# **Targeting RNA Structures with Multivalent Branched Peptide Libraries**

# David Irby Bryson, Jr.

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

# Doctor of Philosophy In Chemistry

# Webster L. Santos, Chair Paul R. Carlier Felicia A. Etzkorn James M. Tanko

March 19, 2012 Blacksburg, VA

Keywords: RNA, HIV-1, Medium-Sized Ligands, Branched Peptides, Combinatorial Libraries, High Throughput Screening, Multivalency, Cell Permeable Peptides, and Boron.

Copyright © 2012 David I. Bryson, Jr.

#### Targeting RNA Structures with Multivalent Branched Peptide Libraries

# David Irby Bryson, Jr.

#### ABSTRACT

RNA is essential for the transfer of genetic information, as the central dogma of biology dictates. The role of RNA, however, is not limited to serving as an information shuttle between DNA and fully functional protein. Indeed, RNA has experienced a surge of interest in the field of chemical biology for its other critical roles in biology including those in control of transcription, translation, splicing, genetic replication, and catalysis. RNA has proven to be a difficult and complex target for the design of small molecular ligands because of its structural heterogeneity and conformational flexibility. Yet, the highly folded tertiary structures of these oligomers present unique scaffolds which designed ligands should be able to selectively target. To that end, two branched peptide libraries ranging in size from 4,096–46,656 unique sequences were screened for their ability to bind HIV-1 related RNA structures, the transactivation response element (TAR) and the Rev response element (RRE). In addition to discovering a mid-nanomolar branched peptide ligand for TAR, the first branched boronic acid peptide library designed to target RNA was screened for binding to RRE. Each of these efforts resulted in the identification of selective binders to their respective RNA targets, and the unnatural branching of these compounds was demonstrated to provide a multivalent binding interaction with the RNA. Furthermore, these compounds were shown to be cell permeable and displayed little to no cytotoxicity in HeLa and A2780 cells.

# Acknowledgements

I would first like to thank my Doctoral Advisor and friend Dr. Webster L. Santos. His guidance throughout my studies at Virginia Tech has been indispensible. I could not have asked for a better mentor. He has done everything possible to provide me with the tools needed to be a successful scientist in the future, and his tireless work has provided my research with the funding required to keep it moving forward. I am constantly amazed and appreciative of his availability to provide answers to my questions or concerns. It is rare that a late-night e-mail goes unanswered for more than an hour or two, and he has always made me feel welcome to knock on his office door for a brief chat. He has even been there for me through deeply personal issues, and I am sincerely grateful for that. I must also thank my current and previous Ph. D. Committee Members Drs. Carlier, Etzkorn, Reineke, Smith and Tanko. Your positive support and critiques throughout my five years at Virginia Tech have helped fuel my desire for research and bolstered my confidence as a scientist.

I would also like to thank my friends and colleagues in the Santos Group – Dr. Philippe Bissel, Dr. Ming Gao, Ken Knott, Mithun Raje, Michael Perfetti, Jason Crumpton, Jing Sun, Brandon Thorpe, Wenyu Zhang, Xi Guo, Jessica Wynn, Amanda Nelson, Molly Congdon, Joe Calderone, Emily Morris, Marietou Paye, Ryan Stephens, Evan Gilius, Leah Heist, Wes Morris, and Caitlyn Criss. I would also like to thank our collaborators on this project – Dr. W. Keith Ray, Dr. Patrick M. McLendon, Dr. David M. Rekosh, and Dr. Takashi Morii.

Prior to attending Virginia Tech, I was mentored by someone who I hold in the highest esteem for directing me onto the path I find myself on today. As an

iii

undergraduate, I had the opportunity to work for Dr. Alice Haddy at the University of North Carolina at Greensboro over the course of three years. Working in Dr. Haddy's lab showed me that not only did I have an interest in science and research, but that I was smart enough and skilled enough to pursue it as a career. At that time, I was once shocked to hear that she believed I should continue my education toward a Ph.D. That was the first time I had ever considered that I might have the capacity to achieve such an honor. For her belief in me, I am forever indebted. I must also take this opportunity to thank my Research Advisor during my Masters Degree, Dr. Gregory M. Raner, as well as Dr. Jason J. Reddick for their guidance and support.

Finally, I would like to thank my Mother, Father, and family. Your support and encouragement has been unwavering throughout my life, and you have always believed in me even when I didn't believe in myself. I love you all so much. Dedication

To Mom and Dad

# **Table of Contents**

List of Schemes	viii
List of Figures	ix
List of Tables	xii
Chapter 1 RNA as a Target for Molecular Binding	1
1.1 RNA as a Therapeutic Target	1
1.2 Targeting RNA	4
1.3 RNA–Protein Interactions	5
1.4 RNA–RNA Interactions	8
1.5 RNA–DNA Interactions	11
1.6 RNA Binding Interactions	12
1.7 Small Molecules	14
1.8 Large Molecules	18
1.9 RNA Targets	22
References	22
Chapter 2 Targeting Structured RNAs Using "Medium-Sized" Molecules	28
2.1 Introduction	28
2.2 The Third Molecular Class of RNA Binders: Medium-Sized Molecules	32
2.3 Medium-Sized Neomycin B-Arginine Conjugates and Aminoglycoside	
Dimers	34
2.4 Medium-Sized Peptides and Peptidomimetics	
2.5 Targeting RNA Using Modified Peptide Backbones	47
2.6 Medium Sized Cyclic Peptides	56
2.7 Conclusions	64
References	65
Chapter 3 Screening of Branched Peptide Library with HIV-1 TAR RNA	75
Abstract	75
3.1 Introduction	
3.2 Results and Discussion	77
3.3 Materials and Methods	

Chapter 4 Toward Targeting RNA Structure: Branched Peptides as Cell-	
Permeable Ligands to TAR RNA	98
Abstract	99
4.1 Introduction	99
4.2 Results and Discussion	103
4.3 Materials and Methods	121
References	128
Chapter 5 Compounds Discovered from a Branched Peptide Library Featur	ring
Unnatural Boronic Acid Side Chains Selectively Bind to HIV-1 RRE RNA	
Abstract	134
5.1 Introduction	135
5.2 Results and Discussion	138
5.3 Materials and Methods	162
References	171
Chapter 6 Conclusions	178
References	182
Appendix 1 Citations of Copyrighted Works	

# List of Schemes

3.1 On-bead screening of the branched peptide library against 5'-DY547	-labeled TAR
RNA and hit identification using MALDI/MS	80
4.1 Synthesis of fluorescein-labeled branched peptides by solid-phase pe	ptide synthesis
(SPPS)	105
5.1 Synthesis of the 3.3.4 Branched Boronic Acid Peptide Library by Sol	lid-Phase
Peptide Synthesis (SPPS)	140

# List of Figures

1.1 Hydrogen bonding interactions of spectinomycin, streptomycin, and paromomycin
with the 30S ribosomal subunit
1.2 Structural motifs available to RNA molecules
1.3 The structure of the HIV-1 Rev-response element (RRE)7
1.4 Crystal structure data of the hammer head ribozyme10
1.5 Crystal structures of <i>thi</i> -box riboswitch bound with metabolite and analogs17
1.6 Biogenesis of miRNAs
2.1 The structure and molecular weight of several aminoglycoside small molecules33
2.2 The structure of HIV-1 TAR (A) and the neomycin B-hexaarginine conjugate (B)
2.3 The structure of HIV-1 RRE-IIB (A) and neomycin dimer (B)
2.4 The structure of RRE66 (A) and affinities toward aminoglycoside dimers (B)38
2.5 The structure of the 16S A-site of <i>E. coli</i> and HCV IRES domain IV (IRES) with
their binding affinity to peptide <b>16</b> 44
2.6 Structural comparison of peptide and peptoid backbones (A), the structure of
peptoid/ <sub>D</sub> -peptide hybrid CGP6422 screened from a combinatorial library (B), and the
structures of peptoid amides and peptoid esters (C)48
2.7 The sequence of <b>Tat(47–57</b> ) from which oligocarbamate <b>35</b> (A) oligourea <b>36</b> (B)
and β-peptides (C) were derived51
2.8 The structure and binding affinity to TAR RNA of <b>Tat</b> (48–57), $\gamma$ -AA1, and $\gamma$ -AA2
2.9 The structure and sequence of HIV-1 TAR RNA and BIV TAR RNA (A) and
structure of BIV Tat peptide variant with binding affinities for BIV TAR
2.10 The structure, sequence, and affinity of <b>KP-Z-41</b> to HIV-1 TAR compared to L-
22
3.1 Sequence and secondary structure of HIV-1 TAR RNA (A), small molecule ligands
(neomycin B (B) and argininamide (C)) and 11-mer Tat-derived peptide (D) that inhibit
Tat-TAR interaction
3.2 Combinatorial library of 4,096 branched peptides synthesized by split and pool
technique

3.3 Visualization of hit beads under confocal microscope	81
3.4 Synthesis of FITC-labeled peptides	84
3.5 Fluorescence polarization binding curves of branched peptides with HIV-	I TAR
RNA and mutant (U24>C) TAR RNA	84
3.6 DY547 fluorescence intensity over auto-fluorescence of Tentagel Macrobe	ead-NH <sub>2</sub>
resin	89
3.7 Organic washed control branched peptide (KYR) <sub>2</sub> FDS	90
3.8 Optimizing screening conditions with control branched peptide	91
3.9 The 16 hits screened from the 3.3.3 library	93
3.10 The 10 hits from the 3.3.3 library, which were re-screened under stringer	nt
screening conditions	93
4.1 High-throughput screening of a branched peptide library with TAR RNA	eveals
peptide hits that are sequenced by MALDI-TOF	104
4.2 LC/ESI-MS of <b>FL15</b>	106
4.3 Dot blot assays of labeled peptides and HIV-1 TAR RNA	109
4.4 Titration curves comparing EMSA and dot blot assay with or without com	peting
tRNA	112
4.5 Competition of <b>FL4</b> with $10 \times$ and $1000 \times$ tRNA using EMSA used to gene	rate
binding curves observed in Figure 4.4	112
4.6 Sequence and secondary structure of TAR RNA and variants (A) and titra	tion
curves of <b>FL4</b> with these RNAs (B)	113
4.7 Dot blot assays of FL4 with TAR variants used to obtain binding curves in	ı
Figure 4.6	114
4.8 Hill plots of <b>FL4</b> dot blot data with native TAR RNA and TAR RNA varia	ints
	116
4.9 Cell permeability and cytotoxicity of FITC-labeled hit peptides	118
4.10 Confocal microscopy of fluorescent BPs in HeLa cells	120
5.1 Combinatorial (3.3.4) branched-peptide library featuring boronic acid side	chains
	139
5.2 Sequence and structure of RRE-IIB	141

5.3 Representative structures (from top to bottom) of <b>BB1-1</b> , <b>CBB1-1</b> , <b>CFBB1-1</b> ,
<b>BB1-8</b> , and <b>CBB1-8</b> (A) and dot blot assay curves with RRE-IIB (B)146
5.4 Sequence of <b>BB1-1</b> and linear variants (A) and titration curves of RRE-IIB with
these peptides (B)148
5.5 Sequence and structure of RRE-IIB and variants (A) and titration curves of <b>BB1-1</b>
with these RNAs (B)149
5.6 Hill plot analysis of <b>BB1-1</b> dot blot data
5.7 Titration curves of <b>BB1-1</b> with RRE-IIB and in the presence of competing tRNA
5.8 RNase protection assays of RRE-IIB using <b>BB1-1</b>
5.9 RNase V1 protection assay of RRE-IIB at G10, G17, and C18157
5.10 Cell microscopy from initial incubation of peptides in the presence of mounting
media158
5.11 Cellular uptake of <b>FBB1-1</b> into (A) Hela cells and (B) A2780 cells159
5.12 MTT assays with <b>BB1-1</b> and <b>BB1-8</b>
5.13 A2780 cell MTT assay 24 h time study as a function of the concentration of
<b>BB1-6</b>
5.14 A2780 cell MTT assay 24 h time study of <b>BB1-6</b> at 30 μM161

# List of Tables

2.1 Sequences of peptides featuring <i>N</i> , <i>N</i> -dimethyl Lys and their binding affinity to
several RNAs
2.2 Sequences of peptides featuring <i>N</i> -acridinyl Lys and their binding affinity to
several RNAs
2.3 The sequences, percent helicity, and $K_d$ values of covalently cross-linked peptides
against hairpin RNA structures
2.4 The sequences, percent helicity, and $K_d$ values of covalently cross-linked peptide
dimers against hairpin RNA structures
2.5 Dissociation constant for <b>Tat</b> ( <b>47–57</b> ) and compounds <b>37–39</b> from wild-type or
bulgeless TAR
2.6 The sequence of <b>BIV Tat(65–81</b> ) and the backbone cyclized mimetics with their
binding affinities to BIV TAR RNA60
2.7 The sequences of lead peptides L-22, L-50, and L-51 and their affinities toward
HIV-1 and BIV TAR RNA
3.1 Sequence of branched peptides <b>BP1–BP16</b> derived from MALDI/MSMS82
3.2 Conditions of first incubations for controls A–C90
3.3 Conditions for incubations of controls D–E91
4.1 Binding constants and molecular weights of hit BPs108
5.1 Dissociation Constants ( $K_d$ ) and Molecular Weights of Hit BB Compounds143

# Chapter 1RNA as a Target for MolecularBinding

# **1.1 RNA as a Therapeutic Target**

Ribonucleic acid (RNA) plays an essential role in the transfer of genetic information from deoxyribonucleic acid (DNA) to the molecular machinery that synthesizes proteins. The function of RNA, however, is not limited to serving as an information shuttle in the form of messenger RNA (mRNA) for the sole purpose of gene expression. In fact, RNA molecules are utilized in nature for a large number of biological processes including genetic replication, regulation of transcription, RNA stabilization, RNA localization, RNA transport, pre-mRNA splicing, metal chelation, and translation of mRNA into the encoded amino acid sequence in the form of transfer RNA (tRNA) and ribosomal RNA (rRNA).<sup>1,2</sup> Because RNA is involved in so many biological processes, it is no wonder that the goal of being able to design synthetic molecules that can selectively target a specific type of RNA is highly desired for antiviral, antibiotic, and gene silencing capabilities.<sup>3</sup>

The history of targeting RNA begins with the ribosome, the biological macromolecule responsible for the translation of mRNA into protein. The ribosome is a 2.5 MDa ribonucleoprotein, of which two-thirds is comprised of RNA.<sup>4</sup> Aminoglycoside antibiotics bind to the RNA component of the ribosome in prokaryotes. The aminoglycoside streptomycin was simultaneously the first antibiotic to be discovered in the aminoglycoside family, and the first drug ever to be discovered through screening

methods.<sup>5</sup> Although streptomycin was discovered in 1943, it was not until 2000 that the first X-ray crystal structure of a bacterial 30S ribosomal subunit-antibiotic complex was solved for three different antibiotics, including spectinomycin as well as the aminoglycosides streptomycin and paromomycin. Interestingly, all three of the antibiotics targeted different regions of the 30S subunit, and functioned in different ways.<sup>6</sup>

The prokaryotic ribosome contains three tRNA binding sites in the 30S subunit, which also bind the mRNA that is being translated. The A-site is named for its aminoacyl-tRNA substrate. Peptidyl-tRNA binds to the P-site, and the E-site is the final binding site before the tRNA exits the ribosome. Spectinomycin inhibits the translocation of peptidyl-tRNA from the A-site to the P-site by binding to the minor groove of 16S rRNA helix 34 through H-bonding (Figure 1.1a). Streptomycin was shown to stabilize the ribosomal ambiguity form of the helix 27 accuracy switch, which increases affinity for non-cognate tRNAs and impedes proof-reading in the A-site. This is accomplished through H-bonding and salt bridges with rRNA from several helices (Figure 1.1b). These interactions ultimately cause bases A1492 and A1493 to flip out of helix 44, and these bases are involved in codon-anticodon recognition.<sup>6,7</sup> Paromomycin also impacts the ability of the ribosome to carry out faithful translation by binding to helix 44 of the A-site (Figure 1.1c), which aids in flipping out bases A1492 and A1493.<sup>6</sup> Not only did these studies help to unlock mysteries of translation, but they also furthered the knowledge of how small molecules can bind to RNA.



**Figure 1.1.** Hydrogen bonding interactions of spectinomycin, streptomycin, and paromomycin with the 30S ribosomal subunit. A) Diagram of interactions between spectinomycin and specific residues of the 30S. B) Diagram of interactions between streptomycin and specific residues of the 30S. C) Diagram of interactions between paromomycin and specific residues of the 30S. [Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **2000**, *407*, 340-348], Copyright 2000; [Used with permission from Nature Publishing Group]

# **1.2 Targeting RNA**

Although there have been drugs that target rRNA since the 1940's, the rational design of molecules that will bind with specificity to RNA is still in its infancy.<sup>8</sup> Targeting RNA with synthetic molecules is, in some ways, like targeting proteins more than it is like targeting DNA. Both RNA and proteins can adopt unique secondary and tertiary structures that can be utilized as targets (Figure 1.2).<sup>3,5,9</sup> RNA, however, only has 4 molecular building blocks, whereas proteins have 20, many of which can also be posttranslationally modified.<sup>10</sup> In that respect, RNA is much more limited than proteins, but the purine and pyrimidine bases of RNA are unique from the amino acids because they are each capable of  $\pi$ -stacking interactions, as well as Watson-Crick, Hoogsteen, and non-canonical base pairing.<sup>1,3,8</sup> The phosphate backbone of RNA molecules also provides a point for electrostatic interactions to occur between the target and ligand.<sup>3</sup> If the structure of a particular target RNA can be elucidated, then these types of three-dimensional structures could be taken advantage of when designing molecules.



**Figure 1.2.** Structural motifs available to RNA molecules. [Zaman, G. J. R.; Michiels, P. J. A.; van Boeckel, C. A. A. Targeting RNA: New opportunities to address drugless targets. *Drug Discov. Today* **2003**, *8*, 297-306], Copyright 2003; [used with permission from Elsevier]

When designing molecules with selectivity in mind, there is much to be learned from the molecular interactions between RNA and endogenous molecules. *In vivo*, there are interactions between RNA and proteins, RNA and other RNA molecules, as well as interactions between RNA and DNA molecules. The general trends of binding used in nature can be utilized to develop binding strategies for a target RNA.

# **1.3 RNA–Protein Interactions**

Thymidylate synthase is an enzyme that is essential for the biosynthesis of the DNA precursor thymidylate. In 1991, it was discovered that this enzyme is inhibited by its own mRNA, which could be a way of regulating its own translation.<sup>11</sup> Later, the precise binding sites on the thymidylate synthase mRNA (TS mRNA) were found via binding assays with a series of the truncated TS mRNA.<sup>12</sup> The first binding site was in the first 188 nucleotides containing the AUG start codon as well as the 5' untranslated region (UTR). The AUG start codon and the preceding three nucleotides GCC were found to be essential for binding. This GCCAUG sequence was found on the loop domain of a 36-nucleotide stem-loop structure, and the loop domain was shown to be the main contributor to binding, while the stem structure was less important. The second binding site was suggested to also have a stem-loop structure and is located within the coding region of the mRNA.<sup>12</sup>

Another example in nature of RNA structure that is essential to activity is found in Human Immunodeficiency Virus type-1 (HIV-1) RNA. HIV-1 encodes many different genes, some of which overlap each other. Therefore, some genes in HIV-1 require the RNA transcript to be spliced before they are functional for translation.<sup>13</sup> In this process, intervening regions of RNA, called introns, that separate the beginning and end of a gene are cut out so that the gene can be pieced together to become a functional mRNA, which is then exported from the nucleus to the cytoplasm. The full-length RNA transcript has several functions as an unspliced molecule. First, it functions as the viral genome in retroviruses such as HIV-1. Second, the full-length transcript is used in the translation of the proteins Gag and Gag-Pol. The unspliced RNA must be transported out of the nucleus into the cytoplasm without being spliced to function in these capacities.

A 116 amino acid protein encoded by HIV-1, called Rev, is responsible for the activation of RNA export from the nucleus to the cytoplasm.<sup>13</sup> The full length HIV-1 RNA target contains a 351-nucleotide Rev-response element (RRE) located in the env gene of HIV-1. The RRE possesses a single high affinity site in stem loop IIB to which Rev can bind (Figure 1.3).<sup>13-16</sup> NMR studies demonstrated that a 34-nucleotide high affinity site forms a hairpin stem region with an A-form RNA helix.<sup>16</sup> The major groove of the stem structure is where the  $\alpha$ -helical arginine-rich binding domain of Rev can interact with the substrate RNA. Because the RNA A-form helix has a deep and narrow groove, the side chains of the  $\alpha$ -helical binding domain of Rev would be too sterically hindered to penetrate into the groove and interact with the ribonucleotide bases. The RRE, however, utilizes non-Watson-Crick base pairs; one between G48 and G71 and another between G47 and A73. These base pairs cause a 5Å widening of the major grove at an internal loop/bulge, which accommodates three to four turns of the bulky Rev  $\alpha$ helix. There are four essential residues for binding. Arg35 and Arg39 are found within hydrogen bonding distance of three bases, U66, G67, and G70 on one side of the major groove. The two other essential residues, Asn40 and Arg44, are in hydrogen bonding distance with U45, G46, G47, and A73 on the other side of the groove. Additional residues, including one Thr and six Arg residues, are within hydrogen bonding or electrostatic distances of the phosphate backbone.



**Figure 1.3.** The structure of the HIV-1 Rev-response element (RRE). [Yuan, C.; Xinyong, L.; De Clercq, E. Cessation of HIV-1 transcription by inhibiting regulatory protein Rev-mediated RNA transport. *Curr. HIV Res.* **2009**, *7*, 101-108], Copyright 2009; [used with permission from Bentham Science Publishers Ltd]

A peptide fragment (RSG-1.2), which is rich in Arg and has been selected for binding to the HIV-1 RRE, causes a conformational change in the RRE that is unique from the conformation imparted by the natural Rev binding domain.<sup>17</sup> This is accomplished by binding U72, which is not utilized by Rev. Trapping the target RNA into unnatural conformations by utilizing flexible portions of target RNA, especially by

binding strongly to bases that are not bound by the natural ligand, could be effective as a general RNA targeting strategy.

Most RNA binding proteins take advantage of the direct read-out of base sequences, as well as the three dimensional structure of the target RNA.<sup>18</sup> Often, conformational changes in the RNA are also observed upon binding that further accommodate the ligand and increase the binding affinity.

# **1.4 RNA–RNA Interactions**

An example of RNA-RNA interaction that has revolutionized gene silencing is RNA interference (RNAi). This is an effective method of mRNA degradation and suppression of gene expression. In the case of mRNA degradation, double stranded RNA (dsRNA) molecules called small interfering RNAs (siRNAs) initiate mRNA cleavage. siRNAs are between 21 and 25 nucleotides in length and are generated by the cleavage of longer dsRNAs. These long dsRNAs are transcribed from exogenous plasmid DNA after transfection of the cell. Cleavage of the long dsRNA is achieved by an enzyme called Dicer in the cytoplasm.<sup>19</sup> The siRNA strand that is complementary to the target mRNA is recruited into the RNA-induced silencing complex (RISC). The newly incorporated siRNA then binds to its target mRNA sequence, whereupon the ribonucleoprotein complex RISC cleaves the complementary mRNA sequence. This form of RNA-RNA interaction is based purely on Watson-Crick base pairing, although few mismatches are tolerated.<sup>19</sup>

Another type of RNAi is that of the microRNAs (miRNAs). These oligomers begin as endogenous hairpin RNAs called pre-miRNA, and contain non-Watson-Crick base pairing in their stem structures resulting in internal bulges.<sup>19</sup> The mature miRNAs function differently than siRNAs, and in animals, base complementarity is limited to ~6 nucleotides between the 5' region of the miRNA and the 3' region of its target mRNA.<sup>20</sup> Like with siRNAs, RISC incorporates miRNA, but because the miRNA has incomplete complementarity to the target mRNA, only translation is blocked and cleavage is not induced.<sup>20</sup>

Another example of an RNA-RNA interaction is that of the RNA catalysts, ribozymes. Just as proteinaceous enzymes adopt a tertiary structure in order to function, ribozymes also fold into complex structures. Because of the phosphate backbone's electrostatic charge, neutralizing metal ions are incorporated into the structure, which helps stabilize the RNA folds by eliminating the repulsive forces.<sup>21</sup> The hammerhead ribozyme catalyzes the site-specific reversible cleavage of its own phosphate backbone; the 2'-hydroxyl attacks its adjoined 3'-phosphate to release 5'-hydroxyl and results in a 2',3'-cyclic phosphate.<sup>22</sup> This ribozyme is comprised of three helical stem structures conjoined at a single helical intersection. The helical stem-I contains a bulge structure that, when magnesium ion concentration is optimal, interacts with a loop structure at the end of stem-II.<sup>21,23</sup> The contacts between these two structures are comprised of noncanonical base pairing and base stacking, winding stems I and II onto one another with bases interdigitating between the two structures (Figure 1.4a). The folding of the ribozyme induces a conformational change in the catalytic core at the three-stem intersection. It is suggested that this brings the necessary G12 and G8 bases into proximity and correct alignment with the scissile phosphate for general acid-base catalysis to be possible (Figure 1.4b).<sup>23</sup>



**Figure 1.4.** Crystal structure data of the hammer head ribozyme. A) Interactions of stems I, II, and III. B) Positioning of nucleotides for general acid-base catalysis. [Martick, M.; Scott, W. G. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **2006**, *126*, 309-320], Copyright 2006; [used with permission from Elsevier]

The molecular interactions of RNAi and the ribozyme, taken together, show that RNA-RNA interactions can depend on the sequence readout of the target, as well as the tertiary structure. Furthermore, the structure function-relationship in the catalytic RNAs, such as the hammerhead ribozyme, rivals that of proteins. Like proteins, these RNA molecules fold into unique tertiary structures and switch between active and inactive conformations upon binding of effectors. The similarity in RNA-RNA interactions and those of the protein-RNA interactions underlines how congruent strategies can be applied toward RNA and protein targets. Like proteins, local charges, hydrogen bonds, and folds exploited binding structural be in many RNA targets. can to

# **1.5 RNA–DNA Interactions**

RNA–DNA interactions are essential to life because it is through such contacts that the RNA polymerase transcribes the genetic information encoded in DNA into RNA. Viruses also utilize RNA-DNA interaction through the action of reverse transcriptase, which transcribes viral RNA into DNA. Other vital enzymes such as telomerase require similar interactions.<sup>24</sup> In these examples, RNA and DNA single strands form double-helical structures held together by Watson-Crick base-pairing.<sup>25</sup> Triple helical structures have also been shown as possible formations, where an RNA single strand forms a right-handed triple helix with the DNA double strand.<sup>26</sup> Watson-Crick base pairing stabilizes the DNA strand, but the RNA strand binds in the major groove of the DNA double helix and is stabilized by Hoogsteen-type hydrogen bonds. Thus, complementation of the three strands is the major factor for formation of the triple helix hybrid structure.

Other structures of RNA-DNA complexes have also emerged. In one instance, the end base pairs of two identical RNA-DNA hybrid helixes intercalate into one another resulting in a dimer of RNA-DNA helixes stacked end-on-end. Although this has been observed in the crystal structure, the biological implications need to be resolved.<sup>27</sup>

These examples of RNA-DNA hybridization show the relatively limited range of intermolecular interactions available between these two types of molecules. Tertiary structures of RNA that have been previously discussed do not seem to be involved with DNA binding. Base compatibility between the two oligomers is what ultimately determines binding affinity, whether by Watson-Crick base-pairing, non-canonical basepairing, or by base intercalation.

11

# **1.6 RNA Binding Interactions**

RNA–RNA and RNA–Protein interactions are able to take advantage of the direct read-out of the RNA strand. They also exploit the unique tertiary structures of RNA, where pockets, loops, and bulges become viable targets. This opens avenues to researchers for binding a single RNA target selectively and with binding specificity. The aminoglycosides are one of the oldest known RNA binding molecules, and knowledge of their intermolecular interactions has grown to the point that they can provide an adequate model for researchers when designing new RNA binding molecules.

#### Electrostatic Interactions

Electrostatic interactions are one of the major binding interactions found between the aminoglycosides and their target rRNA. One study suggests that electrostatic interactions comprise at least half of the total binding energy observed for paromomycin and geneticin. This conclusion was drawn from the dependence of binding on Na<sup>+</sup> concentration.<sup>28</sup> Salt concentration in solution can mediate molecular binding to the negatively charged phosphate backbone of target RNA. Other researchers' observations corroborate these results, where increasing the number of amine groups on many aminoglycosides also increases affinity for target RNA.<sup>29</sup> Although, electrostatic interactions may be involved largely in tight binding of aminoglycosides, it can also be detrimental. All RNA molecules have negatively charged backbones, so too much reliance on electrostatics can lead to promiscuous binding as in the case of neomycin, which can bind to multiple regions of its target RNA.<sup>30</sup>

#### Non-ionic Interactions

The negatively charged backbone is a ubiquitous feature of RNA. Targeting that feature with positively charged molecules may seem like an obvious route to binding. Exploitation of RNA's non-ionic features, however, is a less apparent tactic, and it has been demonstrated that aminoglycosides can take advantage of this. Crystal structures of several aminoglycosides showed that they were able to bind inside of the helix and form stacking interactions with a guanine base.<sup>31</sup> Binding in the groove of helical RNA was also described above in a similar interaction between the HIV-1 RRE RNA and the Rev protein.

#### Pseudo-Base Pair Interactions

Much like nuclotide bases can hydrogen bond in Hoogsteen type interactions, aminoglycosides can also use similar contacts for binding to ribonucleobases. The sugar ring I of nine aminoglycosides were shown to form hydrogen bonds with Watson-Crick sites of A1408 in the minimal A site helix of bacterial ribosome 16S rRNA.<sup>31</sup> These types of interactions were also discussed above between the HIV-1 RRE RNA and the binding domain side-chains of the Rev protein, which hydrogen bonded with bases in the groove of the helix.<sup>16</sup>

#### Structural Complementarity and Conformational Adaptation

There is some amount of controversy surrounding the notion of designing a molecule to fit a specific conformation of target RNA, or inducing a conformational switch in both molecules upon a binding event. It has been proposed that any three-

dimensional structure of RNA would be difficult to target with a great degree of specificity because of the electrostatic interactions that dominate binding. Additionally, most pockets and folds in RNA structures are likely to be anionic in nature.<sup>3</sup> One could envision how small molecules such as aminoglycosides would be at a disadvantage for utilizing secondary structure. These molecules' size is a limiting factor by reducing their sampling area of the target. Thus, individual pockets of charge, dispersed over a distance on a conformationally rigid RNA, could not be exploited simultaneously. Additionally, many of the secondary structures of RNA are relatively flexible, so conformational changes can be induced in the target upon binding.<sup>32</sup> This is especially seen when larger proteins bind to target RNAs.<sup>1</sup> This inherent flexibility may allow for different types of RNA to adjust in order to bind tightly to the same effector molecule with a cost of reduced selectivity.

# **1.7 Small Molecules**

The use of small molecules (MW < ~1000) provides an alluring route to targeting RNA. Synthetically, these molecules have less functional groups to deal with than larger molecules, resulting in a relative ease of synthesis. These molecules can also be tailored to allow for cell permeability, increasing potential to be pharmacologically effective. A large number of RNA binding small molecules have been developed, and there are several classes of binders.<sup>2,3,5</sup> The aminoglycosides have been discussed above, and this class of small molecules causes miscoding during translation by binding at the A-site of the decoding region of rRNA. Interestingly, all of these molecules except streptomycin are specific for A1408, which is the only base in the A-site of eubacteria that is different

in eukaryotes (a guanine).<sup>1,31</sup> Furthermore, the A-site has a secondary structure comprised of a helix, with two asymmetric internal loops, and A1408 is located on one of the internal loops. Asymmetric loops often have no secondary structure in solution, but their flexibility probably plays a role in binding of aminoglycosides.

Another class of small molecule binders is the intercalators. Classical DNA binding intercalators will bind to the RNA helix, but these are preferential toward DNA. Other molecules bind tightly in the RNA stem structure.<sup>3</sup> Targeting the stem structure of an RNA target is not limited to intercalation. Ionic association is also an important route of RNA binding with molecules of this size. Such molecules are thought to bind to the RNA helix preferentially over DNA due to the inability of the ligand to pack into the minor groove of the DNA double helix.<sup>33,34</sup>

Riboswitches are an interesting class of genetic switches found in the 5'-UTR of selective mRNAs that bind to small molecules for the purpose of regulating gene expression.<sup>35</sup> Most riboswitches function as feedback loops, where the same metabolite that the translation product is responsible for generating is the ligand that binds to the mRNA. This binding event prevents gene expression and ultimately stops the biosynthesis of the metabolite.<sup>36</sup> These genetic switches can be modular, where the riboswitch contains one or more metabolite binding aptamers, and a second module called a platform that undergoes a structural change. In some cases, the induced structural change can prevent RNA polymerase from continuing transcription elongation, while in other cases, the translation of the mRNA into protein is shut down at the ribosome. Riboswitches are not limited to a modular form. The platform and aptamer are not always distinct from one another.<sup>36</sup> mRNAs are also not limited to a single

riboswitch in the 5'-UTR. Having tandem riboswitches or a single riboswitch containing multiple aptamers improves responsiveness to lower concentrations of metabolite.<sup>35</sup> The mRNA of the *metE* gene from *Bacillus clausii* contains tandem riboswitches that are unique from most other examples–they bind independently to two different ligands, *S*-adenosylmethionine and coenzyme  $B_{12}$  to trigger transcription termination.<sup>37</sup> After ligand binding, single hairpin structures are formed, called terminator stems located immediately downstream from each one of the aptamers.

Many aptamers of riboswitches are finely tuned to recognize their substrate. The thi-box was one of the first riboswitches to be discovered, and it binds thiamine pyrophosphate (TPP) with high affinity and specificity. Many thi-box aptamers can distinguish between TPP and thiamine monophosphate (TMP) with up to 1000-fold greater affinity, and they can also distinguish between TPP and positively charged thiamine to the same extent.<sup>38</sup> Crystal structures of the *thi*-box with TPP, TMP, or thiamine have uncovered intermolecular interactions that allow for tight and selective binding.<sup>39</sup> Phylogenetically conserved residues of the aptamer bury most of the natural substrate. The aminopyrimidine ring of TPP and TMP base stacks between A43 and G42, and hydrogen bonds to G40 of the pyrimidine sensor helix. The pyrophosphate portion of TPP is recognized by two divalent metal cations coordinated to water and several bases of the pyrophosphate sensor helix. Many of these coordinated water molecules are lost in the TMP bound structure, although distances between ligand and metal cations are mostly unchanged (Figure 1.5a–b). To accomplish this, however, the RNA must become more compact to achieve binding with TMP. Additionally, TMP becomes more elongated to bridge the gap between the two sensor helices. Finally,

binding to the pyrophosphate sensor helix is completely lost in the case of pyrithiamine (PT), which has no phosphates (Figure 1.5c). This observation would account for poor binding of thiamine to the *thi*-box.



**Figure 1.5.** Crystal structures of *thi*-box riboswitch bound with metabolite and analogs. A) TPP. B) TMP. C) PT. [Edwards, T. E.; Ferre-D'Amare, A. R. Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure* **2006**, *14*, 1459-1468], Copyright 2006; [used with permission from <u>Elsevier</u>]

Although there have been many small compounds reported that bind to RNA, there have been relatively few studies on their intermolecular interactions. Some small compounds have been identified that bind to RNA hairpin loop structures, which account for the most RNA-protein interactions second only to duplex regions.<sup>3</sup> Of the molecules that do bind hairpin loops, the contribution of electrostatic interactions to binding is only ~25% or less, in contrast to > 50% for aminoglycosides.<sup>40</sup> Riboswitches can achieve mid-picomolar binding affinity ( $K_d$ ) and excellent specificity for their ligands because these systems have evolved naturally over immense lengths of time.<sup>35,41,42</sup> Designing small molecules that selectively bind to a target RNA that has no natural small ligand is

an arduous task. Thus, binding trends for individual RNA structures are still being brought into focus, and much work lies ahead in this field.<sup>3</sup>

# **1.8 Large Molecules**

Over the past decade there has been an increased interest in targeting RNA with large molecules in the form of RNAi. As mentioned above, RNAi is a method of using dsDNA for silencing genes. The single target of this technique is mRNA in the cytoplasm, where the mode of action is through endogenous gene silencing machinery. This phenomenon was initially uncovered in 1998 by experiments with *Caenorhabditis elegans*.<sup>43</sup> These researchers found the unexpected result that using dsRNA was ten times more effective for gene silencing than using just the sense or antisense strand alone. The initiating step in the pathway is cutting of the dsRNA trigger into siRNAs of ~22 nucleotides in length by the evolutionarily conserved RNase III enzyme named Dicer. These antisense strands can be incorporated into RISC to bind to the target mRNA template and activate degradation by the ribonucleoprotein complex. This mode of action depends simply on base-pairing between the two interacting strands, and while it may seem that siRNAs would be very specific, the situation is confounded by another form of RNAi, the miRNAs.

