Targeting RNA Structures that Control Protein Synthesis and RNA Stability

Matthew J. Walker

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Gabriele Varani

Jesse Zalatan

Andrew Hsieh

Program Authorized to Offer Degree:

Department of Chemistry

University of Washington

Abstract

Targeting RNA Structures that Control Protein Synthesis and RNA Stability

Matthew J. Walker

Chair of the Supervisory Committee:

Gabriele Varani

Department of Chemistry

RNA structures play a pivotal role in many biological processes and the progression of human disease, making them an unexploited target for therapeutic development. The first chapter of this thesis reviews features related to targeting RNA structures and to the subsequent topics of this thesis: disease-associated RNA classes, advantages and disadvantages of different RNA-binding chemistries (e.g. small molecules, peptides, engineered proteins), NMR-based methods for RNA screening and structure determination, and a future perspective on the field.

Chapter 2 describes the results of my research on specialized translation initiation, whereby protein synthesis is controlled by eukaryotic initiation factor 3 (eIF3) recognizing RNA structures within the 5'-untranslated regions of certain genes to regulate translation rates of specific mRNAs. I examined this mechanism by establishing the structural basis for eIF3 recognition of the c-JUN 5'-UTR cis-regulatory element (SL1). SAXS modeling and NMR structure determination identified similarities to the way eIF3 recognizes RNA motifs within internal ribosomal entry sites (IRES) in the Hepatitis C Virus (HCV) RNA, suggesting

mechanistic similarities. This work establishes RNA structural features involved in c-JUN specialized translation initiation and provides a foundation to search for small molecules inhibitors of aberrant expression of the proto-oncogenic c-JUN protein.

Chapter 3 reviews strategies for designing cyclic β-hairpin peptidomimetics targeting pharmaceutically relevant structured RNAs such as HIV-TAR and pre-microRNA-21 (pre-miR-21). This rational is built into several pre-microRNA targeting projects discussed in chapter 4, where I describe the development of a pre-miR-21 processing assay to characterize inhibitors ranging from cyclic peptides to engineered RNA-binding motifs (RRM*). My results demonstrate that many cyclic peptides disrupt efficient pre-miR-21 processing, as demonstrated by the formation of novel reaction intermediates only produced in the presence of cyclic peptide inhibitors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First, I would like to recognize Dr. Matt Shortridge for acting as my post-doc mentor throughout my PhD. Thanks for teaching me techniques around the lab, guiding me through NMR experiments and structure calculations, editing my NSF proposal, and providing valuable career advice.

Second, I would like to express my appreciation for Dr. Gaberiele Varani for funding my research, editing my manuscripts and exams, and hilarious banter ranging from basketball to hip-hop. Thank you for giving me a PhD experience where I could master research, but also learn how to think effectively and independently.

I would also like to recognize other lab members such as Tom Pavelitz for being an invaluable lab manager by always having answers for molecular biology related questions and great conversation about politics. Many thanks go to Dr. Wen Yang for acting as my protein purification guru, and to Rajan Paranji for late nights on the 800 troubleshooting pulse programs that I would have been completely lost figuring out on my own.

Over these past 5 years my friends have been an essential source of motivation and reasons to escape from the lab. Fellow graduate students Josh Cochran, Alisha Jones, Giuseppina Pisignano and post-bachelor fellows Dreycey Albin and Angelique Amado were always there for great company and beers at Shultzy's. Many thanks to my softball team, Frank's Little Sluggers, for being great friends and not getting too mad when I went yard after we met the HR cap. Last but not least, an enormous amount of gratitude for Devin Colyer for being an incredible friend, visiting Seattle even during its rainiest months, and listening to my hours of ranting about grad school woes.

I could have never accomplished my degree without the help and support from my family. I cannot thank my mom and dad enough for providing the occasional extra bit of cash to survive an expensive Seattle or offering advice in tough situations like what to do when dealing with frustrating landlords. I want to thank my sister as well for being a great friend and someone who I could always call up for answers to non-science questions; it is also great that we get to share 2019 as our graduation year. Many family members have provided motivational and financial support over the years for which I am thankful for as well: Tante Elfriede, Oma and Opa, Nanny, Aunt Pam, Aunt Carmen and Jen.

This PhD could not have been accomplished without the day-in and day-out support from my girlfriend, Elise Matza. You were an enormous help by listening to my practice talks months before my defense and I appreciate that you always let me bounce ideas or rant about troublesome experiments. Thank you for your love, patience, comic relief, and astute advice during my most challenging moments.

DEDICATION

To my Opa for years of clever jokes, stories from Holzmaden, and the support that has made my education and especially this PhD possible.

Chapter 1 – Targeting Structured RNA in Human Disease

Section 1: Why pursue RNA as a drug target?

As pharmaceutical companies progressively build drug discovery programs into the 21st-century, the field has welcomed novel targeting strategies including the prospect of using structured, disease-associated RNA molecules as viable drug targets. RNA molecules regulate healthy cellular processes (e.g. transcription, splicing, mRNA transport, translation, etc) and are therefore strictly regulated transcriptionally and post-transcriptionally, typically through binding with other RNAs and RNA binding proteins. RNAs and RNA-protein interactions regulate viral replication or the expression of proto-oncogenes, and are mis-regulated in many infectious and chronic diseases (Cooper, Wan, & Dreyfuss, 2009; Esteller, 2011) making these RNA structures and RNA-protein surfaces an untapped source of potential drug targets (Burnett, & Rossi, 2012; Ling, Fabbri, & Calin, 2013). Thus, when considering the growing appreciation for the number of functionally relevant non-coding and functional RNAs in the human genome (Li et al, 2013), combined with their impact on disease, targeting RNA potentially yields a wealth of therapeutically relevant drug targets (Cooper et al, 2009; Matsui et al, 2017).

1.1 Non-coding RNAs

Only 2-3% of the human genome codes for proteins, while non-coding RNAs (ncRNA) represent nearly 40% of the genome with >>10,000 ncRNA genes already identified (Li et al, 2013), yet how many of these RNAs are actually functional is still widely debated. Multiple examples have established various forms of ncRNAs are functional and associated with human disease, as demonstrated by the use of microRNAs (miR) as effective diagnostics for cancer and

viral infections based on specific patterns of down or up-regulation in disease (Giordano et al, 2013; Nana-sinkam et al, 2013). For example, the non-coding microRNA-21 (miR-21) has been shown to be highly expressed in almost all cancers, where it promotes disease progression by, among other mechanisms, silencing of the PTEN/PI-3 K/Akt pathway (Bullock et al, 2013). Target validation efforts for pre-miR-21 have shown reversal of tumor formation in mouse models of cancer as well (Shortridge et al, 2017; Naro et al, 2018; Cheng et al, 2015), demonstrating its therapeutic potential as a target for cancer treatment.

Long non-coding RNAs (lncRNA), a large class of ncRNAs that extend beyond 200 nucleotides in length, have been associated with various developmental defects and cancer pathways (Schimitt et al, 2016). For instance, chromosome 8 is highly enriched with cancer-associated single nucleotide polymorphisms (SNP), several of which are implicated in the expression of cancer-associated lncRNAs, including CCAT2 in colorectal cancers and PCAT-1 in prostate cancer (Ling et al., 2015). Interestingly, despite their lack of sequence conservation across species (Taft et al, 2007), their RNA structure is arguably more highly conserved than their sequence (Mathews et al, 2010), supporting the exploration for compounds that can redirect misfiring biological pathways caused by these SNPs.

These two classes of ncRNAs are some of the many varieties of ncRNAs continuously being discovered, demonstrating the growing role of ncRNA in disease and highlighting the potential of ncRNAs for therapeutic targeting.

1.2 5'-Regulatory Elements within mRNA UTRs

5'-UTRs, the regions of mRNA that precede the protein coding sequence, provide numerous examples of the formation of cis-regulatory elements that serve regulatory functions

by activating or repressing translation rates of their downstream coding sequence (Wilkie et al, 2003), while 3'-UTRs, the regions of genes that follow, often control mRNA stability by providing landing sites for microRNAs and RNA-binding proteins.

Riboswitches provide some of the best characterized examples of translational control elements, with often well-established structure-function relationships, located in bacterial and, more occasionally, eukaryotic 5'-UTRs (Cheah et al, 2007; Vitreschak et al, 2004). They function by binding small molecules ligands, typically metabolites important for maintaining bacterial homeostasis (e.g. adenine, thiamine pyrophoshphate, tetrahydrogfolate, Mg^{2+}); small molecule binding induces a conformational change in the RNA structure, revealing or hiding the AUG start codon thereby regulating translation of the downstream coding sequence (Serganov et al, 2012). Riboswitches are an example of exploitation of mRNA structure through evolutionary processes to tailor translation rates based on an organism's surroundings, which also provide proof of concept that RNA-small molecule interactions can regulate translation by targeting the 5'-UTRs of cellular mRNAs. Thus, it is hypothesized that important regulatory mechanisms that harness RNA structure could be perturbed by ligand binding to induce therapeutically favorable effects.

G-quadruplexes are another example of inhibition of translation through stabilization of cis-regulatory elements, where G-repeat sequences form translation-impeding, ultra-stable structures that are often found in mRNAs coding for proliferative transcripts, such as the proto-oncogenic NRAS mRNA (Kumari et al, 2007). Evidence showing that trans-acting factors are implicated in RNA G-quadruplex regulation supports the conjecture that these structures could be targeted by small molecule or peptide targeting efforts (Bugaut et al, 2012).

Alternative translation initiation pathways are an area of particular interest, as these mechanisms drive translation rates for viruses and proliferative genes associated with cancer, which are controlled by mRNA structures interacting with ribosomal subunits and eukaryotic translational initiation factors (eIF). Viral internal ribosomal entry sites (IRES), for example, have been targeted based on the aforementioned rational. Eukaryotic mRNAs, such as the proto-oncogenic c-Myc 5'-UTR, have become some of the most sought-after targets for new RNA drug companies (Dibrov et al, 2012; Shi et al, 2016). The recently discovered specialized translation initiation mechanism is another example, which requires similar 5'-UTR structures as IRES transcripts, but differs from IRES-mediated initiation by its cap-dependence and its common occurrence in cellular proliferative transcripts (e.g. c-JUN) (Lee et al, 2015; Lee et al, 2016). With each transcript differing from others in sequence and secondary structure, cis-regulatory RNA structures that participate in specialized translation initiation provide many new potential RNA drug targets.

Section 2: The Chemistry of RNA Binding Molecules

High-throughput screening of RNA structures allows the exploration of vast amounts of chemical space, accelerating the process of elucidating what chemical moieties are most effective for targeting particular RNAs. Rational, structure guided design of candidate molecules based on a physical model offers insight and prediction for what drives an RNA-targeted small molecule's affinity and specificity. Therefore, implementation of each may be an effective strategy for pursuing RNA drug target.

In this section, chemistries using high-throughput screening and structure-based design for targeting three-dimensional RNA structures are reviewed. Oligonucleotide chemistry for targeting RNA primary sequence is briefly covered as well.

2.1 Oligonucleotides

Decades of research efforts focused on targeting RNA sequence with antisense oligonucleotides have recently seen breakthroughs and some as FDA approved therapeutics (Mullard, 2019; Bennett et al, 2019). Oligonucleotides therapeutics target RNA sequence using short, modified ribonucleotide stretches (8-50 residues) that bind to RNA through Watson-Crick base pairing; mechanisms of inhibition involve either blocking RNA interference (RNAi) seeds, blocking of splice sites to redress unfavorable patterns of pre-mRNA splicing, promoting degradation of the RNA through endogenous enzymes such as RNaseH-mediated cleavage (Bennett et al, 2010), or even treating lncRNAs associated diseases (Wahlestedt et al, 2013). Although oligonucleotides suffer from bioavailability and stability issues (Yu et al, 2016), it is clear that when cellular uptake and targeting occurs, oligonucleotide chemistries can functionally inhibit their target RNA sequence.

2.2 Small Molecules

The discovery of small molecules targeting RNA in disease dates back to the mid-20th century with the discovery of the antibacterial streptomycin, one of several aminoglycoside-based antibacterials that reduce translational accuracy, resulting in bacterial death (Jones et al. 1944; Wang et al, 1995). However, not until breakthroughs in RNA foot-printing techniques occurred, were aminoglycosides shown to directly interact with rRNA (Moazed, Noller, 1987). Around the same time, a tremendous amount of enthusiasm was generated for *in vitro* selection and evolution of RNA sequences, called "aptamers" to generate a wide range of RNA structures that bound to proteins and small-molecules (Ellington et al, 1990; Robertson et al, 1990). In 2002, riboswitches were discovered and soon after ligands-bound X-ray crystal structures provided some of the first structure-based rationales for RNA recognition of small molecules (Serganov et al, 2013). Various RNA targeting strategies are now being explored, ranging from entirely structure-based efforts to high-throughput screening approaches (Fatemi et al, 2015; Shortridge et al, 2015; Rizvi et al, 2017; Donlic et al, 2018; Warner et al, 2018).

2.2.1 Structure-Based RNA-Targeted Drug Discovery

Three-dimensional structure-based design methods such as NMR and X-ray crystallography provide atomic resolution, revealing small molecule interactions with the RNA which can be exploited for optimization. RiboTargets, Inc., a pharmaceutical company from the late 1990's located in Cambridge, UK, embarked on one of the first large-scale small molecule optimization projects by implementing a combination of NMR structure determination/characterization and design, in conjunction with a sustained medicinal chemistry effort, to target the Tat-TAR interaction of human immunodeficiency virus (HIV) (Davis et al,

2003). At the time, the importance of charge interactions in the binding of small molecules to RNA had long been recognized, and it had been generally assumed that electrostatic contacts with the backbone phosphates were the dominant interactions. By using a series of structures of TAR with small molecule ligands of increasing activity, (Davis et al, 2003) identified electrostatic "hot spots" associated with steep gradients of strong electronegativity in the major groove of the RNA, near hydrogen bond donors such as uracil O4s, adenosine N7's, etc. This led to an effective structure activity relationship (SAR) with a heterocycle-focused library, and eventually a compound, RBT550, was discovered with low nM (K_i = 40 nM) activity for inhibiting the Tat-TAR interaction (Murchie et al, 2004) and attractive pharmacological characteristics (Matsson et al 2016). These observations demonstrated that electrostatic phosphate interactions only partially describe what drives RNA-targeted ligand affinity and potentiates RNA-targeted specificity.

The Tat-TAR interaction is still commonly pursued, as demonstrated by (Abulwerdi et al, 2016) where a small molecule microarray was used to discover series of molecules specifically targeting the trinucleotide bulge (U23, C24, U25) to investigate functional groups to be considered during the RNA-small molecule design process: amine positioning required for binding, oxadiazole analogs being key for Tat displacement, the preference for small aliphatic groups. Key to the success of this project was the combined use of NMR and biochemical assays. NMR TOCSY experiments were used to quickly screen a library of compounds for binding while also providing rational for SAR development based on nucleotide-specific chemical shift perturbations detected only from high affinity ligands. IC_{50} HIV replications assays that followed provided means to validate TAR binders discovered in NMR studies as inhibitors by identifying compounds that inhibit *in vitro* HIV replication in addition to binding HIV-TAR.

X-ray crystallography can provide rapid atomic resolution for lead compounds when RNAs and RNA-small complexes form rigid structures that crystallize. Antiviral discovery efforts by Thomas Hermann's group at the University of California San Diego targeting the hepatitis C virus (HCV) IRES subdomain IIa showed that benzimidazole derivatives induce a conformation change that essentially bend the RNA at an enlarged bulge, and this change locks subdomain IIa and IIb into a novel structure that separates subdomain IIb from the ribosome and ultimately leads to inhibition of IRES-driven translation in HCV-infected cells (Parsons et al, 2009; Dibrov et al, 2012). Here is an example where RNA is targeted by manipulation of its secondary structure. Instead of competing with the IIb recognition site for the 40S ribosomal subunit, high affinity benzimidazole derivatives inhibit allosterically by binding to the proximal IIa subdomain. This induces the IIb stem-loop structure to bend significantly, resulting in a population of ligand-bound IIa/IIb with an altered structure unrecognizable by the 40S ribosomal subunit.

2.2.2 High-Throughput Screening

High-throughput screening efforts involve assaying large small molecule libraries (10^3-10^6) and detect binding *in vitro* or in cells to a target of interest (Wigglesworth et al, 2015). Discovery of a compound with inhibitory properties even by qualitative detection or phenotypic screens initiates a targeting project by giving medicinal chemists and biophysicists a compound to optimize. Merck recently disclosed ribocil, a highly selective chemical modulator of bacterial riboflavin riboswitches, which was discovered in a 57,000 synthetic small molecule phenotypical antibacterial screen, which measured *E. coli* growth inhibition due to perturbed riboswitch function (Howe et al, 2015).

Fluorescent reporter systems (e.g. FRET-based) have been implemented for *in vitro* screening, as shown by (Murchie et al, 2003) in the discovery of RBT550 which used a solution state *in vitro* FRET competition assay. However, fluorescent assays suffer from drawbacks such as compound interference, quenching, and autofluorescence (Janzen et al, 2014; Imbert et al, 2007).

An alternative to FRET for assaying inhibition of miRNA processing was recently reported in (Lorenz et al, 2015) using a catalytic enzyme-linked click chemistry assay or cat-ELCCA. Their high-throughput cat-ELCCA was implemented for discovery of selective small molecule ligands for pre-miR-21 involving 5'-end immobilization of a biotinylated pre-miR-21 substrate containing a click chemistry handle in the terminal loop. Mixing with Dicer cleaves the loop and no chemiluminescent is detected, whereas inhibition by small molecule prevents cleavage and chemiluminescent signal is observed. ~33,000 natural product extract libraries were assayed using this approach. A family of tetracycline-like molecules were discovered with low μM IC_{50} concentrations when targeting pre-miR-21.

One the largest RNA-focused screening technologies involves combinations of automation and mass spectrometry. Merck's automated ligand detection system (ALIS) coupled with affinity-selection mass spectrometry (AS-MS) technique allows for selective detection of small molecule–RNA interactions (Rizvi et al, 2018). By coupling fast (< 20 s) size exclusion chromatography to separate free compounds from RNA–compound complexes, screening of 500,000 compounds/day was achieved; this report emphasized the advantage over other techniques in its ability to cover a large range of chemical space avoiding bias in chemical libraries by sheer increase in the size of the library screen.

2.2.3 Fragment-Based Screening

Fragment based screening utilizes low molecular weight compounds (MW <250 Da) to facilitate larger explorations of chemical space through the discovery of low affinity hits, which are often abundant owing to their chemical simplicity and low entropic penalty upon binding (Hann et al, 2001).

Fragment-based methods are effective in targeting proteins and have had some success for targeting RNA as well (Congreve et al, 2008). NMR is a powerful tool for screening fragments due to its medium-to-high-throughput (10^2-10^3 compounds per screen (Carr et al, 2005)) and ability to detect even weak ligands (Erlanson et al, 2016). WaterLOGSY experiments for instance have been implemented to screen thiamine base fragments against the thiM riboswitch (Chen et al, 2010). Another approach is utilizing saturation transfer difference, as shown by (Davidson et al, 2011) where an inverse-pharmacophore strategy was applied involving screening fragments against HIV-TAR already bound to a low μM-binding chemical probe. The probe locks the RNA into a conformation capable of binding other fragments, while simultaneously allowing the identification of proximal binding fragments by ligand-based NMR detected through a 10% minimum required different STD change. Success has also been found in studies elucidating riboswiches suitable for reliable crystallization and optimizing native ligands to understand their recognition and potentially develop new inhibitors (Deigan et al 2011). Co-crystallization of fragments-bound structures involves either soaking unbound or natively bound riboswitches with fragments of interest, allowing rapid structure generation to infer rationale for future design of focused libraries (Warner et al, 2014).

2.3 Peptides

Targeting the large surface area of RNA-protein interactions with small molecules faces issues similar to targeting protein-protein interactions (Smith et al, 2012). RNA-protein interaction sites are wide and often shallow, commonly involving multiple RNA binding motifs to ensure affinity and specificity (Lunde et al, 2007). Thus, effective targeting for some RNA targets requires maximizing multiple functional groups to compete with the higher molecular weight native binding protein. RNA-binding peptides provide an alternative solution with greater molecular weight and surface area, and the added capability of using information on the native binding protein for rational optimization of binding activity.

2.3.1 Structure-Based Rational

Using the structure of native binding protein to guide the discovery of peptides that bind to RNA is surprisingly uncommon, yet it is a highly effective strategy to discover even ultra-potent ligands for RNA. The Varani Lab at the University of Washington has spent a decade using peptidomimetics that bind to viral elements critical for replication and cancer associated pre-microRNAs. Short peptides sequences can be grafted onto cyclic β-hairpin scaffolds, and side-chain optimization can be carried out with small-to-medium sized libraries. This approach has led to several successes, including the discovery of JB-181, a 30 pM binding ligand for HIV associated TAR. Chapter 3 reviews this topic in-depth (Walker et al, 2019).

2.3.2 High-Throughput Screening

High-throughput screening for the discovery of RNA binding peptides has had some success, enabling investigators to rapidly analyze molecular interactions between peptide and

RNA on larger scale. Microarrays for example can be powerful tools by acting as an organizing approach for developing structure-activity relationship information and rapidly identifying critical amino acid residues (Fernandes et al 2009).

Peptoids provides the first example of peptide-focused high-throughput targeting of RNA. Peptoids are advantageous from a therapeutic point of view as their sidechains branch from the amide nitrogen instead of the alpha carbon, preventing proteolysis. (Hamy et al, 1997) generated very large libraries of short peptoids targeting HIV-TAR using a "split and mix" process involving step-wise additions of different residues to a growing peptoid on resin, and pooling the peptoids into subdivisions based on the residue added. This deconvolution approach produced millions of 5-8 residue peptoids with sequence designed to partially mimick the TAR protein binding partner, Tat. The subdivisions of these peptoids could be assayed by electrophoretic mobility shift assays, guiding the pooling process to avoid residues additions that were unfavorable for binding.

Through advances in solid-state coupling peptide chemistries, many high-throughput projects targeting RNAs with peptides have implemented microarrays for screening (Chirayil et al, 2009; Pai et al, 2012). However, a drawback is the limited number of peptides that can be screened relative to higher-throughput techniques and potential misleading information due to peptide immobilization on solid surfaces masking part of the potential binding interface (Fernandes et al 2009).

In an attempt to target the cancer associated pre-miR-21, (Chirayil et al, 2009) utilized a combinatorial peptoid synthesis scheme similar to the aforementioned TAR-targeted study. Cysteine-terminated peptoids were printed on slides pretreated with N-(p-maleimidophenyl) isocynate; cysteine oxidation couples the peptoids to the maleimido moiety readily. Peptides

were organized by sequence subdivisions and detection of RNA binding was done through straightforward 5'-end labeling with the fluorescent dye Dy547. This qualitative screen produced hits that were further optimized to 1.9 μM affinity for pre-miR-21 with 20-fold discrimination from similar RNAs. (Pai et al, 2012) introduced a more sophisticated peptide microarray system to estimate hundreds of dissociation constants of RNA-peptide complexes in a single assay. Conjugation chemistry with hydrazide-linkages limited the number of peptides per plate, but the quanatative capability of the assay facilitated profiling interactions of six hairpin RNAs (HCV IRES subdomain IV, HIV RRE, HIV TAR, 16S-rRNA, and thymidylate synthase mRNA) with 111 peptides, containing systematic amino acid replacements or deletions.

RNA-targeting with branched peptides is another strategy that has met some success when applied in a high-throughput screening format (Bryson et al, 2009; Dai et al, 2018). Branched peptides contain naturally occurring amino acids that are displayed in unique molecular conformations by side-chains acting as nucleation points for new peptide backbones to grow. (Bryson et al, 2009) demonstrated the synthetic methodology for preparing thousands of branched peptides by a "mix and pool" deconvolution approach. A key design step in their synthetic scheme is to retain bead coupling while deprotecting sidechains, using a photoactivatable linker. Using this approach for library synthesis, (Bryson et al, 2009) implemented a fluorescence polarization assays using a 5'-fluorophore labeled HIV-TAR and identified several low μM binders. Almost a decade later, this same synthesis scheme was scaled up to target the HIV associated Rev-RRE interaction with a 46,656 member library composed of unnatural amino acids, revealing several hits against RRE IIB RNA (Dai et al, 2018). By utilizing sidechains that mimic nucleotide base chemistries (pyrimidine, purine heterocycles) a low nM peptide was discovered with cellular inhibitory activity. SAR was also developed by

utilizing the chemical probing technique Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) (Siegfried et al, 2014), which showed secondary structure changes and nucleotides stabilized upon binding.

2.4 Engineered Proteins

Effectively designing new RNA-binding proteins has the potential to generate stable proteins that specifically recognize one of the thousands of disease-associated RNAs using a sequence-based approach built on developing a recognition code for protein-RNA interactions (Chen et al, 2013). Multiple studies have focused on bioinformatic analysis of recognition sequences, binding affinities, and correlated specificity rankings to identify recognition codes (Alipanahi et al, 2015; Jankowsky et al, 2015); however, in terms of rationally designing proteins that bind RNA de novo, the field is still at a very early stage.

Puf proteins are foundational in the field of RNA-binding protein design as their stable scaffold can be designed to recognize any RNA sequence (Chen et al, 2013). (Cheong et al, 2006) used crystal structures of human Puf1 bound to various RNA sequences to elucidate a recognition code for this class of protein. The protein recognizes RNA through hydrogen bonding and π-stacking to the Watson-Crick faces for a minimum of 8 consecutive nucleotides (Chen et al, 2013). The designed proteins achieve sub-picomolar binding affinities and have the potential for antisense oligonucleotide equivalent therapeutics. A caveat is that Puf1 recognition is limited to completely unfolded RNA, and even partially base-paired nucleotides are non-amenable to Puf1 recognition.

One of the few examples of an engineered RNA-binding protein capable of recognizing RNA three-dimensional structure involves the human splicing factor RbFox2. This is a unique

RNA recognizing motif (RRM) protein, due to its high affinity; typically, RRM domains exhibit high nM to low μM K_d's for their target RNA requiring multiple motifs to accomplish efficient recognition, whereas Rbfox2 binds pre-miR-20b with low nM affinity (Lunde et al, 2007; Kuroyanagi, 2009). This unique property is what gave (Chen et al, 2016) the idea to redirect Rbfox specificity to a therapeutically valuable microRNA while still retaining the high affinity. Structure determination of wildtype Rbfox bound to pre-miR-20b acted as a starting place for structure-based design. Inspection of the Rbfox—pre-miR-20b interface provided the atomic resolution required to rationally explain what drives Rbfox's high affinity and specificity for the pre-miR-20b apical loop. Using this insight, a series of engineered proteins were developed with reduced affinity for pre-miR-20b, and increased specificity for oncogenic pre-miR-21. "Fox-RRM*" was considered the final design which bound pre-miR-21 with a ~13 nM K_d and exhibited inhibition of pre-miR-21 processing by Dicer when assayed *in vitro* at uM concentrations. This demonstrated that through structure-based design and rigorous mutational analysis, specificity can be engineered into RNA-binding proteins, suggesting therapeutic potential for targeting other diseased associated miRNAs and other types of ncRNA with this approach.

RRMs evolved in the laboratory produce ligands with high affinity, suggesting this approach could be applied the discovery of RNA-binding proteins (Blakeley et al, 2013, Crawford et al, 2016). (Belashov et al, 2018) utilized the structure of the complex (K_d = 32 pM) between hairpin II of the U1 snRNA and RRM1 of the U1A as starting point and retargeted the protein towards HIV-TAR using saturation mutagenesis and phage display. The crystal structure of the resulting low nM binding lead provided high enough resolution for researchers to realize a portion of the RRM β-sheet contributed to the affinity. With this rational, they synthesized β-

hairpin mimics with low μM affinity and capable of inhibiting HIV transcription. This seminal effort demonstrates the success in combining high-throughput techniques to screen large sequence space for establishing hit discovery for proteins, and for combining a protein engineering approach with structure-based peptidomimetic peptide designs.

Section 3: Characterizing RNA-Ligand Interactions by NMR Spectroscopy

RNA is intrinsically flexible resulting in motion occurring on multiple times scales; thus, its structural characterization can be tasking, and in some cases unlikely to be achieved. High resolution X-ray crystal structure are in some cases obtainable for highly stable RNAs such as tRNA or rRNAs (Correll et al, 1998; Vicens et al, 2003; Korostelev et al, 2006), ribozymes (Ferré-D'Amaré et al, 1998; Scott et al, 1995; Rupert et al 2001) and riboswitches (Batey et al, 2004; Garst et al, 2008); however, the technique often lacks resolution for parts of the molecule that are unpaired or at binding interfaces and many RNAs simply cannot be crystallized. Solution-state NMR is better suited to probe flexible RNAs, with wide selection of experiments available to facilitate structure determination of somewhat flexible regions and to measure dynamics on multiple timescales (Barnwal et al, 2017; Marušič et al, 2019).

3.1 Assaying RNA-Ligand Interactions

One of the simplest methods for assessing binding of a ligand to an RNA is chemical shift perturbation (Williamson et al, 2013). Straightforward 1D proton spectra can qualitatively indicate conditions of fast, intermediate, and slow exchange kinetics between the free and bound state of a molecule based on inspection of peak broadening and appearance (Kovrigin et al, 2012). Imino protons located downfield in the spectrum (15-10 ppm) in a region free of significant spectral overlap are key indicators of subtle to large RNA conformational changes due to their participation in base-pairing (Mayer et al, 2006). This is advantageous from a screening standpoint as peak broadening for individual nuclei reports on changes in chemical environment due to structural rearrangements induced by ligand binding. In (Shortridge et al, 2018), a cyclic peptide bound HIV-TAR with a 30 pM K_d by forming a stabilizing base-triplet;

inspection of the 1D proton spectrum when the peptide is titrated to a 1:1 molar ratio with TAR demonstrates slow-exchange kinetics for imino nuclei near the interaction, and included a novel imino resonance as a result of the triple base-pair formation.

Many examples have been reported of the use of NMR saturation techniques to detect and characterize the binding of low molecular weight compounds to RNA (Dalvit et al, 2001; Marcheschi et al, 2009; Chen et al, 2010; Davidson et al, 2011). Saturation transfer difference (STD) experiments work by selectively applying a saturating radiofrequency pulse to a narrow frequency window in a region where no ligand resonances reside, but RNA resonances are present (Calabrese et al, 2019). The magnetization is then transferred to the ligand, if bound, and results in broadening of ligand resonances local to the interaction site. However, this approach is limited by the large MW of some RNAs (> 20kDa) resulting in shorter correlation times (Tc) and poor saturation transfer. The WaterLOGSY experiment differs slightly by involving indirect saturation of the RNA and selective saturation of bulk water protons (Calabrese et al, 2019). Saturation of water molecules is transferred to the RNA and then to the ligand bound to the RNA. Spectra for this experiment differ from STD experiments in that binding is detected by inversion of the peak phase as opposed to intensity attenuation.

The Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence allows measurement of T2 relaxation times (Hajduk et al, 1997). It is frequently applied for studying RNA structural dynamics (Bothe et al, 2011), but is also a reliable approach for screening low molecular weight ligand binding to RNA (Kloiber et al, 2011; Garavís et al, 2014; Calabrese et al, 2019). CPMG techniques utilize pulse programs to measure a range of differences in transverse relaxation times (T_2), which are directly dependent on the molecular rotational correlation time (τ_c) or molecular tumbling, representing the molecular weight of an RNA or RNA-ligand complex (Levitt, 2001).

Small molecules relax slowly (long T_2, seconds) exhibiting sharp peaks, but if bound to the RNA, which is inherently larger in molecular weight, broader peaks with decreased intensity are observed. Pulse programs are tailored on the qualitative knowledge of how well the ligand bind the RNA as the longer a ligand is bound, the faster its relaxation will be. For instance, shorter CPMG delay times are used for high affinity ligands, since their relaxation is more strongly increased upon binding. Comparison of the resulting NMR spectra in the absence and presence of the RNA can be used to validate binding, as observed by altered peaks with decreased signal intensities and/or broadened line-widths.

3.2 RNA-Ligand Structure Determination

NMR structure determination allows the application of structure-based approaches to achieve the synthesis of highly optimized ligands. Due to the intrinsic flexibility of RNA, X-ray crystallography is often difficult to apply and sometimes results in poor resolution at binding interfaces (Reyes et al, 2009; Stark et al, 2011). NMR-focused RNA structure determination parallels proteins in many ways, but an additional layer of difficulty is introduced when pursuing a ligand-bound structure as intermolecular nuclear Overhauser effect (NOE) interactions between RNA and ligand may be transient and difficult to detect, and introduction of the ligand adds additional spectral complexity and overlap (Varani et al, 1991). NOE interactions are T1-dependent; therefore, significant changes in molecular weight directly correlate with increased molecular rotational correlation times (τ_c) that attenuate NOE signal (Levitt, 2001). Therefore, for ligands that reach comparable molecular weight to the RNA under study, such as peptides or proteins, binding reduces NOE detection resulting in highly overlapped and attenuated cross-peak intensities for both molecules (Stark et al, 2011). Intermolecular NOEs are often prominent

based on their chemical shift perturbation patterns, but separate data sets and assignment for the high molecular weight ligand need to be collected as well.

A bottleneck in RNA structure determination is resonance assignment due to the severe overlapped of ribose nuclei (Wüthrich, 1986). H_2O NOESY spectra are analyzed to assign key inter-residue NOE interactions involving imino and amino nuclei indicative of base-pairing that "walk" up the helix. This provides NMR-validated secondary structures and valuable starting points for assigning nearby nuclei. For example, Ade-H2 nuclei are 2-3 Å away from the UA base-pair imino proton, exhibiting a strong NOE interaction, which in turn facilitates assignment of D_2O NOESY spectra. H2's exhibit intense strong-medium intensity NOE interactions with i+1 and cross-strand i+1 H1' nuclei that act as starting points for establishing helical walk patterns that occur between ribose and base nucleic all the way up an A-form helix. H1' protons for instance exhibit weak NOE interactions with base H6/8 both intra-residue and with the i+1 residue; whereas H2' protons exhibit strong NOE interactions with their respective i+1 H6/H8 and weak interactions with intra-residue H6/H8's. By exploiting these patterns and validating assignments using ^{13}C filtered 3D NOESY experiments, assignments for nearly all nuclei can be obtained and typically 20-25 NOE interactions for each residue can be collected. Detailed explanation for the assignment process is covered in (Varani et al, 1996).

Structure determination of an RNA is an interactive process that begins when at least a majority of assignments are confidently assigned. NOE assignments from H_2O and D_2O spectra are binned into restraint tables based on peak intensities indicative of inter-proton distances and classified as strong, medium or weak (Varani et al, 1991; Varani et al, 1996; Allain et al, 1995). Each NOE restraint is compiled into a table that is input into the NIH-XPLOR package for biomolecular structure determination from experimental NMR data supplemented with

stereochemical properties (Schwieters et al, 2006). Initial calculations generate a global fold for the RNA by simulated annealing calculations that start with a linearized form of the RNA of interest where force field weight is slowly increased to simulate decrease in temperature while simultaneously guiding the molecular dynamics simulation that is restrained by the compiled NOE restraints (Schwieters et al, 2006). Base-pairs validated by H_2O NOESY can be enforced by providing hydrogen bond and planarity restraints for each base pair. Typically, during initial calculations, large number of structures are generated (200-1000 models) to produce a reliable ensemble of converged structures. Restraint violations are reviewed and reevaluation of spectral assignments and binning limits conducted several times. The lowest energy structure of the ensemble is finally input into an NIH-XPLOR refinement script as a guide for a extended simulated annealing calculation that also incorporates statistical-based torsional angle restraints and base-base contact parameters (Schwieters et al, 2006). The 10 lowest scoring structures from this final refinement are collected and input into VMD for RMSD calculations and scored for validation using servers such as MolProbity (Chen et al, 2010; Humphrey et al, 1996).

