

THE MINIMAL PRIMARY STRUCTURES OF RNA APTAMERS SELECTED TO BIND  
HIV-1 REVERSE TRANSCRIPTASE

---

A Thesis

presented to

the Faculty of the Graduate School  
at the University of Missouri-Columbia

---

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

JOSHUA D. FRANKEN

Dr. Donald Burke, Thesis Supervisor

© 2010

© Copyright by Joshua D. Franken 2010

All Rights Reserved

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

THE MINIMAL PRIMARY STRUCTURES OF RNA APTAMERS SELECTED TO BIND  
HIV-1 REVERSE TRANSCRIPTASE

presented by Joshua D. Franken,

a candidate for the degree of master of biochemistry,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Donald Burke

---

Professor Brenda Peculis

---

Professor Charlotte Phillips

---

Professor Stefan Sarafianos

---

.....Thanks, to my family and the Bullpen.

## ACKNOWLEDGEMENTS

This work was made possible by Dr. Donald Burke, Professor of Biochemistry and Molecular Microbiology and Immunology at the University of Missouri. A debt for this work also goes out to the University of Missouri and its Department of Biochemistry and Department of Molecular Microbiology and Immunology. Funding for this work was provided by the NIH.

Thank you to my committee members, Dr. Donald Burke, Dr. Brenda Peculis, Dr. Charlotte Phillips, and Dr. Stefan Sarafianos, for guiding me through the process of research and thesis preparation and defense. I would also like to thank everyone I worked with, including but not limited to, Dr. Elisa Biondi, Dr. Maggie Lang, Dr. Mark Ditzler, Ms. Terri Lyddon, Mr. Jin-Geol Kim, and Ms. Angela Whatley. You have all proven instrumental in the development of my graduate career.

Finally, I would like to thank my graduate class. We started together and though we will end separately and all follow our own paths, you made my first year of graduate study unforgettable and a little more fun. Here's to Ying Wan, Sha Zhu, Sara Drenkhan, Kirby Swatek, and Maura Bates. Good luck to you all!

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	ii
LIST OF TABLES AND FIGURES.....	iv
ABSTRACT .....	vi
Chapter	
1. INTRODUCTION .....	1
2. RESULTS AND DISCUSSION .....	16
80.104 .....	22
80.62 .....	36
80.33 .....	44
70.60 .....	49
70.07 .....	54
80.89 .....	57
80.55 .....	61
3. CONCLUSIONS.....	65
4. MATERIALS AND METHODS.....	73
REFERENCES .....	79
VITA.....	82

## LIST OF TABLES AND FIGURES

Figures	Page
1. SELEX procedure.....	3
2. Family I, Family II, and Family III motifs.....	9
3. HIV phylogenetic tree.....	11
4. RT crystal structure as determined by Sarafianos.....	12
5. RT activity assays.....	18
6. Secondary structure predicted for 80.104 by mfold.....	24
7. DDDP inhibition by 80.104.....	25
8. Boundary determination of 80.104 and predicted pseudoknot.....	28
9. Binding studies of 80.104 to RT.....	30
10. Native gel analysis of structure population of aptamers.....	32
11. Secondary structure by mfold and DDDP inhibition for 80.62.....	37
12. Binding studies of 80.62 to RT.....	38
13. S1, V1, T1 digestion of 80.62.....	40
14. DDDP inhibition by 80.33.....	45
15. Binding studies of 80.33 to RT.....	47
16. DDDP inhibition by 70.60.....	50
17. Binding studies of 70.60 to RT.....	51
18. Binding studies of 70.07 to RT and pseudoknot prediction.....	55
19. Binding studies of 80.89 to RT and S1, V1, T1 digestion.....	58
20. 3' boundary experiment and published pseudoknot of 80.55.....	62
21. 3' boundary study of various aptamers.....	63
22. Flow chart for determination of aptamer primary and secondary structure.....	70

Tables	Page
1. 70N Family I sequences.....	7
2. 80N Family I sequences.....	8
3. Full length and truncated aptamer sequences .....	23
4. Summary of results .....	66



# THE MINIMAL PRIMARY STRUCTURES OF RNA APTAMERS SELECTED TO BIND HIV-1 REVERSE TRANSCRIPTASE

Joshua D. Franken

Dr. Donald Burke, Thesis Supervisor

## ABSTRACT

Human Immunodeficiency Virus (HIV) reverse transcriptase (RT) is the most common molecular target of current HIV treatments. Oligonucleotide aptamers bind and inhibit the RNA- and DNA-dependent polymerization activities of HIV RT. Libraries consisting of aptamers including 32, 70 or 80 nucleotide variable regions were previously screened by Systematic Evolution of Ligands by Exponential Enrichment (SELEX) against RT. Roughly half of the resulting aptamers were represented by pseudoknots with well defined signature sequences (the Family I), but also additional pseudoknots with little sequence convergence (Family II), and non-pseudoknot aptamers (Family III).

Nucleic acid aptamers bind RT in the primer/template binding site. Aptamers are generally non-toxic and non-immunogenic molecules making them enticing drug prospects. Many aptamers inhibit DNA dependent DNA polymerization by RT from several phenotypically different recombinant viruses, but inhibition depends on a single amino acid mutation at position 277 for other aptamers. Aptamers that are un-reactive to the identity of this amino acid represent a group which may inhibit RT from other viruses as well.

In this work I present seven aptamers with unknown secondary structure which in some cases inhibit polymerization activity by RT from two HIV subtypes with different polymorphisms at position 277. I identify the minimal primary structure containing a pseudoknot in many of these aptamers sequences. I present evidence that in at least one aptamer, the structure responsible for binding RT is not a pseudoknot, which is highly uncommon in RNA anti-RT aptamers. For at least one other aptamer I show that a pseudoknot is the important binding element, and binding is increased in the presence of additional flanking sequence.

## **Chapter 1: Introduction**

Nearly 40 million humans are infected with HIV around the globe and more than 4 million new infections occur every year (1). Highly Active Anti-Retroviral Therapy (HAART), the current standard of care for HIV/AIDS patients in developed countries, has decreased the mortality and morbidity of patients infected with HIV-1 drastically. The lives of these patients are prolonged and infection is converted to a chronic, but manageable disease. However, the current antiviral drugs put selective pressure on the virus to evolve drug-resistant viral strains and the treatments do not eradicate the virus from the host. Additionally, prolonged treatment can cause side effects that are mild to serious in nature (2). Finally, HAART is not readily available in all parts of the world. Of these aspects of HIV infection, I am most interested in therapeutics that potentially overcome the evolution of drug resistant strains. It remains imperative that new therapeutics and strategies for treatment of HIV/AIDS are developed.

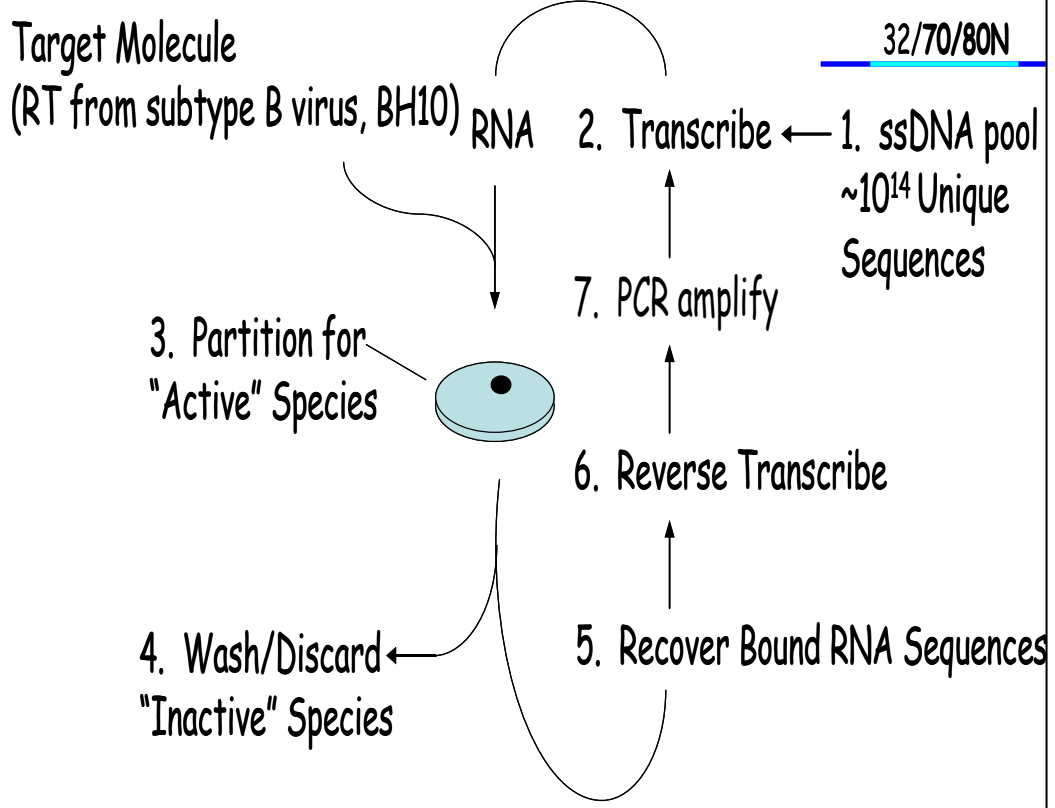
Nucleic acid aptamers are short DNA or RNA oligonucleotides generated by SELEX (Systematic Evolution of Ligands by Exponential Enrichment) (3) from a synthetic library. SELEX is a method for isolating high affinity ligands to a specific target(s) from a large pool of variable oligonucleotides. The basic procedure for selecting RNA aptamers begins with production of a large library of single-stranded DNA oligonucleotides. The primary structure of these molecules includes a fixed 5' region containing sequence important for annealing the cDNA to the 5' PCR primer as well as sequence that signals initiation of transcription by T7 RNA polymerase. There is also a fixed 3' region for annealing the cDNA to the 3' PCR primer. Between the fixed regions

is a variable region that accounts for the diversity in the oligonucleotide library. The library is transcribed into RNA and the RNA is incubated with the target molecule(s), often on a column or nitrocellulose filter. Aptamers with affinity for the target bind the filter and the other oligonucleotides are removed in a wash step. The bound aptamers are then recovered from the partitioning device, transcribed into DNA, and PCR amplified to produce a library enriched for binding the target molecule. The new pool is cycled through the affinity selection again, and this is continued until affinity of the pool for the target reaches a plateau (Figure 1, provided by Dr. Donald Burke).

Aptamers derived from SELEX have been shown to recognize and bind target molecules as well as, or better than, monoclonal antibodies. Dissociation constants ( $K_d$ ) for aptamer target binding have been measured in the nM (4, 5) and even the pM ranges (4, 6, and 7). Aptamers are candidates for diagnostic and therapeutic applications. One aptamer derived by SELEX binds vascular endothelial growth factor (VEGF), inhibits binding of VEGF<sub>165</sub> to KDR and Flt-1 VEGF receptors, reduces choroidal neovascularization, and is approved by the FDA for treatment of age-related wet macular degeneration (4). Aptamers selected for binding HIV-1 reverse transcriptase (RT) protein (8, 9) have been found to bind RT and demonstrate anti-RT activity *in vitro* (10). RNA aptamers have also been shown to reduce RT activity and HIV-1 infectivity *in vivo* (11, 12). These facts have led to this work investigating the structure of RNA aptamers that bind to HIV RT.

The RT protein from HIV-1 is the most common target of HAART and seventeen of thirty-two FDA approved drugs are inhibitors of RT ([www.fda.gov/](http://www.fda.gov/)). The two primary classes of RT inhibitors are nucleoside analog RT inhibitors (NRTIs), which are

**Figure 1:** The process of SELEX. Step 1 is the production of a library of unique DNA sequences. The upper right contains a depiction of a member of the DNA library. Dark blue represents regions of fixed primary sequence important for PCR amplification and transcription, while cyan represents the variable region that accounts for the  $10^{14}$  unique sequences. Previous anti-RT selections have incorporated variable regions of 32, 70, and 80 nucleotides and both fixed and variable regions can play roles in affinity for the target. The aptamers presented in this work originated from libraries including  $10^{14}$  unique sequences and contained variable regions of either 70 or 80 nucleotides. Step 2 is transcription of DNA into RNA. The RNA is then partitioned for the molecules with affinity for the target in step 3. In this case the RT came from a subtype B virus, BH10. Partitioning occurred via binding the protein to a nitrocellulose filter and binding the RNA to the protein. In step 4 the molecules that do not bind the target are washed away and step 5 is recovery of the bound RNA sequences. The RNA is reverse transcribed to DNA in step 6 and finally PCR amplified in step 7 to produce a new library with fewer unique sequences but increased overall affinity for the protein target. Steps 2 through 7 can then be looped in a cycle to further increase the affinity of the library by removing more and more lower affinity molecules.



mostly chain terminators, and non-nucleoside RT inhibitors (NNRTIs), which act by allosteric inhibition. RNA and DNA aptamers that bind and inhibit RT are referred to as TRTIs (template/primer analog RT inhibitors) because many of these aptamers compete with template/primer for access to enzyme binding (13, 14, and 15). Furthermore, aptamers have been found to inhibit RNA dependent DNA polymerization (RDDP) at levels comparable to NRTIs. The concentration of inhibitor required to achieve fifty percent of the maximum inhibition ( $IC_{50}$ ) values for aptamers and aptamer mutants ranged from 1.4 to 82  $\mu$ M while  $IC_{50}$  values for NRTIs ranged from 0.1 to 4.37  $\mu$ M (16). Synergistic effects have also been observed when inhibiting DNA Dependent DNA Polymerization (DDDP) using aptamers and NRTIs together (13). Although DNA and RNA anti-RT aptamers are currently useful research tools, anti-RT aptamers are also promising as therapeutics in the future. Unlike current HAART medication, nucleic acids pose little cellular toxicity threat and they are also not likely to cause an immune response (17 and 18). Additionally, clinical application is likely to employ RNA aptamers in the context of gene therapy. For example, genes that direct expression of a particular anti-RT RNA aptamer could be transduced into  $CD34^+$  hematopoietic progenitor cells as has been performed previously using three other anti- HIV RNA elements by Li et al. (19).

Aptamers are not the only nucleic acid compounds being investigated as anti-HIV therapeutics. Small hairpin RNA and small duplex RNA (shRNA and siRNA) utilize cellular machinery in a process known as RNA interference (RNAi) and have proven effective against HIV-1 (20-24). It has been proposed that use of an intrinsic mechanism

which could amplify siRNA introduced into a cell is a potential benefit of RNAi (25). It was a problem though that most RNAi studies had been performed at low multiplicity of infection (moi) of one or less because this does not represent the true environment in which infection occurs. Cells in an infected individual are likely to undergo many infection events. Also, it has also been shown that shRNAs are effective at inhibiting HIV-1 replication at low moi of one, but not a higher moi of 50 (25). RNA aptamers can block HIV-1 infection across this moi range as well as inhibiting HIV through two successive rounds of infection (25). Furthermore, anti-RT aptamers expressed in T-cells protected against high-dose infection by chimeric RT-SHIV viruses (25). The advantages aptamers have over RNAi may actually be due to “their lack of dependence on cellular machinery and their encapsidation into virion particles” (25). These advantages make aptamers the most promising oligonucleotide therapeutic.

Using the process of SELEX, both Tuerk (1992) and Burke (1996) selected hundreds of unique RNA aptamers that bind HIV-1 RT (8, 9). Most of these aptamers contain a sequence, which indicates the presence of a pseudoknot, a structure in which the aptamer forms a stem-loop and pairs nucleotides in the loop with nucleotides in a proximal sequence. The original selection by Tuerk (8) was found to contain only RT aptamers that conformed to the Family I pseudoknot (8, 26), which begins with the sequence UCCG. Green et al. described the sequence as UCCS (S equals C or G, 26). Subsequent selections by Burke against RT from the same protein preparation used in the selection by Tuerk, produced over 2000 aptamers in the enriched pool, of which 169 total were sequenced, and 136 unique sequences were found (9). The important difference in these selections was that Tuerk utilized a variable region 32 nucleotides long and Burke

utilized variable regions of 70 and 80 nucleotides in length. Greater than half of the aptamers from Burke's selection contained sequence which indicated they could form a Family I pseudoknot (Table 1 and 2, and Figure 2). Other aptamers contained sequence indicating they could form a pseudoknot that does not fall under the original constraints. Any RNA aptamer which forms a pseudoknot, but does not fit the stringent criteria of Family I pseudoknots, including nucleic acid identity, stem, loop, and linker length, was termed Family II. A third group of aptamers includes all aptamers that do not fit the description of Family I or II, and are termed Family III.

RT performs three main functions in the HIV life cycle. RDDP and DDDP are necessary to convert the viral RNA genome into double stranded DNA for incorporation into the host genome, while RNase H activity degrades the viral RNA after it has been reverse transcribed. Potential aptamer therapies should have some biologically relevant effect, such as inhibition of enzymatic function, pertaining to one or all of the RT functions. Furthermore, there are many diverse subtypes of HIV and potential therapies need to inhibit RT from the most common subtypes. Held et al. characterized inhibition of DDDP, RDDP, and RNase H activity by several pseudoknot aptamers against a panel including "10 recombinant RT's from phylogenetically diverse isolates" (27).

Family I aptamers inhibited RT from subtypes B and C well, but not from any other subtype including subtype A. The three subtypes listed are the three most common together making up over 86% of the world's infections. Family II aptamers inhibited RT from a broader set of subtypes, including HIV-2 and SIV<sub>cpz</sub>, to varying degree. Inhibition by Family I aptamers also depended on the amino acid identity of the RT at position 277. The existence of a lysine (K) at this position provided the RT resistance to inhibition and

**Table 1:** Unique 70N DNA sequences that contain Family I pseudoknots. Pseudoknots are in bold and stems are underlined. The sequences are aligned to accentuate the pseudoknot. Capital letters represent nucleotides from the variable region of the aptamer while under-case letters represent nucleotides in the constant region. Nucleotides outside of pseudoknot structure that are conserved are highlighted in yellow.

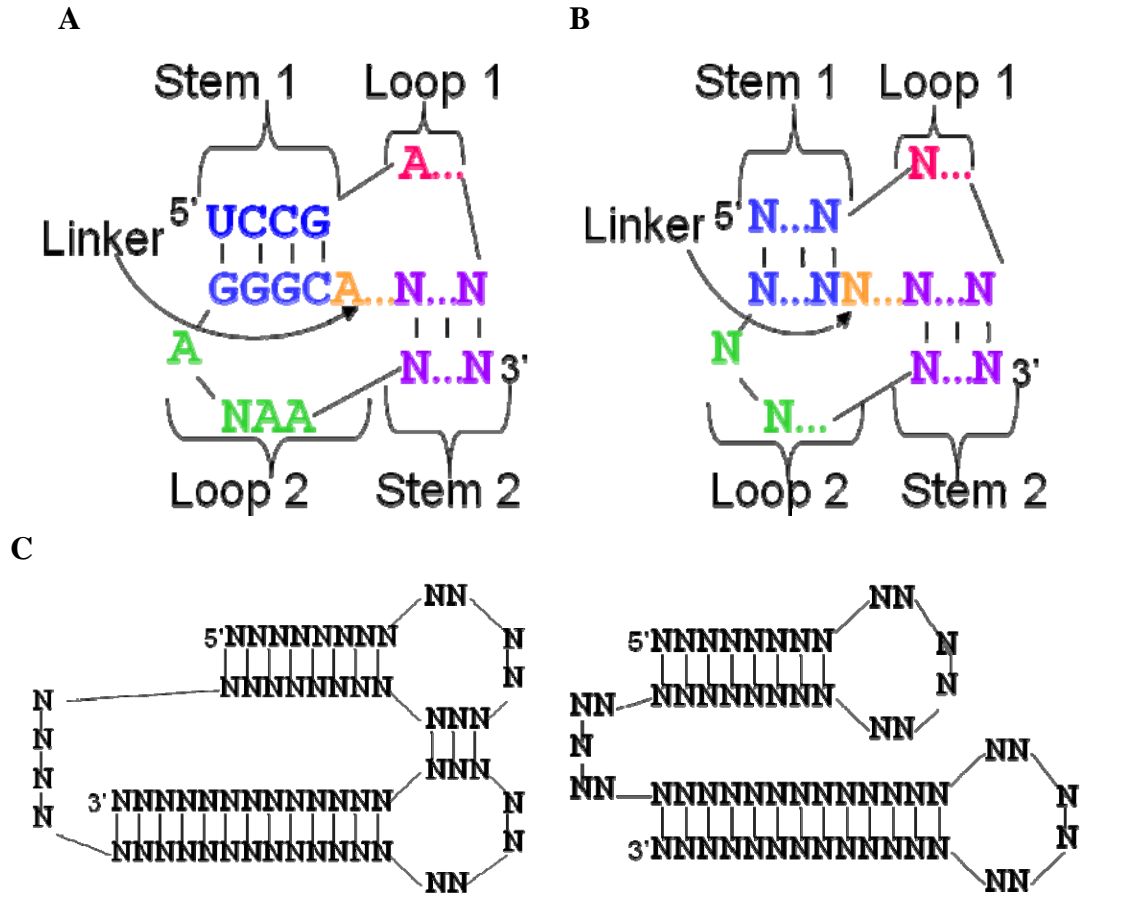
70.3	gggaaaaggtaagtcatacaacaaga	AGATGAATCAGCCCTCAACGAGATGCG	AAGGAAAA	CGGG	ATAACC	AAATAGATGGAA	CGAAGACC	ggggcataaag...
70.76a		AGATGAATCAGCCCTCAACGAGATGCG	AAGGAAAA	CGGG	ATAACC	AAATAGATGGAA	CGAAGACC	
70.50	gggaaaaggtaagtcatacaacaaga	TGAAAATGACACCAAGAAAATCCG	AGGTGATAACGGGAAAAACAC	AGAGAAA	CAATAAC	AGTTGCAAGTC	ggggca...	
70.64	gggaaaaggtaagtcatacaacaaga	TTGAAAAACCGAATGATCCG	ATAAGCAA	CGGG	AGAATGCTTAAAGATGGATAC	CCCGATGCTCACGg...		
70.9	gggaaaaggtaagtcatacaacaaga	GTGGAGAAATCGCCTATAATAACCCAGAAATCGATCCG	AAGTGA	CGGG	AGAAATCACCGACTCGAT	Cgggcataaaggtatttaa...		
70.11	gggaaaaggtaagtcatacaacaaga	CCTCGGTCCAGCAATCCG	AAGGAAAA	CGGG	ATAACCAACACACAAATGAAAACAGCCAGTAAGCGT...			
70.48	gggaaaaggtaagtcatacaacaaga	TGAACGTACGACGTAGACCAAGATCCG	AAGATGCA	CGGG	AGAAGCATCAATACCAAGTATGGCCgggcataaaggt...			
70.70	gggaaaaggtaagtcatacaacaaga	TGAACGTACGACGTAGTACCAAGATCCG	AAGATGCA	CGGG	AGAAGCATCAATACCAAGTATGGCC			
70.12	gggaaaaggtaagtcatacaacaaga	TATCCG	AGCCAAAA	CGGG	AAAAGATGGAAAAATGGAATTAGTCTCTAAACCCAA...			
70.21	CTGTGCTGGCTAGACTATCCAGAAAGTTGAAATCCG	AGCGAAA	CGGG	AAAAGATGGAAAAATGGAATTAGTCTCTAAACCCAA...				
70.27	gggaaaaggtaagtcatacaacaaga	CCCTCCTGTGATGACGCTAATCGAGATCCG	AAGTCCAA	CGGG	AGAAAGGACACTTATGACGTGGCGGGgggcataaaggt...			
70.17	gggaaaaggtaagtcatacaacaaga	CCCTCCTGTGATGACGCTAATCGAGAAATCCG	AAGTCCAA	CGGG	AGAGAGGACACTTATGACGGCGCGG			
70.26	gggaaaaggtaagtcatacaacaaga	TGCTAAACCAAGTAAGAAATCCG	TGMACTCACAGCAGGATAAACTGTG	TCAAAAC	CGCCATAGCT			
70.69	gggaaaaggtaagtcatacaacaaga	AGCGGGACCCAAATCGAAATCCG	AAGCGAA	CGGG	AGAAAGGACCAAGATACCTGTGAAATGGCGgggcat...			
70.30	gggaaaaggtaagtcatacaacaaga	CTAACCGCCATCCTAGCGATCGACAAACAGGAACTCATCCG	AGGCCTAA	CGGG	ACAACGGgcataaaggtatttaattccata			
70.31	catacaacaagaCGAAAGAACTTAGTGTGATAGACGATTGACCAAACCAATGACAAATCCG	AGCCGCA	CGGG	ATACGGgcataaaggtatttaattccata				
70.74	tcatacaacaagaCGAAAGAACTTAGTGTGATAGACGATTGACCAAACCAATGACAAATCCG	AGCCGCA	CGGG	ATACGGgcataaaggtatttaattccata				
70.75	gggaaaaggtaagtcatacaacaaga	CCTTACCAGACTGATCCG	AAGGCAA	CGGG	ACAAAAGCC	AAAAGAAAAACCTAAC...		
70.5	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AGGCAGAA	CGGG	AAAATCTGC	AAAAGTAACTGTGAAATCCGTGAC...		
70.6	catacaacaagaCCTACTGTAGCGCCTTGGCAGATTACCGCGATCCATGTTACTGATCCG	AACGCA	A	CGGG	ATAATGCGgggcataaaggtatttaattccata			
70.10	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AATAATCCG	CGGG	AAAAATGGGTGGT	AAAAATCTGACCC...		
70.14	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AAATATCCG	CGGG	ATAACCT	TCAAAAGCAAGGGACTTTAAAGTA...		
70.20	gggaaaaggtaagtcatacaacaaga	CTTTTCGTTGTAACCCCAACGAAATCCG	AAGGCAA	CGGG	AAAATAGCCTAGACCAATGTCCGACCGgggcat...			
70.23	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AGCTACGA	CGGG	AGAAAGGG	AAAGCAACCCAAAT...		
70.49	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AGCGGGA	CGGG	AAAATGTCAAAACGACCCgggcataaaggt...			
70.57	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AAGCTTGA	CGGG	AGAAACAAGC	AAAGCACTGATTAAGCACCCACCA...		
70.58	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AACGTTAA	CGGG	ACAAATGCG	AAATGGAACATACGG...		
70.61	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AAATGAGGC	AGAGCCGAA	CCAAAAT			
70.63	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AAGGAAA	CGGG	ATAACC	AAAGATGGGAAACAACTTACA...		
70.68	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AATCTCA	CGGG	AAAAGCAAAAAAAGGCAAC...			
70.73	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AAGGCAA	CGGG	AGAAACTTGCAGAACTACCAAAATGTCGT...			
70.77	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AAGGCAA	CGGG	AAAATCAGTC	CGAGAAAAGGGAGA...		
70.78	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AACGAGGA	CGGG	ACAAACTCG	AGCAATGATCACTATCTCCCT...		
70.44	gggaaaaggtaagtcatacaacaaga	gggaaaaggtaagtcatacaacaaga	AGACCA	A	CGGG	AAAATGGTC	ACACCAAGGAATAGATGACCGCC...	