The same molecular machinery is utilized in processing the pre-miRNA hairpin into miRNAs, but the mode of action is slightly different from that of siRNA (Figure 1.6). As discussed above, miRNAs function with RISC by binding to the 3' region of the mRNA target with incomplete complementarity. Through this interaction, translation of mRNA is blocked. Thus, base complementarity between RNA target and the fully processed RNAi molecule is ultimately the determining factor of whether mRNA is cleaved, or if translation of mRNA is blocked. Therefore, siRNA and miRNA can be functionally interchangeable, where an siRNA may have incomplete complementarity to an off-target RNA and block translation just as an miRNA could have complete complementarity to a region in some off-target RNA to induce cleavage by RISC (Figure 1.6).<sup>44,45</sup>



**Figure 1.6.** Biogenesis of miRNAs. Mature miRNA can bind to RISC and repress translation of targets with incomplete complementarity, or it can induce cleavage of sequences with complete complementarity. [Esquela-Kerscher, A.; Slack, F. J. Oncomirs – microRNAs with a role in cancer. *Nat. Rev. Cancer* **2006**, *6*, 259-226], Copyright 2006; [used with permission from <u>Nature Publishing Group</u>]

Single stranded oligonucleotides are also used for gene silencing in antisense technology. Typically these strands range from 13 to 25 nucleotides in length, and can easily reach molecular weights over 2000 Da. These oligomers bind to their target mRNA through base sequence complementarity in order to induce enzymatic degradation of the RNA-DNA double helix. This degradation is accomplished through hydrolysis by the ubiquitous enzyme, ribonuclease H (RNase H). Alternatively, antisense strands can function via steric blockade, where splicing or translation associated proteins cannot proceed.<sup>47</sup> In this technique, the oligonucleotides can be comprised of natural or modified bases, although unnatural bases have significant advantages over their natural counterparts. Natural oligonucleotides are extremely sensitive to endonucleases and exonucleases that can degrade the oligomer, and the products of these reactions can have cytotoxic properties.<sup>48</sup> The unnatural antisense oligomers are generally comprised of methylphosphonates,<sup>49-51</sup> phosphorothioates,<sup>52,53</sup> peptide nucleic acids (PNAs),<sup>54-56</sup> phosphorodiamidate morpholino oligomers,<sup>57</sup> N3' $\rightarrow$ P5' polynucleotides,<sup>58</sup> and 2'-Omethyloligoribonucleotides.59-61

The methylphosphonates and other 2'-O-alkyl antisense oligomers are not widely useful.<sup>47</sup> The backbone of these compounds has increased cellular stability, but also results in bending of the double helix; causing reduced affinity of the hybridized structure toward RNase H.<sup>61,62</sup> This class of antisense molecules utilizes an RNase H independent mechanism of action.<sup>47</sup> The phosphorothioates are the most well studied antisense oligonucleotides since they have good resistance to nucleases as a single strand. There is some difficulty associated with these molecules, because replacement of non-bridging oxygen with sulphur introduces a chiral center. This can be problematic because only

one diastereomer hybridizes well with target mRNA. Additionally, their negatively charged backbone assists in solubility, but nonspecific interactions with other cellular components is a common problem that can lead to cytotoxic effects.<sup>47</sup> PNAs. phosphorodiamidate morpholino oligomers,  $N3' \rightarrow P5'$  polynucelotides, and 2'-Omethyloligoribonucleotides are second-generation antisense oligonucleotides. **PNAs** work through steric blockade so that translation of a tightly bound mRNA target cannot occur. These unique molecules lack a phosphate ribose backbone and tight binding is attributed to resulting lack of charge. Like PNAs, phosphorodiamidate morpholino oligomers have an uncharged backbone, and a morpholine ring replaces the ribose sugar.<sup>57</sup> This molecule also works through steric blockade, but its shortcoming is that it has poor cellular uptake. Finally, N3' $\rightarrow$ P5' polynucleotides have a ribose 3' oxygen replaced by an amine.<sup>47</sup> All of these antisense oligonucleotides work by either backbone RNase H compatibility or steric blockade of one or more macromolecular biomolecules involved in protein expression. The major limitation with these oligos is that they have poor cellular uptake as "naked" molecules, so delivery vectors are required in order for cells to internalize them.<sup>47</sup>

Targeting mRNA with large molecules has a number of shortcomings. In the case of RNAi, there is a risk of silencing off target sequences, as was described with the miRNAs and siRNAs. Antisense technology also presents a similar risk, where a binding event could occur with a RNA sequence of incomplete complementarity. The effectiveness of antisense gene silencing also depends heavily on RNase activity. Finally, these classes of large molecules suffer from poor cellular uptake, and in some cases have problems with cytotoxicity.

# **1.9 RNA Targets**

The selective binding of synthetic molecules to RNA targets is primarily sought after for antibacterial, gene silencing, or antiviral capabilities. In many ways, targeting RNA is a better alternative than binding DNA. The DNA is tightly wrapped around histones and is sequestered to the nucleus in eukaryotes, but RNA is more accessible as it is translated in the cytoplasm. Targeting RNA can also be complementary to protein targeting or employed to manipulate biological functions where proteins are not utilized, as is the case with ribozymes and riboswitches.

# References

1. Hermann, T. Strategies for the design of drugs targeting RNA and RNA-protein complexes. *Angew. Chem. Int. Ed.* **2000**, *39*, 1890-1904.

2. Luedtke, N. W.; Tor, Y. Small molecule DNA and RNA binders: From synthesis to nucleic acid complexes; Wiley-VCH: Weinheim, **2003**; Vol. 1.

3. Thomas, J. R.; Hergenrother, P. J. Targeting RNA with small molecules. *Chem. Rev.* **2008**, *108*, 1171-1224.

4. Ramakrishnan, V. What we have learned from ribosome structures. *Biochem. Soc. Trans.* **2008**, *036*, 567-574.

5. Zaman, G. J. R.; Michiels, P. J. A. *Trends in RNA research*; Nova Science Publishers: New York, **2006**.

6. Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **2000**, *407*, 340-348.

7. Ogle, J. M.; Murphy, F. V.; Tarry, M. J.; Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **2002**, *111*, 721-732.

8. Gallego, J.; Varani, G. Targeting RNA with small-molecule drugs: Therapeutic promise and chemical challenges. *Acc. Chem. Res.* **2001**, *34*, 836-843.

9. Zaman, G. J. R.; Michiels, P. J. A.; van Boeckel, C. A. A. Targeting RNA: New opportunities to address drugless targets. *Drug Discov. Today* **2003**, *8*, 297-306.

10. Uy, R.; Wold, F. Post-translational covalent modification of proteins. *Science* **1977**, *198*, 890-896.

11. Chu, E.; Koeller, D. M.; Casey, J. L.; Drake, J. C.; Chabner, B. A.; Elwood, P. C.; Zinn, S.; Allegra, C. J. Autoregulation of human thymidylate synthase messenger RNA translation by thymidylate synthase. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8977-8981.

12. Chu, E.; Voeller, D.; Koeller, D. M.; Drake, J. C.; Takimoto, C. H.; Maley, G. F.; Maley, F.; Allegra, C. J. Identification of an RNA binding site for human thymidylate synthase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 517-521.

13. Pollard, V. W.; Malim, M. H. The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **1998**, *52*, 491-532.

14. Frankel, A. D.; Young, J. A. T. HIV-1: Fifteen proteins and an RNA. *Annu. Rev. Biochem* **1998**, 67, 1-25.

15. Ippolito, J. A.; Steitz, T. A. The structure of the HIV-1 RRE high affinity Rev binding site at 1.6 Å resolution. *J. Mol. Biol.* **2000**, *295*, 711-717.

16. Battiste, J. L.; Mao, H.; Rao, N. S.; Tan, R.; Muhandiram, D. R.; Kay, L. E.; Frankel, A. D.; Williamson, J. R.  $\alpha$  helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* **1996**, *273*, 1547-1551.

17. Gosser, Y.; Hermann, T.; Majumdar, A.; Hu, W.; Frederick, R.; Jiang, F.; Xu, W.; Patel, D. J. Peptide-triggered conformational switch in HIV-1 RRE RNA complexes. *Nat. Struct. Biol.* **2001**, *8*, 146-150.

18. De Guzman, R. N.; Turner, R. B.; Summers, M. F. Protein-RNA recognition. *Biopolymers* **1998**, *48*, 181-195.

19. Dykxhoorn, D. M.; Lieberman, J. The silent revolution: RNA interference as basic biology, research tool, and therapeutic. *Annu. Rev. Med.* **2005**, *56*, 401-423.

20. Du, T.; Zamore, P. D. Microprimer: The biogenesis and function of microRNA. *Development* **2005**, *132*, 4645-4652.

21. Penedo, J. C.; Wilson, T. J.; Jayasena, S. D.; Khvorova, A.; Lilley, D. M. J. Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *RNA* **2004**, *10*, 880-888.

22. Hoogstraten, C. G.; Sumita, M. Structure-function relationships in RNA and RNP enzymes: Recent advances. *Biopolymers* **2007**, *87*, 317-328.

23. Martick, M.; Scott, W. G. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **2006**, *126*, 309-320.

24. Rich, A. Discovery of the hybrid helix and the first DNA-RNA hybridization. *J. Biol. Chem.* **2006**, *281*, 7693-7696.

25. Wang, A. H. J.; Fujii, S.; van Boom, J. H.; van der Marel, G. A.; van Boeckel, S. A. A.; Rich, A. Molecular structure of r(GCG)d(TATACGC): A DNA-RNA hybrid helix joined to double helical DNA. *Nature* **1982**, *299*, 601-604.

26. van Dongen, M. J. P.; Heus, H. A.; Wymenga, S. S.; van der Marel, G. A.; van Boom, J. H.; Hilbers, C. W. Unambiguous structure characterization of a DNA-RNA triple helix by 15N- and 13C-filtered NOESY spectroscopy. *Biochemistry* **1996**, *35*, 1733-1739.

27. Han, G. W.; Kopka, M. L.; Langs, D.; Sawaya, M. R.; Dickerson, R. E. Crystal structure of an RNA–DNA hybrid reveals intermolecular intercalation: Dimer formation by base-pair swapping. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9214-9219.

28. Kaul, M.; Barbieri, C. M.; Pilch, D. S. Defining the basis for the specificity of aminoglycoside-rRNA recognition: A comparative study of drug binding to the a sites of escherichia coli and human rRNA. *J. Mol. Biol.* **2005**, *346*, 119-134.

29. Means, J. A.; Hines, J. V. Fluorescence resonance energy transfer studies of aminoglycoside binding to a T box antiterminator RNA. *Biorg. Med. Chem. Lett.* **2005**, *15*, 2169-2172.

30. Hendrix, M.; Priestley, E. S.; Joyce, G. F.; Wong, C. H. Direct observation of aminoglycoside-RNA interactions by surface plasmon resonance. *J. Am. Chem. Soc.* **1997**, *119*, 3641-3648.

31. Francois, B.; Russell, R. J. M.; Murray, J. B.; Aboul-ela, F.; Masquida, B.; Vicens, Q.; Westhof, E. Crystal structures of complexes between aminoglycosides and decoding a site oligonucleotides: Role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucl. Acids Res.* **2005**, *33*, 5677-5690.

32. Hermann, T. Rational ligand design for RNA: The role of static structure and conformational flexibility in target recognition. *Biochimie* **2002**, *84*, 869-875.

33. McConnaughie, A. W.; Spychala, J.; Zhao, M.; Boykin, D.; Wilson, W. D. Design and synthesis of RNA-specific groove-binding cations: Implications for antiviral drug design. *J. Med. Chem.* **1994**, *37*, 1063-1069.
34. Zhao, M.; Ratmeyer, L.; Peloquin, R. G.; Yao, S.; Kumar, A.; Spychala, J.; Boykin, D. W.; David Wilson, W. Small changes in cationic substituents of diphenylfuran derivatives have major effects on the binding affinity and the binding mode with RNA helical duplexes. *Biorg. Med. Chem.* **1995**, *3*, 785-794.

35. Breaker, R. R. Complex riboswitches. Science 2008, 319, 1795-1797.

36. Coppins, R. L.; Hall, K. B.; Groisman, E. A. The intricate world of riboswitches. *Curr. Opin. Microbiol.* **2007**, *10*, 176-181.

37. Sudarsan, N.; Hammond, M. C.; Block, K. F.; Welz, R.; Barrick, J. E.; Roth, A.; Breaker, R. R. Tandem riboswitch architectures exhibit complex gene control functions. *Science* **2006**, *314*, 300-304.

38. Winkler, W.; Nahvi, A.; Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **2002**, *419*, 952-956.

39. Edwards, T. E.; Ferre-D'Amare, A. R. Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure* **2006**, *14*, 1459-1468.

40. Thomas, J. R.; Liu, X.; Hergenrother, P. J. Biochemical and thermodynamic characterization of compounds that bind to RNA hairpin loops: Toward an understanding of selectivity. *Biochemistry* **2006**, *45*, 10928-10938.

41. Sudarsan, N.; Hammond, M. C.; Block, K. F.; Welz, R.; Barrick, J. E.; Roth, A.; Breaker, R. R. Tandem riboswitch architectures exhibit complex gene control functions. *Science* **2006**, *314*, 300-304.

42. Welz, R.; Breaker, R. R. Ligand binding and gene control characteristics of tandem riboswitches in bacillus anthracis. *RNA* **2007**, *13*, 573-582.

43. Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C. Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. *Nature* **1998**, *391*, 806-811.

44. Doench, J. G.; Petersen, C. P.; Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev.* **2003**, *17*, 438-442.

45. Saxena, S.; Jonsson, Z. O.; Dutta, A. Small RNAs with imperfect match to endogenous mRNA repress translation: Implications for off-target activity of small inhibitory RNA in mammalian cells. *J. Biol. Chem.* **2003**, *278*, 44312-44319.

46. Esquela-Kerscher, A.; Slack, F. J. Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer* **2006**, *6*, 259-269.

47. Dias, N.; Stein, C. A. Antisense oligonucleotides: Basic concepts and mechanisms. *Mol. Cancer Ther.* **2002**, *1*, 347-355.

48. Vaerman, J. L.; Moureau, P.; Deldime, F.; Lewalle, P.; Lammineur, C.; Morschhauser, F.; Martiat, P. Antisense oligodeoxyribonucleotides suppress hematologic cell growth through stepwise release of deoxyribonucleotides. *Blood* **1997**, *90*, 331-339.

49. Strauss-Soukup, J. K.; Vaghefi, M. M.; Hogrefe, R. I.; Maher, L. J. Effects of neutralization pattern and stereochemistry on DNA bending by methylphosphonate substitutions. *Biochemistry* **1997**, *36*, 8692-8698.

50. Shoji, Y.; Akhtar, S.; Periasamy, A.; Herman, B.; Juliano, R. L. Mechansism of cellular uptake of modified oligodeoxynucleotides containing methylphosphonate linkages. *Nucleic Acids Res.* **1991**, *19*, 5543-5550.

51. Kurreck, J. Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem.* **2003**, *270*, 1628-1644.

52. Eckstein, F. Phosphorothioate oligodeoxynucleotides: What is their origin and what is unique about them? *Antisense Nucleic Acid Drug Dev.* **2000**, *10*, 117-121.

53. Matsukura, M.; Shinozuka, K.; Zon, G.; Mitsuya, H.; Reitz, M.; Cohen, J. S.; Broder, S. Phosphorothioate analogs of oligodeoxynucleotides: Inhibitors of replication and cytopathic effects of human immunodeficiency virus. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7706-7710.

54. Nielsen, P. E. Antisense properties of peptide nucleic acid. *Methods Enzymol.* **2000**, *313*, 156-164.

55. Elayadi, A. N.; Corey, D. R. Application of pna and lna oligomers to chemotherapy. *Curr. Opin. Invest. Drugs* **2001**, *2*, 558-561.

56. Braasch, D. A.; Corey, D. R. Novel antisense and peptide nucleic acid strategies for controlling gene expression. *Biochemistry* **2002**, *41*, 4503-4510.

57. Heasman, J. Morpholino oligos: Making sense of antisense? *Dev. Biol.* **2002**, *243*, 209-214.

58. Gryaznov, S. Oligonucleotide N3' $\rightarrow$ P5' phosphoramidates as antisense agents. *Nucleic Acids Res.* **1996**, *24*, 1508-1514.

59. Zamaratski, E.; Pradeepkumar, P. I.; Chattopadhyaya, J. A critical survey of the structure-function of the antisense oligo/RNA heteroduplex as substrate for RNase H. *J. Biochem. Bioph. Methods* **2001**, *48*, 189-208.

60. Kurreck, J.; Wyszko, E.; Gillen, C. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res.* **2002**, *30*, 1911-1918.

61. Crooke, S. T.; Lemonidis, K. M.; Neilson, L. Kinetic characteristics of escherichia coli RNase H1: Cleavage of various antisense oligonucleotide-RNA duplexes. *Biochem. J.* **1995**, *312*, 599-608.

62. Strauss-Soukup, J. K.; Vaghefi, M. M.; Hogrefe, R. I.; Maher, L. J. Effects of neutralization pattern and stereochemistry on DNA bending by methylphosphonate substitutions. *Biochemistry* **1997**, *36*, 8692-8698.

### Chapter 2 Targeting Structured RNAs Using "Medium-Sized" Molecules

#### 2.1 Introduction

RNA has been catapulted to the forefront of chemical biology research in recent years because of its key roles in transcription, translation, splicing, replication, and catalysis.<sup>1-5</sup> Indeed, RNA's biological repertoire is far more expansive than what was previously thought. Due to the ubiquitous nature of RNA in its control of vital biological functions, researchers have aimed to harness what has been found to be an elusive target at times. Although RNA is not as structurally complex as many proteins, it is capable forming of an array of secondary structures such as stems, loops, bulges and turns, which can be folded upon one another to generate unique tertiary structures.<sup>6</sup> These structures are possible in part because RNA, like protein, is synthesized in vivo as a single stranded macromolecule. In principle, it should be possible to design ligands that can selectively target the pockets and folded surfaces of unique RNA three-dimensional structures, much in the same way rational design is employed for the preparation of small molecules capable of docking with a protein. In reality, it has proven to be a herculean task to generate RNA binding molecules that achieve high affinity, while simultaneously maintaining high selectivity for the target, good cell-permeability, and low cytotoxicity profiles.

A major challenge in designing ligands to bind RNA involves the ability of RNA to fold into complex three-dimensional structures. These conformations are flexible, and

it is possible for the RNA to become locked into misfolded yet stable conformers. Therefore, it is not currently possible to accurately predict the shape of an RNA from the sequence information alone.<sup>6</sup> Due to their conformational flexibility, many important RNA structures that are currently viewed as high-value targets either do not have published crystal structures, or only short fragments of the RNA have been successfully crystallized.<sup>7-11</sup> Thus, researchers must rely on the structures predicted from solution NMR experiments or nuclease and chemical probing methods. Because of this deficit of information, *in silico* design of small molecular ligands for RNA has often been a less bountiful exercise than protein-structure based drug design; however, some valuable functional insights have been gleaned from these endeavors, and significant progress has recently been made in this area.<sup>12-15</sup> Taking into account that a single RNA can have more than one stable conformation, *research* has relied heavily on high-throughput screening of chemical libraries *in vitro* or structure-activity based design from known, strongly binding lead compounds such as aminoglycosides and other antibiotics.<sup>16-24</sup>

RNA binding ligands have traditionally been divided into two classes, namely small or large molecules. The progress of small molecular ligand design and discovery over the past few decades has been extensively reviewed previously in excellent detail.<sup>25-27</sup> This class of RNA binding molecules has been broadly successful in achieving high affinities for RNA, and these ligands have also displayed excellent cell permeability in many cases.<sup>28,29</sup> The major hurdle with these compounds has been their general poor selectivity. The flagship for small ligands that bind RNA have traditionally been antibiotics such as the aminoglycosides, macrolides, oxazolidinones, and tetracyclines, which are known for binding to the ribosomal RNA (rRNA), and these examples have

helped to demonstrate that RNA is a druggable target.<sup>28</sup> However, these successes must be taken within the context of the target structure, where it has been suggested that the conformation of rRNA is highly stabilized in part by the surrounding ribosomal proteins that constrain it.<sup>30</sup> This is a form of RNA that is highly structured and in high quantities within the cell.<sup>31</sup> In contrast, many high-value targets such as the human immunodeficiency virus type-1 (HIV-1) transactivation response element (TAR), the HIV-1 Rev response element (RRE), the hepatitis-C virus internal ribosomal entry site (HCV-IRES), and others are not as conformationally restricted as the rRNA. As a testament to their conformational adaptability, it has been well-demonstrated that many different RNA hairpin shapes including HIV-1 RRE, HIV-1 TAR, and others will bind non-specifically to aminoglycosides.<sup>14,32</sup> Additional issues arise with these small molecules due to their heavy reliance on electrostatic interactions with the RNA target, and this highlights a second hurdle to overcome. Small molecules have been broadly designed in the past with positively charged moieties to increase affinity because the RNA phosphate backbone is polyanionic, and some RNA structures display pockets of negative charge within the major grooves.<sup>12</sup> However, a significant price is paid in the selectivity of these molecules because all RNAs are polyanionic. This problem has led to electrostatic interactions being described as the "double-edged sword" of RNA targeting, where affinity is increased at the expense of selectivity.<sup>14</sup>

Large molecules are the other traditional class of RNA-binding ligands, and these typically encompass the oligonucleic acid binders of RNA such as those used with antisense DNA technology and RNA interference (RNAi) technology, which have both been extensively reviewed.<sup>33-37</sup> Binding of this molecular class relies on the direct

sequence read-out of the target, where Watson-Crick base pairing is the dominant mode of intermolecular interaction, and this allows selectivity for the target to be quite high although incomplete complementarity to the target can result in off-target effects.<sup>38,39</sup> However, cell permeability is the major hurdle in the case of large molecules because they are typically polyanionic and have high molecular weights by definition. Currently, much effort is being made to design transfection agents, typically composed of polycationic polymers, to aid in cellular delivery of these large molecules, and other delivery methods are being explored as well.<sup>34,40,41</sup> Additional issues with this class of compounds have been their poor serum stability, cytotoxicity, and non-specific binding with cellular proteins in some cases.<sup>34,35</sup> Furthermore, many RNAs are inaccessible to these larger inhibitors because of the highly folded nature of the target.

Clearly, molecular design of RNA binding ligands is far from straightforward. Both small and large molecules have beneficial properties, but their successes as RNA binders are severely tempered by their undesirable qualities, which result in a catch-22. Small molecules can simultaneously achieve high affinity and cell permeability, yet they are notoriously non-specific RNA binders. These ligands are also hampered by their low molecular weights because they rely on a relatively few number of intermolecular interactions with the target structure, and using a small molecule to preclude the binding of a protein at the interface of a RNA–protein complex is a truly daunting task. Large molecules, on the other hand, are capable of multiple points of contact with RNA, yet they do not recognize the three-dimensional architecture of the target because they heavily rely on base pairing. Additionally, they are poorly taken-up by cells, and in some cases they have poor cytotoxicity profiles. These drawbacks have plagued on-going research, and it has driven many in the field to explore a third class of molecules, so called "medium-sized" molecules. In this review, we aim to define a separate class of molecular ligands for RNA, which fall between the range of large and small molecules. Examples from this class of compounds have the capability of making multiple points of contact with the RNA target while also retaining cell permeability. Research where this type of ligand has been employed will also be highlighted.

# 2.2 The Third Molecular Class of RNA Binders: Medium-Sized Molecules

Medium-sized molecules can be defined as those that are between ~1200–2500 Da. This mass window provides a better chance for these molecules to remain cell permeable while also presenting an ample scaffold with which to recognize and bind to unique RNA surfaces and folds. In contrast, molecules classified as large in the field of RNA targeting, such as antisense DNA and small interfering RNAs (siRNA), typically have molecular weights above 4000 Da. As previously discussed, they broadly display poor cell permeability and do not recognize the unique tertiary structure of RNA. Small molecules are typically thought of as compounds near or below 500 Da by Lipinski's rule of five.<sup>42</sup> However, within the context of RNA binding ligands we have classified small molecules as those with molecular weights below ~1100 Da, and these include the aminoglycoside antibiotics (Figure 2.1), as well as the oxazolidinone, macrolide, and tetracycline antibiotics.



Figure 2.1. The structure and molecular weight of several aminoglycoside small molecules.

## 2.3 Medium-Sized Neomycin B–Arginine Conjugates and Aminoglycoside Dimers

The aminoglycoside small molecule neomycin B (Neo) has been shown to bind non-selectively to various hairpin RNA structures including HIV-1 TAR, which is a span of 59-nucleotides located at the 5'-end of all nascent HIV-1 mRNA.<sup>9,43,44</sup> This portion of RNA, which folds into a highly conserved hairpin structure that features a hexanucleotide loop and a trinucleotide bulge (Figure 2.2a), binds to the virally encoded Tat protein through an arginine rich motif. This Tat-TAR interaction transactivates transcription of the virus allowing for efficient production of mRNA.<sup>9,45</sup> Thus, TAR RNA has become a popular and potentially important target for the development of new anti-HIV drugs because inhibition of the TAR-Tat interaction substantially lowers the rate of viral transcription.<sup>43,44</sup> Neo has a modest affinity for TAR in the low micromolar range, and binding occurs at the trinucleotide bulge and flanking stem on the 5'-side.<sup>46</sup> Lapidot and coworkers prepared a neomycin B-hexaarginine conjugate (NeoR) with the intention of creating a Tat peptide mimetic (Figure 2.2b). Appending Neo with six Arg residues effectively increased the overall positive charge of the compound while also increasing the molecular weight to fall within the medium-sized molecular class.<sup>23</sup> The affinity for TAR was greatly improved ( $K_d = 5.8$  nM) by fluorescence anisotropy experiments, and cellular uptake was also greatly improved. However, NeoR was also found to be a nonselective binder with another RNA structure, the HIV-1 RRE stem-loop IIB ( $K_d = 18.1$ nM). Computational studies in the same report suggested that NeoR binding was more akin to that of the Tat–TAR complex, where a single Arg residue was docked in the TAR cavity that is formed by folding of the bulge region. This article also showed that NeoR

was able to inhibit binding of HIV-1 virions to MT2 cells, possibly through blockade of the CXCR4 co-receptor. In addition, NeoR was found to inhibit several extracellular functions of the Tat protein including the upregulation of CXCR4 expression, suppression of human peripheral blood mononuclear cell (PBMC) proliferation, and induction of CD8 expression. This neomycin conjugate and similar compounds are now being explored primarily for their role as viral entry inhibitors and less so for their RNA binding capabilities.<sup>47,48</sup>



**Figure 2.2.** The structure of HIV-1 TAR (A) and the neomycin B–hexaarginine conjugate (B).

Another popular target for the aminoglycoside antibiotics has been the HIV-1 RRE RNA. This RNA, which is a portion of the HIV-1 viral genome, is responsible for the egress of full-length and incompletely spliced HIV-1 transcripts from the nucleus to the cytoplasm when bound to the virally encoded Rev protein.<sup>49-51</sup> In this process, Rev binds to the high-affinity site of RRE at stem-loop IIB (RRE-IIB) (Figure 2.3a), which is located within the *env* gene of HIV-1 and present in all incompletely spliced

transcripts.<sup>52,53</sup> Because the accumulation of these transcripts in the cytoplasm is required for the translation of several viral structure proteins and genomic material for new viruses, inhibitors of the Rev–RRE interaction could potentially become valuable anti-HIV therapeutics.<sup>53-55</sup>



Figure 2.3. The structure of HIV-1 RRE-IIB (A) and neomycin dimer (B).

Neo has been demonstrated to non-specifically bind at multiple sites on RRE with an apparent binding affinity of ~0.2  $\mu$ M.<sup>24,56-60</sup> Tok and coworkers sought to take advantage of the multiple Neo binding sites on RRE through the use of bivalent ligands.<sup>60</sup> To that end, they prepared a dimer of Neo (**11**) (Figure 2.3b) that had previously been shown by Wang *et al.* to inhibit the hammerhead ribozyme (HH16).<sup>61</sup> As predicted, a roughly 17-fold increase in binding affinity was observed by fluorescence anisotropy for the dimeric construct **11** ( $K_d = 0.01 \pm 0.001 \mu$ M) compared to Neo ( $K_d = 0.18 \pm 0.01$  $\mu$ M). Unfortunately, however, these measurements were made using concentrations of RNA that were far in excess of the binding constants that were obtained; therefore, their results were called into question.<sup>24</sup> Later, Tor and co-workers prepared and tested **11** as well as several similar aminoglycoside dimers (12–15) for their affinity and selectivity to a larger RRE RNA structure (RRE66) (Figure 2.4).<sup>24</sup> Using a solid phase assay, they measured a 100-fold increase in binding to RRE66 using 11 compared to the undimerized Neo. Although a strong binding affinity of 11 ( $K_i = 2.6$  nM) was observed for RRE66 using fluorescence anisotropy, the compound also displayed strong affinity to a poly r[A]–r[U] RNA duplex structure ( $K_i = 12$  nM) as well as calf thymus (CT) DNA ( $K_i =$ 390 nM). Further interrogation of the compound's selectivity using a solid-phase assay displayed a dramatic loss of affinity (96-fold) when conducted in the presence of 100fold excess tRNA bases, although only a 2-fold loss of affinity was observed in the presence of 100-fold excess of CT DNA bases. These results indicated that the dimer 11 preferentially binds RNA over DNA, but the selectivity for a particular RNA tertiary structure is poor. This trend was broadly maintained across all of the dimers they tested.



Figure 2.4. The structure of RRE66 (A) and affinities toward aminoglycoside dimers (B).

#### 2.4 Medium-Sized Peptides and Peptidomimetics

Many of the high-value RNA targets such as HIV-1 TAR, HIV-1 RRE, and HCV-IRES function through their interaction with key proteins or ribonucleoprotein binding partners.<sup>9,11,43,44,62-71</sup> Furthermore, it is has been well demonstrated that only short fragments of these proteins are required for binding to the target RNA structure.<sup>72-78</sup> Since nature has provided a blue print for RNA ligand design, it is unsurprising that peptidic compounds and peptidomemetics are the most well explored medium-sized molecular species designed for RNA binding. Peptidic ligands are also particularly attractive due to their synthetic accessibility, functional modularity, and amenability to the rapid preparation of combinatorial libraries. This review will cover the general approaches to targeting RNA with medium-sized peptides and peptidomemetics as well as some of the classic examples of each type where they have been utilized.

Yu and coworkers, over the course of four studies, have extensively explored short,  $\alpha$ -helical peptides with amphiphilic properties for their ability to bind hairpin RNA targets with good selectivity.<sup>79-81</sup> In this series of papers, the helical peptides were all based on the sequence, LKKLLKKLLKKLLKKG (**16**). This peptide was originally targeted toward calmodulin and was predicted to have a structure that displayed the Lys and Leu residues on the opposite sides of an  $\alpha$ -helix.<sup>82,83</sup> In their first report, they explored the notion that methylated Lys and Arg residues play an important role in the selective and specific RNA binding of endogenous proteins.<sup>79,84,85</sup> To that end,  $\alpha$ -helical peptides featuring  $N^{e}$ , $N^{e}$ -dimethyl Lys residues were synthesized and tested for their ability to bind to HIV-1 RRE RNA by fluorescence anisotropy (Table 2.1).<sup>79</sup> This study revealed that methylation of lysine at positions 3 and 9 in peptide **22** resulted in a ~2.4-

fold increase in affinity to RRE compared to **16**, and that the specificity to RRE over two other RNA structures was also roughly doubled. This was an improvement over the selectivity displayed by a Rev related peptide that was also tested in the same study. It was also shown that the methylated peptides displayed lower  $\alpha$ -helical content measured by circular dichroism (CD) spectroscopy compared to **16**. The authors did note that in a separate experiment, the  $\alpha$ -helical content of **22** was increased from 7.5% to 11% and 18% when it was titrated with 0.1 and 0.3 equivalents of RRE, respectively. This result was indicative of a conformational change of the peptide resulting in an induced-fit binding to RRE. RNA footprinting studies also suggested that both **22** and the Rev peptide share a similar binding site on RRE, which is located at the internal loop and nearby stem regions.

Peptide	Sequence (position(s) of K*) <sup>a</sup>	α-Helicity (%) <sup>c</sup>	K <sub>d</sub> vs RRE [nM]	$K_{\rm d}$ vs TAR $[{\rm nM}]^{\rm d}$	$K_{\rm d}$ vs tRNA <sup>mix</sup> [nM] <sup>d</sup>
16	LKKLLKLLKKLLKLKG	26/57	22	62 (2.8)	55 (2.5)
17	LKKLLKLLKKLLK*LKG	5/45	79	-	-
	(13)				
18	LKKLLKLLK*KLLKLKG (9)	9/48	74	-	-
19	LKK*LLKLLKKLLKLKG (3)	8/50	75	-	-
20	LKKLLKLLK*KLLK*LKG	7/49	30	-	-
	(9, 13)				
21	LKK*LLKLLKKLLK*LKG	6/43	69	-	-
	(3, 13)				
22	LKK*LLKLLK*KLLKLKG	8/52	9.1	53 (5.8)	42 (4.6)
	(3, 9)				
23	LKK*LLKLLK*KLLK*LKG	6/47	87	—	-
	(3, 9, 13)				
Rev <sup>b</sup>	TRQARRNRRRRWRERQRA	33/73	8.5	21 (2.5)	20 (2.4)
	AAAR				

**Table 2.1.** Sequences of peptides featuring *N*,*N*-dimethyl Lys and their binding affinity to several RNAs.

<sup>&</sup>lt;sup>a</sup> K\* =  $N^{e}$ ,  $N^{e}$ -dimethyl Lys. <sup>b</sup> The N-terminus of the peptide is succinvlated. <sup>c</sup> In 10 mM H<sub>3</sub>PO<sub>4</sub> / 50% 2,2,2-trifluoroethanol (TFE) in 10 mM H<sub>3</sub>PO<sub>4</sub>, pH 7.4. <sup>d</sup> Discrimination ratios ( $K_{d}$  against other RNA /  $K_{d}$  against RRE) are given in parenthesis.

Later, Yu and coworkers reasoned that using intercalating moieties rather than methyl groups in their  $\alpha$ -helical peptides would aid in binding RNA through potential  $\pi$ - $\pi$  interactions.<sup>80</sup> In this study, variants of peptide **16** were initially prepared featuring Lys residues acridinylated at the  $N^{\varepsilon}$  atoms. These amphiphilic peptides were evaluated for their selectivity toward the hairpin structures of RRE-IIB and TAR compared to tRNA<sup>mix</sup> using fluorescence anisotropy. Of the monoacridinylated peptides sequences, 24, 27, and 28 had high binding affinity to RRE and TAR hairpin structures (Table 2.2), but their selectivity between these two RNA structures was quite poor. It was also noted that the position of the acridine moiety caused changes in the observed binding affinities; thus, these groups may be positioned differently once in complex with the RNA. They tested this hypothesis by generating bis-acridinylated (30, 31, and 32) and tris-acridinylated (33) variants of **16** to determine if these changes resulted in a "multivalency effect". Every peptide featuring more than one acridine moiety achieved increased affinity for all RNA types tested, where the bis-acridinylated compounds had better discrimination ratios for the hairpin RNAs in general. Although tris-acridinylated peptide 30 displayed the highest affinity of all compounds tested, a heavy price was paid in destruction of any selectivity across all three RNA types. Further analysis of these peptides using several additional RNA sequences showed that these amphiphilic peptides highly preferred stembulges and internal loop structures to apical loops and purely double-helical stems. This study demonstrated that although intercalating moieties can highly improve binding affinity to RNA structures, care must be taken in their utilization to avoid large decreases in selectivity.

Peptide	Sequence (position(s) of K*) <sup>a</sup>	α-Helicity	$K_{\rm d}$ vs RRE	$K_{\rm d}$ vs TAR	$K_{\rm d}$ vs tRNA <sup>mix</sup>
		(%) <sup>b</sup>	[nM] <sup>c</sup>	[nM] <sup>c</sup>	[nM]
16	LKKLLKLLKKLLKLKG	26/57	22 (2.5)	62 (0.89)	55
24	LK*KLLKLLKKLLKLKG (2)	14/52	3.2 (4.4)	2.5 (5.6)	14
25	LKK*LLKLLKKLLKLKG (3)	18/57	11	17	16
26	LKKLLK*LLKKLLKLKG (6)	18/55	14	8.3	19
27	LKKLLKLLK*KLLKLKG (9)	36/62	4.3 (3.0)	1.3 (10)	13
28	LKKLLKLLKK*LLKLKG	25/63	6.8 (2.1)	1.5 (9.3)	14
	(10)				
29	LKKLLKLLKKLLK*LKG	15/65	13	16	9.3
	(13)				
30	LK*KLLKLLK*KLLKLKG	35/57	0.61 (8.0)	0.55 (8.9)	4.9
	(2,9)				
31	LK*KLLKLLKK*LLKLKG	15/48	0.72 (8.5)	0.64 (9.5)	6.1
	(2, 10)				
32	LKKLLKLLK*K*LLKLKG	12/54	0.92 (7.5)	0.37 (18)	6.9
	(9, 10)				
33	LK*KLLKLLK*K*LLKLKG	15/49	0.25 (0.92)	0.20 (1.1)	0.23
	(2, 9, 10)				

**Table 2.2.** Sequences of peptides featuring *N*-acridinyl Lys and their binding affinity to several RNAs.

<sup>a</sup> K\* =  $N^{\epsilon}$ -acridinyl-Lys. <sup>b</sup> In 10 mM H<sub>3</sub>PO<sub>4</sub> / 50% TFE in 10 mM H<sub>3</sub>PO<sub>4</sub>, pH 7.4. <sup>c</sup> Discrimination ratios ( $K_{d}$  against tRNA<sup>mix</sup>/ $K_{d}$  against RRE or TAR) are given in parenthesis.