Key RNA-ligand interactions can be identified from a well resolved structure, such as electrostatic interactions between charged functional groups with phosphate backbone and hydrogen bonding with nucleotides on their Hoogsteen and Watson-Crick nucleotide faces (Murchie et al, 2004; Gallego et al, 2001). In (Davidson et al, 2009), inspection of the structure of a cyclic peptide bound to RNA shows examples of both specific and non-specific interactions: Arg-1 and Arg-3 side chains form specific polar contacts with guanosine N7 hydrogen bond acceptors, while non-specific electrostatic interactions occur between Arg-9 and Arg-11 and phosphate groups. The Hoogsteen face of base-paired residues contain several exposed hydrogen

bond acceptors which are available to provide both affinity and selectivity to RNA targeting ligands.

Section 4: Future Perspective

4.1 Targeting RNA Sequences

The past decade has seen growing enthusiasm and support for RNA-targeting drug companies. Currently, the five RNA targeted drugs that are FDA approved are antisense oligonucleotides, targeting RNA sequence rather than structure (Mullard, 2019). Oligonucleotide antisense therapeutics are designed to complement their target sequence, which range from mRNA spice sites to RNAi mimetics (Bennett et al, 2019), by Watson-Crick base pairing. Ionis Pharmaceuticals, Inc. has spearheaded the field by developing a multitude of oligonucleotide modifications (e.g. 2'-O-methoxyethylation) aimed at increasing bioavailability, nuclease resistance and biodistribution (Yu et al, 2016). After almost 30 years of research and billions of dollars in development costs, in 2016, with their subsidiary company Akecea, Ionis released their first oligonucleotide-based FDA approved drug, Spiranza, used to treat spinal muscular atrophy (SMA), a rare neuromuscular disorder (Glascok et al, 2017). Despite FDA approval, this milestone highlights the weaknesses in this RNA targeting strategy, the most notable being cost: Spiranza and other similar oligonucleotide therapies cost annually up to $750,000, placing it as "among the most expensive drugs in the world" (Thomas, 2016). When pricing is coupled with oligonucleotide pharmacological limitations, still limiting with current oligonucleotide chemistry (e.g. negative recognition by the immune system, collection in liver during distribution, nuclease susceptibility, etc), these considerations bring into question the breadth of the potential of this form of RNA-targeted therapy (Bosgra et al, 2019). Will antisense oligonucleotides ever grow beyond a hyperspecialized, difficult to deliver, expensive, yet customizable treatment, suitable perhaps for only a few extremely difficult cases (e.g. rare genetic disorders)? And will the next

Spinraza equivalent therapeutic require another 30 years of expensive research and development costs?

Moderna, Inc. has developed an alternative targeting strategy for RNA primary sequence, founded on the premise of the efficient delivery of entire mRNAs by transfection for reprograming cells (Altounian, 2017). A key feature of their approach is using 1-methylpseudouridine modifications to avoid receptor binding at the surface of immune cells (Altounian, 2017). Many of the Moderna pipeline projects have reached Phase 1 clinical trials as of 2019, supported by continuous rounds of big pharma funding and by a 2018 IPO (Altounian, 2017). Most recently, Moderna announced successful results from a Phase 1 cytomegalovirus (CMV) vaccine and progress toward phase 2 and 3 trials (Moderna Inc., 2019), demonstrating momentum for this novel technology.

4.2 Targeting Structured RNAs

Recent industrial endeavors targeting folded, structured RNAs have primarily focused on small molecules, discovered by various screening methods. Ribometrix, Inc. has taken this approach based on Kevin Week's selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) technology. SHAPE probes the flexibility of each nucleotide in an RNA molecule by using compounds that acylate ribose 2'-hydroxyls only when the hydroxyl is solvent exposed, typically when nucleotides are flexible and unpaired. Commonly, activities are measured for each nucleotide which can be used for RNA secondary structure determination. Ribometrix has modified the assay to be used for high-throughput small molecule discovery by allowing SHAPE reactions to occur in the presence of a small molecule. Loss of activity that was originally detected in a control reaction indicates small molecule binding at the stabilized nucleotide

(Warner et al, 2018). The technique is capable of analyzing large parts of the transcriptome in parallel to identify secondary structures likely to bind to small molecules, giving the technique the ability to identify novel RNAs with potential favorable therapeutic targeting opportunities. Another company taking a similar approach is Arrakis Therapeutics, which is implementing a proprietary platform for the systematic discovery and design of RNA-targeted small molecules, "rSMs" (Mullard et al, 2017). Their approach involves expected features of a RNA drug discovery company (i.e. RNA-specific chemical and biological assays, and RNA-directed medicinal chemistry), but also leverages PEARL-Seq, an exclusive sequencing technology that identifies small molecules that modulate RNA function on a high-throughput scale (Mullard et al, 2017). Arrakis founders have hinted at using this platform for targeting IRES-like structures in the 5'-UTR of proliferative genes as a possible cancer targeting strategy; compounds that can induce conformational changes in IRES' could potentially thwart their recognition by key initiation factors resulting in repressed translation rates for transcripts coding for proliferative genes.

Section 5: References

1. Abulwerdi, F. A., Shortridge, M. D., Sztuba-Solinska, J., Wilson, R., Le Grice, S. F., Varani, G., & Schneekloth Jr, J. S. (2016). Development of small molecules with a noncanonical binding mode to HIV-1 trans activation response (TAR) RNA. *J Med Chem*, *59*(24), 11148-11160.
2. Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. *Nat Biotechnol*, *33*(8), 831-838.
3. Barnwal, R. P., Yang, F., & Varani, G. (2017). Applications of NMR to structure determination of RNAs large and small. *Arch Biochem Biophys*, *628*, 42-56.
4. Batey, R. T., Gilbert, S. D., & Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, *432*(7015), 411-415.
5. Belashov, I. A., Crawford, D. W., Cavender, C. E., Dai, P., Beardslee, P. C., Mathews, D. H., ... & Wedekind, J. E. (2018). Structure of HIV TAR in complex with a Lab-Evolved RRM provides insight into duplex RNA recognition and synthesis of a constrained peptide that impairs transcription. *Nucleic Acids Res*, *46*(13), 6401-6415.
6. Bennett, C. F. (2019). Therapeutic antisense oligonucleotides are coming of age. *Annu Rev Med*, *70*, 307-321.
7. Blakeley, B. D., Shattuck, J., Coates, M. B., Tran, E., Laird-Offringa, I. A., & McNaughton, B. R. (2013). Analysis of Protein–RNA Complexes Involving a RNA Recognition Motif Engineered To Bind Hairpins with Seven-and Eight-Nucleotide Loops. *Biochemistry*, *52*(28), 4745-4747.

8. Bosgra, S., Sipkens J., de Kimpe S., den Besten C., Datson N., van Deutekom J. (2019). The Pharmacokinetics of 2'--Methyl Phosphorothioate Antisense Oligonucleotides: Experiences from Developing Exon Skipping Therapies for Duchenne Muscular Dystrophy. *Nucleic Acid Ther*. *29*(6), 305-322.
9. Bothe, J. R., Nikolova, E. N., Eichhorn, C. D., Chugh, J., Hansen, A. L., & Al-Hashimi, H. M. (2011). Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat Methods*, *8*(11), 919-931.
10. Bryson, D. I., Zhang, W., Ray, W. K., & Santos, W. L. (2009). Screening of a branched peptide library with HIV-1 TAR RNA. *Mol Biosyst*, *5*(9), 1070-1073.
11. Bugaut, A., & Balasubramanian, S. (2012). 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res*, *40*(11), 4727-4741.
12. Bullock, M. D., Pickard, K. M., Nielsen, B. S., Sayan, A. E., Jenei, V., Mellone, M., ... & Mirnezami, A. H. (2013). Pleiotropic actions of miR-21 highlight the critical role of deregulated stromal microRNAs during colorectal cancer progression. *Cell Death Dis*, *4*(6), e684.
13. Burnett, J. C., & Rossi, J. J. (2012). RNA-based therapeutics: current progress and future prospects. *Chem Biol*, *19*(1), 60-71.
14. Calabrese, D. R., Connelly, C. M., & Schneekloth Jr, J. S. (2019). Ligand-observed NMR techniques to probe RNA-small molecule interactions. *Methods Enzymol*, *623*, 131-149.
15. Carr, R. A., Congreve, M., Murray, C. W., & Rees, D. C. (2005). Fragment-based lead discovery: leads by design. *Drug Discov Today*, *10*(14), 987-992.

16. Cheah, M. T., Wachter, A., Sudarsan, N., & Breaker, R. R. (2007). Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, *447*(7143), 497-500.

17. Chen, L., Cressina, E., Leeper, F. J., Smith, A. G., & Abell, C. (2010). A fragment-based approach to identifying ligands for riboswitches. *ACS Chem Biol*, *5*(4), 355-358.

18. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., ... & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*(1), 12-21.

19. Chen, Y., & Varani, G. (2013). Engineering RNA-binding proteins for biology. *FEBS J*, *280*(16), 3734-3754.

20. Chen, Y., Yang, F., Zubovic, L., Pavelitz, T., Yang, W., Godin, K., ... & Varani, G. (2016). Targeted inhibition of oncogenic miR-21 maturation with designed RNA-binding proteins. *Nat Chem Biol*, *12*(9), 717-723.

21. Cheng, C. J., Bahal, R., Babar, I. A., Pincus, Z., Barrera, F., Liu, C., ... & Saltzman, W. M. (2015). MicroRNA silencing for cancer therapy targeted to the tumour microenvironment. *Nature*, *518*(7537), 107-110.

22. Cheong, C. G., & Hall, T. M. T. (2006). Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci U S A*, *103*(37), 13635-13639.

23. Chirayil, S., Chirayil, R., & Luebke, K. J. (2009). Discovering ligands for a microRNA precursor with peptoid microarrays. *Nucleic Acids Res*, *37*(16), 5486-5497.

24. Congreve, M., Chessari, G., Tisi, D., & Woodhead, A. J. (2008). Recent developments in fragment-based drug discovery. *J Med Chem*, *51*(13), 3661-3680.

25. Cooper, T. A., Wan, L., & Dreyfuss, G. (2009). RNA and disease. *Cell*, *136*(4), 777-793.

26. Correll, C. C., Munishkin, A., Chan, Y. L., Ren, Z., Wool, I. G., & Steitz, T. A. (1998). Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc Natl Acad Sci U S A*, *95*(23), 13436-13441.

27. Crawford, D. W., Blakeley, B. D., Chen, P. H., Sherpa, C., Le Grice, S. F., Laird-Offringa, I. A., & McNaughton, B. R. (2016). An evolved RNA recognition motif that suppresses HIV-1 Tat/TAR-dependent transcription. *ACS Chem Biol*, *11*(8), 2206-2215.

28. Dai, Y., Wynn, J. E., Peralta, A. N., Sherpa, C., Jayaraman, B., Li, H., ... & Santos, W. L. (2018). Discovery of a branched peptide that recognizes the rev response element (RRE) RNA and blocks HIV-1 replication. *J Med Chem*, *61*(21), 9611-9620.

29. Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M., & Stockman, B. (2001). WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J Biomol NMR*, *21*(4), 349-359.

30. Davidson, A., Begley, D. W., Lau, C., & Varani, G. (2011). A small-molecule probe induces a conformation in HIV TAR RNA capable of binding drug-like fragments. *J Mol Biol*, *410*(5), 984-996.

31. Davidson, A., Leeper, T. C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J. A., & Varani, G. (2009). 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*, *106*(29), 11931-11936.

32. Davis, B., Afshar, M., Varani, G., Murchie, A. I., Karn, J., Lentzen, G., ... & Swarbrick, T. (2004). Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic “hot spots”. *J Mol Biol*, *336*(2), 343-356.

33. Deigan, K. E., & Ferré-D’Amaré, A. R. (2011). Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs. *Acc Chem Res*, *44*(12), 1329-1338.

34. Di Giorgio, A., & Duca, M. (2019). Synthetic small-molecule RNA ligands: future prospects as therapeutic agents. *Medchemcomm*, *10*(8), 1242-1255.

35. Dibrov, S. M., Ding, K., Brunn, N. D., Parker, M. A., Bergdahl, B. M., Wyles, D. L., & Hermann, T. (2012). Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. *Proc Natl Acad Sci U S A*, *109*(14), 5223-5228.

36. Donlic, A., & Hargrove, A. E. (2018). Targeting RNA in mammalian systems with small molecules. *Wiley Interdiscip Rev RNA*, *9*(4), e1477.

37. Ellington, A. D., & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, *346*(6287), 818-822.

38. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., & Jhoti, H. (2016). Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov*, *15*(9), 605-619.

39. Fernandes, T. G., Diogo, M. M., Clark, D. S., Dordick, J. S., & Cabral, J. M. (2009). High-throughput cellular microarray platforms: applications in drug discovery, toxicology and stem cell research. *Trends Biotechnol*, *27*(6), 342-349.

40. Ferré-D'Amaré, A. R., Zhou, K., & Doudna, J. A. (1998). Crystal structure of a hepatitis delta virus ribozyme. *Nature*, *395*(6702), 567-574.

41. Gallego, J., & Varani, G. (2001). Targeting RNA with small-molecule drugs: therapeutic promise and chemical challenges. *Acc Chem Res*, *34*(10), 836-843.

42. Garavís, M., López-Méndez, B., Somoza, A., Oyarzabal, J., Dalvit, C., Villasante, A., ... & González, C. (2014). Discovery of selective ligands for telomeric RNA G-quadruplexes (TERRA) through 19F-NMR based fragment screening. *ACS Chem Biol*, *9*(7), 1559-1566.

43. Garst, A. D., Héroux, A., Rambo, R. P., & Batey, R. T. (2008). Crystal structure of the lysine riboswitch regulatory mRNA element. *J Biol Chem*, *283*(33), 22347-22351.

44. Giordano, S., & Columbano, A. (2013). MicroRNAs: new tools for diagnosis, prognosis, and therapy in hepatocellular carcinoma?. *Hepatology*, *57*(2), 840-847.

45. Glascock, J., Lenz, M., Hobby, K., & Jarecki, J. (2017). Cure SMA and our patient community celebrate the first approved drug for SMA. *Gene Ther*, *24*(9), 498-500.

46. Hajduk, P. J., Olejniczak, E. T., & Fesik, S. W. (1997). One-dimensional relaxation-and diffusion-edited NMR methods for screening compounds that bind to macromolecules. *J Am Chem Soc*, *119*(50), 12257-12261.

47. Hamy, F., Felder, E. R., Heizmann, G., Lazdins, J., Aboul-Ela, F., Varani, G., ... & Klimkait, T. (1997). An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. Proc Natl Acad Sci U S A, *94*(8), 3548-3553.

48. Hann, M. M., Leach, A. R., & Harper, G. (2001). Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci*, *41*(3), 856-864.

49. Howe, J. A., Wang, H., Fischmann, T. O., Balibar, C. J., Xiao, L., Galgoci, A. M., ... & Murgolo, N. (2015). Selective small-molecule inhibition of an RNA structural element. *Nature*, *526*(7575), 672-677.

50. Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph Model*, *14*(1), 33-38.

51. Imbert, P. E., Unterreiner, V., Siebert, D., Gubler, H., Parker, C., & Gabriel, D. (2007). Recommendations for the reduction of compound artifacts in time-resolved fluorescence resonance energy transfer assays. *Assay Drug Dev Technol*, *5*(3), 363-372.

52. Jankowsky, E., & Harris, M. E. (2015). Specificity and nonspecificity in RNA–protein interactions. *Nat Rev Mol Biol*, *16*(9), 533-544.

53. Janzen, W. P. (2014). Screening technologies for small molecule discovery: the state of the art. *Chem Biol*, *21*(9), 1162-1170.

54. Jones, D., Metzger, H. J., Schatz, A., & Waksman, S. A. (1944). Control of gram-negative bacteria in experimental animals by streptomycin. *Science*, *100*(2588), 103-105.

55. Kloiber, K., Spitzer, R., Tollinger, M., Konrat, R., & Kreutz, C. (2011). Probing RNA dynamics via longitudinal exchange and CPMG relaxation dispersion NMR spectroscopy using a sensitive 13C-methyl label. *Nucleic Acids Res*, *39*(10), 4340-4351.

56. Korostelev, A., Trakhanov, S., Laurberg, M., & Noller, H. F. (2006). Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell*, *126*(6), 1065-1077.

57. Kovrigin, E. L. (2012). NMR line shapes and multi-state binding equilibria. *J Biomol NMR*, *53*(3), 257-270.

58. Kumari, S., Bugaut, A., Huppert, J. L., & Balasubramanian, S. (2007). An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol*, *3*(4), 218-221.

59. Kuroyanagi, H. (2009). Fox-1 family of RNA-binding proteins. *Cell Mol Life Sci*, *66*(24), 3895-3907.

60. Lee, A. S., Kranzusch, P. J., & Cate, J. H. (2015). eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature*, *522*(7554), 111-114.

61. Lee, A. S., Kranzusch, P. J., Doudna, J. A., & Cate, J. H. (2016). eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature*, *536*(7614), 96-99.

62. Levitt, M. H. (2001). *Spin dynamics: basics of nuclear magnetic resonance*. John Wiley & Sons.

63. Li, C. H., & Chen, Y. (2013). Targeting long non-coding RNAs in cancers: progress and prospects. *Int J Biochem Cell Biol*, *45*(8), 1895-1910.

64. Ling, H., Fabbri, M., & Calin, G. A. (2013). MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat Rev Drug Discov*, *12*(11), 847-865.

65. Ling, H., Vincent, K., Pichler, M., Fodde, R., Berindan-Neagoe, I., Slack, F. J., & Calin, G. A. (2015). Junk DNA and the long non-coding RNA twist in cancer genetics. *Oncogene*, *34*(39), 5003-5011.

66. Lorenz, D. A., Song, J. M., & Garner, A. L. (2014). High-throughput platform assay technology for the discovery of pre-microRNA-selective small molecule probes. *Bioconjug Chem*, *26*(1), 19-23.

67. Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*, *8*(6), 479-490.

68. Marcheschi, R. J., Mouzakis, K. D., & Butcher, S. E. (2009). Selection and characterization of small molecules that bind the HIV-1 frameshift site RNA. *ACS Chem Biol*, *4*(10), 844-854.

69. Marušič, M., Schlagnitweit, J., & Petzold, K. (2019). RNA dynamics by NMR. *ChemBioChem*, *20*(21), 2683-2799.

70. Mathews, D. H., Moss, W. N., & Turner, D. H. (2010). Folding and finding RNA secondary structure. *Cold Spring Harb Perspect Biol*, *2*(12), a003665.

71. Matsson, P., Doak, B. C., Over, B., & Kihlberg, J. (2016). Cell permeability beyond the rule of 5. *Adv Drug Deliv Rev*, 101, 42-61.

72. Matsui, M., & Corey, D. R. (2017). Non-coding RNAs as drug targets. *Nat Rev Drug Discov*, *16*(3), 167-179.

73. Mayer, M., Lang, P. T., Gerber, S., Madrid, P. B., Pinto, I. G., Guy, R. K., & James, T. L. (2006). Synthesis and testing of a focused phenothiazine library for binding to HIV-1 TAR RNA. *Chem Biol*, *13*(9), 993-1000.

74. Moazed, D., & Noller, H. F. (1987). Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature*, *327*(6121), 389-394.

75. Moderna Announces Positive Interim Results from Phase 1 Cytomegalovirus (CMV) Vaccine (mRNA-1647) Study and Progress Toward Phase 2 and Pivotal Trials. (Sept 2019). Retrieved from https://investors.modernatx.com/node/7706/pdf

76. Mullard A. (2019). 2018 FDA drug approvals. *Nature Rev Drug Discov*, *18*(2), 85-89.

77. Mullard, A. (2017). Small molecules against RNA targets attract big backers. *Nature Rev Drug Discov*, *16*, 813-815.

78. Murchie, A. I., Davis, B., Isel, C., Afshar, M., Drysdale, M. J., Bower, J., ... & Prescott, C. D. (2004). Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. *J Mol Biol*, *336*(3), 625-638.

79. Nana-Sinkam, S. P., & Croce, C. M. (2013). Clinical applications for microRNAs in cancer. *Clinical Pharmacology & Therapeutics*, *93*(1), 98-104.

80. Naro, Y., Ankenbruck, N., Thomas, M., Tivon, Y., Connelly, C. M., Gardner, L., & Deiters, A. (2018). Small molecule inhibition of MicroRNA miR-21 rescues chemosensitivity of renal-cell carcinoma to topotecan. *J Med Chem*, *61*(14), 5900-5909.

81. Pai, J., Yoon, T., Kim, N. D., Lee, I. S., Yu, J., & Shin, I. (2012). High-throughput profiling of peptide–RNA interactions using peptide microarrays. *J Am Chem Soc*, *134*(46), 19287-19296.

82. Parsons, J., Castaldi, M. P., Dutta, S., Dibrov, S. M., Wyles, D. L., & Hermann, T. (2009). Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat Chem Biol*, *5*(11), 823-825.

83. Pedram Fatemi, R., Salah-Uddin, S., Modarresi, F., Khoury, N., Wahlestedt, C., Faghihi, M.A. (2015). Screening for small-molecule modulators of long noncoding RNA-protein interactions using AlphaScreen. *J Biomol Screen*, *20*(9), 1132-41.

84. Reyes, F. E., Garst, A. D., & Batey, R. T. (2009). Strategies in RNA crystallography. In *Methods Enzymol* (Vol. 469, pp. 119-139). Academic Press.

85. Rizvi, N. F., Howe, J. A., Nahvi, A., Klein, D. J., Fischmann, T. O., Kim, H. Y., ... & Chamberlin, C. (2018). Discovery of selective RNA-binding small molecules by affinity-selection mass spectrometry. *ACS Chem Biol*, *13*(3), 820-831.

86. Rizvi, N. F., & Smith, G. F. (2017). RNA as a small molecule druggable target. *Bioorg Med Chem Lett*, *27*(23), 5083-5088.

87. Robertson, D. L., & Joyce, G. F. (1990). Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, *344*(6265), 467-468.

88. Rupert, P. B., & Ferre-D'Amare, A. R. (2001). Crystal structure of a hairpin ribozyme–inhibitor complex with implications for catalysis. *Nature*, *410*(6830), 780-786.

89. Schmitt, A. M., & Chang, H. Y. (2016). Long noncoding RNAs in cancer pathways. *Cancer Cell*, *29*(4), 452-463.

90. Schwieters, C. D., & Clore, G. M. (2014). Using small angle solution scattering data in Xplor-NIH structure calculations. *Prog Nucl Magn Reson Spectrosc*, *80*, 1-11.

91. Scott, W. G., Finch, J. T., & Klug, A. (1995). The crystal structure of an AII-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell*, *81*(7), 991-1002.

92. Serganov, A., & Nudler, E. (2013). A decade of riboswitches. *Cell*, *152*(1-2), 17-24.

93. Serganov, A., & Patel, D. J. (2012). Metabolite recognition principles and molecular mechanisms underlying riboswitch function. *Annu Rev Biophys*, *41*, 343-370.

94. Servick, K. (2016). This mysterious $2 billion biotech is revealing the secrets behind its new drugs and vaccines. *Science* (New York, NY). AAAS.

95. Shi, Y., Yang, Y., Hoang, B., Bardeleben, C., Holmes, B., Gera, J., & Lichtenstein, A. (2016). Therapeutic potential of targeting IRES-dependent c-myc translation in multiple myeloma cells during ER stress. *Oncogene*, *35*(8), 1015-1024.

96. Shortridge, M. D., & Varani, G. (2015). Structure based approaches for targeting non-coding RNAs with small molecules. *Curr Opin Struct Biol*, 30, 79-88.

97. Shortridge, M. D., Walker, M. J., Pavelitz, T., Chen, Y., Yang, W., & Varani, G. (2017). A macrocyclic peptide ligand binds the oncogenic microRNA-21 precursor and suppresses dicer processing. *ACS Chem Biol*, *12*(6), 1611-1620.

98. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A., & Weeks, K. M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods*, *11*(9), 959-965.

99. Silvers, R., Keller, H., Schwalbe, H., & Hengesbach, M. (2015). Differential scanning fluorimetry for monitoring RNA stability. *Chembiochem*, *16*(7), 1109-1114.

100. Smith, M. C., & Gestwicki, J. E. (2012). Features of protein–protein interactions that translate into potent inhibitors: topology, surface area and affinity. *Expert Rev Mol Med*, e16.

101. Stark, J. L., & Powers, R. (2011). Application of NMR and molecular docking in structure-based drug discovery. In NMR of proteins and small biomolecules (pp. 1-34). Springer, Berlin, Heidelberg.

102. Taft, R. J., Pheasant, M., & Mattick, J. S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays*, *29*(3), 288-299.

103. Thomas, K. (2016). Costly drug for fatal muscular disease wins FDA approval. The New York Times.

104. Varani, G., Tinoco, I. (1991). RNA structure and NMR spectroscopy. *Q Rev Biophys*, *24*(4), 479-532.

105. Varani, G., Aboul-ela, F., & Allain, F. H. T. (1996). NMR investigation of RNA structure. *Prog Nucl Magn Reson Spectrosc*, *29*(1-2), 51-127.

106. Varani, G., Cheong, C., & Tinoco Jr, I. (1991). Structure of an unusually stable RNA hairpin. *Biochemistry*, *30*(13), 3280-3289.

107. Vicens, Q., & Westhof, E. (2003). Crystal structure of geneticin bound to a bacterial 16 S ribosomal RNA A site oligonucleotide. *J Mol Biol*, *326*(4), 1175-1188.

108. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., & Gelfand, M. S. (2004). Riboswitches: the oldest mechanism for the regulation of gene expression?. *Trends Genet*, *20*(1), 44-50.

109. Walker, M. J., & Varani, G. (2019). Design of RNA-targeting macrocyclic peptides. *Methods Enzymol*, *623*, 339-372.

110. Wang, Y., & Rando, R. R. (1995). Specific binding of aminoglycoside antibiotics to RNA. *Chem Biol*, *2*(5), 281-290.

111. Warner, K. D., & Ferré-D'Amaré, A. R. (2014). Crystallographic analysis of TPP riboswitch binding by small-molecule ligands discovered through fragment-based drug discovery approaches. In *Methods Enzymol* (Vol. 549, pp. 221-233). Academic Press.

112. Warner, K. D., Hajdin, C. E., & Weeks, K. M. (2018). Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov*, *17*(8), 547-558.

113. Wigglesworth, M. J., Murray, D. C., Blackett, C. J., Kossenjans, M., & Nissink, J. W. M. (2015). Increasing the delivery of next generation therapeutics from high throughput screening libraries. *Curr Opin Chem Biol*, *26*, 104-110.

114. Wilkie, G. S., Dickson, K. S., & Gray, N. K. (2003). Regulation of mRNA translation by 5′-and 3′-UTR-binding factors. *Trends Biochem Sci*, *28*(4), 182-188.

115. Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc*, *73*, 1-16.

116. Wüthrich, K. (1986). NMR with proteins and nucleic acids. *Europhysics News*, 17(1), 11-13.

117. Yu, R. Z., Warren, M. S., Watanabe, T., Nichols, B., Jahic, M., Huang, J., ... & Wang, Y. (2016). Lack of interactions between an antisense oligonucleotide with 2′-O-(2-methoxyethyl) modifications and major drug transporters. *Nucleic Acid Ther*, *26*(2), 111-117.

Chapter 2 – Design of RNA-Targeting Macrocyclic Peptides*

Section 1: Introduction

RNA molecules play crucial roles in regulating healthy and diseased cellular processes (e.g. transcription, splicing, mRNA transport, translation, etc) and are therefore tightly regulated transcriptionally and post-transcriptionally, often through their interactions with other RNAs and with RNA binding proteins. RNAs and RNA-protein interactions regulate viral replication or the expression of proto-oncogenes, and are mis-regulated in many infectious and chronic diseases (Cooper, Wan, & Dreyfuss, 2009; Esteller, 2011) making these RNA structures and RNA-protein surfaces an untapped source of potential drug targets (Burnett, & Rossi, 2012; Ling, Fabbri, & Calin, 2013). However, RNA-protein interactions are much more challenging to target with small molecules than traditional enzymatic active sites (Warner, Hajdin, & Weeks, 2018). These interactions span large surface areas and often lack structural complexity (Jones, Daley, Luscombe, Berman, & Thornton, 2001; Lunde, Moore, & Varani, 2007). Therefore, it is more challenging, though not impossible (Afshar et al, 1999; Bower et al, 2003; Murchie et al, 2004; Davis et al, 2004; Howe et al, 2015; Palacino et al, 2015; Ratni et al, 2016), to discover small molecules that compete with much larger proteins and stabilize the often dynamic single stranded regions of the RNA. Intermediate molecular weight (1.5-2 kDa) peptides can provide much greater surface area and therefore have greater potential to form high affinity and specific complexes (Puglisi, Chen, Blanchard, & Frankel, 1995; Battiste et al, 1996). Thus, our group has

* This chapter appeared in a volume of Methods in Enzymology, 2019;623:339-372.

used peptides to discover RNA-binding ligands for probing structural and mechanistic aspects of RNA-protein interactions and investigating possible new RNA inhibitors.

Here, we discuss our approach to engineering peptides that bind RNAs by highlighting methods and design strategies **(Fig. 2.1)**. In Section 2, we describe limitations of targeting RNA with linear peptides (Leulliot, & Varani, 2001) and how conformationally constrained peptide mimetics address some of these issues (Robinson, 2008). Section 3 discusses the design process for building mimetics from protein structure and sequence, based on our successful targeting of the interaction between viral trans-activator of transcription (Tat) and trans-activating response element (TAR) (Athanassiou et al., 2004; Leeper, Athanassiou, Dias, Robinson, & Varani, 2005). Section 4 describes the use of positional scanning libraries to discover a high affinity peptide capable of binding human immunodeficiency virus TAR RNA (Athanassiou et al, 2007; Davidson et al., 2009). Section 5 is dedicated to explaining how structure-based optimization can lead to the discovery of peptides with low pM affinity and exquisite specificity (Davidson, Patora-Komisarska, Robinson, & Varani, 2011; Shortridge et al, 2018). In Section 6, we evaluate how β-turn mimetics can be adapted to target other pharmaceutically relevant RNA stem-loop structures by two cases as examples (Moehle et al, 2007; Shortridge et al, 2017).

Section 2: Targeting RNA with Macrocyclic β-Hairpin Peptide Mimetics

2.1 Challenges in Targeting RNA with Peptides

Engineering peptides to bind RNA requires an understanding of several fundamental challenges inherent to targeting RNA. First, RNA structures are not abundant in the PDB. Crystallization can be difficult and sometimes impossible because of RNA's intrinsic flexibility (Leulliot, & Varani, 2001; Williamson, 2000). Our group uses NMR spectroscopy to circumnavigate this issue as it provides characterization of RNA and peptide conformation, even when structures are flexible (Varani, Aboula-ela, & Allain, 1996). Structure determination of peptide-RNA complexes produce atomic resolution accuracy and provides critical insight regarding key peptide-RNA interactions, yet only becomes accessible once a potent enough peptide is identified. Furthermore, overlapped NMR spectra reduce resolution, making unambiguous chemical shift assignment impossible for weakly bound peptides discovered early on in the design process. Second, binding interfaces are sometimes so large that it can be difficult to target them with peptides, particularly for larger multimeric RNA-protein complexes. Third, for any RNA binding ligand, it is difficult to differentiate using only biochemical assays (K_d) how stable the RNA-ligand complex actually is. Even when a molecule is bound, conformational flexibility might remain, or the molecule might bind in multiple locations on the RNA, boosting affinity through avidity and giving the illusion of potency. It is often the case that the resulting affinity results from non-specific interactions from electrostatic interactions between the phosphate backbone or electronegative major groove with basic side chains of peptides, or charged small molecules. Maintaining a single RNA-ligand conformation is also inherently energetically unfavorable because of the entropic penalty associated with rigidification of flexible RNA molecules by induced fit (Leulliot, & Varani, 2001; Williamson, 2000). The

penalty is magnified when RNA is targeted with linear peptides with poor conformational stability, since short peptide fragments lack the tertiary interactions that buttress protein binding sites in folded proteins. This is the reason we prefer to use highly structured peptide mimics with stable secondary structure in our targeting efforts.

2.2 Macrocyclic β-Hairpin Scaffolds as an Effective Strategy for Targeting RNA

Structure-based peptidomimetics is a peptide design strategy where inspection of a protein-bound structure provides a rational to design a minimal peptide that mimics the activity of the full protein (Pelay-Gimeno, Glas, Koch, & Grossmann, 2015; Mason, 2010; Avan, Hall, & Katritzky, 2014). The protein structure provides a starting point for the design of structurally rigid peptides that recapitulate intermolecular interactions within a minimal framework and inhibit biological activity (Werder, Hauser, Abele, & Seebach, 1999; Rosenström et al, 2006; Whitby et al, 2011; Muppidi et al, 2012). It is crucial for the mimetic to match both sequence and structure, and this is done most effectively when using secondary structure motifs that stabilize a peptide fold through local intramolecular interactions while allowing incorporation of the desired sequence into small peptide fragments (Fujii et al, 2003; Pelay-Gimeno, Glas, Koch, & Grossmann, 2015). The two most common protein secondary structures are β-hairpins and β-helices, whose backbones act as scaffolds upon which energetically favorably sidechains can be pre-presented for binding in position precisely defined by intramolecular hydrogen bonding and packing interactions. β-hairpins are of particular interest for targeting RNA as many RNA-binding proteins (RBP's) exploit β-sheet structures (Cléry, Blatter, & Allain, 2008; Sibanda, & Thornton, 1985).

Our group has targeted RNA using macrocyclic β-hairpin peptidomimetic chemistry. Head-to-tail cyclization rigidifies the structure (Pelay-Gimeno, Glas, Koch, & Grossmann, 2015), and cyclized peptides are advantageous from a pharmacological standpoint as well because they are exonuclease resistant (Robinson, 2000) and have increased cellular uptake when arginine-rich (Lättig-Tünnemann et al, 2011). Their structure is comprised of two anti-parallel β-strands stabilized by two β-hairpin favoring turn inducers located at the tip of both strands. β-hairpins strongly favor Type I' ($\varphi_{i+1} = 60$, $\psi_{i+1} = 30$, $\varphi_{i+2} = 90$, $\psi_{i+2} = 0$) and Type II' ($\varphi_{i+1} = 60$, $\psi_{i+1} = -120$, $\varphi_{i+2} = -80$, $\psi_{i+2} = 0$) turn inducing residues (Madan, Seo, & Lee, 2014). Most of our published work uses 14-mer β-hairpins with a Type II' D-Pro/L-Pro turn inducer as it readily favors β-hairpin folding **(Fig. 2.2A)** (Nair, Vijayan, Venkatachalapathi, & Balaram, 1979; Bean, Kopple, & Peishoff, 1992). The D-Pro/L-Pro turn developed in the Robinson group acts as a template for grafting mimicked sequences onto the β-hairpin scaffold (Robinson, 2008), favoring formation of stable β-hairpins. The second turn varies in sequence depending on the protein being mimicked with the only requirement that it favors Type I' or Type II' (Robinson, 2008).