**Table 2:** Unique 80N DNA sequences that contain Family I pseudoknots. Pseudoknots are in bold and stems are underlined. The sequences are aligned to accentuate the pseudoknot. Capital letters represent nucleotides from the variable region of the aptamer while under-case letters represent nucleotides in the constant region. No constant region nucleotides are represented here. Nucleotides outside of the pseudoknot structure that are conserved are highlighted in yellow.

80.100 • 46 GTCGCGCAATCCG TTTCCAGCAGTCITT CGGGA TAAA CC CTGCTACTCGAGGTCGTTTGATTCGGGCCAATTA<sup>†</sup>TCCA  
80.2•105b TGCAAGCCTGATGA TCCG AA GGTGA CACC AAACAGTATGCTAATCCACATCCGAAACCGACTGTAC<sup>†</sup>A  
80.8=90 GTTAAGCAGTGTCCGCTACCTTCA TCCG TTG GCGAT ATCCG AAGAGGAGCACCAGGCTCCGGTATCAT  
80.7 AAGTGGATGTATGCCACCTTCCGTAAATA<sup>†</sup>CCG TTACT TGGCTT AGGGCC TACGAAGT  
80.21 AGCCTAGTACCCTAAGTGTATCCG TTCC TGTATAGT ACTATACA ACTAACTACCGCTATATTCA  
80.37 CGCACCGCACACCGCACCTCAAATCCG TTC CGTAAAT CCGGA TAA ATTTAGC ACTTTTACTCCCAACACGATTACCA  
80.39 CCGTCCAAITAAA<sup>†</sup>CCG TTTT GACGAT CCGGA AAA ATCGTC GGACCGGGCCTAAGTTTCTTCGCTGGACATTTCAGT  
80.49 TGCCTTATCGAACATCCG TTTSCTGGCGT CCGGA AAAAT GCC ATCGTC GGACCGGGCCTAAGTTTCTTCGCTGGACATTTCAGT  
80.52 GCGACCCACGACTGATCCG TTT GCGGC CCGGA TAA AGCTCGC GGCCATTGCTGTGCACTGGTCCACGAGAATCTA  
80.54 CAGTCCATCAAAA<sup>†</sup>CCG TCAA GCCG TAGA CCGGA GACAAAAT CCGC ATGACCCAGAAAAGAGAGATGTGAAAAGT  
80.72 ACCGGTTAAGCAGTGTCCGCTACTTTTCA<sup>†</sup>CCG TTG GCGAT CCGGA AAA ATCGC AAGACGAGGACCGGCTCCGTATCGT  
80.76 GTCCATGATCCG TGT CCGAA ACA CCGGA GACAAA TTCCGG CTAGAAGACACGGCCCTGATTA<sup>†</sup>TCCCCTTGGGATGGA  
80.80 TGCAGTTGTACAA<sup>†</sup>CCG TTTTA CTG CCGGA CAA AACACAG CAACAATAGTACTATCGCTCAACCGGACATGCAAGA  
80.18CGTACAAAATGAGTCTGGCCCGAATTAACGTATCCG AAA GCTG CAT CCGGA CACAAGCA CAGC GATAATGAG  
80.35 CGTATAGTCCAAACAAAGTATCCG AT GTCC A CCGGA GTAA GGAC CGATCTCGCCCTGAAGCACCGAACTTGAATTGTATG  
80.57 TCGTTCITGGGATCCG A TGGCAA CCGGA TAA TGGCA ACTTGACCACGCTAGGCTAAGATCAAATTTCCCTTCTAT  
80.83 TCGCTCTGAATGCATTTGCCATCGAAATCCG AG GCCA AA CCGGA AAAGA TGGC ACACGCGACTGGCACACGGATATAAGGTT  
80.17 TTGGTCCGAGACGATGTCTCCCA TCCG TTTT GTTCT CCGG AAAA AAGCC GGACATGGCCCGACGCTGTCATCAAATTCAGTCT  
80.74 TTGGTCCGAGACGATGTCTCCCA TCCG T TTTCT TCTCCGG AGAAC TAAAATTGGAGCTATCCACCCCAACG  
80.75 TCCATCCCAATPTGA TCCG AC GGC GA CCGG ACAAT ACATA GAACTAAATTTGGAGCTATTTCCACCACAAG  
80.78 CGTCCACCCGTGAATACGGTCCAGATA TCCG AC GGCA AA CCGG AGAAC AACCTGTACCACACCGTTCTCGGATACCCCTCAG  
80.82 ACCCAAACCGAATA TCCG GCGAGTC CCATCCGG ATAA GA ATAGAGGACAATGAAAAGTGGAGGCAATAGAAAACCG  
70.30=72 ACCCAAACCGAATA TCCG GCGAGTCC CACTCCGG ATAGAATAGA GGA CAATGAAAAGTGGAGGCAATAGAAAACCG  
80.86 CTGCTCAAACCGCAAACA TCCG TC GAGG AT CCGG AAAAA CCTC CATATGCAAGAAAACCGCAACCTGAGATGTACTA  
80.102 CCAATGGCACAGACATCGTAA TCCG TATAGACCTGAGTA CCGG ACAAA ACT ATCAGGTCATGAGAACAGAGACTGTGCTG

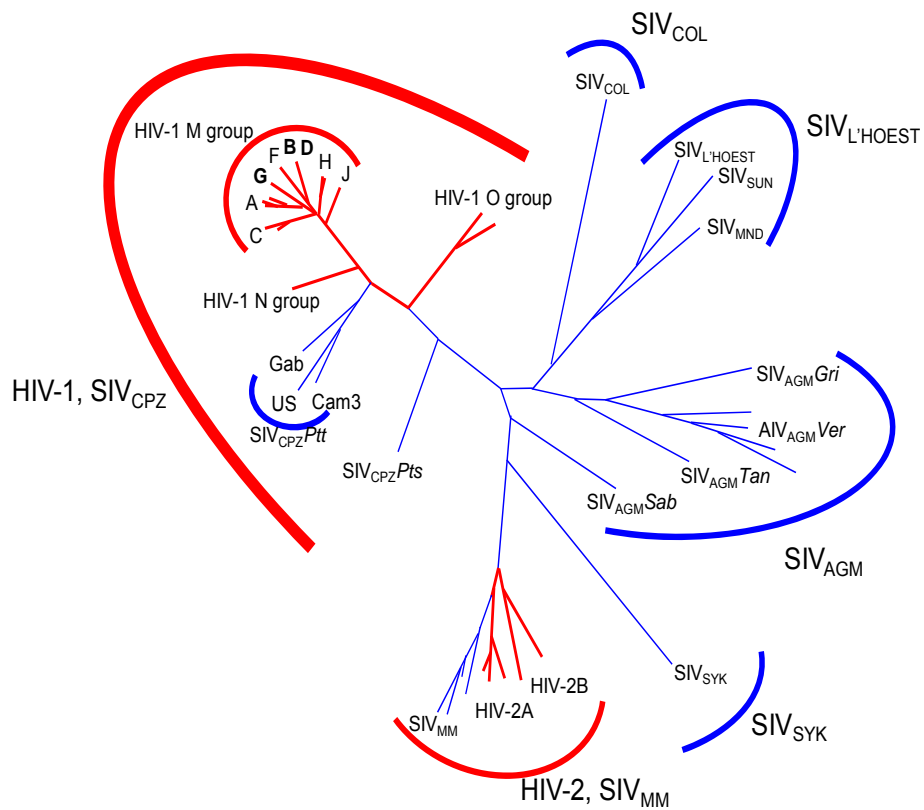
**Figure 2:** A) The Family I pseudoknot adapted from Tuerk et al. (Tuerk, McDougal and Gold, *Biochemistry* 1992). Pseudoknot conformations include two complementary regions that overlap in the primary sequence and base-pair in the secondary structure. The base-paired regions are connected by two loops and a linker. Family I pseudoknots contain the UCCG paired with CGGG in stem 1, colored blue. Stem 2, colored purple, contains any set of base pairs, two to six base pairs long. Loop 1 colored red, consists of one to four unpaired nucleotides and loop 2, colored green, consists of three to six or seven unpaired nucleotides. The linker region is short, often one nucleotide long and is colored orange. B) Family II pseudoknots contain the same general secondary structure as Family I pseudoknots. The Family II pseudoknot is color coded as in A. There is no requirement for nucleic acid identity, nor for number of nucleotides present in stem 1, stem 2, loop 1, loop 2, or the linker. A linker may not be required and loop 2 is likely to be longer than loop 1 since the former crosses the minor groove and the latter crosses the major groove. C) Family III aptamers are any anti-RT RNA aptamer that does not appear to form a pseudoknot as they are defined, or has unknown structure. Two hypothetical examples are depicted. The aptamer may form any number of stem-loop structures including the examples below, a double stem-loop, a kissing complex or other RNA structure. Family III aptamers may or may not contain a pseudoknot.



an arginine (R) caused susceptibility to inhibition (27). Furthermore, NRTIs were found to inhibit DDDP from all the recombinants and NNRTIs are strong inhibitors of DDDP in most recombinants except for HIV-2 (27). Like NNRTI's, Family II aptamers inhibit DDDP activity by RT in many, but not all, subtypes of the virus tested. Inhibit of RTs from viruses from evolutionarily different families suggests these aptamers are "broad-spectrum" inhibitors and may be effective against many common mutant forms of the virus, although this has not been demonstrated directly. These broad spectrum inhibitors are interesting for their potential as gene-therapy agents that could limit the potential of viral escape (27).

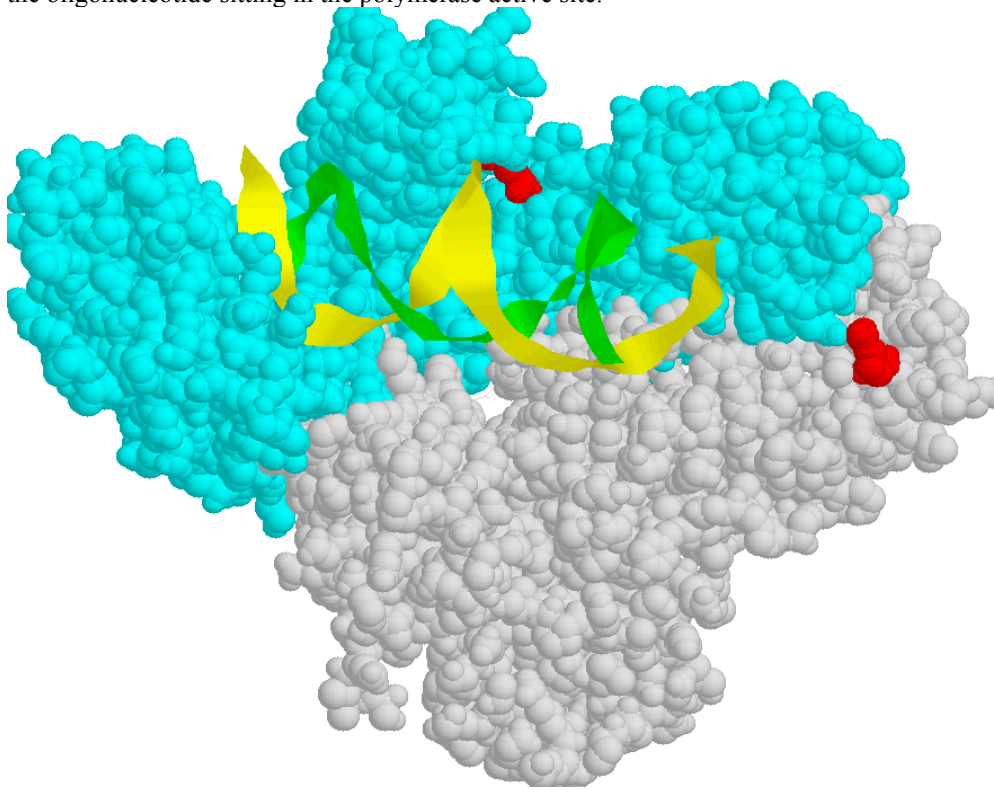
Two subtypes of HIV are subtype B and subtype A. They are both from M group in the phylogenetic tree of HIV (Figure 3). RT from subtype B virus contains R at position 277 while RT from subtype A virus contains K at position 277 (Figure 4). The aptamers studied here were selected against RT from BH10, a subtype B virus. HXB2's RT protein is greater than 99.4 % identical to BH10's RT as shown "using the maximum likelihood approach for analysis of the 1,680 nt of the *pol* gene, which encodes RT" (27). For this reason I considered them experimentally interchangeable. HIV 94CY017 is the subtype A virus studied here that contains the important K277 phenotype. In Held's experiments (27), aptamers that could inhibit RT that contained either R or K at position 277 represented aptamers that could inhibit RT from the most "phylogenetically diverse isolates" tested (27). One hundred unique aptamers were screened for inhibition of the polymerase activity of a subtype B RT and  $IC_{50}$  values were determined (Frankie Rose and Robin Cutler, unpublished observations). Sixty aptamers inhibiting RT with an  $IC_{50}$  value less than 100 nM were further considered for this study. These aptamers were

Redrawn based on Reeves and Doms, 2002



**Figure 3:** HIV phylogenetic tree representing HIV and its close relatives. Arcs and branches in red represent viruses that infect humans and blue represents viruses that do not infect humans. The tree is divided into evolutionarily related families. These include HIV-1/SIV<sub>CPZ</sub> (Simian Immunodeficiency Virus), HIV-2/SIV<sub>MM</sub>, SIV<sub>SYK</sub>, SIV<sub>AGM</sub>, SIV<sub>L'HOEST</sub>, and SIV<sub>COL</sub>. The vast majority of human infections are HIV-1 M group. RT investigated in these works comes from HIV-1, M group, subtypes B and A. RT from subtype B is specifically from BH10 and HXB2 virus. The RT's from these viruses are virtually identical and interchangeable in experiments. RT from subtype A is specifically from 94CY017.

**Figure 4:** Crystal structure of HIV-1 RT for reference. RT is complexed with RNA:DNA oligonucleotide as determined by Sarafianos (2001). Red space fill denotes the position of amino acid 277. Blue space fill is the p66 subunit responsible for polymerase and RNase H activity. Grey space fill represents the p51 subunit. Green and yellow ribbon represents the oligonucleotide sitting in the polymerase active site.



tested further for inhibition of a genetically and phenotypically different subtype A RT from 94CY017. It was found that most did not inhibit the subtype A RT. However, at least two from each family had  $IC_{50}$  values lower than 100 nM, despite the presence of K277 (Judy Gondrohne, Rebecca Chitima-Matsiga, and Angela Whatley, unpublished observations). Inhibition of the polymerase activity of RT from both of these viral subtypes is “interesting biochemistry” and I am interested in discerning the structure of any aptamer that falls in this category.

The aptamers tested for inhibition of subtypes B and A RTs range in length from 116 to 134 nucleotides. It has been shown that for several of these aptamers only a portion of their primary structure is required for binding and inhibition of RT (8, 9, and 10). Elucidation of these minimal binding elements is important for characterization of the aptamer and its interaction with RT, including crystallization. Crystallization with large oligonucleotides is often not possible because they often contain large single stranded regions. The single stranded portions of an oligonucleotide contain too much variable structure to crystallize well and the entropic penalty of ordering single stranded RNA or DNA into a crystal is great. Aptamers 116 to 134 nucleotides in length also often contain multiple conformations (Figure 10), which makes crystallization of the binding structure very difficult.

For Family I and some Family II pseudoknots, it is easy to identify the portion of aptamer responsible for RT recognition by manual sequence inspection and it is likely that all these aptamers need only the sequence including the pseudoknot domain to bind RT. Ying Wan and I verified the existence of many of these Family I and Family II pseudoknots by comparing DDDP inhibition of the full length aptamer and predicted

minimal structure (Results included in Table 4). It is also likely that any aptamer which does not include a known pseudoknot, the Family III aptamers investigated in this work, will contain RT binding in only a segment of the published aptamer sequence. This sequence segment will form a stable secondary structure responsible for the RT binding and elucidation of this secondary structure would be revealing.

I established two important criteria for investigating aptamers. The first criterion was “interesting biochemistry.” Some aptamers inhibited RT from subtype A and B viruses, as discussed above. The behavior of these aptamers was reminiscent of the “broad-spectrum” inhibitors that Held investigated (27), and I was interested in the structure of any aptamers that inhibit RT from broad range of viruses because these might be the most promising future therapeutics. The second criterion was “interesting structure,” which was Family III aptamers, or unknown secondary structure. It is useful to determine the structure of any anti-RT aptamer to facilitate further investigations, including crystallization and bioactivity assays. Furthermore, non-pseudoknot anti-RT RNA aptamers were rare and investigation of aptamers that did not fit the current models where important for discerning if there are more anti-RT RNA aptamer motifs.

In this thesis, I present data from investigations into the minimal primary and secondary structure of aptamers with “interesting biochemistry” and “interesting structure.” Data confirm that RNA aptamers selected by SELEX with previously unrecognized minimal primary and secondary structure contain minimal binding elements which retain full activity against RT comparable to the full length form. I discuss the aptamers systematically, one at a time. For each aptamer I begin with results including the techniques used to determine the minimal core of the aptamer and its

secondary structure and the results of those studies. These data are presented in the results and discussion chapter. For the sake of clarity, any analysis of the results is presented in the next section following results, discussion. Discussion not only includes the full analysis of the results for the individual aptamers, but also what are the implications of the results. A more in depth analysis of my results as it pertains to the field of anti-RT RNA aptamers and HIV is presented in the chapter entitled conclusions. I find that many of the aptamers are likely to contain a pseudoknot that was originally missed by manual sequence inspection, while others may rely on non-pseudoknot structure. We also explore the possibility that one aptamer may be bicistronic. Aptamers characterized in this work may help in diagnosis of HIV subtype and treatment as an inhibitor of reverse transcription (25).



## Chapter 2: Results and Discussion

The following presents the results of biochemical experiments aimed at elucidating the minimal primary and secondary structure of anti-RT aptamers selected against an RT from a subtype B HIV viral strain. Results are presented one aptamer at a time, starting with the two aptamers that include the most complete set of results and ending with the aptamer with the most incomplete and inconsistent data. Results are followed by discussion for each individual aptamer. The implications of the results as a whole on the field of anti-RT aptamers is then summarized in the subsequent chapter.

In the original selection, all RNA aptamers that bound to HIV RT conformed to the Family I Pseudoknot (6). Many aptamers from subsequent selections did not contain the signature UCCG sequence of this pseudoknot. For these aptamers something other than the Family I pseudoknot was responsible for binding. In some cases the structure for binding appeared to be another type of pseudoknot (Family II) or possibly a non-pseudoknot structure (Family III). My experimental approach to discerning the structure of these aptamers has led to surprising insights. Aptamers are not always what they seem at first glance. Discovery of a sequence with the potential to fold into a pseudoknot does not guarantee discovery of the functional core. The inability of a sequence surveyor to locate a pseudoknot signature by visual surveillance of the aptamer sequence does not guarantee lack of a pseudoknot. The accurate determination of structure requires that the aptamer is probed by biochemical means.

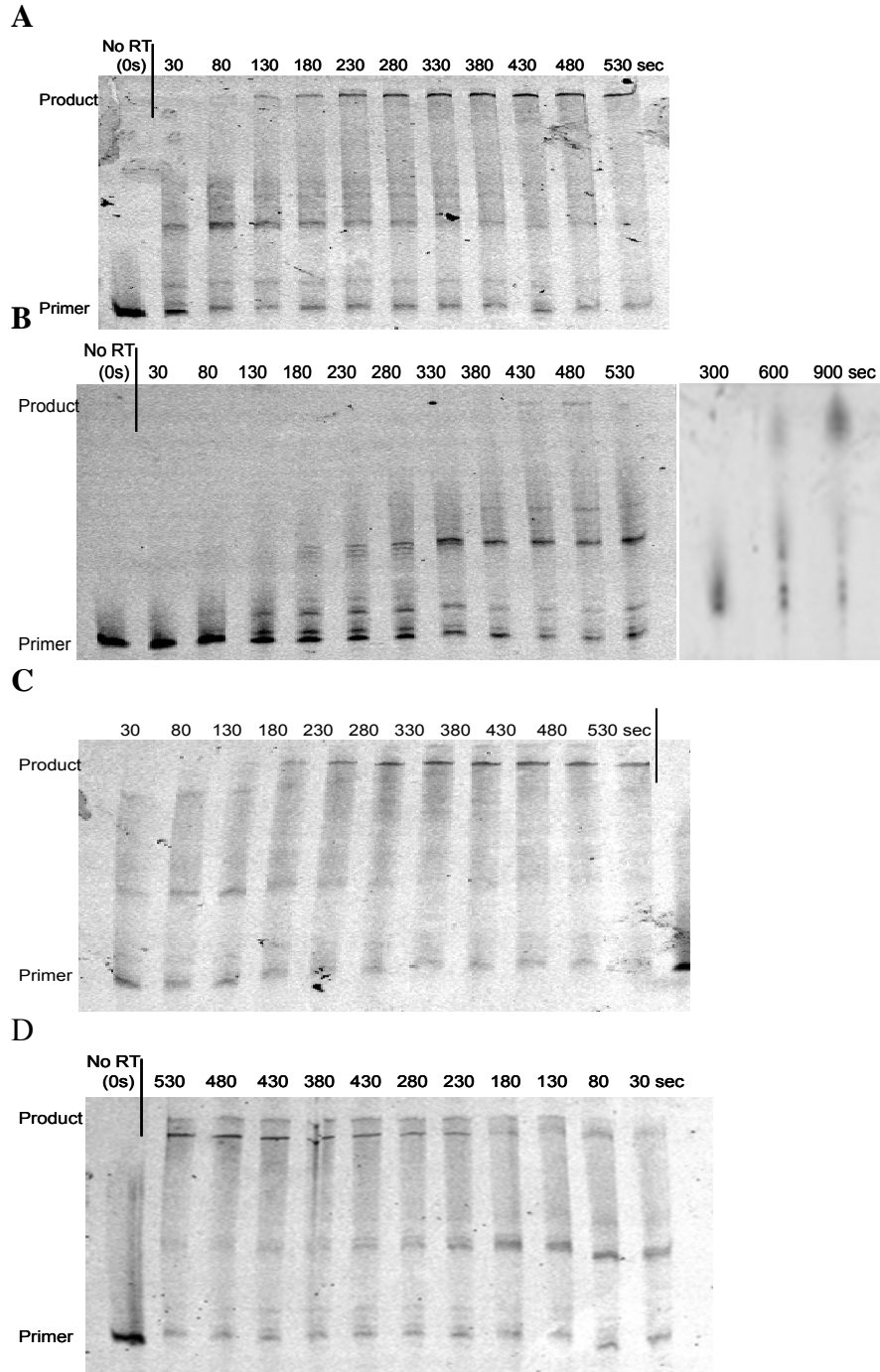
Even though data for aptamers are presented one aptamer at a time, my strategy was to perform similar experiments on several different aptamers simultaneously.