In their next iteration of amphiphilic  $\alpha$ -helical peptides, Yu and co-workers owed their previous difficulties in selectively targeting the TAR and RRE hairpins to the conformational flexibility and inherent promiscuity of these RNA structures.<sup>12,15,81</sup> To get around this problem, they used their compounds to identify hairpin RNA structures that are capable of specific binding and thus, have a greater chance of being druggable targets. First, the researchers performed an alanine scan of peptide **16**, where each Lys residue was systematically replaced with Ala (7 total). These peptides were tested for their affinity to RRE-IIB, TAR, 16S rRNA A-site from *E. coli* (16S A-site), and HCV IRES domain IV (IRES) (Figure 2.5). Based on the results from this experiment, IRES was found to be the most specific binder of the four hairpin structures tested, as it had the highest standard deviation in binding affinity across all seven Ala modified peptides and **16**. In the second generation of peptides used in this report, further modifications were

made at the three Lys residues of peptide 16 that most profoundly decreased binding to IRES upon replacing them with Ala. These three locations were each replaced by the progressively shorter amine containing side chains of ornithine (Orn), 2,4-diaminobutyric acid (Dab), and 2,3-diaminopropionic acid (Dap) in order to preserve the critical amine functionalities, while also making the peptides smaller to allow binding more deeply in the RNA groove. In addition to these modifications, the final Lys residue of these peptides was also modified to Ala (Lys15Ala) in all subsequent experiments because this peptide, 16h (16h = 16 (Lys15Ala)), had the highest affinity for all RNAs tested in the Ala scan. Gratifyingly, each modification to Orn, Dab, or Dap increased binding affinity relative to 16h in all cases tested, where the most improved binder contained two mutations in the sequence (Lys9Orn; Lys13Dap). This sequence had a  $K_d$  of 680 ± 60 pM for IRES and displayed the best discrimination ratio (DR = 25), which was defined in this case as the average  $K_d$  of non-specific RNAs (RRE, TAR, and 16S A-site) divided by the  $K_{\rm d}$  of IRES. Finally, they were able to further improve the affinity and discrimination ratio to IRES in an additional experiment by performing a tryptophan scan of all Leu This experiment provided the optimal sequence, 16h (Leu7Trp; Lys9Orn; residues. Lys13Dap), which yielded a  $K_d$  for IRES of 550  $\pm$  50 pM and provided a discrimination ratio of 45.



Figure 2.5. The structure of the 16S A-site of *E. coli* and HCV IRES domain IV (IRES) with their binding affinity to peptide 16.

In their most recent report, Yu and coworkers investigated covalently cross-linked helical peptides as RNA binders.<sup>86</sup> The utility of cross-linked  $\alpha$ -helical peptides had been previously established when Guy and coworkers targeted the HIV-1 RRE RNA using macrolactam constrained peptides with induced  $\alpha$ -helicity.<sup>87</sup> In addition, dimeric  $\alpha$ -helical peptides had been shown to form and retain coiled-coil structures even in aqueous media.<sup>88,89</sup> Verdine and coworkers had also demonstrated that the stability of cross-linked  $\alpha$ -helical peptides containing Lys was increased 41-fold in the presence of trypsin compared to uncross-linked controls.<sup>90</sup> Thus, Yu and coworkers reasoned that cross-linked helical peptides based on the sequence of **16h** would be amenable for targeting hairpin RNA structures, and the metabolic stability of their compounds would also be improved.<sup>86</sup> Through cross-linking, they hypothesized that the  $\alpha$ -helical character of their amphiphilic peptides would be increased by inducing conformational rigidity, which could help reduce the entropic cost of binding to the RNA target. Therefore, residues at positions 5 and 12 of **16h** were changed to Cys and various maleimido linkers

were installed through those side chains (Table 2.3). Although the helicity of the peptides in buffered aqueous solution was improved in each case, all of the cross-linked peptides failed to achieve binding affinities to hairpin RNAs higher than that of the parent sequence 34. In an attempt to overcome these poor affinites, they prepared five different  $\alpha$ -helical dimers through disulfide linkages provided by Cys residues in the peptides. These Cys residues were systematically positioned at unique i and i + 7 locations in each dimers (Table 2.4). We must note that these dimerized peptides fall outside the mass range of medium-sized molecules defined above, with weights well in excess of 2500 g mol<sup>-1</sup>. However, these constructs did not recognize the RNA through Watson-Crick interaction or direct base readout, in contrast to other large-molecules used in antisense or RNAi technologies. Thus, they are significant to the scope of this review. Dimerization of the peptides greatly improved the  $\alpha$ -helical character in aqueous buffered solution as well as under membrane-like conditions. Moreover, the position of the cross-linker in the sequence allowed the hydrophilic side of each chain to be positioned differently in each construct, where each dimer had C2 symmetry when viewed from the top or bottom along the helical axes. This property was responsible for the difference in affinities across the 5 dimers examined. Although **dimers 1–5** achieved sub-nanomolar affinities for the RNA targets, selectivity between RRE, TAR, or IRES was poor. This suggests these molecules may in fact have too many non-specific intermolecular interactions with the targets, presumably from the multiple positively charged amine functionalities presented in each peptide chain.

**Table 2.3.** The sequences, percent helicity, and  $K_d$  values of covalently cross-linked peptides against hairpin RNA structures.

Peptide	Sequence	α-Helicity (%) <sup>a</sup>	K <sub>d</sub> vs RRE [nM]	<i>K</i> <sub>d</sub> vs TAR [nM]	<i>K</i> <sub>d</sub> vs IRES [nM]
16h	LKKLLKLLKKLLKLAG	9/43	25	63	49
34	LKKLCKLLKKLCKLAG	18/73	13	9.6	34
<b>34</b> (BMB) <sup>b</sup>	LKKLCKLLKKLCKLAG	50/61	90	67	47
<b>34</b> (BMOE) <sup>b</sup>	LKKLCKLLKKLCKLAG	13/23	46	43	31
<b>34</b> (BMDB) <sup>b</sup>	LKKLCKLLKKLCKLAG L Linker	15/53	56	8.4	25
<b>34</b> (BMH) <sup>b</sup>	LKKLCKLLKKLCKLAG	24/44	85	52	27

<sup>a</sup> In 10 mM H<sub>3</sub>PO<sub>4</sub>/ 50% TFE in 10 mM H<sub>3</sub>PO<sub>4</sub>, pH 7.4. <sup>b</sup> 1,4-(Bismaleimido)butane (BMB), bis(maleimido)ethane (BMOE), 1,4-bismaleimidyl-2,3-dihydroxybutane (BMDB), or bis(maleimido)hexane (BMH).

**Table 2.4.** The sequences, percent helicity, and  $K_d$  values of covalently cross-linked peptide dimers against hairpin RNA structures.

Peptide	Sequence	α-Helicity (%) <sup>a</sup>	<i>K</i> <sub>d</sub> vs RRE [nM]	K <sub>d</sub> vs TAR [nM]	K <sub>d</sub> vs IRES [nM]
Dimer 1	CKKLLKLCKKLLKLAG	77/80	0.21	0.17	0.14
Dimer 2	LKKCLKLLKKCLKLAG IIII LKKCLKLLKKCLKLAG	79/88	0.18	0.16	0.068
Dimer 3	LKKLCKLLKKLCKLAG LKKLCKLLKKLCKLAG	74/79	0.18	0.061	0.060
Dimer 4	LKKLLKCLKKLLKCAG	80/83	0.040	0.021	0.024
Dimer 5		73/82	0.24	0.12	0.093

<sup>a</sup> In 10 mM H<sub>3</sub>PO<sub>4</sub> / 50% TFE in 10 mM H<sub>3</sub>PO<sub>4</sub>, pH 7.4.

#### 2.5 Targeting RNA Using Modified Peptide Backbones

A major limitation of using natural peptide backbones and amino acids is that they are highly susceptible to enzymatic hydrolysis *in vivo*. An approach to solve this issue is by modifying peptide backbones such that proteolytic enzymes no longer recognize them. Simon and coworkers were the first to detail the general synthesis of Nsubstituted glycine oligomers, named peptoids, for drug discovery (Figure 2.6a).<sup>91</sup> They demonstrated that these compounds have increased stability to proteinases, and their computational studies of these compounds suggested that peptoids would also have greater conformational diversity compared to similar peptides. In addition, they showed that a peptoid sequence corresponding to the RNA binding domain of Tat (Ac-YKKKRKKKA-NH<sub>2</sub>) had a binding affinity for TAR that was similar to the natural Tat peptide sequence (less than or equal to 2  $\mu$ M). They were also able to demonstrate that the peptoid was selective for the wild-type TAR RNA over a TAR mutant containing a point mutation in the tri-nucleotide bulge region (U23>C).



**Figure 2.6.** Structural comparison of peptide and peptoid backbones (A), the structure of peptoid/<sub>D</sub>-peptide hybrid **CGP6422** screened from a combinatorial library (B), and the structures of peptoid amides and peptoid esters (C).

The utility of peptoids was further revealed the when Hamy and coworkers developed the first demonstrated inhibitor of the Tat–TAR interaction, which was discovered from screening a combinatorial library comprised of peptoid/D-peptide hybrids.<sup>92</sup> The library was designed with limited complexity, where the first five residues

were randomized with 20 unique peptoid and D-amino acid building blocks. The remaining four residues on the C-terminus in each sequence were held constant and were composed of all D-amino acids (D-Lys-D-Lys-D-Arg-D-Pro-amide). The optimal compound from the library was **CGP64222** (Figure 2.6b). This peptoid/D-peptide hybrid was able to inhibit the Tat-TAR interaction by 50% at 12 nM in vitro, and NMR studies demonstrated it interacted with the tri-nucleotide bulge as well as the two flanking base pairs on either side of the bulge. CGP64222 was also able to inhibit Tat mediated transactivation in cell based assays with an IC<sub>50</sub> of  $3-5 \mu$ M. Furthermore, reverse transcriptase production and HIV-1 replication were both undetectable when infected primary human lymphocytes were incubated with the compound at 30 µM; meanwhile, cell viability and proliferation were both unaffected. Rana and coworkers later increased the complexity of peptoids when they appended the side chains of peptoids with methyl esters or amides to create a new class of RNA-binding oligomers named peptoid esters and peptoid amides, respectively (Figure 2.6c).<sup>93</sup> In this study, the peptoid ester and amide were each modeled from the arginine rich region of the Tat peptide (amino acids 47–57), and these compounds were evaluated for their ability to bind TAR RNA. The peptoid amide achieved a binding affinity ( $K_d$ ) of ~155 nM while the peptoid ester was an even stronger binder with a  $K_d$  of ~ 68 nM.

Rana and coworkers first used oligocarbamates to determine their feasibility as RNA ligands.<sup>94</sup> It was expected that the hydrogen-bonding properties, lipophilicity, and degree of conformational restriction of the oligocarbamate backbone would differ slightly from the natural peptide backbone. Here, they used a Tat-derived oligocarbamate (**35**) corresponding to the basic arginine-rich region of Tat (Figure 2.7a). This compound was

bound to TAR RNA with a  $K_d$  of 1.13 µM by electrophoretic mobility shift assay (EMSA), and no affinity was observed for a TAR mutant lacking the tri-nucleotide bulge region up to 4 µM in competitive binding experiments. Unfortunately, the affinity of **35** for wild-type TAR was poorer than the natural Tat-derived peptide, **Tat(47–57)**, which displayed a  $K_d$  of 0.78 µM. This provided a  $K_{rel}$  of 0.69 ( $K_{rel} = K_d$  of **35** /  $K_d$  of **Tat(47–57)**), which was used to normalize the data because binding affinities can vary depending on experimental conditions.<sup>74,94</sup> The binding site of **35** on TAR was then probed using a photo-cross-linking experiment, which demonstrated that the site of binding was similar to Tat and located in the widened major grove at the bulge region of TAR. Further investigation of the RNA–oligocarbamate complex through treatment with proteinase K demonstrated that the unnatural backbone of **35** was highly resistant to enzymatic degradation.

Tat(47-57)



**Oligocarbamate Backbone Structure** 





**Oligourea Backbone Structure** 

С

- $\begin{array}{l} \textbf{37} = \beta^{3}\text{HTyr} \beta^{3}\text{HGly} \beta^{3}\text{HArg} \beta^{3}\text{HLys} \beta^{3}\text{HArg} \beta^{3}\text$
- **38** =  $\beta^{3}$ HTyr- $\beta^{3}$ HGly- $\beta^{3}$ HLys- $\beta^{3}$ HLy

39 = Tyr-Gly-Lys-Lys-Lys-Lys-Lys-Gln-Lys-Lys-Lys



**Figure 2.7.** The sequence of **Tat(47–57)** from which oligocarbamate **35** (A) oligourea **36** (B) and  $\beta$ -peptides (C) were derived.

Later, Rana and coworkers prepared an oligourea compound (36) that was also based on the sequence of Tat(47–57) (Figure 2.7b).<sup>95</sup> Oligourea backbones, like oligocarbamates and peptides, have hydrogen-bonding groups, chiral centers, and are

conformationally restricted. Unlike the previous attempt using oligocarbamates, compound **36** displayed a higher affinity to TAR ( $K_d = 0.11 \pm 0.07 \mu M$ ) than the natural **Tat**(47-57) peptide ( $K_{\rm d} = 0.78 \pm 0.05 \ \mu$ M) by EMSA. In this case, a  $K_{\rm rel}$  of 7.09 was obtained. Oligourea 36 also displayed selectivity for the wild-type TAR structure over two mutant sequences in competition experiments at concentrations up to 200 nM of competitor RNA. These TAR mutants were prepared such that one of the RNAs lacked the tri-nucleotide bulge, and the second TAR mutant featured a mononucleotide bulge. The HIV-1 Tat peptide is known to interact with the two base pairs above the trinucleotide bulge of TAR. To determine if **36** interacted with TAR in a similar way, an additional experiment was conducted where the G26-C39 base pair immediately above the bulge was inverted to C26-G39. This mutant RNA was not able to outcompete binding of **36** to wild-type TAR RNA at concentrations up to 200 nM of the mutant TAR competitor, suggesting that the native RNA sequence is required for binding at that location. The oligourea **36**-TAR complex was also found to be completely inert when treated with proteinase K demonstrating its potential metabolic stability.

More recently, Rana and coworkers investigated  $\beta$ -peptides (Figure 2.7c) as binders to TAR RNA due to their ability to adopt protein-like structures that can have biological activity as well as their known stability to proteases.<sup>96-99</sup>  $\beta$ -Peptide **37** was also based on **Tat(47–57**). They noted that this region of Tat had been shown previously to adopt an extended conformation and that peptide **37** was also known to be unstructured in solution.<sup>96,100</sup> Thus, it was reasoned that the conformational flexibility of **37** could allow it to bind TAR. Unfortunately, the affinity of  $\beta$ -peptide **37** for TAR was 15-fold less than the wild-type peptide control ( $K_{rel} = 0.071$ ) by fluorescence anisotropy measurements (Table 2.5). Surprisingly, however, another  $\beta$ -peptide (**38**) where all Arg residues were converted to Lys had a higher affinity to TAR than its corresponding peptide control **39** as well as  $\beta$ -peptide **37**. Although the TAR binding affinity of compound **37** was poorer than anticipated, its selectivity for the TAR bulge was far greater than any of the other three peptides tested in this study. The authors also noted that the altered spacing of the  $\beta$ -peptide side chains could not be the only explanation for the altered affinity to TAR; their previous experiments with the oligocarbamate backbone provided similarly poor affinities, but using the oligourea backbone had provided improved binding affinity compared to **Tat(47–57)**, as discussed above.

Table 2.5. Dissociation constant for Tat(47–57) and compounds 37–39 from wild-type or bulgeless TAR.

Peptide	Wild-Type TAR (K <sub>d</sub> )	Bulgeless TAR (K <sub>d</sub> )	K <sub>rel</sub>
37	$29 \pm 4 \text{ nM}$	$281 \pm 68 \text{ nM}$	10
Tat(47–57)	$2.1 \pm 0.8 \text{ nM}$	$1.1 \pm 0.2 \text{ nM}$	0.5
38	$16 \pm 2 \text{ nM}$	$33 \pm 2$ nM	2
39	$32 \pm 4 \text{ nM}$	$72 \pm 8 \text{ nM}$	2

The newest example of backbone modified peptides was provided recently by Cai and coworkers.<sup>101</sup> These compounds, named  $\gamma$ -AApeptides, were based on the  $\gamma$ -peptide nucleic acid (PNA) backbones typically used in antisense technology,<sup>34</sup> and like the other backbone modified peptides discussed earlier, were also found to be resistant to enzymatic degradation.<sup>101</sup> In a recent report, Varani, Cai, and coworkers demonstrated the utility of these compounds for targeting TAR RNA.<sup>102</sup> They tested two Tat peptide (**Tat(48-57**)) mimetics (Figure 2.8) for their ability to bind HIV-1 TAR RNA and a closely related bovine immunodeficiency virus (BIV) TAR RNA structure (*vide infra*). **Tat(48–57**) and the full length mimetic ( $\gamma$ -AA1) had nearly identical affinity to both

HIV-1 and BIV TAR structures, while the truncated sequence ( $\gamma$ -AA2) displayed dramatically reduced affinity. Furthermore, the  $\gamma$ -AA1-TAR complex was stable in the presence of 25,000-fold excess of competing tRNA. It appears that these  $\gamma$ -AApeptides are especially adept for functioning as  $\alpha$ -peptide mimetics partly due to the identical amino acid side chain length in both structures. In fact, Cai and coworkers have already demonstrated the ability of these compounds to disrupt protein-protein interactions; thus,  $\gamma$ -AApeptides may prove to be an important new class of compounds for the disruption of RNA–protein interactions in the future.<sup>103</sup>

Tat(48–57)



γ-ΑΑ1



**K<sub>d</sub> (HIV), nM K<sub>d</sub> (BIV), nM** 166 300

γ**-ΑΑ2** 



**Figure 2.8.** The structure and binding affinity to TAR RNA of **Tat(48–57)**,  $\gamma$ -**AA1**, and  $\gamma$ -**AA2**.

#### 2.6 Medium Sized Cyclic Peptides

Over the last decade, cyclic peptides of medium size have been studied extensively for their capacity to bind various folded RNA structures.<sup>104-112</sup> Their rise in popularity has largely been due to the concomitant decrease in entropy achieved through conformational restriction of the compounds, also known as preorganization.<sup>110</sup> An additional benefit to cyclized peptides is their resistance to enzymatic degradation.<sup>111</sup>

In their 2001 report, Tok and coworkers were able to enhance binding of a 17amino acid BIV Tat peptide to the BIV TAR RNA structure (Figure 2.9a) by using a cyclic peptide variant of BIV Tat.<sup>112</sup> At the time of the study, there was no structural information available for the HIV-1 Tat/TAR complex; however, the structurally similar BIV Tat/TAR complex had been solved in solution by NMR.<sup>113</sup> Thus, it was reasoned that information gained through targeting the BIV complex may be able to be exported to the HIV-1 complex after initial development. In contrast to the TAR RNA construct associated with HIV-1, the BIV TAR structure features two unpaired nucleotides that constitute two mononucleotide bulges on a 28-nucleotide hairpin. In this study, Cys residues were appended to the N- and C-terminus of the wild-type linear BIV Tat peptide, and cyclization was achieved through disulfide bond formation between these terminal side chains (Figure 2.9b). This simple modification resulted in a 5-fold decrease in the measured dissociation constant ( $K_d$ ) by fluorescence anisotropy.

### Α

G	G		
U	G	Α	U
С	A25	С	U
С	G	U	Α
G	С	С	G
A <sub>27</sub>	U	G	С
G	С	Α	U
U		U	
C <sub>24</sub> U		G U	С
A <sub>22</sub>	U	G	С
G	С	С	G
Α	U	U	Α
С	G <sub>43</sub>	С	G
G	C	G	С
G <sub>17</sub>	С	$G_4$	C <sub>31</sub>
5'-	-3'	5'-	-3'
HIV-1	TAR	BIV	TAR

В

	<i>K</i> <sub>d</sub> , nM
Wild-Type Linear BIV Tat Peptide:	
SGPRPRGTRGKGRRIRR	1.321 ± 0.078
Control Linear BIV Tat Pentide 1	
	1 277 + 0 097
Cournendingkanninn	1.277 ± 0.007
Control Linear BIV Tat Peptide 2:	
SGPRPRGTRGKGRRIRRC	1.332 ± 0.121
Cyclic Peptide Variant of BIV Tat:	
SGPRPRGTRGKGRRIRR	0.273 ± 0.011
CC	

**Figure 2.9.** The structure and sequence of HIV-1 TAR RNA and BIV TAR RNA (A) and structure of BIV Tat peptide variant with binding affinities for BIV TAR.

In 2003, Puglisi and coworkers detailed their work providing the first example where backbone cyclic peptides had been used for targeting an RNA structure.<sup>111</sup> Here again, a cyclic BIV Tat peptide was used, and strong binding of the cyclic peptide in a 1:1 complex with BIV TAR was demonstrated through NMR spectroscopy. In this study, a 14-amino acid cyclic peptide (cyclo-<sup>D</sup>YGRGTRGKGRRIVN) derived from the arginine-rich motif of BIV Tat was designed. The wild-type sequence (underlined) was appended at the N-terminus with a D-Tyr and a Gly to promote the formation of a  $\beta$ hairpin-like turn. Although the amino acid of choice for inducing the conformation of this desired structure is D-Pro, D-Tyr provided a better spectroscopic handle for their experiments. In addition, Val and Asn were appended to the C-terminus to promote stability of the  $\beta$ -sheet. Unfortunately, neither the cyclized nor the linear wild-type peptide adopted a preformed structure in solution. Instead, each was reminiscent of a random-coil, as determined by circular dichroism spectroscopy as well as HN-HN NOESY and ROESY experiments. However, in both cases the chemical shifts of unbound RNA and peptide were dramatically different from the corresponding peaks in the bound complexes. In addition, the NOE spectra of the cyclic peptide–RNA complex were strikingly similar to those of the Tat-derived linear control peptide–RNA complex in all NMR experiments conducted. This suggested that the cyclic peptide was able to retain recognition of BIV TAR with a similar binding affinity as than the natural peptide. However, no evidence was provided to demonstrate the selectivity of the cyclic compound.

At the same time that Puglisi and coworkers were disseminating their work, Varani and coworkers were preparing the report of their first experiments using a small library of medium-sized cyclic peptides.<sup>110</sup> In their study, the design of nine peptides was based on the structure of BIV Tat mounted on a D-Pro-L-Pro template to induce a β-turn structure (Table 2.6). Additional hydrophobic/aromatic or cationic residues were added to peptides **BIV1–8** such that they flanked each side of the D-Pro-L-Pro template in order to stabilize the  $\beta$ -hairpin turn, and binding affinities to BIV TAR were determined by EMSA. Impressively, compound **BIV2** bound to BIV TAR selectively ( $K_d = 0.15 \mu M$ ) in the presence of 10,000-fold excess of tRNA, which approached the affinity of the natural peptide sequence **BIV Tat(65–81**). Additionally, **BIV5** was able to bind BIV TAR as well as HIV-1 TAR with similar affinities ( $K_d = 1-2 \mu M$ ). In contrast to the constructs prepared by Puglisi and coworkers earlier, BIV2 was found to have a single stable conformation by NMR studies when free in solution. This compound preferred a regular 2:2 β-hairpin with a predominant type-II' β-turn (for a review of hairpin nomenclature see Sibanda *et al.*<sup>114</sup>). **BIV5** was found to have three structural populations in an 8:1:1 ratio. Although exchange broadening of signals in <sup>1</sup>H NMR studies of the bound complex suggested that BIV5 weakly bound to the BIV TAR RNA, experiments with **BIV2** in complex with BIV TAR demonstrated strong interactions through slow exchange between bound and unbound populations.

Peptide	Sequence	$K_{\rm d}, \mu { m M}$
BIV Tat(65-81)	Ac-SGPRPRGTRGKGRRIRR-NH <sub>2</sub>	0.050
BIV0	<i>cyclo</i> -RG <b>TRGK</b> G <b>RRI</b> R- <sup>D</sup> P-P	40 <sup>b</sup>
BIV1	<i>cyclo</i> -IRG <b>TRGKRRI</b> RV- <sup>D</sup> P-P	30
BIV2	<i>cyclo</i> -RVR <b>TRGKRRI</b> RV- <sup>D</sup> P-P	0.15
BIV3	<i>cyclo</i> -IYR <b>TRGKRRI</b> RT- <sup>D</sup> P-P	nd <sup>c</sup>
BIV4	<i>cyclo</i> -YRG <b>TRGKRRI</b> YV- <sup>D</sup> P-P	>50
BIV5	<i>cyclo</i> -RRG <b>TRGKRRI</b> GR- <sup>D</sup> P-P	1–2
BIV6	<i>cyclo</i> -VRG <b>TRGKRRI</b> KY- <sup>D</sup> P-P	>50
BIV7	<i>cyclo</i> -VRR <b>TRGKRRI</b> KY- <sup>D</sup> P-P	nd <sup>c</sup>
BIV8	<i>cyclo</i> -KRG <b>TRGKRRI</b> GY- <sup>D</sup> P-P	>50

**Table 2.6.** The sequence of **BIV Tat(65–81)** and the backbone cyclized mimetics with their binding affinities to BIV TAR RNA.<sup>a</sup>

<sup>a</sup> Amino acids in bold were held constant in all peptides. <sup>b</sup> Value is the  $IC_{50}$  of BIV0 determined by inhibition assay. <sup>c</sup> nd = not determined.

In a follow-up study, Varani and coworkers reported the structure of **BIV2** in complex with BIV TAR RNA, which was solved using heteronuclear NMR techniques.<sup>109</sup> Interestingly, the orientation of **BIV2** bound to TAR was flipped upside down compared to the natural peptide, but the induced conformational change in the RNA was similar upon binding of either peptide. The cyclic **BIV2** peptide was found to bind the major grove of TAR with the D-Pro-L-Pro end pointed at the bottom of the binding region, toward the 5' and 3' end of the RNA helix. Key interactions were made with the RNA through the side chains of Arg1, Arg3, Arg5, Ile10, and Val12. Meanwhile, residues 2, 4, 9, and 11 were solvent exposed. Thus, it was reasoned that these residues and those in proximity could be tuned to increase binding affinity in the subsequent generation of a cyclic peptide library.

In their next study, Varani and coworkers used the NMR structure of the TAR RNA bound cyclic peptide to guide the rational design of new larger libraries.<sup>108</sup> In the first step, an Ala scan was conducted at all amino acids of **BIV2** except those of the D-Pro-L-Pro template and Gly6 in an attempt to identify the side chains that were most
significant for binding while also preserving the  $\beta$ -turn structure. Each of the eleven mutations made in this study resulted in reduced binding affinity to TAR, and each mutation of a positively charged residue resulted in complete loss of binding affinity. Following this experiment, a library of 86 compounds was constructed where all residues of the RNA-contacting face of **BIV2** were mutated to side chains that might optimize the intermolecular interactions with the target. In addition, the residues of the solvent exposed face were mutated such that the total charge of the peptide would be reduced and so that the affinity might also be increased through indirect effects on the peptide's conformation. From this effort, several peptides were identified that achieved  $K_d$  values as low as 50 nM. Consensus sequence analysis and binding affinities revealed that Arg1, Arg3, and Arg5 were required for binding. Every attempt to mutate these residues to another side chain, including replacement with another amine such as Lys, resulted in total loss of affinity. This highlights the requirement for the hydrogen bond donor/positive charge provided through a guanidinium group.

Later, it was shown that several cyclic peptides from the 86 compound library, L-22, L-50, and L-51, had binding affinity for HIV-1 TAR with low nanomolar  $K_d$  values in the presence of a 10,000-fold excess of tRNA (Table 2.7).<sup>106</sup> Equally impressive was the selectivity displayed by L-22 and L-51 for BIV TAR and HIV-1 TAR, respectively. Furthermore, when the U23 nucleotide of HIV-1 TAR was either mutated to cytidine or removed completely, no binding was observed for L-22. The structure for the L-22– HIV-1 TAR complex was solved by NMR and revealed that the peptide induced a conformational change in the RNA, where the peptide was partially enclosed by the bulge and apical loop residues. Thus, a deep binding pocket was created in the major groove (residues 26–29 and 36–39) of the RNA upon association of **L-22** and buried nearly 51% of the peptide surface. Importantly, the apical loop region of TAR is responsible for binding to P-TEFb, an essential cellular cofactor for Tat mediated transcriptional elongation. However, only bases G33–A35 of the TAR loop were drawn in toward the peptide. Therefore, it was reasoned that further modifications in the peptide sequence could be made to increase interaction with this region using rational design.

**Table 2.7.** The sequences of lead peptides L-22, L-50, and L-51 and their affinities toward HIV-1 and BIV TAR RNA.<sup>a</sup>

Peptide	Sequence	$K_{\rm d}$ (HIV), nM	$K_{\rm d}$ (BIV), nM
L-22	<i>cyclo</i> -RVRTRKGRRIRI - <sup>D</sup> P-P	30	5
L-50	<i>cyclo</i> -RVRTRGKRRIRR- <sup>D</sup> P-P	1	1
L-51	<i>cyclo</i> -RTRTRGKRRIRV- <sup>D</sup> P-P	5	50

<sup>a</sup> Dissociation constants determined by EMSA.

Varani and coworkers recently made 25 additional cyclic peptides featuring 1, 2, or 4 additional amino acid residues.<sup>105</sup> Information gained from these compounds led to the preparation of another 12 peptide library. Unfortunately, the compound that provided the highest quality NOE data (**KP-Z-41**) in complex with HIV-1 TAR RNA was found to bind further down the major groove than **L-22** (Figure 2.10). Thus, additional interaction with the TAR RNA apical loop was not achieved. However, this study revealed critical interactions between the peptides and the interhelical junction of the RNA. Indeed, residues Arg7 and Lys8 of **KP-Z-41** have a nearly identical role in binding to HIV-1 TAR as Arg5 and Lys6 in **L-22**. In the case of **KP-Z-41**, Arg7 is essential for stabilizing a base triple between U23 and the A27–U38 base pair, which is observed upon binding of **L-22** as well. The positively charged N<sup>*e*</sup> of Lys8 is directed into a pocket of negative charge presented by the phosphate backbone between C24–G28, and the methyl groups of Ile12 are buried against U23. This lipophilic interaction is similar to that of **L-22** 

where Ile10 interacts with TAR. With this new information, it may be possible to design ligands that can bind closer to the loop region while also preserving these three critical interactions with TAR.



**Figure 2.10.** The structure, sequence, and affinity of **KP-Z-41** to HIV-1 TAR compared to **L-22**. Arrows point to residues that make key interactions with the RNA target determined by NMR.

Compounds from the original library (**L-22**, **L-50**, and **L-51**) have also been tested for activity in cell-based assays.<sup>104</sup> Arts and coworkers reported that compound **L-50** was the most potent inhibitor (IC<sub>50</sub> = 4.1  $\mu$ M) of replication in HIV-1<sub>NL4-3</sub> infected U87 cells expressing the CD4 receptor and CXCR4 coreceptor. The potency of **L-50** against HIV-1<sub>NL4-3</sub> was further increased in human PBMCs (IC<sub>50</sub> = 0.25  $\mu$ M), within the concentration range of the FDA-approved drugs 3TC and nevirapine. Interestingly, the potency of each compound mirrored their relative affinity for HIV-1 TAR, where **L-22**, **L-50**, and **L-51** had affinities (*K*<sub>d</sub>) of 30, 1, and 5, respectively.<sup>106</sup> Furthermore, the cytotoxicity profile of **L-51** showed good cell viability for incubation of up to ten days at 500  $\mu$ M. In order for an antiviral candidate to have utility in practice, it must be able to

inhibit many subtypes of HIV-1 such as clades of the virus. Gratifyingly, the researchers demonstrated that L-50 was capable of antiviral activity against both CCR5-tropic and CXCR4-tropic primary HIV-1 isolates that were representative of each of subtypes A, B, C, D, and CRF01 AE of the virus. This highlights the broad activity of this compound across many clades of HIV-1. Furthermore, because the L-50 was active in viruses that utilize CCR5 or CXCR4 co-receptors, it is likely that the compound does not inhibit viral entry, in contrast to the NeoR Tat mimetic developed by Lapidot and coworkers (vide *supra*).<sup>23,47,48</sup> Indeed, a cell-to-cell fusion assay confirmed that **L-50** did not inhibit entry of the virus. However, it was determined through cell-free transcription assays and several independent in vivo assays that L-50 reduced HIV-1 transcription by only twofold, which could not account for the 10-100-fold reduction in HIV-1 replication observed in their earlier viral replication assays. This discrepancy in HIV-1 inhibition was accounted for when further experiments revealed L-50 to also be a potent inhibitor of HIV-1 reverse transcription, making this cyclic peptide a dual inhibitor of HIV-1. The mechanism of reverse transcription inhibition is currently under further investigation.

#### 2.7 Conclusions

The development of ligands to control the function of RNA is a daunting task. Although they are structurally complex biomolecules, they are significantly more flexible than proteins. This is apparent through the association of cognate protein ligands, which favor binding to the RNA through mutually induced fit mechanisms.<sup>110,115</sup> Small molecules are at a disadvantage, because of their limited sampling area when bound to the RNA target. In order to affect RNA function, these molecules must be able to out compete much larger proteins that often bind through the interaction of large surface areas with the target. Large molecules that bind to the RNA primarily through Watson-Crick base pairing rules can achieve high selectivity, and they also have a greater potential to outcompete the endogenous protein ligands. However, the major flaw with this class of molecules broadly is that they are not cell permeable. In contrast, compounds of medium size have demonstrated good to excellent cell permeability in many cases. Furthermore, several medium-sized compounds have been developed with low nanomolar binding affinity to their RNA target and have been shown to inhibit RNA controlled processes in cell-based assays. Despite the advantages that medium-sized molecules have over their counterparts, no one has developed an RNA binding inhibitor that has made it to clinic. This further demonstrates the inherent difficulty of designing RNA binding ligands. Although the task is a daunting one, the successes and failures of prior attempts to provide solutions will guide the future of RNA inhibitor design.

# References

1. Bayne, E. H.; Allshire, R. C. RNA-directed transcriptional gene silencing in mammals. *Trends Genet.* **2005**, *21*, 370-373.

2. Fedor, M. J.; Williamson, J. R. The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 399-412.

3. Jovanovic, M.; Hengartner, M. O. miRNAs and apoptosis: RNAs to die for. *Oncogene* **2006**, *25*, 6176-6187.

4. Sen, G. L.; Blau, H. M. A brief history of RNAi: The silence of the genes. *FASEB J.* **2006**, *20*, 1293-1299.

5. Coppins, R. L.; Hall, K. B.; Groisman, E. A. The intricate world of riboswitches. *Curr. Opin. Microbiol.* **2007**, *10*, 176-181.

6. Schroeder, R.; Barta, A.; Semrad, K. Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 908-919.

7. Ippolito, J. A.; Steitz, T. A. The structure of the HIV-1 RRE high affinity Rev binding site at 1.60 Å resolution. *J. Mol. Biol.* **2000**, *295*, 711-717.

8. Battiste, J. L.; Mao, H.; Rao, N. S.; Tan, R.; Muhandiram, D. R.; Kay, L. E.; Frankel, A. D.; Williamson, J. R. A helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* **1996**, *273*, 1547-1551.

9. Aboul-ela, F.; Karn, J.; Varani, G. The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* **1995**, *253*, 313-332.

10. Paulsen, R. B.; Seth, P. P.; Swayze, E. E.; Griffey, R. H.; Skalicky, J. J.; Cheatham, T. E.; Davis, D. R. Inhibitor-induced structural change in the HCV IRES domain IIA RNA. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 7263-7268.

11. Berry, K. E.; Waghray, S.; Mortimer, S. A.; Bai, Y.; Doudna, J. A. Crystal structure of the HCV IRES central domain reveals strategy for start-codon positioning. *Structure* **2011**, *19*, 1456-1466.

12. Davis, B.; Afshar, M.; Varani, G.; Murchie, A. I. H.; Karn, J.; Lentzen, G.; Drysdale, M.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T.; Aboul-ela, F. Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *J. Mol. Biol.* **2004**, *336*, 343-356.