If properly designed, cyclic β-hairpins will possess a rigid backbone structure, stabilized by multiple NH to CO backbone hydrogen bonds. A rigid β-hairpin provides a reliable scaffold for mimicking a protein and makes amino acid substitution less likely to affect the β-hairpin character. A 14-mer cyclic β-hairpin favors a 2:2 sidechain orientation (Sibanda, Blundell, & Thornton, 1989) presenting 6 residues in one direction (for the purpose of this chapter, positions: 1, 3, 5, 8, 10, 12) and 4 residues in the opposite direction (positions: 2, 4, 9, 11) **(Fig. 2.2B)**. Considering this positioning is essential during optimization and highlighted in Section 4. This class of cyclic β-hairpins also has size and shape to match major groove RNA features found at

apical loop and/or bulge junctions **(Fig. 2.2C)**, allowing sidechains to make stabilizing contacts with RNA bases and the phosphate backbone. The combination of major groove fit and optimizable intermolecular interactions gives the β-hairpin a higher probability of inducing RNA conformation changes that favor formation of both canonical and non-canonical base-pairs, even when these are not present in the unbound state of the RNA (Bardaro et al, 2009, Borkar et al, 2016). Formation of these base-pairs is influential in developing high affinity RNA-targeted peptides.

A detailed description of synthetic methods used to prepare these peptides can be found in previous publications (Athanassiou et al, 2004; Athanassiou et al, 2007). Briefly, cyclic β-hairpin peptides can be generated following solid-state peptide synthesis followed by a standard cyclization procedure, using Fmoc chemistry on acid sensitive 2-chlorotrityl resin in order to retain protecting groups during resin cleavage and HoBt/HBTU head-to-tail cyclization. HPLC purification is necessary after cyclization and before side-chain deprotection in order to separate cyclized product from unreacted linear peptide. It should be noted that due to the highly constrained nature of β-hairpins, some sequences have unexpected poor yields, likely due to unfavorable reaction geometries. In addition, the cyclization reaction is sensitive to sequences that produce flexible backbones (e.g. incorporation of glycine, backbone kinks) which may inhibit the cyclization reaction and/or require longer reaction times.

Section 3: Peptidomimetic Targeting the BIV-TAR–Tat Interaction

Mimicking the TAR–Tat interaction of the Bovine Immunodeficiency Virus (BIV) provides an illustration of how to mimic an RBP by grafting amino acids onto a cyclic β-hairpin peptidomimetic scaffold (Athanassiou et al., 2004; Leeper, Athanassiou, Dias, Robinson, & Varani, 2005), and in turn shows the feasibility of using β-hairpin scaffolds to target RNA.

3.1 Design Rational

Regulatory RNAs in viral RNA genomes provide attractive targets for discovery of new antiviral drugs because these elements are essential for viral replication, structured, and highly conserved between isolates (Gallego, & Varani, 2001; Wilson, & Li, 2000; Kirk et al, 1998). A well-studied example is provided by the Tat protein, which plays a critical role lentiviral life cycle and contributes significantly to HIV pathogenesis by increasing transcription of the integrated pro-virus by recognition of TAR RNA, a structured *cis*-regulatory element in viral 5' UTR mRNA (Gonda, Luther, Fong, & Tobin, 1994; Gene, 1994; Wei, Garber, Fang, Fischer, & Jones, 1998). The TAR-Tat interaction is essential for viral replication and driven by TAR recognition of the highly basic Tat Arginine Rich Motif (ARM), which, in the BIV protein (though not in the human counterpart) forms a β-hairpin that binds to BIV TAR within the major groove at the apical loop-stem junction (Ye, Kumar, & Patel, 1995; Puglisi, Chen, Blanchard, & Frankel, 1995). Single nucleotide or amino acid substitutions within either the protein or the RNA severely affect transcriptional levels and hence viral replication (Fischer, Huber, Boelens, Mattaj, & Luhrmann, 1995). BIV and HIV-TAR have similar structure and function, but at the time structural information was limited to the BIV TAR–Tat complex, to which we directed our mimicking efforts.

The starting point for this project was the structure of a Tat 14-mer linear peptide (residues 68-RPRGTRGKGRRIRR-81) bound to the BIV TAR stem-loop (residues 4-31) **(Fig. 2.3A)** (Puglisi, Chen, Blanchard, & Frankel, 1995). In the structure, TAR grasps the Tat peptide in a glove-like manner within a widened major groove pocket, sandwiching it between the apical loop and thumb-like bulge **(Fig. 2.3B)**. The 14-mer peptide bound to the RNA forms a β-hairpin that is kinked due to flexible Gly-71, pushing Thr-72 to bulge upwards towards G22 within the BIV-TAR apical loop. The β-hairpin is favored by a Gly-Lys-Gly γ-turn facing downward and away from the apical loop. Significant contact points include: hydrophobic contacts between the methyl group of Thr-72 and the ribose ring of G22; phosphate H-bonding and π-stacking interactions between Arg-70/G14 and Arg-77/C8 and G9; a triple base pair induced by Arg-73 positioned between G9 and U10 and stabilized by hydrophobic packing of Ile-79 neighboring U11 **(Fig. 2.3C)**. Together, sandwiching of three arginines (Arg-77, Arg-73, Arg-70) between the neighboring bulge nucleotides (C8, G9, U10, G11) are critical elements for BIV-TAR recognition forming arginine sandwich motifs (ASM) (Puglisi, Chen, Blanchard, & Frankel, 1995).

Since BIV Tat forms a β-hairpin when bound to BIV TAR, design of the first mimetic (BIV-0) was straightforward. Mutagenesis data shows that residues critical for binding (Arg-70, Gly-71, Thr-72, Arg-73, Gly-74, Arg-77, and Ile-79) are all located close to the reverse turn formed by residues Gly-74, Lys-75, and Gly-76 (Chen, & Frankel, 1995). Therefore, the peptide was cyclized by replacing Arg-68, Pro-69, and Arg-81 with the D-Pro/L-Pro Type II' turn template **(Fig. 2.3D)**. However, this mutagenesis and sequence-based approach to graft the wildtype Tat peptide fragment onto a β-hairpin scaffold was unsuccessful, since BIV-0,

containing an almost identical sequence to wild-type Tat showed very poor, probably non-specific binding (K_d ~ 40μM) (Athanassiou et al., 2004).

This result prompted us to re-consider the importance of peptide stability for β-hairpin design. When we investigated by NMR the structure of BIV-0, we observed essentially a random coil conformation (Athanassiou et al., 2004). In retrospect, a regular 2:2 β-hairpin conformation (Sibanda, Blundell, & Thornton, 1989) requires a particular number of residues on each strand for ideal hydrogen bonding (Griffiths-Jones, Maynard, & Searle, 1999). Turn chemistry may have also destabilized the peptide since the rigid turn-inducing template D-Pro/L-Pro near Gly-71 can induce an irregular backbone conformation. We therefore redesigned the peptide mimic by using 12-residue peptides to promote 2:2 β-hairpin formation and mounting the sequence on a D-Pro/L-Pro template (Favre, Moehle, Jiang, Pfeiffer, & Robinson, 1999; Jiang, Moehle, Dhanapal, Obrecht, & Robinson, 2000). Eight peptides were synthesized (BIV-1–BIV-8, **Table 2.1**) with residues Thr-4, Arg-5, Gly-6, Lys-7, Arg-8, Arg-9, and Ile-10 kept constant, corresponding to residues Thr-72, Arg-73, Gly-74, Lys-75, Arg-77, Arg-78, and Ile-79 in BIV Tat, which make critical interactions with BIV-TAR (Athanassiou et al., 2004). The sequence near the apical loop was trimmed by changing positions 1-3, 11, and 12 (positions 68-70, 80, and 81 in wildtype Tat) to residues with charged and aromatic sidechains. Since β-hairpins prefer type I' and type II' chemistry, Gly-76 from the β-turn was also removed to favor a Type II' turn of Gly-74 and Lys-75. Although all peptides are basic and bind with low μM affinity, only of the eight mimetics (BIV-2) bound with nM affinity (K_d = 150nM) (Athanassiou et al., 2004). BIV-2 is arginine-rich and lacks a mid-strand glycine between turns, further supporting our hypothesis that formation of a stable and regular β-hairpin is essential for peptides to bind RNA.

Next, we investigated which BIV-2 structural features made it bind BIV-TAR strongly by evaluating each peptide structure by NMR. BIV-0, for example had random coil characteristics, but NMR data for BIV-2 showed it had a single set of peaks (Athanassiou et al., 2004). Structures were determined for BIV-2 and BIV-3, and both adopted β-hairpin structures, whereas the other mimetics adopted multiple conformations, yet BIV-2 had a lower RMSD compared to BIV-3, suggesting BIV-2 was more rigid and that peptide stability contributed to binding (Athanassiou et al., 2004).

3.2 Structure and SAR Development

The structure of HIV-TAR bound to BIV-2 was determined using NMR (Leeper, Athanassiou, Dias, Robinson, & Varani, 2005). Inspection of the BIV-TAR–BIV-2 complex reveals that BIV-2 is bound in the same major groove pocket as BIV Tat, but in a completely different orientation, being upside down relative to what was expected from the targeted structure with the D-Pro/L-Pro template positioned downward into the major groove and the Gly-6/Lys-7 turn upwards into the apical loop **(Fig. 2.4A)**. Yet, despite the flipped orientation, many sidechain interactions with the RNA are similar to BIV Tat. Arg-1, Arg-3, and Arg-5 are sandwiched between bases forming ASMs mimicking BIV Tat sidechain positioning for Arg-70, Arg-73, and Arg-77, respectively. Intercalation occurs between C8, G9, U10, G11, and G14 through a combination of cation-π interactions and hydrogen bonding with purine N7's and guanosine O2's **(Fig. 2.4B)**. Arg-8 is also buried relative to the β-hairpin, yet positioned near flexible apical loop nucleotides and only making transient contacts with multiple apical loop nucleotides. Mimicry is nearly exact between the two structures for Ile-10 and Ile-79, both are positioned deep within the base stack. Ile-10 stabilizes the U10-A13-U24 base triple through van

der Waal interactions with A13 and U24 and is a major driving force for TAR recognition **(Fig. 2.4C, D)**. Solvent exposed Arg-9 and Arg-11 make no hydrogen bonds or cation-π interactions, yet contacts with the phosphate backbone of A21 and G22 favor the apical loop to fold over the peptide (**Fig. 2.4E**). Val-2 and Thr-4 point away from the RNA and into the solvent, making little contact with the RNA.

In summary, the structure revealed that, once a stable β-hairpin conformation is achieved, residue changes that increase or decrease affinity with the RNA are likely to arise from changes in side-chain to RNA interactions. We identified key stabilizing interactions, which contributed to future efforts to characterize the different binding potentials for each position (see Section 4).

Section 4: Discovery of Peptide Mimetics Targeting HIV-TAR

The long-term focus of targeting BIV-TAR project discussed in Section 3 was to discover peptide inhibitors targeting Human Immunodeficiency Virus (HIV) TAR, located at positions 17-45 within the 5'-UTR of the HIV genome **(Fig. 2.5A)**. Its trinucleotide bulge (U23, C24, U25) nucleotides are functionally analogous to BIV TAR U10, G11, and U12 in that they are necessary for Tat recognition and have been shown to reduce replication rates in mutational studies (Keen, Gait, & Karn, 1996; Dingwall et al, 1990; Roy, Delling, & Rosen, 1990). We used the SAR developed for the BIV-TAR–BIV-2 complex to generate a large positional scanning peptide library, synthesized both to optimize binding affinity, but also to discover new HIV-TAR specific peptides (Athanassiou et al, 2007). Our aim of discovering a peptide which bound to HIV-TAR strongly enough for structure determination was successful (Davidson et al., 2009), which was used for developing an HIV-TAR specific SAR utilized in future rounds of optimization.

4.1 Discovery of Low nM Peptide Ligands for HIV-TAR Using a Large Positional Scanning Peptide Library

The structure and initial SAR for the BIV-2 peptide led to the following summary of its requirement for high affinity binding:

- Positions 1, 3, 5, 8 are buried against the RNA and exhibit ASM sandwiching with C8, G9, G11, U10, and G14. Arginines are likely favored at these positions because of their multiple hydrogen bond acceptors and charged protonation state. Hydrogen bonding was observed between purine N7's and guanosine O2's, along with cation-π and π-π interactions with various bases. Arg-5 contributes

significantly to triple base-pair formation based on its location in the binding pocket.

- Positions 6 and 7 are non-template turn inducing residues required to be Type I' and Type II' turn inducers, since these turns favor β-hairpin formation.
- Positions 2 and 4 are solvent exposed and make negligible interactions with the RNA and minor contacts with neighboring side-chains.
- Positions 9 and 11 are solvent exposed but make interactions with A21 and G22 that support the RNA conformation through induced fit
- Positions 10 and 12 are deeply buried into the RNA interface and form hydrophobic interactions. Ile-10 contributes significantly to triple base-pair formation by packing against A13 and U24.

These interactions are represented on a schematic of a β-hairpin **(Fig. 2.5B)** and provide a design rational for the positional scanning peptide library (L-1 to L-86) that led to the discovery of HIV-TAR binding peptides L-22, L50 and L-51 (Athanassiou et al, 2007). Since optimization was carried out for both BIV and HIV-TAR, patterns discussed are explained as observations for both RNAs unless otherwise specified.

4.1.1 Alanine Scanning

In a pre-panel screen, each position except the D-Pro/L-Pro template was changed to alanine to identify which residues contributed to binding. Remarkably, every alanine substitution reduced binding significantly, with the exception of solvent exposed Val-2 and Thr-4 where K_ds only increased 2-fold relative to BIV-2 (Athanassiou et al, 2007). Changing the buried Ile-10 and Val-12 to alanine also led to low μM binding, likely because of a reduced number of

hydrophobic interactions near the base triple. Decreased binding affinity for each residue was also proposed to be correlated to the loss of cross-strand packing with arginine methylene chains. This observation invited us to consider peptide intramolecular stability through side-chain to side-chain interactions when making design decisions.

4.1.2 ASM-forming Residues: Arg-1, Arg-3, Arg-5, Arg-8

Arg-1, Arg-3, Arg-5 essentially mimic BIV Tat Arg-77, Arg-73 and Arg-70, while Arg-8 interacts with both A21 associated base and phosphate in the apical loop. Arginines at positions 1,3,5, and 8 were substituted with both charged (Lys, L-ornithine Orn); polar (Asn, Gln, L-citrulline Cit); and aromatic (Tyr, Trp) residues. Arginines are less likely to make non-specific phosphate salt bridges than lysine and its analogue ornithine (Rohs et al, 2010). Therefore, when affinity is retained by mutating an arginine to lysine, the interaction is potentially non-specific. Positions Arg-1, Arg-3, Arg-5 all show no binding when replaced with Lys, but substitution of position 8 to lysine and ornithine only increased its K_d to 2.5μM suggesting some degree of non-specific electrostatic backbone interaction (Athanassiou et al, 2007).

4.1.3 Non-template B-Hairpin Turn Inducers: Gly-6, Lys-7

The D-Pro/L-Pro template and Gly-6, Lys-7 form Type II' turns. Their identity is important (Chen, Lin, Lee, Jan, & Chan, 2001); BIV-0 bound poorly likely due to conformational instability from dis-favorable turn chemistry (Athanassiou et al, 2004). This hypothesis was tested with L-22 and L-51 by switching Gly-6 and Lys-7 to residues favoring a Type I' turn. This resulted in a significant increase in affinity and selectivity for HIV-TAR over BIV-TAR, and equally important, produced cleaner NMR data for structure determination even

when compared to peptides with higher binding affinities (Athanassiou et al, 2007). Lys-7 was also changed to polar, aromatic, and charged residues to improve binding, yet lysine appeared to be the preferred amino acid here. How turn sequences can be used to modulate specificity to other RNAs is an ongoing subject of investigation.

4.1.4 Solvent Exposed Val-2, Thr-4

Solvent exposed residues were expected to make negligible contributions to binding. Changing Val-2 and Thr-4 to hydrophobic (Thr, Leu); polar (Asn, Gln); and aromatic (Tyr, Trp) produced mixes results. Val-2 to Thr and Thr-4 to Gln, Val, Tyr increased binding by 2-3-fold (as ordered by largest to smallest increase in affinity) (Athanassiou et al, 2007). Increase in affinity was speculated to be due to changes in peptide conformational stability or lower solvation free energy, but careful analysis of these unbound peptides would be needed to confirm this conjecture.

4.1.5 Apical Loop Binding Residues Arg-9, Arg-11

Residues that might make contacts with the apical loop because of their location on the peptide structure are difficult to optimize. Structures tend to converge poorly in these regions due to the inherent apical loop flexibly and non-specific interactions with the phosphate backbone might drive any energetic contribution. Arg-9 and Arg-11 were substituted, similarly to Arg-1, Arg-3, Arg-5, and Arg-8, by changing arginine to lysine. For BIV-TAR, lysine and ornithine substitutions retained binding (K_d = 150nM) (Athanassiou et al, 2007). When the same peptides were assayed against HIV-TAR, changes at position 9 retained binding, yet changes at position

11 led to 2-3-fold loss in binding affinity suggesting a greater and unexpected dependence on arginine at position 11.

4.1.6 Major Groove Buried Hydrophobic Residues: Ile-10 and Val-12

Inspection of the BIV-2 bound structure shows both of these hydrophobic residues are buried deep within the major groove and contribute a significant faction of the buried interfacial surface. There residues were altered to optimize hydrophobic packing, and for Ile-10, increased stabilization for the triple base pair. Both were substituted to hydrophobic (Leu, cyclohexylalanine); aromatic (Phe, Tyr); and polar (Asn, Gln). Cyclohexylalanine was used as a control due to its hydrophobicity, yet sterically unfavorable size resulting in low μM affinity at each position. All hydrophobic changes retained or increased K_d 2-3-fold, likely due to their increased size leading to a looser fit of the peptide in the major groove (Athanassiou et al, 2007). Val-12 to Ile-12 increased binding affinity (K_d = 100nM), however, and was incorporated in conjunction with changes at other positions (e.g. in L-22). Surprisingly, Ile-10 could be changed to different hydrophobic residues for BIV-TAR without significant changes in binding, yet for HIV-TAR substitutions to other residue led to loss of binding, revealing that isoleucine is optimal at this position for HIV-TAR recognition.

4.1.7 Combining Changes at Multiple Positions

Residue changes that resulted in increased binding affinity were introduced in combination aiming for additive effects**. Table 2.2** provides examples of peptides with several single residue changes (L-42, L-46, L-51, L-59) with improved binding affinity against BIV-TAR. L-72 was designed to incorporate both modifications at position 11 (Arg-to-Orn, K_d =

100nM) and 12 (Val-to-Ile, K_d = 100nM), which together increased binding 2-fold (K_d = 50nM). L-76 incorporated Thr at position 2 for L-72, decreasing the K_d further (K_d = 20nM) (Athanassiou et al, 2007). Not all changes were additive, however, since L-78 incorporated a glutamate into position 4 and reduced binding affinity 2.5-fold (K_d = 50nM). L-76 and L-78 together demonstrates incorporation of multiple positional changes for additivities effects requires careful step-wise design as well, since the effects are not simply additive.

4.2 HIV-TAR–L-22 Structure and SAR

Peptides L-22, L50 and L-51 were discovered as the highest-affinity peptides targeting HIV-TAR (Davidson et al., 2009) **(Table 2.3)**. L-51 also demonstrated 10-fold specificity, with binding affinities of 5 nM for HIV-TAR and 50 nM for BIV-TAR, but competition assays showed that L-22 inhibited a 37-mer Tat-derived linear peptide better, since its IC_{50} was only 2-3-fold above its K_d (Davidson et al., 2009). This promising inhibition data, combined with cleaner NMR data compared to L50 and L-51, prompted us to determine the structure of HIV-TAR bound to L-22, even if L50 binds most strongly and L-51 is specific. It is commonly the case that the most potent or highest affinity ligand is chosen as the optimal lead, but we strongly believe that the optimal starting structure is that which provides for structure determination, because of more facile optimization. Not only because of the obvious advantage of the availability of a structure, but also because the quality of the NMR data indicates that the peptide binds in a unique rigid orientation, which is not the case for L50 and L-51.

The structure of the HIV-TAR–L-22 complex was determined using a very high-quality NMR dataset (Davidson et al., 2009). Positioning of L-22 is similar to BIV-2 on BIV TAR, with the D-Pro/L-Pro template facing down into the major groove and a base triple forming between

U23, A27, and U38 **(Fig. 2.6A)**. However, L-22 bound is buried even deeper into the major groove than BIV-2 into BIV TAR, sequestering ~50% of the peptide surface from solvent, including most of its hydrophobic side chains. L-22 SAR with HIV-TAR is as follows: residues 1, 3, 5, 8 make hydrogen bond and/or stacking interactions with the RNA; 10 and 12 make hydrophobic interactions; positions 9 and 11 bind through transient electrostatic interactions near the backbone of G34 and A35 **(Fig. 2.6B)**. Arg-1 and Arg-3 stabilized the stem-bulge interface, with Arg-1 contacting the RNA backbone and making specific hydrogen bonds with G21 N7 and A22 N7 and Arg-3 with A22 **(Fig. 2.6C)**. Both side chains are proximal to the phosphate backbone making electrostatic interactions as well. As for the BIV Tat and BIV-2 structures, Arg-5 and Ile-10 stabilize the HIV-TAR base triple, with Arg-5 stacking over the three nucleotides. Very high-quality NMR data provided reliable identification of hydrogen bonding between the Arg-5 guanidium group and G28 N7 and O2 **(Fig. 2.6D)**. Ile-10 is packed deep into the groove and makes hydrophobic interactions with the H5 and H6 side of U23 **(Fig. 2.6E)**. A hypothesis for why the KG turn was preferred over the GK turn in L50 was a newly formed electrostatic interaction between Lys-6 and the phosphate backbone of G28. This long-distance interaction with the RNA was considered in future design and optimized in the later generation peptide JB-181 (Section 5.2).

Section 5: Structure-Based Optimization of L-22 Binding Affinity and Specificity

Further improvements in affinity and specificity were achieved by generating reliable SAR and the structure of the complex as a basis for design hypothesis. In this section, we will dissect these strategies by discussing both a blind alley (Davidson, Patora-Komisarska, Robinson, & Varani, 2011) and success (Shortridge, et al., 2018).

5.1 KP-Z-41: Extending Cyclic β-Hairpin Peptides to 15, 16, and 18-mers

Synthesis of the KP-Z peptide series sought to improve specificity and affinity by extending the peptide to fully engage the apical loop. Inspection of the 14-mer L-22 bound to HIV-TAR shows that several apical loop nucleotides make no interactions with it. Peptide libraries based on the L-22 sequence were designed to take advantage of these available interaction sites by extending the peptides to 15, 16 and 18 amino acids (Davidson, Patora-Komisarska, Robinson, & Varani, 2011). Essential structural requirements for specific recognition of HIV-TAR were observed based on the structure of KP-Z-41 bound to TAR, which revealed an unexpected position of the peptide distinct from our original design.

We synthesized a medium sized cyclic β-hairpin peptide library (~25-50) of 15, 16, and 18 residues in length (Davidson, Patora-Komisarska, Robinson, & Varani, 2011). The L-22 sequence was conserved in these peptides, yet new amino acids were mainly based on our previously established understanding of buried and solvent exposed residues and model-based predictions of new possible contacts in the apical loop by extension of the peptide. Cross-strand disulfide linkages were also introduced to test whether increased rigidity from additional covalent linkage would enhance binding. 15 and 16 amino acid β-hairpins showed poor binding

in all cases, perhaps not unexpectedly since these peptides should be unstably folded, whereas 18 amino acid β-hairpins fold similarly to the 14-mers, following the n+4 periodicity for 2:2 β-hairpins (Griffiths-Jones, Maynard, & Searle, 1999). Binding affinity was enhanced for some 18-mer β-hairpins, improving binding relative to L-22, likely due to an additional arginine positioned on the major groove side of the peptide (positions 1, 3, 5, 7). KP-Z-40 displayed the highest binding affinity for HIV-TAR with a K_d in the pM range without excess tRNA present, and a 25 nM K_d even in the presence of 10,000-fold excess of tRNA (Davidson, Patora-Komisarska, Robinson, & Varani, 2011). However, quality of NMR data for bound KP-Z-40 was poor. We hypothesized that Ala-12 may be occupying the triple base pair pocket normally occupied by Ile-10. KP-Z-41 was therefore designed to have this position changed to an isoleucine and we observed an unexpected 3-fold drop in affinity, yet spectral quality improved allowing for structure determination to be executed (Davidson, Patora-Komisarska, Robinson, & Varani, 2011).

The structure for KP-Z-41 bound to HIV-TAR shows that the peptide is positioned completely opposite to our expectations and farther down into the major groove with the D-Pro/L-Pro template sticking out of the complex **(Fig. 2.7)**. This sliding might be due to the requirement to form arginine-mediated interaction at or near the UGU bulge. Combined with the upside down BIV-2 mimicry of BIV Tat, this result exemplifies how deviations from a scaffold that might seem conservative, can induce drastic changes in binding position and orientation.

5.2 JB-181: Non-Canonical Amino Acids

Peptides L-22, L50 and L-51 bound to RNA potently, with low nM K_ds, but discriminated effectively only against structurally unrelated RNAs, such as tRNA present in all

binding assays to reduce non-specific binding (Davidson, A., 2009). When comparing closely-related stem-loop structures such as HIV-TAR mutants or BIV-TAR, binding affinity only decreased by a few to at most 10-folds (Davidson, A., 2009). A new generation of peptides was rationally designed by inspection of the structure of L-22. Since the KP-Z series indicated that extension of the peptide to make additional specific contacts with the apical loop would not be straightforward or possible, new peptides were designed with the goal of allowing deeper penetration into the major groove by incorporating truncated, non-canonical side chains. Using this approach, JB-181 was discovered, a cyclic β-hairpin mimetic of Tat with low picomolar affinity for HIV-TAR and remarkable specificity.

A new library of ~200 peptides was synthesized based on the L-22 sequence, with single residue changes comprising natural L-amino acids, mixing chirality and non-canonical side chains **(Table 2.5, top panel)** (Shortridge et al, 2018). The inclusion of L-2,4-diaminobutryric (Dab) at position 1 (JB-59), and L-2-amino-4-guanidinobutyric acid (noR) (JB-82) at position 11 led to the greatest improvement in binding. When both changes were combined into a single peptide, JB-181, it produced pM binding (K_d = 28±4 pM, as measured by a fluorescence assay) and NMR data of unprecedented quality **(Table 2.5)** (Shortridge et al, 2018). Although the original β-hairpin mimetics were designed using the BIV structure as a model (Puglisi, Chen, Blanchard, & Frankel, 1995; Athanassiou et al, 2004), binding of JB-181 to BIV-TAR occurs at low nM affinity (K_d = 4 nM) and low μM binding affinity is observed towards the non-coding RNA 7SK-SL4 (Shortridge et al, 2018) that also binds to HIV-Tat, demonstrating that JB-181 has 10^2-10^6 fold discrimination against closely related structures.

The quality of the NMR data revealed new interactions responsible for the large increase in binding affinity. JB-181 binds in the same major groove pocket and orientation as the previous

Tat-mimetics **(Fig. 2.8A)** (Shortridge et al, 2018). The backbone φ and ψ angles for JB-181 and L-22 were nearly identical with the exception of Lys-6, yet their positioning in the TAR binding pocket are different **(Fig. 2.8B)**. The introduction of truncated side chains (Dab-1, noR-11) essentially allows the D-Pro/L-Pro template and neighboring residues to fit deeper into the major groove pocket. Inspection of the structure shows that substitution to an amine at positioning one maintains a salt bridge with the RNA backbone (2.4Å from O1P of G21 and 3.6Å from O1P of A22) **(Fig. 2.8C)**, similar to Arg-1, but truncation of the side chain pulls the peptide closer to the major groove by about 2.4Å and allows the KG turn to pitch upwards with respect to the double helical axis. NoR-11 reduces affinity more than 2-fold relative to L-22 (JB-81, K_d = 75 pM), yet we detected more intermolecular NOEs from position noR-11 to A35, which now occupies the *syn* conformation allowing formation of a more favorable cation-π interaction with the new sidechain **(Fig. 2.8D)**. Changes at positions 1 and 11 affected other parts of the structure as well. Deeper burial of the peptide favors improved Arg-3 and Arg-5 sandwiching between bases, increasing polar contacts and closer packing of Ile-10 against the base triple. Ile-10 is stabilized by many new NMR observables never observed by us in >20 years of investigation of multiple complexes of TAR such as a 2'-OH in the base triple protected from exchange with solvent. One of the most notable changes is a new hydrogen bond formed between the amine of Lys-6 and the O4 of U25 **(Fig. 2.8E)**. This novel interaction is supported by mutational analysis such as swapping of the turn from KG to GK in JB-190 **(Table 2.5)**, or mutating U25 to C25, which both decrease binding from low pM to low nM (Shortridge et al, 2018). Altogether, the remarkable structural and binding properties of JB-181 demonstrate that high affinity and specific RNA recognition of relatively simple RNA secondary structures, hairpin loops, can be achieved using a structure guided approach and macrocyclic β-hairpin scaffolds.

Section 6: Targeting other RNAs Using Macrocyclic β-Hairpin Peptides

In this section, we show how the same class of peptides can be used to target RNAs other than TAR, namely the Rev-RRE (Rev Response Element) interaction in HIV (Moehle et al, 2007) and pre-microRNA-21 (pre-miR-21) the precursor to the potent oncomir miRNA-21 (miR-21) (Shortridge et al, 2017). These additional studies demonstrate the versatility of cyclic β-hairpin scaffolds for targeting RNA.

6.1 Targeting the Rev-RRE Interaction

The virally encoded Rev protein plays a critical role in viral replication by regulating the transport of unspliced and partially spliced viral RNA from the nucleus to the cytoplasm of infected cells (Emerman, & Peden, 1989; Malim et al, 1989). Rev functions by binding through an α-helical ARM to stem loop IIb **(Fig. 2.9A)** of the RRE a structured ~300 mRNA segment located within the Env coding region of HIV-1 (Mann et al, 1994). Following this high affinity binding event, approximately 10 additional Rev molecules oligomerize through protein–protein and protein–RNA interactions and coat the entire RRE (Mann et al, 1994). Because the Rev-RRE interaction is analogous to the TAR-Tat interaction in that it involves a high-affinity RNA binding protein that binds by a major groove recognizing ARM. We elected to target it with our BIV-TAR mimetics even if this interaction is distinct from the TAR-Tat interactions in that Rev binds the RRE major groove in an α-helical conformation **(Fig. 2.9B) (**Battiste et al, 1996**)**. Similar to previous studies where β-hairpin scaffolds were used to mimic an α-helices and block HDM2 binding to p53 (Fasan et al, 2004), we reasoned our mimetics could be adapted to position sidechains in similar orientations as in the Rev α-helix. Indeed, the β-hairpins can be

modeled to show that arginines can be positioned on the peptide to moderately mimic the HIV Rev α-helical ARM **(Fig. 2.9C)**.

We screened the small peptide library develop for BIV TAR (**Table 2.1**) for binding to the RRE. BIV-5, which had only a low μM affinity for HIV-TAR (K_d = 1-2μM) bound to the RRE much more strongly (K_d = 0.3μM) (Moehle et al, 2007). Interestingly, BIV-7 which showed no measurable affinity for TAR, bound with ~1μM affinity as well, further suggesting that specificity can be found in arginine rich peptides if the basic side chains are presented on a rigid scaffold like a β-hairpin. BIV-5 showed effective inhibition potential in a competition-based assay as well, by displacing the Rev peptide with IC_{50} similar to its K_d (Moehle et al, 2007). Similar optimization as for BIV Tat-TAR was carried out using a medium-size library, which led to the discovery of R-27, a low nM RRE binding peptide (K_d = 2nM) that was 50-fold more selective for RRE compared to HIV-TAR (K_d = 100nM) (Moehle et al, 2007). This project demonstrated in a proof of principle manner that molecular interactions observed on the Rev α-helix could be successfully transplanted onto a β-hairpin scaffold by rational design, leading to rapid discovery of cyclic peptides that are more potent than the Rev ARM in binding to the RRE high affinity site.

6.2 Pre-microRNA 21

Lastly, we used this chemistry to develop peptides targeting the potent oncogene miR-21. MicroRNAs (miRNA) are short non-coding RNAs that silence gene translation through sequence specific binding in most cases at complementary 3' UTR loci inducing mRNA degradation or translational suppression (Carthew, & Sontheimer, 2009). Their dysregulation is also often observed in human disease (Esteller, 2011; Ling, Fabbri, & Calin, 2013). The biogenesis of

miRNAs begins with expression of a transcript sometimes thousands of nucleotides long called the pri-microRNA (pri-miRNA). The primary transcript undergoes two exonucleolytic cleavage reactions, first in the nucleus by Drosha-DGCR8 to generate the pre-miRNA of 55-75 nucleotides, and second in the cytoplasm by Dicer-TRBP to form the functional mature miRNA of 21-24 nucleotides (Carthew, & Sontheimer, 2009). We focused on miR-21 which is upregulated in many human cancers (Ribas, & Lupold, 2010; Kjaer-Frifeldt et al, 2010; Yan et al, 2008; Asangani et al, 2008; Meng et al, 2007) and sought to discover a peptide inhibitor of miR-21 maturation by screening our cyclic peptide libraries to find an initial scaffold to subsequently construct pre-miR-21 specific cyclic β-hairpins.

Targeting pre-miR-21 was unlike previous projects because of the lack of structural information and because pre-miR-21, our target, has a large apical loop, that NMR data indicate is largely unstructured and flexible **(Fig. 2.10A)**. The Dicer cleavage sites are located above the A29 bulge and at the i+1 cross-strand phosphate backbone, between G45 and C46. Therefore, we focused our screen on this apical section of the structure to avoid discovering peptides that bind to other sections of this RNA. Thus, we truncated the full length pre-miR-21 sequence by removing the bottom half of the helix and used this RNA in both NMR experiments and EMSAs when screening our peptide library (Shortridge et al, 2017). Although EMSAs are cumbersome, they allow direct visualization of complexes and distinguish multimeric and non-specific complexes. We prioritized monomeric hits from screening rather than just affinity because of much easier follow-up to optimize affinity and specificity.

From the initial screening, L50 was identified to bind to pre-miR-21 with nanomolar binding affinity even in the presence excess tRNA (K_d = 200nM) (Shortridge et al, 2017), the only peptide to bind with sub-μM affinity. Follow-up studies revealed that L50 does not bind the

A29 deletion mutant, suggesting L50 bound near the Dicer cleavage site and it out-competed streptomycin (K_d = 100nM) (Bose et al, 2012). These results collectively provided enough evidence to warrant structure determination.

