Samples were assayed for radiation damage by denaturing 10% PAGE after data collection. No radiation damage was detected (data not shown).

**A1.2.4 Ab initio structure calculation**

All SAXS data were processed with GNOM software [330] to extrapolate the scattering curve to intensity at $q = 0 \text{ Å}$ and obtain the pair distance distribution function (PDDF). $D_{\text{max}}$ was calibrated in increments of 2 Å until the PDDF curve fell smoothly to zero. The GNOM output was then used with DAMMIF software [332] to calculate 10 dummy atom models. Models were averaged using the program DAMAVER [334], with a resulting normalized spatial
discrepancy (NSD) of between 0.7 and 0.8, indicating good agreement between the individual models. Ab initio models were superimposed using the Supcomb program [486].

A1.2.5 Molecular modeling and filtration of models by SAXS and RDCs

PKI three-dimensional (3D) models consistent with the NMR-determined secondary structure were created using the MC-Fold/MC-Sym pipeline [337]. The small-angle X-ray scattering amplitudes of the 629 models were predicted by using the FoXS web server [342]. Agreement with the experimental SAXS amplitudes was evaluated by $\chi^2$ goodness-of-fit analysis. The top 25% of models, those with the lowest $\chi^2$ values, were then sorted by fit to 16 experimental imino $^1$H-$^{15}$N RDC measurements as determined using the PALES/DC software [487]. The models with a Q factor of <0.31 were chosen for future refinement. The backbone RMSD for these structures was ~8.2 Å. Covalent connectivity was restored to these five models using the AMBER force field in GROMACS (Open MM Zephyr software) [488]. Additionally, all-atom models for the 12 base-pair extensions of helix I and III were generated using the MC-Fold/MC-Sym pipeline [337] for comparison to ab initio density.

A1.2.6 Molecular modeling and filtration of PKIΔ56-70 models by SAXS and RDCs

Three-dimensional (3D) models of the PKIΔ56-70 consistent with the NMR-determined secondary structure were created using the MC-Fold/MC-Sym pipeline [337]. The small-angle X-ray scattering amplitudes of ~1000 PKI were predicted using the FOXS web server [342]. Agreement to the experimental SAXS amplitudes was measured using $\chi^2$ goodness-of-fit analysis. Based on lowest $\chi^2$ values, the top 25% of models were sorted using the PALES/DC software [487] by 15 experimental RDC measurements from imino $^1$H-$^{15}$N couplings. The models with a Q factor of <0.31 were chosen for analysis. The best seven models have an overall backbone RMSD no better than the backbone RMSD for randomly selected structures (data not shown). We hypothesize that this resulted from not obtaining enough RDCs to define
the helical junction. Because this RNA is simply a truncation of the biologically relevant PKI
IRES domain, we will not pursue further refinement of the PKIΔ56-70 structure.
A1.3 Results

A1.3.1 Global structure of the IAPV IGR IRES PKI domain

To investigate the structure of the IAPV IGR IRES PKI domain (hereafter referred to as PKI), we used a 70-nt construct that contains the entire base-paired region in the pseudoknot domain (Figure A1-1B). The overall fold of the PKI RNA was determined by small-angle X-ray scattering (SAXS) (Figure A1-2). The Kratky profile shows one well defined peak (Figure A1-2B), suggestive of a well-folded RNA structure [489]. A similar profile was observed when nucleotides 56-70 were eliminated from the PKI construct (Figure A1-2B). The p(r) plot shows a major peak at 20 Å indicative of A-form RNA helical width (Figure A1-2C). A minor peak is observed at ~45 Å, which may correspond to helical length. Models of a coaxially stacked PKI helix I, II, and IV or helix III generated with the MC-Fold/MC-Sym pipeline [337] measured ~49 Å and ~40 Å in length, respectively. Both peaks are present in the p(r) plot for PKIΔ56-70; however, the minor peak shifted closer to ~40 Å. This distance is now consistent with the modeled length of helix III, as well as the length of a coaxially stacked helix I & II. The maximum dimension (D_{max}) and radius of gyration (R_g) of PKI are 90 Å and 26.4 Å, respectively. PKIΔ56-70 has a 15 Å reduction D_{max} (75 Å) and an R_g of 22.8 Å, both consistent with its expected reduction in size.

Based on SAXS data collected in 200 mM KCl, we calculated ab initio models of PKI and PKIΔ56-70 using the DAMMIF program [332] (Figure A1-3). The low resolution models reveal that PKI forms a planar structure with an asymmetrical distribution of RNA (Figure A1-3B). The shape of the model is consistent with the Kratky profile and the p(r) plots. In the model we see features consisted with the approximate lengths of individual domains. To determine the location of the helices within the PKI ab initio envelope, ab initio envelopes were calculated from SAXS data from PKIΔ56-70 (Figure A1-3B) and two constructs with 12 base-pair
extensions on either helix I or helix III (Figure A1-4A, B). The loss of density in the PKIΔ56-70 envelope (Figure A1-4B) is indicative of the location of helix IV and linker nucleotides (Figure A1-1C). The helical extensions in PKIΔ56-70 manifested as additional density next to their corresponding helix (Figure A1-4). Therefore, the truncated and extended constructs allowed identification of helix I, helix III, helix IV and the single stranded linker (Figure A1-3B, Figure A1-4). The predicted lengths for MC-SYM [337] generated models of the 12 base-pair extensions in helix I and helix II agreed with the additional density length (Figure A1-4).
Figure A1-2 SAXS of IAPV PKI and PK1Δ56-70.
IAPV IGR PK1 and PK1Δ56-70 (A) experimental scattering profiles, (B) Kratky profiles, and (C) Pair distance distribution function plots. Experiments were conducted in 20 mM potassium phosphate (pH 6.3), 200 mM KCl, 0.5 μM EDTA.
10 ab initio structures were generated for PK1 and PK1Δ56-70 (grey) using the program DAMMIF and then separately averaged with DAMAVER, yielding a normalized spatial discrepancy (NSD) of 0.73 (PKI) and 0.74 (PK1Δ56-70). (A) The ab initio structure of PKI is planar with an asymmetrical RNA density. A 90 Å maximum dimension was determined from the $p(r)$ plot in Figure A1-2. (B) The ab initio models of PK1Δ56-70 and PKI are overlaid. The difference in density between the two identifies the location of the linker nucleotides and PKI helix IV.
Figure A1-4

Identification of helices within the PK1Δ56-70 envelope.