Understanding this strategy is paramount to understanding the data presented in this thesis. It is also important for anyone wishing to continue this work to understand this strategy. I began with a list of aptamers derived from SELEX against RT from the BH10 virus, a subtype B virus. The aptamers I chose had “interesting biochemistry” and/or “interesting structure.”

These studies required active RT protein for binding studies, boundary determinations, and DNA dependent DNA polymerization (DDDP) inhibition assays. The first step was to verify the activity of RT from various viral subtypes and mutants (Figure 5). Activity was determined as the production of fully extended, 103 nucleotide, fluorescently labeled, DNA primer on a long terminal repeat template. Two RTs from subtype B viruses BH10 and HXB2 were tested. These two RTs are highly similar and contain an R at position 277. I also tested an RT from subtype A, 94CY017, and a subtype B RT mutant from Tibotec (a gift from Dr. Stefan Sarafianos). The subtype A RT contains K at position 277 and the mutant RT contains several mutations that render the RNase H domain inactive as well as making the RT easier to crystallize. The Tibotec mutant was the most active, followed by the RT from HXB2, 94CY017, and finally BH10. They all reach a plateau in primer extension no later than nine minutes (530 seconds) into the reaction.

The next step was to apply the mfold algorithm ([mfold.bioinfo.rpi.edu/](http://mfold.bioinfo.rpi.edu/)) to the aptamer sequence or manually inspect the sequence for pseudoknots. Manual inspection is achieved by searching sequence for overlapping regions with the potential to base pair or alternatively searching mfold predicted stem-loops that contain sequence in the loops which can base-pair with sequence outside of the stem-loop. Mfold can only predict

**Figure 5:** RT activity assays. A) HXB2 RT incubated with pre-annealed primer/template, dNTPs, and native buffer temperature conditions as described in materials and methods. Time points at 50 second intervals from 30 to 530 seconds. B) BH10 RT treated as in A, but with additional 300, 600, and 900 second time points. C) 94CY017 RT treated as in A. D) Tibotec subtype B mutant RT treated as in A



stem-loops because algorithm cannot base-pair overlapping regions, so it is important to manual inspect mfold predictions for possible pseudoknots as well. I subsequently designed truncated aptamers either based on the model or based on a system of truncating the aptamer after every X number of nucleotides. I then began to compare the DDDP inhibition properties of the full length and truncated aptamers when incubated with RT from HXB2, the subtype B virus with an RT virtually identical to BH10. If the inhibition pattern of a truncated aptamer was as good as or better than full length, that truncated aptamer was considered to contain the minimal primary structure necessary for binding. This method was used confirm pseudoknot predictions or to rule out the pseudoknot as the minimal secondary structure that binds RT.

To get a more precise picture of the minimal primary structure, I moved on to methods of determining the 5' and 3' boundaries of the aptamer described in materials and methods. These methods required use of radioactively labeled material. The methods were capable of determining the last nucleotide necessary for binding on either the 5' or 3' end to the exact nucleotide. They often yielded approximate boundaries, but they are capable of yielding both accurate and precise boundaries. In the future, it may prove more beneficial to start with these experiments before developing models or using any truncated aptamers in experiments.

Often the exact boundary of an aptamer was not completely clear at this point. When that was the case, I made truncations of the aptamer near that boundary and tested them for binding of RT from HXB2. When I defined the minimal primary structure, I re-examined the possible models for secondary structure and then compared the truncated form to full length in binding studies against RT from BH10 and 94CY017. These

binding studies confirmed or ruled out the truncated aptamer as containing the minimal primary structure.

The result I endeavored to produce from these studies was elucidation of the secondary structure of the aptamers of interest. For those that had a verified minimal primary structure and are from Family III there was one more set of steps. First, the aptamer was separated by native acrylamide gel electrophoresis to determine how many potential structures were in the population for that aptamer. If the truncated aptamer produced only one band it was likely that only one structural conformation was present, although it is possible multiple conformations may co-migrate. If one band did represent one structure, the aptamer could then be successfully probed in a secondary structural analysis. There are many options for probing the secondary structure of RNA, but S1 nuclease, Ribonuclease (RNase) V1, and Ribonuclease (RNase) T1 enzymatic digestion is straight-forward and similar to many other methods I had already performed. Thus, these digestions were my method of choice.

Not every aptamer presented in this thesis was subjected to every step. For some aptamers secondary structural probing was not required to learn about the secondary structure. For instance, if you predict that the aptamer forms a pseudoknot, you can truncate that aptamer to the primary structure that only contains the pseudoknot sequence and compare it to full length aptamer in binding or inhibition studies. If the results for the pseudoknot sequence are similar or better than the full length sequence, you can conclude the pseudoknot is likely the secondary structure that binds RT. For each aptamer, I suggest where this project can lead in the future. One thing is certain though, in the next two chapters you will see that RNA aptamers can form pseudoknots even

when that had not been predicted, they may contain non-pseudoknots, and they can even have interesting additions to the pseudoknots.

## 80.104

### Results for 80.104

The aptamer designated 80.104 (Table 3 – a table containing the sequences of all the aptamers evaluated) is an example of an aptamer appearing to be something it is not. Frankie Rose and Robin Cutler found this aptamer to be associated with the “best” ( $IC_{50} < 3$  nM) category in regards to inhibition of DDDP by RT from HXB2, a subtype B virus. Judy Gondrohne and Rebecca Chitima-Matsiga found that it inhibits the polymerase activity of RT from 94CY017, a subtype A virus, with an  $IC_{50}$  value between 30 and 100 nM and Angela Whatley calculated an  $IC_{50}$  value of 149 nM against this same RT (data not shown). Manual inspection of the sequence by several lab members including myself did not identify a Family I or Family II pseudoknot. The aptamer was assigned to Family III, an indication of unknown structure.

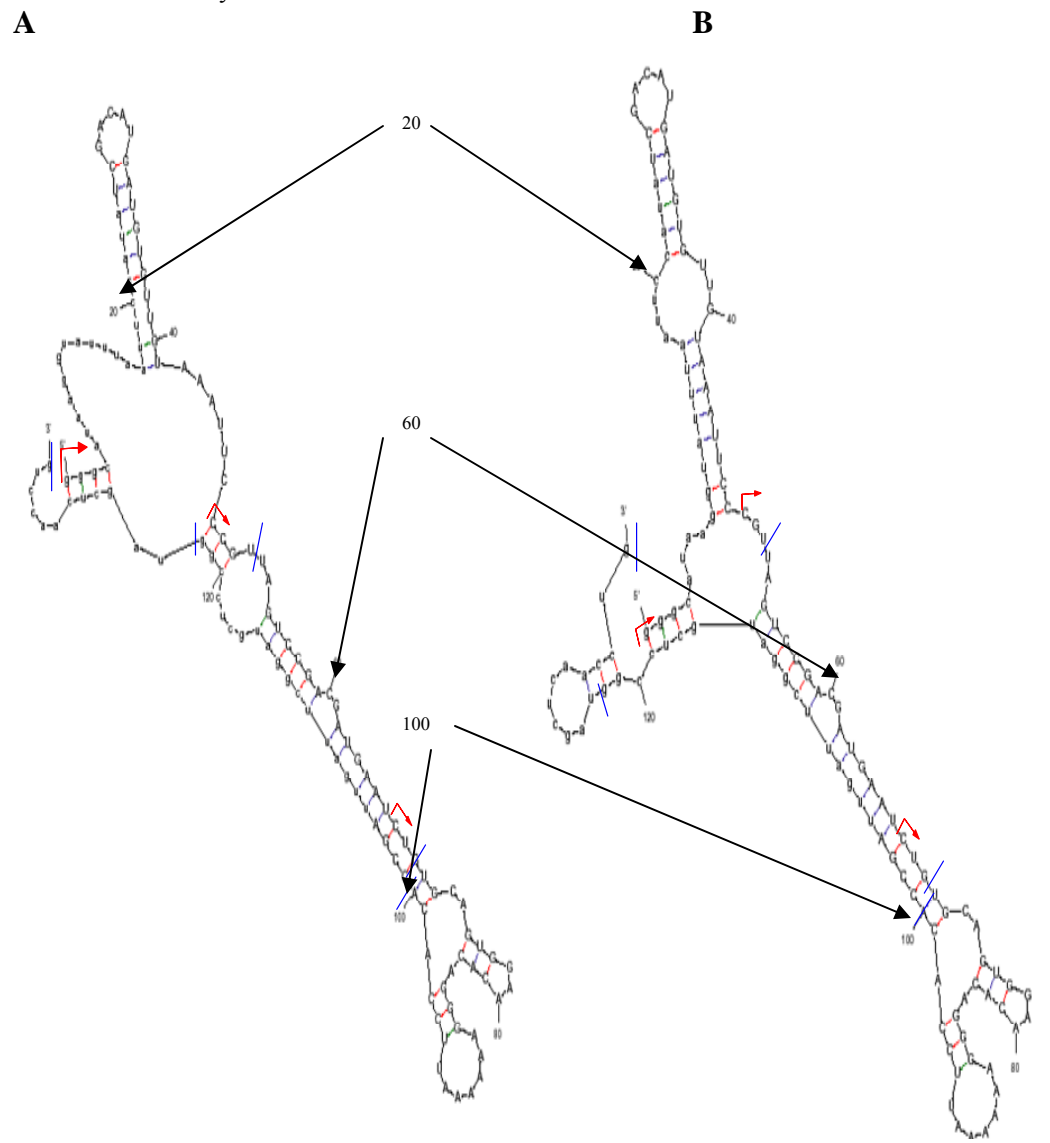
Initially, I probed the effect of truncating the primary sequence of the aptamer on DDDP inhibition against RT from HXB2. I chose the truncation end-points by a combination of systematic truncation and based on stem-loop structures predicted by mfold (Figure 6). The truncated forms I tested include nucleotides 1-134 (full length), 1-50, 1-70, 1-100, 48-122, 48-134, 68-100, and 68-134. The 5' and 3' boundaries of these truncations are represented by red arrows and blue lines respectively in figure 5. DDDP inhibition of RT was observed for full length aptamer as well as truncated aptamers 1-100, 48-122 and 48-134 (Figure 7A). The  $IC_{50}$  value is apparently less than 3 nM for all four aptamers. Weaker inhibition, with an  $IC_{50}$  value between 10 and 30 nM, was observed for 68-100 and 68-134.

**Table 3:** Sequences of full length aptamers and truncated forms tested in this paper in binding assays vs BH10 and 94CY017. Blue nucleotides indicate 5' constant region shared among aptamers. Green indicates 3' constant region shared among aptamers. Orange nucleotides are the variable region, different in every aptamer. Pink nucleotides were added to the sequence for transcription start sites.

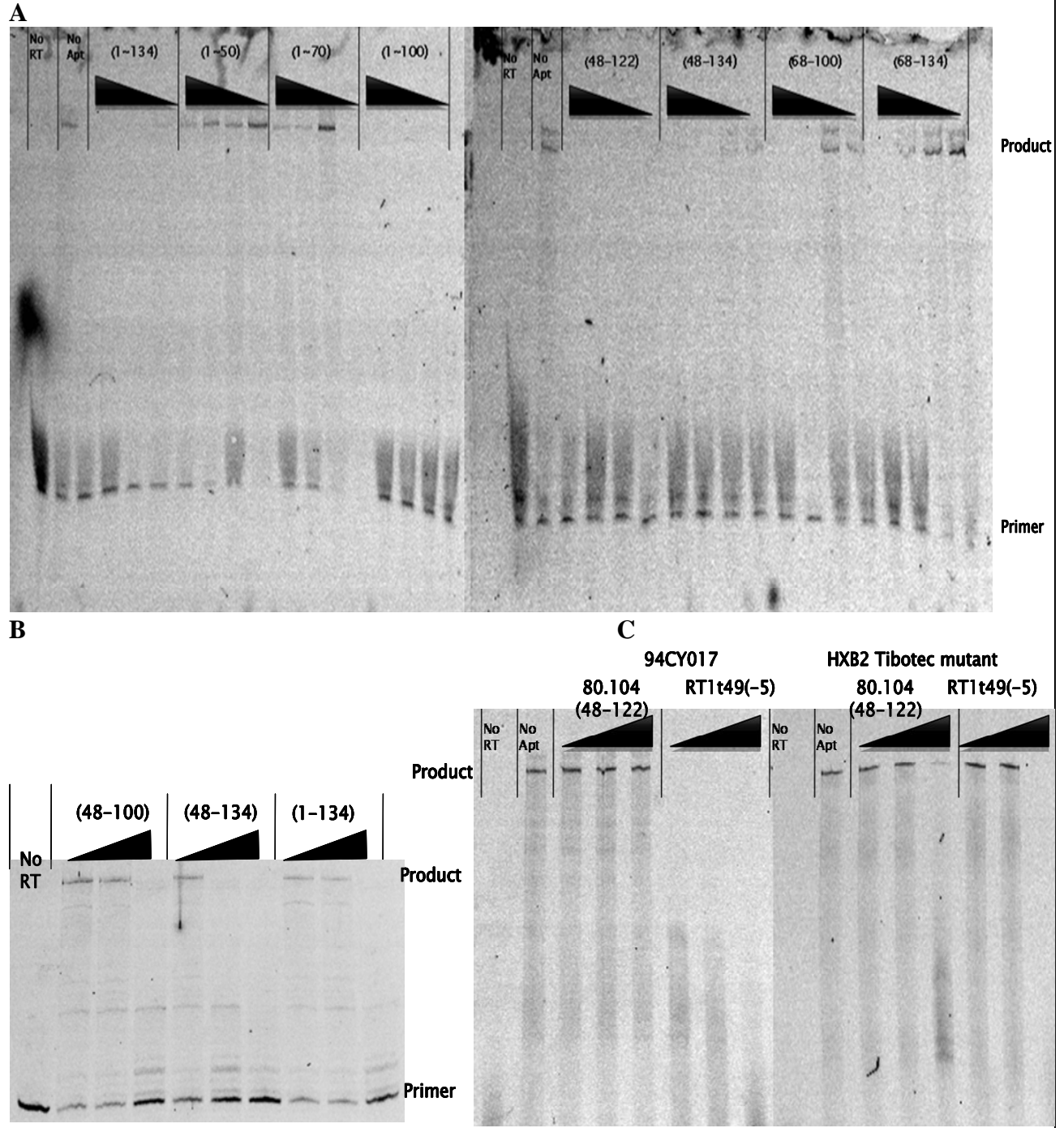
Aptamer	Sequence
70.07(1-118)	5'gggaaaagcgaaucauacacaagaUGUAAGUAUAGCCGUUUUCCAAUGUAAAACCGGAAGCUCC GAACCGAGAAAAAGAGCGAACAAACAUGCGAgggcauaagguuuuuauuccaua3'
70.07(45-86)	5'ggAAUGUAAAACCGGAAGCUCCGAACCGAGAAAAAGAGCGAACAA3'
70.60(1-117)	5'gggaaaagcgaaucauacacaagaCACCAUACCAAUGGCAUCGCUAUGAUCGAGCGCUACGCUA GACCAAUCUGGAGCUGACCACAGGGAACGgggcauaagguuuuuauuccaua3'
70.60(1-90)	5'gggaaaagcgaaucauacacaagaCACCAUACCAAUGGCAUCGCUAUGAUCGAGCGCUACGCUA GACCAAUCUGGAGCUGACCACAGGGA3'
80.33(1-133)	5'gggcauaagguuuuuauuccauaGCGUCGUUUUGCGGUUGAGAGAAGACUUGACACAAGACAU GGAUCCGAUAGCCAAACACCAUGACCCAAUCCGUUACAAuugauucggauaucuccgguagcucaacc ug3'
80.33(1-63)	5'gggcauaagguuuuuauuccauaGCGUCGUUUUGCGGUUGAGAGAAGACUUGACACAAGACA3'
80.55(1-132)	5'gggcauaagguuuuuauuccauaAUGGCUCACCACAAGGGGAACGUUGAUGAAAUAGAGUUUA UCCCUUGGACUCACGCCGGCCGUGCUCACACAAUCCAuugauucggauaucuccgguagcucaacc ug3'
80.62(1-129)	5'gggcauaagguuuuuauuccauaGGCCUUUAGACCUCUCUAAUCCUGGUAGCAUGGACCAGU GAGAAAACUCCAUGUUUAACAAAACCUUAACGCGAuugauucggauaucuccgguagcucaaccug3'
80.62(41-109)	5'gggCUAAUCCUGGUAGCAUGGACCAGUGAGAAAACUCCAUGUUUAACAAAACCUUAACG CGAuugauucgga3'
80.89(1-129)	5'gggcauaagguuuuuauuccauaGCCAACUGCAUCCGCGAGCGUUACGUGGGACAUCAUAGCG CAACGAACCUACCGCUUCAGCACACUCAUCUGACAuugauucggauaucuccgguagcucaaccug3'
80.89(1-65)	5'gggcauaagguuuuuauuccauaGCCAACUGCAUCCGCGAGCGUUACGUGGGACAUCAUAGCGC3'
80.104(1-134)	5'gggcauaagguuuuuauuccauaUCGACAUGAUGUGUUGTAAAUUCCCGUUAGUCCGACGAUGAA UCUGUGCAGUGGAACACAGGGAAAAAUUCCACACCGAuugauucggauaucuccgguagcucaaccug3'
80.104(68-103)	5'gggCUGUGCAGUGGAACACAGGGAAAAAUUCCACACCG3'



**Figure 6:** The secondary structure of RNA aptamer 80.104(1-134) predicted by mfold for A) the most energetically favorable form, and B) the second most energetically favorable form. Red arrows depict 5' boundaries tested by RDDP inhibition while blue lines depict 3' boundaries tested by RDDP inhibition.



**Figure 7:** Inhibition of DDDP activity of HXB2 RT by RNA aptamer 80.104 measured by Cy3 fluorescently labeled primer extension assay. Controls include no addition of reverse transcriptase (No RT) and no addition of aptamer (No Apt.). Truncated forms of aptamers were produced by PCR amplification of desired sequence followed by transcription with T7 RNA Polymerase. Aptamer sequences tested include A) 1-134(full length), 1-50, 1-70, 1-100, 48-122, 48-134, 68-100 and 68-134; B) 1-134, 48-134 and 48-100; or C) 48-122 and DNA aptamer RT1t49(-5) against 94CY017 and Tibotec HXB2 mutant. Aptamer concentrations in each inhibition assay are from left to right, (A) 100, 30, 10 and 3 nM; or (B and C) 10, 30 and 100 nM.



These data led to the hypothesis that either nucleotides 48-100 or 48-122 contain the minimal fragment necessary to bind RT. The hypothesis that 48-100 contains the minimal primary structure was tested by inhibiting DDDP activity of HXB2 RT using RNA truncated aptamer 48-100 and it was determined in preliminary results that this aptamer inhibits as well as full length (data not shown). Upon repeating this experiment, I found contradicting data. Truncated aptamer 48-100 inhibits HXB2 RT with an  $IC_{50}$  value between 30 and 100 nM. This is weaker than inhibition measured for full length 80.104, which inhibits DDDP activity with an  $IC_{50}$  value of approximately 10 nM (Figure 7B).

Truncated aptamer 48-122 was also tested for DDDP inhibition of RT 94CY017 and an HXB2 mutant from Tibotec. Inhibition was compared with that of DNA aptamer RT1t49(-5), a known inhibitor of RT from both subtype A and subtype B (16). No inhibition was observed against the subtype A RT by the 80.104 truncation. Inhibition was observed against the subtype B mutant for both the 80.104 truncation and RT1t49(-5) with an  $IC_{50}$  value observed between 30 and 100 nM (Figure 7C).

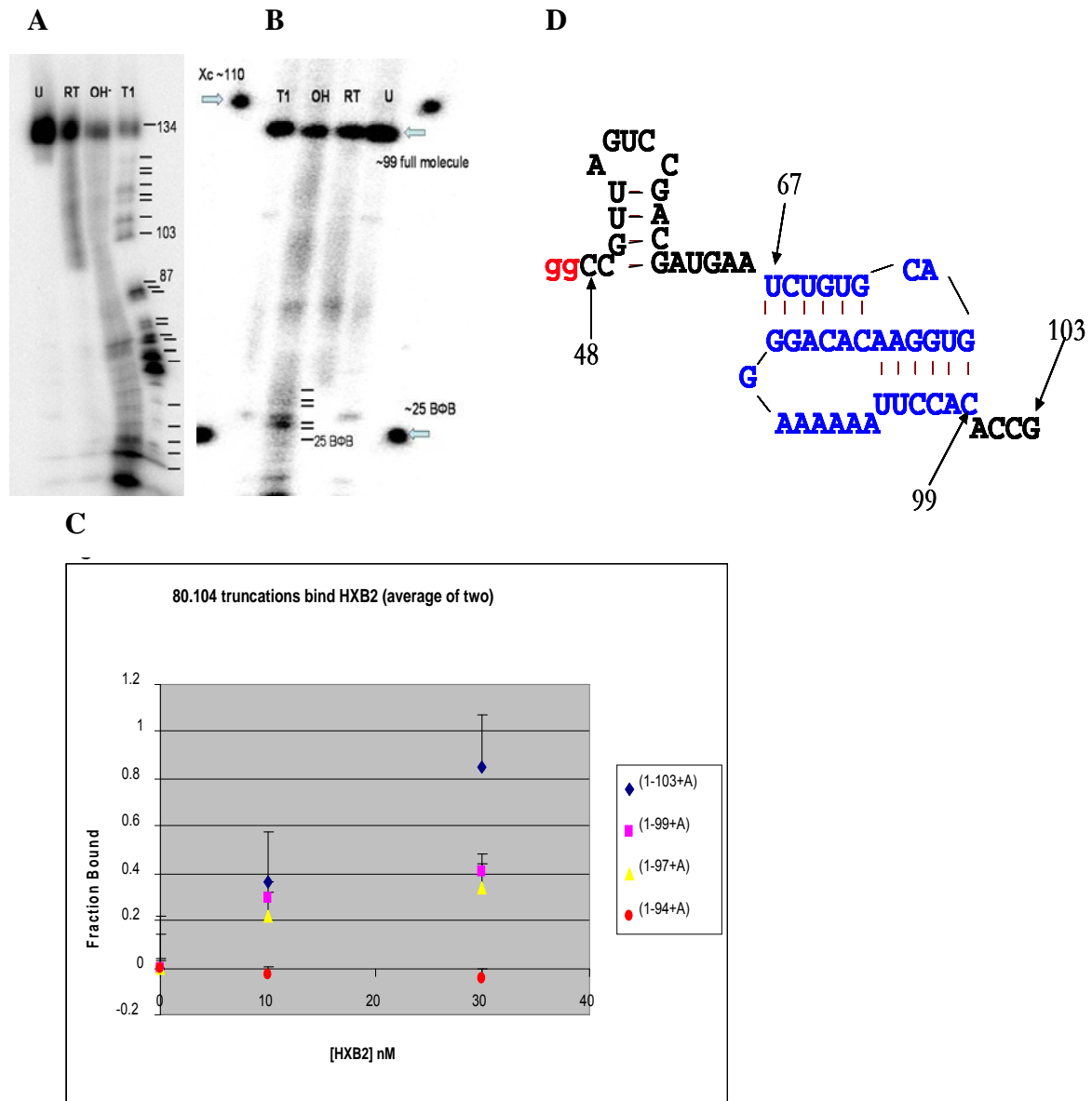
To precisely establish the minimal sequence necessary for 80.104 to bind RT I performed three prime (3') and five prime (5') boundary experiments (Figure 8A and B respectively). As described in Materials and Methods, these experiments are performed by binding radio-labeled, alkaline ( $OH^-$ ) digested aptamers to RT, washing away the unbound aptamer fragments and using gel electrophoresis to separate the bound fragments. Boundaries for the 3' end of the aptamer fragments were determined using 5' labeled RNA, and 5' boundaries were determined with 3' labeled RNA. Signal was

visualized on the gel as bands representing each individual nucleotide, ending abruptly at the location representing the last nucleotide required on that end of the molecule to bind. The identity of the nucleotide was determined by comparison to separate RNase T1 and  $\text{OH}^-$  digested ladders.