Despite poor definition at this early stage of discovery, the structure revealed critical sites of contacts and contributed an initial SAR. Inspection of the lowest energy structure reveals that L50 binds in the opposite orientation compared to TAR binding predecessors, in that the D-Pro/L-Pro template is facing upward into the apical loop **(Fig. 2.10B)**, and charged residues appear to make non-specific contacts with the phosphate backbone (Shortridge et al, 2017). Interestingly, L50 is the only peptide in our screening library with an arginine at position 12, located near the phosphate backbone of G28, suggesting Arg-12 is necessary for pre-miR-21 binding; both Ile-10 and Arg-12 contact the RNA near the Dicer cleavage site **(Fig. 2.10C)**. When we tested for activity in an *in vitro* dicer assay using *in-house* purified Dicer-TRBP, we observed anomalous Dicer products; cellular assays showed modest but specific inhibition of miR-21 production in HEK293 cell (Shortridge et al, 2017). These preliminary results motivated efforts currently ongoing to optimize L50 for better pre-miR selectivity, and to incorporate new turn chemistry that deviates from the D-Pro-L-Pro template. Current unpublished results have identified peptides with new turn chemistry and sequences that bind much more strongly to pre-miR-21 compared to TAR, inhibit Dicer processing *in vitro,* and reduce miR-21 levels in cells with greatly improved potency compared to the original hit.

Section 7: Discussion

Peptides provide an appealing solution to inhibiting large intermolecular interfaces, such as protein-protein or protein-RNA interactions, that are less amenable to small molecule inhibitors than traditional enzyme targets. While less pharmaceutically attractive than small molecules, they represent a growing area of pharmacology (Pelay-Gimeno, Glas, Koch, & Grossmann, 2015; Fosgerau, & Hoffmann, 2015; Teixido, & Giralt, 2017) that can sometimes lead to bioactive molecules and even, in favorable cases, oral bioavailability (Renukuntla, Dutt, Patel, Boddu, & Mitra, 2013). Past efforts to use peptides to target RNA focused on peptoids (Hamy et al, 1997) or other backbone modification (Niu, Jones, Wu, Varani, & Cai, 2011) invariably led to highly flexible linear structure from which specific binding activity was very difficult to generate, or very short fragments whose binding activity was too weak to observe any biochemical or cellular effects. We show here that cyclic β-hairpin mimics can yield very potent (pM K_d) binding with strong specificity, and that these can be rapidly identified from a combination of conventional library-generated SAR and structure-based design. Success can be found either by mimicking the known interfaces of RNA-binding protein, or by an agnostic process of library screening to identify imperfect hits that can then be rapidly optimized, due to the relative ease of peptide chemistry compared to small molecules. Conformationally restrained peptides are particularly important for RNA, which often have flexible structure. For this reason, peptide folding and binding must be established. Critical information on the characteristics of the intermolecular interface that transcend and extend binding constants can be obtained by NMR. Namely, whether the complex is formed at a unique rigid binding site on the RNA, as opposed to sliding on the structure as observed, for example, with aminoglycosides like streptomycin (Bose et al, 2012; Maiti, Nauwelaerts, & Herdewijn, 2012; Shortridge, & Varani, 2015). Inspection of

the structure of the complex, which can be generated by NMR even for initial relatively low affinity hits if the peptide is rigid, reveals how the peptide is positioned and identifies the location of key sidechains identified through library-based SAR.

While peptides do not generally match small molecules with regards to *in vivo* application, they are nevertheless more conductive to optimization of pharmacological properties than oligonucleotide analogues (Juliano, 2016; Wang et al, 2015; Stein et al, 2011). Combined with *in vitro* evolution, our methods rapidly generate leads that can be optimized to generate peptides with improved pharmacological potential through the introduction of side chain and backbone modifications, leading even, in favorable cases, to oral bio-availability (Witt et al, 2001; Adessi, Soto; 2002; Chatterjee, Gilon, Hoffman, & Kessler, 2008; Boehm et al, 2017). We consider this family of peptides to be very useful alternatives to small molecules in discovering RNA targeting bioactive structures and provide herein protocols for their discovery, characterization and optimization of their binding activity to RNA.

Section 8: References

1. Aboul-ela, F., Karn, J., Varani, G. (1996). Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res*, *24*(20), 3974-3981.
2. Adessi, C., & Soto, C. (2002). Converting a peptide into a drug: strategies to improve stability and bioavailability. Curr Med Chem, *9*(9), 963-978.
3. Afshar M, Prescott CD, Varani G. (1999). Structure-based and combinatorial search for new RNA-binding drugs. *Curr Opin Biotechnol*, *10*(1), 59-63.
4. Asangani, I.A., Rasheed, S.A., Nikolova, D.A., Leupold, J.H., Colburn, N.H., Post, S., Allgayer, H. (2008). MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene*, *27*(15), 2128-2136.
5. Athanassiou, Z., Dias, R., Moehle, K., Dobson, N., Varani G., & Robinson J. (2004). Structural Mimicry of Retroviral Tat Proteins by Constrained B-Hairpin Peptidomimetics: Ligands with High Affinity and Selectivity for Viral TAR RNA Regulatory Elements. *J Am Chem Soc*, *126*(22), 6906-6913.
6. Athanassiou, Z., Patora, K., Dias, R.L., Moehle, K., Robinson, J. A., & Varani, G. (2007). Structure-Guided Peptidomimetic Design Leads to Nanomolar B-Hairpin Inhibitors of the Tat-TAR Interaction of Bovine Immunodeficiency Virus. *Biochemistry*, *46*(3), 741-751.
7. Avan, I., Hall, C. D., Katritzky, A.R. (2014). Peptidomimetics via modifications of amino acids and peptide bonds. *Chem Soc Rev*, *43*(10), 3575-3594.

8. Bardaro Jr, M. F., Shajani, Z., Patora-Komisarska, K., Robinson, J. A., & Varani, G. (2009). How binding of small molecule and peptide ligands to HIV-1 TAR alters the RNA motional landscape. *Nucleic Acids Res*, *37*(5), 1529-1540.
9. Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D., Williamson, J.R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*, *273*(5281), 1547-1551.
10. Bean, J. W., Kopple, K. D., & Peishoff, C. E. (1992). Conformational analysis of cyclic hexapeptides containing the D-Pro-L-Pro sequence to fix. beta.-turn positions. *J Am Chem Soc, 114*(13), 5328-5334.
11. Boehm, M., Beaumont, K., Jones, R., Kalgutkar, A. S., Zhang, L., Atkinson, K., ... & Holder, B. R. (2017). Discovery of Potent and Orally Bioavailable Macrocyclic Peptide–Peptoid Hybrid CXCR7 Modulators. *J Med Chem*, *60*(23), 9653-9663.
12. Borkar, A. N., Bardaro, M. F., Camilloni, C., Aprile, F. A., Varani, G., & Vendruscolo, M. (2016). Structure of a low-population binding intermediate in protein-RNA recognition. *Proc Natl Acad Sci U S A*, *113*(26), 7171-7176.
13. Bose, D., Jayaraj, G., Suryawanshi, H., Agarwala, P., Pore, S. K., Banerjee, R., and Maiti, S. (2012). The tuberculosis drug streptomycin as a potential cancer therapeutic: inhibition of miR-21 function by directly targeting its precursor. *Angew. Chem., Int. Ed.*, *51*(4), 1019−1023.
14. Bower J, Drysdale M, Hebdon R, Jordan A, Lentzen G, Matassova N, Murchie A, Powles J, Roughley S. (2003). Structure-based design of agents targeting the bacterial ribosome. *Bioorg Med Chem Lett.*, *13*(15), 2455-2458.

15. Burnett JC, Rossi JJ. (2012). RNA-based therapeutics: current progress and future prospects. *Chem Biol.*, *19*(1), 60-71.

16. Carthew, R. W. & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, *136*(4), 642–655.

17. Chatterjee, J., Gilon, C., Hoffman, A., & Kessler, H. (2008). N-methylation of peptides: a new perspective in medicinal chemistry. *Acc Chem Res*, *41*(10), 1331-1342.

18. Chen, L., & Frankel, A. D. (1995). A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA. *Proc. Natl. Acad. Sci. U.S.A.*, *92*(11), 5077-5081.

19. Chen, P., Lin, C., Lee, C., Jan, H., & Chan, S. I. (2001). Effects of turn residues in directing the formation of the β-sheet and in the stability of the β-sheet. *Protein Sci*, *10*(9), 1794–1800.

20. Cléry, A., Blatter, M., Allain, F.H. (2008). RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol*, *18*(3), 290-298.

21. Cooper, T. A., Wan, L., & Dreyfuss, G. (2009). RNA and Disease. *Cell*, *136*(4), 777–793.

22. Davidson, A., Leeper, T. C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J. A., & Varani, G. (2009). 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*, *106*(29). 11931–11936.

23. Davidson, A., Patora-Komisarska, K., Robinson, J. A., & Varani, G. (2011). Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of Tat protein. *Nucleic Acids Res, 39*(1), 248–256.

24. Davis B, Afshar M, Varani G, Murchie AI, Karn J, Lentzen G, … & Aboul-ela F. (2004). Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *J Mol Biol*, *336*(2), 343-356.

25. Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., & Skinner, M. A. (1990). HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. *EMBO J. 9*(12), 4145–4153.

26. Emerman, M., & Peden, K. (1989). The rev Gene Product of the Human Immunodeficiency Virus Affects Envelope-Specific RNA Localization. *Cell*, *57*(7), 1155-1165.

27. Esteller, M. (2011). Non-coding RNAs in human disease. *Nat Rev Genet*, *12*(12), 861–874.

28. Fasan, R., Dias, R. L. A., Moehle, K., Zerbe, O., Vrijbloed, J. W., Obrecht, D., & Robinson, J. A. (2004). Using a B-Hairpin To Mimic an Alpha-Helix: Cyclic Peptidomimetic Inhibitors of the p53–HDM2 Protein–Protein Interaction**. *Angew Chem Int Ed Engl Ed*, *43*(16), 2109-2112.

29. Favre, M., Moehle, K., Jiang, L., Pfeiffer, B., & Robinson, J. A. (1999). Structural Mimicry of Canonical Conformations in Antibody Hypervariable Loops Using Cyclic Peptides Containing a Heterochiral Diproline Template. *J Am Chem Soc*, *121*(12), 2679-2685.

30. Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W., & Luhrmann, R. (1995). The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell*, *82*(3), 475-483.

31. Fosgerau, K., & Hoffmann, T. (2015). Peptide therapeutics: current status and future directions. *Drug Discov Today, 20*(1), 122–128.

32. Fujii, N., Oishi, S., Hiramatsu, K., Araki, T., Ueda, S., Tamamura, H., ... & Broach, J. A. (2003). Molecular-size reduction of a potent CXCR4-chemokine antagonist using orthogonal combination of conformation-and sequence-based libraries. *Angew Chem Int Ed Engl, 42*(28), 3251-3253.

33. Gallego, J., & Varani, G. (2001). Targeting RNA with Small-Molecule Drugs: Therapeutic Promise and Chemical Challenges. *Acc Chem Res, 34*(10), 836-843.

34. Gonda, M. A., Luther, D. G., Fong, S. E., & Tobin, G. J. (1994). Bovine immunodeficiency virus: molecular biology and virus-host interactions. *Virus Res, 32*(2), 155-181.

35. Griffiths-Jones, S.R., Maynard, A.J., Searle, M.S. (1999). Dissecting the stability of a β-hairpin peptide that folds in water: NMR and molecular dynamics analysis of the β-turn and β-strand contributions to folding. *J Mol Biol, 292*(5), 1051-1069.

36. Hamy, F., Felder, E.R., Heizmann, G., Lazdins, J., Aboul-ela, F., Varani, G., Karn, J., Klimkait, T. (1997). An inhibitor of the Tat ¯TAR RNA interaction that effectively suppresses HIV-1 replication. *Proc Natl Acad Sci U S A., 94*(8), 3548–3553.

37. Howe, J.A., Wang, H., Fischmann, T.O., Balibar, C.J., Xiao, L., Galgoci, A.M., … & Roemer, T. (2015). Selective small-molecule inhibition of an RNA structural element. *Nature*, 526(7575), 672-677.

38. Jiang, L., Moehle, K., Dhanapal, B., Obrecht, D., & Robinson, J. A. (2000). Combinatorial Biomimetic Chemistry: Parallel Synthesis of a Small Library of β-Hairpin

Mimetics Based on Loop III from Human Platelet-Derived Growth Factor B. *Helvetica Chimica Acta, 83*(12), 3097-3112.

39. Gene, L. D. (1994). Bovine immunodeficiency virus: molecular biology and virus-host interactions. *Virus Research, 32*(2), 155-181.
40. Jones, S., Daley, D. T., Luscombe, N. M., Berman, H. M., & Thornton, J. M. (2001). Protein–RNA interactions: a structural analysis. *Nucleic Acids Res*, *29*(4), 943-954.
41. Juliano, R. L. (2016). The delivery of therapeutic oligonucleotides. *Nucleic Acids Res*, *44*(14), 6518-6548.
42. Keen, N. J., Gait, M. J., & Karn, J. (1996). *Human immunodeficiency virus type-1 Tat is an integral component of the activated transcription-elongation complex. Proc Natl Acad Sci. U S A. 93*(6), 2505-2510.
43. Kirk, S. R., Luedtke, N. W., & Tor, Y. (2000). Neomycin− acridine conjugate: a potent inhibitor of rev-rre binding. *J. Am. Chem. Soc, 122*(5), 980-981.
44. Kjaer-Frifeldt, S., Hansen, T. F., Nielsen, B. S., Joergensen, S., Lindebjerg, J., Soerensen, F. B., ... & Jakobsen, A. (2012). The prognostic importance of miR-21 in stage II colon cancer: a population-based study. *Br J Cancer*, *107*(7), 1169-1174.
45. Kozomara, A., and Griffiths-Jones, S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, 42, D68−73.
46. Leeper, T, Athanassiou, Z., Dias, R., Robinson, J. A., & Varani, G. (2005). TAR RNA Recognition by a Cyclic Peptidomimetic of Tat Protein. *Biochemistry, 44*(37), 12362-12372.

47. Leulliot, N., Varani, G. (2001). Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry, 40*(27), 7947-7956.

48. Ling, H., Fabbri, M., Calin, G.A. (2013). MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat Rev Drug Discov, 12*(11), 847-865.

49. Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*, *8*(6), 479-490.

50. Lättig-Tünnemann, G., Prinz, M., Hoffmann, D., Behlke, J., Palm-Apergi, C., Morano, I., Herce, H.D., Cardoso, M.C. (2011). Backbone rigidity and static presentation of guanidinium groups increases cellular uptake of arginine-rich cell-penetrating peptides. *Nat Commun*, 2, 453.

51. Madan, B., Seo, S.Y., Lee, S.G. (2014). Structural and sequence features of two residue turns in β-hairpins. *Proteins., 82*(9), 1721-33.

52. Maiti, M., Nauwelaerts, K., & Herdewijn, P. (2012). Pre-microRNA binding aminoglycosides and antitumor drugs as inhibitors of Dicer catalyzed microRNA processing. *Bioorg Med Chem Lett, 22*(4), 1709-1711.

53. Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V., Cullen, B.R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature, 338*(6212), 254–257.

54. Mann, D. A., Mikaélian, I., Zemmel, R. W., Green, S. M., Lowe, A. D., Kimura, T., ... & Karn, J. (1994). 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*, *241*(2), 193-207.

55. Mason, J.M. (2010). Design and development of peptides and peptide mimetics as antagonists for therapeutic intervention. *Future Med Chem*, *2*(12), 1813-1822.

56. Mei, H. Y., Cui, M., Heldsinger, A., Lemrow, S. M., Loo, J. A., Sannes-Lowery, K. A., ... & Czarnik, A. W. (1998). Inhibitors of protein− RNA complexation that target the RNA: specific recognition of human immunodeficiency virus type 1 TAR RNA by small organic molecules. *Biochemistry*, 37(40), 14204-14212.

57. Meng, F., Henson, R., Wehbe–Janek, H., Ghoshal, K., Jacob, S. T., & Patel, T. (2007). MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology, 133*(2), 647-658.

58. Moehle, K., Athanassiou, Z., Patora, K., Davidson, A., Varani, G., & Robinson, J. A. (2007). Design of β-Hairpin Peptidomimetics That Inhibit Binding of α-Helical HIV-1 Rev Protein to the Rev Response Element RNA. *Angew Chem Int Ed Engl, 46*(47), 9101-9104.

59. Muppidi, A., Doi, K., Edwardraja, S., Drake, E. J., Gulick, A. M., Wang, H. G., & Lin, Q. (2012). Rational design of proteolytically stable, cell-permeable peptide-based selective Mcl-1 inhibitors. *J Am Chem Soc, 134*(36), 14734-14737.

60. Murchie AI, Davis B, Isel C, Afshar M, Drysdale MJ, Bower J, … & Karn J. (2004). Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. *J Mol Biol, 336*(3), 625-638.

61. Nair, C. M., Vijayan, M., Venkatachalapathi, Y. V., & Balaram, P. (1979). X-Ray crystal structure of pivaloyl-D-Pro-L-Pro-L-Ala-N-methylamide; observation of a consecutive β-turn conformation. *J Chem Soc Chem Commun*, (24), 1183-1184.

62. Niu, Y., Wu, H., Varani, G., & Cai, J. (2011). γ-AApeptides bind to RNA by mimicking RNA-binding proteins. *Org Biomol Chem, 9*(19), 6604-6609.

63. Palacino, J., Swalley, S. E., Song, C., Cheung, A. K., Shu, L., Zhang, X., ... & Beibel, M. (2015). SMN2 splice modulators enhance U1–pre-mRNA association and rescue SMA mice. *Nat Chem Biol, 11*(7), 511-517

64. Pelay-Gimeno, M., Glas, A., Koch, O., Grossmann, T.N. (2015). Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. *Angew Chem Int Ed Engl, 54*(31), 8896-8927.

65. Puglisi, J. D., Chen, L., Blanchard, S., & Frankel, A. D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science, 270*(5239), 1200-1203.

66. Renukuntla, J., Dutt, A., Patel, A., Boddu, S. H. S., & Mitra, A. K. (2013). Approaches for enhancing oral bioavailability of peptides and proteins. *Int J Pharm X, 447*(1-2), 75–93.

67. Ribas, J., & Lupold, S. E. (2010). The transcriptional regulation of miR-21, its multiple transcripts and their implication in prostate cancer. *Cell Cycle, 9*(5), 923-929.

68. Ratni, H., Karp, G. M., Weetall, M., Naryshkin, N. A., Paushkin, S. V., Chen, K. S., ... & Zhang, X. (2016). Specific correction of alternative survival motor neuron 2 splicing by small molecules: discovery of a potential novel medicine to treat spinal muscular atrophy. *Int J Med Chem, 5*9(13), 6086-6100.

69. Robinson, J. A. (2000). The design, synthesis and conformation of some new β-hairpin mimetics: novel reagents for drug and vaccine discovery. *Synlett, 2000*(04), 429-441.

70. Robinson, J.A. (2008). B-hairpin peptidomimetics: design, structures and biological activities. *Acc Chem Res, 41*(10),1278-1288.

71. Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., & Mann, R. S. (2010). Origins of Specificity in Protein-DNA Recognition. *Annu Rev Biochem, 79*, 233–269.

72. Rosenström, U., Sköld, C., Lindeberg, G., Botros, M., Nyberg, F., Karlén, A., & Hallberg, A. (2006). Design, synthesis, and incorporation of a β-turn mimetic in angiotensin II forming novel pseudopeptides with affinity for AT1 and AT2 receptors. *Int J Med Chem, 49*(20), 6133-6137.

73. Roy, S., Delling, U., Chen, C. H., Rosen, C. A., & Sonenberg, N. (1990). A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev, 4*(8), 1365-1373.

74. Shortridge, M. D., Walker, M. J., Pavelitz, T., Chen, Y., Yang, W., & Varani, G. (2017). A Macrocyclic Peptide Ligand Binds the Oncogenic MicroRNA-21 Precursor and Suppresses Dicer Processing. *ACS Chem Biol, 12*(6), 1611−1620.

75. Shortridge, M. D., Wille, P. T., Jones, A. N., Davidson, A., Bogdanovic, J., Arts, E., ... & Varani, G. (2018). An ultra-high affinity ligand of HIV-1 TAR reveals the RNA structure recognized by P-TEFb. *Nucleic Acids Res, 47*(3), 1523-1531.

76. Sibanda, B. L., Blundell, T. L., & Thornton, J. M. (1989). Conformation of β-hairpins in protein structures: A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J Mol Biol, 206*(4), 759-777.

77. Sibanda, B. L., & Thornton, J. M. (1985). β-Hairpin families in globular proteins. *Nature, 316*(6024), 170-174.

78. Stein, C. A., & Goel, S. (2011). Therapeutic oligonucleotides: the road not taken. *Clin Cancer Res, 17*(20), 6369-6372.

79. Teixido, M., & Giralt, E. (2017). Jumping Hurdles: Peptides Able To Overcome Biological Barriers. *Acc Chem Res, 50*(8), 1847-1854.

80. Varani, G., Aboul-ela, F., & Allain, F. H. T. (1996). NMR investigation of RNA structure. *Prog Nucl Magn Reson Spectrosc, 29*(1-2), 51-127.

81. Wang, J., Lon, H. K., Lee, S. L., Burckart, G. J., & Pisetsky, D. S. (2015). Oligonucleotide-Based Drug Development: Considerations for Clinical Pharmacology and Immunogenicity. *Ther Innov Regul Sci, 49*(6), 861-868.

82. Warner, K. D., Hajdin, C. E., & Weeks, K. M. (2018). Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov*, *17*(8), 547-558.

83. Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H., & Jones, K. A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell, 92*(4), 451-462.

84. Werder, M., Hauser, H., Abele, S., & Seebach, D. (1999). β-Peptides as Inhibitors of Small-Intestinal Cholesterol and Fat Absorption. *Helvetica Chimica Acta, 82*(10), 1774-1783.

85. Whitby, L. R., Ando, Y., Setola, V., Vogt, P. K., Roth, B. L., & Boger, D. L. (2011). Design, synthesis, and validation of a β-turn mimetic library targeting protein–protein and peptide–receptor interactions. *J Am Chem Soc, 133*(26), 10184-10194.

86. Williamson, J. R. (2000). Induced fit in RNA–protein recognition. *Nat Struct Mol Biol, 7*(10), 834-837.

87. Wilson, W. D., & Li, K. (2000). Targeting RNA with small molecules. *Curr Med Chem, 7*(1), 73-98.

88. Witt, K. A., Gillespie, T. J., Huber, J. D., Egleton, R. D., & Davis, T. P. (2001). Peptide drug modifications to enhance bioavailability and blood-brain barrier permeability. *Peptides, 22*(12), 2329-2343.

89. Yan, L.X., Huang, X.F., Shao, Q., Huang, M.Y., Deng, L., Wu, Q.L., Zeng, Y.X., Shao, J.Y. (2008). MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA, 14*(11) 2348-2360.

90. Ye, X., Kumar, R. A., & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem Biol, 2*(12), 827-840.

BIV Tat R_{68} P R G **T R G K G R R I** R R_{81}

BIV-0 ^{D}P R G **T R G K G R R I** R ^{L}P

	1	2	3	4	5	6	7	8	9	10	11	12	K_d (uM)
BIV-1	I	R	G	**T**	**R**	**G**	**K**	**R**	**R**	**I**	R	V	30
BIV-2	R	V	R	**T**	**R**	**G**	**K**	**R**	**R**	**I**	R	V	0.15
BIV-3	I	Y	R	**T**	**R**	**G**	**K**	**R**	**R**	**I**	R	T	nb
BIV-4	Y	R	G	**T**	**R**	**G**	**K**	**R**	**R**	**I**	Y	V	>50
BIV-5	R	R	G	**T**	**R**	**G**	**K**	**R**	**R**	**I**	G	R	1-2
BIV-6	V	R	G	**T**	**R**	**G**	**K**	**R**	**R**	**I**	K	Y	>50
BIV-7	V	R	R	**T**	**R**	**G**	**K**	**R**	**R**	**I**	K	Y	nb
BIV-8	K	R	G	**T**	**R**	**G**	**K**	**R**	**R**	**I**	G	Y	>50

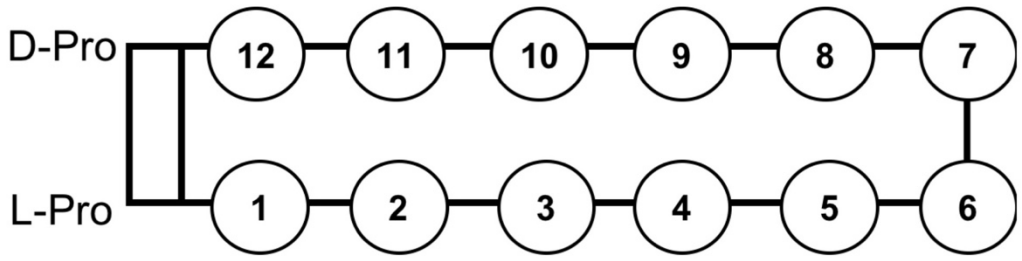


Table 2.1 Sequences of the BIV-Tat ARM (68-81), peptide mimic BIV-0, and β-hairpin mimetics BIV-1 – BIV-8. Reported K_ds were measured by electromobility shift assays (EMSA) (nb = no binding) (Athanassiou et al, 2004). Position numbering based on the schematic representation shown below.

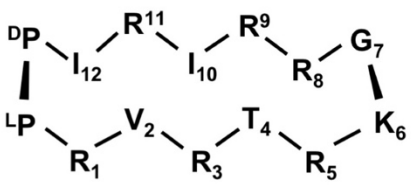
	1	2	3	4	5	6	7	8	9	10	11	12	K_d BIV-TAR (nM)
BIV-2	R	V	R	T	R	G	K	R	R	I	R	V	150
L-42	R	V	R	T	R	G	K	R	R	I	R	**I**	100
L-46	R	V	R	T	R	G	K	R	R	I	**O**	V	100
L-51	R	**T**	R	T	R	G	K	R	K	I	R	V	50
L-59	R	V	R	**Q**	R	G	K	R	R	I	R	V	50
L-72	R	V	R	T	R	G	K	R	R	I	**O**	**I**	50
L-76	R	**T**	R	T	R	G	K	R	R	I	**O**	**I**	20
L-78	R	**T**	R	**Q**	R	G	K	R	R	I	**O**	**I**	50

Table 2.2 Combination of sequences and K_ds of peptides from the L-1 to L-86 peptide where single residue changes improved binding affinity relative to BIV-2 (L-42, L-46, L-51, L-59). Single residue changes were incorporated together as peptides with multiple position changes (L-72, L-76, L-78) (Athanassiou et al, 2007).

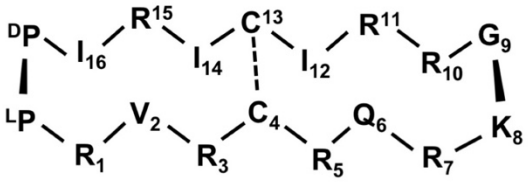
	1	2	3	4	5	6	7	8	9	10	11	12	K_d (nM), HIV	K_d (nM), BIV
L-22	R	V	R	T	R	K	G	R	R	I	R	I	30	5
L-50	R	V	R	T	R	G	K	R	R	I	R	R	1	1
L-51	R	T	R	T	R	K	G	R	R	I	R	V	5	50

Table 2.3 Sequences and K_ds for peptides with highest-affinity targeting HIV-TAR from the L-1 to L-86 library (Davidson et al., 2009).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	K_d (nM)
L-22	R	V	R	T	R	K	G	R	R	I	R	I					30
KP-Z-40	R	V	R	C	R	Q	R	K	G	R	R	**A**	C	I	R	I	25
KP-Z-41	R	V	R	C	R	Q	R	K	G	R	R	**I**	C	I	R	I	75



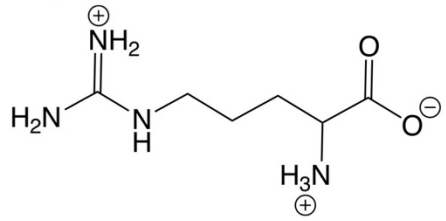
L-22



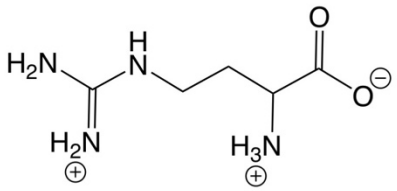
KP-Z-41

Table 2.4 Sequences and K_ds for peptides with highest-affinity to HIV-TAR from the KP-Z library (Davidson, Patora-Komisarska, Robinson, & Varani, 2011).

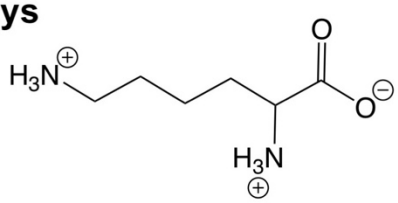
Arg



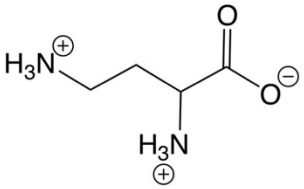
noR



Lys



Dab



	1	2	3	4	5	6	7	8	9	10	11	12	K_d (nM)
L-22	R	V	R	T	R	K	G	R	R	I	R	I	30
JB-59	**Dab**	V	R	T	R	K	G	R	R	I	R	I	1
JB-82	R	V	R	T	R	K	G	R	R	I	**noR**	I	75
JB-181	**Dab**	V	R	T	R	K	G	R	R	I	**noR**	I	0.03
JB-190	Dab	V	R	T	R	**G**	**K**	R	R	I	noR	I	1.5

Table 2.5 (top) Most effective non-canonical amino acids used in the JB peptide series. **(bottom)** Sequences and K_ds for peptides with the highest HIV-TAR affinities from the JB library (Shortridge et al, 2018).

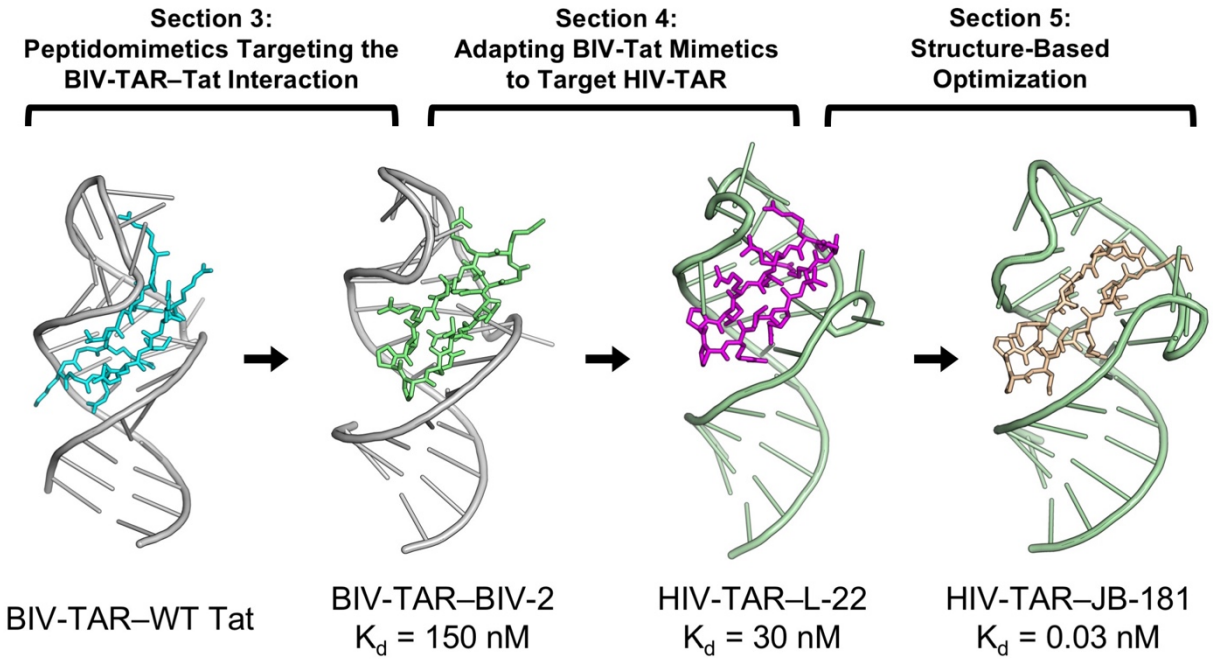


Figure 2.1 Schematic depicting the three design stages described in Sections 3-5 that led to the discovery of very potent RNA-targeting macrocyclic peptides. **From left to right:** Bovine immunodeficiency virus trans-activation response (BIV-TAR) is a structured *cis*-regulatory element in the BIV 5' UTR mRNA that is recognized by BIV Tat to stimulate viral transcription (PDB ID: 1MNB; Puglisi, Chen, Blanchard, & Frankel, 1995); **Section 3** overviews the process of using peptidomimetics to target the BIV-TAR—Tat interaction by grafting the wildtype Tat sequence onto a β-hairpin scaffold. This strategy led to the discovery of peptide BIV-2 and its structure determination by NMR (PDB ID: 2A9X; Leeper, Athanassiou, Dias, Robinson, & Varani, 2005); **Section 4** describes how a large positional scanning peptide library was organized to generate structure activity relationship (SAR), from which a BIV-Tat mimetic was adapted to target human immunodeficiency virus TAR (HIV-TAR) (PDB ID: 2KDQ; Davidson et al., 2009); **Section 5** describes the structure-based optimization process and peptide design strategies that incorporated non-canonical amino acid substitution (e.g. backbone extension, side-chain truncation) to achieve low picomolar binding, JB-181 (PDB ID: 6D2U; Shortridge et al, 2018).