10 ab initio structures were generated for PK1 and PK1Δ56-70 (grey) using the program DAMMIF and then separately averaged with DAMAVER, yielding a normalized spatial discrepancy (NSD) of 0.69 (helix I extension) and 0.82 (helix III extension). Helical extensions of 12 base-pairs on PK1Δ56-70 helix I (A, in green) or helix III (B, in blue) were added to the original 55-nt RNA. Lengths of helical extensions were estimated based on 3D models generated using MC-Fold/MC-Sym pipeline [337]. The ab initio model containing extended helix I (C, green); the model containing extended helix III (D, blue); the original RNA PK1Δ56-70 (grey).
A1.3.2 NMR spectroscopy of PKI

PKI secondary structure was determined from two-dimensional (2D) $^1$H-$^1$H NOESY (Figure 1A-5) and $^1$H-$^{15}$N HMQC NMR spectra (Figure A1-6) in 20 mM KPO$_4$, 200 mM KCl, and 0.5 μM EDTA (pH 6.3). Aside from the expected loss of signals for nucleotides 56-70, deletion of PKI nucleotides 56-70 did not significantly alter the $^1$H-$^1$H NOESY (Figure 1A-7) and $^1$H-$^{15}$N HMQC (Figure A1-8) spectra, indicating that helix I, helix II, and helix III are folded in similar manner in PKI and PK1Δ56-70.

Nearly all base-paired imino resonances in PKI (Figure A1-5) and PK1Δ56-70 (Figure A1-7) were assigned, excluding helical termini that are rapidly exchanging with solvent. Sequential NOEs indicate formation of helix I, helix II, helix III, and helix IV in PKI (Figure A1-5). Imino resonances in helix II, helix III, and helix IV were unambiguously assigned. Helix I was ambiguously assigned due to the symmetrical G-U-U-G NOE pattern. The ambiguity in these assignments could be eliminated by reference to a single base-pair substitution in helix I.

In helix II, observation of the NOE cross-peak between G20-U22 (Figure A1-5) indicates that U21 is flipped out of the helix, allowing its neighboring base-pairs to stack. This conformation is consistent with reactivity levels obtained by SHAPE chemical probing [466] for U21 in the context of the PKI structure (personal communication E. Jan).

In addition to all expected NOEs within helix III given the originally proposed base-pairing (Figure A1-1A), two unexpected cross peaks were detected between G36-G37 and G37-U38 (Figure A1-5). These cross peaks are shifted upfield into a non-Watson-Crick imino region. The nucleotides within the helix III terminal loop are minimally reactive when probed by SHAPE [466] (personal communication E. Jan), suggesting that nucleotides are involved in base-paired interactions. Although the PKI helix IV NOE resonances are observable, the decreased intensity for many of these NOE cross peaks suggests that the helix may be only weakly formed in
solution. Resonances for U14 and G70 were observable in the $^1$H NMR spectrum (Figure A1-5) and in the $^1$H-15N HMQC (Figure A1-6); however these resonances were not visible in the NOESY spectrum indicating that they exchange with water during the 100 msec mixing time. Therefore, under these solution conditions, U14 and G70 are not involved in a stable base-pair. Their chemical shifts are diagnostic of U-G wobble pairs [490], which suggested that they may interact transiently with one another. SHAPE probing on the PKI RNA showed high and modest reactivity for U14 and G70 (personal communication E. Jan), consistent with transient base-pair formation. The stability of this base-pair in solution may affect the frequency of translation initiation in the +1 reading frame (Figure A1-1).
Figure A1-5 Secondary structure of PKI as determined by NMR.

1D $^1$H and 2D $^1$H-$^1$H NOESY NMR spectra of PKI in 20 mM potassium phosphate (pH 6.3), 200 mM KCl, 0.5 μM EDTA. Assignments and connecting lines are color-coded according to secondary structure, as in Figure 5-1. The NOE peaks between U69-G66, G2-U53, and G10-U11 are only visible at lower contour level and are therefore indicated with dashed circles. Base-pairs confirmed by $^1$H-$^1$H 2D NOESY are indicated in Figure A1-1B by black lines or circles.
Figure A1-6

Figure A1-6 $^1$H and $^{15}$N imino chemical shift assignments for PKI.

$^1$H-$^{15}$N HMQC spectrum of PK1 in 20 mM potassium phosphate (pH 6.3), 200 mM KCl, 0.5 μM EDTA. Resonance assignments for helix I (green), helix II (purple), helix III (blue), and helix IV (orange) are indicated. The G70 resonance is only visible at lower contour levels and is therefore indicated with a dashed circle.
Figure A1- 7

**Figure A1- 7 Secondary structure of the 55-nt PK1Δ56-70 RNA as determined by NMR.**

1D $^1$H and 2D $^1$H-$^1$H NOESY NMR spectra of PK1Δ56-70 in 20 mM potassium phosphate (pH 6.3), 200 mM KCl, and 0.5 μM EDTA. Assignments and connecting lines are color-coded according to secondary structure, as in Figure A5-1. The NOE peak between G2-U53 and G10-U11 is only visible at lower contour level and is therefore indicated with a dashed circle. Base-pairs confirmed by $^1$H-$^1$H 2D NOESY are indicated in Figure A1-1C by black lines or circles.
Figure A1-8 1H and 15N imino chemical shift assignments for PK1Δ56-70. 