13. Stelzer, A. C.; Frank, A. T.; Kratz, J. D.; Swanson, M. D.; Gonzalez-Hernandez, M. J.; Lee, J.; Andricioaei, I.; Markovitz, D. M.; Al-Hashimi, H. M. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* **2011**, *7*, 553-559.

14. Gallego, J.; Varani, G. Targeting RNA with small-molecule drugs: Therapeutic promise and chemical challenges. *Acc. Chem. Res.* **2001**, *34*, 836-843.

15. Murchie, A. I. H.; Davis, B.; Isel, C.; Afshar, M.; Drysdale, M. J.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T. M.; Mirza, S.; Prescott, C. D.; Vaglio, P.; Aboul-ela, F.; Karn, J. Structure-based drug design targeting an inactive RNA conformation: Exploiting the flexibility of HIV-1 TAR RNA. *J. Mol. Biol.* **2004**, *336*, 625-638.

16. Luedtke, N. W.; Baker, T. J.; Goodman, M.; Tor, Y. Guanidinoglycosides: A novel family of RNA ligands. *J. Am. Chem. Soc.* **2000**, *122*, 12035-12036.

17. Chapman, R. L.; Stanley, T. B.; Hazen, R.; Garvey, E. P. Small molecule modulators of HIV Rev/Rev response element interaction identified by random screening. *Antiviral Res.* **2002**, *54*, 149-162.

18. Hwang, S.; Tamilarasu, N.; Kibler, K.; Cao, H.; Ali, A.; Ping, Y.-H.; Jeang, K.-T.; Rana, T. M. Discovery of a small molecule Tat-trans-activation-responsive RNA antagonist that potently inhibits human immunodeficiency virus-1 replication. *J. Biol. Chem.* **2003**, *278*, 39092-39103.

19. Xie, B.; Calabro, V.; Wainberg, M. A.; Frankel, A. D. Selection of TAR RNAbinding chameleon peptides by using a retroviral replication system. *J. Virol.* **2004**, *78*, 1456-1463.

20. Bryson, D. I.; Zhang, W. Y.; Ray, W. K.; Santos, W. L. Screening of a branched peptide library with HIV-1 TAR RNA. *Mol. BioSyst.* **2009**, *5*, 1070-1073.

21. Zapp, M. L.; Stern, S.; Green, M. R. Small molecules that selectively block RNA binding of HIV-1 Rev protein inhibit Rev function and viral production. *Cell* **1993**, *74*, 969-978.

22. Hamasaki, K.; Ueno, A. Aminoglycoside antibiotics, neamine and its derivatives as potent inhibitors for the RNA-protein interactions derived from HIV-1 activators. *Biorg. Med. Chem. Lett.* **2001**, *11*, 591-594.

23. Litovchick, A.; Lapidot, A.; Eisenstein, M.; Kalinkovich, A.; Borkow, G. Neomycin b–arginine conjugate, a novel HIV-1 Tat antagonist: Synthesis and anti-HIV activities. *Biochemistry* **2001**, *40*, 15612-15623.

24. Luedtke, N. W.; Liu, Q.; Tor, Y. RNA-ligand interactions: Affinity and specificity of aminoglycoside dimers and acridine conjugates to the HIV-1 Rev response element. *Biochemistry* **2003**, *42*, 11391-11403.

25. Chow, C. S.; Bogdan, F. M. A structural basis for RNA-ligand interactions. *Chem. Rev.* **1997**, *97*, 1489-1514.

26. Thomas, J. R.; Hergenrother, P. J. Targeting RNA with small molecules. *Chem. Rev.* **2008**, *108*, 1171-1224.

27. Guan, L.; Disney, M. D. Recent advances in developing small molecules targeting RNA. *ACS Chem. Biol.* **2012**, *7*, 73-86.

28. Hermann, T. Drugs targeting the ribosome. *Curr. Opin. Struct. Biol.* **2005**, *15*, 355-366.

29. Foloppe, N.; Matassova, N.; Aboul-Ela, F. Towards the discovery of drug-like RNA ligands? *Drug Discov. Today* **2006**, *11*, 1019-1027.

30. Favaretto, P.; Bhutkar, A.; Smith, T. F. Constraining ribosomal RNA conformational space. *Nucleic Acids Res.* **2005**, *33*, 5106-5111.

31. Yusupov, M. M.; Yusupova, G. Z.; Albion, B.; Lieberman, K.; Earnest, T. N.; Cate, J. H. D.; Noller, H. F. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **2001**, *292*, 883-896.

32. Verhelst, S. H. L.; Michiels, P. J. A.; van der Marel, G. A.; van Boeckel, C. A. A.; van Boom, J. H. Surface plasmon resonance evaluation of various aminoglycoside–RNA hairpin interactions reveals low degree of selectivity. *ChemBioChem* **2004**, *5*, 937-942.

33. Achenbach, T. V.; Brunner, B.; Heermeier, K. Oligonucleotide-based knockdown technologies: Antisense versus RNA interference. *ChemBioChem* **2003**, *4*, 928-935.

34. Dias, N.; Stein, C. A. Antisense oligonucleotides: Basic concepts and mechanisms. *Mol. Cancer Ther.* **2002**, *1*, 347-355.

35. Dykxhoorn, D. M.; Lieberman, J. The silent revolution: RNA interference as basic biology, research tool, and therapeutic. *Annu. Rev. Med.* **2005**, *56*, 401-423.

36. Hannon, G. J. RNA interference. Nature 2002, 418, 244-251.

37. Zamore, P. D. Ancient pathways programmed by small RNAs. *Science* **2002**, *296*, 1265-1269.

38. Peacock, H.; Fostvedt, E.; Beal, P. A. Minor-groove-modulating adenosine replacements control protein binding and RNAi activity in siRNAs. *ACS Chem. Biol.* **2010**, *5*, 1115-1124.

39. Jackson, A. L.; Bartz, S. R.; Schelter, J.; Kobayashi, S. V.; Burchard, J.; Mao, M.; Li, B.; Cavet, G.; Linsley, P. S. Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **2003**, *21*, 635-637.

40. Li, C. X.; Parker, A.; Menocal, E.; Xiang, S.; Borodyansky, L.; Fruehauf, J. H. Delivery of RNA interference. *Cell Cycle* **2006**, *5*, 2103-2109.

41. Xu, L.; Anchordoquy, T. Drug delivery trends in clinical trials and translational medicine: Challenges and opportunities in the delivery of nucleic acid-based therapeutics. *J. Pharm. Sci.* **2011**, *100*, 38-52.

42. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **1997**, *23*, 3-25.

43. Bannwarth, S.; Gatignol, A. HIV-1 TAR RNA: The target of molecular interactions between the virus and its host. *Curr. HIV Res.* **2005**, *3*, 61-71.

44. Brigati, C.; Giacca, M.; Noonan, D. M.; Albini, A. HIV Tat, its targets and the control of viral gene expression. *FEMS Microbiol. Lett.* **2003**, *220*, 57-65.

45. Rana, T. M.; Jeang, K.-T. Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA. *Arch. Biochem. Biophys.* **1999**, *365*, 175-185.

46. Wang, S.; Huber, P. W.; Cui, M.; Czarnik, A. W.; Mei, H. Y. Binding of neomycin to the TAR element of HIV-1 RNA induces dissociation of Tat protein by an allosteric mechanism. *Biochemistry* **1998**, *37*, 5549-5557.

47. Berchanski, A.; Kalinkovich, A.; Ludin, A.; Lapidot, T.; Lapidot, A. Insights into the mechanism of enhanced mobilization of hematopoietic progenitor cells and release of CXC112 by a combination of AMD3100 and aminoglycoside-polyarginine conjugates. *FEBS J.* **2011**, 278, 4150-4165.

48. Berchanski, A.; Lapidot, A. Computer-based design of novel HIV-1 entry inhibitors: Neomycin conjugated to arginine peptides at two specific sites. *J. Mol. Model.* **2009**, *15*, 281-294.

49. Felber, B. K.; Hadzopoulou-Cladaras, M.; Cladaras, C.; Copeland, T.; Pavlakis, G. N. Rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 1495-1499.

50. Hammarskjold, M. L.; Heimer, J.; Hammarskjold, B.; Sangwan, I.; Albert, L.; Rekosh, D. Regulation of human immunodeficiency virus env expression by the Rev gene product. *J. Virol.* **1989**, *63*, 1959-1966.

51. Malim, M. H.; Hauber, J.; Le, S. Y.; Maizel, J. V.; Cullen, B. R. The HIV-1 Rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **1989**, *338*, 254-257.

52. Heaphy, S.; Dingwall, C.; Ernberg, I.; Gait, M. J.; Green, S. M.; Karn, J.; Lowe, A. D.; Singh, M.; Skinner, M. A. HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element region. *Cell* **1990**, *60*, 685-693.

53. Malim, M. H.; Tiley, L. S.; McCarn, D. F.; Rusche, J. R.; Hauber, J.; Cullen, B. R. HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* **1990**, *60*, 675-683.

54. Dayton, A. I.; Sodroski, J. G.; Rosen, C. A.; Goh, W. C.; Haseltine, W. A. The transactivator gene of the human T cell lymphotropic virus type III is required for replication. *Cell* **1986**, *44*, 941-947.

55. Kjems, J.; Brown, M.; Chang, D. D.; Sharp, P. A. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 683-687.

56. Hendrix, M.; Priestley, E. S.; Joyce, G. F.; Wong, C. H. Direct observation of aminoglycoside-RNA interactions by surface plasmon resonance. *J. Am. Chem. Soc.* **1997**, *119*, 3641-3648.

57. Cho, J.; Rando, R. R. Specificity in the binding of aminoglycosides to HIV-RRE RNA. *Biochemistry* **1999**, *38*, 8548-8554.

58. Luedtke, N. W.; Tor, Y. Fluorescence-based methods for evaluating the RNA affinity and specificity of HIV-1 Rev-RRE inhibitors. *Biopolymers* **2003**, *70*, 103-119.

59. Lacourciere, K. A.; Stivers, J. T.; Marino, J. P. Mechanism of neomycin and Rev peptide binding to the Rev responsive element of HIV-1 as determined by fluorescence and NMR spectroscopy. *Biochemistry* **2000**, *39*, 5630-5641.

60. Tok, J. B.; Dunn, L. J.; Des Jean, R. C. Binding of dimeric aminoglycosides to the HIV-1 Rev responsive element (RRE) RNA construct. *Biorg. Med. Chem. Lett.* **2001**, *11*, 1127-1131.

61. Wang, H.; Tor, Y. Dimeric aminoglycosides: Design, synthesis and RNA binding. *Biorg. Med. Chem. Lett.* **1997**, *7*, 1951-1956.

62. Daugherty, M. D.; Booth, D. S.; Jayaraman, B.; Cheng, Y.; Frankel, A. D. HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 12481-12486.

63. Daugherty, M. D.; Liu, B.; Frankel, A. D. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1337-1342.

64. DiMattia, M. A.; Watts, N. R.; Stahl, S. J.; Rader, C.; Wingfield, P. T.; Stuart, D. I.; Steven, A. C.; Grimes, J. M. Implications of the HIV-1 Rev dimer structure at 3.2 Å resolution for multimeric binding to the Rev response element. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5810-5814.

65. Suhasini, M.; Reddy, T. R. Cellular proteins and HIV-1 Rev function. *Curr. HIV Res.* **2009**, *7*, 91-100.

66. Frankel, A. D.; Young, J. A. T. HIV-1: Fifteen proteins and an RNA. *Annu. Rev. Biochem* **1998**, 67, 1-25.

67. Stevens, M.; De Clercq, E.; Balzarini, J. The regulation of HIV-1 transcription: Molecular targets for chemotherapeutic intervention. *Med. Res. Rev.* **2006**, *26*, 595-625.

68. Pestova, T. V.; Shatsky, I. N.; Fletcher, S. P.; Jackson, R. J.; Hellen, C. U. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev.* **1998**, *12*, 67-83.

69. Sizova, D. V.; Kolupaeva, V. G.; Pestova, T. V.; Shatsky, I. N.; Hellen, C. U. Specific interaction of eukaryotic translation initiation factor 3 with the 5' nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. *J. Virol.* **1998**, *72*, 4775-4782.

70. Kolupaeva, V. G.; Pestova, T. V.; Hellen, C. U. An enzymatic footprinting analysis of the interaction of 40S ribosomal subunits with the internal ribosomal entry site of hepatitis C virus. *J. Virol.* **2000**, *74*, 6242-6250.

71. Kieft, J. S.; Zhou, K.; Jubin, R.; Doudna, J. A. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **2001**, *7*, 194-206.

72. Weeks, K. M.; Ampe, C.; Schultz, S. C.; Steitz, T. A.; Crothers, D. M. Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science* **1990**, *249*, 1281-1285.

73. Cordingley, M. G.; LaFemina, R. L.; Callahan, P. L.; Condra, J. H.; Sardana, V. V.; Graham, D. J.; Nguyen, T. M.; LeGrow, K.; Gotlib, L.; Schlabach, A. J.; et al. Sequence-specific interaction of Tat protein and Tat peptides with the transactivation-responsive sequence element of human immunodeficiency virus type 1 in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8985-8989.

74. Calnan, B. J.; Biancalana, S.; Hudson, D.; Frankel, A. D. Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. *Genes Dev.* **1991**, *5*, 201-210.

75. Calnan, B. J.; Tidor, B.; Biancalana, S.; Hudson, D.; Frankel, A. D. Argininemediated RNA recognition: The arginine fork. *Science* **1991**, *252*, 1167-1171.

76. Kjems, J.; Frankel, A. D.; Sharp, P. A. Specific regulation of mRNA splicing in vitro by a peptide from HIV-1 Rev. *Cell* **1991**, *67*, 169-178.

77. Kjems, J.; Calnan, B. J.; Frankel, A. D.; Sharp, P. A. Specific binding of a basic peptide from HIV-1 Rev. *EMBO J.* **1992**, *11*, 1119-1129.

78. Tan, R.; Chen, L.; Buettner, J. A.; Hudson, D.; Frankel, A. D. RNA recognition by an isolated  $\alpha$  helix. *Cell* **1993**, *73*, 1031-1040.

79. Hyun, S.; Kim, H. J.; Lee, N. J.; Lee, K. H.; Lee, Y.; Ahn, D. R.; Kim, K.; Jeong, S.; Yu, J. A-helical peptide containing N,N-dimethyl lysine residues displays low-nanomolar and highly specific binding to RRE RNA. *J. Am. Chem. Soc.* **2007**, *129*, 4514-4515.

80. Lee, Y.; Hyun, S.; Kim, H. J.; Yu, J. Amphiphilic helical peptides containing two acridine moieties display picomolar affinity toward HIV-1 RRE and TAR. *Angew. Chem. Int. Ed.* **2008**, *47*, 134-137.

81. Lee, S. J.; Hyun, S.; Kieft, J. S.; Yu, J. An approach to the construction of tailormade amphiphilic peptides that strongly and selectively bind to hairpin RNA targets. *J. Am. Chem. Soc.* **2009**, *131*, 2224-2230.

82. Cox, J. A.; Comte, M.; Fitton, J. E.; DeGrado, W. F. The interaction of calmodulin with amphiphilic peptides. *J. Biol. Chem.* **1985**, *260*, 2527-2534.

83. Usui, K.; Ojima, T.; Takahashi, M.; Nokihara, K.; Mihara, H. Peptide arrays with designed secondary structures for protein characterization using fluorescent fingerprint patterns. *Biopolymers* **2004**, *76*, 129-139.

84. Liu, Q.; Dreyfuss, G. In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.* **1995**, *15*, 2800-2808.

85. Tan, R.; Frankel, A. D. A novel glutamine-RNA interaction identified by screening libraries in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4247-4252.

86. Hyun, S.; Na, J.; Lee, S. J.; Park, S.; Yu, J. RNA grooves can accommodate disulfide-bridged bundles of α-helical peptides. *ChemBioChem* **2010**, *11*, 767-770.

87. Mills, N. L.; Daugherty, M. D.; Frankel, A. D.; Guy, R. K. An α-helical peptidomimetic inhibitor of the HIV-1 Rev–RRE interaction. *J. Am. Chem. Soc.* **2006**, *128*, 3496-3497.

88. Eyre, D. R. Collagen: Molecular diversity in the body's protein scaffold. *Science* **1980**, 207, 1315-1322.

89. Barthe, P.; Rochette, S.; Vita, C.; Roumestand, C. Synthesis and NMR solution structure of an  $\alpha$ -helical hairpin stapled with two disulfide bridges. *Protein Sci.* **2000**, *9*, 942-955.

90. Schafmeister, C. E.; Po, J.; Verdine, G. L. An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides. *J. Am. Chem. Soc.* **2000**, *122*, 5891-5892.

91. Simon, R. J.; Kania, R. S.; Zuckermann, R. N.; Huebner, V. D.; Jewell, D. A.; Banville, S.; Ng, S.; Wang, L.; Rosenberg, S.; Marlowe, C. K. Peptoids: A modular approach to drug discovery. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9367-9371.

92. Hamy, F.; Felder, E. R.; Heizmann, G.; Lazdins, J.; Aboul-ela, F.; Varani, G.; Karn, J.; Klimkait, T. An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3548-3553.

93. Kesavan, V.; Tamilarasu, N.; Cao, H.; Rana, T. M. A new class of RNA-binding oligomers: Peptoid amide and ester analogues. *Bioconjugate Chem.* **2002**, *13*, 1171-1175.

94. Wang, X.; Huq, I.; Rana, T. M. HIV-1 TAR RNA recognition by an unnatural biopolymer. *J. Am. Chem. Soc.* **1997**, *119*, 6444-6445.

95. Tamilarasu, N.; Huq, I.; Rana, T. M. High affinity and specific binding of HIV-1 TAR RNA by a Tat-derived oligourea. *J. Am. Chem. Soc.* **1999**, *121*, 1597-1598.

96. Gelman, M. A.; Richter, S.; Cao, H.; Umezawa, N.; Gellman, S. H.; Rana, T. M. Selective binding of TAR RNA by a Tat-derived  $\beta$ -peptide. *Org. Lett.* **2003**, *5*, 3563-3565.

97. Cheng, R. P.; Gellman, S. H.; DeGrado, W. F. B-peptides: From structure to function. *Chem. Rev.* **2001**, *101*, 3219-3232.

98. Werder, M.; Hauser, H.; Abele, S.; Seebach, D. B-peptides as inhibitors of smallintestinal cholesterol and fat absorption. *Helv. Chim. Acta* **1999**, *82*, 1774-1783.

99. Gademann, K.; Hintermann, T.; Schreiber, J. V. β-peptides: Twisting and turning. *Curr. Med. Chem.* **1999**, *6*, 905-925.

100. Gregoire, C.; Peloponese, J. M., Jr.; Esquieu, D.; Opi, S.; Campbell, G.; Solomiac, M.; Lebrun, E.; Lebreton, J.; Loret, E. P. Homonuclear (1)H-NMR assignment and structural characterization of human immunodeficiency virus type 1 Tat mal protein. *Biopolymers* **2001**, *62*, 324-335.

101. Niu, Y.; Hu, Y.; Li, X.; Chen, J.; Cai, J.  $\gamma$ -AApeptides: Design, synthesis and evaluation. *New J. Chem.* **2011**, *35*.

102. Niu, Y.; Jones, A. J.; Wu, H.; Varani, G.; Cai, J. γ-AApeptides bind to RNA by mimicking RNA-binding proteins. *Org. Biomol. Chem.* **2011**, *9*, 6604-6609.

103. Niu, Y.; Padhee, S.; Wu, H.; Bai, G.; Harrington, L.; Burda, W. N.; Shaw, L. N.; Cao, C.; Cai, J. Identification of  $\gamma$ -AApeptides with potent and broad-spectrum antimicrobial activity. *Chem. Commun.* **2011**, *47*, 12197-12199.

104. Lalonde, M. S.; Lobritz, M. A.; Ratcliff, A.; Chamanian, M.; Athanassiou, Z.; Tyagi, M.; Wong, J.; Robinson, J. A.; Karn, J.; Varani, G.; Arts, E. J. Inhibition of both HIV-1 reverse transcription and gene expression by a cyclic peptide that binds the Tattransactivating response element (TAR) RNA. *PLoS Path.* **2011**, *7*, e1002038.

105. Davidson, A.; Patora-Komisarska, K.; Robinson, J. A.; Varani, G. Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of Tat protein. *Nucleic Acids Res.* **2011**, *39*, 248-256.

106. Davidson, A.; Leeper, T. C.; Athanassiou, Z.; Patora-Komisarska, K.; Karn, J.; Robinson, J. A.; Varani, G. Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11931-11936.

107. Chaloin, L.; Smagulova, F.; Hariton-Gazal, E.; Briant, L.; Loyter, A.; Devaux, C. Potent inhibition of HIV-1 replication by backbone cyclic peptides bearing the Rev arginine rich motif. *J. Biomed. Sci.* **2007**, *14*, 565-584.

108. Athanassiou, Z.; Patora, K.; Dias, R. L. A.; Moehle, K.; Robinson, J. A.; Varani, G. Structure-guided peptidomimetic design leads to nanomolar  $\beta$ -hairpin inhibitors of the Tat–TAR interaction of bovine immunodeficiency virus. *Biochemistry* **2007**, *46*, 741-751.

109. Leeper, T. C.; Athanassiou, Z.; Dias, R. L.; Robinson, J. A.; Varani, G. TAR RNA recognition by a cyclic peptidomimetic of Tat protein. *Biochemistry* **2005**, *44*, 12362-12372.

110. Athanassiou, Z.; Dias, R. L.; Moehle, K.; Dobson, N.; Varani, G.; Robinson, J. A. Structural mimicry of retroviral Tat proteins by constrained  $\beta$ -hairpin peptidomimetics: Ligands with high affinity and selectivity for viral TAR RNA regulatory elements. *J. Am. Chem. Soc.* **2004**, *126*, 6906-6913.

111. Runyon, S. T.; Puglisi, J. D. Design of a cyclic peptide that targets a viral RNA. *J. Am. Chem. Soc.* **2003**, *125*, 15704-15705.

112. Tok, J. B.; Des Jean, R. C.; Fenker, J. Binding of a cyclic BIV  $\beta$ -Tat peptide with its TAR RNA construct. *Biorg. Med. Chem. Lett.* **2001**, *11*, 43-46.

113. Puglisi, J. D.; Chen, L.; Blanchard, S.; Frankel, A. D. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science* **1995**, *270*, 1200-1203.

114. Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. Conformation of  $\beta$ -hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol.* **1989**, *206*, 759-777.

115. Leulliot, N.; Varani, G. Current topics in RNA-protein recognition: Control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **2001**, *40*, 7947-7956.

# Chapter 3 Screening of Branched Peptide Library with HIV-1 TAR RNA

# Attributions

This chapter was taken from Bryson, D. I.; Zhang, W. Y.; Ray, W. K.; Santos, W. L. Screening of a branched peptide library with HIV-1 TAR RNA. *Mol. BioSyst.* **2009**, *5*, 1070-1073. Most of the work described in this chapter was performed by the author of this dissertation. He was solely responsible for the synthesis of the peptide library, RNA synthesis, high throughput screening, screening optimization, and all biophysical characterizations. Wenyu Zhang of the Santos Group aided in the preparation and purification of several hit peptide compounds. Dr. W. Keith Ray of the VT mass spectrometry incubator performed MALDI analysis of hit compounds and performed the majority of sequence deconvolution. The author significantly contributed to the writing and editing of the manuscript. The final manuscript was prepared by Dr. Webster L. Santos. [reproduced by permission of the Royal Society of Chemistry]

# Abstract

A series of HIV-1 TAR RNA-selective ligands were discovered using an on-bead screening of a library of 4,096 branched peptides.

# 3.1 Introduction

The recognition that RNA is more than just an intermediate in the information transfer from genetic code to fully functional protein has placed it in the forefront of chemical research. RNA is important because of its vital role in regulating transcription, translation, splicing, replication and catalysis.<sup>1</sup> Consequently, molecules that can bind to RNA and control its function have potential as powerful tools in biology and medicine.

In principle, the tertiary structure of RNA can provide a scaffold in which molecules can dock and form the necessary intermolecular interactions to become highly selective. The formation of unique three dimensional architecture, often generated through formation of local structures such as bulges, stem-loops, pseudoknots and turns, differentiates RNA from DNA and allows it to become a therapeutic target.<sup>2</sup> In fact, antibacterial drugs (aminoglycoside, macrolide, oxazolidinone and tetracycline) have cemented the drug-feasibility of RNA.<sup>3</sup> These molecules exert their antibacterial activity by binding to specific regions of rRNA. However, it is clear that targeting generic RNA with small molecules is difficult and requires further studies.<sup>4</sup> Still, the difficult task is designing ligands independent of the canonical Watson-Crick base-pairing on RNA primary structure. Thus, we became interested in the widely recognized problem of developing RNA-selective molecules.

Many important functions of RNA result from specific proteins binding to complex tertiary structures of RNA. One such example is evident in the HIV-1 transactivation response element region (TAR) RNA, a conserved 59-nucleotide stem loop located at the 5' end of all nascent transcribed HIV-1 mRNA.<sup>5,6</sup> The TAR structure is comprised of a highly conserved hexanucleotide loop and a three-nucleotide bulge flanked by two double stranded stems (Figure 3.1). Binding of the transcriptional activator protein Tat and the cyclin T1/cdk1 kinase complex promotes

efficient transcriptional elongation. Tat is an 86 amino acid protein that contains an arginine rich motif (ARM), which specifically interacts with the TAR tri-nucleotide bulge (U23, C24, and U25). Disruption of Tat-TAR interaction results in the blockade of viral replication and thus represents a viable strategy in developing new anti-HIV therapeutics. Several ligands (aminoglycosides,<sup>7</sup> argininamide,<sup>8</sup>\_Epeptides,<sup>9</sup> peptidomimetics,<sup>10</sup> small molecules,<sup>11,12</sup> and others<sup>13</sup>) are continuously being developed to inhibit Tat-TAR interaction (Figure 1).



**Figure 3.1.** Sequence and secondary structure of HIV-1 TAR RNA (A), small molecule ligands (neomycin B (B) and argininamide (C)) and 11-mer Tat-derived peptide (D) that inhibit Tat-TAR interaction.

# **3.2 Results and Discussion**

Owing to the complexity, the lack of designable ligands and a growing interest in developing new molecules that can selectively interact with RNA, we aimed to develop a strategy that allows for the rapid screening of a short branched peptide library and determine their capacity as possible ligands for HIV-1 TAR RNA (Figure 3.2). Branched peptides have found extensive use in biological systems including synthetic peptide vaccine (multiple antigen peptides), drug delivery vehicles and therapeutics for a variety

of disease states.<sup>14-16</sup> To date, the potential of using branched peptides for the selective targeting of structured RNA targets has not been realized. Because multivalency is often described to increase the affinity of a ligand to a particular receptor.<sup>17</sup> we wanted to determine whether branched molecules are capable of forming multivalent interactions with RNA and whether presentation of structurally diverse functional groups can enhance selectivity with the target RNA. Branched peptides are a good starting point because naturally occurring amino acids contain a wide array of substituents that can display unique molecular architectures, and their synthetic accessibility is straightforward. An attractive feature of this strategy is the synthesis of the library on solid support, allowing for rapid generation of a large number of peptides using the split and pool technique, and subsequent on-bead screening against TAR.<sup>9,18-21</sup> We designed our initial branched peptide library with biased parameters that enhance binding and selectivity to RNA (Figure 3.2). These interactions include electrostatics (positively charged residues interacting with the negatively charged phosphate),  $\pi$ - $\pi$  interactions and hydrogen bonding. In principle, the diversity of possible peptide structures can be immense because of the commercial availability of unnatural amino acid monomers.

$A_1 - A_2 - A_3$ $A_1 - A_2 - A_3$ $A_1 - A_2 - A_3$ $A_1 - A_2 - A_3$							
	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	
	Asp	Arg	Tyr	His	Tyr	Phe	
	Arg	Asn	Leu	Val	Arg	Leu	
	Tyr	His	Trp	Asn	Ala	Asp	
	His	Val	Ala	Trp	Gln	Ser	

**Figure 3.2.** Combinatorial library of 4,096 branched peptides synthesized by split and pool technique. Peptides are attached to TentaGel resin via a photocleavable linker, and the branching unit used was lysine.

The branched peptide library was prepared such that each variable position contained one of four possible amino acid monomers (Figure 3.2). Branches  $A_1$ — $A_3$  had an identical sequence in order to simplify the *de novo* sequencing by MALDI/mass spectrometry (MS) and were branched through the  $\alpha$ - and  $\varepsilon$ -amino groups on lysine. Using orthogonal protecting groups, 4,096 compounds were generated by the split and pool technique in a few peptide coupling steps. It was important that the peptides were not released from the beads after deprotection of the side chain residues because the on-bead screening required resin-tethered probes. Thus, peptides were attached to TentaGel NH<sub>2</sub> beads via a photocleavable linker (3-amino-3-(2-nitrophenyl)propionic acid, ANP),<sup>22</sup> which allowed the efficient release of hit molecules by UV-irradiation, post screening (Scheme 3.1).



**Scheme 3.1.** On-bead screening of the branched peptide library against 5'-DY547-labeled TAR RNA and hit identification using MALDI/MS.

Previous high-throughput screening methods suggested that the on-bead screening method can be adapted with our branched peptide library.<sup>9,18</sup> Following this protocol, we transferred three copies of the library in an Eppendorf tube and washed them with water  $(5\times)$  and buffer  $(3\times)$  containing 50 mM Tris·HCl, 20 mM KCl, and 0.1% Triton X-100 at pH 7.4.<sup>23</sup> In order to address the issue of selectivity at the outset, we decided to preclude the beads containing promiscuous RNA binders. A simple process that can increase assay selectivity involves the addition of competitor RNA. In this case, unlabeled  $\alpha$ -synuclein mRNA, a 340-nucleotide long RNA, was added in the incubation mixture. We hypothesized that longer RNA sequences are likely to bind nonspecifically to our library, and abolishing promiscuous binders can decrease the potential for off-target binding. Thus, we performed a preliminary incubation of a small subset of beads (~2000) in BSA (1 mg/mL) and 500 nM unlabeled  $\alpha$ -synuclein mRNA at 4 °C for 3 h and washed (3×) the beads to remove unbound BSA and RNA in solution. Finally, TAR RNA screening was effected by incubating the 29-nucleotide TAR RNA-labeled with DY547 fluorescent tag on the 5'-end at 3.4 nM concentration for 3 h at 4 °C. The solution was then filtered,

washed successively with buffer, placed in a 96-well plate and analyzed using confocal microscopy. It was evident from the initial experiment that the stringency of the assay was inadequate because out of ~2000 beads, six beads were sufficiently fluorescent to isolate them.

The results of the preliminary incubation studies suggested that a more stringent protocol is necessary to decrease the number of hit compounds that will have to be analyzed. To achieve this goal, the incubation with DY547 labeled TAR RNA was reduced to 1 h. Following this protocol for the remaining library members, 10 additional beads were selected. Figure 3.3 shows three representative fluorescent beads visualized under laser irradiation (3A) and transmitted light with laser irradiation (3B). These beads were washed with buffer and organic solvents to remove the bound RNA until fluorescence was no longer visible under the confocal microscope. As expected, rescreening of the same beads with DY547-TAR RNA under a more stringent condition confirmed RNA binding (*vide infra*). Finally, labeled RNA was removed with copious buffer and organic solvent washes, which was followed by photolytic release of branched peptides from the resin by exposure to UV irradiation (365 nm, handheld UV lamp, 4 W) for 1 h.



**Figure 3.3.** Visualization of hit beads under confocal microscope. Transmitted light (A) and laser irradiation (B); arrows point to additional beads. Sequence alignment of 16 hits compounds generated a LOGO shown in (C).

The identity of the peptide hits were determined by *de novo* sequencing of the resulting solution by MALDI/MS (Table 3.1). Sequence homology analysis of the hit peptides revealed a LOGO shown in Figure 3C.<sup>24</sup> Of the 16 branched peptides, twelve contained Arg-Arg on the N-terminus—an unsurprising outcome given that RNA binding can be mediated by electrostatic interactions. This result suggests that a positive charge is favored both in the  $A_1$  and  $A_2$  positions. While no overwhelming preference for a specific amino acid was observed at any other positions, interestingly, arginine was not the preferred amino acid in position  $A_5$ . Because the propensity for RNA binding is expected to increase in the presence of an electrostatic interaction, albeit nonselectively, the preference for tyrosine in position  $A_5$  is noteworthy. It appears that a favorable electrostatic force can be adequately replaced with potential hydrogen bonding and hydrophobic interactions. Consistent with this observation, hydrophobic residues, Phe and Leu, dominate position A<sub>6</sub> when possible amino acids such as Ser and Asp that can hydrogen bond are present. It was also gratifying to find that one of the beads collected from the library screening, BP16, contained the sequence with the highest amino acid frequency as indicated in the LOGO illustration in Figure 3.3C.

	Sequence		Sequence
BP1	(RRL) <sub>2</sub> *WYL	BP9	(RRA) <sub>2</sub> *NYF
BP2	(YRA) <sub>2</sub> *HRF	BP10	(DNL) <sub>2</sub> *HYF
BP3	(HRW) <sub>2</sub> *WAS	<b>BP11</b>	(RRA) <sub>2</sub> *VYF
BP4	(RRW) <sub>2</sub> *HAL	BP12	(RRW) <sub>2</sub> *HAS
BP5	(RRL) <sub>2</sub> *NRF	BP13	(RRY) <sub>2</sub> *NQD
BP6	(RRY) <sub>2</sub> *VRL	BP14	(RRY) <sub>2</sub> *VQL
BP7	(RRW) <sub>2</sub> *HYD	BP15	(RRA) <sub>2</sub> *VRD
BP8	(YRL) <sub>2</sub> *WRL	<b>BP16</b>	(RRL) <sub>2</sub> *HYL

Table 3.1. Sequence of branched peptides BP1–BP16 derived from MALDI/MSMS.

To evaluate whether the branched peptide hits bound TAR RNA, biophysical characterization was undertaken. We selected **BP15**, **BP16** and **BP17** ([RRL]<sub>2</sub>\*HRF) for analysis. Validation of the activities of these compounds was paramount because false positives were possible, as in any high throughput screening assay. **BP15** was randomly selected from the hit peptides, and although **BP17** was not a sequence derived from the library screening, we hypothesized that this sequence should have good binding affinity for HIV-1 TAR based on the consensus sequence analysis. For this purpose, the dissociation constants for these compounds were determined using fluorescence polarization (FP). This technique required the resynthesis of select peptides with FITC on one of the N-terminal branches, which was accomplished using orthogonally protected lysine (Fmoc-Lys(ivDde)-OH) as the branching unit (Figure 3.4). The resulting branched peptides contained a N-terminal acetyl group on one branch. After HPLC purification and characterization by MALDI/MS, FP experiments were performed in triplicate for each peptide (Figure 3.5). In each case, 0.5 nM FITC-peptide was incubated with increasing concentration of TAR RNA (up to 100  $\mu$ M). The results reveal that **BP15** has a  $K_d$  of 27  $\mu$ M while **BP16–17** have affinity for TAR RNA in the  $\mu$ M range (Figure 3.5). Because the FP technique required the use of high concentration of RNA to reach full saturation, in this case >100  $\mu$ M, we were not able to determine the accurate dissociation constants for **BP16–17**. Gratifyingly, when a mutant version of TAR (U24>C) was titrated against **BP15–17**, it is clear that their binding affinity is severely decreased, suggesting that these peptides are selective towards the TAR RNA used in the screening. As a control, we also synthesized a linear peptide version of **BP15**, **T15**, to investigate whether branching has an effect on the binding affinity towards TAR RNA. T15 has a sequence of RRAGVRD, where the branching unit Lys was replaced with Gly. The data indicates a marked loss of binding affinity (> 10 fold) suggesting that the additional functional groups present on the branch aid in anchoring to TAR RNA.



**Figure 3.4.** Synthesis of FITC-labeled peptides. SPPS, solid-phase peptide synthesis; Ac, acetyl; FITC, fluorescein isothiocyanate.



**Figure 3.5.** Fluorescence polarization binding curves of branched peptides with HIV-1 TAR RNA and mutant (U24>C) TAR RNA. **T15** is a linear peptide (RRAGVRD) version of **BP15**. Each experiment was done in triplicate.

In conclusion, we discovered selective binders for HIV-1 TAR RNA using an on-bead screening of a focused library of branched peptides. The use of branched structures with diverse functional groups and capacity to form multivalent interactions is important in defining new molecular entities that can be selective for the tertiary structure of target RNA. In addition to synthesizing a more complex branched peptide library, future work in our laboratory will focus on extensive characterization including mapping of TAR-branched peptide interactions, cell permeability and *in-vitro* inhibition of Tat-TAR RNA interactions.