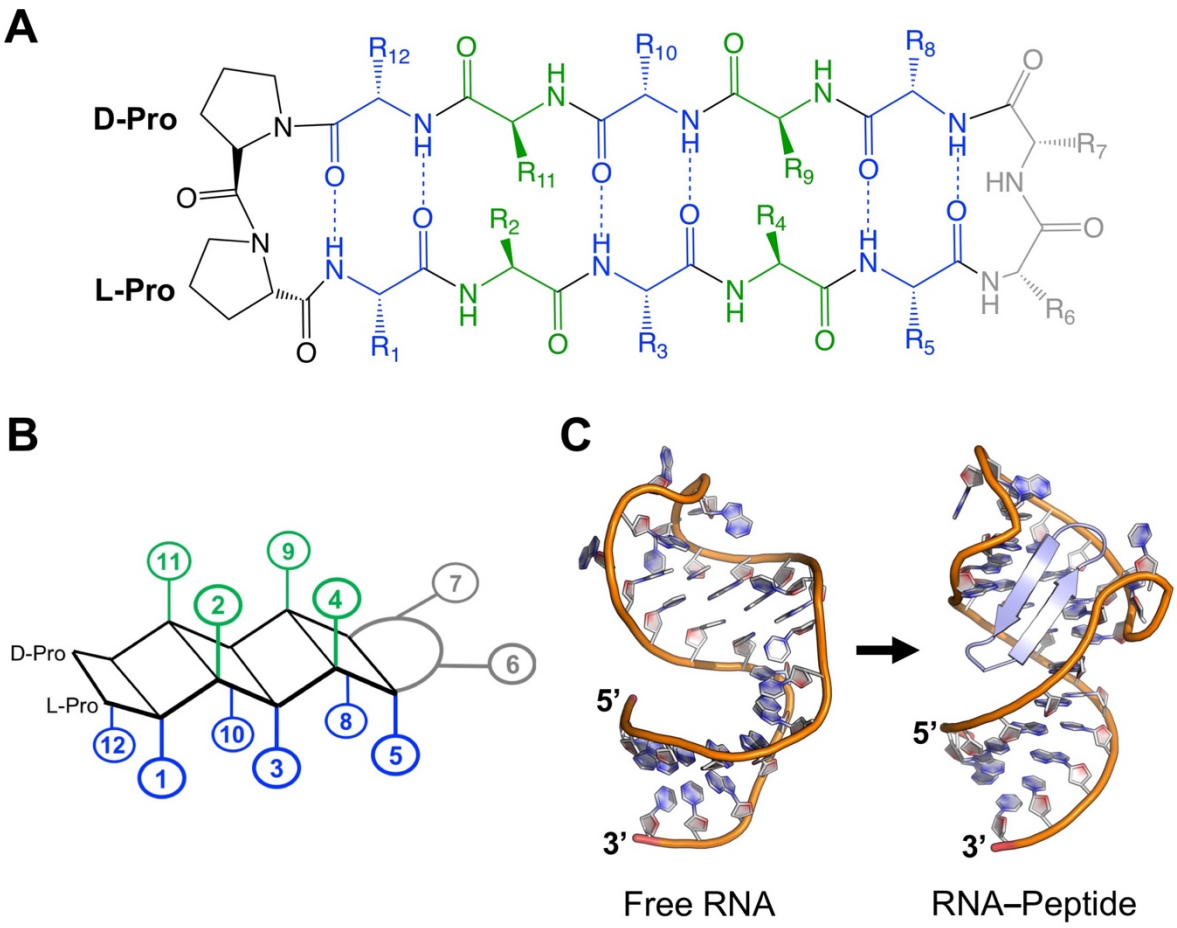


Figure 2.2 (A) Generic 2D structure of a 12-mer β-hairpin macrocycle with a D-Pro/L-Pro template. **(B)** Representation of a 12-mer β-hairpin macrocycle with sidechains grouped by orientation relative to the backbone plane. Positions 2, 4, 9, 11 point 'up' (green); while positions 1, 3, 5, 8, 10, 12 point 'down' (blue). **(C)** β-hairpin scaffolds match the shape and size of the RNA major groove and mold the RNA structure. From left to right: (PDB ID: 1ANR; Aboul-ela, Karn, & Varani, 1996), (PDB ID: 6D2U; Shortridge et al, 2018).

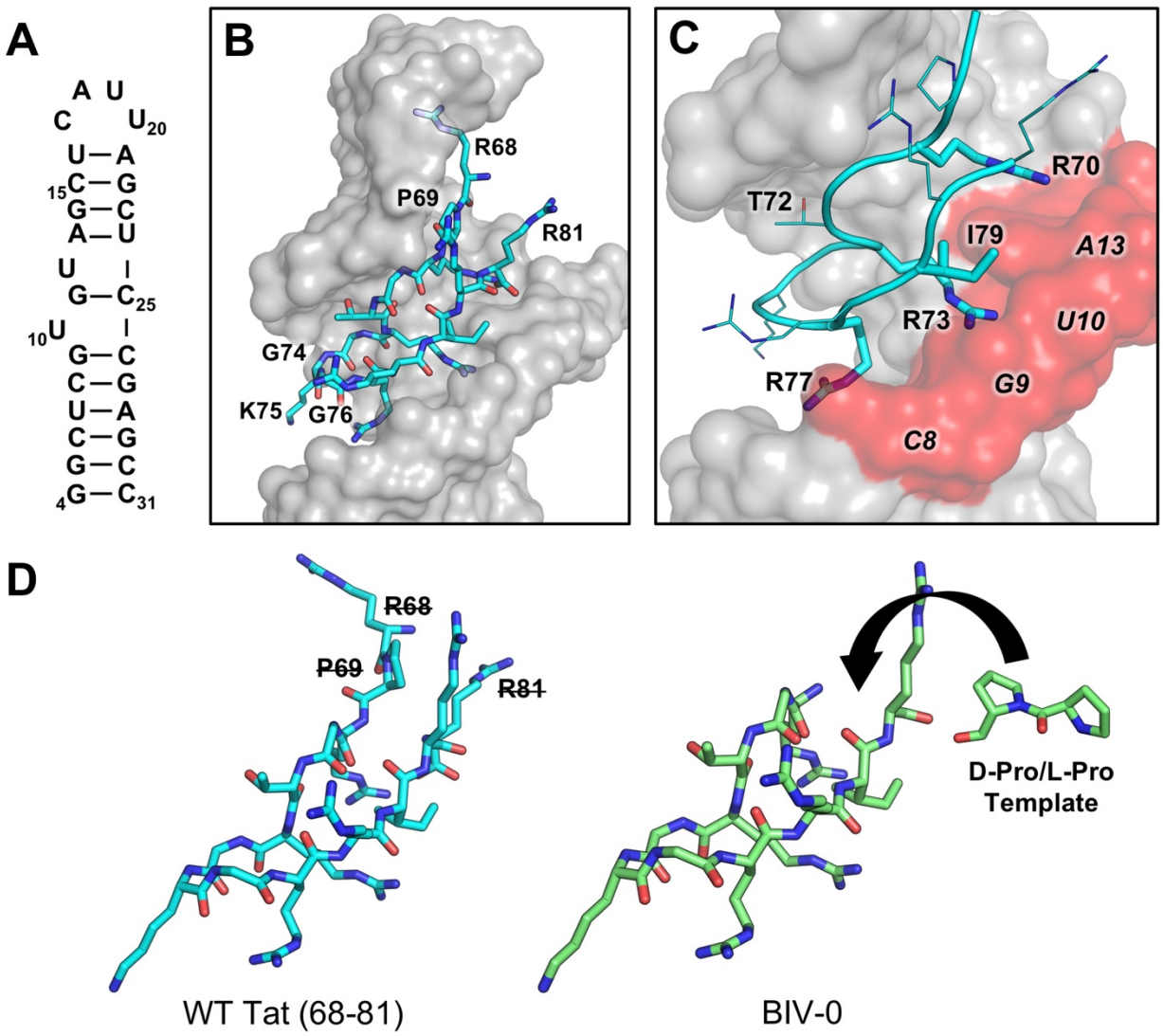


Figure 2.3 (A) Secondary structure of BIV-TAR. **(B)** BIV Tat arginine-rich motif (cyan, 68-81) binds to within the BIV-TAR (grey) major groove at the stem-bulge and stem-apical loop interface (PDB ID: 1MNB; Puglisi, Chen, Blanchard, & Frankel, 1995). **(C)** Arg-77, Arg-73, and Arg-70 are critical recognition elements for BIV-TAR binding by sandwiching between nucleotides C8, G9, U10, G11 forming arginine sandwich motifs (ASM). **(D)** BIV-0 cyclization was designed by eliminating R68, P69, and R81 and grafting a D-Pro/L-Pro template (Athanassiou et al., 2004).

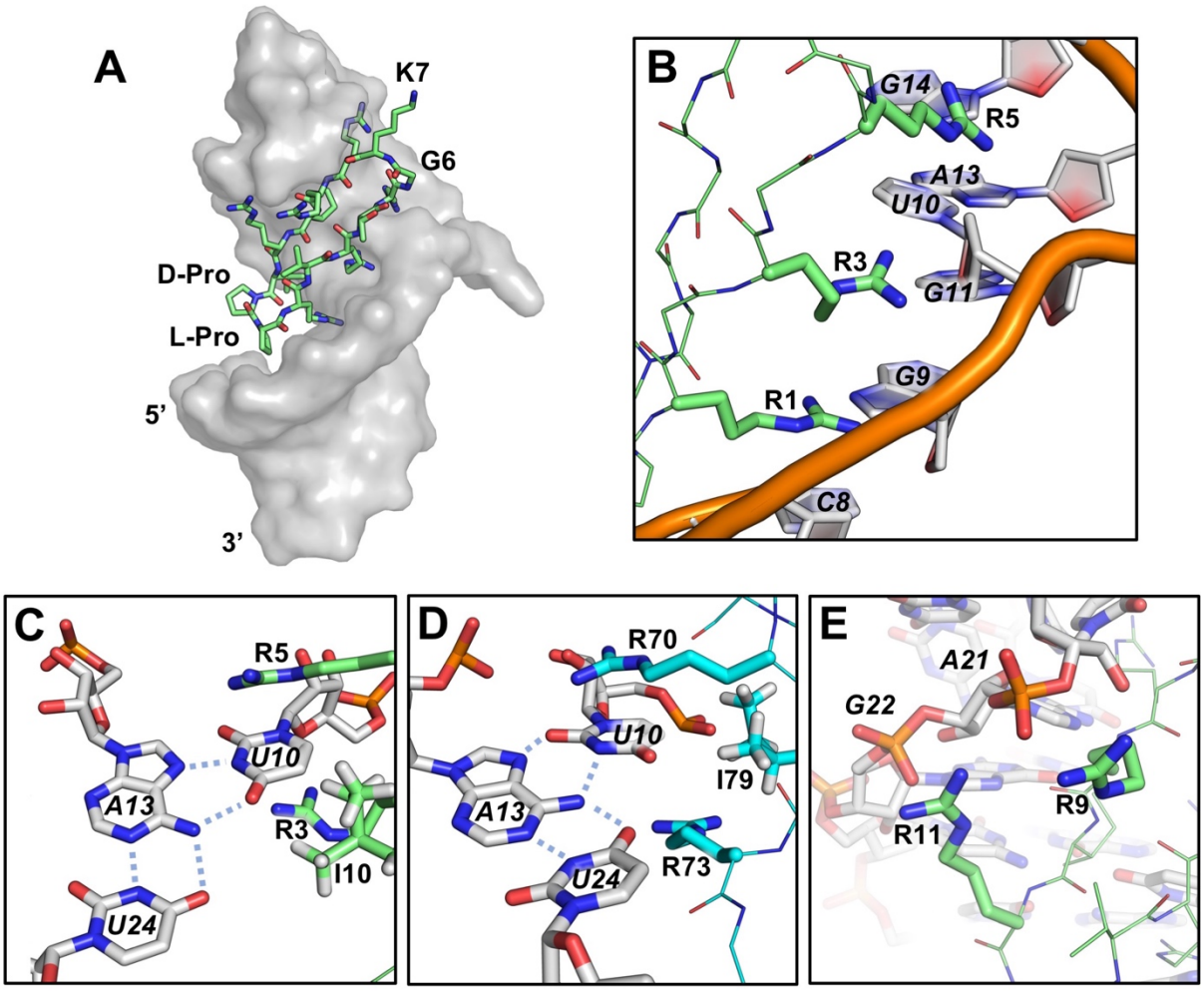


Figure 2.4 (A) BIV-2 (green) bound to BIV-TAR (grey) upside down relative to wildtype BIV-Tat. (PDB ID: 2A9X; Leeper, Athanassiou, Dias, Robinson, & Varani, 2005) **(B)** ASM formed by Arg-1, Arg-3 and Arg-5 intercalating between C8, G9, U10, and G11. **(C&D)** The base triple induced by BIV-2 mimics that induced by BIV Tat (cyan) (PDB ID: 1MNB; Puglisi, Chen, Blanchard, & Frankel, 1995); Arg-5, Arg-3, Ile-10 mimic Arg-70, Arg-73, Ile-79, respectively. (**E)** Solvent exposed Arg-9 and Arg-11 make electrostatic interactions with backbone phosphates at A21 and G22.

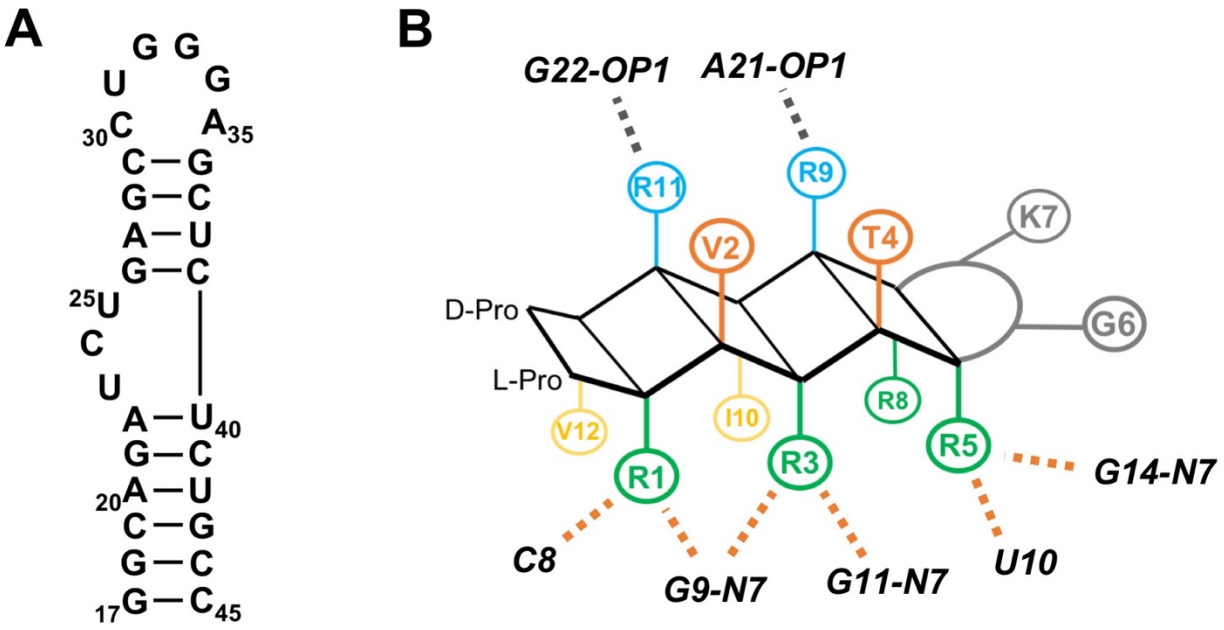


Figure 2.5 (A) Secondary structure of HIV-TAR. **(B)** Cartoon depicting SAR for BIV-2 binding to BIV-TAR. Orange dotted lines indicate hydrogen bonding and/or π-stacking interactions, grey dotted indicate electrostatic interactions (Athanassiou et al, 2007).

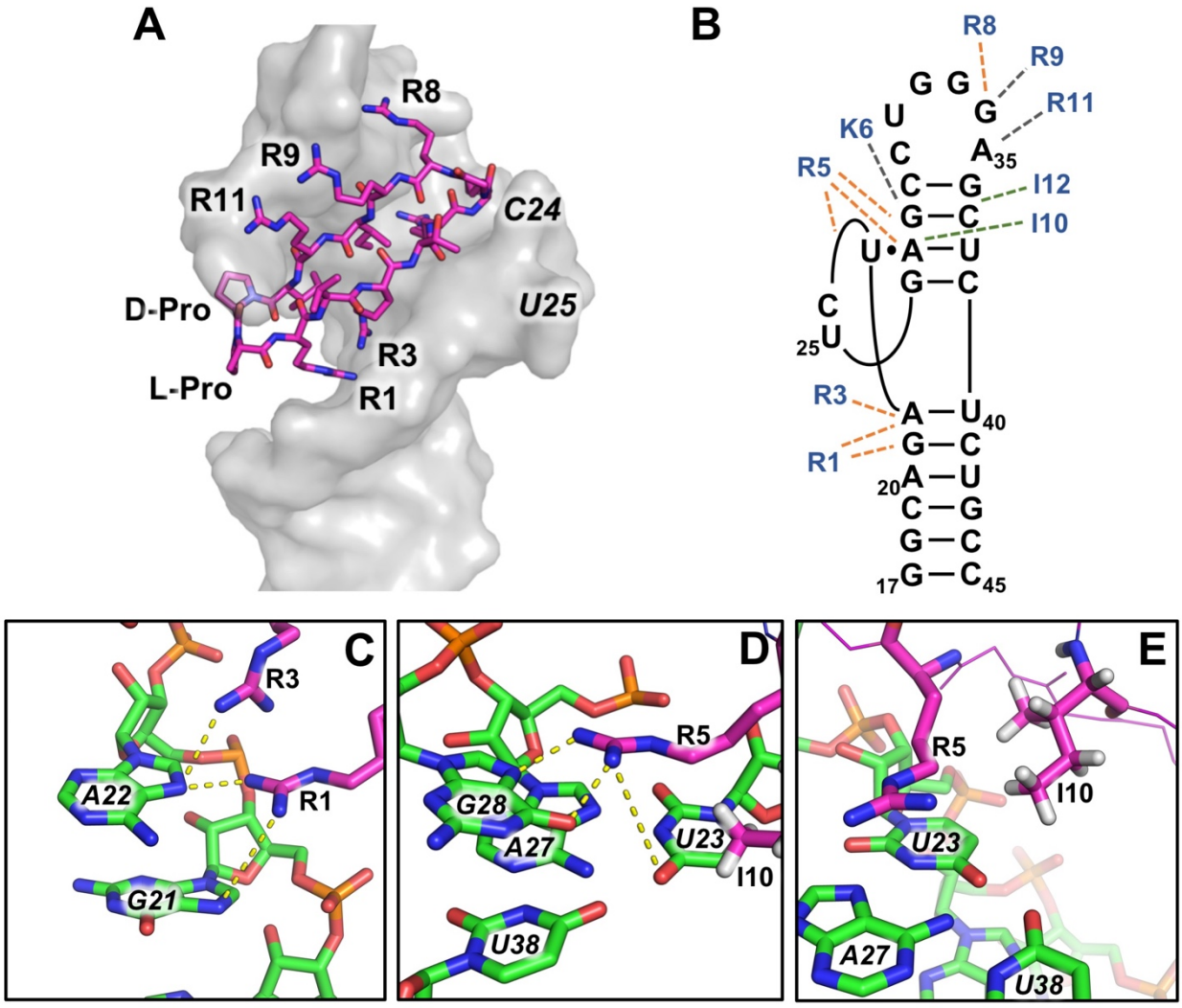


Figure 2.6 (A) L-22 (magenta) bound HIV-TAR (grey). (PDB ID: 2KDQ; Davidson et al, 2009) **(B)** SAR for L-22 binding to HIV-TAR. Orange dotted lines indicate hydrogen bonding and/or π-stacking interactions, green dotted lines indicate hydrophobic interactions, grey dotted indicate electrostatic interactions. **(C)** Hydrogen bonds between Arg-1 with G21 N7 and Arg-3 with A22 N7 stabilize the stem-bulge interface. **(D)** Hydrogen bonds between R5 and U23 and G28 induce the U23-A27-U38 base triple. **(E)** Ile-10 contributes to base triple stabilization by hydrophobic interactions with the H5 and H6 side of U23.

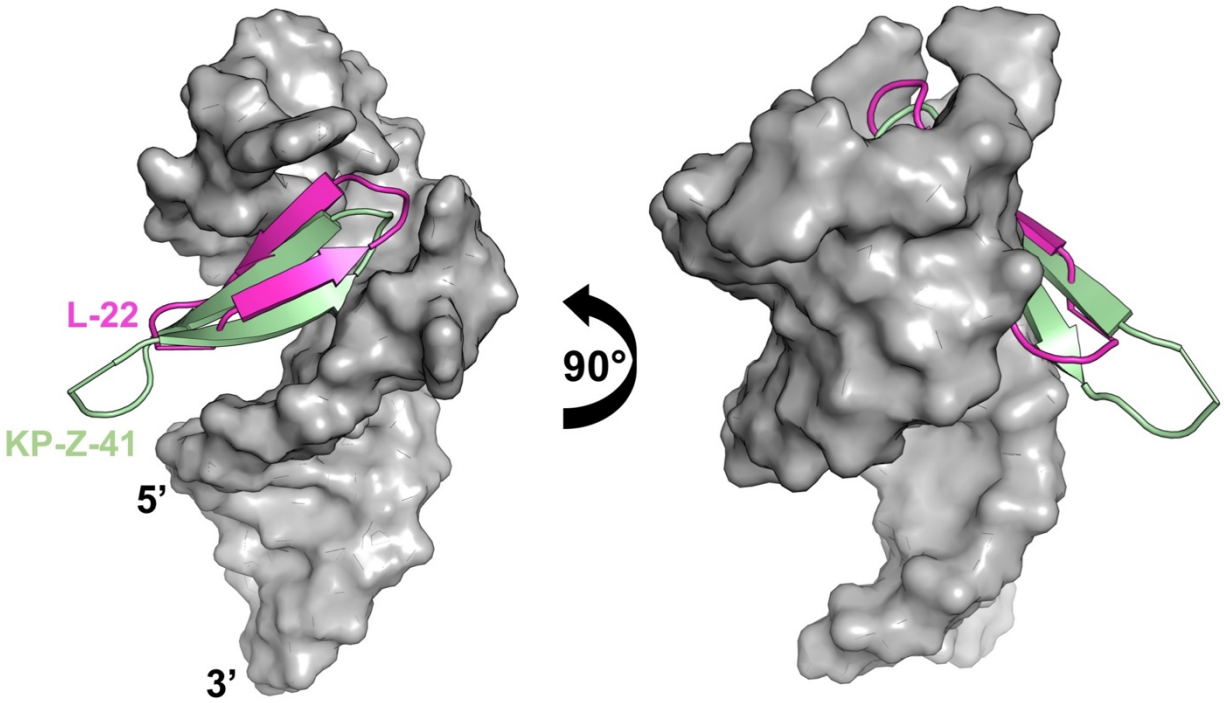


Figure 2.7 Overlaid HIV-TAR-bound structures (grey) for L-22 (magenta) and KP-Z-41 (pale green) (PDB ID: 2KX5, Davidson, Patora-Komisarska, Robinson, & Varani, 2011). HIV-TAR backbone phosphates were used as anchor points for the superposition.

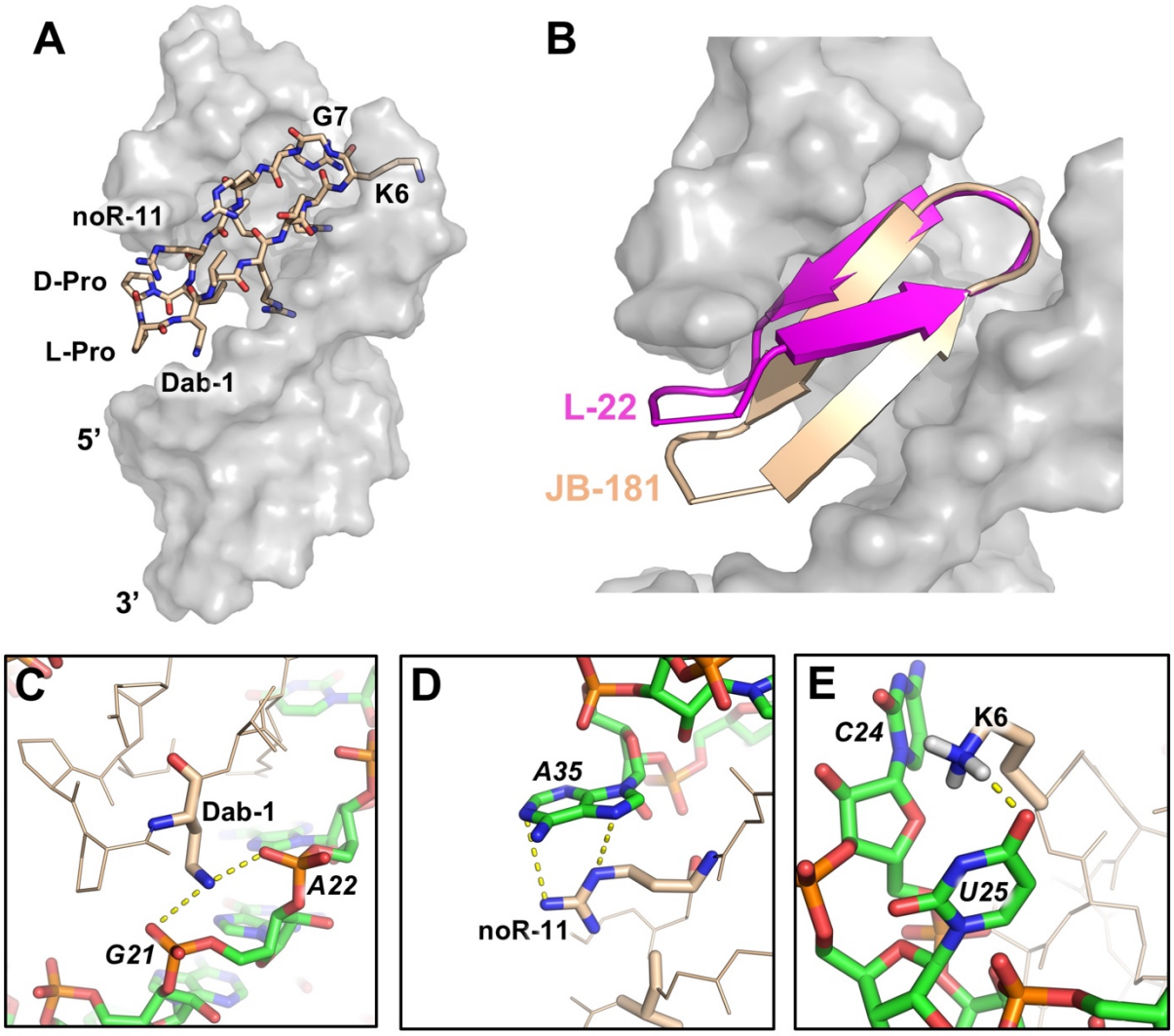


Figure 2.8 (A) JB-181 (wheat) bound to HIV-TAR (grey) (PDB ID: 6D2U; Shortridge et al, 2018). **(B)** Overlaid of HIV-TAR-bound structures (grey) for L-22 (magenta) (PDB ID: 2KDQ; Davidson, 2009) and JB-181 (wheat). HIV-TAR backbone phosphates were used as anchor points for the superposition. **(C)** Dab-1 forms salt bridges with the G21 and A22 phosphates. **(D)** noR-11 makes a cation-π interaction with A35 favoring a *syn* conformation for the A35 base. **(E)** Change in peptide position favors formation of a novel hydrogen bond between Lys-6 and U25 O4.

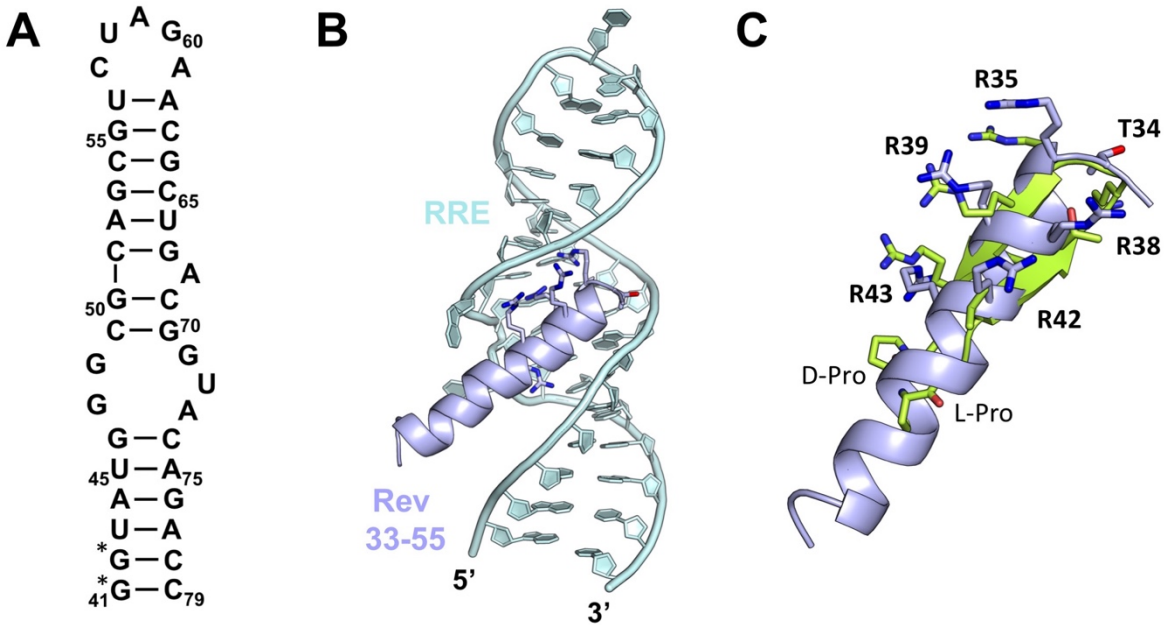


Figure 2.9 (A) Secondary structure of stem loop IIb of the Rev Response Element (RRE). (*) indicated next to residues changed for increased T7 transcription efficiency. **(B)** α-helical Rev ARM (33-55, light blue) bound to the RRE major groove (light cyan) (PDB ID: 1ETF; Battiste et al, 1996). **(C)** A standard 2:2 β-hairpin mimetic (residues 1-12 with a D-Pro/L-Pro template) (green) superimposed onto the Rev residues critical for RRE binding demonstrating how β-hairpin scaffolds can be adapted to mimic similar α-helix sidechain positioning.

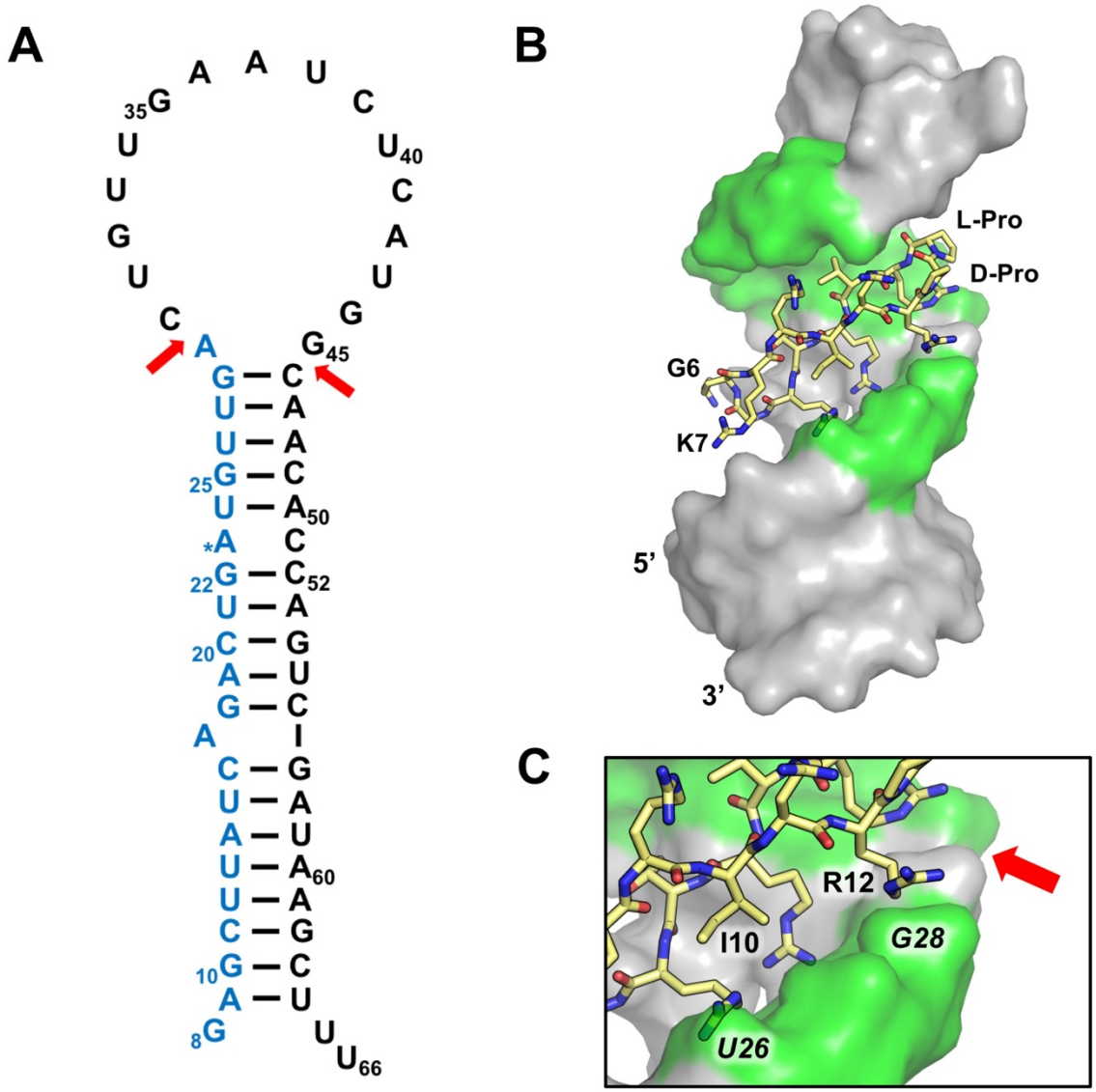


Figure 2.10 (A) Secondary structure of pre-miRNA-21 (pre-miR-21; numbers based on miRBASE v21 (Kozomara, & Griffiths-Jones, 2014), the Dicer cleavage sites are indicated by red arrows to generate the ~22 nucleotide mature miR-21 (blue). **(B)** Structure of pre-miR-21-L50 complex with nucleotides that make intermolecular NOEs highlighted in green (PDB ID: 5UZZ; Shortridge et al, 2017). **(C)** Both Ile-10 and Arg-12 contact the RNA near one of the Dicer cleavage sites.

Chapter 3 – Development of microRNA Processing Assays for Evaluating Inhibition of microRNA Maturation

Section 1: Introduction

1.1 Inhibition of Pre-miRNA Processing

MicroRNAs (miRNA) have a major impact on healthy cell regulation and are implicated in various diseases (Bentwich et al, 2005; van Rooij et al, 2014; Cheng et al, 2015; Hsu et al, 2006; Schetter et al, 2008). Thousands of miRNA genes exist in the human genome, with each undergoing the same intricate processing pathway to produce functional mature miRNAs through their biogenesis (Wilson et al, 2013). MiRNA biogenesis involves processing of the native miRNA primary transcript through a series of cleavage reactions executed by two ribonuclease complexes: Drosha-DGCR8 and Dicer-TRBP (**Fig. 3.1**) (Wilson et al, 2013; Carthew et al, 2009; Nguyen et al, 2015). The final product of both reactions is a double-stranded 'mature' miRNA, which then associates with the RNA-inducing silencing complex (RISC) to target mRNAs by complementary base pairing (MacRae et al, 2008; Paroo et al, 2007; Wang et al, 2009). Targeting involves specific base-pairing with partially complementary elements usually within the 3' UTR of target genes, which results in suppression of translation by inducing mRNA degradation or disruption of ribosomal complexes (Wilson et al, 2013; Carthew et al, 2009; Nguyen et al, 2015; Wang et al, 2009).