The 3' boundary experiment established a boundary between nucleotide 87 and nucleotide 103. I estimated the boundary at  $98 \pm 3$  due to its location relative to nucleotides 87 and 103 in the T1 ladder, and to the  $\text{OH}^-$  ladder. The uncertainty of  $\pm 3$  was not a true quantitative or statistical error, but it was my personal level of certainty of the assignment of the boundary due to the quality of these data. The main use of the uncertainty estimation was in designing subsequent experiments to clarify the boundary. The same method of establishing uncertainty was also performed in other 3' and 5' boundary experiments to guide further experiments. The lack of digestion in the T1 ladder of the 5' boundary experiment, along with the lack in clarity of the RNA bands, made these data more difficult to interpret. After gel electrophoresis was performed on the sample I manually applied radio-labeled RNA in the 6% polyacrylamide gel at the positions of xylene cyanol (110 nucleotides) and bromophenol blue (25 nucleotides). This helped estimate the 5' boundary as  $69 \pm 3$  when compared to the  $\text{OH}^-$  digestion (Figure 8B). The boundary which appeared on the gel was  $30 \pm 3$  but I was using 3' labeled RNA for this experiment. This meant I had to count from the 3' end of the molecule, which was 99 nucleotides long. Thus to determine the actual boundary I subtracted the observed boundary from 99 and calculated 69.

To reduce the uncertainty of the 3' boundary, I tested a series of truncations near the predicted boundary for binding to RT from a subtype B HIV. Truncations 1-94+A,

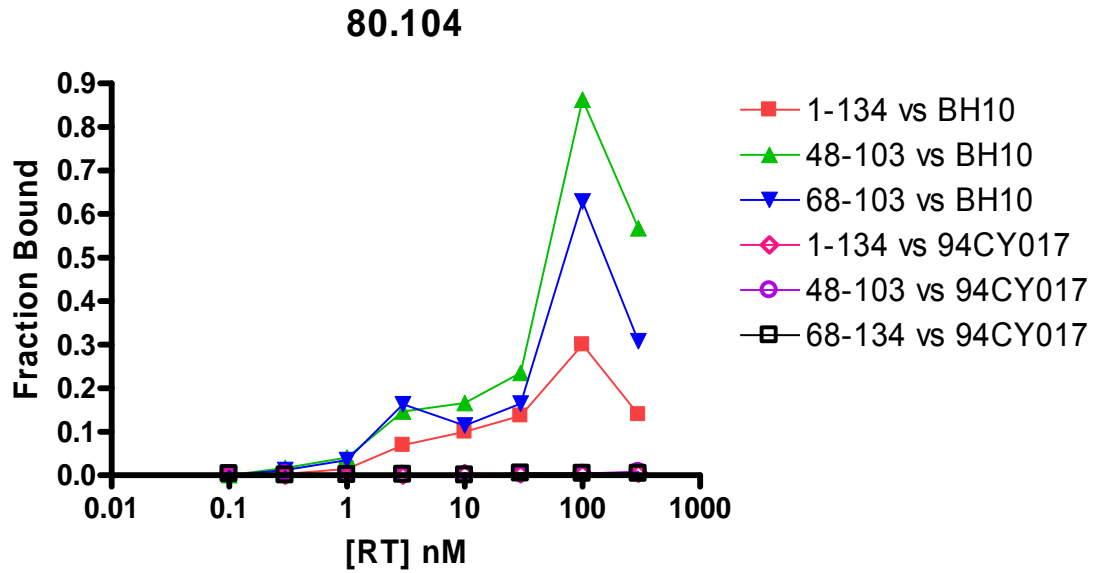
**Figure 8:** Boundary determination of minimal structure necessary to bind HXB2 RT. A) 3' boundary determination by binding 5' radio-labeled and OH<sup>-</sup> digested RNA to 30 nM RT. U is unreacted RNA, RT is sample bound to RT, OH<sup>-</sup> is high pH digested ladder, and T1 is RNase T1 digested ladder. B) 5' boundary determination by binding 3' radio-labeled and OH<sup>-</sup> digested RNA to 30 nM RT. Lane designations are identical to A. Xc stands for xylene cyanol and BΦB stands for bromophenol blue. C) Results of duplicate binding assays. Truncations 1-103+A, 1-99+A, 1-97+A, and 1-94+A were bound to 0, 10, and 30 nM HXB2 RT. They are represented by blue, pink, yellow, and red respectively. D) Newly recognized pseudoknot and stem loop predicted as the secondary structure of truncation 48-103. Red nucleotides are not found in the original 80.104 sequence and are an addition to aid transcription start. Blue nucleotides are involved in the pseudoknot and black are not involved in the pseudoknot.



1-97+A(1-98), 1-99+A(1-100) and 1-103+A were PCR amplified using internal primer, transcribed, and 3' labeled. The additional A on each sequence was only present after the 3' labeling event. Nucleotides 98 and 100 are both A in aptamer 80.104, thus truncation 1-97+A is equivalent to 1-98, and 1-99+A is equivalent to 1-100. The 3' label was incorporated using an overhang primer and  $\alpha^{32}\text{P}$ -dATP. The aptamer fragments were then incubated with 0, 10, and 30 nM HXB2 RT and binding of each truncated aptamer to RT was compared as the fraction of input radio-labeled aptamer retained on nitrocellulose filter (Figure 8C). Truncated aptamer 1-103+A bound the best while 1-100 and 1-98 bound but not as well as 1-103. Truncated aptamer 1-94+A did not bind. These data led me to further investigate the mfold model predicted for nucleotides 48-103. The model predicted two stem-loop structures, the second of which contained a large loop, and a large unpaired linker region. After manual inspection of the sequence in the context of the mfold models, I confirmed the loop of the second stem-loop can potentially base pair with nucleotides in the linker region to form a pseudoknot. The structure predicted is a 12 base pair pseudoknot from nucleotides 67-99 (Figure 8D). Truncated aptamers 1-103+A and 1-100 contain all of stem 2, while 1-98 contains a 4 base pair stem 2, and 1-94+A contains one base pair of stem 2.

Binding was measured as described in materials and methods with the intent of determining dissociation constants for full length, 48-103, and 68-103 against BH10 and 94CY017 RT (Figure 9). Briefly, binding was measured as the fraction of radio-labeled aptamer that bound RT blotted to a nitrocellulose filter membrane and measured using a Fujifilm FLA-5000 phosphoimager. Binding was measured against 0, 0.1, 0.3, 1.0, 3.0, 10, 30, 100, and 300 nM RT. Truncated aptamer 68-103, which contains the entire

**Figure 9:** Binding of two 80.104 truncations was compared to full length against BH10 (subtype B) and 94CY017 (subtype A) RT. Concentrations of RT at which binding was measured are 0, 0.1, 0.3, 1.0, 3.0, 10, 30, 100, and 300 nM. Samples are: red squares, 1-134 binding BH10; green triangles, 48-103 binding BH10; blue inverted triangles, 68-103 binding BH10; pink hollow diamonds, 1-134 binding 94CY017; purple hollow circles, 48-103 binding 94CY017, black hollow rectangles, 68-103 binding 94CY017.



predicted pseudoknot except for the 5'-most nucleotide, bound better than full length consistently across all concentrations of RT tested. Truncated aptamer 48-103 bound even better than 68-103. Both truncations bind the subtype B RT as well as full length, but none of the samples bind the RT from subtype A. A phenomenon caused binding at 300 nM to drop compared to binding at 100 nM. The glycerol content of the reaction is approximately 10% at 100 nM and approximately 30% at 300 nM RT. I propose that 30% glycerol disrupts aptamer RT binding. This phenomenon was present not only in these assays, but in every aptamer tested against both RTs. This interpretation is consistent with previous observations by Dr. Kamendra Singh. In Dr. Singh's experience, when glycerol exceeds 10% of the solution for a binding reaction, binding of DNA to RT is hampered (personal comm.). This prevented accurate determinations of dissociation constants, however a comprehensive view of binding of the aptamers from 0.1 to 100 nM RT are presented.

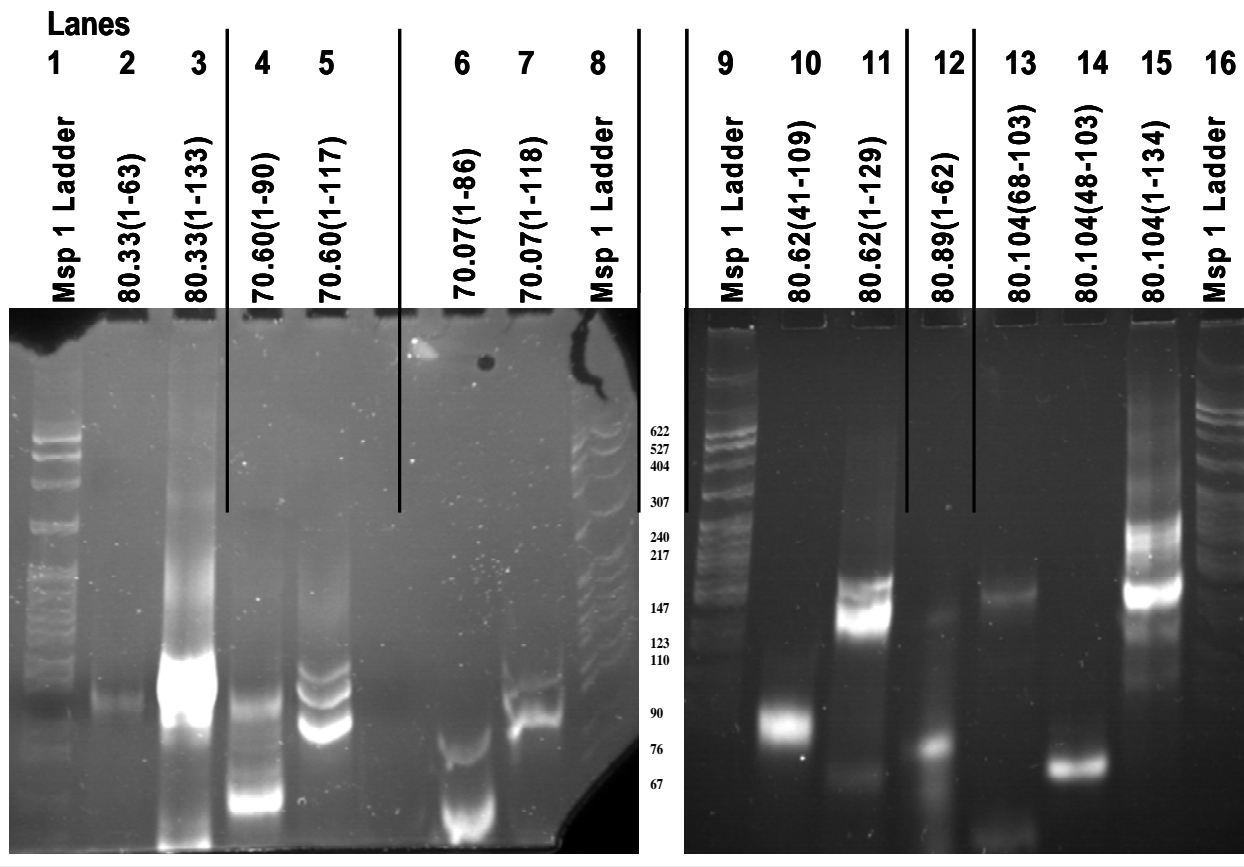
The RNA implicated in binding was run on a native gel (Figure 10; lanes 13, 14, and 15). One major band was viewed for truncated aptamer 48-103 while multiple bands were apparent in the full length sample and truncated aptamer 68-103. This indicates that 48-103 forms one major structural species while the shorter 68-103 may form multiple structures in a population, which could cause a reduction in binding and inhibition by the aptamer becoming trapped in alternative conformation.

#### Discussion of 80.104

Aptamer 80.104 is a strong inhibitor of HXB2 RT, exhibiting behavior of an  $IC_{50}$  value below 3 nM against 4.5 nM final active site concentration RT (Figure 7A). It is not nearly as strong as an inhibitor of 94CY017 or the Tibotec HXB2 mutant (Figure 7C).



**Figure 10:** Gel electrophoresis of native folded RNA aptamers on 6% native polyacrylamide gels. Aptamer identity and length is indicated in each lane as well as the Msp 1 DNA digest ladder. RNA aptamers and DNA ladder were visualized by incubation in Ethidium Bromide and exposure to UV light. Structures in the population are represented by individual bands in each lane. The purpose of these data is to represent complexity within the samples and thus equal volumes and not equal weight of nucleotides was added.



Originally it was classified as a Family III aptamer due to lack of an easily identifiable pseudoknot by visual inspection. Interest in structures other than pseudoknots that inhibit RT led to further investigation of this anti-RT aptamer.

Mfold directed truncations (Figure 6) inhibiting HXB2 (Figure 7), 3' boundary studies, 5' boundary studies, and binding studies (Figure 8A-C) focused attention on a region of the aptamer from nucleotides 48-103 as containing the minimal inhibitory fragment of the aptamer. Close investigation of the stem-loop structures predicted by mfold for this portion of the molecule led to the prediction of a Family II pseudoknot. In the nucleotide sequence from 67-99, I was able to see that the loop predicted by mfold base-paired with a sequence predicted to be single stranded. Every truncation that includes these nucleotides, save 48-100, inhibited RT from HXB2 as well as full length, whereas every truncation that removed any of these nucleotides was found to be a weaker inhibitor, or lose inhibitory capacity. This includes truncation 68-100, which only removed the first nucleotide of stem 1. This indicates that all of stem 1 may be important to binding.

For stem 2, binding studies showed that 5 of 6 base pairs, ending at position 98, bound as effectively as having all 6 base pairs, but additional nucleotides out to position 103 actually increased binding. This could mean these unpaired nucleotides stabilized the formation of this stem, or they could have participated in some unknown RNA-protein interaction. This stabilization of the structure might have increased the overall affinity of the population of aptamers for RT. This may also be the reason that truncations 48-100 and 68-100 did not inhibit RT as well as full length.

The comprehensive binding studies (Figure 9) prove that the minimal element needed to inhibit subtype B RT is contained in truncation 48-103. This includes the pseudoknot depicted in figure 8D. It is apparent that 68-103 binds better than full length but not as well as 48-103. This could be because 68-103 is without the first base-pair of stem 1, or it could be due to multiple structures competing in this truncation (Figure 10, lane 13, two bands). It could also be that the stem-loop preceding the pseudoknot on the 5' end has some interaction with the RT or the pseudoknot that stabilizes the interaction. The most interesting speculation I can imagine is that this stem-loop creates additional contacts to the RT or even that there is a second site of interaction. It is important to note that there are not currently any characterized inhibitors of RT that form both a pseudoknot and another stem loop. Although, it is arguable that the inhibiting structure is the predicted pseudoknot because 68-103 does inhibit better than full length, these studies suggest the truncated aptamer including the additional stem-loop inhibits better. The 48-103 RNA collapses to one major band when compared to full length on native gel (Figure 10) and the only other secondary structure elements predicted to form for this truncation are each individual stem of the pseudoknot and the stem-loop preceding the knot. The fact that 68-103 binds better than full length does suggest this stem-loop may not be necessary either. This aptamer proves how elusive the secondary structure can be to find and prove. An aptamer that appears to be something novel may be just like what came before it.

The most interesting discovery of this aptamer is not that it is a pseudoknot, but that unpaired nucleotides at the end of the structure seem to confer stability in RT binding. Truncated aptamer 48-100 did not inhibit HXB2 as well as 48-122 but 1-103

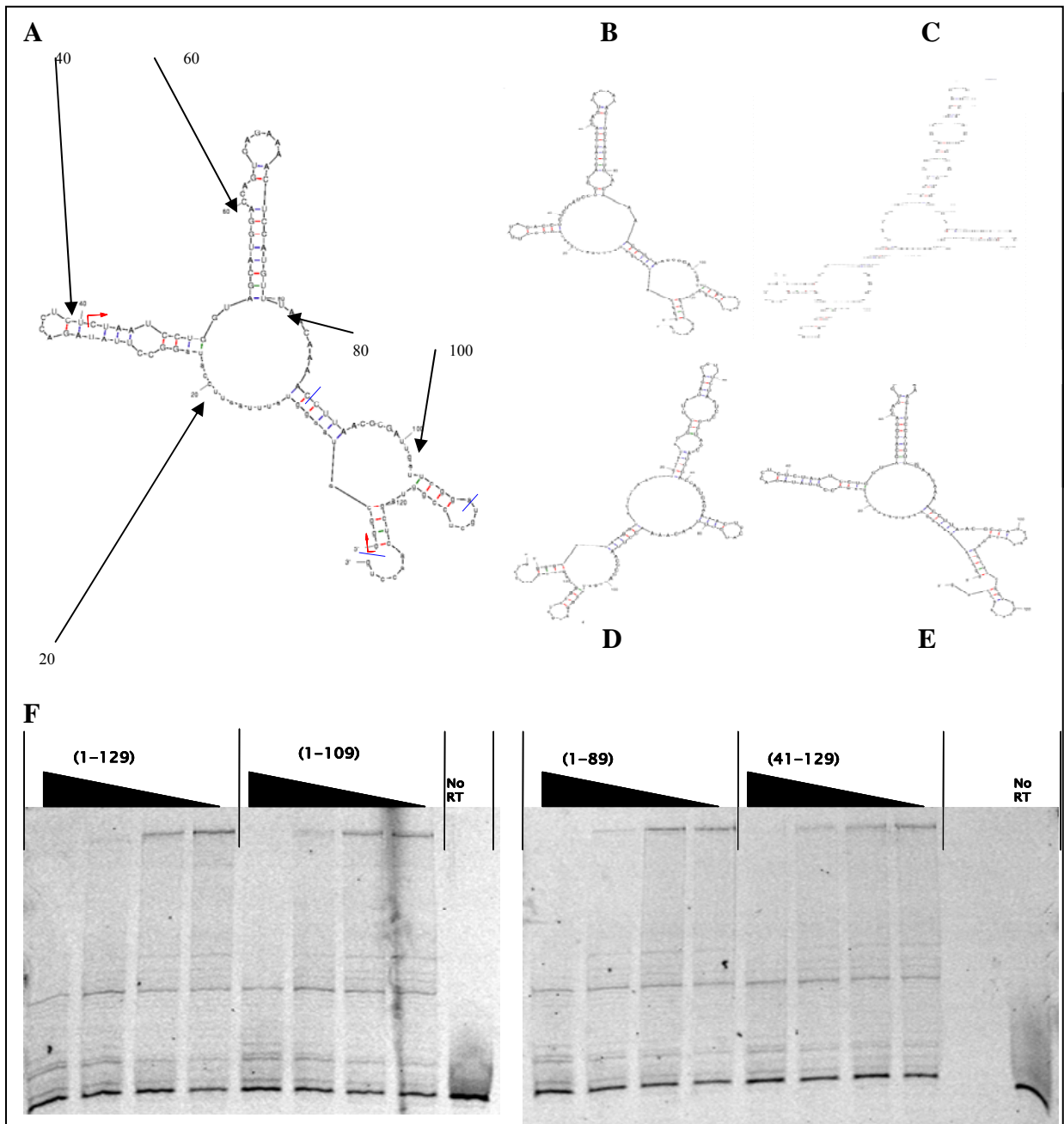
binds better than 1-100. Other aptamers have also displayed more impressive inhibition of RT when short single stranded regions are left unchanged flanking the binding secondary structure (data not shown) and it would be interesting if a pattern of stability of binding could be observed for varying numbers of unpaired nucleotides.

## 80.62

### Results for 80.62

Aptamer 80.62 (Table 3) was previously found to inhibit subtype B RT strongly ( $IC_{50} < 3$  nM against HXB2, data not shown) and RT from subtype A poorly ( $IC_{50} > 100$  nM against 94CY017, data not shown). Manual inspection of the sequence did not identify a pseudoknot model and it was assigned to the Family III group of aptamers. Mfold predicted 5 separate stable structures (Figure 11A-E). Due to the large number of predicted structures at least nine separate aptamer truncations would be necessary to probe the mfold structures fully. The decision was made to probe 80.62 by systematic truncation using 5' and 3' deletion. Truncated aptamers were made removing twenty, then forty nucleotides from the 3' end and forty from the 5' end for the purpose of further refining the location of the minimal primary structure. Truncated aptamers 1-109, 1-89, and 41-129 were tested against 1-129 (full length) in DDDP inhibition assays (Figure 11F). Truncated aptamers 1-109, 1-89, and 41-129 appeared to inhibit as well as full length, all with  $IC_{50}$  values apparently between 10 and 30 nM. This prompted comparison of the DDDP inhibition qualities of truncated aptamer 41-89 and full length. Full length aptamer inhibited much better than 41-89 (data not shown). Binding experiments were used as an alternative method of determining if an aptamer truncation contains the binding element and also showed that 41-89 does not bind to RT (Figure 12A). Therefore the 3' boundary is between nucleotide 89 and 109.