# **3.3** Materials and Methods

#### Synthesis of branched peptides on resin

Synthesis of the branched peptides was achieved by solid phase peptide synthesis using N- $\alpha$ -Fmoc protected L-amino acids (Novabiochem), HCTU (Peptides International) in DMF as coupling reagent, and DIEA (Aldrich). Tentagel Macrobead-NH<sub>2</sub> resin (Peptides International) with 0.3 mmol/g loading was coupled with photocleavable linker Fmoc-ANP<sup>25</sup> before synthesizing the control branched peptide (KYR)<sub>2</sub>\*FDS and the 3.3.3 library. Solid phase synthesis was done on a vacuum manifold (Qiagen) outfitted with 3-way Luer lock stopcocks (Sigma) in either Poly-Prep columns or Econo-Pac polypropylene columns (Bio-Rad). The resin was mixed in solution by bubbling argon during all coupling and washing steps.

Coupling ANP and N- $\alpha$ -Fmoc protected L-amino acids in control branched peptide

Tentagel Macrobead-NH<sub>2</sub> resin (28 mg) with 0.3 mmol/g loading (Peptides International) was swollen in DCM (2 mL,  $2 \times 15$  min) followed by DMF (2 mL, 15 min). Tentagel Macrobead-NH<sub>2</sub> resin was first coupled for 3 h in DMF with 3 equiv. of Fmoc-ANP in presence of HCTU (3 equiv) and DIEA (5 equiv). N-Fmoc amino acids (3 equiv) were then coupled to resin bearing ANP linker in DMF in the presence of HCTU (3 equiv) and DIEA (5 equiv). Fmoc-Lys(Fmoc)-OH was used as the branching unit. Molar equivalences of all reagents were doubled in coupling reactions after incorporation of the branching unit. Beads were washed extensively with DMF between reactions, and were tested for completeness via Kaiser test.

## Coupling ANP and N- $\alpha$ -Fmoc protected L-amino acids in 3.3.3 library

Tentagel Macrobead-NH<sub>2</sub> resin (200 mg) with 0.3 mmol/g loading was swollen and coupled with ANP as was done in the preparation of the control branched peptide. N-Fmoc amino acids (3 equiv) were then coupled to resin bearing ANP linker in NMP in the presence of HCTU (3 equiv) and DIEA (5 equiv). Molar equivalences of reagents were doubled after incorporation of the Lys branching unit. The standard split and pool method was utilized between reactions in the preparation of the 3.3.3 library. Beads were washed extensively with NMP between reactions, and were tested for completeness via Kaiser test.

#### Coupling N- $\alpha$ -Fmoc protected L-amino acids in FITC-peptides

Rink amide MBHA resin (40 mg) with 0.7 mmol/g loading (Novabiochem) was swollen in DCM (2 mL,  $2 \times 15$  min) followed by DMF (2 mL, 15 min). N-Fmoc amino acids (3 equiv) were then coupled directly to resin in DMF in the presence of HCTU (3 equiv) and DIEA (5 equiv). Fmoc-Lys(ivDde)-OH (Novabiochem) was used as the branching unit. The orthogonally protected Lys was selectively deprotected with 20% piperidine, and the first branch was synthesized with an acetyl cap at the N-terminus. Capping was accomplished by treating the Fmoc deprotected terminal amino acid with 1:1 acetic anhydride:DMF for 30 min. The ivDde protecting group was then removed with 2% hydrazine in DMF (2 mL,  $2 \times 5$  min). The second branch was then synthesized, and the final amino acid was left Fmoc protected. The resin was washed with DMF, DCM, and MeOH before being dried and stored in a desiccator. Beads were washed extensively with DMF between reactions, and were tested for completeness via Kaiser test.

#### Coupling with fluorescein-5-isothiocyanate in FITC-peptides

Once peptide synthesis was complete on the Rink amide MBHA resin, the uncapped N-terminus of the branched peptides was Fmoc deprotected and washed extensively with DMF. The peptides were then fluorescently labeled by mixing the resin with 6 equiv of fluorescein-5-isothiocyanate (Sigma) and DIEA (14 equiv) in minimal DMF for 12 h in the dark. The resin was then washed with DMF, DCM, and MeOH. The beads were dried and immediately treated with 94:1:2.5:2.5 TFA/TIS/H<sub>2</sub>O/EDT (v/v).

#### Cleavage of Fmoc protecting groups

N- $\alpha$ -Fmoc protected amino acids were deprotected by mixing the resin in a 20% solution of piperidine in DMF (2 mL, 2 × 10 min). Resin was washed extensively with DMF after filtering off cleaving solution.

Side-chain deprotection and peptide cleavage with 94% TFA, 1% TIS, 2.5%  $H_2O$ , 2.5% EDT

The amino acid side-chains were deprotected by a 3-h treatment with 94:1:2.5:2.5 TFA/TIS/H<sub>2</sub>O/EDT (v/v). After deprotection, resin was washed extensively with DMF, DCM, and MeOH before drying and storing it at -20 °C protected from the light. Treatment of Rink amide MBHA resin with the 94% TFA cocktail also afforded cleavage of the peptide when preparing the FITC-peptides. After peptide cleavage, TFA was removed under reduced pressure. The resulting yellow peptide was washed several times with cold diethyl ether and was finally dried to a yellow powder overnight under nitrogen flow. Dried peptide was stored at -20 °C in the dark until it was purified.

## DY557 fluorescence intensity over auto-fluorescence of Tentagel Macrobead-NH<sub>2</sub> resin

Due to the auto-fluorescence displayed by the Tentagel Macrobead-NH<sub>2</sub> resin, the ability to visualize the fluorescence of DY547 labeled HIV-1 TAR RNA (5'- DY547-GCC-CGA-UUU-GAG-CCU-GGG-AGC-UCU-CGG-GC-3', Dharmacon) needed to be ensured. The control branched peptide (KYR)<sub>2</sub>\*FDS was incubated in 1  $\mu$ M DY547 labeled HIV-1 TAR RNA for 1 h in phosphate buffer (0.2 M sodium phosphate, pH 7.45 + 10 mM NaCl). These beads were then washed extensively with the same buffer before

examining them in a sterile 96 well plate (Nunc), with a Zeiss LSM 510 microscope set to longpass 585. The fluorescence intensity of these RNA treated beads was compared with the auto-fluorescence of untreated control branched peptide resin, and the detector sensitivity was adjusted to remove auto-fluorescence of the resin as shown in Figure 3.6. This study showed that resin bound with DY547 labeled HIV-1 TAR RNA could be distinguished from the auto-fluorescence of the resin.



**Figure 3.6.** DY547 fluorescence intensity over auto-fluorescence of Tentagel Macrobead-NH<sub>2</sub> resin. A: Control branched peptide  $(KYR)_2*FDS$  without treatment with 1 µM DY547 labeled HIV-1 TAR RNA. B: Control branched peptide  $(KYR)_2*FDS$  with treatment of 1 µM DY547 labeled HIV-1 TAR RNA for 1 h in phosphate buffer.

## Organic washing conditions to remove bound RNA

Control branched peptide with bound DY547 HIV-1 TAR RNA (shown in Figure 6B) was washed with DMF ( $5\times$ ) and MeOH ( $5\times$ ) to rinse away the fluorescent RNA. The washed resin was then examined by confocal microscopy in a 96 well plate. Reduced fluorescence was observed after this procedure as is seen in Figure 3.7. This procedure was also used to remove RNA from screened hits in the 3.3.3 library prior to the second round of screening, and prior to photocleavage of branched peptides from the resin.



**Figure 3.7.** Organic washed control branched peptide  $(KYR)_2*FDS$ . A: Washed resin shown as transmitted light + fluorescence image. B: Washed resin shown as fluorescence image only.

## Optimizing screening conditions with control branched peptide (KYR)<sub>2</sub>\*FDS

Screening conditions were optimized using the control branched peptide (KYR)<sub>2</sub>\*FDS. About 2000 beads were divided evenly among 6 non-stick 1.5 mL microfuge tubes. All incubations were performed in 300 µL of TK buffer (50 mM Tris•HCl, pH 7.4/ 20 mM KCl/ 0.1% Triton X-100) at 4 °C. A Barnstead/Thermolyne Labquake rotisserie shaker was used to agitate the incubation mixtures. The first set of 3 controls was done in two incubation steps. The first incubation was 3 h and the components are shown in Table 3.2. The beads were then washed 3 times with TK buffer before being incubated in 3.4 nM DY547 labeled HIV-1 TAR RNA for 3 h. The second set of 3 controls was incubated for 6 h at 4 °C with the components shown in Table 3.3. Alpha-synuclein mRNA was prepared by standard cloning and transcription procedures. Bovine serum albumin (BSA) was obtained from New England BioLabs Inc.

Table 3.2. Conditions of first incubations for controls A	<b>A-C</b>
---	------------

First Incubation Components	Control-A	Control-B	Control-C
1 mg/mL BSA	Х		X
0.5 mM alpha-synuclein mRNA		Х	Х

Incubation Components	Control-D	Control-E	Control-F
1 mg/mL BSA	X		Х
0.5 mM alpha-synuclein mRNA		Х	Х
3.4 nM HIV-1 TAR RNA	X	X	X

Table 3.3. Conditions for incubations of controls D–E.

After the final incubation, a Zeiss LSM 510 microscope set to longpass 585 was used to image the beads. The most stringent conditions were found to be when the resin was incubated simultaneously with DY547 HIV-1 TAR RNA, BSA, and alpha-synuclein mRNA (Figure 3.8).



**Figure 3.8.** Optimizing screening conditions with control branched peptide. A1–F1: Images of beads from different incubation conditions. A2–F2: Same as A1–F1 with detector sensitivity adjusted for better visualization.

#### Screening of the 3.3.3 library for binding to HIV-1 TAR

Initially, ~2000 beads by volume of the 3.3.3 library were taken into a 1.5  $\mu$ L non-stick microfuge tube. The beads were allowed to mix for 3 h with 1 mg/mL BSA and 0.5  $\mu$ M alpha-synuclein mRNA in 500  $\mu$ L of TK buffer at 4 °C. After the first incubation, the beads were washed (5×) with TK buffer and were then incubated for 3 h at 4 °C in 500  $\mu$ L of 3.4 nM DY547 HIV-1 TAR mRNA in TK buffer. The beads were then washed (5×) with TK buffer before being evenly distributed among the wells of a 96 well plate and analyzing them by confocal microscopy. This initial screening procedure yielded 6 beads with elevated fluorescence over background (Figure 3.9A–E). These first 6 hits were rinsed with the organic wash and stored at 4 °C until they were photocleaved and sequenced by MALDI-TOF.

When screening the remaining beads of the 3.3.3 library, more stringent conditions were used in an attempt to reduce the total number of hit beads. The first incubation proceeded as done previously, but the second incubation time was limited to 1 h. This procedure yielded an additional 10 beads with elevated fluorescence (Figure 3.9F–N). These beads were rinsed with the organic wash and were then re-screened under even more stringent conditions.

The re-screening procedure was comprised of two incubations. The 10 hits were first mixed for 3 h in 1 mg/mL BSA and 0.5  $\mu$ M alpha-synuclein mRNA in 300  $\mu$ L of TK buffer at 4 °C. The beads were then extensively washed with TK buffer before incubating them a second time for 30 min in 300  $\mu$ L of TK buffer with 3.4 nM DY547 HIV-1 TAR mRNA at 4 °C. The beads were extensively rinsed with TK buffer prior to imaging them in a 96 well plate with the confocal microscope (Figure 3.10). The rescreened beads were finally treated by the organic wash to remove any bound RNA prior to photocleavage and MALDI-TOF sequencing.



**Figure 3.9.** The 16 hits screened from the 3.3.3 library. A–E: Initial 6 hits. F–N: Additional 10 hits found with more stringent screening conditions.



Figure 3.10. The 10 hits from the 3.3.3 library, which were re-screened under stringent screening conditions.

#### Photocleavage of peptides from the ANP resin and MALDI-TOF sequencing

The 3.3.3 library peptides and control peptides were released from the resin by photocleavage of the ANP photolinker. Beads selected for photocleavage were taken up individually into clear non-stick 0.5 mL microfuge tubes with 20  $\mu$ L of 1:1 MeOH:H<sub>2</sub>O.

The microfuge tubes containing resin were placed into a foil-lined container and were irradiated for 1 h with light at 365 nm in using a 4 W handheld UV lamp. The resin was immediately removed from the supernatant after photocleavage. The peptide sequences were obtained from MALDI-TOF MSMS.

Synthesis, HPLC purification and MALDI-TOF identification of FITC-peptides (BP15– BP17, T15)

Branched peptides 15–17 were resynthesized as described above with a single FITC label to be used in fluorescence polarization experiments. A linear peptide, T15 RRAGVRD, was also synthesized using Rink Amide MBHA resin utilizing standard Fmoc chemistry. T15 was also coupled to FITC as described above. Each synthesized FITC-peptide was purified by HPLC. Fractions containing the peptide of interest were identified by MALDI-TOF MS.

#### MALDI/MS

FITC-BP15: FITC-(RRA)<sub>2</sub>\*VRD; Calculated: 1712.83, Found: 1713.86 FITC-BP16: FITC-(RRL)<sub>2</sub>\*HYL; Calculated: 1840.93, Found: 1841.07 FITC-BP17: FITC-(RRL)<sub>2</sub>\*HRF; Calculated: 1867.95, Found: 1868.99 T15: FITC-RRAGVRD; Calculated: 1216.52, Found: 1217.70

#### Fluorescence polarization experiments

Fluorescence polarization experiments were conducted to obtain  $K_d$  values using the FITC-peptides that had been HPLC purified and identified by MALDI-TOF MS.

Experiments were conducted in black, flat-bottom 384 well plates (Corning) and an Analyst AD (LJL BioSystems). The FITC-peptides (0.5 nM) were individually titrated with HIV-1 TAR RNA (up to 100  $\mu$ M) used during the on-bead assay or mutTAR (U24>C, synthesized in-house using a Mermade IV oligonucleotide synthesizer) at 25 °C in a buffer containing 20 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES), 1 mM MgCl<sub>2</sub>, 5 mM KCl, and 140 mM NaCl at pH 7.4. Samples were excited at 485 nm and emission was monitored at 530 nm. Each data point was an average of 5 scans, and all fluorescence polarization experiments were conducted in triplicate.

## References

1. Draper, D. E. Protein-RNA recognition. Annu. Rev. Biochem. 1995, 64, 593-620.

2. Schroeder, R.; Barta, A.; Semrad, K. Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 908-919.

3. Hermann, T. Drugs targeting the ribosome. *Curr. Opin. Struct. Biol.* **2005**, *15*, 355-366.

4. Thomas, J. R.; Hergenrother, P. J. Targeting RNA with small molecules. *Chem. Rev.* **2008**, *108*, 1171-1224.

5. Aboul-ela, F.; Karn, J.; Varani, G. The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* **1995**, *253*, 313-332.

6. Rana, T. M.; Jeang, K. T. Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA. *Arch. Biochem. Biophys.* **1999**, *365*, 175-185.

7. Mei, H. Y.; Mack, D. P.; Galan, A. A.; Halim, N. S.; Heldsinger, A.; Loo, J. A.; Moreland, D. W.; Sannes-Lowery, K. A.; Sharmeen, L.; Truong, H. N.; Czarnik, A. W. Discovery of selective, small-molecule inhibitors of RNA complexes–I. The Tat protein/TAR RNA complexes required for HIV-1 transcription. *Bioorg. Med. Chem.* **1997**, *5*, 1173-1184.

8. Tao, J.; Frankel, A. D. Specific binding of arginine to TAR RNA. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2723-2726.

9. Hwang, S.; Tamilarasu, N.; Ryan, K.; Huq, I.; Richter, S.; Still, W. C.; Rana, T. M. Inhibition of gene expression in human cells through small molecule-RNA interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 12997-13002.

10. Athanassiou, Z.; Patora, K.; Dias, R. L.; Moehle, K.; Robinson, J. A.; Varani, G. Structure-guided peptidomimetic design leads to nanomolar beta-hairpin inhibitors of the Tat-TAR interaction of bovine immunodeficiency virus. *Biochemistry* **2007**, *46*, 741-751.

11. Davis, B.; Afshar, M.; Varani, G.; Murchie, A. I.; Karn, J.; Lentzen, G.; Drysdale, M.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T.; Aboul-ela, F. Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *J. Mol. Biol.* **2004**, *336*, 343-356.

12. Pang, R.; Zhang, C.; Yuan, D.; Yang, M. Design and SAR of new substituted purines bearing aryl groups at N9 position as HIV-1 Tat-TAR interaction inhibitors. *Bioorg. Med. Chem.* **2008**, *16*, 8178-8186.

13. Froeyen, M.; Herdewijn, P. RNA as a target for drug design, the example of Tat-TAR interaction. *Curr. Top. Med. Chem.* **2002**, *2*, 1123-1145.

14. Tam, J. P. Synthetic peptide vaccine design: Synthesis and properties of a highdensity multiple antigenic peptide system. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5409-5413.

15. Singh, D.; Kiarash, R.; Kawamura, K.; LaCasse, E. C.; Gariepy, J. Penetration and intracellular routing of nucleus-directed peptide-based shuttles (loligomers) in eukaryotic cells. *Biochemistry* **1998**, *37*, 5798-5809.

16. Pini, A.; Falciani, C.; Bracci, L. Branched peptides as therapeutics. *Curr. Protein Pept. Sci.* **2008**, *9*, 468-477.

17. Kitov, P. I.; Bundle, D. R. On the nature of the multivalency effect: A thermodynamic model. *J. Am. Chem. Soc.* **2003**, *125*, 16271-16284.

18. McNaughton, B. R.; Gareiss, P. C.; Miller, B. L. Identification of a selective smallmolecule ligand for HIV-1 frameshift-inducing stem-loop RNA from an 11,325 member resin bound dynamic combinatorial library. *J. Am. Chem. Soc.* **2007**, *129*, 11306-11307.

19. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res.* **1991**, *37*, 487-493.

20. Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, *354*, 82-84.
21. Lescrinier, T.; Hendrix, C.; Kerremans, L.; Rozenski, J.; Link, A.; Samyn, B.; Aerschot, A. V.; Lescrinier, E.; Eritja, R.; Beeumen, J. V.; Herdewijn, P. DNA-binding ligands from peptide libraries containing unnatural amino acids. *Chem. Eur. J.* **1998**, *4*, 425-433.

22. Brown, B. B.; Wagner, D. S.; Geysen, H. M. A single-bead decode strategy using electrospray ionization mass spectrometry and a new photolabile linker: 3-amino-3-(2-nitrophenyl)propionic acid. *Mol. Divers.* **1995**, *1*, 4-12.

23. Churcher, M. J.; Lamont, C.; Hamy, F.; Dingwall, C.; Green, S. M.; Lowe, A. D.; Butler, J. G.; Gait, M. J.; Karn, J. High affinity binding of tar rna by the human immunodeficiency virus type-1 Tat protein requires base-pairs in the RNA stem and amino acid residues flanking the basic region. *J. Mol. Biol.* **1993**, *230*, 90-110.

24. Schneider, T. D.; Stephens, R. M. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **1990**, *18*, 6097-6100.

25. Tan, D. S.; Foley, M. A.; Stockwell, B. R.; Shair, M. D.; Schreiber, S. L. Synthesis and preliminary evaluation of a library of polycyclic small molecules for use in chemical genetic assays. *J. Am. Chem. Soc.* **1999**, *121*, 9073-9087.

# Chapter 4 Toward Targeting RNA Structure: Branched Peptides as Cell-Permeable Ligands to TAR RNA

# Attributions

This chapter was taken from Bryson, D. I.; Zhang, W.; McLendon, P. M.; Reineke, T. M.; Santos, W. L. Toward targeting RNA structure: Branched peptides as cell-permeable ligands to TAR RNA. ACS Chem. Biol. 2012, 7, 210-217. Most of the work described in this chapter was performed by the author of this dissertation. He was solely responsible for all RNA design and preparation, all electrophoretic mobility shift experiments, interpretation of all binding affinity data, Hill plot analyses, and liquid chromatography coupled with electrospray ionization mass spectrometry. The author also performed the majority of the dot blot experiments. Wenyu Zhang of the Santos Group aided in the preparation and purification of several peptide compounds and performed some of the dot blot assays. Patrick M. McLendon of the Reineke Group performed all cellular assays. The author of this dissertation significantly contributed to the writing and editing of the manuscript. The final manuscript was prepared by Dr. [Reprinted with permission from Bryson, D. I.; Zhang, W.; Webster L. Santos. McLendon, P. M.; Reineke, T. M.; Santos, W. L. Toward targeting RNA structure: Branched peptides as cell-permeable ligands to TAR RNA. ACS Chem. Biol. 2012, 7, 210-221. Copyright 2012.]

## Abstract

Rational design of RNA ligands continues to be a formidable challenge, but the potentially powerful applications in biology and medicine catapults it to the forefront of chemical research. Indeed, small molecule and macromolecular intervention are attractive approaches, but selectivity and cell permeability can be a hurdle. An alternative strategy is to use molecules of intermediate molecular weight that possess large enough surface area to maximize interaction with the RNA structure but are small enough to be cell permeable. Herein, we report the discovery of non-toxic and cell permeable branched peptide (BP) ligands that bind to TAR RNA in the low micromolar range from on-bead high throughput screening of 4,096 compounds. TAR is a short RNA motif in the 5'-UTR of HIV-1 that is responsible for efficient generation of full RNA transcripts. We demonstrate that BPs are selective for the native TAR RNA structure and that 'branching' in peptides provides multivalent interaction, which increases binding affinity to RNA.

## 4.1 Introduction

RNA-protein interactions, along with protein-protein interactions, control many functions in a living cell such as transcription, splicing, replication, transport and catalysis. Due to the ubiquity of RNA-mediated biological processes, molecules that can selectively bind and regulate the function of RNA have enormous potential application in biotechnology and therapeutics. Despite the considerable effort in utilizing RNA as a drug target, the discovery of molecules with desirable drug-like properties remains challenging and continues to be a subject of intense investigation.<sup>1</sup> One of the main

issues for RNA targeting is the involvement of large surface area for recognition and tight binding—outcompeting the endogenous protein partner using a small molecule is a herculean task. With RNA, the problem is exacerbated by the conformational dynamics often resulting in structural heterogeneity making it difficult for the *de novo* design of RNA ligands.<sup>2</sup>

RNA is often characterized by a variety of secondary structures, including hairpins, bulges, stems, loops, pseudoknots, and turns. The folding of these local structures can give rise to tertiary structures that are unique to specific RNA constructs and potentially allow the RNA to be targeted selectively.<sup>3</sup> The unique three dimensional architectures present in RNA make it possible to target at a level that is not presented by DNA and in a manner that is not solely dependent on Watson-Crick base pairing. For example, the ability of small molecules to target the tertiary structures of ribosomal RNA (rRNA) has been well demonstrated by aminoglycoside, macrolide, oxazolidinone, and tetracycline antibiotics.<sup>4-9</sup> However, the development of RNA-binding small molecules is far from a straightforward process, where poor selectivity is a common hurdle to overcome, and a general RNA-targeting paradigm remains an elusive goal.<sup>1</sup> The most desired approach of rational, structure-based RNA-targeted drug discovery is still in its infancy, making the design of RNA ligands difficult.<sup>10</sup> RNA-targeted gene repression can be attained using antisense or RNA interference technologies.<sup>11</sup> However, despite significant efforts from academia and the pharmaceutical industry, drug characteristics such as cellular delivery, stability and off-target effects remain a challenge, although several siRNAs are in clinical trials.<sup>11,12</sup>

The human immunodeficiency virus-1 (HIV-1) transactivation response element (TAR) is one of the most studied RNA targets because of its key role in viral replication.<sup>13</sup> HIV-1 TAR is characterized by a 59-nucleotide stem-bulge-loop secondary structure, which is found at the 5'- end of all nascent HIV-1 transcripts and is a highly conserved region of the virus.<sup>14,15</sup> The region from +19 to +42 bases of HIV-1 TAR is characterized by a hexanucleotide loop at the end of a helical stem containing a single, trinucleotide pyrimidine bulge. HIV-1 TAR RNA is the target of the 101-amino acid Tat protein, which is the virally encoded *trans*-activator of HIV-1 transcription.<sup>14</sup> When the Tat protein is not associated with HIV-1 TAR, basal viral transcription is very low, and short RNA transcripts are generated.<sup>16</sup> When the Tat–TAR RNA complex is formed, cofactors such as cyclin T1 and its cognate kinase CDK9 stimulate efficient transcription from the long terminal repeat (LTR).<sup>15</sup> Blocking this Tat–TAR interaction is therefore a potent strategy for controlling the proliferation of the virus and provides a potential anti-HIV therapy.

Given the global problem of HIV-1 infection, drugs that exhibit novel modes of action are of the utmost importance since resistance to current HIV therapies has been observed.<sup>17</sup> Indeed, the inhibition of Tat–TAR interaction can fill this gap. Notwithstanding the significant efforts committed to discovering compounds capable of disrupting the function of TAR, no drug has made it to the clinic. Several strategies have been reported to inhibit the Tat–TAR interaction. Although attractive, small molecules such as aminoglycosides, argininamide, purine analogs, bis-guanidine compounds, tripeptides and others appear to suffer from selectivity issues.<sup>1,18-23</sup> Large macromolecules such as TAR RNA decoys<sup>24,25</sup> are currently being investigated, but

compounds of intermediate size (oligomeric amines<sup>26</sup> and  $\beta$ -hairpin peptidomimetics<sup>27,28</sup>) are subject to increased attention in part because of better overlap with the large surface area of RNA. In particular, cyclic peptide L50 has recently been shown to have anti-viral activity by inhibiting the Tat dependent transcription process and the reverse transcription step.<sup>29</sup>

While interesting leads are emerging from prior studies, molecular scaffolds that can recognize RNA structures still need to be developed. In an effort to provide a general platform to targeting RNA structures, we focused on inhibiting RNA-protein interactions involving Tat and TAR RNA (vide supra). Recently, we disclosed an approach that uses branched peptides (BPs) because of the potential for multivalent interaction on various regions of TAR-a desirable property that can increase selectivity and affinity.<sup>30</sup> Because the TAR structure is expected to be a dynamic ensemble of many conformations in solution, high throughput screening of a library of BPs allows the highly populated TAR conformations to selectively bind compounds with diverse chemical architecture projected on beads. We reason that BPs are excellent candidates because natural and unnatural amino acids contain a wide array of functional groups that can display unique molecular architectures and their synthetic accessibility is straightforward. Indeed, rapid generation of a large number of peptides on beads using the split and pool technique can provide access to libraries of varying molecular weight that can be fine-tuned to increase pharmacological properties such as cell permeability. Herein, we report the discovery of BPs of intermediate molecular weight binding to TAR RNA in the low micromolar regime. In addition, we show that 'branching' in peptides can lead to multivalent interactions that increase affinity to TAR. Further, we demonstrate that BPs are cell

permeable and non-toxic making them excellent chemical biology tools for targeting well-defined RNA structures and useful in anti-HIV drug discovery.

# 4.2 **Results and Discussion**

## Design of branched peptides and on-bead high throughput screening

We recently reported the synthesis of a 4,096-membered BP library<sup>30</sup> on Tentagel beads by split and pool synthesis.<sup>20,31-34</sup> The amino acid composition of the library was biased to increase intermolecular interaction with RNA. For example in position 1, amino acids with functional groups that can induce attractive forces such as electrostatics (the positively charged Arg interacting with the negatively charged phosphate backbone),  $\pi$ - $\pi$  interactions (Tyr) and hydrogen bonding (Asp, His) were introduced (Figure 4.1). Presentation of diverse epitopes on the bead surface would allow DY547-labeled TAR RNA to sample all possible binding modes. Binding of the RNA to the bead resulted in increased fluorescence, which was monitored by confocal microscopy. The BPs in the screened library featured two identical N-termini branched by a Lys residue and a single C-terminus, which was linked to the bead via a photocleavable linker (3-amino-3-(2nitrophenyl)propionic acid, ANP).<sup>35</sup> During the incubation, nonspecific binding was minimized with the addition of excess unlabeled alpha-synuclein mRNA (340-nt) and bovine serum albumin (BSA) in the incubation buffer. Seventeen beads were selected as putative hits, photocleaved and identified by matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) de novo sequencing.<sup>36</sup>



**Figure 4.1**. High-throughput screening of a branched peptide library with TAR RNA reveals peptide hits that are sequenced by MALDI-TOF.

## Synthesis of Fluorescein-labeled Branched Peptides and Truncated Variants

The BPs were synthesized following standard solid-phase peptide synthesis techniques using Rink Amide MBHA resin. Each BP was prepared such that a single Nterminus was labeled with fluorescein using fluorescein isothiocyanate (FITC, Scheme 4.1). The branching unit was attached to the  $\varepsilon$ -nitrogen of Lys that was orthogonally protected with ivDde; therefore, although the two branches were the same sequence, the spacing was different. Over the course of this study, we found that directly coupling the N-terminal amino acid to FITC resulted in poor isolated yields after HPLC purification. Upon further analysis of the crude peptides, we observed a strong signal in the MALDI-TOF mass spectrum corresponding to truncated peptide, where fluorescein and the adjacent amino acid were cleaved. This result suggested that acid mediated formation of fluorescein thiohydantoins was occurring upon cleavage from solid-support resin with trifluoroacetic acid (TFA).<sup>37</sup> Installation of aminohexanoic acid (Ahx) as spacer between the N-terminal amino acid and fluorescein resolved the problematic autocleavage and resulted in increased isolated yields. Further, one particular peptide, FL15, epimerized resulting in two diastereomers of equal intensity that were separated by HPLC and confirmed by LC-ESI/MS (Figure 4.2A). **FL15** was finally prepared as the single diastereomer by using [ethyl cyano(hydroxyimino)acetato-O<sup>2</sup>]tri-1-pyrrolidinylphosphonium hexafluorophosphate (PyOxim) as coupling reagent (Figure 4.2B).



Scheme 4.1. Synthesis of fluorescein-labeled branched peptides by solid-phase peptide synthesis (SPPS).

A.



Figure 4.2. LC/ESI-MS of FL15. A) FL15 synthesized using HCTU as peptide coupling agent. B) FL15 synthesized using PyOxim as peptide coupling agent.

## Binding Affinities of Hit Branched Peptides to HIV-1 TAR

Consensus sequence analysis of BP hits from the HTS revealed seventeen peptide sequences that appear to have a preference for Arg moieties. This result is not surprising due to the strong electrostatic attraction between the positively charged side chain of the peptides and the negatively charged phosphate backbone of the RNA target. To validate binding of BP hits to TAR RNA and eliminate false positives, dissociation constants ( $K_d$ values) were measured using dot blot assay (Table 4.1 & Figure 4.3). Our results indicate binding affinities in the low micromolar range, with one peptide hit (**FL10**) determined as a false positive. In particular, FL4, FL6 and FL7 contained four positively charged Arg residues in the N-termini and had  $K_d$  values of 600 nM, 1.2  $\mu$ M and 1.2  $\mu$ M, respectively. Gratifyingly, **FL4** bound with similar affinity to its native protein counterpart, Tat ( $K_d$  = 780 nM). The sequence of **FL4** ([RRW]<sub>2</sub>\*HAL) was different from **FL7** in that the last 2 amino acid residues were changed to YD—the slight decrease (2 fold) in binding affinity might be due to an anionic Asp residue that imparts repulsive interaction with the phosphate backbone of TAR. Surprisingly, the additional Arg moiety in FL6 did not result in increased binding affinity, but suggested that a putative electrostatic interaction with a positive charge could be compensated with other modes of interaction. FL12 was nearly identical in sequence with **FL4** except for the Leu $\rightarrow$ Ser change; the switch from hydrophobic to hydrophilic group resulted in a 3-fold decrease in binding affinity. In general, however, a hydrophobic moiety was desired in amino acid position 6 since hydrophobic groups in this position tended to have higher binding affinity for TAR (Table 4.1).

Entry	Peptide	Sequence	$K_{\rm d}(\mu{\rm M})$ MW (g mol <sup>-1</sup> )	
1	FL4	(RRW)2*HAL	$0.6 \pm 0.1$	1463.74
2	FL6	(RRY) <sub>2</sub> *VRL	$1.2\pm0.2$	1464.77
3	FL7	(RRW) <sub>2</sub> *HYD	$1.2\pm0.2$	1557.77
4	FL12	(RRW) <sub>2</sub> *HAS	$1.9 \pm 0.3$	1437.66
5	FL5	(RRL) <sub>2</sub> *NRF	$2.0 \pm 0.3$	1413.72
6	FL1	(RRL) <sub>2</sub> *WYL	$2.5\pm0.3^{\mathrm{a}}$	1458.80
7	FL17	(RRL) <sub>2</sub> *HRF	$3.1 \pm 0.3$	1436.76
8	FL9	(RRA) <sub>2</sub> *NYF	3.7 ± 1.9	1336.55
9	FL2	(YRA) <sub>2</sub> *HRF	$7.5\pm0.9$	1366.58
10	FL16	(RRL) <sub>2</sub> *HYL	$7.7 \pm 1.1$	1409.73
11	FL8	(YRL) <sub>2</sub> *WRL	$7.8 \pm 1.9^{a}$	1465.79
12	FL11	(RRA) <sub>2</sub> *VYF	8.1 ± 1.3	1321.58
13	FL3	(HRW) <sub>2</sub> *WAS	$8.7\pm3.1^{a}$	1448.64
14	FL14	(RRY) <sub>2</sub> *VQL	$20.2\pm2.9$	1436.71
15	FL10	(DNL) <sub>2</sub> *HYF	NB <sup>a</sup>	1277.38
16	FL13	(RRY) <sub>2</sub> *NQD	> 75	1453.61
17	FL15	(RRA) <sub>2</sub> *VRD	> 75	1282.51
18	T4-1	RRWGHAL	> 75	
19 <sup>38</sup>	Tat <sub>47-57</sub>	GRKKRRQRRR	0.78	

Table 4.1. Binding constants and molecular weights of hit BPs.

<sup>a</sup>Performed with a final concentration of 5% DMSO. No binding observed (NB). \* = Lysine branching unit. Each value is an average of at least three experiments.



**Figure 4.3.** Dot blot assays of labeled peptides and HIV-1 TAR RNA. Assays were run in triplicate with 400 pM  $^{32}$ P-labeled HIV-1 TAR RNA and increasing peptide concentration. From left to right: RNA only, 0.001  $\mu$ M, 0.01  $\mu$ M, 0.03  $\mu$ M, 0.1  $\mu$ M, 0.3  $\mu$ M, 10  $\mu$ M, 30  $\mu$ M, 100  $\mu$ M. **BP15** is a non-FITC labeled peptide.

From the pool of seventeen BPs, three peptides (FL2, FL3 & FL8) did not contain the consensus Arg-Arg in the N-termini. These peptides had similar dissociation constants at around 8  $\mu$ M. Compounding the elevation in  $K_d$  relative to **FL4** was the corresponding decrease in solubility in aqueous condition—DMSO was added to the buffer used in the dot blot assay—that may contribute to decrease in affinity. Finally, it is noteworthy that the electrostatic interaction mediated by the two Arg moieties in the N-termini did not contribute to the majority of the binding affinity with TAR since **FL13** and **FL15** had  $K_d$  values > 75  $\mu$ M.

## The Branched N-terminus Imparts Multivalency

To determine whether the additional "branch" in our peptides results in complementary increase of binding affinity as a consequence of multivalency, we synthesized **T4-1** as a truncated variant of the strongest binder, **FL4**. **T4-1** (RRWGHAL) featured a single N-terminus and a C-terminus that were identical to **FL4**. We substituted the branching Lys unit with Gly in **T4-1** to preserve the spacing between the N- and Ctermini, and to avoid incorporating any additional functionality compared to **FL4**.

To our delight, we observed a > 125-fold decrease in binding affinity for **T4-1** ( $K_d$  > 75  $\mu$ M) compared to **FL4** when measured by dot blot assay (Table 4.1). While it is tempting to draw a conclusion that **T4-1** displayed such a dramatic loss of affinity because two positively charged side-chains were omitted, the presence of Arg-Arg in the N-termini did not necessarily result in effective RNA binding (*vide supra*). We also note that the decrease in binding affinity was not a result of the decrease in the number of positive charges in the molecule because BPs with five positive charges (**FL5**, **FL6**, **FL15**, and **FL17**) had lower  $K_d$  values. These results suggest that electrostatic interactions were not solely responsible for high affinity binders, but that all branches of

the 3.3.3 peptide were responsible for tight binding with the RNA. This conclusion is intriguing because a central problem with designing RNA binding molecules is that inclusion of multiple positive charges generally increases affinity, but a significant price is paid in the reduced selectivity of such compounds.<sup>26</sup> Thus, it stands to reason that the incorporation of diverse binding interactions in RNA targeting molecules will prove to be an important feature when designing effective binders that attain high selectivity.

## Selectivity Determination and Identification of Requisite TAR Binding Elements

In order to interrogate the selectivity of **FL4**, dot blots were performed in the presence of excess tRNA and a series of modified TAR RNA structures. First, we confirmed that the dissociation constants determined from the dot blot assay were reliable by performing electrophoretic mobility shift assay (EMSA) with **FL4**. As shown in Figure 4, the dissociation constants determined through EMSA were in excellent agreement with the dot blot data ( $0.5 \pm 0.1 \mu$ M and  $0.6 \pm 0.1 \mu$ M, respectively). Consistent with previous reports using EMSA, we also noticed non-selective binding to TAR RNA in the presence and absence of competitor tRNA (Figure 4.5).<sup>26,39</sup> Addition of 10- and 1000-fold excess of tRNA caused a shift in the binding affinity to <sup>32</sup>P-labeled TAR RNA suggesting that **FL4** was partially selective.