\(^1\)H-\(^{15}\)N HMQC NRM spectrum of PK1Δ56-70 in 20 mM potassium phosphate (pH 6.3), 200 mM KCl, and 0.5 µM EDTA. Resonances associated with spin-pairs in helix I (green), helix II (purple), and helix III (blue) are indicated.
A1.3.3 Modeling the structure of PKI in solution

All-atom structural models of PKI were generated with MC-SYM software [337] based on the secondary structure determined by NMR. Six hundred and twenty nine models were filtered against SAXS data by comparison of the goodness-of-fit ($\chi^2$ agreement) between the experimental data for each model and the predicted small-angle X-ray scattering amplitudes. The top 25% of structures (based on lowest $\chi^2$ values) were selected for further filtering by NMR data. These models were tested for agreement with $^1$H-$^{15}$N RDC measurements. Models with a Q factor of less than 0.31 were accepted (Table A1-1), resulting in 5 models that approximate the NMR and SAXS data (Figure A1-9) and backbone RMSD of 8.2 Å. Although the RMSD between these five models was significantly smaller than the RMSD between 10 randomly selected models, $\sim$12 Å, further refinement of the models will be necessary to bring the RMSD to publishable levels ($\sim$1-3 Å). The predicted SAXS scattering profile for the five models diverged from the experimental data at low q values (less than 0.04). Low q values are reflective of molecular size. The radius of gyration ($R_g$) and maximum dimension ($D_{max}$) describing each of these models is consistently smaller (Table A1-1) than the experimentally determined $R_g$ (26.4 Å) and $D_{max}$ (90 Å). This difference is expected given the difference in predicted and experimental scattering intensity at low q values (Figure A1-9).

When the 5 model ensemble was fit into the PKI ab initio envelope (Figure A1-10) there was an extra area of density proximal to the single stranded linker, resulting from the large experimental $D_{max}$. If helix IV is only transiently forming in solution, the RNA may appear to be larger in size than is predicted (Table A1-1).
Figure A1-9 Filtering structural models by SAXS and NMR data.

(A) The 5 best models of 500 generated by MC-SYM were selected based on agreement with SAXS and RDC measurements. Experimental (black circles) and predicted (solid lines) scattering profiles of the unrefined models diverge considerably at \( q < 0.04 \). (B) Experimentally measured RDC measurements and calculated RDC values for the unrefined models with Q factors < 0.31 modestly agree.
Table A1 - 1 Filtering of structural models of PKI

<table>
<thead>
<tr>
<th>Model</th>
<th>RDC Q</th>
<th>RDC $R^2$</th>
<th>SAXS $\chi^2$</th>
<th>$R_g$ (Å)</th>
<th>$D_{\text{max}}$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.100</td>
<td>0.776</td>
<td>1.88</td>
<td>22.6</td>
<td>72</td>
</tr>
<tr>
<td>2</td>
<td>0.191</td>
<td>0.852</td>
<td>1.50</td>
<td>23.9</td>
<td>76</td>
</tr>
<tr>
<td>3</td>
<td>0.244</td>
<td>0.870</td>
<td>1.87</td>
<td>23.4</td>
<td>76</td>
</tr>
<tr>
<td>4</td>
<td>0.285</td>
<td>0.944</td>
<td>1.34</td>
<td>23.1</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>0.306</td>
<td>0.857</td>
<td>2.37</td>
<td>21.6</td>
<td>75</td>
</tr>
</tbody>
</table>
Figure A1-10 Preliminary model of the PKI structure filtered using SAXS and RDC measurements.

(A) The 5 best models had an overall backbone RMSD of 8.2 Å. Structural features are color-coded as in Figure A1-1B. Except for an extra area of density, the models agree well with the ab initio structure of PKI (grey). (B) The model with the lowest RDC Q value. Helices I, II, and IV form a coaxial stack.
A1.4 Conclusions and future directions

A1.4.1 Conclusions

A preliminary structural model of the 70 nucleotide IAPV IGR IRES PKI domain was determined using NMR and SAXS. The combination of SAXS and NMR to determine these models was critical, because structural determination by NMR alone for such a large RNA would be challenging, although possibly feasible (discussed in section 3.2.3). We present the experimentally determined secondary structure of PKI (Figure A1-5, Figure A1-6), which is in agreement with SHAPE chemical probing experiments on the IRES domain (personal communication E. Jan). While our structural models appear to agree with the structure of class I PKI domains [476], refinement of these models is necessary before any legitimate conclusions can be reached regarding the relationship between structure and function of this IRES domain.

A1.4.2 Future directions

*Molecular refinement of the PKI structure*

The top five PKI models will be refined against SAXS data, RDC measurements, and base-pairing restraints in XPLOR-NIH as previously described [273, 484, 485]. Distance restraints based on P-P distances between base-pairs in helices I-IV will be incorporated for helices I-IV. P-P distances (5.4 ± 0.5 Å) will be included for single-stranded regions to help maintain pseudo-A-form geometry during simulated annealing in the absence of other restraints. P-P envelope size restraints will be incorporated based on the overall size of the ab initio structure. Finally, base-pair restraints and RDC restraints will be incorporated on the basis of our NMR data.