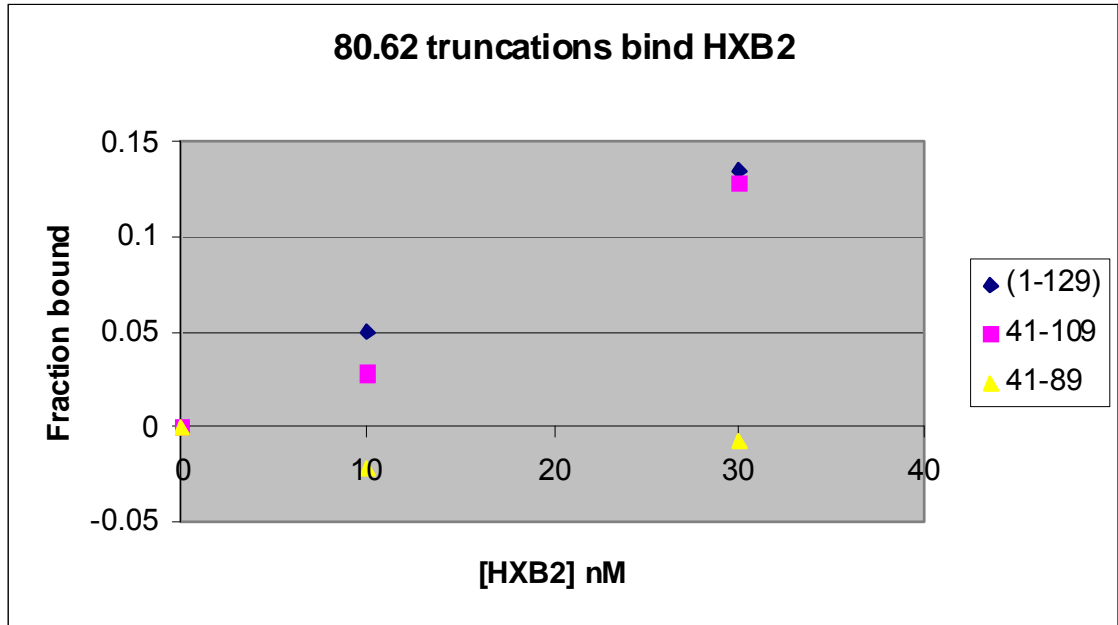
The full length 80.62 sequence was run through vfold (Chen, personal comm.). Vfold is an RNA and DNA folding algorithm much like mfold in development by Shi-Jie Chen (30). It is different from mfold such that it can predict tertiary, double folding, and



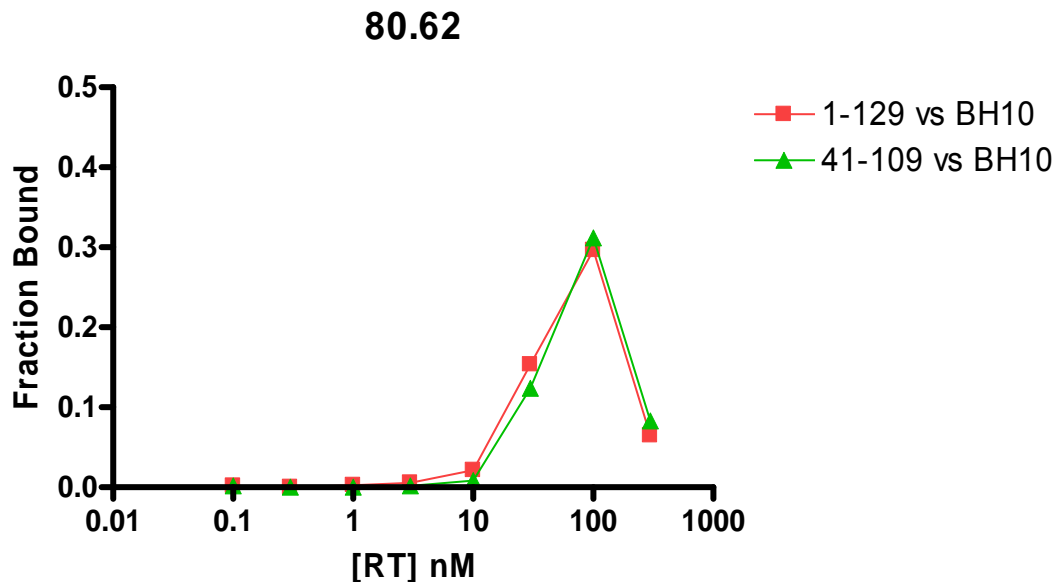
**Figure 11:** Systematic truncation of 80.62 produced three aptamer fragments that inhibited DDDP activity in a similar quantity to full length. A) The most stable stem-loop structure predicted by mfold. Location of 5' and 3' ends of aptamers tested in inhibition assays are represented by red arrows and blue lines respectively. B-E) Other structures predicted by mfold in order higher to lower stability. F) Inhibition assay of HXB2 DDDP activity. Aptamer truncations 1-109, 1-89, and 41-129 were compared with full length. No RT represents no reverse transcriptase control. Aptamer concentrations are from left to right 100, 30, 10, and 3 nM respectively for each aptamer.

**Figure 12:** A) Internally C labeled full length 80.62 and two truncations were incubated with 0, 10, and 30 nM HXB2 and binding was measured. The graph represents one experiment. The X axis is RT concentration and the Y axis is the fraction of aptamer bound. Full length aptamer (1-129), 41-109, and 41-89 are represented by blue diamonds, pink squares and yellow triangles respectively. B) 5' labeled full length and truncation 41-109 were bound to 0, 0.1, 0.3, 1.0, 3.0, 10, 30, 100, and 300 nM RT from viral strain BH10. Red squares represent full length bound to BH10 while green triangles represent the truncation bound to BH10.

**A**



**B**



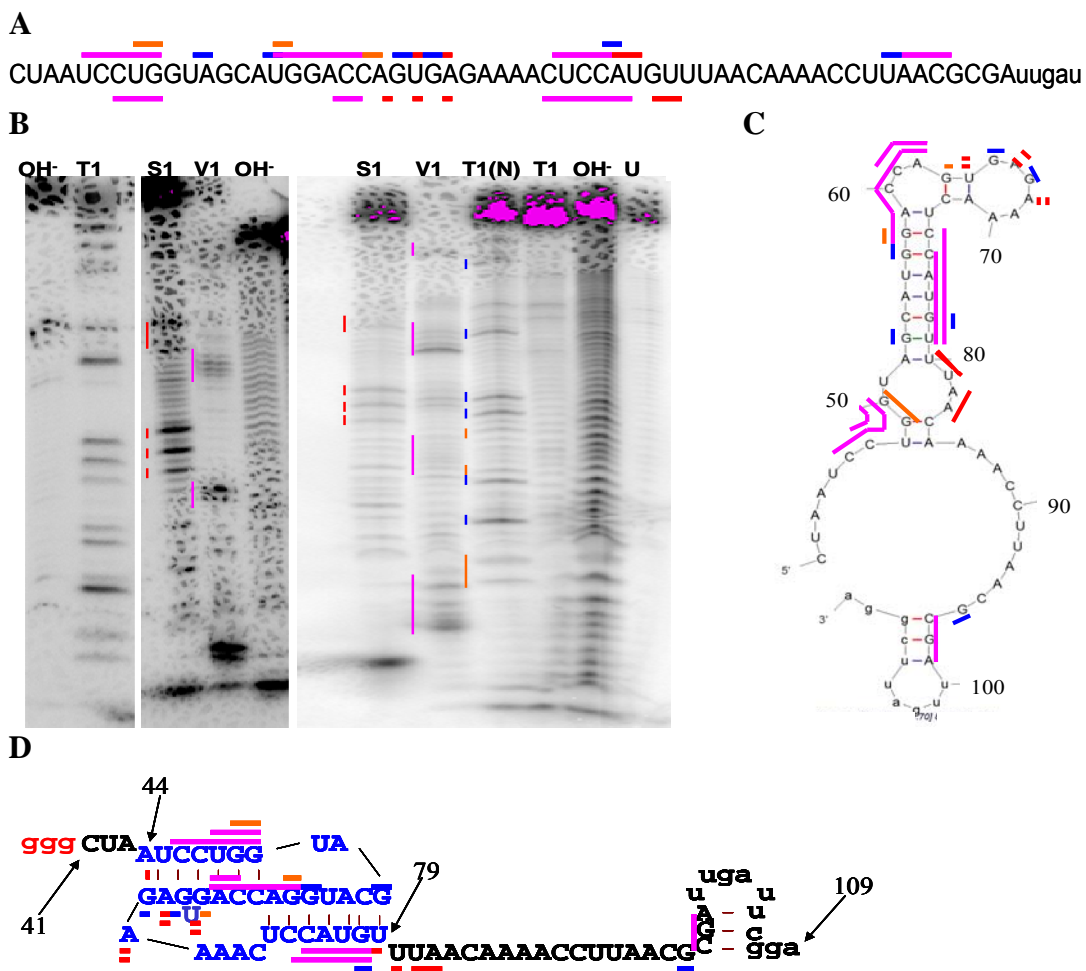
pseudoknot structures. A pseudoknot was predicted from nucleotide 44 to nucleotide 79 (Figure 13D). This model was not considered likely since the truncation 41-89 did not inhibit RT. Subsequently, both truncations 41-89 and 41-109 were compared to full length in binding of HXB2 RT. The 80.62 aptamers were transcribed with internal label and bound to 0, 10, and 30 nM RT. Truncation 41-109 bound as well as 1-129, but 41-89, which contains the vfold predicted pseudoknot, did not bind RT from HXB2 at all (Figure 12A).

Binding was measured with the intent of determining dissociation constants for full length and 41-109. The same problem with the 300 nM RT sample that prevented dissociation constant determination for all other aptamers occurred. Stock protein concentration was too low in comparison to glycerol concentration (500 nM RT with 50% glycerol) with the result that glycerol concentration was too high at 300 nM RT. Binding was measured at 0, 0.1, 0.3, 1.0, 3.0, 10, 30, 100, and 300 nM BH10. Truncation 41-109 bound RT with almost identical values to the full length across all measured RT concentrations. These experiments focus the secondary structural analysis in the region 41-109. Mfold predicts a stem-loop structure in this region (Figure 13C) which contains a bulged stem loop from nucleotide 52-80 and a stem loop from nucleotide 97-107. Vfold predicts a pseudoknot from nucleotide 44-79 and the stem-loop from 97-107 is included in the structural prediction (Figure 13D) because we have seen that a truncation including this pseudoknot alone does not bind RT.

Truncation 41-109 was subjected to native acrylamide gel electrophoresis and only one band was observed (Figure 10), consistent with the existence of a one structure population. The nucleases S1, V1, T1 were utilized to digestion 5' labeled RNA



**Figure 13:** Secondary structure evaluation of 80.62. A) Primary sequence of 80.62(41-109) including color coded secondary structural analysis. Pink lines represent nucleotides cut by RNase V1, indicating involvement in secondary structure. Red lines represent nucleotides cut by S1 nuclease, indicating the nucleotide may not be involved in secondary or tertiary structure. Blue lines represent nucleotides cleaved by RNase T1 under native conditions, indicating no involvement in secondary structure. Orange lines represent nucleotides not cleaved by RNase T1 under native conditions, indicating involvement in secondary structure. B) S1, V1, and T1 nuclease digestions exposed to denaturing polyacrylamide gel electrophoresis. Secondary structure determination is represented by colored lines using the same pattern as in A. Structure determined from these gels was then mapped onto the primary structure in A and secondary structure predictions in C and D. Samples are RNA digested by: OH-, alkaline digestion; T1, RNase T1 under denaturing conditions; S1, S1 nuclease; V1, RNase V1; T1(N), RNase T1 under native conditions; U, unreacted aptamer. C) Stem-loop structure predicted by mfold. Secondary structure determined from S1, V1, T1 nuclease digestions are mapped onto the structure as lines color-coded as in A. D) Pseudoknot predicted by vfold (Chen, personal com) with an additional mfold predicted stem-loop. Secondary structure determination is modeled onto the prediction as it is in C. Color coded as in A.



truncation 41-109 and an interpretable cleavage pattern was observed (Figure 13B). S1 nuclease generally cleaves after nucleotides in single-stranded regions but can cleave some nucleotides involved in tertiary structures and irregular helices (17 and 28). The nucleotides cleaved by S1 nuclease are considered likely single-stranded, but the author recognizes that this enzyme is not perfectly precise. S1 nuclease cleaved after nucleotides corresponding to number 64, 66, 68, 79, and 80 in one instance and 64, 66, 68, 81, and 82 in a second experiment.

RNase V1 is considered highly precise and all nucleotides cleaved by exposure to this enzyme are considered to be involved in secondary or tertiary structure, or within one nucleotide of structure. RNase V1 cleaved the aptamer at positions 46-50, 58-62, 75-78, and 97-99 in the former experiment. The aptamer was also exposed to RNase V1 digestion on a second occasion and was cleaved at positions 61, 62, and 74-79 during this digestion.

RNase T1 cleaves after every G nucleotide in ssRNA. Under native conditions it is much more likely to cleave G residues not involved in any structure. Any cleavage event in this reaction indicates the G residue resides in an unstructured portion of the aptamer. Lack of cleavage at a G indicates this portion of the aptamer is structured. The aptamer was digested with RNase T1 under native conditions on one occasion. This produced cleavage events at positions 53, 57, 65, 67, 78, and 96 but lacked cleavage events for each G nucleotide at positions 49, 50, 58, and 63. These structural data have been mapped onto mfold and vfold predicted secondary structures, as well as the primary structure in figure 12.

## Discussion of 80.62

The minimal primary structure of 80.62 which binds and inhibits RT from subtype B virus is contained within the truncation 41-109 (Figure 12). Inhibition data, and binding data, against RT from two separate viral strains support this conclusion. This truncation bound and inhibited RT from HXB2 as well as full length and bound RT from BH10 as well as full length.

It seemed likely that the secondary structure is not characterized by a pseudoknot from nucleotide 44-79 or a stem-loop from nucleotide 52-80 alone, since the truncation 41-89 did not inhibit or bind (Figure 12A) to HXB2 RT. S1, V1, and T1 nuclease digestions support both structural predictions strongly, but when you consider the accuracy of each enzyme, the pseudoknot predicted by vfold is supported more strongly. The RNase V1 digestions occur only in base paired regions when they are plotted on the pseudoknot model (Figure 13D). When these digestions are plotted on the mfold model there are conflicts between the data and the model at nucleotides 46, 47, 50, 60, 61, and 62. The digestion of the aptamer by S1 nuclease actually supports the mfold model better (Figure 13C). Most of the S1 cleavage events are concentrated in an area predicted to be a loop by mfold, but predicted to be part of stem 1 of the pseudoknot in the vfold model. This may not discredit the pseudoknot though, since S1 does cleave irregular helices. The RNase T1 digestion data support each model equally well, leading me to conclude that the pseudoknot structure is more likely, but neither structure is ruled out completely.

It is certain that the short stem-loop from nucleotide 97-107 (Figure 13C and D) is important to binding. It is the only structural element predicted from nucleotide 86-109, it is supported by V1 and T1 cleavage events, and truncating it off of the aptamer

destroys affinity for RT. It is highly significant that binding, and structural data show the need of such a small stem-loop in addition to other structure in an RT binding motif, as this has not been seen before.

### 80.33

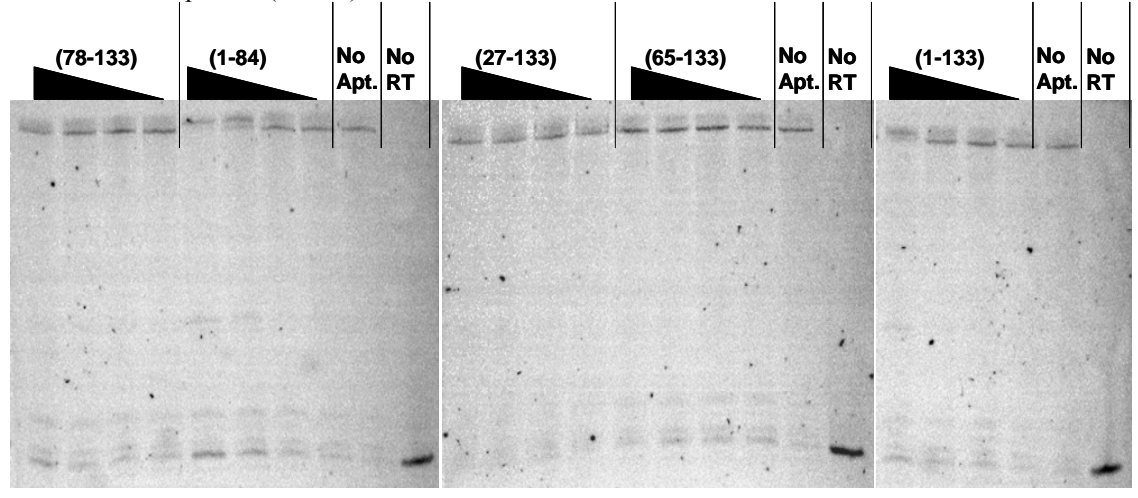
#### Results for 80.33

Aptamer 80.33 (Table 3) inhibits both the subtype B RT and the subtype A RT.  $IC_{50}$  values of 3-10 nM and 54 nM were measured for RT from HXB2 and 94CY017 respectively. It was assigned as a Family II pseudoknot based on manual inspection of the sequence. The pseudoknot was proposed from nucleotides 94-130. Ying Wan tested this model and found the 80.33 truncated aptamer from nucleotide 88-130 does not inhibit HXB2 RT as well as full length (data not shown).

The minimal primary structure was investigated further, as was done with 80.104 and 80.62 in the preceding sections. Systematic truncations were made testing the effect of deletion of the 5' and 3' ends of the aptamers on DDDP inhibition of HXB2. DNA for truncated aptamers 27-133, 65-133, 78-133 and 1-84 were PCR amplified, transcribed and compared against inhibition produced by the aptamer 1-133 (full length). Only the truncated aptamer 1-84 inhibited RT as well as full length (Figure 14) and the observed inhibition was much less than previously measured  $IC_{50}$  values predicted.  $IC_{50}$  values for full length and 1-84 were both between 30 and 100 nM, and the other truncations did not inhibit at all. This may indicate that the minimal binding element is within truncated aptamer 1-84, but the reduced inhibition compared with previous data means further investigation is necessary.

Next, 5' radio-labeled full length aptamer entered 3' boundary studies (Figure 21). There is a strong change in signal intensity for the RT bound sample that depicts a 3' boundary of  $67 \pm 4$ . The boundary was assigned relative to  $OH^-$  and partial T1

**Figure 14:** Inhibition of DDDP activity of HXB2 RT by aptamer 80.33 full length (1-133) and truncations 78-133, 1-84, 27-133, and 65-133. Aptamer concentration was 100, 30, 10, and 3 nM from left to right for each aptamer. No aptamer control (No Apt.) was set as the 100% complete extension control. No RT protein (No RT) was set as the 0% extension control.



digestions for both 70.07 and 80.55. Truncated aptamers 1-63, 1-67, 1-71, and full length were PCR amplified and internally labeled during transcription. These were tested for binding to HXB2 at 0, 10, and 30 nM protein concentration. At 10 nM, full length bound best and at 30 nM each truncation bound better than full length (Figure 15A).

Full length and 1-63 were transcribed again and 5' labeled for a more comprehensive binding study. They were tested against BH10 (a subtype B the selections were originally made against) and 94CY017 at 0, 0.1, 0.3, 1.0, 3.0, 10, 30, 100, and 300 nM. Full length bound to both the subtype A and subtype B RT, but truncation 1-63 did not bind either subtype (Figure 15B).

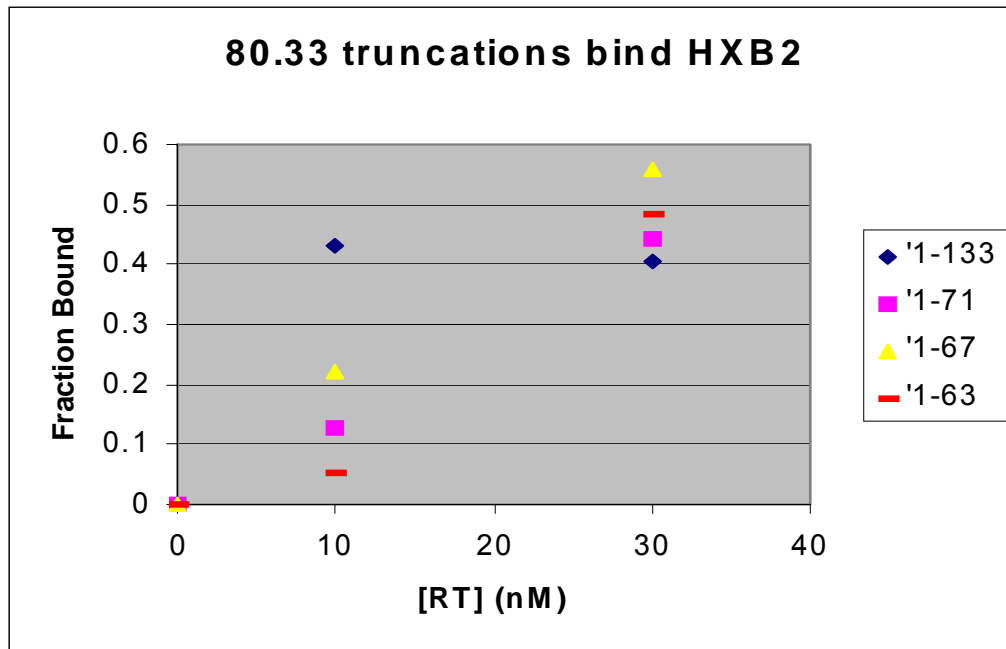
80.33 was not subjected to secondary structural analysis, but samples of the full length and truncation 1-63 were subjected to native gel electrophoresis (Figure 10). The truncation folds to one band on the gel. The full length aptamer folds to one major band, but other bands are present. If later studies were to support that 1-63 does in fact contain the minimal binding fragment, the native gel suggests that it contains one structural species and would be a good candidate for secondary structural analysis. The full length seems to contain several structural species in a population and may produce structural analysis results representative of several structures.

#### Discussion of 80.33

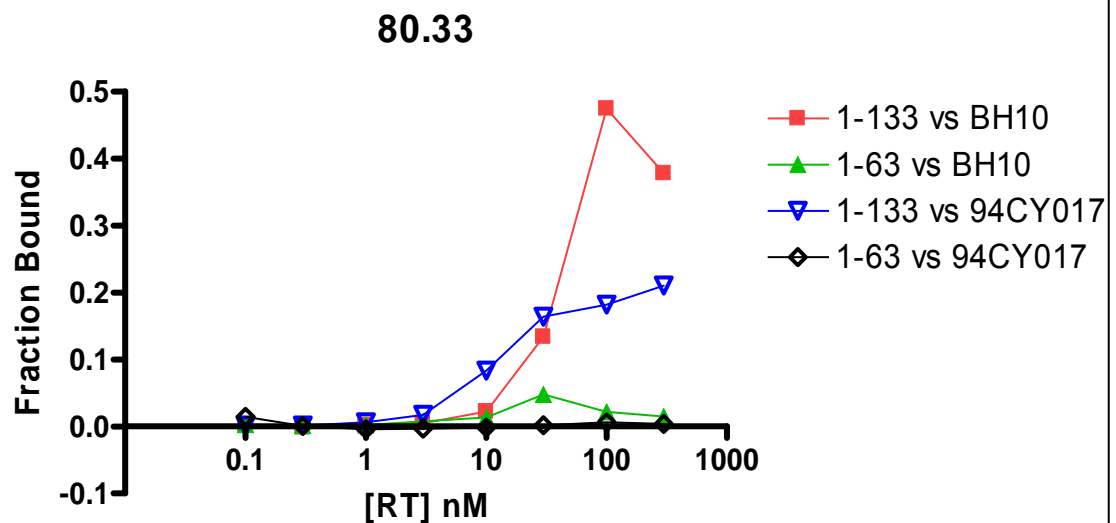
The best model for 80.33 primary structure, determined by DDDP data, currently includes a 3' boundary between nucleotides 63 and 84 and a 5' boundary between nucleotides 1 and 27 or 1 and 65. The behavior of the three truncations in the 3' boundary binding assay was similar (Figure 15A). Binding measured for 1-63 indicates weaker binding of the truncated form to both RT subtypes when compared to full length

**Figure 15:** A) 3' boundary binding study. Three truncations near the predicted 3' boundary were tested against full length for binding of HXB2 RT at 0, 10, and 30 nM. Aptamer samples are represented by: blue diamonds, 1-133; pink squares, 1-71; yellow triangles, 1-67; red lines, 1-63. B) Binding of truncation 1-63 compared to full length for both BH10 and 94CY017 at 0, 0.1, 0.3, 1.0, 3.0, 10, 30, 100, and 300 nM each. Samples are represented by: red squares, 1-133 vs BH10; green triangles, 1-63 vs BH10; blue hollow inverted triangles, 1-133 vs 94CY017; black hollow diamonds, 1-63 vs 94CY017.

**A**



**B**





(Figure 15B) but it is important to note that binding of 1-63 in 15B is much less than in 15A. This indicates the minimal structure is not located in nucleotides 1-63 but the data should be repeated to verify this. Since all three truncations tested in binding showed similar behavior, it stands to reason not one of them contains the minimal structure for binding and the 3' boundary is beyond nucleotide 71.

Inhibition data (Figure 14) suggest that truncated aptamer 1-84 inhibited as well as full length but 27-133 and 65-133 may not have inhibited as well as full length. The inhibition data regarding the 3' boundary is more easily interpreted than that regarding the 5' boundary. The 5' boundary must be determined more precisely through 5' boundary experiments as done previously for 80.104. The 5' and 3' boundary could then be refined if needed by binding truncated aptamers with 5' and 3' ends near the boundaries and full length. Measurement of a dissociation constant for a truncated aptamer that is similar or better than that of full length could verify the location of the binding element.

The truncation 1-63 does bind, but not as well as full length. This truncation may contain a portion of the binding element or may experience non-specific binding. There is also currently no predicted pseudoknot in 1-84. Further analysis of this aptamer may yield proof of a non-pseudoknot RNA aptamer that can bind and inhibit HIV RT. Furthermore, it binds both the subtype B and subtype A RT up to 30 nM similarly. Its binding to this point does not discriminate based on the differences between these two types of RT, including the amino acid identity of position 277. This aptamer is a great candidate for further study.