Figure 4.4. Titration curves comparing EMSA and dot blot assay with or without competing tRNA.



**Figure 4.5.** Competition of **FL4** with  $10 \times$  and  $1000 \times$  tRNA using EMSA used to generate binding curves observed in Figure 4.4. Peptide concentrations increase from left to right: 0.001  $\mu$ M, 0.01  $\mu$ M, 0.03  $\mu$ M, 0.1  $\mu$ M, 0.3  $\mu$ M, 1  $\mu$ M, 3  $\mu$ M, 10  $\mu$ M, 30  $\mu$ M, 100  $\mu$ M.

In order to determine the selectivity of **FL4** with the native TAR RNA structure, we synthesized TAR sequences containing a point mutation (24U>C), bulgeless TAR, tetraloop TAR, and bulgeless tetraloop TAR (Figure 4.6A). The measured  $K_d$  using TAR (24U>C) on the 3-nt bulge, a site where Tat is known to bind, was  $0.8 \pm 0.2 \mu$ M (Figure 4.6B & Figure 4.7).



**Figure 4.6**. Sequence and secondary structure of TAR RNA and variants (A) and titration curves of **FL4** with these RNAs (B).



**Figure 4.7.** Dot blot assays of **FL4** with TAR variants used to obtain binding curves in Figure 4.6. A) TAR(24U>C), B) Bulgeless TAR, C) Tetraloop TAR, and D) Bulgeless Tetraloop TAR.

This value is within experimental error of the native TAR  $K_d$  in this assay and suggests that **FL4** may not directly interact with the nucleobase in U or C. Indeed, TAR (24U>C) is a mutation that is present in several clades of HIV-1 and is expected to retain the native structure of TAR.<sup>2640</sup> In contrast, a more dramatic increase in  $K_d$  was observed when the TAR RNA tertiary structure was modified as a consequence of removing the 3-nt bulge region (bulgeless TAR) or decreasing the size of the apical loop region by 2-nt (tetraloop TAR). The resulting  $K_d$  for these TAR variants were  $6.6 \pm 1.3 \mu$ M and  $5.5 \pm 2.0 \mu$ M, respectively (Figure 4.7B). The approximately ten-fold decrease in binding affinity is exciting because it suggests that **FL4** is selective for the three dimensional structure of native TAR RNA and that **FL4** may interact with both of these structural elements when bound to the native TAR RNA.

We anticipated that modifying both structural elements simultaneously (bulgeless tetraloop TAR) would further decrease the binding affinity if **FL4** indeed spans the bulge and apical loop of TAR RNA. However, implementing both modifications resulted in a  $K_d$  value of 3.5 ± 1.0  $\mu$ M, which was within error of tetraloop TAR. It is possible that

the additional decrease in affinity is not observed with bulgeless tetraloop TAR because the tertiary structure of these RNAs is sufficiently altered to preclude **FL4** from binding specifically. The Hill analyses of our dot blot data for native TAR RNA and TAR (24U>C) yield Hill coefficients (n) of 1.4 and 1.2 respectively (Figure 4.8) suggesting non-cooperative binding of **FL4**, where n is near 1. This supports our hypothesis of a single binding site that spans the bulge and apical loop, as the binding affinity is clearly decreased when these structural elements are removed individually. Furthermore, cooperative binding ( $n \ge 1.5$ ) is observed for bulgeless TAR, tetraloop TAR and bulgeless tetraloop TAR, which presented Hill coefficients of 1.5, 2.1, and 1.6 respectively. Taken together, these results suggest that modifications on TAR result in multiple, low-affinity binding sites for **FL4**.



**Figure 4.8.** Hill plots of **FL4** dot blot data with native TAR RNA and TAR RNA variants. A) TAR B) TAR(24U>C), C) Bulgeless TAR, D) Tetraloop TAR, and E) Bulgeless Tetraloop TAR.

# Branched Peptides are cell permeable and exhibit no cytotoxicity

An attractive feature of developing BPs as RNA ligands is the ability to control the molecular weight, a property that can have significant influence on cellular uptake. Although arginine-rich BPs (Arg > 8) have been shown to be cell permeable,<sup>41</sup> the spacing and number of Arg residues present in our BP hits are significantly different than previously reported examples. We biased our investigation to 3.3.3-BP in part because of their medium molecular weight (i.e., larger than 500 Da but less than typical oligonucleotides/peptide nucleic acids), which is around 1400 Da. Based on the hit sequences revealed through HTS, we predicted that our BPs would be cell permeable due to the high content of basic residues and because they were less than 20-amino acids in size.<sup>42-45</sup>

Cellular uptake in HeLa cells was determined upon incubation with FITC-labeled BPs (1 µM) in the culture medium for 4 h at 37 °C. After washing, cells were fixed with 4% paraformaldehyde and imaged using a confocal microscope. Cells incubated with **FL3** contained fluorescence that was diffuse throughout the nucleus and cytoplasm (Figure 4.9A). Punctate structures were also observed in cells incubated with other BPs, wherein BPs were suspected to be internalized via endocytosis and sequestered in vesicles (Figure 4.10). Flow cytometry studies with HeLa cells also provided supporting evidence that BPs were cell permeable (Figure 4.9B). Gratifyingly, the majority of the peptides were internalized by >95% of the counted cells; although **FL9**, **FL15** and **FL16** showed a lesser degree of internalization (~20–55% increased fluorescence compared to background). To our delight, the majority of cells were viable in a (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay suggesting that BPs

that penetrate cells were nontoxic (Figure 4.9C). Overall, medium-sized BPs were cell permeable and demonstrated low cytotoxicity.



**Figure 4.9.** Cell permeability and cytotoxicity of FITC-labeled hit peptides. A) Cellular uptake of branched peptides into HeLa cells, top left: fluorescence image of cells; top right: DAPI staining of the nucleus; bottom left: phase contrast image; bottom right: overlay of all images. White scale bar is 25  $\mu$ m. B) Cell permeability of FITC-labeled branched peptides by flow cytometry in HeLa cells. C). MTT cell toxicity assay.





Figure 4.10. Confocal microscopy of fluorescent BPs in HeLa cells. Column one: fluorescence image of cells; second column: DAPI staining of the nucleus; third column: phase contrast image; fourth column: overlay of all images in the row. T15 (RRAGVRD) is a linear FITC-labeled BP. White ruler bars in the lower corners of overlay images measure  $25 \mu m$ .

In summary, a general platform for generating selective RNA-binding ligands based on multivalent, branched peptides has been developed. In targeting the Tat–TAR RNA interaction, non-toxic and cell permeable compounds binding to the RNA component in the low micromolar regime were discovered. Selective binding of branched peptide **FL4** to the tertiary structure of TAR RNA against other variants was demonstrated, where a single binding site that spans the bulge and apical loop is likely involved. Targeted localization of BP in specialized regions of the cell can be important in realizing their therapeutic potential in various disease states such as HIV-1. Current efforts are aimed at introducing localization signal modules that can deliver BPs into the nucleus. The selective targeting of RNA structures remains a formidable task; the current effort steps towards the realization of a possible solution.

# 4.3 Materials and Methods

#### Peptide Synthesis, Purification, and Characterization

All reagents were purchased from Novabiochem unless stated otherwise. N- $\alpha$ -Fmoc protected L-amino acids (3 equiv.) were used at all steps except when Fmoc-Lys(ivDde)-OH was used as the branching unit in the BPs. All coupling steps were 15 min at room temperature and performed in dimethylformamide (DMF) bubbling with argon. Deprotection of Fmoc was performed with two consecutive 5 min washes with 20% piperidine in DMF. Peptides were synthesized on Rink amide MBHA resin (100-200 mesh) by hand using the previously described apparatus.<sup>30</sup> 2-(6-chloro-1-Hbenzotriazole-1-yl)-1,1,3,3-tetramethylaminiumhexafluorophosphate (HCTU) (Peptides International) was used as coupling reagent (3 equiv.) and N,N-diisopropylethylamine as base (5 equiv.) in all coupling steps except for the preparation of **FL15**, where PyOxim (3) equiv.) was required as the coupling reagent to avoid epimerization in the sequence. Acetic anhydride (3 equiv.) was used to cap the first N-terminus in all BPs using DIEA (5 equiv.) as base. The ivDde protecting group was then removed using 2% hydrazine in DMF, and the second N-terminus was synthesized off of the Lys side-chain in the BPs. Fmoc-6-Ahx-OH (AnaSpec) was installed at the uncapped N-terminus in the samples used in dot blot and EMSA experiments. Fluorescein 5-isothiocyanate (FITC) (Sigma) (5 equiv) was coupled last to the uncapped N-terminus of all peptides using DIEA (14 equiv) as base (3 h). Peptides were protected from the light at all subsequent steps.

Side chains were deprotected and peptides cleaved from resin simultaneously by stirring for 3hin 95:2.5:2.5 (v/v/v)mixture of trifluoroacetic a acid (TFA):water:triisopropylsilane. The supernatant was dried under reduced pressure, and the crude peptide was washed with cold diethyl ether. Peptides were purified to  $\geq 95\%$ purity by HPLC (Agilent 1200 series) using a Jupiter 4 µ Proteo 90 Å semi-prep column (Phenomenex), and a binary solvent gradient composed of 0.1% TFA in MilliQ water and HPLC grade acetonitrile. Purity was determined under the same conditions using a Jupiter 4 µ Proteo 90 Å analytical column (Phenomenex). All peptides were finally characterized by MALDI-TOF MS analysis. Peptide stock concentrations were prepared by spectrophotometry, by monitoring absorbance of FITC at 495 nm,  $\varepsilon = 77,000 \text{ mol}^{-1} \text{cm}^{-1}$  in 100 mM glycine, pH 9.0.

## Cell Culture

All cell culture products, unless otherwise noted, were purchased from Gibco/Invitrogen. HeLa cells (human adenocervical carcinoma) were purchased from American Type Culture Collection (ATCC), and were subcultured once per week. Cells were grown in 75 cm<sup>2</sup> flasks (Corning) in Advanced DMEM, supplemented with 2% heat-inactivated fetal bovine serum (FBS), 1% L-glutamine, 100 units/mg penicillin, 100  $\mu$ g mL<sup>-1</sup> streptomycin, and 0.25  $\mu$ g mL<sup>-1</sup> amphotericin. Cells were plated for internalization in Dulbecco's Modified Eagle's Medium (DMEM) containing GlutaMAX<sup>TM</sup>, supplemented with 10% FBS, 100 units mg<sup>-1</sup> penicillin, 100  $\mu$ g mL<sup>-1</sup> streptomycin, and 0.25  $\mu$ g mL<sup>-1</sup> amphotericin. Cells were determined free of mycoplasma contamination using MycoAlert<sup>TM</sup> Mycoplasma Detection Kit (Lonza).

## Cellular Internalization of Peptides

HeLa cells were plated at  $1 \times 10^5$  cells/well in DMEM containing 10% FBS in 12well tissue culture plates (Corning) and allowed to attach at 37 °C in a humid 5% CO<sub>2</sub> atmosphere for 24 h. After removing media and washing cells with PBS, 600 µL of FITC-labeled peptide in Opti-MEM (1 µM) was added to each well. Cells were incubated with peptides for 4 h, then 1.5 mL DMEM was added and incubated for 30 min. Cells were detached with 500 µL of trypsin-EDTA, quenched with internalization media (1 mL), and the contents of each well were collected into Falcon Tubes (BD Biosciences). Cells were centrifuged at 4 °C and 1250 rpm for 10 min. The supernatant was removed, the cell pellets rinsed with PBS, and centrifuged again at identical conditions. Supernatant was removed and cell pellet was again suspended in 2% FBS in PBS. Propidium iodide (PI; 5 µg mL<sup>-1</sup>, Molecular Probes) was added to each tube 2– 5 min prior to analysis. Cellular uptake of FITC-labeled peptides was measured on a FACS Canto II flow cytometer (BD Biosciences). FITC was excited using a 488 nm solid state laser, and detected at  $530 \pm 30$  nm bandpass filter, and PI was excited using a 488 nm solid state laser and detected with a 670 nm longpass filter. Appropriate gating was done against the untreated cells control to ensure that autofluorescence was not measured as cellular uptake, and only live (PI-negative) cells were included in subsequent analysis. 10,000-20,000 gated events were collected for each sample, and experiments were done in duplicate unless noted otherwise.

## Cellular uptake of peptides into HeLa cells using confocal microscopy

HeLa cells were plated at  $1.5 \times 10^4$  cells/well in 10% FBS-containing DMEM in 12-well tissue culture plates containing a sterile no. 1 PLL-coated glass coverslip. Cells were incubated at 37 °C in a humid 5% CO<sub>2</sub> atmosphere for 48 h to facilitate adherence to the coverslip. Medium was removed and cells rinsed with PBS before adding 600 µL FITC-labeled peptide (5 µM final concentration) in Opti-MEM. Peptides were allowed to internalize for 4 h at 37 °C. Cells were fixed in 4% PFA for 10 min at room temperature, stained with DAPI (600 nM; 150 µl) for 2 min and mounted onto glass slides with Prolong antifade mounting media (11 µl, Invitrogen). Cells were imaged with a 40 × oil immersion objective (N.A. = 1.3) on an LSM510 confocal system fitted onto Axiovert 100 inverted microscope (Zeiss). Minimal modifications to image brightness, contrast, and noise reduction were performed using Image J software (NIH). Each sample was measured against a control of untreated cells using identical acquisition parameters to ensure no reported fluorescence was derived from detector noise visible in the green channel.

#### MTT Toxicity Assay

HeLa cells were plated at 10,000 cells/well in 10% FBS-containing DMEM in 48well tissue culture plates and incubated for 24 h at 37 °C in a humid 5% CO<sub>2</sub> atmosphere. Culture medium was removed and cells rinsed with PBS. Peptides (175  $\mu$ L in Opti-MEM, 1  $\mu$ M) were added to each well and incubated for 4 h. Peptides were removed and cells were rinsed with PBS. DMEM (300  $\mu$ L), containing 10% FBS and 0.5 mg mL<sup>-1</sup> MTT (Invitrogen) was added to each well. Cells were incubated for one additional h before removing MTT-containing media, rinsing with PBS, and dissolving cells and formazan product in 300  $\mu$ L DMSO. Absorbance was measured at 570 nm for each sample, and toxicity was normalized against DMSO only and untreated cells.

# Preparation of <sup>32</sup>P-Labeled RNA

RNA was prep	pared by in vitro transcri	iption with the Ri	ibomax T7 Exp	ress System		
(Promega) using prev	iously reported techniqu	ues. <sup>46</sup> All steps w	vere done using	RNase free		
conditions. The	antisense template	and sense cor	mplementary	strand, 5'-		
ATGTAATACGACT	CACTATAGG (Integr	ated DNA Tech	nnologies), we	re annealed		
prior to transcription	by heating an equimola	r mixture of the	ssDNA strands	in water at		
65 °C for 2 min foll	owed by a 2 min incu	bation in an ice	bath. Antisen	ise template		
sequences used were	e as follows with loca	ations of mutati	on in bold an	d deletions		
underlined:	HIV-1	TA	AR,	5'-		
GCCCGAGAGCTCCCAGGCTCAAATCGGGCCTATAGTGAGTCGTATTACAT;						
TAR(24U>C),				5'-		
GCCCGAGAGCTCCCAGGCTCATATCGGGCCTATAGTGAGTCGTATTACAT;						
bulgeless	r.	ΓAR,		5'-		
GCCCGAGAGCTCC	CAGGCT <u>CT</u> CGGGCC	TATAGTGAGT	CGTATTACA	AT;		
tetraloop		TAR		5'-		
GCCCGAGAG <u>CC</u> GA	A <u>AG</u> CTCAAATCGGG	CCTATAGTGA	GTCGTATTA	CAT;		
bulgeless	tetraloop	TA	AR,	5'-		
GCCCGAGAG <u>CCGAAAG</u> CT <u>CT</u> CGGGCCTATAGTGAGTCGTATTACAT. After						
transcription, the DN	A was degraded by D	Nase. The newl	ly transcribed 1	HIV-1 TAR		

RNA was purified by 20% denaturing PAGE. The topmost band was excised and eluted from the gel overnight in 300 mM sodium acetate, 10 mM Tris•HCl, pH 7.4, and 10 mM EDTA. The sample was desalted using a NAP-25 column (GE Healthcare) and lyophilized to dryness before treating with calf intestinal phosphatase in NEBuffer 3 (New England BioLabs). The dephosphorylated RNA was recovered by phenol extraction followed by ethanol precipitation. Preparation of the full-length transcript was confirmed by MALDI-TOF. Dephosphorylated RNA was stored as a pellet at -80 °C.

HIV-1 TAR RNA was labeled at the 5'-end by treating 100 pmoles of dephosphorylated RNA with 20 nmoles of  $[\gamma^{-32}P]ATP$  (111 TBq mol<sup>-1</sup>) and 20 units of T4 polynucleotide kinase in 70 mM Tris•HCl, pH 7.6, 10 mM MgCl<sub>2</sub>, and 5 mM dithiothreitol. The mixture was incubated at 37 °C for 1 h followed by a 10 min incubation at 65 °C. The labeled RNA was recovered by ethanol precipitation. A 20% denaturing PAGE run at 25W for 2.25 h followed by autoradiography was used to ensure that the RNA sample was  $\geq$  95% pure prior to use in EMSA and dot blot assays. The <sup>32</sup>P-labeled RNA was stored at a concentration of 500 nM in water at -20 °C for up to two months without measurable degradation.

## Dot Blot Assay

Dot blot assays were performed in triplicate using a Whatman Minifold I 96 well Dot Blot system and Whatman 0.45 µm pore size Protran nitrocellulose membranes. Assays were carried out by first refolding 800 pM <sup>32</sup>P-labeled HIV-1 TAR RNA in 2× TK buffer (100 mM Tris•HCl, pH 7.4, and 200 mM KCl) by heating the sample at 95 °C for three min and then allowing it to cool slowly at room temperature for 20 min. Peptides were diluted with water or 10% DMSO from 200  $\mu$ M stocks to prepare each series at 2× concentrations ranging from 0.002  $\mu$ M to 200  $\mu$ M at half-log intervals. 25  $\mu$ L aliquots of refolded <sup>32</sup>P-labeled HIV-1 TAR RNA in 2× TK buffer were added to 25  $\mu$ L aliquots of the peptides at 2× concentration to give the desired final concentration of each component. The mixtures were incubated at room temperature for 20 min prior to filtering each 50  $\mu$ L sample through the nitrocellulose membrane, which had been pre-equilibrated in 1× TK buffer. Two consecutive 50  $\mu$ L washes with 1× TK buffer followed each filtration. Peptide binding was measured by autoradiography using a storage phosphor screen (GE Healthcare), which was imaged on a Typhoon Trio (GE Health Care). Densitometry measurements were taken in ImageQuant TL (Amersham Biosciences). Binding curves were generated using Kaleidagraph (Synergy Software). Error bars represent the standard deviation calculated for three replicates.

#### Electrophoretic Mobility Shift Assay

EMSA were performed in duplicate by first refolding 4 nM <sup>32</sup>P-labeled HIV-1 TAR RNA in 2× TK buffer (100 mM Tris•HCl, pH 7.4, and 200 mM KCl) using the previously described method (*vide supra*). Peptides were diluted with water from 200  $\mu$ M stocks to prepare each series at 2× concentration. Aliquots (10  $\mu$ L) of refolded <sup>32</sup>Plabeled HIV-1 TAR RNA in 2× TK buffer were added to 10  $\mu$ L aliquots of the peptides at 2× concentration to give the desired final concentration of each component. The mixtures were incubated at room temperature for 20 min followed by the addition of 2  $\mu$ L of 30% glycerol for loading. Each sample (15  $\mu$ L) was loaded on to a 10% native PAGE, which had pre-run for 1 h at 300 V. The samples electrophoresed for 1 h at 300 V. Gels were dried to filter paper prior to autoradiography. Data was measured as the percentage of bound RNA in each lane and error bars represent the standard deviation calculated for two replicates. Selectivity studies were performed in the presence of 10-fold or 1000-fold excess by mass of tRNA from *E. Coli* MRE 600 (Roche). The tRNA was refolded separately from HIV-1 TAR RNA prior to use.

## References

1. Thomas, J. R.; Hergenrother, P. J. Targeting RNA with small molecules. *Chem. Rev.* **2008**, *108*, 1171-1224.

2. Lu, J.; Kadakkuzha, B. M.; Zhao, L.; Fan, M.; Qi, X.; Xia, T. Dynamic ensemble view of the conformational landscape of HIV-1 Tar RNA and allosteric recognition. *Biochemistry*. **2011**, *50*, 5042-5057.

3. Zaman, G. J. R.; Michiels, P. J. A.; van Boeckel, C. A. A. Targeting RNA: New opportunities to address drugless targets. *Drug Discov. Today* **2003**, *8*, 297-306.

4. Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **2000**, *407*, 340-348.

5. Tor, Y. The ribosomal A-site as an inspiration for the design of RNA binders. *Biochimie* **2006**, *88*, 1045-1051.

6. Brodersen, D. E.; Clemons Jr, W. M.; Carter, A. P.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin b on the 30S ribosomal subunit. *Cell* **2000**, *103*, 1143-1154.

7. Wilson, D. N.; Harms, J. M.; Nierhaus, K. H.; Schlünzen, F.; Fucini, P. Species-specific antibiotic-ribosome interactions: Implications for drug development. *Biol. Chem.* **2005**, *386*, 1239-1252.

8. Hermann, T. Drugs targeting the ribosome. *Curr. Opin. Struct. Biol.* **2005**, *15*, 355-366.

9. Leach, K. L.; Swaney, S. M.; Colca, J. R.; McDonald, W. G.; Blinn, J. R.; Thomasco, Lisa M.; Gadwood, R. C.; Shinabarger, D.; Xiong, L.; Mankin, A. S. The site of action of oxazolidinone antibiotics in living bacteria and in human mitochondria. *Mol. Cell* **2007**, *26*, 393-402.

10. Lee, M. M.; Pushechnikov, A.; Disney, M. D. Rational and modular design of potent ligands targeting the RNA that causes myotonic dystrophy 2. *ACS Chem. Biol.* **2009**, *4*, 345-355.

11. Davidson, B. L.; McCray, P. B. Current prospects for RNA interference-based therapies. *Nat. Rev. Genet.* **2011**, *12*, 329-340.

12. Watts, J. K.; Corey, D. R. Clinical status of duplex RNA. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3203-3207.

13. Stevens, M.; De Clercq, E.; Balzarini, J. The regulation of HIV-1 transcription: Molecular targets for chemotherapeutic intervention. *Med. Res. Rev.* **2006**, *26*, 595-625.

14. Rana, T. M.; Jeang, K.-T. Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA. *Arch. Biochem. Biophys.* **1999**, *365*, 175-185.

15. Bannwarth, S.; Gatignol, A. HIV-1 TAR RNA: The target of molecular interactions between the virus and its host. *Cur. HIV Res.* **2005**, *3*, 61-71.

16. Keen, N. J.; Gait, M. J.; Karn, J. Human immunodeficiency virus type-1 Tat is an integral component of the activated transcription-elongation complex. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 2505-2510.

17. Hirsch, M. S.; Brun-Vézinet, F.; Clotet, B.; Conway, B.; Kuritzkes, D. R.; D'Aquila, R. T.; Demeter, L. M.; Hammer, S. M.; Johnson, V. A.; Loveday, C.; Mellors, J. W.; Jacobsen, D. M.; Richman, D. D. Antiretroviral drug resistance testing in adults infected with human immunodeficiency virus type 1: 2003 recommendations of an international aids society-USA panel. *Clin. Infect. Dis.* **2003**, *37*, 113-128.

18. Tao, J.; Frankel, A. D. Specific binding of arginine to TAR RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 2723-2726.

19. Pang, R.; Zhang, C.; Yuan, D.; Yang, M. Design and SAR of new substituted purines bearing aryl groups at n9 position as HIV-1 Tat-TAR interaction inhibitors. *Biorg. Med. Chem.* **2008**, *16*, 8178-8186.

20. Hwang, S.; Tamilarasu, N.; Ryan, K.; Huq, I.; Richter, S.; Still, W. C.; Rana, T. M. Inhibition of gene expression in human cells through small molecule-RNA interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 12997-13002.

21. Mei, H. Y.; Mack, D. P.; Galan, A. A.; Halim, N. S.; Heldsinger, A.; Loo, J. A.; Moreland, D. W.; Sannes-Lowery, K. A.; Sharmeen, L.; Truong, H. N.; Czarnik, A. W. Discovery of selective, small-molecule inhibitors of RNA complexes–I. The Tat protein/TAR RNA complexes required for HIV-1 transcription. *Bioorg. Med. Chem.* **1997**, *5*, 1173-1184. 22. Davis, B.; Afshar, M.; Varani, G.; Murchie, A. I.; Karn, J.; Lentzen, G.; Drysdale, M.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T.; Aboul-ela, F. Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *J. Mol. Biol.* **2004**, *336*, 343-356.

23. Murchie, A. I.; Davis, B.; Isel, C.; Afshar, M.; Drysdale, M. J.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T. M.; Mirza, S.; Prescott, C. D.; Vaglio, P.; Aboul-ela, F.; Karn, J. Structure-based drug design targeting an inactive RNA conformation: Exploiting the flexibility of HIV-1 TAR RNA. *J. Mol. Biol.* **2004**, *336*, 625-638.

24. Michienzi, A.; Li, S.; Zaia, J. A.; Rossi, J. J. A nucleolar TAR decoy inhibitor of HIV-1 replication. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14047-14052.

25. Bohjanen, P. R.; Colvin, R. A.; Puttaraju, M.; Been, M. D.; GarciaBlanco, M. A. A small circular TAR RNA decoy specifically inhibits Tat-activated HIV-1 transcription. *Nucleic Acids Res.* **1996**, *24*, 3733-3738.

26. Wang, D.; Iera, J.; Baker, H.; Hogan, P.; Ptak, R.; Yang, L.; Hartman, T.; Buckheit, R. W., Jr.; Desjardins, A.; Yang, A.; Legault, P.; Yedavalli, V.; Jeang, K. T.; Appella, D. H. Multivalent binding oligomers inhibit HIV Tat-TAR interaction critical for viral replication. *Bioorg. Med. Chem. Lett.* 2009, *19*, 6893-6897.

27. Davidson, A.; Leeper, T. C.; Athanassiou, Z.; Patora-Komisarska, K.; Karn, J.; Robinson, J. A.; Varani, G. Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11931-11936.

28. Athanassiou, Z.; Patora, K.; Dias, R. L.; Moehle, K.; Robinson, J. A.; Varani, G. Structure-guided peptidomimetic design leads to nanomolar beta-hairpin inhibitors of the Tat-TAR interaction of bovine immunodeficiency virus. *Biochemistry* **2007**, *46*, 741-751.

29. Lalonde, M. S.; Lobritz, M. A.; Ratcliff, A.; Chamanian, M.; Athanassiou, Z.; Tyagi, M.; Wong, J.; Robinson, J. A.; Karn, J.; Varani, G.; Arts, E. J. Inhibition of both HIV-1 reverse transcription and gene expression by a cyclic peptide that binds the Tat-transactivating response element (TAR) RNA. *PLoS Path.* **2011**, *7*, e1002038.

30. Bryson, D. I.; Zhang, W.; Ray, W. K.; Santos, W. L. Screening of a branched peptide library with HIV-1 TAR RNA. *Mol. BioSyst.* **2009**, *5*, 1070-1073.

31. McNaughton, B. R.; Gareiss, P. C.; Miller, B. L. Identification of a selective smallmolecule ligand for HIV-1 frameshift-inducing stem-loop RNA from an 11,325 member resin bound dynamic combinatorial library. *J. Am. Chem. Soc.* **2007**, *129*, 11306-11307.

32. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res.* **1991**, *37*, 487-493.

33. Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, *354*, 82-84.

34. Lescrinier, T.; Hendrix, C.; Kerremans, L.; Rozenski, J.; Link, A.; Samyn, B.; Aerschot, A. V.; Lescrinier, E.; Eritja, R.; Beeumen, J. V.; Herdewijn, P. DNA-binding ligands from peptide libraries containing unnatural amino acids. *Chem. Eur. J.* **1998**, *4*, 425-433.

35. Brown, B. B.; Wagner, D. S.; Geysen, H. M. A single-bead decode strategy using electrospray ionization mass spectrometry and a new photolabile linker: 3-amino-3-(2-nitrophenyl)propionic acid. *Mol. Divers.* **1995**, *1*, 4-12.

36. Crumpton, J. B.; Zhang, W.; Santos, W. L. Facile analysis and sequencing of linear and branched peptide boronic acids by MALDI mass spectrometry. *Anal. Chem.* **2011**, *83*, 3548-3554.

37. Jullian, M.; Hernandez, A.; Maurras, A.; Puget, K.; Amblard, M.; Martinez, J.; Subra, G. N-terminus FITC labeling of peptides on solid support: The truth behind the spacer. *Tetrahedron Lett.* **2009**, *50*, 260-263.

38. Wang, X.; Huq, I.; Rana, T. M. HIV-1 TAR RNA recognition by an unnatural biopolymer. *J. Am. Chem. Soc.* **1997**, *119*, 6444-6445.

39. Gelman, M. A.; Richter, S.; Cao, H.; Umezawa, N.; Gellman, S. H.; Rana, T. M. Selective binding of TAR RNA by a Tat-derived beta-peptide. *Org. Lett.* **2003**, *5*, 3563-3565.

40. http://www.hiv.lanl.gov

41. Futaki, S.; Nakase, I.; Suzuki, T.; Zhang; Sugiura, Y. Translocation of branchedchain arginine peptides through cell membranes: Flexibility in the spatial disposition of positive charges in membrane-permeable peptides. *Biochemistry*. **2002**, *41*, 7925-7930.

42. Park, S. H.; Doh, J.; Park, S. I.; Lim, J. Y.; Kim, S. M.; Youn, J. I.; Jin, H. T.; Seo, S. H.; Song, M. Y.; Sung, S. Y.; Kim, M.; Hwang, S. J.; Choi, J. M.; Lee, S. K.; Lee, H. Y.; Lim, C. L.; Chung, Y. J.; Yang, D.; Kim, H. N.; Lee, Z. H.; Choi, K. Y.; Jeun, S. S.; Sung, Y. C. Branched oligomerization of cell-permeable peptides markedly enhances the transduction efficiency of adenovirus into mesenchymal stem cells. *Gene Ther.* **2010**, *17*, 1052-1061.

43. Frankel, A. D.; Pabo, C. O. Cellular uptake of the Tat protein from human immunodeficiency virus. *Cell* **1988**, *55*, 1189-1193.

44. Green, M.; Loewenstein, P. M. Autonomous functional domains of chemically synthesized human immunodeficiency virus Tat trans-activator protein. *Cell* **1988**, *55*, 1179-1188.

45. Torchilin, V. P. Tat peptide-mediated intracellular delivery of pharmaceutical nanocarriers. *Adv. Drug Del. Rev.* **2008**, *60*, 548-558.

46. Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **1987**, *15*, 8783-8798.
# Chapter 5 Compounds Discovered from a Branched Peptide Library Featuring Unnatural Boronic Acid Side Chains Selectively Bind to HIV-1 RRE RNA

# Attributions

This chapter was taken from a manuscript that is in preparation for submission to a peer-reviewed journal. The author of this dissertation performed a major portion of the work described in this chapter. He was solely responsible for all RNA footprint assays and footprint optimization. He also performed a significant amount of dot blot assays, electrophoretic mobility shift assays, binding data interpretation, and he prepared and designed the majority of the RNA constructs. Design of the branched peptide library was a collaborative effort between the author of this dissertation, Wenyu Zhang, and Jason B. Crumpton of the Santos Group. Jason B. Crumpton performed all MALDI analysis of peptides and synthesized the unnatural amino acids containing boron. Wenyu Zhang synthesized the peptide library, performed the majority of the high throughput screening, synthesized the majority of the hit peptides for biophysical characterization, and performed a significant amount of dot blot and electrophoretic mobility shift assays. Deconvolution of hit sequences was a collaborative effort between the author of this dissertation, Wenyu Zhang, Jason B. Crumpton, and Jessica Wynn of the Santos Group. Jessica Wynn performed all cell-based assays, and aided in the synthesis of several peptides. The author of this dissertation performed the majority of the writing and editing of the manuscript along with Wenyu Zhang. Jason B. Crumpton and Jessica Wynn also participated in preparation of the manuscript.

### Abstract

The Rev Response Element (RRE) within the env gene of the HIV-1 RNA genome is a highly conserved region that is recognized by the HIV-1 Rev protein, and activates nuclear export of unspliced and singly spliced HIV-1 mRNA to the cytoplasm. The RRE/Rev export pathway is attractive for drug design due to its critical role of allowing the translocation of HIV-1 late transcripts, translation of viral structure proteins, and proliferation of the virus. Early studies have identified the stem-loop structure, known as RRE stem IIB (RRE-IIB), as the initial binding site of the Rev protein. Herein, we report the discovery of the first cell-permeable and nontoxic branched boronic acid peptides (BB peptides), which bind selectively to RRE-IIB RNA in the low micromolar regime. These BB peptides were obtained from on-bead high-throughput screening of a 3.3.4-library comprised of 46,656 unique sequences. We demonstrate that our highest affinity BB peptide is a selective binder for RRE-IIB in the presence of 1000-fold molar excess of tRNAs as well as six RRE-IIB variant structures. We also show that the boronic acid moieties are tunable and have critical effects on binding affinity. In addition, we successfully demonstrate that the "branch" in these peptides provides strong, multivalent interactions with the target RNA.

# 5.1 Introduction

The human immunodeficiency virus type 1 (HIV-1) continues to be a global issue despite significant gains made in research, treatment, and prevention efforts. An estimated 2.7 million new cases of infection occurred in 2010 bringing the total number of people living with HIV-1 to 34-million according to the World Health Organization.<sup>1</sup> HIV-1 remains a problem today due in part to the emergence of drug-resistant viral strains driven by selective pressure. Indeed, the high rate of mutation common among retroviruses provides for rapid evolution of the drug target, which allows it to escape purview of the medication being administered.<sup>2-5</sup> Presently, combination therapies are required for effective treatment of HIV-1, yet discovery of new drug combinations that reduce the development of drug resistance remains a formidable challenge. The majority of drugs approved for the treatment of HIV-1 target either the reverse transcriptase or protease. However, fusion inhibitors, entry inhibitors, and integrase inhibitors have also been approved.<sup>6</sup> Thus, the variety of HIV-1 targets being exploited presently is relatively limited. Furthermore, each of the FDA approved drugs is vulnerable to mutations of HIV-1, as they function through inhibition of a specific enzyme, receptor, or co-receptor.<sup>6</sup> In order to keep pace with the rapidly evolving HIV-1, there is a great need for development of drugs that target novel viral mechanisms, which are genetically well conserved and less prone to mutation under selective pressure.