*Structural investigation of IAPV IGR IRES +1 initiation*

The IAPV IGR IRES typically initiates translation on a non-AUG codon (GGC). We have developed a preliminary structural model of the PKI domain responsible for positioning the
GGC start codon in the A-site. Interestingly, the same domain initiates translation in the +1 frame on a GCG codon, which occurs with a 20% frequency [483]. Our collaborator, Eric Jan (University of British Columbia), has identified several point mutations that alter the relative frequency of +0 and +1 initiation. Two mutants are especially interesting. Deletion of U21 (Figure A1-1B) eliminates initiation on the GGC start codon and increased +1 initiation 3-fold. Second, deletion of A6 (Figure A1-1B) reduced +0 initiation without impacting +1 initiation. Importantly, ribosome positioning on the IRES was unchanged by both mutations, as inferred from toe-printing experiments (personal communication, E. Jan).

Similar to our work presented herein, NMR and SAXS could be used to investigate the differences in structure in PKIΔU21 and PKIΔA6 RNAs. Based on our preliminary structural models of PKI, deletion of U21 is not predicated to cause a significant change in overall pseudoknot fold (A1-10B). U21 deletion may impact initiation by removing a critical contact with the ribosome. Contrastingly, A6 is predicted to form an A:A wobble pair. Deletion of U6 could significantly alter the geometry of the three-helix junction (A1-10B).
Appendix II: HIV-1 frameshift site stem-loop domain motions in the presence of cobalt-hexamine and doxorubicin

The material presented in this chapter is the result of research performed by the author with contributions from Elizabeth A. Dethoff, Marco Tonelli, and Prof. Samuel E. Butcher.

E.A.D. contributed to RDC data analysis.

M.T. assisted with NMR data collection.

The remainder of the work was by K.D.M.
A2.1 Overview

In Chapter 4, the dynamic interhelical motions of the HIV-1 frameshift site stem-loop were investigated as a function of counter-ion concentration by NMR and order tensor analysis. Additionally, the fluorescent purine analogue 2-aminopurine (2AP) was utilized to probe base stacking at each position in the bulge in three different cation solution conditions (20 mM potassium, 150 mM potassium, and 20 mM potassium with 2 mM magnesium). The domain motions of this RNA were investigated in two additional solution conditions (20 mM potassium phosphate (pH 6.8) with 0.5 mM cobalt-hexamine or 1.35 mM doxorubicin). Results with these compounds were not included in the main data chapter for two reasons. First, no difference in dynamics was detected, and second, the results were likely influenced by compound solubility and thus, may not be reflective of actual changes in dynamics at saturating compound concentrations.

Cobalt-hexamine is an exchange-inert metal that mimics magnesium hexahydrate and is frequently used in structural studies [108, 338, 491-493]. However, its impact on RNA dynamics is unknown. Doxorubicin was identified in a high-throughput screen designed to find small molecules that bind the FS-SL [199]. Doxorubicin binds with low micromolar affinity to the FS-SL three-purine bulge and surrounding area [199].

It has been suggested that tertiary contacts and intermolecular interactions between ligands and RNA stabilize specific conformations that an RNA structure is predisposed to sample [220]. Until now, this idea of conformational trapping had not been investigated using ligands specific for the three-purine bulge motif within a dynamic RNA structure. Understanding doxorubicin’s impact on the FS-SL’s interhelical bend angle and domain motions would expand our understanding of conformational trapping.
A2.2 Materials and methods

A2.2.1 DNA template design and synthesis

Using the HIV-1 FS-SL as a starting model [108], the consensus sequence for the lower helix was modified for optimal alignment in Pf1 phage (Figure 4-1B). The synthetic oligonucleotide, (5’TTCTAATACGACTCACTATAGGCGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAAT CGCC-3’), and its compliment were purchased (Integrated DNA Technologies (IDT), Inc.) and utilized as a substrate for in vitro transcription.

A2.2.2 RNA synthesis and purification

RNA for NMR was transcribed in vitro using purified His6-tagged T7 RNA polymerase and synthetic DNA oligonucleotides (IDT), as previously described [108, 189, 199]. 13C/15N-labeled RNA samples were prepared using 13C/15N-labelled rNTPs. Target RNA was purified by denaturing (15%) polyacrylamide gel electrophoresis, identified by UV absorbance, and excised from the gel. RNA was recovered by diffusion into 0.3 M sodium acetate, precipitated with ethanol, purified on a High Q anion exchange column (Bio-Rad), again precipitated with ethanol, and desalted on a sephadex G-15 (Sigma) gel filtration column. The purified RNA was lyophilized, resuspended in 20 mM KH2PO4 and brought to pH 6.8 before exchange into specific solution conditions. NMR samples were referenced using 2 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Partial alignment of RNA for RDC measurements was achieved by adding Pf1 filamentous bacteriophage at a final concentration of ~10-15 mg/mL (ASLA Ltd., Riga, Latvia) to 13C/15N-labeled samples. RNA used in fluorescence-monitored titrations (Figure 4-1C) was purchased from Dharmacon (Thermo Scientific).
A2.2.3 NMR spectroscopy

RDCs were measured in two different solution conditions: 20 mM potassium phosphate and 2 μM sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) supplemented with either 0.5 mM cobalt-hexamine (1:1 RNA:compound) or 1.35 mM doxorubicin (1:1.5 RNA:compound), all at pH 6.8. RNA was exchanged into each solution by dialysis against 2L of buffer at 4° C. 13C-1H-S3A-HSQCs, 13C-1H TROSY-HSQC, and 15N-1H TROSY-HSQC experiments were used to collect scalar coupling values in non-aligned and partially aligned (~10-15 mg/ml buffer exchanged Pf1 filamentous bacteriophage (ASLA Ltd., Riga, Latvia)) samples. All NMR spectra were obtained on a Varian 900 MHz spectrometer at the National Magnetic Resonance Facility at Madison (NMRFAM). The spectrometer was equipped with a z-axis pulsed field gradient proton, carbon, and nitrogen cryogenically cooled probe.