## 70.60

### Results for 70.60

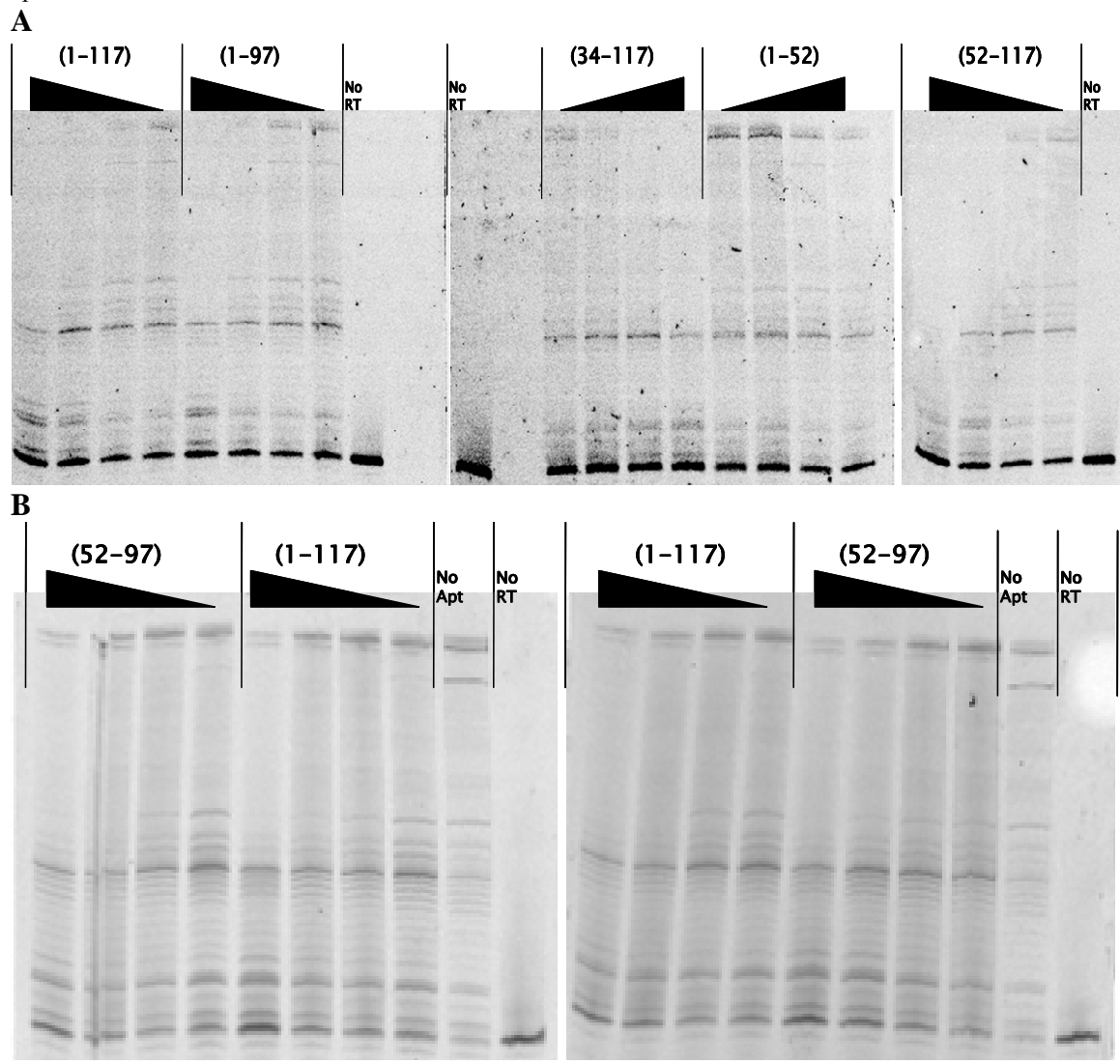
Aptamer 70.60 was originally assigned as a Family II pseudoknot from nucleotide 36-66 by manual sequence inspection. This aptamer also inhibited both subtype B and subtype A RT in preliminary surveys.  $IC_{50}$  values were previously determined as less than 10 nM against HXB2 and 43.6 nM against 94CY017. The sequence from nucleotide 34-66 plus three 5' Gs (Table 3) was made and tested for inhibition of HXB2. This truncation did not inhibit HXB2 compared with full length, ruling out the proposed pseudoknot (data not shown).

Full length aptamer (1-117) was next compared to truncated aptamers 1-97, 1-52, 34-117, and 52-117 in inhibition assays (Figure 16A). All truncations, except for 1-52, inhibited HXB2 DDDP similar to full length with  $IC_{50}$  values between 3 and 10 nM. Subsequent comparison of the truncation 52-97 and full length also provided similar inhibition profiles in one instance but stronger inhibition for full length in a second (Figure 16B). Binding of full length and 52-97 were also compared against 0, 10, and 30 nM HXB2. Binding for full length was greater than binding for the truncated form at both 10 and 30 nM RT (Figure 17A).

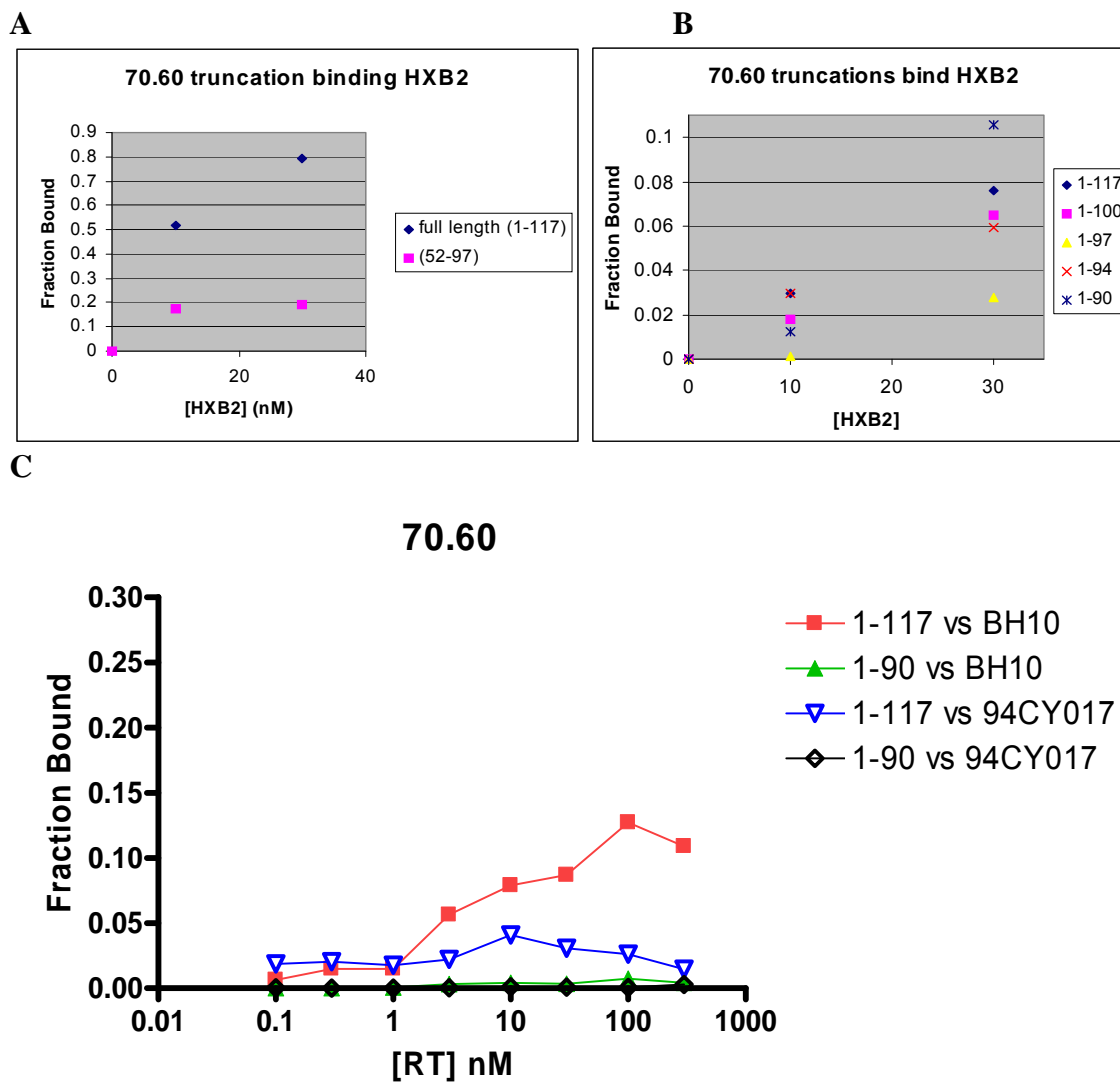
The 3' boundary of the aptamer was tested by binding truncated aptamers ending near nucleotide 97. Full length, 1-100, 1-97, 1-94, and 1-90 were bound to 0, 10, and 30 nM HXB2 (Figure 17B). Truncated aptamer 1-97 was the worst binder in the group at all tested RT concentrations. All other truncations bound similar or better than full length, including the shortest truncated aptamer 1-90.

Binding was measured for 0 through 300 nM RT for BH10 subtype B virus, and

**Figure 16:** A) 5' and 3' truncations compared to full length 70.60 for DDDP inhibition of 70.60. Concentrations of aptamer include 100, 30, 10, and 3 nM. Samples tested include 1-117 (full length), 1-97, 1-52, 34-117, and 52-117. B) Same as A, direct comparison of truncation 52-97 and full length aptamer for RDDP inhibition.



**Figure 17:** A and B) Binding of 70.60 truncations to 0, 10, and 30 nM HXB2 RT. A) Samples are: blue diamonds, full length (1-117); pink squares, 52-97. B) Samples are: blue diamonds, 1-117; pink squares, 1-100; yellow triangles, 1-97; red crosses, 1-94; purple asterisks, 1-90. C) Binding of 70.60 truncations to 0, 0.1, 0.3, 1.0, 3.0, 10, 30, 100, and 300 nM BH10 and 94CY017. Samples are: red square, 1-117 vs BH10; green triangle, 1-90 vs BH10; blue hollow inverted triangle, 1-117 vs 94CY017; black hollow diamond, 1-90 vs 94CY017.



94CY017 subtype A virus for full length and truncated aptamer 1-90 (Figure 17C). Binding is apparent for full length incubated with BH10. Measureable binding also occurs when full length is incubated with 94CY017. No measureable binding was observed when truncation 1-90 was incubated with either subtype A or B RT.

Native gel electrophoresis of full length and truncation 1-90 folded RNA, shows at least three major bands and at least two major bands respectively for each 70.60 aptamer (Figure 10). The multiple structure population contained within each sample makes 70.60 a poor candidate for secondary structural analysis, so no such analysis was performed on the full length or the truncated aptamer.

#### Discussion of 70.60

Much data regarding 70.60 is in conflict with other data. Inhibition data suggests that truncated aptamers 1-97, 34-117, and 52-117 all inhibit RT similarly to full length. It would not be a stretch to hypothesize that 52-97 contains the minimal RT binding element, but inhibition data is inconclusive regarding the properties of this truncation. Furthermore, binding of this truncation is not nearly as strong as full length. It seems that this truncation alone does not contain the minimal binding element.

Oddly, 1-52 also inhibits RT with an  $IC_{50}$  value near 30 nM. This is not as impressive as the other truncations and full length, which inhibit with  $IC_{50}$  values between 3 and 10 nM but it is still measureable inhibition. This is a case in which two fragments from the same aptamer, which are almost exclusive of one another, both inhibit the same RT. Truncated aptamers 1-52 and 52-117 only share one nucleotide in common. This suggests that both of these truncations could contain separate RT binding elements. In other words, this aptamer may be naturally bicistronic.

The native gel contains important information on what may be occurring with the aptamer when truncated to 90 nucleotides. Full length 70.60 has three distinct bands in the structure population visible on the gel, and they all run close to each other. Truncated aptamer 1-90 contains at least two distinct bands on the same gel. One band appears to be in common with both truncations of 70.60 (Figure 10, lanes 4 and 5). That means that upon truncation to 1-90, 70.60 lost the ability to make two conformations and gained the ability to fold to one new conformation. Reduced binding and inhibition activity could be due to loss of the inhibiting structure or gain of a competing non-binding structure.

Binding was not as strong for the truncations and full length tested in 17B and C as it was in 17A. I can conclude for certain that full length 70.60 does bind BH10 but to make other conclusive statements about these data without further testing is difficult. The binding tested in B needs to be repeated first as well as new 5' and 3' boundary assays. The minimal structure of this aptamer, which tested positive at inhibiting both subtype A and B RT, is still unknown and is worth pursuing.

## 70.07

### Results for 70.07

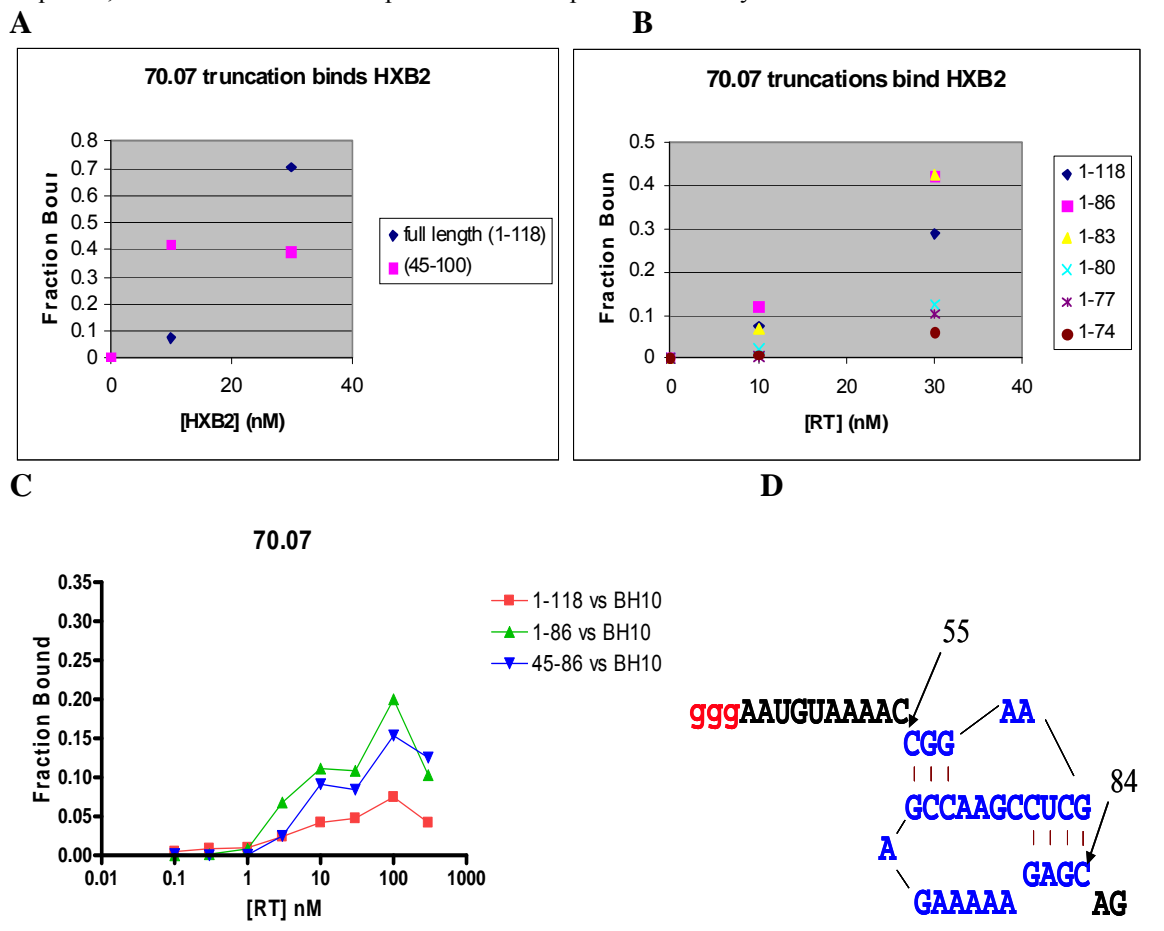
Aptamer 70.07 was originally assigned as a Family III aptamer that inhibited HXB2 RT well with an  $IC_{50}$  less than 3 nM. Inhibition of HXB2 DDDP revealed that truncated aptamers 1-100 and 45-118, but not 1-50, inhibited RT as well as 1-118 (full length, data not shown). This led me to test binding of truncation 45-100 and full length against HXB2. Binding assays produced mixed results. At 0 nM RT the background binding of the full length aptamer and the truncated form were equivalent. At 10 nM RT, truncation 45-100 bound a much larger fraction than full length, but at 30 nM, full length aptamer bound the larger fraction (Figure 18A).

Three prime boundary assays provided an estimated 3' boundary of  $80 \pm 6$  (Figure 21). This boundary was tested by binding full length, 1-86, 1-83, 1-80, 1-77, and 1-74 with 0, 10, or 30 nM HXB2. Truncated aptamers 1-86 and 1-83 bound a similar fraction as full length at 10 nM, and a larger fraction at 30 nM. The other truncated aptamers bound smaller fractions than full length at both concentrations (Figure 18B).

Binding was determined for 0.1 to 300 nM BH10 for full length, 1-86, and 45-86 9 (Figure 18C). Data became distinguishable from background at 3 nM RT. From 3 nM and higher, 1-86 consistently bound the best. Truncation 45-86 consistently bound better than full length over this range as well.

Native gel electrophoresis demonstrates a single primary band and one secondary band for full length and 1-86 (Figure 10). Mfold predicts two structures for 45-83. Each structure contains one stem-loop. Combination of the two structures produces a possible pseudoknot structure depicted in figure 18D.

**Figure 18:** A) Full length and 45-100 bind HXB2 at 0, 10, and 30 nM RT. Samples are: blue diamonds, full length; pink squares, 45-100. B) Probing 3' boundary by comparing binding of truncations to full length. RT identity and concentration are the same as A. Samples are: blue diamonds, full length; pink squares, 1-86; yellow triangles, 1-83; cyan crosses, 1-80; purple lines, 1-77; brown circles, 1-74. C) Binding of full length compared to truncations against BH10 at 0, 0.1, 0.3, 1.0, 3.0, 10, 30, 100, and 300 nM RT. Samples are: red squares, full length vs BH10; green triangles, 1-86 vs BH10; blue inverted triangles, 45-86. D) Pseudoknot structure predicted for nucleotides 55-84 (45-86 with 5' ggg shown). Red nucleotides were added to the sequence for transcription start, black and blue nucleotides are original to the sequence, and blue nucleotides are predicted to take part in secondary structure.





## Discussion of 70.07

Binding suggest that 1-83, 1-86, 45-100, 45-86 all bind as well as full length to subtype B RT, but 1-80 and any truncated aptamer smaller does not. This suggests that the minimal fragment for binding RT is contained in truncation 45-83 and in 45-86.

Binding against BH10 shows 45-86 binds better than full length. Truncated aptamer 1-86 binds slightly better than 45-86 and I believe the better binding demonstrated by 1-86 is due to unknown stabilizing effects of the additional nucleotides on the secondary structure of the aptamer, additional contacts, or possibly even a stem-loop that has not been recognized. The minimal binding element likely is the pseudoknot predicted from nucleotides 55-84 (Figure 18D) and that the extra nucleotides 85 and 86 are not necessary to form the minimal binding element.

The native gel electrophoresis confirms multiple folding structures for full length and truncation 1-86, making these poorer candidates for secondary structural analysis. Mfold predicts two structures, which can be combined to predict a pseudoknot within truncation 45-86. The first nucleotide of stem 1 of the pseudoknot is 55 and the last nucleotide of stem 2 of the pseudoknot is nucleotide 84. The only predicted structure in this portion of the aptamer is a pseudoknot and its components. Furthermore, pseudoknots are the most commonly recognized RNA aptamer motifs that bind RT. Also, truncating the aptamer to position 80 fully disrupts the pseudoknots ability to form stem 2. It is highly likely the pseudoknot depicted in figure 17D is the RT binding element.

## 80.89

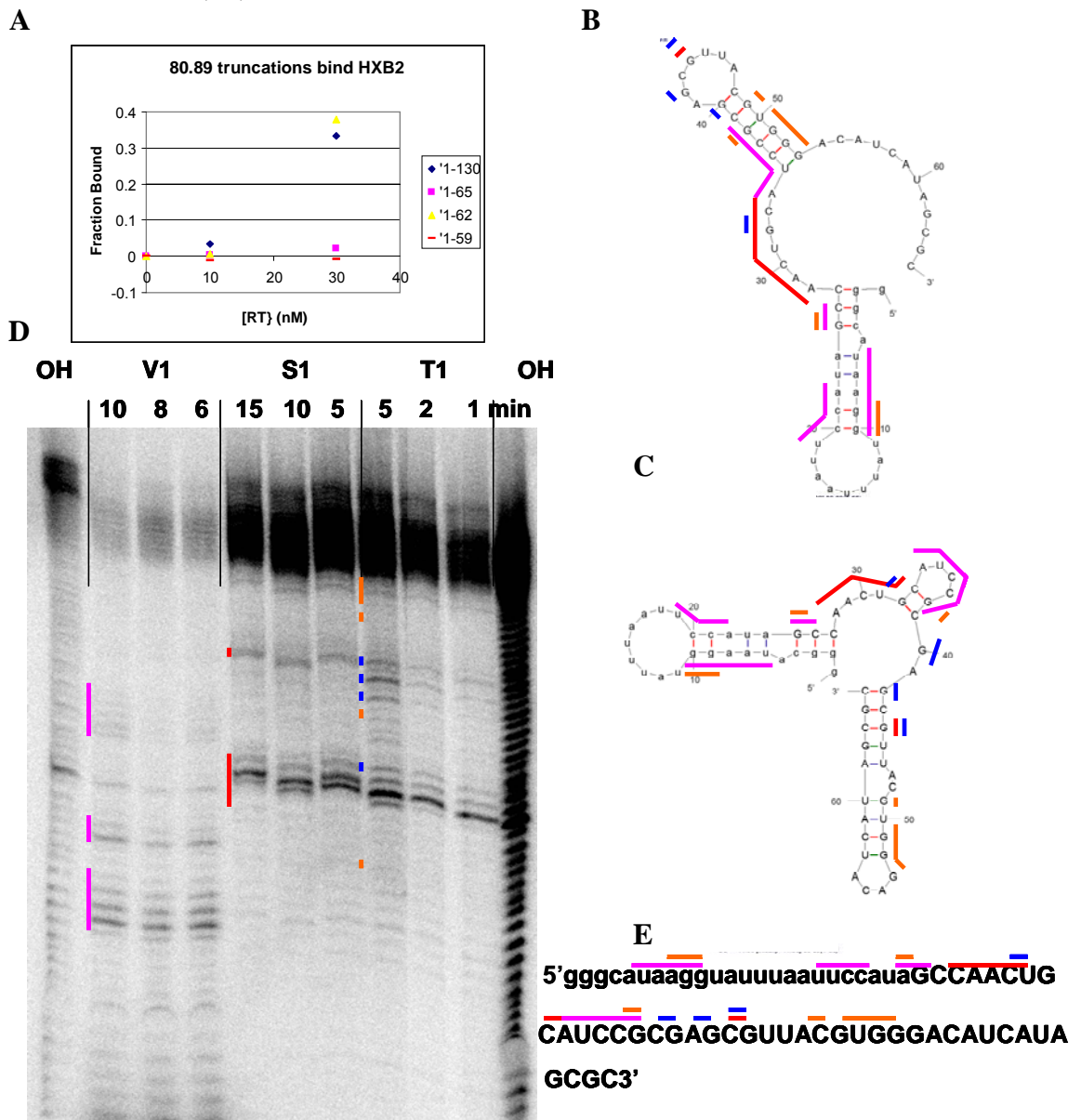
### Results for 80.89

This aptamer was found to inhibit RT from HXB2 ( $IC_{50} < 3$  nM). There was no inhibition observed against RT from 94CY017, but no pseudoknot was predicted by manual inspection of primary sequence, inspection of mfold, or vfold. Full length (1-129) was tested in inhibition of DDDP and compared to truncated aptamers 1-112, 1-82, 32-129, and 51-129. Only 1-82 inhibited as well as full length aptamer (data not shown).