Lentiviruses, including HIV-1, use a unique pathway to evade the well-controlled and tightly coupled processes of mRNA splicing and nuclear export in eukaryotes. This alternative pathway involves interaction between the virally encoded Rev protein and a well-defined sequence of viral RNA called the Rev response element (RRE), which is a span of ~240-nucleotides located in the *env* gene of all singly spliced and unspliced HIV-1 transcripts.<sup>7-9</sup> The Rev-RRE interaction is essential for these RNA transcripts to escape the nucleus without first being fully spliced; therefore, this interaction is necessary for the synthesis of viral proteins that are vital to the HIV-1 life cycle such as Gag, Pol, and Env as well as the accumulation of HIV-1 genomic RNA in the cytoplasm.<sup>8,10,11</sup> In absence of this process, the structural proteins necessary for viral particle construction are not produced because the intron containing mRNA transcripts are sequestered to the nucleus. Thus, the life cycle of HIV-1 cannot continue.<sup>12</sup>

The Rev/RRE export pathway is mediated by constant shuttling of Rev in and out of the nucleus, and although many key steps have been extensively reviewed, the process has not yet been fully elucidated.<sup>13,14</sup> Initially, HIV-1 encoded Rev protein consisting of 116-amino acid residues is imported into the nucleus by binding to multiple importin proteins using the nuclear localization signal located at the N-terminus of Rev.<sup>15-17</sup> Once inside the nucleus, Rev binds cooperatively to RRE with high affinity in monomeric increments to form a multimeric complex that ultimately involves up to six Rev monomers.<sup>8,11,18-24</sup> The stem-loop IIB of RRE (RRE-IIB) has been recognized as the high affinity site, where Rev initially binds.<sup>9,22,24-29</sup> This Rev-RRE ribonucleoprotein complex binds the host Crm1 through the Rev C-terminal nuclear export signal and is then shuttled out of the nucleus through the nuclear pore after the larger complex binds to Ran-GTP.<sup>30-34</sup>

The Rev/RRE export pathway has become a high profile drug target precisely for its critical role in proliferation of HIV-1.<sup>12</sup> Indeed, the Rev-RRE interaction is completely viral in nature, which provides a high value therapeutic target that is completely independent from the natural cellular processes of the host. This is a huge advantage that could allow the interaction to be targeted selectively with minimal risk of side effects. Additionally, targeting RRE may provide an important opportunity to develop drugs that do not induce drug resistance in the virus because this is a highly conserved region of the *env* gene.<sup>35</sup> Due to the therapeutic potential of the Rev/RRE export pathway, many ligands have been designed to interrupt the Rev-RRE interaction with limited clinical success. Small molecules such as neomycin B as well as other aminoglycosides are demonstrated submicromolar binding ligands of RRE; however, their lack of binding specificity, poor cell permeability, and toxicity make them therapeutically undesirable.<sup>14,36-42</sup> Other inhibitors of the Rev-RRE interaction such as aromatic heterocycles, antisense oligonucleotides, transdominant negative Rev mutant proteins, RRE-based decoys, cyclic peptides,  $\alpha$ -helical peptidomimetics, and others have also been identified, yet none of these have found clinical success.<sup>14,21,40,43-57</sup>

We have previously demonstrated that branching in peptide ligands provides strong multivalent interactions with another HIV-1 related RNA tertiary structure, the transactivation response element (TAR). These branched peptides (BPs) displayed no cytotoxicity, provided excellent cell permeability, and bound to TAR in the submicromolar regime.<sup>58,59</sup> While designing our second generation of BPs, we envisioned a strategy to improve the selectivity and binding affinity to the RNA target through the incorporation of unnatural amino acid side chains featuring boron. We reasoned that the empty p-orbital of boron would aid in ligand binding by acting as an acceptor for the 2'-hydroxyl or other Lewis bases of the RNA. Indeed, peptides displaying boronic acid moieties have been demonstrated elsewhere to form reversible covalent bonds with the alizarin diol and glucose, in addition to being utilized as potent serine protease inhibitors.<sup>60-67</sup> Importantly, boron-containing compounds are well tolerated *in vivo* as is evident from the FDA approval of the first boron-containing drug, Bortezomib (Velcade), which is used in the treatment of multiple myeloma and mantel cell lymphoma.<sup>68,69</sup> Another promising example is that of Tavaborole (AN2690), which is a benzoxaborole small molecule in phase III clinical trials for treatment of onychomycosis.<sup>70</sup>

Herein, we report the discovery and biophysical characterization of the first ever branched boronic acid peptides (BB peptides) as ligands of RRE-IIB and their boron-free analogues, which were obtained through partial screening of a one-bead one-compound library comprised of 46,656 unique sequences. We demonstrate that these multivalent BB peptides can be selective, high-affinity ligands that are both cell-permeable and nontoxic, while removal of the boron moieties has deleterious effects on binding affinity.

# 5.2 **Results and Discussion**

Design of the Branched Boronic Acid Peptide Library and On-Bead High-Throughput Screening

The BB peptide library was synthesized in triplicate on Tentagel beads by split and pool synthesis. The library was prepared such that there were three variable amino acid positions at both the N- and C-termini ( $A_1$ – $A_3$  and  $A_4$ – $A_6$ , respectively), and each variable position was composed of six possible side chains (Figure 5.1).



6<sup>6</sup>=46,656 library members

**Figure 5.1.** Combinatorial (3.3.4) branched-peptide library featuring boronic acid side chains.

Each of the six possible side chains at variable positions  $A_1$ – $A_6$  was chosen for its potential to interact with the RRE-IIB target RNA through a unique mode of binding relative to the other side chains we made available at the same location. In position  $A_1$ for example, we selected amino acids with functional groups that can interact with the RNA through electrostatic attraction (Lys), hydrogen bonding (Ser),  $\pi$ –stacking (Trp), and reversible covalent bonding between boron and a Lewis base presented by the RNA target ( $K_{BBA}/F_{BPA}$ ). The rational for including unnatural boronic acid side chains was three-fold. First, we hypothesized that their inclusion would increase the selectivity of our compounds for RNA over DNA due to the presence of a 2'-hydroxyl that is absent in DNA, which may function as a potential electron donor to the empty p-orbital of boron. Secondly, this introduced an additional mode of binding to the library, which increases the complexity of these compounds. This could improve the chances of discovering ligands that are selective for a single RNA target. Third, we envisioned that the presence of boron would increase overall affinity to the RNA target due to the potential formation of reversible covalent bonds between the RNA and BP ligand. Furthermore, the incorporation of unnatural side chains would likely impart resistance to enzymatic cleavage. We chose to incorporate two different boron-containing side chains at each variable position in the library to examine if RRE-IIB had a preference in the side chain length or electrophility of boron presented by  $F_{BPA}$  and  $K_{BBA}$  (Figure 5.1). Finally, in order to ensure that each unique sequence in the library would contain at least one chromophore, Tyr was included at the C-terminus (position  $A_7$ ) to allow facile determination of peptide concentrations for future biophysical assays. Therefore, we generated three copies of a 3.3.4 branched boronic acid peptide library composed of 46,656 possible amino acid sequences, which were linked to the bead by a photocleavable linker (3-amino-3-(2-nitrophenyl)propionic acid, ANP) at the C-terminus (Scheme 5.1).

**Scheme 5.1.** Synthesis of the 3.3.4 Branched Boronic Acid Peptide Library by Solid-Phase Peptide Synthesis (SPPS).



Half of the entire 3.3.4 library (~70,000 beads) was submitted to high-throughput screening by pre-incubating the library with bovine serum albumin (BSA) (1 mg/mL) and

tRNA (1.25 mg/mL, ~50  $\mu$ M) in phosphate buffer (10 mM potassium phosphate, 100 mM KCl, 1 mM MgCl<sub>2</sub>, 20 mM NaCl, pH=7.0) to block any nonspecific peptide binders. Then, the library was washed and incubated with the 10 nM DY547-labeled HIV-1 RRE-IIB RNA target (Figure 5.2). Specific binding of the target RNA to peptide resulted in increased fluorescence on bead, which was monitored by fluorescent microscopy. Eleven beads demonstrated elevated fluorescence compared to background and were selected as possible hits. These peptides were photocleaved from solid-support resin via UV irradiation and sequenced by matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) MS-MS analysis.<sup>71</sup> Hit compounds were resynthesized in order to perform biophysical characterization assays.

DY547 labeled HIV-1 RRE-IIB RNA

Figure 5.2. Sequence and structure of RRE-IIB.

Binding Affinity Determination of Branched Boronic Acid Peptides Towards HIV-1 RRE-IIB

The dissociation constants ( $K_d$  values) of putative hits and false positives from HTS were identified by dot blot assay. Of the eleven peptides deconvoluted by MALDI-TOF, **BB1-7**, **BB1-10** and **BB1-11** displayed no binding affinity to RRE-IIB, and were deemed false-positive hits. Three other sequences (**BB1-1**, **BB1-6**, and **BB1-8**) stood out for their low micromolar binding affinities (Table 5.1). Gratifyingly, all three of these

sequences contained boronic acid residues. **BB1-1** contained a single  $K_{BBA}$  residue at position  $A_4$  while **BB1-8** displayed  $F_{BPA}$  at positions  $A_1$  and  $A_3$ . In peptide **BB1-6**,  $K_{BPA}$ (positions  $A_1$  and  $A_4$ ) as well as  $F_{PBA}$  (position  $A_3$ ) were represented. Interestingly, no boron containing residues were found at positions  $A_5$  or  $A_6$  in any hit sequences. Consensus sequence analysis also demonstrated that the longer, more Lewis acidic residue of  $K_{BBA}$  was more preferred at positions  $A_1$ ,  $A_2$ , and  $A_4$  over the shorter and less electrophilic  $F_{BPA}$  residue. However, the opposite trend was observed at position  $A_3$ . This demonstrates that both boron containing side chains can provide a positive binding interaction with the RNA target.

entry	peptide	sequence <sup>a</sup>	$K_{\rm d}(\mu{ m M})$	MW(g mol <sup>-1</sup> )
1	BB1-1	(WKK) <sub>2</sub> *K <sub>BBA</sub> YWY	$1.4 \pm 0.4$	1817.99
2	CBB1-1	(WKK) <sub>2</sub> *K <sub>Bz</sub> YWY	8.2 ± 2.3	1773.99
3	CFBB1-1	(WKK) <sub>2</sub> * K <sub>FBA</sub> YWY	$0.8 \pm 0.2$	1835.98
4	BB1-6	$(K_{BBA}KF_{BPA})_2 * K_{BBA}KKY$	3.3 ± 1.2	2031.10
5	BB1-8	(F <sub>BPA</sub> YF <sub>BPA</sub> ) <sub>2</sub> *NKSY	8.7 ± 2.3	1727.78
6	CBB1-8	(FYF) <sub>2</sub> *NKSY	NB	1551.76
7	BB1-5	(KKK <sub>BBA</sub> ) <sub>2</sub> *F <sub>BPA</sub> TSY	$26.8 \pm 4.4$	1751.98
8	BB1-3	(KK <sub>BBA</sub> F) <sub>2</sub> *KKWY	27.2 ± 6.9	1853.04
9	BB1-9	(WYK) <sub>2</sub> *PTWY	28.5 ± 4.4	1646.34
10	BB1-2	(KK <sub>BBA</sub> K) <sub>2</sub> *KLKY	$58.4 \pm 4.0$	1742.09
11	<b>BB1-4</b>	(K <sub>BBA</sub> YK) <sub>2</sub> *HKKY	86.5 ± 10.3	1836.04
12	BB1-7	(VPA) <sub>2</sub> *F <sub>BPA</sub> LAY	NB	1513.81
13	BB1-10	(KNT) <sub>2</sub> *NKK <sub>BBA</sub> Y	NB	1512.82
14	BB1-11	(VPA) <sub>2</sub> *NF <sub>BPA</sub> AY	NB	1219.64
<sup>a</sup> $*$ = Lysine branching unit. Each value is an average of at least three experiments.				

**Table 5.1.** Dissociation Constants ( $K_d$ ) and Molecular Weights of Hit BB Compounds.

Further sequence analysis showed that Lys was highly preferred at most positions, and the number of Lys in each confirmed hit ranged from one to six residues. A preference for positively charged side chains, however, is not unexpected because the Rev protein binds RRE-IIB through the polyarginine nuclear localization signal (NLS) located in its N-terminal region.<sup>22-24,28</sup> The positively charged Lys side chains of our hits likely function similarly to the Arg residues of the Rev NLS and may provide strong electrostatic attraction with the negatively charged phosphate backbone of RRE-IIB. Significantly, the number of Lys residues in the hit sequences did not provide additive

effects on binding. For example, both **BB1-1** and **BB1-6** displayed four Lys residues, and the dissociation constants were lower than 5  $\mu$ M. **BB1-2**, however, contained more Lys residues (6 total) with a weaker  $K_d$  value near 60  $\mu$ M. Additionally, **BB1-8** contained only a single Lys, yet the  $K_d$  remained near 8.7  $\mu$ M. Indeed, we have observed this phenomenon in our previous generation of BPs designed to bind HIV-1 TAR RNA, where higher numbers of Arg side chains did not generally result in increased binding affinity, and other modes of binding could compensate for the loss of an electrostatic interaction.<sup>58,59</sup> Thus, the low dissociation constants of the hit peptides for RRE-IIB are not entirely due to electrostatic attraction. Aliphatic residues were the least preferred side chains in our pool of hit compounds. This suggests that hydrophobic interactions are not generally beneficial for strong binding between these BB peptides and RRE-IIB.

#### The Role of Boronic Acid in Hit Branched Peptides

In order to determine what roles the boronic acids had in binding to the RRE-IIB target RNA, we designed control peptide variants of **BB1-1** and **BB1-8**, without the boronic acid moiety (Figure 5.3). The removal of boronic acid from **BB1-1** (**CBB1-1**) was accompanied by a 4-fold increase in the observed  $K_d$  value. Furthermore, no binding was observed when the boronic acids were removed from **BB1-8** (**CBB1-8**). The complete loss of binding from **CBB1-8** is likely due to the high number of boron containing side chains represented in **BB1-8** because their corresponding mode of binding may be responsible for a larger percentage of the interactions with RRE-IIB compared to **BB1-1**, which only contains a single boronic acid residue. In the case of **BB1-1**, electrostatic attraction is presumably the major interaction with the RNA

provided by the high density of Lys residues in this compound-the single boronic acid likely has a smaller role in binding. Thus, affinity for the target RNA is less affected after removal of the single boronic acid moiety in **BB1-1**. These results clearly indicate that boronic acid side chains can be utilized in peptides to provide strong interactions with a highly structured RNA target.

We hypothesized that additional modifications of  $K_{BBA}$  to increase the Lewis acidity of the boronic acid moiety would yield increased binding affinity to the RRE-IIB RNA. Therefore, we designed a fluorinated **BB1-1** (**CFBB1-1**) with fluorine ortho to the boronic acid at position A<sub>4</sub>. The  $K_d$  of 0.8 ± 0.2 µM was similar or slightly better than the  $K_d$  of 1.4 ± 0.4 µM for **BB1-1**. This result suggests that the binding affinity of BB peptides might be tunable by manipulating the Lewis acidity of the boronic acid.



**Figure 5.3.** Representative structures (from top to bottom) of **BB1-1**, **CBB1-1**, **CFBB1-1**, **BB1-8**, and **CBB1-8** (A) and dot blot assay curves with RRE-IIB (B).

#### Branched Peptides Bind RRE-IIB via Multivalent Interactions

In order to determine if the additional N-terminus in our peptides improved the binding affinities toward RRE-IIB as a result of multivalent interactions, we designed three linear peptide variants of **BB1-1** (Figure 5.4). **LBB1-1A** (WKKGK<sub>BBA</sub>YWY) featured single N- and C-terminus, and the branching Lys residue in **BB1-1** was substituted with Gly to preserve the distance between positions  $A_3$  and  $A_4$ , while also

avoiding the introduction of additional functional groups relative to **BB1-1**. We observed a  $K_d > 30 \mu M$  with **LBB1-1-A**, which was > 21-fold reduction in binding affinity for the RNA target compared to **BB1-1**. To ensure that the decrease in binding affinity was not simply due to a lower total number of amino acid residues compared to **BB1-1** or from a concomitant loss of strong electrostatic interactions (i.e. - the loss of two Lys residues in the variant), we also designed LBB1-1-B (WKKWKKGBYWY). This linear sequence maintained the number and identity of the side chains represented in its branched counterpart. The  $K_d$  of **LBB1-1-B** was 6.8 ± 2.1  $\mu$ M, which represented a 3-fold decrease in binding affinity compared to the branched parent, **BB1-1**. Finally, to determine the importance of the C-terminus, we truncated the C-terminal branch of **BB1-1** to generate **LBB1-1-C** ((WKK)<sub>2</sub>\*), where a branching Lys unit (\*) connected the N-terminal branches. Intriguingly, no binding was observed for this **BB1-1** variant despite the high density of positively charged Lys residues. Taken together, these results suggested that all three branches in **BB1-1** were responsible for high binding affinity with the RRE-IIB RNA, and that binding occurs through multivalent interactions with the target. These results agree well with and expand upon our previous work demonstrating branched

peptides as multivalent binders of RNA tertiary structures.<sup>58,59</sup>



**Figure 5.4.** Sequence of **BB1-1** and linear variants (A) and titration curves of RRE-IIB with these peptides (B).

#### Selectivity of BB1-1 Toward RRE-IIB vs Mutant RRE-IIB and tRNAs

Several mutant RRE-IIB RNAs were synthesized to determine if **BB1-1** could discriminate between native RRE-IIB RNA and RNAs with similar structural elements using a dot blot assay. Initially we synthesized RRE-IIB Augmented Loop RNA where the size of the tri-nucleotide apical loop (AAU) was increased to a hexa-nucleotide loop (AUGGCC) (Figure 5.5). The measured  $K_d$  for this mutant was  $4.4 \pm 0.8 \mu$ M, which is



60

40

20

0

-10

-9

-8

Figure 5.5. Sequence and structure of RRE-IIB and variants (A) and titration curves of **BB1-1** with these RNAs (B).

-6

Log[BB1-1]

-5

-4

-3

-7

slightly increased (~3-fold) compared to the native RRE-IIB RNA structure. We next examined the RRE-IIB Stem B Deleted RNA structure where we removed 4 base pairs in Stem B region, and we obtained a  $K_d$  of 3.7  $\pm$  0.8  $\mu$ M, which also represented a roughly 3-fold decrease in binding affinity. The minor changes in  $K_d$  with these two mutant RRE RNA structures suggest that **BB1-1** has only a minor interaction with the loop and stem B region. Hill analysis of dot blot data for RRE-IIB Augmented Loop and Stem B Deleted RNA provided Hill coefficients (n) of 1.7 and 1.6, respectively (Figure 5.6) This suggests that the modification of the loop and Stem B region results in cooperative binding  $(n \ge 1.5)$  by **BB1-1**. A third modification, RRE-IIB Stem A/Bulge A Deleted RNA, was made by removing Bulge A and four base pairs of Stem A, with resulted in an increase of  $K_d$  by 11-fold (15.6  $\pm$  0.8  $\mu$ M) compared to RRE-IIB and suggests a major interaction between **BB1-1** and the Bulge A or Stem A region. RRE-IIB Bulge B/C Deleted RNA was designed such that only the bulge A remained. RRE-IIB Bulge A (A-G)/B/C Deleted RNA featured only one bulge structural element formed by the unpaired U/G mismatch from bulge A where the A/G mismatch was removed. Finally, RRE-IIB Bulge A/B/C Deleted RNA was prepared as a bulgeless hairpin structure. We found that when only Bulge A remained, the  $K_d$  increased nearly 6-fold (8.7 ± 2.7  $\mu$ M) compared to RRE-IIB RNA. The  $K_d$  increased dramatically as the size of Bulge A was made smaller, where a  $K_d$  of 15.9  $\pm$  0.9  $\mu$ M was observed for RRE-IIB Bulge A (A-G)/B/C Deleted RNA. Furthermore, removing all bulges resulted in a 70-fold increase in the measured  $K_{\rm d}$  $(91.7 \pm 14.5 \ \mu\text{M})$ . The Hill analysis of our dot blot data from RRE-IIB RNA, Stem A/Bulge A Deleted RNA, Bulge B/C Deleted RNA, Bulge A (A-G)/B/C Deleted RNA, and Bulge A/B/C Deleted RNA yielded a Hill coefficient (n) of 1.0, 1.3, 0.9, 1.3 and 1.3, respectively. This suggested that **BB1-1** binds with each of these structures noncooperatively, where n is near 1. Taken together, **BB1-1** binds with RRE-IIB selectively because all modifications of the native sequences resulted in reduced binding affinity.





**Figure 5.6.** Hill plot analysis of **BB1-1** dot blot data. (A) RRE-IIB RNA, (B) RRE-IIB Augmented Loop RNA, (C) RRE-IIB Stem B Deleted RNA, (D) RRE-IIB Stem A/Bulge A Deleted RNA, (E) RRE-IIB Bulge B/C Deleted RNA, (F) RRE-IIB Bulge A(A-G)/B/C Deleted RNA, and (G) RRE-IIB Bulge A/B/C Deleted RNA.

To further demonstrate the selectivity of our highest affinity BB peptides, binding affinities between **BB1-1** and RRE-IIB were measured by dot blot in the presence of excess tRNAs. Initially, a 10-fold molar excess of competing tRNAs was included during incubation with the peptide and RRE-IIB. This resulted in an observed  $K_d$  (2.1 ± 1.0 µM) that was within the margin of error in the absence of tRNA (Figure 5.7). Encouraged by this result, we increased the amount of competing tRNA to a 1000-fold

molar excess, and the observed  $K_d$  was 5.3 ± 3.1 µM, representing a minor shift in RRE-IIB binding. These results suggest that peptide **BB1-1** is selective for RRE-IIB in the presence of excess competing tRNA. These results are quite improved from our previous generation of branched peptides that did not feature boronic acids and demonstrated a dramatic shift (10-fold) in binding affinity in the presence of only a 10-fold molar excess of tRNAs.<sup>58</sup>



**Figure 5.7.** Titration curves of **BB1-1** with RRE-IIB and in the presence of competing tRNA.

## Determination of RRE-IIB Binding Site of BB1-1 by RNase Protection

Although the results of our binding selectivity experiments with RRE-IIB mutants allowed us to evaluate the structural elements required for binding of **BB1-1**, we proceeded forward with ribonuclease protection assays to determine which ribonucleotides were directly involved in binding. **BB1-1** was incubated with varying concentrations of RRE-IIB, and binding sites were confirmed based on the ability of the peptide to protect the RNA from enzymatic cleavage by RNase T1, RNase A, or RNase

V1 (Figure 5.8). The strongest cleavage bands over all the experiments conducted were observed at G17 and C18 using RNase V1, which preferentially cleaves double stranded regions of RNA, and this agrees well with previous reports obtained by enzymatic hydrolysis of RRE.<sup>20,72</sup> This region was also strongly protected from cleavage by **BB1-1** as the concentration of peptide was increased from 0.2 µM to 20 µM. **BB1-1** also visibly protected G10 from RNase V1 as well as U7, C21, and U36 from cleavage by RNase A, which hydrolyzes at the 3' side of unpaired pyrimidine bases. The pattern for **BB1-1** suggests that it binds in the bulged regions of RRE-IIB as U7, G10, and U36 are each located within these structural elements. This agrees well with our selectivity experiments (vide supra), where deletions in the bulge and lower stem regions of RRE-IIB caused dramatically reduced binding affinities with **BB1-1**. Because **BB1-1** protected the upper stem region between at G17 and C18 from RNase V1 as well as C21 from RNase A, it is likely that **BB1-1** is also interacting with the ribonucleotides in this region and is either disrupting the secondary structure of the stem or preventing enzymatic cleavage by steric blockade. This result also agrees with the RRE-IIB mutation studies, where changes in this region resulted in slightly poorer binding affinity. These results in conjunction with the RRE-IIB mutation studies and the noncooperative mode of binding may suggest that **BB1-1** is bound in a folded groove saddled between the bulge regions and upper stem bases of RRE-IIB, as this compound is clearly interacting with both of these structural elements simultaneously.



**Figure 5.8.** RNase protection assays of RRE-IIB using **BB1-1**. The gel depicts the autoradiogram of RNase protection experiments. Alkaline hydrolysis (AH) and all enzymatic digests were performed on RRE-IIB (C). Protection assays were conducted using **BB1-1** (0 - 20  $\mu$ M) with RNases T1, A, and V1. Colored triangles point to bases protected from cleavage by RNase A (blue) and RNase V1 (red). Brackets highlight the regions of RRE-IIB that interact with **BB1-1**.

The strong inhibition of RNase V1 mediated cleavage of RRE-IIB by **BB1-1** at positions at G10, G17 and C18 provided the opportunity to measure the  $EC_{50}$  values for BB1-1 protection of RRE-IIB (Figure 5.9). The  $EC_{50}$  values were obtained by measuring the percentage of RNase V1 protection at each of these three ribonucleotides as BB1-1

was titrated at concentrations of 0.01–100  $\mu$ M. At G10, an EC<sub>50</sub> of 5.5 ± 2.1  $\mu$ M was obtained and is slightly higher (~ 4-fold) than the dissociation constant measured in the absence RNase V1. Measuring nuclease protection at G17 provided the highest EC<sub>50</sub> value (12.4 ± 3.7  $\mu$ M) of the three measured, which was nearly a 9-fold increase compared to the  $K_d$  obtained without RNase V1. Finally, the EC<sub>50</sub> obtained by measuring cleavage at C18 (2.8 ± 1.0  $\mu$ M) suggested that this position is most strongly affected by BB1-1 of the three positions that were measured, and this value was within experimental error of the  $K_d$  obtained in the absence of RNase V1. In each case, the observed protection by BB1-1 may be due to direct binding with the RNA at the nucleotide where protection is observed. However, given that RNase V1 selectively cleaves base-paired ribonucleotides, it is also possible that the binding of BB1-1 at another location is disrupting the base pair where protection is observed. Further experimentation is required to determine the mode of RNase V1 inhibition at G10, G17 and C18.



**Figure 5.9.** RNase V1 protection assay of RRE-IIB at G10, G17, and C18. (A) The gel depicting RNase V1 protection at each RNA base measured and (B) The inhibition curves obtained by densitometry of each band.

#### Branched Boronic Acid Peptides are Cell Permeable and Non-Cytotoxic

Our previous generation of branched peptides that featured Arg rich N-termini exhibited excellent cell permeability and non-cytotoxic properties. In this new peptide library, Lys residues were included instead of Arg, and two new unnatural amino acids featuring boronic acid moieties were added. Despite these changes, we suspected that our BB peptides would be cell permeable in part because they maintained a medium molecular weight (~1000–2000 Da), and contained multiple basic residues.<sup>73</sup> Cellular uptake was performed in both HeLa and A2780 cells by incubation with the FITC-labeled version of **BB1-1** (**FBB1-1**), which was prepared using an established procedure developed for branched peptides.<sup>58</sup> Initially, intense fluorescent background prevented the acquisition of a confocal image due to poor cell contrast. Comparison of a previously

imaged branched peptide (**FL3**) to **FBB1-1** under identical conditions and acquisition parameters showed a large imaging difference between the two peptides (Figure 5.10).<sup>58</sup>



**Figure 5.10.** Cell microscopy from initial incubation of peptides in the presence of mounting media. Confocal image of peptide **FL3** (A) and **FBB1-1** (B) showed large difference in fluorescent background with FITC channel (top left). Top left: fluorescence image of cells; Top right: DAPI staining of the nucleus; Bottom left: overlay of the two images; Bottom right: Empty

Because **FBB1-1** contained a boronic acid residue and **FL3** did not, it was hypothesized that the boronic acid moiety may have allowed residual peptide, adhered on the coverslip, to leach into the Prolong Gold antifade mounting media, thereby increasing background noise to unmanageable levels. Therefore, mounting media was removed from the procedure. Cells were grown in 35 mm poly-lysine treated glass-bottom dishes containing a no. 1.5 coverglass (MatTek) to allow for submersion of fixed cells in PBS for imaging. Bovine serum albumin (BSA) was employed as a blocking agent to help reduce the FITC background on the coverslip. Both HeLa and A2780 cells were incubated with **FBB1-1** (1  $\mu$ M) in the culture medium for 4 h at room temperature. After washing, cells were fixed with 4% (w/v) paraformaldehyde in PBS followed by incubation with BSA and then DAPI. Cells were then imaged using a confocal

microscope. Both HeLa and A2780 cell lines incubated with **FBB1-1** showed fluorescence evenly distributed throughout the cytoplasm and nucleus (Figure 5.11).



**Figure 5.11.** Cellular uptake of **FBB1-1** into (A) Hela cells and (B) A2780 cells. top left: fluorescence image of cells; top right: DAPI staining of the nucleus; bottom left: DIC image; bottom right: overlay of the three images.

Both HeLa and A2780 cells were viable upon incubation with various BB peptides when monitored *via* MTT assay. **BB1-1 and BB1-8** were shown to be non-toxic in concentrations up to 100  $\mu$ M when incubated with A2780 cells for 4 h (Figure 5.12a). In time studies, **BB1-6** remained non-toxic to the A2780 cells even after a 24 h incubation, with a slight drop-off in viability at 100  $\mu$ M for all incubation times (Figure 5.13). The concentration of **BB1-6** was then reduced to 30  $\mu$ M for the time study, and delightfully the cells treated with peptide showed excellent viability relative to control cells at this concentration for all incubation times (Figure 5.14). Further toxicity analysis revealed that **BB1-1** remained non-toxic to both HeLa and A2780 cells up to 10  $\mu$ M for 24 h (Figure 5.12b). In summary, our highest affinity BB peptides (**BB1-1** and **BB1-8**) were found to be cell permeable and non-toxic up to 100  $\mu$ M for 4 h or up to 10  $\mu$ M for 24 h.



**Figure 5.12.** MTT assays with **BB1-1** and **BB1-8**. (A) MTT cell toxicity assay of **BB1-1** and **BB1-8** at various concentrations in A2780 cells for 4 h and (B) MTT cell toxicity assay of **BB1-1** up to 10  $\mu$ M in both HeLa and A2780 cells for 24 h.



**Figure 5.13.** A2780 cell MTT assay 24 h time study as a function of the concentration of **BB1-6**.



Figure 5.14. A2780 cell MTT assay 24 h time study of BB1-6 at 30 µM.

In this study, we have generated the first example of a BB peptide library, which was designed to bind an RNA target. We successfully discovered several cell permeable and nontoxic BB peptides with low micromolar binding affinities to HIV-1 RRE-IIB. Also, we have demonstrated for the first time that boronic acid moieties can be utilized in peptides to provide strong interactions with a highly structured RNA target, and that in conjunction with an unnatural branch they can provide multivalent interactions with the target structure. Peptide **BB1-1** displayed the highest affinity to RRE-IIB and contained a single boronic acid moiety (K<sub>BBA</sub>) that was shown to be essential to binding the RNA target. The binding affinity of **BB1-1** was also shown to be tunable based on the electrophilicity of boron, which was adjusted through the installation of fluorine at the ortho position with respect to the boronic acid. **BB1-1** was also shown to selectively bind with RRE-IIB in the presence of 1000-fold excess tRNA and displayed reduced binding affinities (up to 70-fold) to each RRE-IIB variant that was tested. We have also provided data indicating that the binding site of **BB1-1** involves the bulge and upper stem bases of RRE-IIB. The results from this study highlight the therapeutic potential of these novel RNA binding compounds as well as the utility of this method for RNA ligand discovery. Current efforts are aimed to further improve the binding affinity and selectivity of branched boronic acid peptides and demonstrate their therapeutic potential in cell-based assays. Taken together, the results from this study will inform our ongoing efforts to target highly structured RNAs with high affinity and selectivity.

# **5.3** Materials and Methods

#### Synthesis of Branched Boronic Acid Peptide Library on Resin

We used standard solid phase peptide synthesis techniques to generate the 3.3.4 library via the split and pool method using the previously described apparatus.<sup>59</sup> N- $\alpha$ -Fmoc protected L-amino acids (Novabiochem), PyOxim (Nobabiochem) and N,N-Diisopropylethylamine (DIEA, Aldrich) were used in all coupling reactions. The synthesis and full characterization of Fmoc protected N-E-(4-boronobenzoyl)-L-lysine  $(K_{BBA})$ , N- $\varepsilon$ -benzoyl-L-lysine  $(K_{Bz})$  and N- $\varepsilon$ -(4-borono-3-fluorobenzoyl)-L-lysine  $(K_{FBA})$ are detailed in a forthcoming publication (manuscript in preparation). Fmoc-L-4boronophenylalanine ( $F_{BPA}$ ) and Fmoc-ANP-OH were synthesized using previous methods.<sup>60,74,75</sup> The resin was mixed in solution by bubbling argon during all coupling and washing steps. Three copies of library were prepared simultaneously by using a three-fold excess of Tentagel Macrobead-NH<sub>2</sub> resin (2.13 g, 0.57 mmol/g) (Peptides International). The resin was swollen in DCM (20 mL,  $2 \times 15$  min) followed by DMF (20 mL, 15 min). The photocleavable linker Fmoc-ANP-OH (739 mg, 1.71 mmol) was first coupled to the resin in DMF for 3 h in the presence of PyOxim (901.53 mg, 1.71 mmol) and DIEA (493 µL, 2.85 mmol). After coupling, the resin was washed by DMF (20 mL, 1 min), DCM (20 mL, 1 min) followed by another DMF (20 mL, 1 min) wash. The same washing procedure was applied after every step. Then, 20% piperidine in DMF (20 mL,  $2 \times 10$  min) was used for Fmoc deprotection. A Kaiser test was used after each coupling and deprotection step to confirm that each reaction went to completion. N-Fmoc amino acids (3 equiv), PyOxim (3 equiv), and DIEA (5 equiv) were added to each reaction vessels in DMF and coupled for 30 min. Fmoc-Lys(Fmoc)-OH was used as a branching unit, and molar equivalences of all reagents were doubled in coupling reactions after installation of the branching unit. After Fmoc deprotection of the N-terminal amino acids, the resin was bubbled in a phenylboronic acid solution (0.2 g/mL) overnight to remove the pinacol groups of boron-containing side-chains. Finally, the resin was treated with 95:2.5:2.5 TFA (Trifluoroacetic acid, Acros)/H<sub>2</sub>O/TIS (Triisopropylsilane, Acros) (v/v/v) for 3 h to deprotect side-chains and any remaining pinacol groups. After deprotection, the resin was washed extensively with DMF, DCM, and MeOH before drying and storing it at -20 °C.

## Library Incubation and Screening for Binding to HIV-1 RRE-IIB

A DY547 labeled HIV-1 RRE-IIB RNA (5'-DY547-GGC-UGG-UAU-GGG-CGC-AGC-GUC-AAU-GAC-GCU-GAC-GGU-ACA-GGC-CAG-CC-3') was purchased from Dharmacon and prepared according to the manufacturer's protocol. In order to account for the auto-fluorescence displayed by the Tentagel Macrobead-NH<sub>2</sub> resin, control peptide (KYR)<sub>2</sub>\*FDS was incubated in 100 nM DY547 labeled HIV-1 RRE-IIB RNA for 1 h in phosphate buffer (10 mM potassium phosphate, 100 mM KCl, 1 mM MgCl<sub>2</sub>, 20 mM NaCl, pH 7.0). These beads were washed extensively and placed into a

sterile 96-well plate (Nunc) to be imaged by a Zeiss Axiovert 200 fluorescent microscope under a rhodamine filter. The fluorescence intensity was adjusted to remove the background auto-fluorescence of untreated beads, while RNA treated beads remained visible.

Screening conditions were initially tested using about 2,400 beads from the library. The beads were put into a 1.5 mL non-stick microfuge tube (Fisher) with a 500 µL final volume of phosphate buffer and mixed by a Barnstead/Thermolyne Labquake rotisserie shaker. The beads were first treated with 1 mg/mL bovine serum albumin (BSA) (New England BioLab) and 0.8 mg/mL tRNA (Roche) (~3,144-fold molar excess to RRE-IIB RNA) for 3 h at room temperature to block any nonspecific binding peptide sequences. Then, the beads were washed 5 times with phosphate buffer and incubated in 500 µL of 10 nM DY547 labeled RRE-IIB RNA in phosphate buffer for 5 h at room temperature. After the final incubation, the resin was extensively washed with buffer, and the beads were imaged under the Zeiss fluorescent microscope in a 96-well plate using the previously optimized settings. The initial screening yielded 7 beads with elevated levels of fluorescence. These 7 beads were isolated, rinsed with DMF (5  $\times$  500  $\mu$ L) and MeOH (5 × 500  $\mu$ L), and photocleaved in clear non-stick 0.5 mL microfuge tubes in 15  $\mu$ L of 1:1 MeOH: H<sub>2</sub>O (v/v) by irradiating at 365 nm with a 4W handheld UV lamp. The supernatant was retained and subjected to MALDI-TOF analysis. After analysis of MALDI-MS/MS fragmentation data, we were able to obtain 6 sequences, where the remaining sample yielded no signal upon analysis and was deemed a false hit. This method provided a hit rate of 0.25%.

The remaining beads of the 3.3.4 library were screened using more stringent conditions in an attempt to reduce the total number of hit beads. First the beads were blocked at room temperature for 3 h in phosphate buffer with 1 mg/mL BSA, and 1.25 mg/mL tRNA (5000-fold molar excess to RRE stem IIB RNA). Next, the beads were washed in buffer and incubated with 10 nM DY547 labeled RRE-IIB RNA in phosphate buffer for 5 h at room temperature. Resin was washed extensively after final incubation prior to screening. Another 5 beads were found with elevated fluorescence using these more stringent conditions. All 5 of these beads were photocleaved individually and successfully sequenced *de novo* using MALDI-TOF.

#### Peptide Synthesis, Purification and Characterization

Unlabeled and fluorescein 5-isothiocyanate (FITC) (Sigma) labeled peptides were synthesized on Rink amide MBHA resin (100-200 mesh) (Novabiochem). In the preparation of the FITC-labeled peptides, Fmoc-Lys(ivDde)-OH (Novabiochem) was used as the branching unit. Acetic anhydride in DMF (1:1 v/v) with 10 equivalents of DIEA was used to cap the first N-terminus. Then, ivDde was removed by treatment with 2% hydrazine in DMF for 1 h, and the second N-terminus was synthesized through the N<sup>e</sup> atom of the Lys side chain. Fmoc-6-Ahx-OH (AnaSpec) was coupled to the N-terminal amino acid to provide a linker for FITC, which prevents autocleavage of FITC under acidic conditions.<sup>58,76</sup> All subsequent steps were protected from light. FITC (Sigma) (5 equiv) was reacted with the deprotected N-terminus of the peptides for 6 h using DIEA (14 equiv). The pinacol exchange with phenylboronic acid and deprotection of side chains followed the same procedure as the peptide library synthesis. The supernatant was

dried under reduced pressure, and the crude peptide was triturated from cold diethyl ether. All the peptides were purified using a Jupiter 4  $\mu$ m Proteo 90 Å semiprep column (Phenomenex), using a solvent gradient composed of 0.1% TFA in Milli-Q water and HPLC grade acetonitrile. Purity was determined under the same conditions using a Jupiter 4  $\mu$ m Proteo 90 Å analytical column (Phenomenex), and all the peptides were purified to  $\geq$  90%. After purification, peptides were characterized by MALDI-TOF analysis. Peptide stock concentrations were prepared by spectrophotometry. Unlabeled peptide concentrations were measured in nuclease free water at 280 nm using their calculated extinction coefficients. FITC-labeled peptide concentrations were monitored at 495 nm using the extinction coefficient of FITC at 77,000 mol<sup>-1</sup> cm<sup>-1</sup> in 100 mM glycine, pH 9.0.