A2.2.4 NMR analysis

RDCs were calculated by determining the difference in scalar coupling values between non-aligned and partially aligned samples (Table A2-1, Table A2-2). RDC errors were calculated using the RMSD for each bond type in both the direct and indirect dimensions and between duplicate experiments, as described [307].

RDCs measured in the Watson-Crick base-pairs in the upper and lower helices were subject to order tensor analysis using modeled A-form helices as described [307]. The use of idealized A-form helices for this analysis has been previously validated [308, 312, 313, 423-425]. Upper and lower helices were constructed using Insight (1998 version). A-form RNA parameters were checked using 3DNA [426]. All parameters were consistent with published A-form RNA characteristics, except for the propeller twist, which had the incorrect sign. The propeller twist was corrected in these structures by using an in-house program as described [307]. This program (HPMod_C6,8) was kindly provided by the Al-Hashimi group. RDCs (Supplementary
Tables 1-3) were fit to the corrected A-form helices and the order tensors determined by RAMAH software [308]. Terminal RDCs were excluded in the fits because of their departure from ideal A-form RNA characteristics [307]. To account for differences in alignment between 100% D2O and 90% H2O/10% D2O NMR experiments, the imino data collected in 90% H2O/10% D2O were uniformly scaled such that an optimal fit of the RDC data was achieved by RAMAH software [308], as previously described [391]. Order tensor errors due to A-form structural noise and RDC error were calculated using Aform-RDC [425], with the error input being the average RDC error. An extremely good order tensor fit was achieved for RDCs measured under all conditions. In all instances, the root-mean-square deviation (rmsd) compared favorably with the RDC measurement uncertainty (Table A2-3).

Order tensor solutions were used to rotate each A-form helix into its principal axis system (PAS) by using EULER software [307]. Once in the PAS, the helices were translated without rotation to satisfy connectivity restraints between the U6 and C7 nucleotides. Average A-form distances were used to satisfy the phosphodiester linkage requirement between C7-P and U6-O3' [427]. Specifically, the upper and lower helix were translated such that the U6-O3' atom was 1.58 Å from the C7-P atom, with a 102° O5'-P-O3' bond angle and a 62° dihedral angle about the O5'-P bond. Order tensor solutions have a $4^{n-1}$ degeneracy, where n is the number of helices, such that the once the helix is in its PAS, rotation by 180° about each principal axis would yield the same order tensor solution. To determine the correct helix orientation, two connectivity restraints were used. The first requires connectivity between C7-P and U6-O3', allowing elimination of 2 of the 4 solutions. Next, the distance between G32-O3' and A36-P, which must be smaller than or equal to the theoretically allowed length of 21 Å [307], eliminated the final degenerate solution. The interhelical bend angle and dynamic parameters GDO and η were calculated as previously described [307]. To specifically compare the motion of one helix
relative to the other, the general degree of order (GDO) for the dynamic helix was normalized by the GDO for the helix dominating the alignment, giving the internal GDO (GDO_{int}).
Table A2 - 1 HIV-1 FS-SL RDCs measured in 20 mM potassium and 0.5 mM cobalt-hexamine.