1.2 Targeting Pre-miRNA 21

MicroRNA-21 (miR-21) is considered oncogenic as it represents as much as 25% of total cellular miRNAs in cancer cells, is overexpressed in most human tumors (heart, lung, liver, pancreas, prostate, mammary gland, lymphnode, brain, etc), silencing translation of various tumor suppressors genes such as: BRG1, PDCD4, ANP32A, SMARCA4 (Medina et al, 2010; Krichevsky et al, 2008; Ribas et al, 2010; Kjaer-Frifeldt et al, 2012; Yan et al, 2008; Asangani et al, 2008; Meng et al, 2007; Sayed et al, 2008; Zhu et al, 2007; Papagiannakopoulos et al, 2008; Wickramashinghe et al, 2009; Chan et al, 2005; Loffler et al 2007; Ma et al, 2011). Many screens using various chemistries (e.g. small molecules, peptides, oligonucleotides) have attempted to target miR-21 by reducing its levels through inhibition of processing (Naro et al, 2015; Bose et al, 2015; Esau et al, 2008; Chirayil et al, 2009; Bose et al, 2012; Gumireddy et al, 2008); however, almost all examples lack clear biophysical evidence to support ligand-RNA complex formation. Therefore, utilizing NMR spectroscopy to determine ligand-RNA structures coupled with biochemical assays for validation may prove to be a novel, effective strategy for developing ligands with high affinity for pre-miR-21 that would inhibit its processing. If successful, a miR-21 inhibitor could suppress cancer cell proliferation, establishing miR-21 as a valid target for treating various forms of cancer.

The pre-miR-21 secondary structure is a 59-nucleotide stem-loop that is the precursor to mature miR-21 and the substrate for the Dicer-TRBP endonuclease complex (**Fig. 3.2**) (Nicholson et al, 2013; Lee et al, 2013; Wilson et al, 2015; Kidwell et al, 2014; Foulkes et al, 2014). Dicer is a 255 kDa ribonuclease that associates with pre-miRNA through a dsRNA binding domain (dsRBD) and a PAZ domain that recognizes the RNA 3'-end, which together function as molecular rulers to facilitate cleavage 22 nts from the RNA 5'-end (Park et al, 2011).

The strategy for inhibiting pre-miR-21 processing seeks to discover ligands that can specifically bind pre-miR-21 and block Dicer cleavage. Efforts to achieve this aim have produced an engineered protein that binds the apical loop of pre-miR-21 (Fox RRM*) (Chen et al, 2016), and several β-hairpin cyclic peptides (Shortridge et al, 2017) with high affinity for the pre-miR-21 major groove. Inspection of the NMR structure of one such peptide, L50, bound to pre-miR-21 suggests both ligands may block the Dicer cleavage sites, which should result in inhibition of pre-miR-21 processing (**Fig. 3.2**). However, due to the lack of a high resolution Dicer crystal structure or a clear mechanism of action for human Dicer, this conjecture requires clear demonstration of *in vitro* inhibition of pre-miR-21 processing by these and other pre-miR-21 binding ligands (Taylor et al, 2013; Heller et al, 2007; Sawh et al 2012).

Section 2: Results

2.1 Dicer Assay using Commercial Enzymes (Genlantis Turbo Dicer™)

Initial *in vitro* Dicer activity assays were carried out using a commercially available kit, Turbo Dicer™ from Genlantis. Though conveniently purchased, the commercially available enzyme is designed to bind and cleave double stranded small interfering RNA (siRNA) more efficiently by fusion of a dsDNA binding domain. This design may explain why previous reports have exhibited reduced pre-miRNA processing efficiency when using the Turbo Dicer™ kit as pre-miRNAs are different in structure compared siRNAs. Pre-miRNAs are double stranded like siRNA, but also flexible and unpaired at their apical loops. Fox RRM*, a RNA binding splicing factor engineered to bind the pre-miR-21 apical loop at low nM affinity (Chen et al, 2013; Chen et al 2016), was assayed to validate whether its binding correlates with inhibition. The assays using commercially available enzyme only provided qualitative assessment for processing inhibition due to inconsistent Turbo Dicer™ activities when assays were repeated at 90 minute reaction times. Smeared mature miRNA bands were also detected indicative of cleavage at multiple sites along the pre-miRNA phosphate backbone. This inefficient cleavage efficiency observed in both the control reactions consisting of RNA and Dicer, and reactions including Fox RRM* is attributed to the aforementioned dsDNA binding domain design making *Turbo Dicer*™ specialized for siRNA, background RNase activity leftover from Genlantis protein preperation, and in addition the lack of stabilizing auxiliary protein TRBP essential for cellular processing.

2.2 Dicer Assay Development using Human Dicer-TRBP

Human Dicer in complex with its essential co-factor TRBP was purified in order to develop a more reliable pre-miR-21 processing assay. Expression required a baculovirus system

to avoid protein aggregation and allow post-translation modifications (Kost et al, 2005). Human Dicer-TRBP was successfully purified from Sf9 cells using a standard baculovirus expression system and purified by nickel column purification, followed by isolation of monomeric protein from aggregates by size-exclusion chromatography.

Preliminary assays with no inhibitor were tested to determine reaction and buffer conditions. Reaction conditions were initially modeled after previous kinetic studies on human Dicer (Chakravarthy et al, 2010; Lee et al, 2012), and assay buffer conditions were optimized further for developing an inhibition assay. Using the established Dicer assay conditions, our in-house Dicer-TRBP complex demonstrated similar activity on pre-miR-21 to previously published Dicer assays, with ~50% conversion of pre-miR-21 to its mature form after 80 minutes at 37°C (**Fig. 3.3**).

Activity with respect to varying NaCl concentration was monitored, since addition of NaCl is a key control for reducing non-specific binding when assaying RNA-binding ligands. Although most assays in the literature use only 25mM NaCl, we found similar activity at NaCl concentrations up to 100mM (**Fig. 3.4**), which is preferred for establishing whether or not a ligand binds to RNA by forming non-specific electrostatic interactions with the RNA phosphate backbone (Xavier et al, 2000).

$MgCl_2$ is essential for endonuclease activity due to Mg^{2+} forming favorable interactions with H_2O molecules and sidechains (e.g. glutamate, glutamine), both of being critical components for endonuclease catalysis (Nowotny et al, 2005). $MgCl_2$ concentration is of particular interest for pre-miR-21 processing since pre-miR-21 is processed much slower compared to most miRNAs (e.g Let-7) likely due to its large apical loop, therefore higher $MgCl_2$ concentrations could potentially amplify its low activity by increasing the fraction of Mg^{2+}

bound Dicer-TRBP (Acker et al, 2014). A $MgCl_2$ screen ranging from 0 to 10mM established several features for Dicer-TRBP processing of pre-miR-21 (**Fig. 3.5**). First, 0 mM $MgCl_2$ produces no detectable Dicer activity. This is favorable for setting up competition assays for probing inhibitor competition with Dicer to pre-miRNAs binding to ensure no pre-miR-21 processing while a binding reaction is equilibrated. Dicer dsRNA binding domains that initially associate with the pre-miRNA are distant from the Mg^{2+} binding sites that cleaves the RNA suggesting Mg^{2+} plays little role in pre-mIR-21 association. Second, our Dicer activity at endogenous levels of $MgCl_2$ (1.25mM) were comparable to previous literature results, confirming adequate purification of the complex. Third, unexpectedly, 90% cleavage of pre-miR-21 was observed at 10-20mM with no sign of Mg-induced non specific cleavage.

RNA concentrations used in inhibition assays were established based on a previous report listing the K_m for Dicer-TRBP processing of pre-miR-21 (Acker et al, 2014). The following relationship of substrate concentration and K_m was considered: if $K_m >> [S]$ then the reaction velocity is sensitive to changes in substrate concentration, whereas if $[S] >> K_m$, then the velocity in insensitive to changes in substrate concentration (Brooks et al, 2012, Strelow et al, 2012). Therefore, holding the substrate concentration around or below the K_m is ideal for assessing competitive inhibition.

Previous reports examining Dicer-TRBP activity using in-house purified enzyme require single turnover concentrations where Dicer-TRBP concentrations are greater than pre-miRNA in order to detect near complete conversion of all pre-miRNA to mature miRNA (Lee et al, 2010; Lee et al, 2012; Acker et al, 2014). This is attributed to the enzyme itself being inherently unstable when assayed *in vitro*, based on loss of activity on the order of hours. We also speculate that a large population of Dicer may be substrate inhibited due to its high yet non-specific

affinity for binding dsRNA (Lee et al, 2010; Lee et al, 2012). Pre-miRNAs bound to Dicer in orientations that disfavor in-line cleavage could possibly hinder proper binding orientation needed for cleavage.

2.3 Pre-miR-21 Processing Inhibition Assays

Following optimization, positive controls were used to develop the assay for measuring inhibition of Dicer activity. A DNA oligonucleotide was designed to base-pair with the pre-miR-21 apical loop, thereby unfolding the RNA to inhibit Dicer processing (Ling et al, 2013). The antisense titration assay with human Dicer showed >90% inhibition activity at low nanomolar concentration, and produced clean mature miR-21 bands, demonstrating a greater reliability of the assay compared to the Turbo Dicer™ assay for evaluating inhibition of processing (**Fig. 3.6**). These results concluded optimization and validated carrying out inhibition assays for Fox RRM* and the series of cyclic β-hairpins discovered in the group.

2.3.1 Inhibition of Dicer-TRBP by Proteins: Fox RRM*

Pre-miR-21 processing inhibition assays were carried out to evaluate if Fox RRM*, which binds pre-miR-21 at nM concentrations (K_d ~10 nM, Chen et al, 2016), inhibits processing at comparable concentrations. Inhibition was similar to the previously mentioned assays using Turbo Dicer™ (**Fig. 3.7**). Inhibition was slightly better using Turbo Dicer™ assays attributed to the difference in NaCl concentrations between the commercial reaction buffer compared to the in-house Dicer reaction buffer. These results suggest that poor correlation between K_d and inhibition for Fox RRM* inhibition of pre-miR-21 processing is not a product of untrustworthy commercial enzyme, and potentially a lack of competition with Dicer. This may be explained by

Fox RRM* being designed to recognize five consecutive nucleotides within the pre-miR-21 apical loop (Chen et al, 2016), which are distant from both of the Dicer dsRNA binding domains and its RNase cleavages sites.

2.3.2 Inhibition of Dicer-TRBP by Peptides: JB72, MDS001, MDS002

β-hairpin cyclic peptides from previous HIV-TAR targeting projects (Gibbs et al, 1998; Robinson et al, 2008; Davidson et al 2011; Lalonde et al, 2011; Davidson et al, 2009; Athanassiou et al, 2007; Athanassiou et al, 2004) were screened for binding to pre-miR-21 and several high affinity binders for pre-miR-21 were identified (**Fig. 3.8**). JB72 was discovered to bind to pre-miR-21 (K_d = 200nM), acting as a starting point for generation of peptides of related sequence, such as MDS001 and MDS002, that would bind better to pre-miR-21 (**Fig. 3.8**). Therefore, *in vitro* Dicer assays were carried out to correlate pre-miR-21 binding affinity with functional inhibition of processing.

Similar to Fox RRM*, JB72 exhibited low μM inhibition of pre-miR-21 processing with poor cleavage efficiency with both Turbo Dicer™ and human Dicer-TRBP assays (**Fig. 3.9, Fig. 3.10**). Although marginal, MDS002 inhibition improved compared to JB72 (**Fig. 3.10**) showing that design decisions that improved MDS002 binding also lead to improved inhibition. However, since pre-miR-21 degradation was detected in the course of these assays, more careful characterization of this secondary activity will be required.

2.3.3 Inhibition of Dicer-TRBP by Peptides: L50

L50 is a previously characterized cyclic peptide from HIV-TAR targeting efforts that binds to pre-miR-21 (K_d = 200nM) (Shortridge et al, 2017) but not to an A29 deletion mutant.

This suggests L50 binding is dependent on A29, an unpaired base located at the Dicer cleavage site.

Turbo Dicer™ assays exhibited 50% disruption of Dicer activity at 10 μM L50 (**Fig. 3.11**). This result was higher than we expected, since the K_d is orders of magnitude lower. With in-house Dicer-TRBP assay inhibition of Dicer processing of pre-miR-21 at 10 μM L50 was also observed, but we also detected cleaner processing and a novel Dicer reaction intermediate only detected in the presence of L50 (not shown). Dicer processing intermediates are formed if pre-miRNAs dissociate from the enzyme prematurely with only a single backbone cleavage complete. In the context of pre-miR-21, one of the Dicer cleavage sites is located above the A29 bulge and at the i+1 cross-strand phosphate backbone, between G45 and C46. This cut site must be perturbed by L50 binding based on its dependence on the A29 bulge, suggesting L50 binding induces formation of this additional intermediate.

L50 inhibition with novel intermediate formation was followed up by running single reaction vessel time course assays to demonstrate that intermediate formation was not an artifact or gel loading error. Dicer reactions with both no L50 present as control and 10μM L50 were monitored over 40 minutes (**Fig. 3.12**). 10 μM L50 was needed to produce the novel intermediate and interestingly to produce a novel minor product likely from L50-induced alternative Dicer cleavage.

Follow-up reactions using Turbo Dicer™ showed no additional intermediate (**Fig. 3.13**). We attribute this result to the Turbo Dicer™ sequence being from *Giardia intestinalis* whereas the in-house Dicer-TRBP being the human sequence, suggesting subtle changes in sequence near the Dicer cleavage site confer different requirements for pre-miRNA positioning for in-line cleavage. However, several technical variables should be ruled out first. For instance,

intermediate formation only occurs at high concentrations of L50 when aggregation of reaction in gels is observed as well (**Fig. 3.12**). L50 may bind Dicer-TRBP rather than pre-miR-21 near the cleavage sites at high enough concentrations impeding expected cleavage. Traces of contaminating RNase in either in-house Dicer-TRBP, Turbo Dicer^{TM}, or L50 aliquots should be tested as well to rule out miR-21 cleavage by endonuclease other than Dicer-TRBP.

Section 3: Discussion

After comparing Dicer assays conducted using commercial Turbo Dicer™ enzyme versus human Dicer-TRBP, it is clear that human Dicer-TRBP processes miR-21 more efficiently. Human Dicer-TRBP should exclusively be used for evaluating miRNA inhibitors to avoid misleading results when using the commercially available Turbo Dicer™. Due to efficient processing and optimization of the assay, the results demonstrate that the human Dicer-TRBP activity assay is an effective tool for determining *in vitro* inhibition of pre-miR-21 processing by ligands and can be used to make ligand design decisions.

Why reactions using commercially available Dicer resulted in multiple sized bands for mature miR-21 is difficult to discern because details regarding Turbo Dicer™ sequence and storage buffer are treated as confidential information. We have speculated based on what has been reported (Myers et al, 2003) that it may be a combination of factors mentioned previously in the results section and summarized here: a dsDNA binding domain designed into Turbo Dicer™ for more efficient siRNA processing simultaneously reduces pre-miR-21 efficiency, Turbo Dicer™ being derived from the *Giardia intestinalis* Dicer sequence rather than human, lack of TRBP for pre-miR-21 stabilization, and possible residual endonuclease activity leftover from Turbo Dicer™ protein preparation.

The disparity of μM inhibition despite nM affinity for Fox RRM* and the cyclic peptides could be a result of use of overly high Dicer concentrations due to its very high affinity for pre-miRNAs. The Dicer-TRBP assay was designed based on reports that the K_m for Dicer-TRBP processing pre-miR-21 is similar to its K_d, both of which being in the low nM range (20-30 nM, Lee et al, 2012). However, K_d calculations for Dicer-TRBP to pre-miR-21 vary significantly likely from varying Dicer purification protocols, Dicer instability *in vitro*, and K_d calculations

using different assays with different limits of detection (e.g. filter binding assays, EMSA). In some of these cases K_ds as low as 100 pM are reported (Lee et al, 2010; Acker et al, 2014). If Dicer does bind pre-miRNAs with a 100 pM K_d, then $K_d << K_m$. This would mean using 35 nM Dicer-TRBP, the K_m, would result in saturating conditions where almost all pre-miRNA would be bound to Dicer-TRBP at equilibrium. Therefore, even if inhibitors do compete with Dicer for pre-miRNA binding and are pre-equilibrated with pre-miR-21 before adding Dicer, Dicer would saturate the pre-miRNA before any detection of cleavage would occur due to catalysis being rate limiting. Only very small changes in activity would be detected relative to the control due to the low population of unbound pre-miR-21. In order to avoid this saturation issue, it may require assaying using pM concentrations of Dicer and buffer conditions to stabilize the enzyme to facilitate the multiple days needed to observe significant amounts of pre-miRNA cleavage.

Another explanation for the differences between K_d and inhibition arises from recent structures of pre-miRNA bound to human Dicer (Lie et al, 2018). Lack of inhibition may result from pre-miR-21 ligands not binding pre-miR-21 at its Dicer binding site. Fox RRM* recognizes the pre-miR-21 apical loop and each of the cyclic peptides bind in the RNA major groove near the cleavage site; whereas as Dicer-TRBP associates with pre-miR-21 through dsRBDs that recognize the pre-miR-21 helix away from the Fox RRM* and cyclic peptide binding sites.

Future directions for developing these assays would involve specificity controls to test for pre-miR-21-specific inhibition rather than global pre-miR-21 inhibition or Dicer inhibition. Specificity could be evaluated by using pre-miRNA sequences with varying secondary structures to develop insight into which secondary structures a peptide binds to; this would aid design efforts for future rounds of pre-miR-21 binding ligand optimization and also better replicate cellular conditions.

Section 4: Materials and Methods

4.1 Peptide Synthesis

Linear peptides were synthesized using solid-phase peptide synthesis (SPPS) on a CEM Liberty Blue peptide synthesizer. Peptides were elongated C-to-N using a pre-loaded 2-chlorotrityl resin and a standard Fmoc protection strategy (Roodbeen et al, 2013). Peptides were then cleaved from the resin using TFE and acetic acid to avoid deprotection of sidechains. The peptides were cyclized overnight using a head-to-tail strategy in DMF using PyBoP/HAOT (3:1) and sym-collidine. After DMF removal, the protected cyclic peptides were treated with a TFA/H2O/TIS mixture (95/2.5/2.5) at room temperature for 5 hours to deprotect sidechains. Cyclized, deprotected peptides were precipitated by cold diethyl ether and purified by reverse phase HPLC (C_{18} column) with gradient from 10% to 50% IPA in H2O with 0.1% TFA. Following purification, cyclic peptides went through a series of lyophilizations to salt exchange TFA for HCl. Final desalting and purification was done using Superdex Peptide 10/300 GL size exclusion peptide column in PBS. Peptide cyclization and identity were confirmed by MSI-MS and quantified by peptide bond count and absorption at 205 nm (Anthis et al, 2013).

4.2 RNA Preparation

RNA was prepared by large scale 10 mL *in vitro* transcriptions using T7 RNA polymerase and synthetic oligonucleotides ordered from IDT. Oligonucleotide templates were initially incubated with an oligonucleotide primer and $MgCl_2$ at 90°C for 5 minutes. The mixture was then added to a transcription cocktail containing NTPs, 40% PEG, transcription buffer (spermidine, Tris-HCl, Triton-X, DTT, RNasin, pyrophosphatase). Reagents were incubated for 30 minutes at 37°C, followed by addition of T7 RNA polymerase (1.5 mg/mL) and incubation at

37°C overnight. Transcription products were isolated by ethanol-sodium acetate precipitation and centrifugation. Pellets were dried, resuspended in TBE and loading buffer, and ran on a denaturing acrylamide gel until single nucleotide separation for gel extraction. Extracted bands were ran through an electro-eluter Schleicher and Schuell Elutrap apparatus. Fractions were quantified by UV absorption at 260 nm then pooled for desalting by illustra NAP-10 columns (GE Healthcare Life Sciences).

Radiolabeled RNA was prepared by first 5' phosphate removal by 1a hour incubation with Calf Intestinal Phosphatase (CIP) at 37°C, isolated by phenol-chloroform extraction, and labeled by 1 hour incubation with T4 Polynucleotide Kinase and gamma-^{32}P-ATP. RNA was then eluted at ~35nM using a miRNAeasy clean-up kit (QIAGEN).

4.3 Protein Preparation

Human Dicer and human TRBP were prepared by Bac-to-Bac Baculovirus Expression System (ThermoFisher). N-terminally His_6-tagged Dicer and TRBP were isolated separately from Sf9 cells infected with baculovirus containing Dicer and TRBP cDNA. Lysates were run on a Ni-NTA column, eluted, and treated with tobacco etch virus (TEV) protease to cleave the His6-tag. Cleaved proteins were then purified through a Superdex-200 column. Fox RRM* and Fox WT were expressed by using standard *E. Coli* expression systems and purified using similar chromatographic steps as for human Dicer and TRBP.

4.4 Turbo Dicer™ Assays

Genlantis Turbo Dicer™ and reaction buffers were purchased as a kit, and methods were adapted from the Genlantis siRNA purification protocol (Myers et al, 2003). Snap-cooled pre-

miR-21 at 500-1000 pM (below the reported Dicer-pre-miR-21 K_m values (Lee et al, 2012)), and peptides at titrated concentrations were incubated at room temperature for 30 minutes prior to adding to a reaction buffer (1 mM ATP, 2mM $MgCl_2$; $MgCl_2$ levels were optimized to adjust for the desired Dicer activity). 30 pM Turbo Dicer™ was added following the binding reaction. Samples were then incubated at 37°C for 90 minutes. Dicer was then inactivated by adding 2X TBE urea sample buffer (45mM Tris, 45mM Boric Acid, 6% Ficol, 3.5M Urea, 0.005% Bromophenol Blue, 0.025% Xylene Cyanol, 1mM EDTA). Samples were heated for 5 minutes at 90°C before loading to 8-12.5% denaturing 8 M urea polyacrylamide gel and electrophoresed at 200-390V for 3 hours at room temperature. The gels were exposed overnight on a phosphor image screen (Amersham Biosciences). Screens were scanned on a phosphor scanner (Molecular Dynamics), and autoradiograms were analyzed with ImageQuant 5.0 software. Band intensities were quantified using ImageJ and graphs were generated using Microsoft Excel.

4.5 Human Dicer-TRBP Assays

5'-end-labeled RNAs were processed with Dicer-TRBP at previously reported K_m concentrations (35nM, Lee et al, 2012) in optimized Dicer buffer (20mM Tris HCl at pH 7.5, 3mM MgCl2, 100mM NaCl, 1mM DTT). RNA and peptides were incubated at room temperature for 30 minutes prior to adding Dicer or Dicer-TRBP to make 10uL final reaction volumes. Samples were then incubated at 37°C for 90 minutes. Dicer was inactivated by adding 2X TBE urea sample buffer. Samples were electrophoresed and analyzed similar to the Turbo Dicer™ assays.

Section 5: References

1. Acker, M. G., & Auld, D. S. (2014). Considerations for the design and reporting of enzyme assays in high-throughput screening applications. *Perspect Sci (Neth)*, *1*(1-6), 56-73.
2. Anthis, N. J., & Clore, G. M. (2013). Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci*, *22*(6), 851-858.
3. Asangani, I. A., Rasheed, S. A., Nikolova, D. A., Leupold, J. H., Colburn, N. H., Post, S., & Allgayer, H. (2008). MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene*, *27*(15), 2128-2136.
4. Athanassiou, Z., Patora, K., Dias, R. L., Moehle, K., Robinson, J. A., & Varani, G. (2007). Structure-Guided Peptidomimetic Design Leads to Nanomolar β-Hairpin Inhibitors of the Tat− TAR Interaction of Bovine Immunodeficiency Virus. *Biochemistry*, *46*(3), 741-751.
5. Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., ... & Sharon, E. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet*, *37*(7), 766-770.
6. Bose, D., Nahar, S., Rai, M. K., Ray, A., Chakraborty, K., & Maiti, S. (2015). Selective inhibition of miR-21 by phage display screened peptide. *Nucleic Acids Res*, *43*(8), 4342-4352.
7. Bose, D., Jayaraj, G., Suryawanshi, H., Agarwala, P., Pore, S. K., Banerjee, R., & Maiti, S. (2012). The tuberculosis drug streptomycin as a potential cancer therapeutic: inhibition

of miR-21 function by directly targeting its precursor. *Angew Chem Int Ed Engl*, *51*(4), 1019-1023.

8. Sittampalam, G. S., Coussens, N. P., & Brimacombe, K. (2012). Assay Guidance Manual [Internet]. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004.
9. Carthew, R. W., & Sontheimer, E. J. (2009). Origins and mechanisms of miRNAs and siRNAs. *Cell*, *136*(4), 642-655.
10. Chakravarthy, S., Sternberg, S. H., Kellenberger, C. A., & Doudna, J. A. (2010). Substrate-specific kinetics of Dicer-catalyzed RNA processing. *J Mol Biol*, *404*(3), 392-402.
11. Chan, J. A., Krichevsky, A. M., & Kosik, K. S. (2005). MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Can Res*, *65*(14), 6029-6033.
12. Chen, Y., & Varani, G. (2013). Engineering RNA-binding proteins for biology. *FEBS J*, *280*(16), 3734-3754.
13. Chen, Y., Yang, F., Zubovic, L., Pavelitz, T., Yang, W., Godin, K., ... & Varani, G. (2016). Targeted inhibition of oncogenic miR-21 maturation with designed RNA-binding proteins. *Nat Chem Biol*, *12*(9), 717-723.
14. Cheng, C. J., Bahal, R., Babar, I. A., Pincus, Z., Barrera, F., Liu, C., ... & Saltzman, W. M. (2015). MicroRNA silencing for cancer therapy targeted to the tumour microenvironment. *Nature*, *518*(7537), 107-110.
15. Chirayil, S., Chirayil, R., & Luebke, K. J. (2009). Discovering ligands for a microRNA precursor with peptoid microarrays. *Nucleic Acids Res*, *37*(16), 5486-5497.

16. Davidson, A., Leeper, T. C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J. A., & Varani, G. (2009). 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*, *106*(29), 11931-11936.
17. Davidson, A., Patora-Komisarska, K., Robinson, J. A., & Varani, G. (2010). Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of Tat protein. *Nucleic Acids Res*, *39*(1), 248-256.
18. Davidson, A., Leeper, T. C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J. A., & Varani, G. (2009). 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*, *106*(29), 11931-11936.
19. Esau, C. C. (2008). Inhibition of microRNA with antisense oligonucleotides. *Methods*, *44*(1), 55-60.
20. Esteller, M. (2011). Non-coding RNAs in human disease. *Nat Rev Genet*, *12*(12), 861.
21. Foulkes, W. D., Priest, J. R., & Duchaine, T. F. (2014). DICER1: mutations, microRNAs and mechanisms. *Nat Rev Can*, *14*(10), 662-672.
22. Gibbs, A. C., Kondejewski, L. H., Gronwald, W., Nip, A. M., Hodges, R. S., Sykes, B. D., & Wishart, D. S. (1998). Unusual β-sheet periodicity in small cyclic peptides. *Nat Struct Biol*, *5*(4), 284-288.
23. Gumireddy, K., Young, D. D., Xiong, X., Hogenesch, J. B., Huang, Q., & Deiters, A. (2008). Small-molecule inhibitors of microrna miR-21 function. *Angew Chem Int Ed Engl*, *47*(39), 7482-7484.

24. Macrae, I. J., Li, F., Zhou, K., Cande, W. Z., & Doudna, J. A. (2006, January). Structure of Dicer and mechanistic implications for RNAi. In *Cold Spring Harb Symp Quant Biol* (Vol. 71, pp. 73-80). Cold Spring Harbor Laboratory Press.

25. Howe, J. A., Wang, H., Fischmann, T. O., Balibar, C. J., Xiao, L., Galgoci, A. M., ... & Murgolo, N. (2015). Selective small-molecule inhibition of an RNA structural element. *Nature*, *526*(7575), 672-677.

26. Hsu, P. W., Huang, H. D., Hsu, S. D., Lin, L. Z., Tsou, A. P., Tseng, C. P., ... & Hofacker, I. L. (2006). miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic Acids Res*, *34*(suppl_1), D135-D139.

27. Kost, T. A., Condreay, J. P., & Jarvis, D. L. (2005). Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol*, *23*(5), 567-575.

28. Kidwell, M. A., Chan, J. M., & Doudna, J. A. (2014). Evolutionarily conserved roles of the dicer helicase domain in regulating RNA interference processing. *J Biol Chem*, *289*(41), 28352-28362.

29. Kjaer-Frifeldt, S., Hansen, T. F., Nielsen, B. S., Joergensen, S., Lindebjerg, J., Soerensen, F. B., ... & Jakobsen, A. (2012). The prognostic importance of miR-21 in stage II colon cancer: a population-based study. *Br J Cancer*, *107*(7), 1169-1174.

30. Krichevsky, A. M., & Gabriely, G. (2009). miR-21: a small multi-faceted RNA. *J Cell Mol Med*, *13*(1), 39-53.

31. Lalonde, M. S., Lobritz, M. A., Ratcliff, A., Chamanian, M., Athanassiou, Z., Tyagi, M., ... & Arts, E. J. (2011). Inhibition of both HIV-1 reverse transcription and gene expression by a cyclic peptide that binds the Tat-transactivating response element (TAR) RNA. *PLoS Pathog*, *7*(5), e1002038.

32. Lee, H. Y., & Doudna, J. A. (2012). TRBP alters human precursor microRNA processing in vitro. *RNA*, *18*(11), 2012-2019.

33. Lee, H. Y., Zhou, K., Smith, A. M., Noland, C. L., & Doudna, J. A. (2013). Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. *Nucleic Acids Res*, *41*(13), 6568-6576.

34. Ling, H., Fabbri, M., & Calin, G. A. (2013). MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat Rev Drug Discov*, *12*(11), 847-865.

35. Liu, Z., Wang, J., Cheng, H., Ke, X., Sun, L., Zhang, Q. C., & Wang, H. W. (2018). Cryo-EM structure of human dicer and its complexes with a pre-miRNA substrate. *Cell*, *173*(5), 1191-1203.

36. Löffler, D., Brocke-Heidrich, K., Pfeifer, G., Stocsits, C., Hackermüller, J., Kretzschmar, A. K., ... & Cvijic, H. (2007). Interleukin-6–dependent survival of multiple myeloma cells involves the Stat3-mediated induction of microRNA-21 through a highly conserved enhancer. *Blood*, *110*(4), 1330-1333.

37. Ma, X., Kumar, M., Choudhury, S. N., Buscaglia, L. E. B., Barker, J. R., Kanakamedala, K., ... & Li, Y. (2011). Loss of the miR-21 allele elevates the expression of its target genes and reduces tumorigenesis. *Proc Natl Acad Sci U S A*, *108*(25), 10144-10149.

38. MacRae, I. J., Ma, E., Zhou, M., Robinson, C. V., & Doudna, J. A. (2008). In vitro reconstitution of the human RISC-loading complex. *Proc Natl Acad Sci U S A*, *105*(2), 512-517.

39. Medina, P. P., Nolde, M., & Slack, F. J. (2010). OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. *Nature*, *467*(7311), 86-90.

40. Meng, F., Henson, R., Wehbe–Janek, H., Ghoshal, K., Jacob, S. T., & Patel, T. (2007). MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology*, *133*(2), 647-658.

41. Myers, J. W., Jones, J. T., Meyer, T., & Ferrell Jr, J. E. (2003). Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. *Nat Biotechnol*, *21*(3), 324-328.

42. Naro, Y., Thomas, M., Stephens, M. D., Connelly, C. M., & Deiters, A. (2015). Aryl amide small-molecule inhibitors of microRNA miR-21 function. *Bioorg Med Chem Lett*, *25*(21), 4793-4796.

43. Nguyen, T. A., Jo, M. H., Choi, Y. G., Park, J., Kwon, S. C., Hohng, S., ... & Woo, J. S. (2015). Functional anatomy of the human microprocessor. *Cell*, *161*(6), 1374-1387.

44. Nicholson, A. W. (2014). Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdisciplinary Reviews: RNA*, *5*(1), 31-48.

45. Nowotny, M., Gaidamakov, S. A., Crouch, R. J., & Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell*, *121*(7), 1005-1016.

46. Papagiannakopoulos, T., Shapiro, A., & Kosik, K. S. (2008). MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells. *Cancer Res*, *68*(19), 8164-8172.

47. Park, J. E., Heo, I., Tian, Y., Simanshu, D. K., Chang, H., Jee, D., ... & Kim, V. N. (2011). Dicer recognizes the 5′ end of RNA for efficient and accurate processing. *Nature*, *475*(7355), 201-205.

48. Paroo, Z., Liu, Q., & Wang, X. (2007). Biochemical mechanisms of the RNA-induced silencing complex. *Cancer Res*, *17*(3), 187-194.

49. Ribas, J., & Lupold, S. E. (2010). The transcriptional regulation of miR-21, its multiple transcripts and their implication in prostate cancer. *Cell Cycle*, *9*(5), 923-929.

50. Robinson, J. A. (2008). β-Hairpin peptidomimetics: design, structures and biological activities. *Acc Chem Res*, *41*(10), 1278-1288.

51. Chen, Y. (2019). Recent Advances in Methylation: A Guide for Selecting Methylation Reagents. *Chemistry–A European Journal*, *25*(14), 3405-3439.

52. Sawh, A. N., & Duchaine, T. F. (2012). Turning Dicer on its head. *Nat Struct Mol Biol*, *19*(4), 365.

53. Sayed, D., Rane, S., Lypowy, J., He, M., Chen, I. Y., Vashistha, H., ... & Abdellatif, M. (2008). MicroRNA-21 targets Sprouty2 and promotes cellular outgrowths. *Mol Biol Cell*, *19*(8), 3272-3282.

54. Schetter, A. J., Leung, S. Y., Sohn, J. J., Zanetti, K. A., Bowman, E. D., Yanaihara, N., ... & Liu, C. G. (2008). MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA*, *299*(4), 425-436.

55. Shortridge, M. D., Walker, M. J., Pavelitz, T., Chen, Y., Yang, W., & Varani, G. (2017). A macrocyclic peptide ligand binds the oncogenic microRNA-21 precursor and suppresses dicer processing. *ACS Chem Biol*, *12*(6), 1611-1620.

56. Strelow, J., Dewe, W., Iversen, P. W., Brooks, H. B., Radding, J. A., McGee, J., & Weidner, J. (2012). Mechanism of action assays for enzymes. In *Assay Guidance Manual [Internet]*. Eli Lilly & Company and the National Center for Advancing Translational Sciences.

57. Taylor, D. W., Ma, E., Shigematsu, H., Cianfrocco, M. A., Noland, C. L., Nagayama, K., ... & Wang, H. W. (2013). Substrate-specific structural rearrangements of human Dicer. *Nat Struct Mol Biol*, *20*(6), 662-670.

58. Van Rooij, E., & Kauppinen, S. (2014). Development of microRNA therapeutics is coming of age. *EMBO Mol Med*, *6*(7), 851-864.

59. Wang, B., Li, S., Qi, H. H., Chowdhury, D., Shi, Y., & Novina, C. D. (2009). Distinct passenger strand and mRNA cleavage activities of human Argonaute proteins. *Nat Struct Mol Biol*, *16*(12), 1259-1266.

60. Wickramasinghe, N. S., Manavalan, T. T., Dougherty, S. M., Riggs, K. A., Li, Y., & Klinge, C. M. (2009). Estradiol downregulates miR-21 expression and increases miR-21 target gene expression in MCF-7 breast cancer cells. *Nucleic Acids Res*, *37*(8), 2584-2595.