# Preparation of <sup>32</sup>P-labeled RNA

Wild-type and mutant RRE-IIB RNAs were transcribed in vitro by T7 polymerase with the Ribomax T7 Express System (Promega) using previously reported techniques.<sup>77</sup> The antisense templates. complementary strand (5'sense ATGTAATACGACTCACTATAGG-3') and RRE-IIB reverse PCR primer (5'-GGCTGGCCTGTAC-3') were purchased from integrated DNA Technologies. Antisense templates were used as follows, where mutations are in **bold** and insertions and deletions 5'are underlined: HIV-1 **RRE-IIB** RNA GGCTGGCCTGTACCGTCAGCGTCATTGACGCTGCGCCCATACCAGCCCTATA GTGAGTCGTATTACAT-3'; 5'-**RRE-IIB** Augmented Loop **RNA** GGCTGGCCTGTACCGTCAGCGTCGGCATTGACGCTGCGCCCATACCAGCCCT

ATAGTGAGTCGTATTACAT-3'; RRE-IIB Stem A/Bulge A Deleted RNA 5'-GGCGTACCGTCAGCGTCATTGACGCTGCGCCCGCCCTATAGTGAGTCGTATTA В CAT-3': RRE-IIB Stem Deleted RNA 5'-GGCTGG CCTGTACCGTCACATTGTGCGCCCATACCAGCCCTATAGTGAGTCGTATTACA 5'-T-3'; **RRE-IIB** Bulge A(A-G)/B/CDeleted RNA GGCTGGCCAGCGTCATTGACGCTGACCAGCCCTATAGTGAGTCGTATTACAT-3': 5'-**RRE-IIB** Bulge B/C Deleted RNA GGCTGGCCTGCAGCGTCATTGACGCTGCATACCAGCCCTATAGTGAGTCGTAT TACAT-3'; **RRE-IIB** Bulge A/B/CDeleted **RNA** 5'-GGCTGGCAGCGTCATTGACGCTGCCAGCCCTATAGTGAGTCGTATTACAT-3'. RRE-IIB Augmented Loop and HIV-1 RRE-IIB were both PCR amplified using HotstarTaq DNA polymerase (Qiagen) followed by a clean-up procedure using a spin column kit (Qiagen). PCR products did not require an annealing step prior to T7 transcription. For the preparation of all other sequences, the antisense DNA template was annealed with the sense DNA complementary strand in reaction buffer at 95 °C for 2 min then cooled on ice for 4 min. T7 transcription of all sequences proceeded at 42 °C for 1.5 h. After transcription, DNA templates were degraded by incubation with DNase at 37 °C for 45 min and the transcribed RNA was purified by a 12% polyacrylamide gel containing 7.5 M urea (denaturing PAGE). The band corresponding to the RNA of interest was excised from the gel and eluted overnight in  $1 \times \text{TBE}$  buffer at 4 °C. The sample was desalted using a Sep-Pak syringe cartridge (Waters Corporation), lyophilized dry, and dephosphorylated with calf intestinal phosphatase (CIP) in NEBuffer 3 (New England Biolabs) according to manufacturer's protocol. The product was recovered by a

standard phenol extraction followed by ethanol precipitation. Purified RNA was stored as a pellet at -80 °C.

HIV-1 RRE-IIB RNA as well as the mutant RNA sequences were labeled at the 5'-end by treating 10 pmol of dephosphorylated RNA with 20 pmol of  $[\gamma^{-32}P]$  ATP (111 TBq mol<sup>-1</sup>) and 20 units of T4 polynucleotide kinase in 70 mM Tris•HCl, 10 mM MgCl<sub>2</sub>, and 5 mM dithiothreitol, pH 7.6. The mixture was incubated at 37 °C for 30 min, and then at room temperature for 20 min. The kinase was finally heat-inactivated at 65 °C for 10 min. The RNA was recovered by ethanol precipitation, and the purity was examined using 12% denaturing PAGE followed by autoradiography.

# Dot Blot Assay

Dot blot assays were performed at room temperature using a Whatman Minifold I 96 well Dot Blot system and Whatman 0.45  $\mu$ m pore size Protran nitrocellulose membranes. To determine the binding affinities, 0.4 nM radiolabeled RNA was titrated with peptide (0.001–100  $\mu$ M). First, a solution of 0.08 nM <sup>32</sup>P-labeled RNA was refolded in 2× phosphate buffer (20 mM potassium phosphate, 200mM KCl, 1mM MgCl<sub>2</sub>, 40mM NaCl, pH 7.0) by heating at 95 °C for 3 min and then slowly cooling at room temperature for 20 min. Next, 25  $\mu$ L of the [<sup>32</sup>P]-RNA solution was added to 25  $\mu$ L of peptide in nuclease free water and incubated at room temperature for 4 h. The 50  $\mu$ L mixtures were filtered through the nitrocellulose membrane, which was immediately followed by two consecutive 50  $\mu$ L washes with 1× phosphate buffer. Peptide binding was visualized by autoradiography using a storage phosphor screen (GE Healthcare) and a Typhoon Trio phosphorimager (GE Healthcare). Densitometry measurements were
taken with ImageQuant TL (Amersham Biosciences). Binding curves were generated using Kaleidagraph (Synergy Software). Each experiment was performed in triplicate and error bars represent the standard deviation calculated over three replicates.

#### Nuclease Protection Procedures

RNA was first refolded by heating a solution of 5<sup>°</sup>-<sup>32</sup>P-labeled RRE-IIB (10 nM) and excess unlabeled RRE-IIB (200 nM) at 95 °C for 3 min and then snap cooling it on ice. The refolded RNA was incubated on ice for 4 h in a solution containing the branched boronic acid peptide and buffer composed of 10 mM Tris, pH 7, 100 mM KCl, and 10 mM MgCl<sub>2</sub>. RNase (Ambion) was then added to the solution, which was further incubated on ice for 10 min (0.002 Units RNase V1), or 1 h (1 Unit RNase T1; 20 ng RNase A). Inactivation/Precipitation Buffer (Ambion) was added to halt digestion, and then RNA was pelleted by centrifugation at 13,200 rpm for 15 min. Pelleted RNA was redissolved into tracking dye and run through a 12 % denaturing polyacrylamide gel containing 7.5 M Urea and imaged by autoradiography.

### Cellular Internalization of Peptides and MTT Toxicity Assay

HeLa cells were grown and plated for peptide internalization using previously established methods.<sup>58</sup> A2780 (ovarian) cells were grown and plated in RPMI 1640 media supplemented with 10% heat-inactivated fetal bovine serum (FBS), 1% L-glutamine, 100 units/mg penicillin, 100  $\mu$ g/mL streptomycin, and 0.25  $\mu$ g/mL amphotericin (Invitrogen). Both cell lines were maintained in a 37 °C incubator containing a 5% CO<sub>2</sub> atmosphere and were subcultured once per week. Cell samples

were prepared in dishes using the manufacturer's protocol. The dishes were preequilibrated with media for 15 min at 37 °C in 5% CO2 atmosphere. Media was then aspirated from the dishes and the cells were plated at  $1.5 \times 10^4$  cells/dish in a 500  $\mu$ L media suspension and incubated for 1 h to allow for initial cell adherence. Additional media (2 mL) was then added and the cells were incubated for a total of 48 h. After the removal of medium and washing of cells with PBS, 600 µL of FITC-labeled **BB1-1** (FBB1-1) in Opti-MEM (1 µM) (Invitrogen/Gibco) was added to the dish; cells were incubated with the peptide for 4 h. Control cell samples were incubated in 600 µL of Opti-MEM for 4 h. After incubation and following each subsequent step, a 15 min PBS wash was applied to all cell samples at room temperature. Cells were fixed with 4% (w/v) paraformaldehyde (Acros) in PBS for 15 min, and then incubated with 3% (v/v) BSA (New England Biolabs) in PBS for 30 min. Nucleus staining was then performed with the addition of 150 µL of 600 nM DAPI to the cell samples for 4 min. Finally, cells were submersed in PBS, and the dishes were sealed with parafilm and stored at 4°C in the dark until imaged. Cells were imaged using a  $40 \times$  water-immersion objective (N.A. = 1.2) on an LSM 510 confocal system mounted on an Axiovert 100 inverted microscope (Zeiss). For each cell line, identical acquisition parameters were used for both peptide and control samples. Brightness and contrast were adjusted for the image processing of DAPI and DIC channels using AxioVisionLE software (Zeiss). For the MTT toxicity assays, 96-well plates (Nunc) were first incubated with 100  $\mu$ L of poly-lysine (Sigma) for 24 h at 4°C. The poly-lysine was removed and the wells were rinsed ten times with sterile nuclease-free water. HeLa and A2780 cells were then plated in the poly-lysine treated wells and analyzed via an MTT toxicity assay using previously published procedure.<sup>58</sup>

## References

1. World Health Organization, UNICEF, and UNAIDS. *Global HIV/AIDS Response: Epidemic Update and Health Sector Progress Towards Universal Access*, Progress Report 2011; Geneva, Switzerland, **2011**.

2. Balfe, P.; Simmonds, P.; Ludlam, C. A.; Bishop, J. O.; Brown, A. J. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: Rate of sequence change and low frequency of inactivating mutations. *J. Virol.* **1990**, *64*, 6221-6233.

3. Domaoal, R. A.; Demeter, L. M. Structural and biochemical effects of human immunodeficiency virus mutants resistant to non-nucleoside reverse transcriptase inhibitors. *Int. J. Biochem. Cell Biol.* **2004**, *36*, 1735-1751.

4. Wei, X.; Decker, J. M.; Liu, H.; Zhang, Z.; Arani, R. B.; Kilby, J. M.; Saag, M. S.; Wu, X.; Shaw, G. M.; Kappes, J. C. Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother.* **2002**, *46*, 1896-1905.

5. Fikkert, V.; Van Maele, B.; Vercammen, J.; Hantson, A.; Van Remoortel, B.; Michiels, M.; Gurnari, C.; Pannecouque, C.; De Maeyer, M.; Engelborghs, Y.; De Clercq, E.; Debyser, Z.; Witvrouw, M. Development of resistance against diketo derivatives of human immunodeficiency virus type 1 by progressive accumulation of integrase mutations. *J. Virol.* **2003**, *77*, 11459-11470.

6. Mehellou, Y.; De Clercq, E. Twenty-six years of anti-HIV drug discovery: Where do we stand and where do we go? *J. Med. Chem.* **2009**, *53*, 521-538.

7. Heaphy, S.; Dingwall, C.; Ernberg, I.; Gait, M. J.; Green, S. M.; Karn, J.; Lowe, A. D.; Singh, M.; Skinner, M. A. HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element region. *Cell* **1990**, *60*, 685-693.

8. Malim, M. H.; Hauber, J.; Le, S. Y.; Maizel, J. V.; Cullen, B. R. The HIV-1 Rev transactivator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **1989**, *338*, 254-257. 9. Malim, M. H.; Tiley, L. S.; McCarn, D. F.; Rusche, J. R.; Hauber, J.; Cullen, B. R. HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* **1990**, *60*, 675-683.

10. Felber, B. K.; Hadzopoulou-Cladaras, M.; Cladaras, C.; Copeland, T.; Pavlakis, G. N. Rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 1495-1499.

11. Hammarskjold, M. L.; Heimer, J.; Hammarskjold, B.; Sangwan, I.; Albert, L.; Rekosh, D. Regulation of human immunodeficiency virus env expression by the Rev gene product. *J. Virol.* **1989**, *63*, 1959-1966.

12. Dayton, A. I.; Sodroski, J. G.; Rosen, C. A.; Goh, W. C.; Haseltine, W. A. The transactivator gene of the human T cell lymphotropic virus type III is required for replication. *Cell* **1986**, *44*, 941-947.

13. Suhasini, M.; Reddy, T. R. Cellular proteins and HIV-1 Rev function. *Curr. HIV Res.* **2009**, *7*, 91-100.

14. Yuan, C.; Xinyong, L.; De Clercq, E. Cessation of HIV-1 transcription by inhibiting regulatory protein Rev-mediated RNA transport. *Curr. HIV Res.* **2009**, *7*, 101-108.

15. Henderson, B. R.; Percipalle, P. Interactions between HIV Rev and nuclear import and export factors: The Rev nuclear localisation signal mediates specific binding to human importin-beta. *J. Mol. Biol.* **1997**, *274*, 693-707.

16. Truant, R.; Cullen, B. R. The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. *Mol. Cell. Biol.* **1999**, *19*, 1210-1217.

17. Arnold, M.; Nath, A.; Hauber, J.; Kehlenbach, R. H. Multiple importins function as nuclear transport receptors for the Rev protein of human immunodeficiency virus type 1. *J. Biol. Chem.* **2006**, *281*, 20883-20890.

18. Rosen, C. A.; Terwilliger, E.; Dayton, A.; Sodroski, J. G.; Haseltine, W. A. Intragenic cis-acting art gene-responsive sequences of the human immunodeficiency virus. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2071-2075.

19. Hadzopoulou-Cladaras, M.; Felber, B. K.; Cladaras, C.; Athanassopoulos, A.; Tse, A.; Pavlakis, G. N. The Rev (trs/art) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a cis-acting sequence in the env region. *J. Virol.* **1989**, *63*, 1265-1274.

20. Kjems, J.; Brown, M.; Chang, D. D.; Sharp, P. A. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 683-687.

21. Legiewicz, M.; Badorrek, C. S.; Turner, K. B.; Fabris, D.; Hamm, T. E.; Rekosh, D.; Hammarskjold, M. L.; Le Grice, S. F. Resistance to RevM10 inhibition reflects a conformational switch in the HIV-1 Rev response element. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14365-14370.

22. DiMattia, M. A.; Watts, N. R.; Stahl, S. J.; Rader, C.; Wingfield, P. T.; Stuart, D. I.; Steven, A. C.; Grimes, J. M. Implications of the HIV-1 Rev dimer structure at 3.2 Å resolution for multimeric binding to the Rev response element. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5810-5814.

23. Daugherty, M. D.; Booth, D. S.; Jayaraman, B.; Cheng, Y.; Frankel, A. D. HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 12481-12486.

24. Daugherty, M. D.; Liu, B.; Frankel, A. D. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1337-1342.

25. Cook, K. S.; Fisk, G. J.; Hauber, J.; Usman, N.; Daly, T. J.; Rusche, J. R. Characterization of HIV-1 Rev protein: Binding stoichiometry and minimal RNA substrate. *Nucleic Acids Res.* **1991**, *19*, 1577-1583.

26. Huang, X. J.; Hope, T. J.; Bond, B. L.; McDonald, D.; Grahl, K.; Parslow, T. G. Minimal Rev-response element for type 1 human immunodeficiency virus. *J. Virol.* **1991**, *65*, 2131-2134.

27. Mann, D. A.; Mikaelian, I.; Zemmel, R. W.; Green, S. M.; Lowe, A. D.; Kimura, T.; Singh, M.; Butler, P. J.; Gait, M. J.; Karn, J. A molecular rheostat. Co-operative Rev binding to stem I of the Rev-response element modulates human immunodeficiency virus type-1 late gene expression. *J. Mol. Biol.* **1994**, *241*, 193-207.

28. Battiste, J. L.; Mao, H.; Rao, N. S.; Tan, R.; Muhandiram, D. R.; Kay, L. E.; Frankel, A. D.; Williamson, J. R. Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* **1996**, *273*, 1547-1551.

29. Tan, R.; Chen, L.; Buettner, J. A.; Hudson, D.; Frankel, A. D. RNA recognition by an isolated  $\alpha$  helix. *Cell* **1993**, *73*, 1031-1040.

30. Askjaer, P.; Jensen, T. H.; Nilsson, J.; Englmeier, L.; Kjems, J. The specificity of the Crm1-Rev nuclear export signal interaction is mediated by rangtp. *J. Biol. Chem.* **1998**, *273*, 33414-33422.

31. Bogerd, H. P.; Echarri, A.; Ross, T. M.; Cullen, B. R. Inhibition of human immunodeficiency virus Rev and human T-cell leukemia virus rex function, but not mason-pfizer monkey virus constitutive transport element activity, by a mutant human nucleoporin targeted to Crm1. *J. Virol.* **1998**, *72*, 8627-8635.

32. Fornerod, M.; Ohno, M.; Yoshida, M.; Mattaj, I. W. Crm1 is an export receptor for leucine-rich nuclear export signals. *Cell* **1997**, *90*, 1051-1060.

33. Zolotukhin, A. S.; Felber, B. K. Nucleoporins Nup98 and Nup214 participate in nuclear export of human immunodeficiency virus type 1 Rev. *J. Virol.* **1999**, *73*, 120-127.

34. Neville, M.; Stutz, F.; Lee, L.; Davis, L. I.; Rosbash, M. The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Curr. Biol.* **1997**, *7*, 767-775.

35. Le, S. Y.; Malim, M. H.; Cullen, B. R.; Maizel, J. V. A highly conserved RNA folding region coincident with the Rev response element of primate immunodeficiency viruses. *Nucleic Acids Res.* **1990**, *18*, 1613-1623.

36. Hendrix, M.; Priestley, E. S.; Joyce, G. F.; Wong, C. H. Direct observation of aminoglycoside-RNA interactions by surface plasmon resonance. *J. Am. Chem. Soc.* **1997**, *119*, 3641-3648.

37. Hamasaki, K.; Ueno, A. Aminoglycoside antibiotics, neamine and its derivatives as potent inhibitors for the RNA-protein interactions derived from HIV-1 activators. *Biorg. Med. Chem. Lett.* **2001**, *11*, 591-594.

38. Lacourciere, K. A.; Stivers, J. T.; Marino, J. P. Mechanism of neomycin and Rev peptide binding to the Rev responsive element of HIV-1 as determined by fluorescence and NMR spectroscopy. *Biochemistry* **2000**, *39*, 5630-5641.

39. Zapp, M. L.; Stern, S.; Green, M. R. Small molecules that selectively block RNA binding of HIV-1 Rev protein inhibit Rev function and viral production. *Cell* **1993**, *74*, 969-978.

40. Zapp, M. L.; Young, D. W.; Kumar, A.; Singh, R.; Boykin, D. W.; Wilson, W. D.; Green, M. R. Modulation of the Rev-RRE interaction by aromatic heterocyclic compounds. *Biorg. Med. Chem.* **1997**, *5*, 1149-1155.

41. Wilson, W. D.; Li, K. Targeting RNA with small molecules. *Curr. Med. Chem.* **2000**, 7, 73-98.

42. Thomas, J. R.; Hergenrother, P. J. Targeting RNA with small molecules. *Chem. Rev.* **2008**, *108*, 1171-1224.

43. Xiao, G.; Kumar, A.; Li, K.; Rigl, C. T.; Bajic, M.; Davis, T. M.; Boykin, D. W.; Wilson, W. D. Inhibition of the HIV-1 Rev-RRE complex formation by unfused aromatic cations. *Biorg. Med. Chem.* **2001**, *9*, 1097-1113.

44. Luedtke, N. W.; Liu, Q.; Tor, Y. RNA-ligand interactions: Affinity and specificity of aminoglycoside dimers and acridine conjugates to the HIV-1 Rev response element. *Biochemistry* **2003**, *42*, 11391-11403.

45. Li, K.; Davis, T. M.; Bailly, C.; Kumar, A.; Boykin, D. W.; Wilson, W. D. A heterocyclic inhibitor of the Rev–RRE complex binds to RRE as a dimer. *Biochemistry* **2001**, *40*, 1150-1158.

46. Hart, R. A.; Billaud, J. N.; Choi, S. J.; Phillips, T. R. Effects of 1,8-diaminooctane on the FIV Rev regulatory system. *Virology* **2002**, *304*, 97-104.

47. DeJong, E. S.; Chang, C. E.; Gilson, M. K.; Marino, J. P. Proflavine acts as a Rev inhibitor by targeting the high-affinity Rev binding site of the Rev responsive element of HIV-1. *Biochemistry* **2003**, *42*, 8035-8046.

48. Prater, C. E.; Saleh, A. D.; Wear, M. P.; Miller, P. S. Allosteric inhibition of the HIV-1 Rev/RRE interaction by a 3'-methylphosphonate modified antisense oligo-2'-o-methylribonucleotide. *Oligonucleotides* **2007**, *17*, 275-290.

49. Prater, C. E.; Saleh, A. D.; Wear, M. P.; Miller, P. S. Chimeric RNase H-competent oligonucleotides directed to the HIV-1 Rev response element. *Biorg. Med. Chem.* **2007**, *15*, 5386-5395.

50. Hamm, T. E.; Rekosh, D.; Hammarskjold, M. L. Selection and characterization of human immunodeficiency virus type 1 mutants that are resistant to inhibition by the transdominant negative RevM10 protein. *J. Virol.* **1999**, *73*, 5741-5747.

51. Plavec, I.; Agarwal, M.; Ho, K. E.; Pineda, M.; Auten, J.; Baker, J.; Matsuzaki, H.; Escaich, S.; Bonyhadi, M.; Bohnlein, E. High transdominant RevM10 protein levels are required to inhibit HIV-1 replication in cell lines and primary T cells: Implication for gene therapy of aids. *Gene Ther.* **1997**, *4*, 128-139.

52. Lee, T. C.; Sullenger, B. A.; Gallardo, H. F.; Ungers, G. E.; Gilboa, E. Overexpression of RRE-derived sequences inhibits HIV-1 replication in CEM cells. *New. Biol.* **1992**, *4*, 66-74.

53. Symensma, T. L.; Baskerville, S.; Yan, A.; Ellington, A. D. Polyvalent Rev decoys act as artificial Rev-responsive elements. *J. Virol.* **1999**, *73*, 4341-4349.

54. Giver, L.; Bartel, D. P.; Zapp, M. L.; Green, M. R.; Ellington, A. D. Selection and design of high-affinity RNA ligands for HIV-1 Rev. *Gene* **1993**, *137*, 19-24.

55. Chaloin, L.; Smagulova, F.; Hariton-Gazal, E.; Briant, L.; Loyter, A.; Devaux, C. Potent inhibition of HIV-1 replication by backbone cyclic peptides bearing the Rev arginine rich motif. *J. Biomed. Sci.* **2007**, *14*, 565-584.

56. Friedler, A.; Friedler, D.; Luedtke, N. W.; Tor, Y.; Loyter, A.; Gilon, C. Development of a functional backbone cyclic mimetic of the HIV-1 Tat arginine-rich motif. *J. Biol. Chem.* **2000**, *275*, 23783-23789.

57. Mills, N. L.; Daugherty, M. D.; Frankel, A. D.; Guy, R. K. An α-helical peptidomimetic inhibitor of the HIV-1 Rev–RRE interaction. *J. Am. Chem. Soc.* **2006**, *128*, 3496-3497.

58. Bryson, D. I.; Zhang, W.; McLendon, P. M.; Reineke, T. M.; Santos, W. L. Toward targeting RNA structure: Branched peptides as cell-permeable ligands to TAR RNA. *ACS Chem. Biol.* **2012**, *7*, 210-217.

59. Bryson, D. I.; Zhang, W. Y.; Ray, W. K.; Santos, W. L. Screening of a branched peptide library with HIV-1 TAR RNA. *Mol. BioSyst.* **2009**, *5*, 1070-1073.

60. Duggan, P. J.; Offermann, D. A. The preparation of solid-supported peptide boronic acids derived from 4-borono-l-phenylalanine and their affinity for alizarin. *Aust. J. Chem.* **2007**, *60*, 829-834.

61. Hoeg-Jensen, T.; Ridderberg, S.; Havelund, S.; Schaffer, L.; Balschmidt, P.; Jonassen, I.; Vedso, P.; Olesen, P. H.; Markussen, J. Insulins with built-in glucose sensors for glucose responsive insulin release. *J. Pept. Sci.* **2005**, *11*, 339-346.

62. Knott, K.; Fishovitz, J.; Thorpe, S. B.; Lee, I.; Santos, W. L. N-terminal peptidic boronic acids selectively inhibit human ClpXP. *Org. Biomol. Chem.* **2010**, *8*, 3451-3456.

63. Kettner, C.; Mersinger, L.; Knabb, R. The selective inhibition of thrombin by peptides of boroarginine. *J. Biol. Chem.* **1990**, *265*, 18289-18297.

64. Kinder, D. H.; Katzenellenbogen, J. A. Acylamino boronic acids and difluoroborane analogues of amino acids: Potent inhibitors of chymotrypsin and elastase. *J. Med. Chem.* **1985**, *28*, 1917-1925.

65. Matteson, D. S.; Sadhu, K. M.; Lienhard, G. E. R-1-acetamido-2phenylethaneboronic acid. A specific transition-state analog for chymotrypsin. *J. Am. Chem. Soc.* **1981**, *103*, 5241-5242.

66. Priestley, E. S.; De Lucca, I.; Ghavimi, B.; Erickson-Viitanen, S.; Decicco, C. P. P1 phenethyl peptide boronic acid inhibitors of HCV NS3 protease. *Biorg. Med. Chem. Lett.* **2002**, *12*, 3199-3202.

67. Snow, R. J.; Bachovchin, W. W.; Barton, R. W.; Campbell, S. J.; Coutts, S. J.; Freeman, D. M.; Gutheil, W. G.; Kelly, T. A.; Kennedy, C. A. Studies on proline boronic acid dipeptide inhibitors of dipeptidyl peptidase IV: Identification of a cyclic species containing a B-N bond. *J. Am. Chem. Soc.* **1994**, *116*, 10860-10869.

68. Baker, S. J.; Ding, C. Z.; Akama, T.; Zhang, Y. K.; Hernandez, V.; Xia, Y. Therapeutic potential of boron-containing compounds. *Future Med. Chem.* **2009**, *1*, 1275-1288.

69. Kane, R. C.; Bross, P. F.; Farrell, A. T.; Pazdur, R. Velcade: U.S. FDA approval for the treatment of multiple myeloma progressing on prior therapy. *The Oncologist* **2003**, *8*, 508-513.

70. Rock, F. L.; Mao, W.; Yaremchuk, A.; Tukalo, M.; Crepin, T.; Zhou, H.; Zhang, Y. K.; Hernandez, V.; Akama, T.; Baker, S. J.; Plattner, J. J.; Shapiro, L.; Martinis, S. A.; Benkovic, S. J.; Cusack, S.; Alley, M. R. An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing site. *Science* **2007**, *316*, 1759-1761.

71. Crumpton, J. B.; Zhang, W.; Santos, W. L. Facile analysis and sequencing of linear and branched peptide boronic acids by maldi mass spectrometry. *Anal. Chem.* **2011**, *83*, 3548-3554.

72. Kjems, J.; Calnan, B. J.; Frankel, A. D.; Sharp, P. A. Specific binding of a basic peptide from HIV-1 Rev. *EMBO J.* **1992**, *11*, 1119-1129.

73. Park, S. H.; Doh, J.; Park, S. I.; Lim, J. Y.; Kim, S. M.; Youn, J. I.; Jin, H. T.; Seo, S. H.; Song, M. Y.; Sung, S. Y.; Kim, M.; Hwang, S. J.; Choi, J. M.; Lee, S. K.; Lee, H. Y.; Lim, C. L.; Chung, Y. J.; Yang, D.; Kim, H. N.; Lee, Z. H.; Choi, K. Y.; Jeun, S. S.; Sung, Y. C. Branched oligomerization of cell-permeable peptides markedly enhances the transduction efficiency of adenovirus into mesenchymal stem cells. *Gene Ther.* **2010**, *17*, 1052-1061.

74. Tan, D. S.; Foley, M. A.; Stockwell, B. R.; Shair, M. D.; Schreiber, S. L. Synthesis and preliminary evaluation of a library of polycyclic small molecules for use in chemical genetic assays. *J. Am. Chem. Soc.* **1999**, *121*, 9073-9087.

75. Malan, C.; Morin, C. A concise preparation of 4-borono-l-phenylalanine (L-BPA) from l-phenylalanine. *J. Org. Chem.* **1998**, *63*, 8019-8020.

76. Jullian, M.; Hernandez, A.; Maurras, A.; Puget, K.; Amblard, M.; Martinez, J.; Subra, G. N-terminus FITC labeling of peptides on solid support: The truth behind the spacer. *Tetrahedron Lett.* **2009**, *50*, 260-263.

77. Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **1987**, *15*, 8783-8798.

## Chapter 6 Conclusions

The research described in Chapters 3–5 has successfully demonstrated the utility of medium-sized branched peptides for selectively binding to unique RNA tertiary structures. Our first attempt to generate a biased 3.3.3 branched peptide library and perform high throughput screening with natural L-amino acids resulted in the discovery of several low micromolar binders to the HIV-1 TAR RNA. A single false positive hit sequence (FL10) was identified from the pool of 16 hit sequences. One compound (FL4) achieved mid-nanomolar affinity for the target. Furthermore, FL4 demonstrated an approximately 10-fold stronger binding to the native TAR RNA sequence compared to three TAR RNA related structural variants (Bulgeless TAR, Tetraloop TAR, and Bulgeless Tetraloop TAR). A single point mutation in the bulge region of TAR (24U>C) that was not predicted to alter the tertiary structure of the RNA did not result in a loss of binding affinity to FL4. Taken together, these experiments suggest that FL4 is able to recognize the tertiary structure of TAR RNA, although the presence of excess tRNA in the binding buffer resulted in an equivalent loss of binding to TAR. Hill analysis of the **FL4** dot blot data suggested a single binding site on TAR that spans the apical loop and tri-nucleotide bulge region. The compounds from the 3.3.3 library were also cell permeable and well tolerated by HeLa cells at 1  $\mu$ M for 4 h.

Additional insights gained from this study included the use of PyOxim as the peptide-coupling reagent in order to avoid epimerization of the branched peptides during synthesis. It was also demonstrated that inclusion of a six-carbon linker (Ahx) adjacent

to fluorescein prevents the problematic acid mediated formation of fluorescein thiohydantoins during TFA deprotection of the branched peptides. Finally we demonstrated the importance of the branching unit. Truncation of **FL4** by removing the N-terminus linked through the  $N^{\epsilon}$  of the branching Lys residue resulted in over a 160-fold loss of binding affinity to TAR RNA.

The next library generation consisted of a 3.3.4 scaffold, and the size of the library was increased more than 10-fold (46,656 possible unique sequences). In addition, the inclusion of two boronic acid containing L-amino acid side chains further increased the diversity of the library. It was hypothesized that the empty p-orbital of boron would be able to interact with the RNA as a Lewis acid, and could potentially provide an additional intermolecular interaction with the RNA (reversible covalent bonding) that is not possible through natural amino acid side chains. Screening the library for binding to HIV-1 RRE-IIB RNA resulted in the discovery of 11 hit sequences; of which three were determined to be false positives by dot blot analysis.

Three 3.3.4 branched boronic acid peptides achieved binding affinities ( $K_d$ ) to RRE-IIB that were below 10 µM. The strongest binder discovered in the library, **BB1-1**, achieved a  $K_d$  of  $1.4 \pm 0.4$  µM. Although, none of the compounds from the 3.3.4 library matched the mid-nanomolar affinity acheived by **FL4** from the 3.3.3 library toward HIV-1 TAR, **BB1-1** of the 3.3.4 library was shown to maintain its binding affinity for HIV-1 RRE-IIB in the presence of up to 1000-fold excess tRNA. Furthermore, **BB1-1** was found to bind to HIV-1 RRE-IIB with up to 70-fold stronger affinity compared to six RRE-IIB related structural variant RNAs. Hill plot analysis demonstrated that the interaction between RRE-IIB and **BB1-1** was non-cooperative. Evidence for a

multivalent interaction between **BB1-1** and RRE-IIB was provided by three structural variants of the branched boronic acid (**LBB1-1-A**, **LBB1-1-B**, and **LBB1-1C**). Each of these peptides exhibited reduced binding affinity to the RRE-IIB, however deletion of the C-terminus (**LBB1-1-C**) completely abolished binding to the RNA target. The role of the boronic acid moiety was also examined in two different hit sequences (**BB1-1** and **BB1-8**). Similar structures that lacked the boronic acid moiety (**CBB1-1** and **CBB1-8**) each displayed poorer binding affinity to the target RNA compared to the parent compound, however, binding was completely lost in **CBB1-8**. These results suggest that boron plays an integral role to the overall affinity for the RRE-IIB. Intriguingly, placement of a fluorine atom in the ortho position to boron in **BB1-1** resulted in compound **CFBB1-1**, which was found to have a measurably improved (~2-fold) affinity toward the RNA target. This result may suggest that the interaction between boron and the RNA can be tuned through electronic effects; however, more experiments are required to verify the scope and generality of this observation.

Through the use of RNase protection assays (RNA footprinting) it was shown that **BB1-1** likely binds through the bulge region of RRE-IIB as well as the stem region near the tri-nucleotide apical loop. Ribonucleotide bases U7, C21, and U36 were each protected by **BB1-1** from cleavage with RNase A, while bases G10, G17, and G18 were protected from RNase V1. In addition, **BB1-1**, **BB1-6**, and **BB1-8** were cell permeable in HeLa and A2780 cells and were non-toxic up to 100  $\mu$ M for 4 h or up to 10  $\mu$ M for 24 h. This study provided the first example of peptides featuring boronic acid moieties designed for RNA binding. Current efforts are toward establishing the therapeutic potential of these compounds in cell-based assays.

Future peptide libraries will be designed to achieve stronger binding affinity to the RNA target in the low nM range. We have demonstrated that additional positive charge does not always provide increased binding affinity with branched peptide libraries, as discussed in Chapters 3–5. One variable that we have yet to explore is conformational restriction in our compounds. The work with  $\beta$ -hairpin cyclic peptides detailed in Chapter 2 demonstrates that increased structure generally results in increased binding affinity due to preorganization.<sup>1</sup> It may be possible to introduce branching side chains off of a  $\beta$ -hairpin peptide to provide a rigid core (or palm) with multivalent "fingers" incorporating boronic acid residues. The main hurdle with this approach would be deconvolution of the hits via MALDI-MS/MS analysis. An alternative approach for introducing rigidity into the branched peptides would be to alter the peptide backbone of these compounds. Chiral oligomers of pentenoic amides (COPAs) have demonstrated conformational restriction due to 1,3-allylic strain, and libraries as large as 160,000 sequences have been screened for binding to proteins.<sup>2</sup> These compounds are also amenable to sequence deconvolution by mass spectrometry and have not been explored as RNA ligands. This approach could be the next logical step to take; resynthesizing the boronobranched 3.3.4 library using the COPA backbone would show the consequence of introducing conformational restriction in branched boronic acid RNA ligands.

## References

1. Athanassiou, Z.; Dias, R. L.; Moehle, K.; Dobson, N.; Varani, G.; Robinson, J. A. Structural mimicry of retroviral Tat proteins by constrained  $\beta$ -hairpin peptidomimetics: Ligands with high affinity and selectivity for viral TAR RNA regulatory elements. *J. Am. Chem. Soc.* **2004**, *126*, 6906-6913.

2. Aquino, C.; Sarkar, M.; Chalmers, M. J.; Mendes, K.; Kodadek, T.; Micalizio, G. C. A biomimetic polyketide-inspired approach to small-molecule ligand discovery. *Nature Chem.* **2012**, *4*, 99-104.

# **Appendix 1** Citations of Copyrighted Works

Figure 1.1 Reprinted by permission from Macmillan Publishers Ltd: [Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **2000**, *407*, 340-348], Copyright 2000.

Figure 1.2 Reprinted by permission from Elsevier: [Zaman, G. J. R.; Michiels, P. J. A.; van Boeckel, C. A. A. Targeting RNA: New opportunities to address drugless targets. *Drug Discov. Today* **2003**, *8*, 297-306], Copyright 2003.

Figure 1.3 Reprinted by permission from Bentham Science Publishers Ltd: [Yuan, C.; Xinyong, L.; De Clercq, E. Cessation of HIV-1 transcription by inhibiting regulatory protein Rev-mediated RNA transport. *Curr. HIV Res.* **2009**, *7*, 101-108], Copyright 2009.

Figure 1.4 Reprinted by permission from Elsevier: [Martick, M.; Scott, W. G. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **2006**, *126*, 309-320], Copyright 2006.

Figure 1.5 Reprinted by permission from Elsevier: [Edwards, T. E.; Ferre-D'Amare, A. R. Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure* **2006**, *14*, 1459-1468], Copyright 2006.

Figure 1.6 Reprinted by permission from Macmillan Publishers Ltd: [Esquela-Kerscher, A.; Slack, F. J. Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer* **2006**, *6*, 259-226], Copyright 2006.