<table>
<thead>
<tr>
<th>Residue</th>
<th>RDC (Hz)</th>
<th>Error (Hz)</th>
<th>Residue</th>
<th>RDC (Hz)</th>
<th>Error (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>C1'H1'</td>
<td>9.47</td>
<td>0.35</td>
<td>G23</td>
<td>C1'H1'</td>
</tr>
<tr>
<td>G2</td>
<td>N1H1</td>
<td>-3.73</td>
<td>3.78</td>
<td>G23</td>
<td>C8H8</td>
</tr>
<tr>
<td>C3</td>
<td>C5H5</td>
<td>10.87</td>
<td>2.11</td>
<td>G23</td>
<td>N1H1</td>
</tr>
<tr>
<td>C3</td>
<td>C6H6</td>
<td>12.60</td>
<td>1.27</td>
<td>G24</td>
<td>N1H1</td>
</tr>
<tr>
<td>G4</td>
<td>C8H8</td>
<td>19.79</td>
<td>2.54</td>
<td>G24</td>
<td>C8H8</td>
</tr>
<tr>
<td>G4</td>
<td>N1H1</td>
<td>-6.39</td>
<td>0.09</td>
<td>A25</td>
<td>C2H2</td>
</tr>
<tr>
<td>A5</td>
<td>C2H2</td>
<td>19.34</td>
<td>1.42</td>
<td>A25</td>
<td>C8H8</td>
</tr>
<tr>
<td>A5</td>
<td>C8H8</td>
<td>21.15</td>
<td>0.64</td>
<td>A26</td>
<td>C2H2</td>
</tr>
<tr>
<td>A5</td>
<td>C1'H1'</td>
<td>10.59</td>
<td>0.24</td>
<td>G27</td>
<td>C8H8</td>
</tr>
<tr>
<td>U8</td>
<td>C1'H1'</td>
<td>-2.83</td>
<td>0.67</td>
<td>G27</td>
<td>N1H1</td>
</tr>
<tr>
<td>U8</td>
<td>C255</td>
<td>25.13</td>
<td>0.30</td>
<td>G28</td>
<td>N1H1</td>
</tr>
<tr>
<td>U8</td>
<td>N3H3</td>
<td>-5.84</td>
<td>0.53</td>
<td>C29</td>
<td>C5H5</td>
</tr>
<tr>
<td>G9</td>
<td>N1H1</td>
<td>-6.05</td>
<td>0.96</td>
<td>C30</td>
<td>C1'H1'</td>
</tr>
<tr>
<td>G10</td>
<td>N1H1</td>
<td>-6.69</td>
<td>0.33</td>
<td>C30</td>
<td>C6H6</td>
</tr>
<tr>
<td>G10</td>
<td>C1'H1'</td>
<td>-9.24</td>
<td>1.79</td>
<td>A31</td>
<td>C1'H1'</td>
</tr>
<tr>
<td>G10</td>
<td>C8H8</td>
<td>22.93</td>
<td>3.18</td>
<td>A31</td>
<td>C2H2</td>
</tr>
<tr>
<td>C12</td>
<td>C1'H1'</td>
<td>-13.72</td>
<td>1.58</td>
<td>A31</td>
<td>C8H8</td>
</tr>
<tr>
<td>U13</td>
<td>C5H5</td>
<td>23.39</td>
<td>2.42</td>
<td>U37</td>
<td>C5H5</td>
</tr>
<tr>
<td>U13</td>
<td>N3H3</td>
<td>-3.50</td>
<td>0.13</td>
<td>U37</td>
<td>N3H3</td>
</tr>
<tr>
<td>U14</td>
<td>C6H6</td>
<td>12.15</td>
<td>3.87</td>
<td>U37</td>
<td>C6H6</td>
</tr>
<tr>
<td>U14</td>
<td>N3H3</td>
<td>-4.00</td>
<td>0.24</td>
<td>C38</td>
<td>C6H6</td>
</tr>
<tr>
<td>U14</td>
<td>C5H5</td>
<td>21.51</td>
<td>2.06</td>
<td>C38</td>
<td>C5H5</td>
</tr>
<tr>
<td>C15</td>
<td>C5H5</td>
<td>19.74</td>
<td>1.23</td>
<td>G39</td>
<td>N1H1</td>
</tr>
<tr>
<td>C16</td>
<td>C1'H1'</td>
<td>-20.02</td>
<td>0.43</td>
<td>C40</td>
<td>C6H6</td>
</tr>
<tr>
<td>C16</td>
<td>C5H5</td>
<td>21.56</td>
<td>1.93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A2 - 2 HIV-1 FS-SL RDCs measured in 20 mM potassium and 1.3 mM doxorubicin.

<table>
<thead>
<tr>
<th>Residue</th>
<th>RDC (Hz)</th>
<th>Error (Hz)</th>
<th>Residue</th>
<th>RDC (Hz)</th>
<th>Error (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>N1H1</td>
<td>-3.46</td>
<td>G23</td>
<td>C1'H1'</td>
<td>-20.58</td>
</tr>
<tr>
<td>C3</td>
<td>C6H6</td>
<td>19.39</td>
<td>G23</td>
<td>C8H8</td>
<td>37.64</td>
</tr>
<tr>
<td>C3</td>
<td>C1'H1'</td>
<td>20.26</td>
<td>G23</td>
<td>N1H1</td>
<td>-11.04</td>
</tr>
<tr>
<td>C3</td>
<td>C5H5</td>
<td>23.43</td>
<td>G24</td>
<td>N1H1</td>
<td>-8.91</td>
</tr>
<tr>
<td>G4</td>
<td>C8H8</td>
<td>32.88</td>
<td>G24</td>
<td>C8H8</td>
<td>34.39</td>
</tr>
<tr>
<td>G4</td>
<td>C1'H1'</td>
<td>16.65</td>
<td>A25</td>
<td>C2H2</td>
<td>35.54</td>
</tr>
<tr>
<td>A5</td>
<td>C1'H1'</td>
<td>19.35</td>
<td>A26</td>
<td>C2H2</td>
<td>21.86</td>
</tr>
<tr>
<td>A5</td>
<td>C8H8</td>
<td>39.43</td>
<td>G27</td>
<td>C8H8</td>
<td>30.86</td>
</tr>
<tr>
<td>U8</td>
<td>N3H3</td>
<td>-10.07</td>
<td>C38</td>
<td>C5H5</td>
<td>-3.87</td>
</tr>
<tr>
<td>G9</td>
<td>N1H1</td>
<td>-11.30</td>
<td>G39</td>
<td>C1'H1'</td>
<td>-16.50</td>
</tr>
<tr>
<td>G10</td>
<td>C8H8</td>
<td>22.26</td>
<td>C40</td>
<td>C5H5</td>
<td>38.61</td>
</tr>
<tr>
<td>G10</td>
<td>C1'H1'</td>
<td>-20.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U13</td>
<td>N3H3</td>
<td>-8.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U14</td>
<td>N3H3</td>
<td>-5.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U14</td>
<td>C5H5</td>
<td>46.62</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A2.2.5 Fluorescence-monitored titrations