61. Wilson, R. C., & Doudna, J. A. (2013). Molecular mechanisms of RNA interference. *Annu Rev Biophys*, *42*, 217-239.

62. Wilson, R. C., Tambe, A., Kidwell, M. A., Noland, C. L., Schneider, C. P., & Doudna, J. A. (2015). Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis. *Mol Cell*, *57*(3), 397-407.

63. Xavier, K. A., Eder, P. S., & Giordano, T. (2000). RNA as a drug target: methods for biophysical characterization and screening. *Trends Biotechnol*, *18*(8), 349-356.

64. Yan, L. X., Huang, X. F., Shao, Q., Huang, M. Y., Deng, L., Wu, Q. L., ... & Shao, J. Y. (2008). MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA*, *14*(11), 2348-2360.

65. Zhu, S., Si, M. L., Wu, H., & Mo, Y. Y. (2007). MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). *J Biol Chem*, *282*(19), 14328-14336.

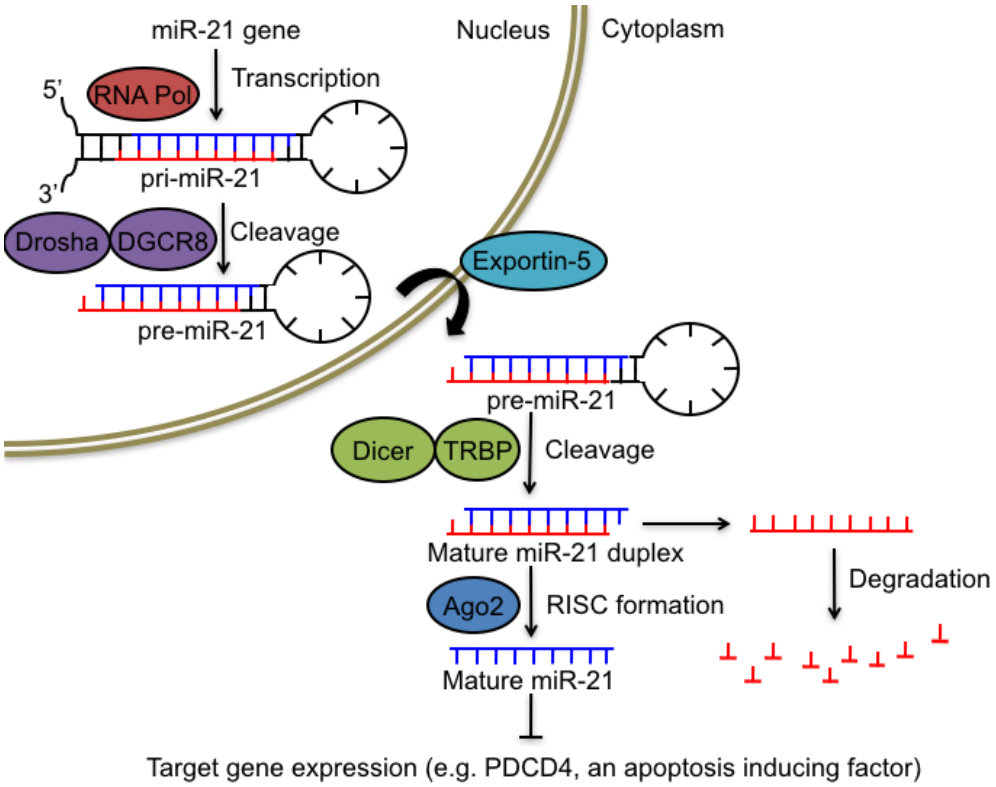


Figure 3.1. Canonical biogenesis pathway of microRNA-21. MicroRNA-21 is processed bin two cleavage steps to produce mature miR-21 which associated with RISC to knock down target gene expression (Wilson et al, 2013; Carthew et al, 2009; Nguyen et al, 2015; MacRae et al, 2008; Paroo et al, 2007; Wang et al 2009).

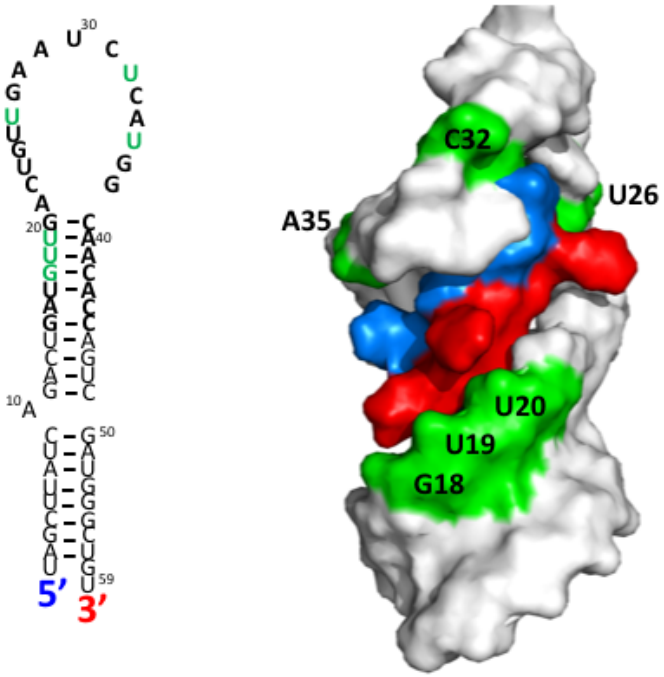


Figure 3.2. Pre-miR-21 structural landscape. (left) The secondary structure of the 59 nucleotide pre-miR-21 is a standard stem-loop with unpaired bases and internal loops. **(right)** A molecular model was generated from NMR NOESY experiments with pre-miR-21 bound to a cyclic peptide. In green are significant peptide-nucleotide interactions that drive binding. (Nicholson et al, 2013; Lee et al, 2013; Wilson et al, 2015; Kidwell et al, 2014; Foulkes et al, 2014).

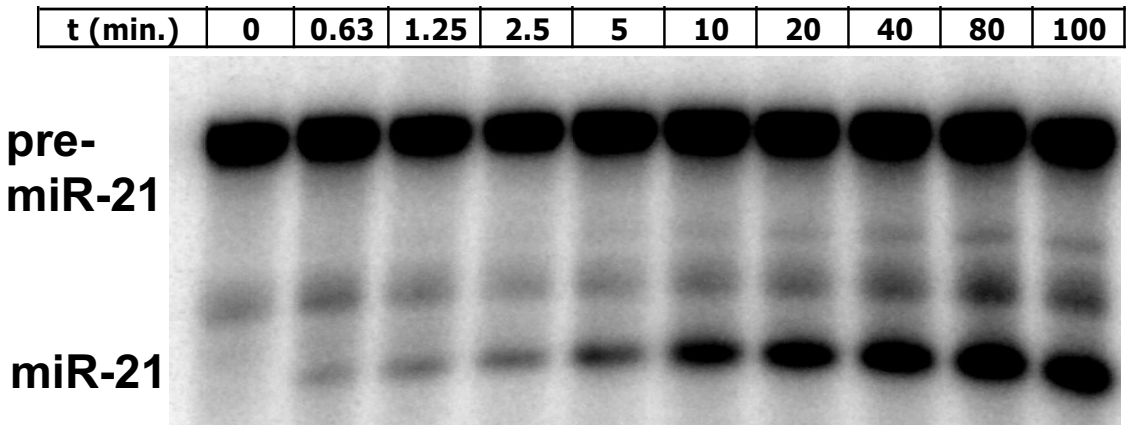


Figure 3.3. Processing activity of human Dicer –TRBP monitored over time. To confirm normal activity of the newly purified Dicer complex, a one-pot reaction was carried with equimolar RNA and enzyme in a reaction buffer identical to previous publications (Lee et al, 2012; Acker et al, 2014). Aliquots were taken out at the times indicated.

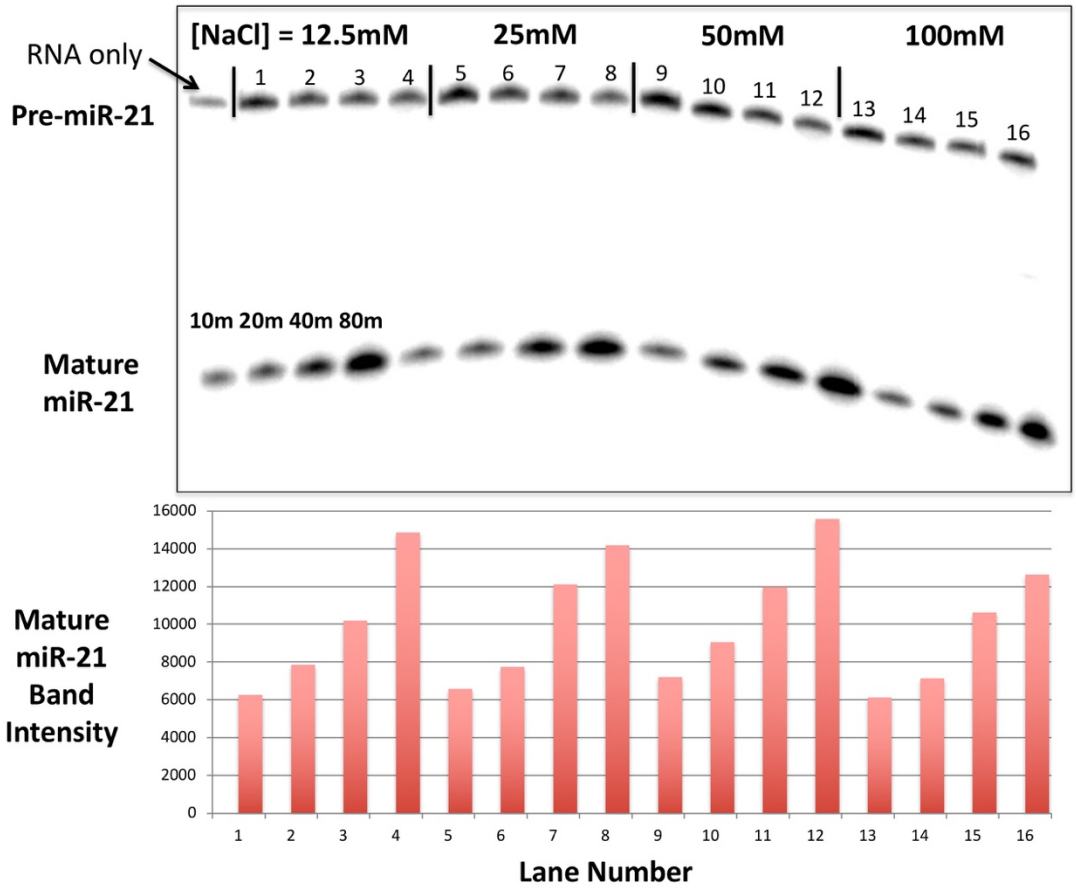


Figure 3.4. Human Dicer-TRBP activity assayed at different NaCl concentrations. Assays were carried out using previously established reaction conditions for Dicer assays, varying only the NaCl concentration. NaCl concentration was 12.5mM (1-4), 25mM (5-8), 50mM (9-12), 100mM (13-16). Time points from a single reaction pot were collected after 10m, 20m, 40m, and 80m.

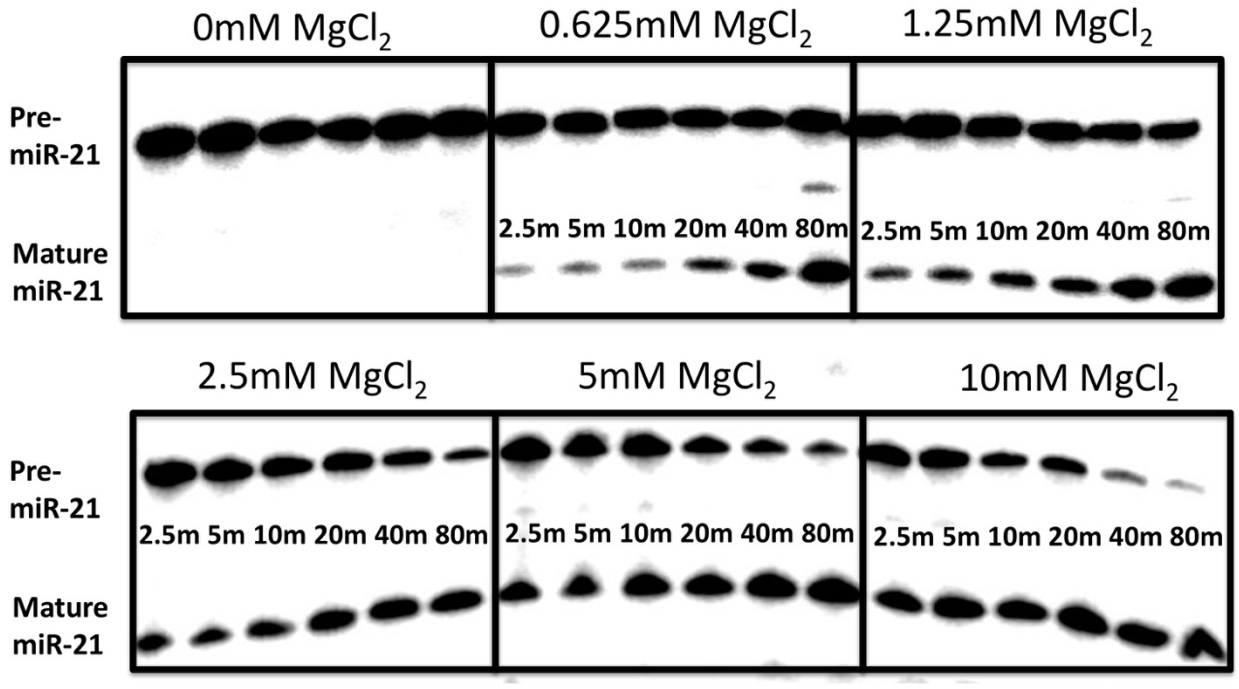


Figure 3.5. Human Dicer-TRBP activity assayed at different $MgCl_2$ concentrations. Assays were carried out using previously established reaction conditions for Dicer assays only varying $MgCl_2$ concentrations. Time points from a single reaction vessel were collected after 2.5m, 5m, 10m, 20m, 40m, 80m. Activity was measured at 0mM $MgCl_2$, 0.625mM $MgCl_2$, 1.25mM $MgCl_2$, 2.5mM $MgCl_2$, 5mM $MgCl_2$, 10mM $MgCl_2$.

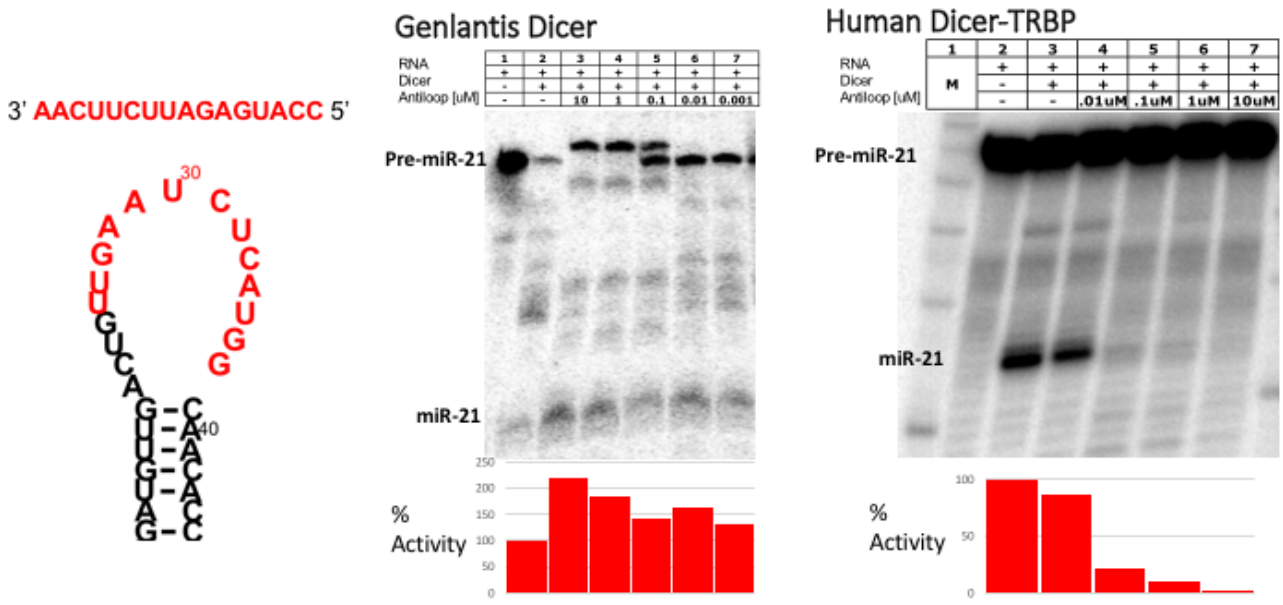


Figure 3.6. **Inhibition of pre-miR-21 processing by anti-loop oligonucleotides by different enzymatic preparations.** A positive control was carried out using a DNA oligonucleotide designed to bind the pre-miR-21 terminal loop; this was assayed using two Dicer preparations.

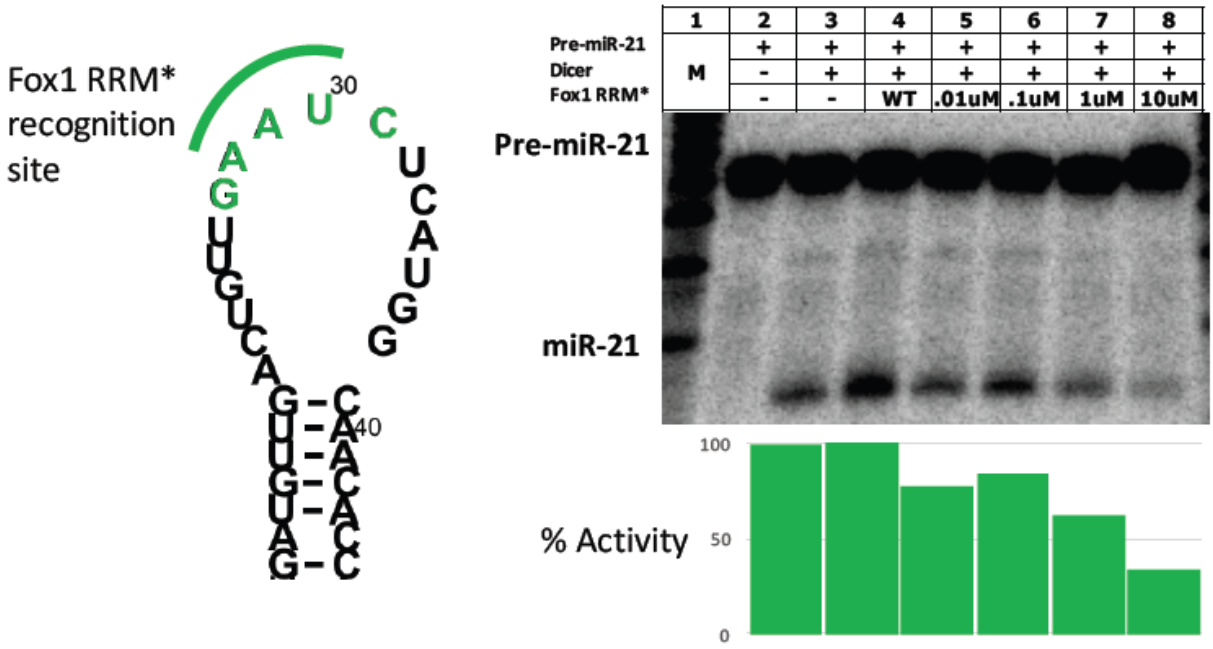
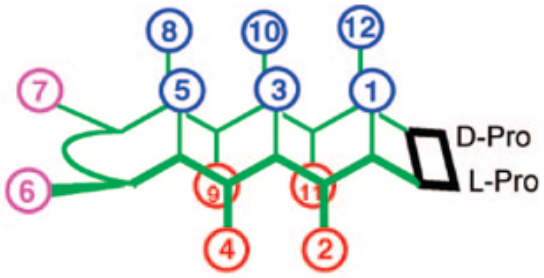


Figure 3.7. Wildtype human Dicer-TRBP activity assayed in the presence of Fox RRM*. Different concentrations of Fox were added to separate Dicer reactions to determine whether high affinity protein binding to pre-miR-21 inhibits processing.

Peptide	HIV TAR	miR21	
L50	1	200	TAR-Tat inhibitors
L22	30	>4000	
JB72	dnb	200	Hit from screen
mds001	dnb	75	Optimized off of JB72 sequence
mds002		50	

EMSA generated Kd values (nM)



12-mer loop

	Position												
		1	2	3	4	5	6	7	8	9	10	11	12	
L22	LP	R	V	R	T	R	K	G	R	R	I	R	I	DP
L50	LP	R	V	R	T	R	G	K	R	R	I	R	R	DP

Figure 3.8. RNA binding activity of cyclic peptides targeting pre-miR-21. Cyclic peptides are derived from a library of previously reported HIV TAR binding peptides (Robinson et al, 2008; Davidson et al, 2009). (dnb = does not bind; all K_d values are reported in nM concentrations).

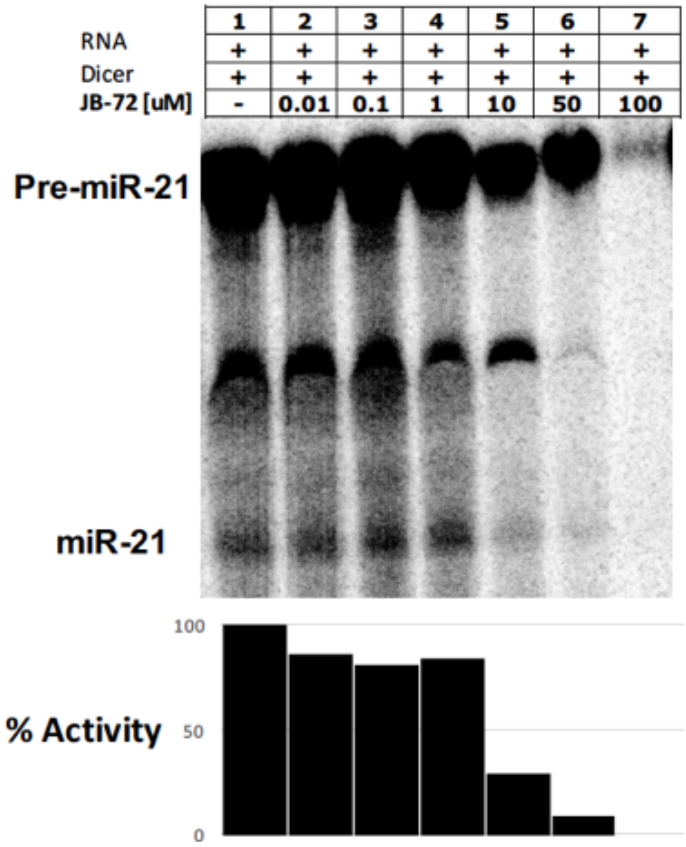


Figure 3.9. Cyclic peptides with high affinity for pre-miR-21 inhibit pre-miR-21 processing in vitro. Human Dicer-TRBP assayed in the presence of peptide JB72. Reactions conditions were set up as previously described for data recorded in the absence of inhibitors.

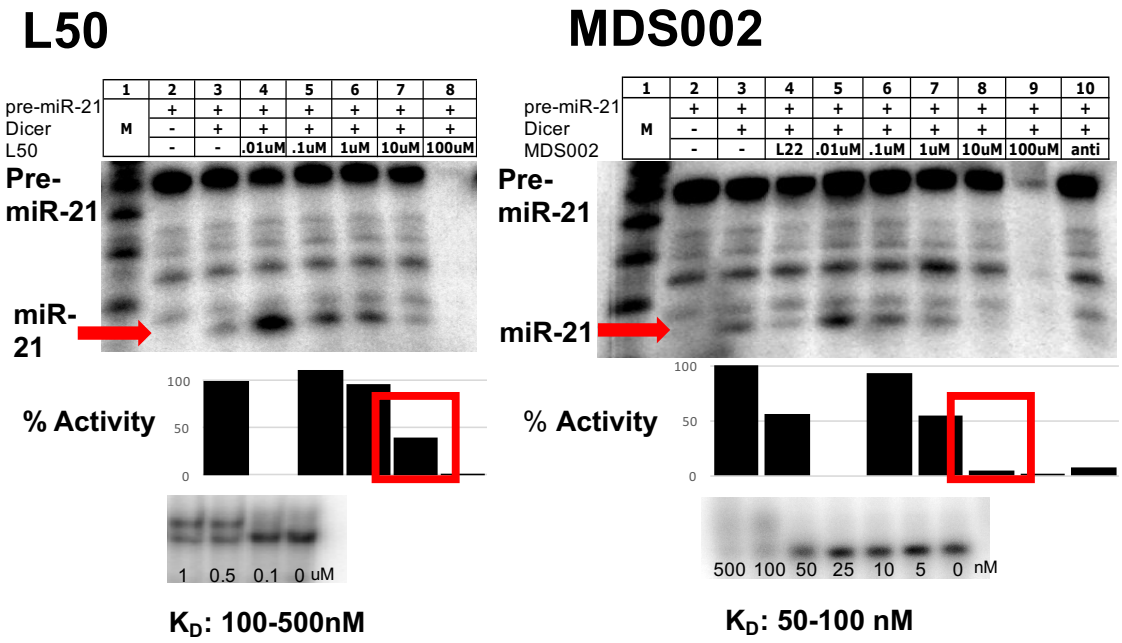


Figure 3.10. In vitro Dicer activity inhibition for cyclic peptides inhibitors of miR-21 processing. **(left)** L50, the first peptide to show sub-μM affinity for pre-miR-21; **(right)** MDS002, a more potent pre-miR-21 binding peptide (Kd = 50-100 nM).

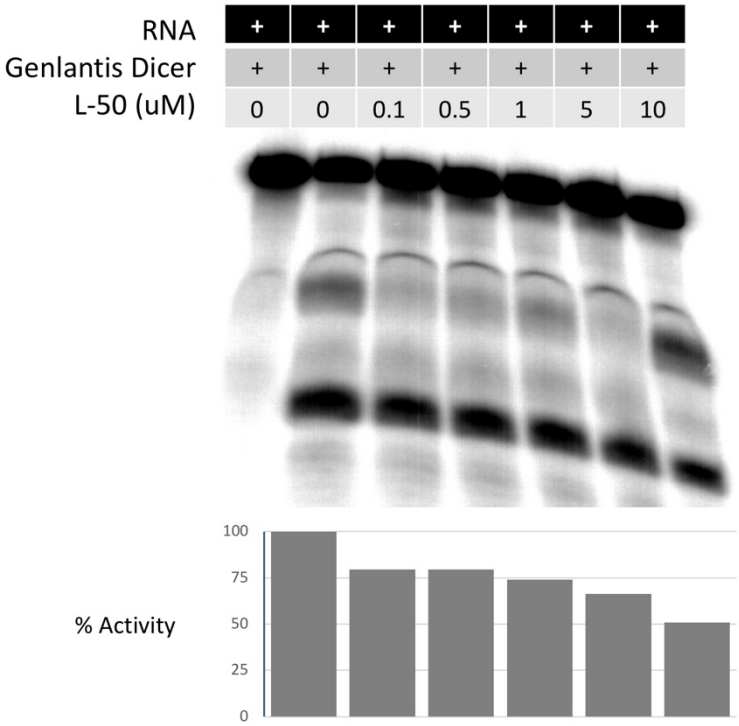


Figure 3.11. Dicer inhibition assay for L50 using Genlantis Turbo Dicer™ enzyme. Dicer reactions were set-up accordingly to manufacturer's conditions and run for a total of 90 minutes before each reaction was halted on ice, then denatured in urea sample buffer.

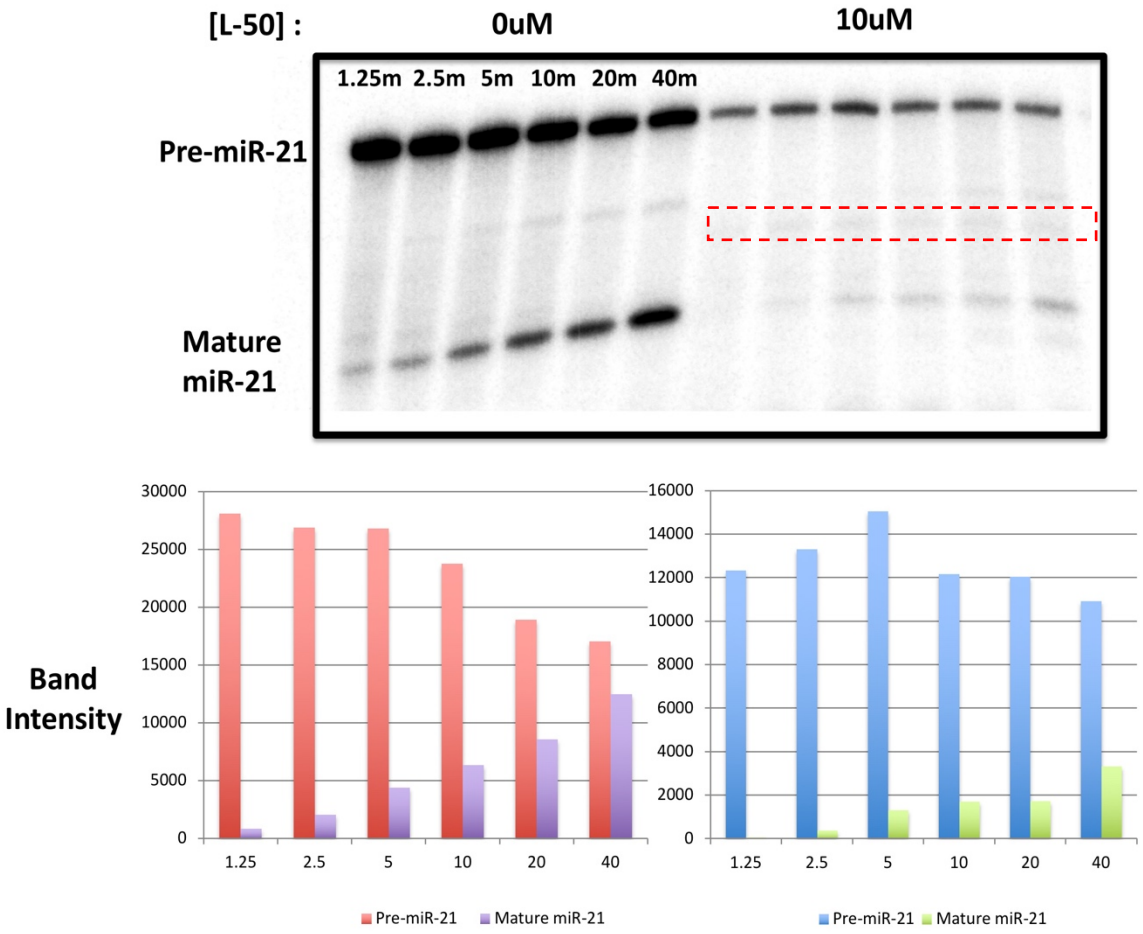


Figure 3.12. Single vessel reaction time course of Dicer activity using in-house purified human Dicer-TRBP. Reactions were set up similarly to previous activity and inhibition assays, with the exception that volume was scaled up to allow for individual time points to be collected and immediately stopped to represent the maturation percentage for each time point. Time points were collected at 1.25, 2.5, 5, 10, 20 and 40 min. Reactions were done in the absence of L50, and in the presence of 10μM L50. Boxed in red is the new intermediate observed only when Dicer processes pre-miR-21 in the presence of L50.

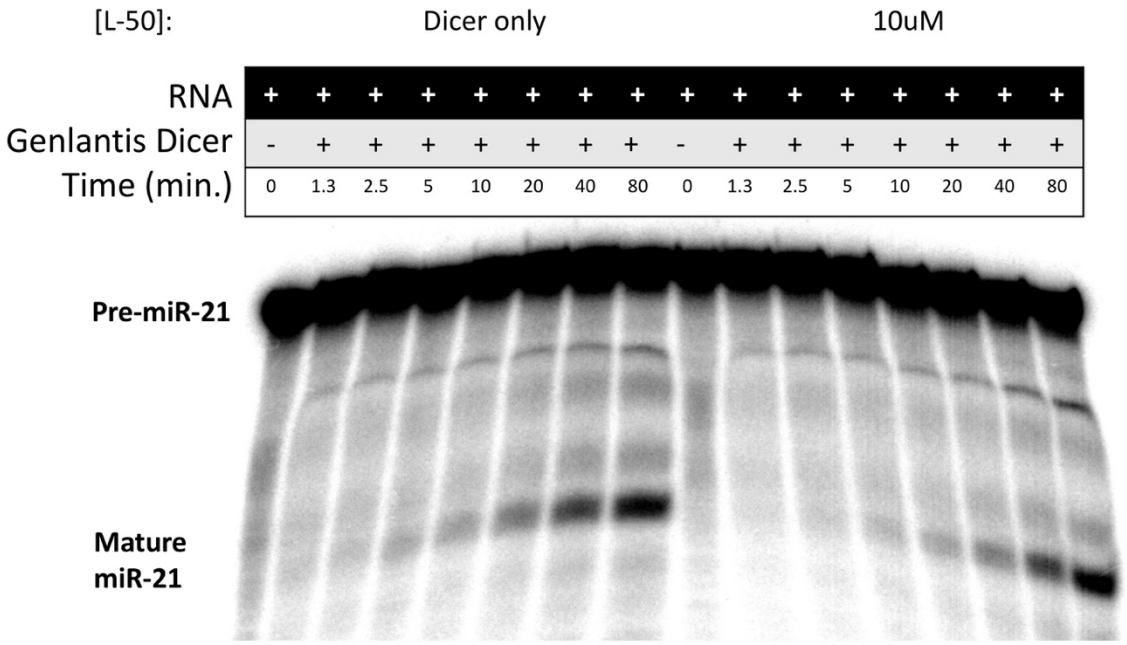


Figure 3.13. Single vessel reaction time course Dicer activity assays using Dicer Turbo™ purified human Dicer-TRBP. Reactions were set-up similarity to time course assays using human Dicer-TRBP. Time points were collected for 1.25m, 2.5m, 5m, 10m, 20m, 40m. Reactions were done in the presence of no L50, and 10μM L50. Difference is activity between reactions omitted as 10μM L50 appeared to not reduce Turbo Dicer™ activity significantly.

Chapter 4 – Structure of the RNA Specialized Translation Initiation Element that Recruits eIF3 to the 5'-UTR of c-Jun*

Section 1: Summary

Specialized translation initiation is a novel form of regulation of protein synthesis whereby RNA structures within the 5'-UTR regulate translation rates of specific mRNAs. Like internal ribosomal entry sites (IRES), specialized translational initiation requires the recruitment of eukaryotic initiation factor 3 (eIF3), but specialized translation initiation requires cap recognition by eIF3d, a new 5'-m^7GTP recognizing protein. How these RNA structures mediate eIF3 recruitment to affect translation of specific mRNAs remains unclear. Here we report the NMR structure of a stem-loop within the c-JUN 5' UTR recognized by eIF3 and essential for specialized translation initiation of this well-known oncogene. The structure exhibits similarity to eIF3 recognizing motifs found in HCV-like IRES, suggesting mechanistic similarities. This work establishes the RNA structural features involved in c-JUN specialized translation initiation and provides a basis to search for small molecules inhibitors of aberrant expression of the proto-oncogenic c-JUN.