Full length aptamer was 5' labeled with  $^{32}P$ , hydrolyzed, and incubated with RT in a binding assay to determine the 3' boundary containing the minimal binding structure. The boundary was determined at nucleotide  $62 \pm 3$  by comparison of the RT bound  $OH^-$  digested lane to the  $OH^-$  ladder and RNase T1 digested ladder (Figure 21). Truncated aptamers 1-59, 1-62, and 1-65 were compared to full length in binding to 0, 10, and 30 nM HXB2 RT. Truncated aptamer 1-62 bound as well as full length at all concentrations, while 1-65 and 1-59 did not (Figure 19A).

Mfold predicts two structures for 80.89 truncated aptamer 1-65 (Figure 19B and C), one of which leaves the 3' most nucleotides free (19B), and a competing structure that includes these nucleotides in a stem loop structure (19C). For truncated aptamer 1-62, only the structure in B is predicted. S1 nuclease, RNase V1, and RNase T1 digestions were performed on truncated aptamer 1-62 (19D) at various incubation times and a model of single-stranded and base-paired nucleotides was mapped onto the primary structure (19E) and then transferred to the secondary structure (19B and C). The RNase V1 digestion predicts that nucleotides 6-10, 19-22, 25, 26, and 34-38 may be involved in secondary structure. The S1 nuclease digestion predicts nucleotides 28-33 and 44 may

**Figure 19:** A) Internally labeled 80.89 and truncated aptamers bind to 0, 10, and 30 nM HXB2. Full length aptamer (1-130), 1-65, 1-62, and 1-59 are represented by blue diamonds, pink squares, yellow triangles, and red lines respectively. B and C) Mfold predicted structures for 80.62(1-65). Truncated aptamer 1-62 has the same secondary structure as B. Secondary structure predictions from S1, V1, T1 digestions are mapped onto the mfold predictions as colored lines. Pink represents nucleotides cut by V1, an indicator of helical structure. Red represents nucleotides cut by S1, an indicator of single stranded structure. Blue represents G residues cut by T1, indicating these G residues are not part of secondary structure. Orange represents G residues not cut by T1, indicating these residues are involved in a secondary structure. D) S1, V1, T1 digestions of 80.89(1-62). Color coded lines are the same as in B and C. Lanes from left to right are: Alkaline digestion, V1 for 10 min, V1 for 8 min, V1 for 6 min, S1 for 15 min, S1 for 10 min, S1 for 5 min, T1 for 5 min, T1 for 2 min, T1 for 1 min, and alkaline digestion. E) Primary structure with S1, V1, T1 analysis mapped by color coded lines. Color coding is the same as in B, C, and D.



be single stranded. The native and denaturing RNase T1 digestions predict that 32, 40, 42, and 44 are not involved in a secondary structure and that 9, 10, 25, 38, 49, and 51-53 are likely part of secondary structure (19D).

One major band is present when truncated aptamer 1-62 is folded and subjected to native gel electrophoresis (Figure 10). Other minor bands may be present, but if they are the signal is so low that not one can be specifically discerned. This indicates that the S1, V1, and T1 nuclease digestion results depict mostly one common structure.

#### Discussion of 80.89

It is likely the minimal fragment that binds RT is contained from nucleotides 1-62. This claim can not be made with absolute certainty since a broader range of concentrations for RT binding data was not collected for 80.89. The observation that truncated aptamer 1-62 can bind RT but 1-65 does not (Figure 19A) indicates there is a competing structure that forms in 1-65 which prevents the binding structure from forming. Mfold does in fact predict two structures for truncation 1-65, one of which can exist in truncation 1-62 (Figure 19B), and the other of which cannot (Figure 19C).

Interestingly, S1, V1, and T1 nuclease digestions of truncation 1-62 strongly support the structure predicted to form by mfold (Figure 19D). Furthermore, native gel electrophoresis depicts one major band for this sample indicating structure homogeneity (Figure 10, lane 12). Even more interesting is that if you examine the structure prediction presented in 19B closely, you can see that the loop of the 6 base-pair stem-loop can potentially base pair with nucleotides on both the 5' and 3' side of the stem-loop. Even though the potential for two pseudoknots to form exists, the structural data does not support either of them. The pseudoknot that would require nucleotides 3' of the stem-

loop also extends to the end of truncated aptamer 1-65 and does not exist in the truncated aptamer 1-62. These data together indicate the structure that binds HIV-1 RT is composed of two short stem-loops and a seven nucleotide linker region contained from nucleotides 2-53 of aptamer 80.89.

## 80.55

### Results for 80.55

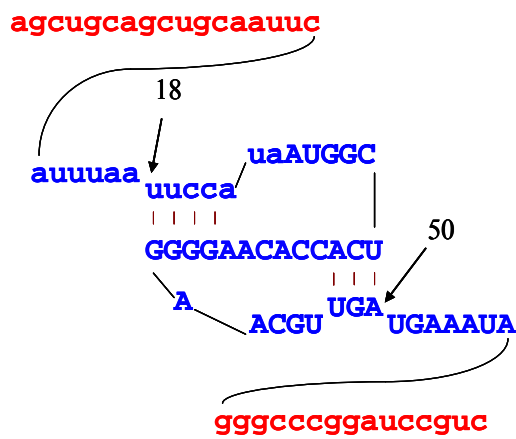
Joshi and Prasad (10) published the truncated sequence of 80.55 from nucleotides 11-57 as binding HIV-1 RT with a dissociation constant of 129 nM. The full sequence was originally published previously by Burke et al. (9 and 13). There is a pseudoknot contained from nucleotides 18-50 in this truncated aptamer (Figure 20A). This pseudoknot is classified as Family II. I tested aptamer 1-134 (full length) and truncated aptamers 52-134, 1-114, 1-90, 1-76, and 1-57 for inhibition of DDDP activity. Results were not fully conclusive, but it did appear that only full length, 1-114, and 1-90 had any effect inhibiting activity (data not shown). This led to further investigation of this aptamer even after Joshi's publication of the pseudoknot sequence. Boundary studies were performed twice to determine the 3' boundary of the minimal binding element. One experiment found the boundary at nucleotide 93 (Figure 20B), while a previous experiment had determined the boundary at nucleotide  $87 \pm 10$  (Figure 21).

### Discussion of 80.55

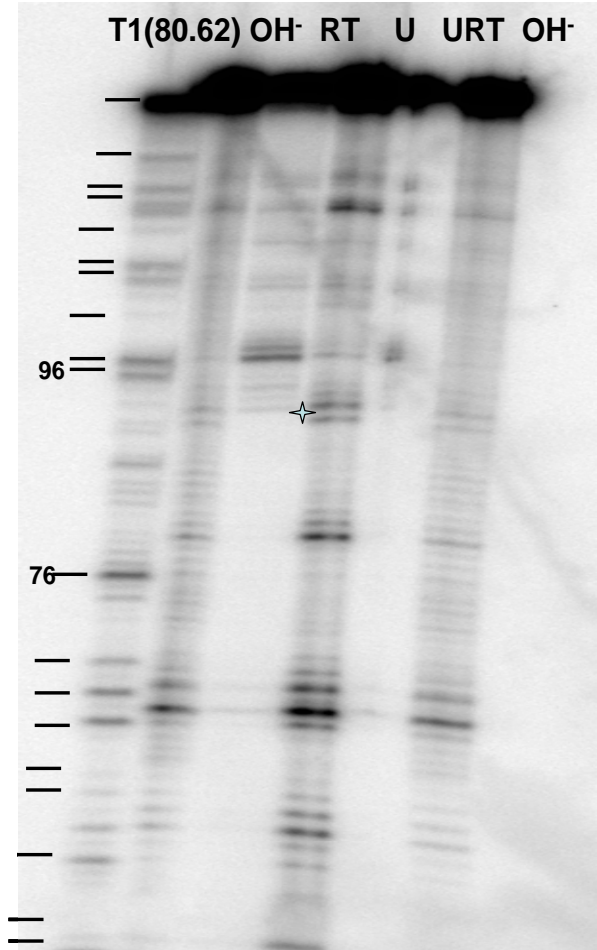
The results of the boundary studies presented here indicate one of two possibilities. The first is that there is an active competing secondary structure that forms in place of the published pseudoknot when the aptamer is kept intact to around nucleotide 93. The second is that a second structure exists beyond the pseudoknot, and that it has a 3' boundary near nucleotide 93. In either case, the competing structure or the supplementary structure must be a stronger binder than the pseudoknot, or else there would be a secondary boundary apparent near nucleotide 50.

**Figure 20:** A) The pseudoknot structure predicted for the truncated 80.55 sequence published by Joshi and Prasad et. al. 2002. Sequence in red is from the cleaved portions of ribozyme used in expression. Sequence in blue is from 80.55. Blue Uppercase letters are from the variable region of the aptamer and blue lower case letters are from the constant region of the aptamer. B) 3' boundary gel for full length 80.55 bound to HXB2 RT. The lanes are from left to right: Denaturing T1 digestion of 80.62, alkaline digestion of 80.55, alkaline digestion of 80.55 bound to and recovered from RT after digestion, unreacted 80.55, unreacted 80.55 bound to and recovered from RT, and alkaline digestion of 80.55. The 3' boundary is indicated in the RT lane and a blue star is placed beside it.

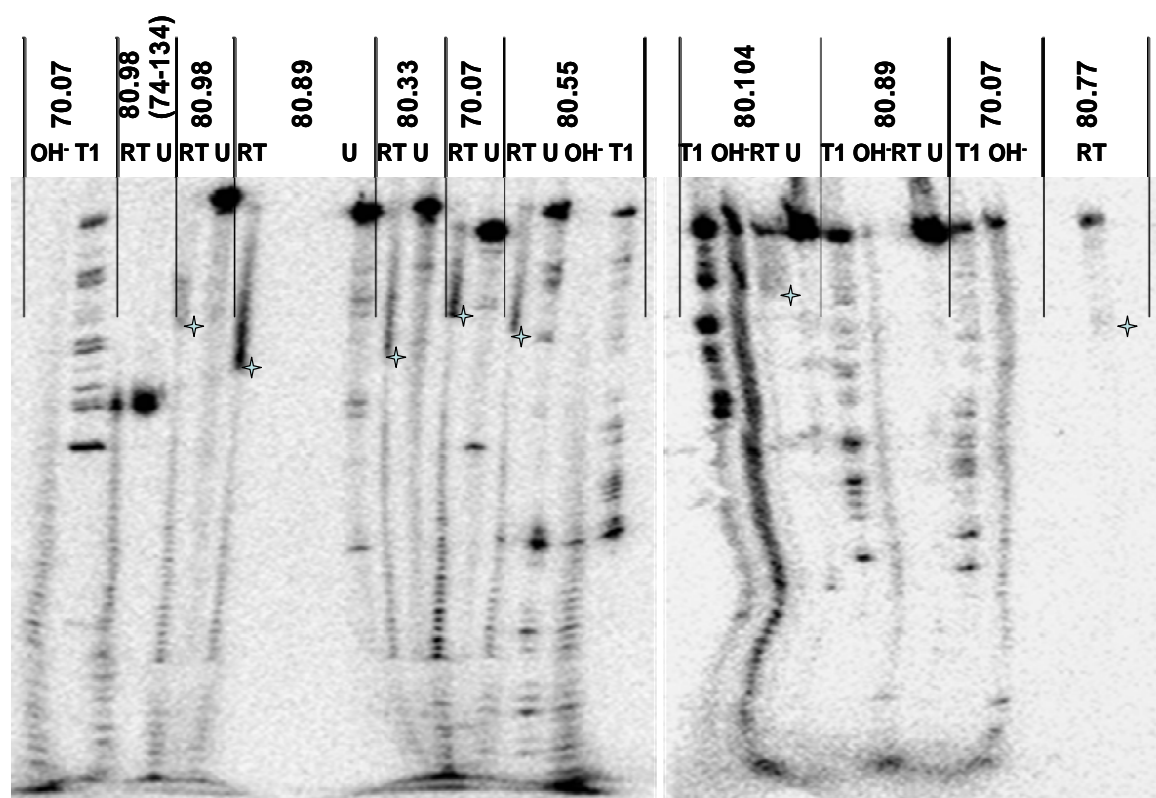
**A**



**B**



**Figure 21:** 3' boundary assay for 80.98, 80.89, 80.33, 80.55, 80.77, 80.104, 70.07 full length aptamers, and 80.98(74-134) truncation aptamer. Lanes are 5' labeled RNA: OH, alkaline digestion; T1, RNase T1 digestion; RT, alkaline digestion incubated with HXB2 RT; U, unreacted. Blue stars represent apparent 3' boundaries for each aptamer. Boundaries determined from this gel are considered to have at least  $\pm 10$  nucleotides in error.





It is hard to draw any solid conclusions from these data. Discerning the 5' boundary could help focus attention in looking for a competing structure, a supplementary structure, or a structure that excludes the sequence including the pseudoknot. Repeating the 3' boundary data would also enhance the data. Then, the aptamer determined by the 3' and 5' boundaries could be generated and compared to full length aptamer and the published pseudoknot in binding. Finally, determining the population of different structures by native gel electrophoresis of each sample and probing secondary structure by S1, V1, and T1 nuclease digestions could determine the structures present and what is actually responsible for binding and inhibiting HIV RT.

Held et al. showed that full length 80.55 is a potent inhibitor of DDDP and RDDP (13) but Joshi published that the  $K_d$  for the 80.55 pseudoknot is 129 nM (10). This binding is weak compared to binding published for other pseudoknots (6, 7, and 10). Also, the full length 80.55 pseudoknot sequence was reported as having 306 percent better binding than a control aptamer with a 5 nM  $K_d$  (9). It seems unlikely that binding would decrease so much if they had used the correct minimal primary structure of 80.55. It is possible that a different structure in 80.55 will bind RT much better than the published structure but to finalize this aptamer's story the experiments outlined above need to take place.

### Chapter 3: Conclusions

Minimal primary structures have been clarified for several aptamers. The data presented in this thesis along with additional data from Ying Wan's and my rotation work are summarized in Table 4. Aptamer 80.104 contains the minimal structure within nucleotides 48-103, and it appears the important structure is a pseudoknot from nucleotides 67-99. Aptamer 70.07 also appears to utilize a pseudoknot from nucleotides 55-84 (contained in truncated aptamer 45-86) to bind RT. Aptamer 80.62 contains its minimal structure between nucleotides 41-109 and 80.89 likely contains its binding element from 1-62. Interestingly, both of these aptamers appear to need structures that at least incorporate more than a pseudoknot to bind RT. Aptamer 80.62 appears to form a pseudoknot but requires additional sequence which forms a stem-loop and 80.89 forms two stem-loops in the primary structure that binds RT. Both 80.33 and 70.60 need further analysis. It is evident that 80.33 does not contain the minimal primary structure in 1-63, and 70.60 may not contain the minimal primary structure in 1-90. The native gel for 70.60, also suggests that cleaving to 1-90 removes two structural conformations and adds one structural conformation which may be responsible for this truncations lack of binding. Even more interesting, there is strong evidence to suggest the pseudoknot published for aptamer 80.55 may not be the important binding element of that aptamer. In fact, the 3' end of the binding element may be present around nucleotide 93, 49 nucleotides away from the end of the published pseudoknot.

The minimal primary structure and the secondary structure of an aptamer binding element is difficult to recognize. Multiple parts of one aptamer may contain inhibitory

**Table 4:** 5' and 3' boundaries, secondary structure, and investigator of individual aptamers. Information in italics represents aptamers that have still unknown secondary structure.

<b>Aptamer</b>	<b>5' boundary</b>	<b>3' boundary</b>	<b>Secondary Structure</b>	<b>Author</b>
80.89	2	62	Two Stem-loops	Josh Franken
80.62	41	109	Pseudoknot plus Stem-loop	Josh Franken
70.21	57	82	Family I	Ying Wan
70.54	51	72	Family I	Ying Wan
80.35	45	68	Family I	Josh Franken
80.80	38	68	Family I	Ying Wan
70.07	45	83	Family II	Josh Franken
70.65	54	77	Family II	Ying Wan
80.63	32	74	Family II	Ying Wan
80.85	35	65	Family II	Josh Franken
80.87	66	101	Family II	Josh Franken
80.93	41	78	Family II	Ying Wan
80.96	67	102	Family II	Josh Franken
80.104	67	99	Family II	Josh Franken
<i>70.60</i>	<i>52</i>	<i>97-117</i>	<i>Family III</i>	<i>Josh Franken</i>
<i>80.33</i>	<i>unknown</i>	<i>84</i>	<i>Family III</i>	<i>Josh Franken</i>
<i>80.55</i>	<i>unknown</i>	<i>93</i>	<i>Family III</i>	<i>Josh Franken</i>
<i>80.77</i>	<i>21</i>	<i>unknown</i>	<i>Family III</i>	<i>Josh Franken</i>
<i>80.96</i>	<i>unknown</i>	<i>unknown</i>	<i>Family III</i>	<i>Josh Franken</i>
<i>80.98</i>	<i>74</i>	<i>107</i>	<i>Family III</i>	<i>Josh Franken</i>
<i>80.105</i>	<i>42</i>	<i>103</i>	<i>Family III</i>	<i>Josh Franken</i>

properties. An element predicted to contain a pseudoknot, may contain no pseudoknot. Aptamer 70.60 is an example of this. Two mutually exclusive parts 1-52 and 52-117, both inhibited DDDP activity. Aptamers 70.07 and 80.104 are both good examples of aptamers that no pseudoknot was predicted for, yet turned out to contain a pseudoknot that became easier to identify as the primary sequence was shortened. Even identification of a pseudoknot that binds and inhibits HIV-1 RT may not always be sufficient. Joshi and Prasad (10) published the pseudoknot in 80.55, along with a dissociation constant, and  $IC_{50}$  values. However, the best binding element may include separate, or at least additional, nucleotides 3' of the published structure.

Even as the minimal structure is honed in on, it can be difficult to elucidate an accurate boundary. 80.104 appears to be stabilized by nucleotides that are 3' of the end of its secondary structure. 70.07 also seems to be stabilized by additional 5' nucleotides. It may be common for aptamers to benefit in stability and binding from additional single stranded nucleotides, or even additional secondary structural elements and tertiary interactions beyond the minimal binding element and this needs further investigation. This means that truncating the aptamer to the minimal binding element may not be completely beneficial.

With the previous consideration in mind, truncating the aptamers to get closer to the minimum binding element is useful, not only to characterize the structure of the binding component to RT, but to increase affinity to RT. The increased affinity may be due to forcing a single conformation on the aptamer that removes non-binding structures. Binding of 80.104 and 70.07 improved as the aptamers were truncated, but binding was reduced slightly upon further truncation. It may be useful to measure  $K_d$  values for each

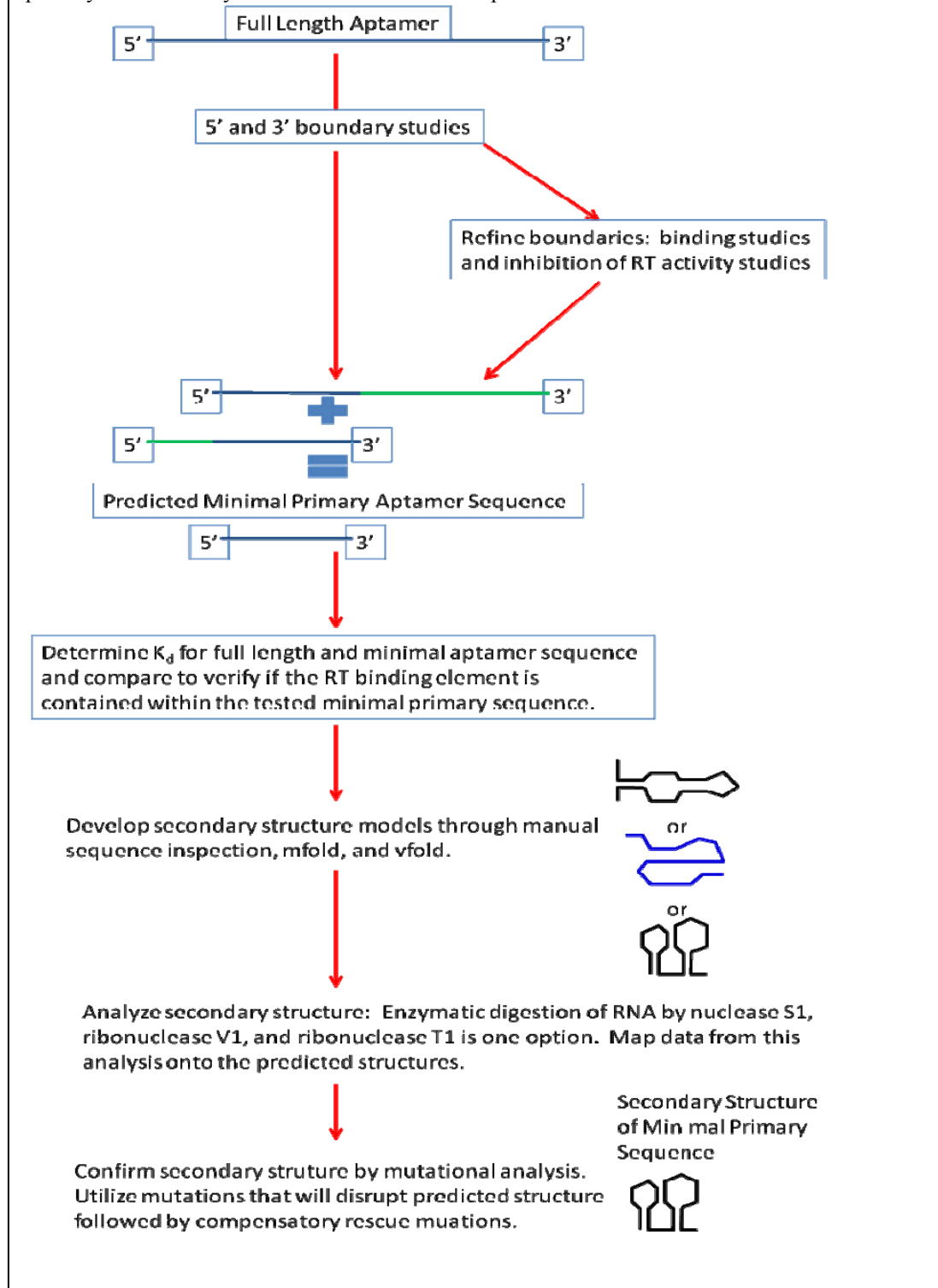
of these aptamers as the 3' and 5' ends were systematically truncated one nucleotide at a time to allow clarification of the true “minimal binding structure.” As you truncate towards the minimal element would binding increase and reach a peak, then decrease slightly? Would several peaks be apparent, or would it just be random? Or, would it be determined by other secondary or tertiary interactions?

There are several directions this project can proceed from this point. The best place to start is the full characterization of 80.62 and 80.89. These two aptamers are well characterized already and represent novel structures. Dissociation constants should be determined for truncated aptamers 41-109 and 1-62 compared with full length of each molecule and paired with the secondary structural analysis. These aptamers comprise the first anti-RT RNA aptamers that require a structure in addition to and other than a pseudoknot to bind RT. Determination of  $K_d$  along with the clear structural analysis is important to the anti-RT RNA aptamer field because this would provide definitive proof that the aforementioned novel structures are the important binding element. Another important finding is to further expand previously published data for aptamer 80.55. If 80.55 does in fact contain a second binding element with more affinity for RT than the originally published pseudoknot; it would be important to determine. If 80.55 contains a second element that increases affinity to RT it may change which part of the primary structure is included in future investigations of 80.55. New 5' and 3' boundary assays should be performed on 80.55 to determine if the published pseudoknot model is supported, or if another model is necessary. The boundaries determined by those assays should be tested against full length and the published pseudoknot to verify if the pseudoknot is contained within the minimal primary structure.

Future investigators of anti-RT RNA aptamers should know of an expanded pseudoknot motif. The idea that 80.55 may contain a second element that increases affinity to RT is not limited to that aptamer. 80.104 contains a stem-loop 5' of the pseudoknot and binding studies suggest that structure increases binding to RT. Furthermore, 80.62 requires an additional stem-loop to bind. It may be more common than is currently known for helical structure adjacent to the pseudoknot structure to be important to binding. In some cases, it may increase affinity while in other cases the additional structure may be necessary for the binding. This observation may be as important as any other as it could affect any or all already characterized pseudoknots; every pseudoknot could potentially be enhanced by an adjacent secondary structure.