Fluorescence measurements were performed in triplicate similarly to those previously described [199]. Briefly, the 2AP of HIV-1 G33-2AP, HIV-1 G34-2AP, HIV-1 A35-2AP, or 2AP-NTP was excited at 309 nm and emission was measured at 360 nm with an 800V detector. Cobalt-hexamine and doxorubicin were titrated in triplicate from 5x10^{-8} M to 0.001 M, into 2 μM RNA (constant) in 10 mM HEPES (pH 7.0). Using a 160 μL sample cell, fluorescence was measured for 10 seconds at 30 °C with a QuantaMaster Model C-60/2000 Spectrofluorimeter. Fluorescence was recorded at 0.5 nM, 100 nM, 240 nM, 480 nM, 950 nM, 1.9 μM, 3.7 μM, 9.9 μM, 35 μM, 74 μM, 190 μM, 370 μM, and 970 μM titrant concentrations. Owing to significant levels of non-specific fluorescence quenching (data not shown), this method could not be used to examine nucleotide stacking in the bulge with doxorubicin or cobalt-hexamine.
A2.3 Results

In 20 mM potassium, the FS-SL was found to adopt a 43° interhelical bend and undergo anisotropic interhelical motions (presented in Chapter 4). Addition of cobalt-hexamine or doxorubicin to the FS-SL did not cause significant changes in its interhelical bend angle or domain motions (Table A2-3). The compound concentrations chosen in these experiments were dictated by the solubility of cobalt-hexamine and doxorubicin in solution. We hypothesize that our results might change significantly if we were able to reach a concentration where the FS-SL was fully bound by doxorubicin or saturated by cobalt-hexamine.
Table A2 - 3 Order tensor analysis of RDCs in the HIV-1 FS-SL RNA

<table>
<thead>
<tr>
<th>Cation</th>
<th>Helix</th>
<th>N</th>
<th>CN</th>
<th>rmsd (Hz)</th>
<th>$R^2$</th>
<th>$\eta$</th>
<th>GDO x 10$^{-3}$</th>
<th>GDOint</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 mM K$^+$</td>
<td>Upper</td>
<td>24</td>
<td>3.3</td>
<td>3.5</td>
<td>0.99</td>
<td>0.12 ± 0.05</td>
<td>1.67 ± 0.05</td>
<td>0.67 ± 0.03</td>
<td>43° ± 2°</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>17</td>
<td>2.8</td>
<td>1.9</td>
<td>0.99</td>
<td>0.31 ± 0.04</td>
<td>1.12 ± 0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 mM K$^+$, 1.35 mM Doxorubicin</td>
<td>Upper</td>
<td>15</td>
<td>3.3</td>
<td>3.8</td>
<td>0.99</td>
<td>0.04 ± 0.05</td>
<td>2.3 ± 0.1</td>
<td>0.61 ± 0.05</td>
<td>39° ± 3°</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>11</td>
<td>2.7</td>
<td>1.7</td>
<td>0.99</td>
<td>0.33 ± 0.02</td>
<td>1.4 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 mM K$^+$, 0.5 mM cobalt-hexamine</td>
<td>Upper</td>
<td>23</td>
<td>2.3</td>
<td>2.4</td>
<td>0.98</td>
<td>0.11 ± 0.03</td>
<td>1.18 ± 0.07</td>
<td>0.67 ± 0.09</td>
<td>38° ± 5°</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>12</td>
<td>3.6</td>
<td>2.3</td>
<td>0.96</td>
<td>0.44 ± 0.09</td>
<td>0.8 ± 0.1</td>
<td>± 0.07</td>
<td></td>
</tr>
</tbody>
</table>

Shown for each helical domain is the number of RDCs (N), the condition number (CN), the root mean square deviation (rmsd), the correlation coefficient ($R^2$) between measured and back-calculated RDCs, the order tensor asymmetry ($\eta = |S_{yy} - S_{xx}| / S_{zz}$), the generalized degree of order (GDO = $\frac{2}{3}(S_{xx}^2 + S_{yy}^2 + S_{zz}^2)$, $S_{xx} \geq S_{yy} \geq S_{zz}$), the internal generalized degree of order (GDOint = GDOI/GDOj; GDOI < GDOj), and the internhelical bend angle ($\theta$). All values were generated by RAMAH software with errors estimated using the program AFORM-RDC.
Appendix III: Structure of the HIV-1 frameshift site three-helix junction

The material presented in this appendix was a result of research performed by the author with contributions from Shruti Waghray and Prof. Samuel E. Butcher.

S.W. assisted with a preliminary screen of 3HJ folding conditions.

The remainder of the work was by K.D.M.
A3.1 Overview

Within viral capsids, the HIV-1 frameshift site RNA is part of a conserved 3HJ secondary structure (Figure 2-9A) [170, 171]. It has been hypothesized that the role of this secondary structure is to slow down the rate of translation [170], which in turn may modulate frameshift efficiency. We found that the conserved 3HJ secondary structure in the HIV-1 genomic RNA [170] causes a significant decrease in frameshift efficiency (Figure 2-9). Here, a preliminary investigated of the 3HJ secondary structure by NMR is presented.
A3.2 Materials and Methods

A3.2.1 Plasmid Construction

The three-helix junction plasmid DNA was prepared through phosphorylation and ligation of short complementary, overlapping oligonucleotides (Integrated DNA technologies) with EcoR1 and BamH1 compatible ends into pUC19 vector (New England Biolabs). A BasI restriction site was included at the end of the template to allow for run-off transcription after digestion with BsaI enzyme (NEB).

A3.2.2 RNA Synthesis and purification

Milligram quantities of RNA were transcribed in vitro by purified His6-tagged T7 RNA polymerase, linearized plasmid DNA (3HJ WT and 3HJ*) or synthetic DNA oligonucleotides (3HJ P2) (IDT), and ribonucleotides (Sigma-Aldrich), as previously described [199]. DNA template sequences are shown in Table A3-1. Briefly, RNA was purified by denaturing 8% (3HJ) or 10% (SipperySL) polyacrylamide gel electrophoresis (PAGE) with 8M urea. Impurities were removed by DEA anion exchange chromatography (Bio-Rad) using a low salt buffer (20 mM Tris-HCL, pH 7.6, 200 mM sodium chloride) to wash and a high salt buffer (20 mM Tris-HCl, pH 7.6, 1.5 M sodium chloride) to elute the RNA. Samples were then ethanol precipitated, and salt was removed using a Sephadex G-15 column (Sigma-Aldrich). Purified 3HJ and 3HJ* RNAs were flash frozen, lyophilized, resuspended in NMR buffer (20 mM potassium phosphate, pH 6.8) to ~50 µM, incubated at 65 °C for 5 minutes, and immediately incubated on ice to promote monomeric folding. Purified 3HJ P2 RNA was flash frozen, lyophilized, and resuspended in NMR buffer to 1 mM. The RNAs were concentrated to 300 µL using Amicon spin-filters (Milipore), and dialyzed for 16-24 hr against 2L of NMR buffer. RNA integrity and homogeneity were checked with denaturing and non-denaturing PAGE, respectively.
Table A3- 1 DNA template sequences used for run off transcription

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Puc19</strong></td>
<td><strong>3HJ WT</strong></td>
</tr>
<tr>
<td></td>
<td>GATCCGGTCTCTGGCTGTGGAAATGTGGAAAGGACACCAATATGAAAGATTGTACTGAGAGACAGGCTAATTTTAGGGAAGATCTGGCCTTCCCACAAGGGAGGCCAGGGAATTTTCTTCAGAGCAGCCAGGACACAGCCACAGCCCTATAGTGAGTCGTATTAGAAG</td>
</tr>
<tr>
<td><strong>Puc19</strong></td>
<td><strong>3HJ</strong></td>
</tr>
<tr>
<td></td>
<td>GATCCGGTCTCTGGAAAGACACCAATATGAAAGATTGTACTGAGAGACAGGCTAATTTCAGAAGATCTGGCCTTCCCACAAGGGAGGCCAGGGAATTTTCTTCAGAGCAGCCAGGACACAGCCCTATAGTGAGTCGTATTAGAAG</td>
</tr>
<tr>
<td><strong>3HJ</strong></td>
<td><strong>P2</strong></td>
</tr>
<tr>
<td></td>
<td>GGTAAAAATTAGCCTGTCTCAGTAACATCTTTTACCTATAGTGAGTCGTATTAGAA</td>
</tr>
</tbody>
</table>

A3.2.3 NMR Spectroscopy

Owing to limited transcription efficiency of the full length 3HJ, two smaller RNA constructs, 3HJ* and 3HJ P2, were utilized to examine 3HJ secondary structure by NMR. All NMR spectra were obtained on a Bruker Avance DMX 750 MHz spectrometer at the National Magnetic Resonance Facility at Madison (NMRFAM). The spectrometer was equipped with a single z-axis gradient proton, carbon, nitrogen cryogenically cooled probe. Spectra were acquired in 90% H$_2$O / 10% D$_2$O, 20 mM potassium phosphate buffer, pH 6.8 at 283 K. Base-pairing was inferred using 2D $^1$H$^1$H NOESY experiments. All data were processed using TopSpin (Bruker BioSpin) software, and assignments were made using Sparky (http://www.cgl.ucsf.edu/home/sparky/).
A3.3 Results

A3.3.1 The 3HJ does not stably form in solution

Three constructs were developed for NMR: WT 3HJ, the full length RNA, 3HJ*, a P1 truncation of WT 3HJ, and 3HJ P2, the minimal P2 domain (Figure A3-1). Resonances were assigned for 3HJ* and 3HJ P2 using ¹H-¹H 2D NOESY experiments with a mixing time of 100 msec in 20 mM potassium phosphate (pH 6.8) and 2 µM DSS, at 10 °C (Figure A3-2). The minimal 3HJ P2 domain was utilized to facilitate identification of NOEs in the 3HJ* P2 helix. While base-pairing was detected in the 3HJ* P1 and P3 helices, base-pairing in 3HJ* P2 was not detected (Figure A3-2, Figure A3-3). From this preliminary NMR analysis we conclude that the 3HJ* structure does not stably form in solution.
Figure A3-1 Three-helix junction (3HJ) constructs used for NMR.

Three constructs were developed for NMR: WT 3HJ, the full length RNA, 3HJ*, a P1 truncation of WT 3HJ, and 3HJ P2, the minimal P2 domain. Base-pairing shown is predicted based on SHAPE probing of the HIV-1 virion RNA [170, 171]. Sequence changes are bold and italicized.
Figure A3-2 Overlay of the 3HJ* and 3HJ P2 $^1$H-$^1$H 2D NOESY and $^1$H 1D experiments.
The 3HJ* and 3HJ P2 spectra are shown in black and grey, respectively. NOE connectivities are shown in red (3HJ P2) and blue (3HJ*). The NOE peak between U48-U47 was only visible at lower contour level and was therefore indicated with a dashed circle. Non-native G iminos in the 3HJ P2 are designated G1* and G2*. NOEs detected in the 3HJ* (blue) and 3HJ P2 (red) are indicated. NOEs characteristic of the base-pairing in the 3HJ P2 helix were not apparent in the context of the larger 3HJ*. However, very weak imino resonances in the 1D $^1$H spectra indicated that the P2 helix may be forming weakly in 3HJ*.
Figure A3-3 Base-pairing inferred from $^1$H-$^1$H NOESY experiments.
References

42. Urbonavicius, J., Q. Qian, J.M. Durand, et al. (2001). Improvement of reading frame maintenance is a common function for several tRNA modifications. The EMBO journal. 20, 4863-73.


236