I have learned a lot from searching for the minimal core of an RNA binding element. In the future I recommend these steps for a scientist beginning with the knowledge that an aptamer binds a protein (Figure 22). First, perform 3' and 5' boundary assays. This is much more straightforward than randomly truncating aptamers and checking for inhibition of RT. Unless you predict a portion of the sequence to form a pseudoknot, do not base any probing on models at this point. RNA aptamers that are 116 to 134 nucleotides long are too long for mfold(version 3.2) to predict a useful structure. Second, if 3' and 5' boundaries need refinement, make a series of truncations near the 3' and 5' boundaries and test these for binding of RT. RT concentrations of 0, 10, 30, and 100 nM should be sufficient for these preliminary findings. Third, when the boundaries have been fully refined, compare an aptamer with the best 3' and 5' boundaries to full length. At this point, put the truncated sequence through mfold and vfold, as well as searching for secondary structure by manual sequence inspection, and try to develop a

**Figure 22:** Flow chart for the path I suggest future researchers take in determining the minimal primary and secondary structure of anti-RT RNA aptamers.



model. Then solve the dissociation constant for binding of full length and the truncated aptamer. Use at least duplicate measurements and you may have to measure binding as high as 1000 nM RT to complete the curve for the  $K_d$ .

Hopefully the truncated aptamer will bind the protein better than the full length. If it does, you know the minimal binding element is contained within the truncation. If not, you'll have to go back a few steps and determine where mistakes were made. You can move on when you have a truncation with the minimal binding element and a model prediction for secondary structure. The fourth step is to run the aptamers on native gel to make sure the population is homogeneous. Finally, put the aptamer through secondary structural digestions, like S1 nuclease, RNase V1, and RNase T1 digestions. Model these data to the primary structure, and then attempt to transfer them to your predicted secondary structure. The most convincing confirmation of the secondary structure can be obtained through mutational analysis. Disruptive mutations of a helical stem or pseudoknot stem that destroy binding followed by compensatory mutations that restore the structure and the binding can prove the importance of the secondary structure that is proposed. This mutational analysis could be performed on any of the aptamers in this thesis with a secondary model prediction, or any aptamers one wishes to study in the future.

As it is in science, answering any question opens the door to more questions. Aptamers are promising therapeutics for HIV/AIDS but the road does not end at characterization of the aptamers structure. A scientist needs to look beyond this and use all the resources around him/her to ask the right questions. Another important step in moving aptamers towards drugs is bioactivity assays in which aptamers are tested for



their ability to inhibit HIV replication and infection in live cells. One of the most daunting tasks may be the expression of aptamers in human cells. Whatever road you choose as a scientist remember that your most important resource is the people around you. Ask, ask, ask, discuss, discuss, discuss and keep those lines of communication with your mentor and fellow lab members wide open.

## Chapter 4: Materials and Methods

**Materials.** Cy3 fluorophore-labeled RNA and unlabeled DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Gamma<sup>32</sup>P-ATP,  $\alpha$ <sup>32</sup>P-dATP, and  $\alpha$ <sup>32</sup>P-CTP were purchased from PerkinElmer (Waltham, MA). RNA aptamers 70.07, 70.60, 80.33, 80.55, 80.62, 80.89, and 80.104 were previously identified by *in vitro* selection from random pools (9), and their complete sequences are presented elsewhere (13). Expression and purification of heterodimeric RT proteins HXB2, BH10, and 94CY017 was performed previously by Dan Held (25). Phosphonucleotide Kinase, Calf Intestinal Phosphatase, and DNA polymerase I Klenow fragment were purchased from New England Biolabs (Ipswich, MA). S1 nuclease was purchased from Promega (Madison, WI). RNase V1 and RNase T1 were both purchased from Ambion (Austin, TX).

**Mfold Structural Predictions.** Full length and truncated aptamer sequences were analyzed using the mfold algorithm ([mfold.bioinfo.rpi.edu/](http://mfold.bioinfo.rpi.edu/)). Predicted structures were utilized to determine further structural analysis, including locating pseudoknots, determining truncations of aptamers to compare to full length, and predicting minimal structures.

**RNA Aptamer Production.** Aptamers were amplified from glycerol plasmid stock by PCR using Taq polymerase and sequence-specific DNA primers at 1  $\mu$ M final concentration. Full-length aptamers were amplified using end primers, and truncated aptamers were amplified using primers complimentary to internal sequence. Twenty cycles were performed including denaturing at 90 °C, annealing at 4 °C below the

predicted melt temperature of the primers provided by IDT, and extension at 72 °C. Production of accurately sized PCR product was verified by gel electrophoresis on 8% native polyacrylamide gel. Sample size was compared to the Msp I of pBR322 digest marker and visualized by intercalation of Ethidium Bromide into the DNA and excitation under UV light. Aptamers were transcribed from PCR product with T7 RNA polymerase for 4 to 8 hours at 37 °C. Transcriptions were performed with unlabeled 1 mM NTPs or 200 nM  $\alpha^{32}\text{P}$ -CTP along with 1 mM unlabeled NTPs final concentrations. Transcripts were purified by gel electrophoresis on 8% denaturing (8 M Urea) polyacrylamide gel, visualized under UV light (or exposure to film when material is radioactive), and extracted with a razor. RNA was eluted from the gel into elution buffer (0.3 M NaOAc, 50 mM EDTA, pH 8.0) on ice. The gel was crushed in the buffer and filtered through a 0.2  $\mu\text{m}$  syringe filter into a fresh tube containing 1.7  $\mu\text{L}$  10 mg/mL glycogen. The original tube, syringe, and filter were rinsed with 200  $\mu\text{L}$  of elution buffer and were followed with 1 mL of 100% ethanol to precipitate RNA. The RNA was precipitated for 30 to 40 minutes at -80 °C (or 2 hours at -20 °C for radioactive samples) and was spun at 13,200 rpm, 4 °C, for 20 minutes in an eppendorf centrifuge 5415R to pellet. Liquid was decanted off the pellet and the pellet was dried in a Jouan RC10.10 vacuum centrifuge at 37 °C. The pellet was suspended in 21  $\mu\text{L}$  water and the concentration was determined using a ThermoScientific Nanodrop 1000.

**Comparison of DNA dependent DNA polymerization (DDDP) inhibition by full length and truncated aptamers.** Inhibition of DDDP by aptamers was monitored as previously described (13), using a 106-nt DNA, long terminal repeat, annealed to an 18-nt Cy3-labeled DNA primer. Annealing was performed in the presence of BSA, RT

buffer, and 20  $\mu$ M dNTP at 90 °C for two minutes followed by cooling at room temperature for ten minutes in the dark. Each aptamer was heated to 80 °C then slowly cooled to room temperature to produce correct folding and added to 0, 3, 10, 30, and 100 nM final concentrations. Reactions were started by addition of HXB2 to 4.5 nM active site concentration (determination of the active site concentration was previously performed by Dan Held) and stopped after 8 minutes by addition of two volumes of 95% formamide 50 mM EDTA. Samples were heated for two minutes at 90 °C before electrophoresis on a 15% denaturing (6 M Urea) polyacrylamide gel. Gels were scanned for Cy3 fluorescence using a Fujifilm FLA5000 imaging system, and RT activity data were collected using Fujifilm Multi Gage V3.0 imaging software. Inhibition was normalized against no aptamer (no inhibition) and no RT (no product formation) samples and results for full length were visually compared to truncated aptamers to determine qualitatively if the truncation contains the minimal binding element.

**RT activity assay.** Performed exactly as DDDP assay but without addition of aptamers for inhibition. Reactions were run for a time course of 30, 80, 130, 180, 230, 280, 330, 380, 430, 480, 530, 600, and 900 seconds. Time for the reaction to run to completion was determined as the first time point at which full length product reaches a plateau.

**5'-labeling of RNA.** The 5' phosphate was removed from RNA transcripts by incubation with Calf Intestinal Phosphatase (CIP) at 37 °C for 1 hour. RNA was recovered by extraction utilizing phenol, chloroform, and isoamyl alcohol at 25:24:1 ratio, and ethanol ppt. Dephosphorylated RNA was incubated with Polynucleotide Kinase (PNK) and  $\gamma^{32}$ P-ATP at 37 °C for 1 hour. RNA was purified by gel electrophoresis on 8% denaturing (8 M Urea) polyacrylamide gel, visualized by exposure to film, and extracted from gel with

a razor. The gel piece was crushed in elution buffer and filtered through a 0.2  $\mu\text{m}$ , followed by an ethanol precipitate as described above. The extent of labeling was quantified on a PerkinElmer TriCarb scintillation counter as  $^{32}\text{P}$  counts per minute (cpm).

**3'-labeling of RNA.** RNA was labeled on the 3' end as previously described (35). DNA primers were ordered as the reverse complement to the 3' end of each aptamer. The sequence 5'CAT3' was added to the 5' end of each primer to allow the addition of one dATP and a two nucleotide overhang. The primers were incubated with  $\alpha^{32}\text{P}$ -dATP, RNA, and DNA polymerase I Klenow fragment for 2 hours at 37 °C, stopped with two volumes of 95% formamide 50 mM EDTA, and purified by gel electrophoresis, filtration, and precipitate as described previously. Extent of labeling was quantified on scintillation counter.

**Functional Boundary Assay.** To determine 3' functional boundaries, 5'-labeled RNA was  $\text{OH}^-$  and T1 digested under single-hit conditions. Briefly, RNA was incubated in 50 mM Sodium Carbonate, 1mM EDTA, pH 9.0, 9 min, at 90 °. This procedure truncated the aptamer randomly producing a sample that contained aptamer of every length, from one nucleotide to full length. T1 digestion was carried out under denaturing single-hit conditions (7 M Urea, 50 °C), using 2 Units (U) T1, in 20 mM Sodium Citrate, 1 mM EDTA, for 10 min, to produce a sample containing aptamers truncated at each G. These two samples were used in conjunction with the known aptamer sequence to assign nucleotide identity when the products were subjected to denaturing acrylamide gel electrophoresis. A fraction of the  $\text{OH}^-$  digestion was heated to 80 °C then slowly cooled to room temperature to produce correct folding and incubated with 30 nM HXB2 RT for 10 min at 37 °C. Unbound aptamer was washed away and bound aptamer was recovered.

Approximately 20,000-40,000 cpm of each sample, and an unreacted labeled RNA sample, were subjected to gel electrophoresis on 10% denaturing (8 M Urea) polyacrylamide gel. The RNA was fixed in the gel with 10% ethanol, 10% acetic acid and dried under vacuum, at 80 °C, for 30 min. The signal was collected using a Fujifilm Imaging Plate and recorded on a Fujifilm FLA5000 imaging system. Data were collected using Fujifilm Multi Gage V3.0 imaging software. 5' functional boundaries were determined using 3'-labeled RNA treated as described for 3' boundary assay above.

**Aptamer Binding to HIV-1 RT.** 5'-labeled aptamers (10,000 cpm) were heated to 80 °C then slowly cooled to room temperature to produce correct folding. They were incubated for 10 min with RT from HIV strains BH10 and 94CY017 at the concentration of 0, 0.1, 0.3, 1.0, 3.0, 10, 30, 100, and 300 nM at 37 °C and then allowed to cool at room temperature for 5 min. They were blotted under vacuum using a Whatman Mini-fold I Dot-Blot System on a pre-wet 0.45 µm nitrocellulose filter pre-soaked in protein binding buffer (200 mM KOAc, 50 mM Tris-HCl, 6 mM MgCl<sub>2</sub>, and 10 mM DTT, pH 8.0). After application of vacuum, the wells were pre-washed with 200 µL of binding buffer, then followed by the 50 µL sample, and washed twice with 300 µL volumes of wash buffer (same as binding buffer but without DTT). Ten thousand cpm was blotted directly on the filter for 100% control. The filter was dried at room temperature for 1 hour and was exposed to a Fujifilm Imaging Plate overnight and recorded on a Fujifilm FLA5000 imaging system. Data were collected using Fujifilm Multi Gage V3.0 imaging software. Alternatively, pre-folded aptamers were incubated for 10 min with RT from HIV strain HXB2 at the concentrations of 0, 10, and 30 nM at 37 °C and then allowed to cool at room temperature for 5 min. Nitrocellulose filters (0.45 µm) were pre-wet and treated

with 500  $\mu$ L binding buffer under vacuum. The 50  $\mu$ L aptamer sample was added and the sample was washed with two 500  $\mu$ L volumes of wash buffer. Samples were measured on a PerkinElmer TriCarb scintillation counter as  $^{32}$ P counts per minute (cpm) and compared with 100% control (10,000 cpm).

**S1, V1, T1 secondary structure analysis.** 5'-labeled RNA (~50,000 cpm) was OH<sup>-</sup> and T1 digested as described above to produce a ladder for assigning nucleotide identity. Equal quantities of RNA also underwent digestion by S1 Nuclease, V1 RNase, and T1 under single-hit native conditions. RNA samples were heated to 80 °C, and then native buffer was added, and slowly cooled to room temperature to produce correct folding. 0.10 U of T1 was added and incubated for 5 min. RNA was incubated with 90.0 U of S1 for 10 min. V1 was incubated with the sample for 10 min at 10  $\mu$ U. All reactions were carried out at 37 °C and stopped with 95% formamide 50 mM EDTA. Reactions were subjected to denaturing acrylamide gel electrophoresis and visualized as described above for boundary reactions.

**Structure Population Analysis.** Aptamer RNA samples were heated to 80 °C then slowly cooled to room temperature to produce correct folding before electrophoresis on 6% native polyacrylamide gel. Gels were incubated in 1 mM Ethidium Bromide for 10 min and visualized under UV light.

## References

1. UNReport. (2007) HIV infection on the rise worldwide. *Clin. Infect. Dis.* **44**, 3-4
2. Abdool Karim, S., Abdool Karim, Q., Gouws, E., and Baxter, C. (2007) Global epidemiology of HIV-AIDS. *Infect. Dis. Clin. North Am.* **21**, 1-17
3. Tuerk, C. and Gold, L. (1990) Systematic evolution of Ligands by exponential enrichment: RNA Ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510
4. Ruckman, J., Green, L., Beeson, J., Waugh, S., Gillette, W., Henninger, D., Claesson-Welsh, L., and Janjic, N. (1998) 2'-fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor. *J. Biol. Chem.* **273**, 20556-20567
5. Mendonsa, S. and Bowser, T. (2004) In vitro evolution of functional DNA using capillary electrophoresis. *J. Am. Chem. Soc.* **126**, 20-21
6. Schneider, D., Feigon, J., Hostomsky, Z., and Gold, L. (1995) High-affinity ssDNA inhibitors of the reverse transcriptase of type 1 human immunodeficiency virus. *Biochemistry* **34**, 9599-9610
7. Kensh, O., Connolly, B., Steinhoff, H., McGregor, A., Goody, R., and Restle, T. (2000) HIV-1 reverse transcriptase-pseudoknot RNA aptamer interaction has a binding affinity in the low picomolar range coupled with high specificity. *J. Biol. Chem.* **275**, 18271-18278
8. Tuerk, C., MacDougal, S., and Gold, L. (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci.* **89**, 6988-6992
9. Burke, D., Scates, L., Andrews, K., and Gold, L. (1996) Bent pseudoknots and novel RNA inhibitors of type 1 human immunodeficiency virus (HIV-1) reverse transcriptase. *J. Mol. Biol.* **264**, 650-666
10. Joshi, P. and Prasad, V. (2002) Potent inhibition of human immunodeficiency virus type 1 replication by template analog reverse transcriptase inhibitors derived by SELEX. *J. Virology* **76**, 6545-6557
11. Nickens, D., Patterson, J., and Burke, D. (2003) Inhibition of HIV-1 reverse transcriptase by RNA aptamers in *Escherichia coli*. *RNA* **9**, 1029-1033
12. Chaloin, L., Lehmann, M., Sczakiel, G., and Restle, T. (2002) Endogenous expression of a high-affinity pseudoknot RNA aptamer suppresses replication of HIV-1. *Nucleic Acids Res.* **30**, 4001-4008
13. Held, D., Kissel, J., Saran, D., Michalowski, D., and Burke, D. (2006) Differential susceptibility of HIV-1 reverse transcriptase to inhibition by RNA aptamers in enzymatic reactions monitoring specific steps during genome replication. *J. Biol. Chem.* **281**, 25712-25722
14. Jaeger, J., Restle, T., and Steitz, T. (1998) The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *EMBO J.* **17**, 4535-4542
15. Kissel, J., Held, D., Hardy, R., and Burke, D. (2007) Active site binding and sequence requirements for inhibition of HIV-1 reverse transcriptase by the RT1 family of single-stranded DNA aptamers. *AIDS Res Human Retroviruses* **23**, 699-708



16. Fisher, T., Joshi, P., and Prasad, V. (2005) HIV-1 reverse transcriptase mutations that confer decreased in vitro susceptibility to anti-RT DNA aptamer RT1t49 confer cross resistance to other anti-RT aptamers but not to standard RT inhibitors. *AIDS Res. Therapy* **2:8**, 1-10
17. White, R., Sullenger, B., and Rusconi, C. (2000) Developing aptamers into therapeutics. *J. Clin. Invest.* **106**, 929-934
18. Jellinek, D., Green, L., Bell, C., Lynott, C., Gill, N., Vargeese, C., Kirschenheuter, G., McGee, D., Abesinghe, P., Pieken, W., Shapiro, R., Rifkin, D., Moscatelli, D., and Janjic, N. (1995) Potent 2'-amino-2'-deoxypyrimidine RNA inhibitors of basic fibroblast growth factor. *Biochemistry* **34**, 11363-11372.
19. Li, M., Kim, J., Li, S., Zaia, J., Yee, J., Anderson, J., Akkina, J., and Rossi, J. (2005) Long-term inhibition of HIV-1 infection in primary hematopoietic cells by lentiviral vector delivery of a triple combination of anti-HIV shRNA, anti-CCR5 ribozyme, and a nucleolar-localizing TAR decoy. *Mol. Therapy* **12**, 900-909
20. Lee, N. (2002) Expression of small interfering RNAs targeted against HIV-1 rev transcriptase in human cells. *Nat. Biotechnol.* **20**, 500-505
21. Novina, C. (2002) SiRNA-directed inhibition of HIV-1 infection. *Nat. Med.* **8**, 681-686
22. Jacque, J., Triques, K., and Stevenson, M. (2002) Modulation of HIV-1 replication by RNA interference. *Nature* **418**, 435-438
23. Coburn, G. and Cullen, B. (2002) Potent and specific inhibition of human immunodeficiency virus type 1 replication by RNA interference. *J. Virol.* **76**, 9225-9231
24. Surabhi, R. and Gaynor, R. (2002) RNA interference directed against viral and cellular targets inhibits human immunodeficiency virus type 1 replication. *J. Virol.* **76**, 12963-12973
25. Joshi, P., North, T., and Prasad, V. (2005) Aptamers directed to HIV-1 reverse transcriptase display greater efficacy over small hairpin RNAs targeted to viral RNA in blocking HIV-1 replication. *Mol. Therapy* **11**, 677-686
26. Green, L., Waugh, S., Binkley, J., Hostomska, Z., Hostomska, Z., and Tuerk, C. (1995) Comprehensive chemical modification interference and nucleotide substitution analysis of an RNA pseudoknot inhibitor to HIV-1 reverse transcriptase. *J. Mol. Biol.* **247**, 60-68
27. Held, D., Kissel, J., Thacker, S., Michalowski, D., Saran, D., Ji, J., Hardy, R., Rossi, J., and Burke, D. (2007) Cross-clade inhibition of recombinant human immunodeficiency virus type 1 (HIV-1), HIV-2, and simian immunodeficiency virus SIVcpz reverse transcriptase by RNA pseudoknot aptamers. *J. Virology* **81**, 5375-5384
28. Martin, S., Ullrich, R. and Meyer, W. (1986) A comparative study of nucleases exhibiting preference for single-stranded nucleic acid. *Biochim. Biophys.* **867**, 76-80
29. Boiziau, C., Dausse, E., Yurchenko, L., and Toulm'e, J. (1999) DNA aptamers selected against the HIV-1 *trans*-activation-response RNA element form RNA-DNA kissing complexes. *J. Biol. Chem.* **274**, 12730-12737

30. Boucard, D., Toulm'e, J., and Di Primo, C. (2006) Bimodal loop-loop interactions increase the affinity of RNA aptamers for HIV-1 RNA structures. *Biochemistry* **45**, 1518-1524
31. Brody, E. and Gold, L. (2000) Aptamers as therapeutic and diagnostic agents. *J. Biotechnol.* **74**, 5-13
32. Bugaut, A., Toulm'e, J., and Raynera, B. (2006) SELEX and dynamic combinatorial chemistry interplay for the selection of conjugated RNA aptamers. *Org. Biomol. Chem.* **4**, 4082-4088
33. Cao, S. and Chen, S. (2006) Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res.* **34**, 2634-2652
34. Ellington, A. and Szostak, J. (1990) *In vitro* selection of RNA molecules that bind specific Ligands. *Nature* **346**, 818-822
35. Famulok, M., Mayer, G., and Blind, M. (2000) Nucleic acid aptamers—From selection *in vitro* to applications *in vivo*. *Acc. Chem. Res.* **33**, 591-599
36. Hariri, S. and McKenna, M. (2007) Epidemiology of human immunodeficiency virus in the United States. *Clin. Microbiol. Rev.* **20**, 478-488
37. Held, D., Kissel, J., Patterson, J., Nickens, D., and Burke, D. (2006) HIV-1 inactivation by nucleic acid aptamers. *Frontiers Biosci.* **11**, 89-112
38. Huang, Z. and Szostak, J. (1996) A simple method for 3'-labeling of RNA. *Nucleic Acids Res.* **24**, 4360-4361
39. James, W. (2001) Nucleic acid and polypeptide aptamers: a powerful approach to ligand discovery. *Curr. Opin. Pharmacol.* **1**, 540-546
40. Liu, Y., Haasnoot, J., ter Brake, O., Berkhout, B., and Konstantinova, P. (2008) Inhibition of HIV-1 by multiple siRNAs expressed from a single microRNA polycistron. *Nucleic Acids Res.* **36**, 2811-2824
41. Milligan, J., Groebe, D., Witherell, G., and Uhlenbeck, O. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783-8798
42. Milligan, J. and Uhlenbeck, O. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enz.* **180**, 51-62
43. Patterson, J., Nickens, D., Burke, D. (2006) HIV-1 reverse transcriptase pausing at bulky 2' adducts is relieved by deletion of the RNase H domain. *RNA Biol.* **3:4**, 163-169
44. Sarafianos, S., Das, K., Tantillo, C., Clark, A., Ding, J., Whitcom, J., Boyers, P., Hughes, S., and Arnold, E. (2001) Crystal structure of HIV-1 reverse transcriptase in complex with the polypurine tract RNA:DNA. *EMBO J.* **20**, 1449-1461
45. Reeves, J. and Doms, R. (2002) Human immunodeficiency virus type 2. *J. of Gen. Virol.* **83**, 1253-1265
46. Chaudhry, M. and Weinfeld, M. (1995) Induction of double-stranded breaks by S1 nuclease, mung bean nuclease and nuclease P1 in DNA abasic sites and nicks. *Nucleic Acids Res.* **23**, 3805-3809

## VITA

Joshua David Franken was born November 16, 1983 to Ronald and Susan Franken at St. Mary's Hospital in Jefferson City, Missouri. He spent his entire adolescence in Linn, Missouri, growing up on the farm and attending the Linn public schools from kindergarten through 12<sup>th</sup> grade. During this time he was very involved in sports and extra-curricular activities, including being a Varsity member of the 1998 class 1A state championship cross-country team and playing the lead snare drum with the Marching Wildcats.

Josh traveled all over the country with his family on vacation. The Franken family has been to every lower 48 state except New Mexico and Arizona. The family also stayed in Canada, Mexico and the Bahamas. Josh's favorite vacation destinations from these travels include Yellowstone National Park, New Orleans and Washington D.C.

Josh chose the University of Missouri to pursue an undergraduate degree in biochemistry. He traveled to Tunisia in 2005 as an international scholar with the "Tunisia – Past and Present" program. He also worked in Dr. Deutscher's lab studying the effects of modified citrus pectin on melanoma cell aggregation. He wrote an Honors Thesis based on his undergraduate research. He continued to work in this lab after graduation until the beginning of graduate school. After slightly more than two years in graduate school, Josh is preparing for graduation and the next step. He plans to pursue work in sales and the life sciences upon receiving his graduate degree.