* This chapter is adapted from an article to be published in *Journal of Molecular Biology* (2019)

Section 2: Introduction

Translational initiation is highly regulated and its deregulation is implicated with diseased cellular states (Silvera et al, 2010, Stumpf et al, 2011). Initiation of protein synthesis involves the orchestrated recruitment of the ribosome to the initiation codon by a multifaceted network of eukaryotic initiation factors (eIFs) that are present in limiting amounts to ensure controlled levels of protein synthesis (Bhat et al, 2015; Jackson et al, 2010; Andaya et al, 2014). Significant up- or down-regulation of eIFs has been linked to the etiology of cancer (Ruggero et al, 2013). In short, canonical translation initiation involves recognition of the mRNA 5'-m^7GTP cap by eIF4F, a trimeric complex with cap binding (eIF4E), mRNA unwinding (eIF4A), and ribosomal loading (eIF4G) activities. The mRNA-bound eIF4F complex is then loaded into the 43S preinitiation complex comprised of the small ribosomal subunit (40S) and several other auxiliary factors (eIF1, eIF2-tRNA, eIF3, and eIF5).

Translation initiation is regulated by secondary and tertiary structures within mRNAs that modulate eIF activity and can increase or decrease translation rates, such as encumbering eIF4A helicase unwinding by stable secondary structures or recruitment of activating auxiliary proteins (Truitt et al, 2016). Internal ribosome entry sites (IRES) are well-known examples of complex RNA structures that facilitate direct entry of the translation machinery, by-passing cap recognition (Terenin et al, 2017; Truitt et al, 2016). Hepatitis C virus (HCV)-like IRES are recognized by the 13-units ~800kDa eIF3 complex (Terenin et al, 2017; Honda et al, 1999; Fraser et al, 2007; Khan et al, 2014), which serves as a scaffold for 43S preinitiation complex formation and assists in mRNA loading during canonical translational initiation. In IRES-mediated translation initiation, eIF3 recognizes the IRES structure directly (Querol-Audi et al, 2013; Valášek et al, 2017). Although a complete physical understanding of eIF3 recognition of

HCV-like IRES' is still lacking, regions of the HCV-like IRESs required for 40S and eIF3 recruitment are known (Kieft et al, 2001; Buratti et al, 1998).

IRESs are rare in eukaryotic transcripts (Holcik et al, 2000; Johannes et al, 1999), but relatively more prevalent in growth factors and house-keeping genes. They facilitate aberrant protein synthesis during cellular stress or cancer, when the eIF4F complex is inhibited (Holcik et al, 2000; Stephane et al, 2001; Hellen et al 2001, Aurelio et al, 2005; Hsieh et al, 2012). mRNAs lacking IRESs can be translated when eIF4F is inhibited through a third form of translation initiation, that requires both recognition of specific mRNA structures by eIF3 and mRNA-cap recognition by eIF3d (Lee et al, 2015; Lee et al, 2016). During specialized translational initiation, the protein DAPI5 binds to eIF3d and might replace the ribosomal loading function of eIF4G (de la Parra et al, 2018). How RNA elements recruit eIF3, which features of their sequence and/or structure are recognized, and whether they somehow resemble IRESs with which they share the ability to interact with eIF3, remain unexplored. These are critical questions to understand specialized translational initiation, which is addressed herein.

One of the most studied mRNAs that undergoes eIF3d-directed specialized translation initiation is the mitotic transcriptional regulator and proto-oncogene c-JUN. The c-JUN 5'-UTR was proposed to harbor a cellular IRES (Johnson et al, 1996; Wisdom et al, 1999; Blau et al, 2012), but specialized translation initiation was discovered instead during a whole cell photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) study, where a region of its mRNA cross-linked to eIF3 and its deletion was shown to be critical for maintaining cap-dependent c-JUN translation rates (Lee, 2015). The cross-linked region mapped to a 56-mer stem-loop (SL1) located at position +93 to +143 relative to the transcription start site, as designated based on cap analysis gene express (CAGE) data deposited in FANTOM

(Lee, 2015) (**Fig. 4.1A**). Both SL1 and a neighboring stem-loop (SL2), located at +151 to +169, exhibit sequence conservation in mammals, and perhaps matching secondary structure based on predictions from the Mfold web server (Zuker et al, 2003) (**Supplementary Fig. S4.1**).

We used nuclear magnetic resonance spectroscopy (NMR) and small angle X-ray scattering (SAXS) to understand structurally what drives eIF3 recognition of SL1. We show that SL1 has a flexible internal loop structurally analogous to the HCV IRES subdomain IIIb internal loop, which has previously been shown to be key for eIF3 binding (Collier et al, 2002) (**Fig. 4.1A, B**). The complete SL1-SL2 element folds into a 3-way stem-loop junction (**Fig. 4.1A)** which is similar in global shape to the IIIabc subdomain found in the HCV IRES (Perard et al, 2013; Quade et al, 2015). Modeling the SL1 and SL2 construct into the eIF3a and eIF3c pockets known to bind the HCV IRES (Sun et al, 2013; Hashem et al, 2013; Erzberger et al, 2014) provides a good fit, suggesting that c-JUN specialized translation initiation and HCV-like IRESs recognize eIF3 in structurally similar ways.

Section 3: Results

3.1 RNA Construct Design and Verification of the Secondary Structure

We determined the NMR structure of a construct shown to cross-link with eIF3 and to be essential for specialized translation initiation of c-JUN (**Fig. 4.1A**). NMR assignments and structure determination were facilitated by preparing a second construct where the highly dynamic, flexible apical loop was replaced with a UUCG tetraloop (TL1) (**Fig. 4.1A**), which provided greatly improved spectral quality (see below). Shorter constructs were generated as well, for assignment confirmation of overlapped regions, corresponding to the bottom (TL1-BOT) and top (TL1-TOP) sections of the structure (**Supplementary Fig. S4.2 & S4.3**).

Before undertaking the NMR analysis, we confirmed that the SL1 sequence retains the proposed secondary structure when isolated from the rest of the 5'-UTR. N-methylisatoic anhydride (NMIA) based SHAPE (selective 2′ hydroxyl acylation analyzed by primer extension) experiments for the SL1 provided chemical reactivities that, when incorporated into the secondary structure prediction program RNAStructure (Mathews et al, 2004), matched the secondary structure of the mRNA obtained from the analysis of the complete 5'-UTR of c-Jun (**Fig. 4.2A**). High reactivities within the apical loop and near the A99 bulge were observed, but the proposed lower and upper internal loops were largely unreactive.

We followed up these results by acquiring small angle x-ray scattering (SAXS) data to generate a low-resolution model of SL1 (**Fig. 4.2B**). The SAXS envelope is ~75Å long and exhibits lower resolution at the apical loop and in the center of the model; this second result was unexpected given the moderate amount of SHAPE reactivity in both the lower (U103, C104, C105, U137, C138, U139) and upper internal loops (C109, C132, C133). We tentatively positioned these internal loops into the low-resolution center of the model (which is 25Å wide),

based on the overall size of the RNA and the attribution of the apical loop to the less resolved region at the top. The model reveals a bend of ~20-30 degrees from the helical axis that coincide with this middle region, suggestive of a kink or bend in the helical structure induced by the internal loop residues, as was also observed for the HCV IRES IIIb internal loop (Kieft et al, 2001; Collier et al, 2002).

NMR was then used to establish the secondary structure unambiguously. Almost all NH protons, the appearance of which is indicative of base-pairing, were assigned (**Fig. 4.3A**, **Supplementary Fig. S4.4**) for both constructs. Imino chemical shifts were assigned for each base-pair, with the exception of U127-H3, predicted to base-pair with A115, located at the stem-apical loop interface and therefore broadened by exchange with solvent. However, a low intensity cross-peak was observed between U127-H3 and A114-H22 suggesting formation of the base-pair.

The chemical shifts and peaks intensities observed for the NH residues of the SL1 RNA map very well to the corresponding resonances for TL1, with the exception of several broad downfield peaks (U128-H3 and U129-H3 which neighbors the apical loop, and U134 which neighbors the upper internal loop) likely due to neighboring unpaired, flexible nucleotides. Additionally, U98-H3 which neighbors the flexible A99 bulge based on SHAPE data was difficult to assign in TL1 spectra.

Since NMR chemical shifts are extremely sensitive to local structure, we conclude that the two RNAs have very similar structure. However, we collected most of the data required for structure determination with the UUCG-containing TL1 construct because of reduced spectral line width, as illustrated in the spectral comparison. We have found in multiple projects that

large unpaired loops lead to reduced spectral quality, perhaps as a result of aggregation under NMR conditions.

A distinct difference in stability between the lower and upper internal loops was observed when examining H_2O NOESY spectra. The upper, cytosine-rich internal loop exhibited no cross-strand NOE interactions. Additionally, the neighboring U110-A131 and G111-A130 base-pairs exhibit a high degree of chemical exchange suggesting the formation of base-pairs, but also exposure to solvent from partial opening of the base pairs. Surrounding unpaired nucleotides C109 and C112 both exhibit high SHAPE activity suggesting their flexibility impart deviation from A-form helical base-pairing for U110-A131 and G111-C130, and globally destabilizing this portion of the structure.

In contrast, the G102-H3 and G106-H3 base-pairs, proximal to the lower internal loop, were assigned unambiguously with G102 exhibiting NOE interactions to both U103-H3 and U139-H3. The non-canonical U103-U139 base-pair was corroborated by strong cross-peaks between U103-H3 and U139-H3 within the non-canonical base-pairing range (12-10 ppm) (**Fig. 4.3B**). U103-H3 and U139-H3 exhibit several intermolecular NOEs with C104 and C138-amino protons as well, suggesting a well-ordered lower internal loop structure. The flexible nucleotides within the upper internal loop, the destabilized U110-A131 and G111-A130 base-pairs, and the non-canonical U103-U139 base-pair are all likely to play a role in lowering resolution, as observed in the center of the SAXS envelope.

3.2 Analysis of the NMR Spectra

Since the loop nucleotides are flexible and devoid of intrinsic structure, and the spectral quality reduced, we elected to determine the structure (93-114, 128-148) using NMR restraints

obtained from TL1 spectra, where the apical loop is replaced with a stable UUCG tetraloop. This is safe, because the SL1 structure is duplicated in the TL1 RNA, based on almost identical chemical shifts (**Supplementary Fig. S4.5**).

Reduced spectral overlap and increased spectral quality in the TL1 D_2O NOESY spectrum, together with recording ^{13}C filtered 3D NOESY experiments, allowed at least one NOE interaction to be obtained for almost all NMR active nuclei in TL1. Only one disconnect in the typical helical "walk" pattern was observed, proximal to the upper internal loop. The C112 bulged nucleotide exhibits the expected H2' cross-peaks with A113-H8, yet its G111-H2'-C112-H8 interaction is absent. Atypical assignments suggest C112 is displaced from the base-stack because of detection of an NOE interaction between G111-H2' and A113-H8, in addition to an unexpected G111-H1' to A113-H8 interaction (**Fig. 4.4A, B**). Further inspection of the D_2O TOCSY data shows that G111 exhibits mixed 2'-endo/3'-endo sugar pucker, based on relatively strong G111-H1' to H2', H3', and H4' cross-peaks (**Fig. 4.4C**). Nucleotides U110 and A113 neighboring G111 exhibited H1' to H2' TOCSY correlations, suggesting a partial 2'-endo conformation and flexibility for this region of the RNA (**Fig. 4.4C**). A131 and C132 also exhibit deviations from A-form helical patterns, where A131-H2 correlations to G111-H1' are of weak intensity, and C132 exhibits mixed 2'-endo/3'-endo TOCSY cross-peaks (**Fig. 4.4C**). These deviations from A-form helical patterns combined with the SHAPE reactivity for upper internal loop residues C109 and A131 indicate that this region of SL1 is flexible, which likely contributes to lowering the resolution in the center of the SAXS envelope (**Fig. 4.2B**)

Almost all expected adenosine H2 correlations to i+1 and cross-strand i-1 H1' nuclei exhibit the expected medium to strong-intensity cross-peaks, with the exception of the bulged A99, A143, and the aforementioned A131. A99-H2 to U100-H1' and A143-H1' could only be

detected in experiments conducted on the smaller fragment as very low-intensity cross-peaks. By contrast, A143 exhibited strong-intensity NOEs with C45-H1' and low to medium intensity cross-peaks to the cross-strand i+1 A99 and unexpectedly, to the i+2 residue U100 as well. This observation suggests A99 might exchange between a stacked-in and a solvent exposed conformation, and that, when A99 is solvent exposed, the U100-A142 base-pair stacks over the U98-A143 base-pair. However, NOEs for all H1', H2', and H3' nuclei to H6 or H8 for residues near A99 were assigned with the expected A-form intensities. Since the SHAPE reactivity data are high for most residues surrounding the A99 bulge, we suspect that A99, along with U98-C45 and U100-A143, are innately flexible with the A99 base fluctuating between a major conformation (stacked between U98 and U100) and a transient minor conformation where A99 is displaced from the base stack. This is sometimes observed in unpaired A's, for example in the pre-miR-21 structure (Shortridge, et al, 2017).

3.3 Structure of the eIF3 binding stem-loop of c-JUN

The structure of TL1 was calculated using a simulated annealing protocol from the NIH-XPLOR package (Schwieters et al, 2005). We describe herein the ensemble of 10 structures chosen based on lowest energy score from a refinement calculation of 200 structures, which converged to a 1.29 Å heavy atom RMSD. Alignment of individual sections of the RNA were all below 1.00 Å, with the exception of internal loop proximal residues 108-113 and 129-134 (**Fig. 4.5A, Supplementary Table S4.1**). Additional refinement was attempted by incorporating residual dipolar coupling (RDC) data into the XPLOR refinement script. However, inclusion of RDC's into the calculations produced no significant improvement in ensemble convergence. We attribute this result to the combination of a high number of NOE restraints per nucleotide and,

more significantly, to the conformational flexibility of the center of the molecule, which prevents an accurate estimation of the alignment tensor vector (Lipsitz et al, 2004).

All models exhibit an unusual narrow to wide major groove width near the center of the stem-loop. Using the 3DNA software suite, we calculated major groove widths at its narrowest location (A101-P to A134-P) to be ~9 Å and at the widest location (C104 to A131) to be ~16Å (Lu et al, 2003) (**Supplementary Fig. S4.6)**. It is likely that the narrow major groove width is a result of the A99 bulge and unpaired C132 and C133 bases, positioned within one helical pitch from each other on opposite strands of the helix, which push the phosphodiester backbones towards each other to accommodate the unpaired nucleotides within the base-stack. We attribute the widened groove to the U103-U139 base-pair. The variation in groove width coincides with the pivot point for the stem-loop to bend ~20-30 degrees from the helical axis, consistent with the SAXS data, while local flexibility and variability at the upper internal loop could explain the low-resolution mid-section of the SAXS model.

Consistent with the SAXS data, the upper internal loop and proximal base pairs (108-113, 129-134) exhibit the largest local RMSD (1.23Å). Inspection of the ensemble of structures shows that this is due to the combined effect of the unpaired C109, C132, C133 residues, and the presence of the C112 bulge that together disrupt base-pair planarity and hydrogen bonding geometries for the neighboring G111-C130 and U110-A131 base-pairs as well (**Fig**. **4.5B**). The unpaired cytosines (C109, C132, C133) converge poorly relative to each structure in the ensemble, yet in all models are positioned within the base-stack. We reason that the variable positions for these bases is indicative of their flexibility, explaining the disrupted the planarity of the neighboring U111-A131 base-pair to some extent. The G111-C130 base-pair may be affected as well, but C112 is also likely to play a role since it is displaced from the base stack in all

models, in a manner reminiscent of a previously submitted structure, PDB 3J3F (Anger et al, 2013) (**Fig. 4.5C)**. Interestingly, high SHAPE reactivity was detected for both C109 and C112, yet only C112 is displaced from the helix. We attribute this property to C112 being sandwiched between two base-pairs, one of which is a rigid GC base-pair, without an unpaired nucleotide on the opposite strand. In contrast, C109 neighbors two AU base-pairs and its stacking between neighboring A108 and U110 is likely facilitated by the flexibility of the unpaired C132 and C133 nearby, suggesting its SHAPE reactivity arises from a transient minor conformation.

The A99 bulge converges well within the ensemble and appears well-ordered (**Fig. 4.5D**), with a large number of -intra and -inter ribose-to-base NOE restraints. The A99 base is inserted between the U98-A143 and U100-A142 base-pairs with the only deviation from A-form helices in that U100 exhibits abnormal backbone angles ($\alpha = 156°$, $\beta = -160°$, $\gamma = 167°$) to accommodate A99. This observation supports our hypothesis that the high SHAPE reactivity near the A99 bulge is due to a transient solvent exposed conformation that affects A99 and its neighboring base-pairs U98-A143 and U100-A142, which is likely to be missed in the NMR refinement that relies on the observation of the more highly populated structure.

U103 and U139 are base-paired in each model in the ensemble even when calculations were executed without inclusion of hydrogen bonding restraints. The base pair is slightly concave and twisted from ideal geometry. Strong intensity NOEs between the H3 nuclei and local amino protons favored the following hydrogen bond pattern: U139-H3 to U103-O2 (2.3-2.5Å) and U103-H3 to U139-O4 (2.1-2.3Å) (**Fig**. **4.5E**). The UU base-pair promotes a highly stacked lower internal loop with long-range electrostatic interactions observed between U137-O4 to C105-H41 (3.0-3.2Å) and U137-H3 to C105-N3 (3.3-3.5Å) (**Fig**. **4.5E**).

When we examined the structure of the SL1 apical loop (nucleotides A115 to U127), the spectra contained significant overlap and broadening, resulting in a more limited number of unambiguous assignments from D_2O NOESY experiments. However, H1' and H2' to H6 and H8 assignments were almost continuously observed in a deuterated sample for apical residues A115 to U127, with the only break observed between U121 and U122 (**Supplementary Figure S4.7)**. Interestingly, intramolecular H1' and H2' to H6 and H8 cross-peaks for nucleotides at the stem-loop junction (A116, U117, C118 and C125, U126, U127) exhibited expected low-intensity NOEs, whereas nucleotides U119 to U124 deviated from the standard helical "walk" cross-peak patterns and exhibited low to medium intensities, commonly seen for nucleotides in anti-syn equilibrium.

We merged restraints derived from the deuterated SL1 D_2O NOESY spectra with restraints for residues 5'-G93 to A114, U128 to 3'-C148 derived from TL1 and intraresidue restraints for A155 to A128. In the 10 structure ensemble of low energy structures, the apical loop is surprisingly well resolved, considering its large size and lack of base-pairs (**Fig. 4.6A**, **Supplementary Table S4.2**). Modeling the lowest energy structure into the SL1 SAXS envelope provides good matching as well, including the lower resolution for the dynamic apical loop (**Fig. 4.6B**). Within it, unpaired nucleotides form an extended helical turn without well-organized base stacking or pairing (**Fig. 4.6C)**, with the exception of A120, U121, and U122 bases, which are in most models completely unstacked and induce the backbone turn with U121 extruded from the loop, consistent with the gap of H1' to H6 and H8 connections in the helical walk. At the stem-loop interface, A115-U127 exhibit regular conformation (**Fig. 4.6D**). During calculations, we initially chose not to enforce the A115-U127 base-pair due to its ambiguous H3 assignments.

Yet, in most models the base-pair stacks above the A115-U127 base-pair and falls within the range and orientation of Watson-Crick hydrogen bond formation.

3.4 Structure of the SL1-SL2 Junction

We next investigated the larger RNA, including the 3-way junction formed by SL1 and SL2, because this extended element is reminiscent of the junctions found in HCV-like IRESs and involving two helices (IIIa and IIIc) that neighbor IIIb (**Fig. 4.1A, B**). The junction and subdomains are critical for 40S recruitment and eIF3 binding to HCV-like IRESs (Kieft et al, 2001).

Constructs were generated containing SL1 and SL2 (SL1-SL2) to assess its ability to fold and confirm the secondary structure (**Fig. 4.1A**). Using an initial low-resolution SAXS model, we observed a ~75 Å envelope similar to the SL1 structure that meets at a possible junction site with a second portion of the envelope that is 50Å long, which we tentatively assigned to be representative of the SL2 and 5'-3' helices (**Fig. 4.7A**). In each SAXS model, subtle extensions located at the presumed junction site suggest formation of the SL2 stem-loop, stacked over the short 5'-3' stem.

A series of constructs containing SL1 and SL2 were generated to establish the SL1-SL2 secondary structure by NMR spectroscopy, with varying length of the 5'-3' stem, starting with 5'-81 to 3'-179, 5'-84 to 3'-176, and 5'-88 to 3'-172. We found that a full stem length (81-179) was required to fold SL1 by tracking the distinctive imino resonances of the U103-U139 base pair, which align well with the SL1 assignments (**Supplementary Fig. S4.8, S4.9**). We then synthesized TL1*-SL2, where SL1 was truncated and capped with a UUCG tetraloop to improve spectral quality for assignments and secondary structure determination (**Supplementary Fig.**

S4.10). All expected base-pair NH's were assigned for this RNA, with the exception of helix termini G81, G90, G93, and G151, which are missing, most likely due to solvent exchange (**Supplementary Fig. S4.11**). Almost all imino resonances had NOE interactions with neighboring imino nuclei, with the exception of SL2 iminos U168, the U155-G165 wobble base-pair, and U164. H_2O NOESY of a double tetraloop construct TL1*-TL2 were used to confirm the identity of U155, U164, and G165 imino resonances by providing weak intensity cross-peaks in this stabilized construct (**Supplementary Fig. S4.12**).

Crystal structures of the HCV IIIabc RNA show that helix IIIa coaxially stacks with helix IIIb, and helix IIIc stacks with the 5'-3' stem instead. We suspected SL1-SL2 would exhibit similar coaxial stacking and that SL2 would fold similarly to helix IIIc because the junction between SL2 and the 5'-3' stem lacks any interhelical nucleotides, as is the case for IIIc and its respective 5'-3' stem. Based on three-way junction topology, lack of junction nucleotides predicts that SL2 would be coaxially stacked with the 5'-3' stem (Lescoute et al, 2006).

We attempted using NMR to identify NOE interactions between the G90-C170 and G151-C169 base-pairs, which would be indicative of the base-pairs being stacked coaxially, but the relevant iminos are unassignable due to their rapid solvent exchange. Thus, we resorted to modeling the junction using simRNA, a coarse-grained molecular dynamics package, to generate representative models of the junction (Boniecki et al, 2016). In total, 24 models were generated using the simRNA platform, guided by the validated base-pairing established using NMR experiments, then followed-up with Rosetta RNA De Novo energy minimization (Das et al, 2010). Global structure convergence was poor (RMSD = ~15Å) due to fluctuation in the orientation of SL1, branching from the 3-way junction. However, when aligned with only SL1 or SL2 individually, the RMSD for the selected stem-loops improves dramatically providing

support for the predictive accuracy of the models (**Supplementary Fig. S4.13**). Several models fit well into the SAXS envelope as well, further supporting our prediction that SL2 and the 5'-3' stem occupy the shorter (50Å) protruding region of the model, while SL1 occupies the longer (75Å) extension (**Fig. 4.7B**). Interestingly, every one of the 24 models demonstrated coaxial stacking between SL2 and the 5'-3' stem, further corroborating the predicted three-way junction topology (**Fig. 4.7C, Supplementary Fig. S4.13**). Short of a formal X-ray structure, the simRNA models fit well enough into the SAXS envelope to warrant modeling eIF3 recognition based on previously reported IRES-eIF3 complexes, as discussed below (Sun et al, 2013; Hashem et al, 2013; Erzberger et al, 2014).

Section 4: Discussion

Specialized translation initiation is a cap-dependent mechanism of protein synthesis that deviates from the eIF4E-dependent canonical cap-dependent pathway and instead utilizes a novel m^7G cap recognition function of eIF3d (Lee et al, 2015; Lee et al, 2016). Specific transcripts such as c-JUN mRNA are programmed to be translated through this pathway by RNA structures within their 5'-UTR that directly recruit eIF3 to the 5'-UTR upstream of the initiation codon (**Fig. 4.1A**). This mechanism has similarities to IRES-driven translational initiation; in both cases, eIF3 recognizes specific RNA structures and these interactions drive recruitment of the mRNA to the ribosome for initiation of protein synthesis (Sizova et al, 1998; Kieft et al, 2001; Collier et al, 2002), but is also distinct because it requires cap recognition. While the interaction of eIF3 with IRESs has been investigated, very little is known regarding how eIF3 recognizes features of mRNAs that drive specialized translation initiation. Here, we report the structure of the RNA element found in the c-JUN 5'-UTR and compare it to HCV IRES subdomains recognized by eIF3.

Inspection of the NMR structure and SAXS-derived model shows that the RNA contains two internal loops, one well-structured and the second more flexible, that create atypical major groove widths and subtle bending of the hairpin from its helical axis. We generated computational models of the SL1-SL2 element that resembles the IRES structure and observed consistent coaxial stacking between the SL2 and the 5'-3' termini. These features exhibit interesting parallels with secondary and three-dimensional structural features of the HCV IRES.

In IRESs, sequences that bind to eIF3 are inconsistently conserved (Smith et al, 1995; Brown et al, 1992; Kieft et al, 2001), suggesting that secondary and tertiary structural features of the RNA, rather than sequence, are recognized by eIF3. For instance, eIF3 recognition is

completely suppressed only when subdomain IIIb, a 50-60 nucleotide stem-loop, is destabilized (Kieft et al, 2001, Collier et al, 2002). Our NMR structure of the c-JUN 5'-UTR and the SAXS models suggest that SL1-SL2 adopts a similar structure, suggesting these RNAs might be recognized similarly by eIF3.

Cryo-EM studies show that HCV IRES IIIabc subdomain binds to eIF3 between the highly basic eIF3a and eIF3c lobes within the eIF3 octameric core (Sun et al, 2013). The size of this pocket (70-75Å) fits the size of SL1-SL2 well, suggesting that the SL1-SL2 element, and not just SL1, is recruited by eIF3 in a manner that is similar to the IIIabc structure (**Fig. 4.8A**). This model provides a unified view of eIF3 in recruitment of IRES and of elements within the 5'-UTRs as occurs in c-JUN specialized translation initiation

In this model, HCV-like IRES and specialized translation initiation mechanisms both share similar modes of eIF3 recognition driven by the molecular shape of the RNA recognition elements, but the similarities between the c-JUN mRNA and HCV IRES IIIabc structures extend beyond size and secondary structure. The middle of the SL1 RNA contains a flexible cytosine-rich internal loop, which causes an enlarged major groove and a kink in the stem-loop, bending it away from its helical axis, much like subdomain IIIb, even if the detailed structural features responsible for the bending differ between the two (**Fig. 4.8B**). We hypothesize this internal loop region serves analogous roles in eIF3 recognition, although functional validation will be required to confirm this hypothesis.

Apical loops vary in size across HCV-like IRES IIIb subdomains making their contribution to specific recognition unclear. For instance, the CSFV IIIb domain harbors a stable tetraloop, whereas SL1 and HCV IIIb are capped by 11-14 nucleotides, flexible pyrimidine-rich loops (Brown et al, 1992). Pyrimidine-rich tracts are found in other IRES elements and have

been shown to bind auxiliary proteins such as hnRNPs, increasing IRES activity (Jang et al, 1990; Luz et al; 1991; Conte et al, 2000; Hellen et al, 1994). This divergence poses the question of whether the SL1 apical loop binds proteins to regulate c-JUN translation rates as well.

Additional structural data of this RNA bound to eIF3 or its subunits, as well as functional data, will be needed to achieve an understanding of eIF3 recognition of c-JUN mRNA and its parallels with HCV-like IRES recognition. However, our chemical probing, SAXS, and NMR spectroscopy approach provides the first insight into three-dimensional RNA structures required for specialized translation initiation. HCV IRES subdomains are anti-viral drug targets (Gallego et al, 2002; Klinck et al, 2000; Dibrov et al, 2014), therefore, targeting the SL1 upper internal loop or the SL1-SL2 3-way junction may prove a viable approach for reducing specialized translation initiation of c-JUN and other proliferative genes that are aberrantly expressed in therapeutic resistant tumors.

In summary, our model suggests that c-JUN mRNA is initially recognized by eIF3 similar to the HCV-like IRES mechanism. However, it remains to be explained how eIF3d cap recognition fits into this model. During specialized translational initiation, mRNA-cap recognition by eIF3d is required for a specialized transcript to bypass canonical machinery (Lee et al, 2015; Lee et al, 2016). The protein DAPI5 binds to eIF3d, replacing the ribosomal loading function of eIF4G (de la Parra et al, 2018). Although requiring further eIF assistance beyond eIF3 for initiation is not a characteristic of an HCV-like IRES commonly found in *Flaviviridae* viruses, this is the case for IRES found in *Picornaviridae* viruses, referred to as “Class I” and “Class II” IRESs (Mailliot et al, 2017). Class I and Class II IRESs typically lack sophisticated and large secondary structures compared to HCV-IRES, and require in most cases almost all initiation factors, including eIF4G, with the exception of canonical cap-recognizer eIF4E

(Mailliot et al, 2017). In the case of c-JUN mRNA, the SL1-SL2 subdomain is initially recognized like that of the HCV IRES subdomain IIIabc, but the rest of the 5'-UTR lacks the several other HCV-like IRES subdomains required for ribosomal delivery. Instead, it may be that c-JUN requires eIF3d and DAPI5 guided cap recognition to compensate for its lack of sophisticated 5'-UTR secondary structure, mimicking Class I and Class II IRES by their necessity for eIF4G function. We even question if all specialized translation initiation transcripts are composed of different compilations of IRES family characteristics based on their 5'-UTR secondary structures. 5'UTR structure for each individual transcript potentially programs different interchangeable IRES subdomains observed in all IRES types, differentiating specialized translation rates for each transcript based on their encoded protein function and impact on the cell.

Section 5: Materials and Methods

5.1 RNA Preparation

All RNAs were transcribed *in vitro* using in-house T7 polymerase using DNA templates (Integrated DNA Technologies) containing the T7 promoter. Transcriptions were concentrated by ethanol precipitation and purified by gel extraction using 12.5-20% denaturing polyacrylamide gel electrophoresis. RNA was removed from gel fragments by electro-elution, concentrated by ethanol precipitation, resuspended in 20 mM sodium phosphate (pH = 6.0), 100mM KCl, 500mM NaCl, and 1mM EDTA buffer, then dialyzed into 20 mM sodium phosphate (pH = 6.0), 0.01mM EDTA for NMR data collection. Deuterated RNA samples were synthesized using partially deuterated (D-H5, H3', H4', H5', H5'') rNTPs (Cambridge Isotope Laboratories). ^{13}C and ^{15}N labeled samples were synthesized using fully labeled rNTPs (Cambridge Isotope Laboratories).

5.2 Selective 2'-OH Acylation Analyzed by Primer Extension (SHAPE)

SHAPE chemical probing on the SL1 construct to verify formation of the same secondary structure observed in the context of the complete 5'-UTR (Wilkinson et al, 2006). The resulting chemical modifications were analyzed by primer extension with a radiolabeled DNA oligonucleotide and separated by gel electrophoresis. In order to accommodate primer binding, SL1 was transcribed with a 3'-extension for priming reverse transcription. Chemical acylation and reverse transcription conditions followed a procedure similar to previous work (Barnwal et al, 2016). The secondary structure of SL1 was calculated using RNAStructure using SHAPE reactivities as soft restraints (Mathews et al, 2004). Secondary structure images were rendered with VARNA and colored based on RNAStructure output (Darty et al, 2009).

5.3 Small Angle X-ray Scattering (SAXS)

RNA samples for small angle X-ray scattering (SAXS) were prepared similarly to samples made for NMR measurements, but at the lower concentration range of 2–4 mg/ml. SAXS experiments were recorded on NANOSTAR, a small angle scattering spectrometer with a Bruker Turbo rotating Cu-Anode X-ray source at the National Magnetic Resonance Facility at Madison. The particle distance distribution function P(r) plots were calculated using GNOM (Svergun et al, 1992). Model generation followed a procedure similar to previous work (Barnwal et al, 2016). In summary, low resolution *ab initio* shape reconstruction was done first using DAMMIN (Svergun et al, 1999). A total of 20 models were generated with DAMMIN using *in house* written scripts. The best and final model was selected with a suite of software (DAMSEL, DAMSUP, DAMAVER and DAMFILT) (http://www.embl-hamburg.de/biosaxs) for comparison and fitting to the NMR structure (Volkov et al, 2003).

5.4 NMR Experiments

All NMR experiments were executed on Bruker Advance 800 or 600 MHz spectrometers equipped with triple resonance cryogenic probes. All NMR samples were prepared to 400 – 1000μM and, before each experiment, freshly annealed by heating to 90°C for 2-3 minutes and immediately snap cooled to -20°C. Exchangeable 2D 1H-1H NOESY and 2D [1H-15N] HSQC spectra were recorded in 95% H_2O/5% D_2O in 20 mM phosphate buffer at pH 6.0 and 0.01 μM EDTA at 5°C. The non-exchangeable 2D 1H-1H NOESY, 1H-1H TOCSY, 2D [1H-13C] HSQC, 1H-13C NOESY-HSQC were recorded in 99.9% D_2O buffer at 25°C. The exchangeable and non-exchangeable NOESY spectra were recorded with various mixing times (100–300 ms) to

facilitate spectral assignments and quantification of cross peak intensities by comparison of with peaks corresponding to known distances (e.g. H5-H6). All NMR data were processed either with Topspin (www.bruker.com) or NMRPipe (Delagio et al, 1995) and analyzed using Sparky (Lee et al, 2015) or CCPNMR (Vranken et al, 2005).

5.5 Experimental Constraints

Assignments of RNA spectra were guided by predicted RNA chemical shift values and using established NOE helical "walk" patterns (Wüthrich et al, 1986; Varani et al, 1996). Experimental constraints for structure calculations of SL1 were compiled by dividing the RNA into two segments: stem (93-114, 128-148) and apical loop (115-127). Stem constraints were obtained by grafting a UUCG tetraloop to replace the dynamic apical loop to generate construct TL1, that improved spectral quality. We have often found that large flexible apical loops, as found in this RNA, lead to loss of spectra quality, perhaps by inducing aggregation (Varani et al, 1991; Barnwal et al 2016). TL1 assignments were used in restrained molecular dynamics calculations of the stem structure. Overlay of TL1 2D 1H-1H NOESY spectra with a spectrum of perdeuterated SL1 showed remarkable similarities in the chemical shifts and allowed the direct transfer of TL1 assignments to SL1 assignments for apical loop calculations. Distance restraints were derived from NOE intensities and organized into 'strong' (2.5Å ± 0.7Å), 'medium' (3.5Å ± 1.5Å), and 'weak' (4.5Å ± 2.0Å) bins based on peak intensities relative to fixed A-form helical distances (e.g. H5-H6 = 2.5Å, H3'-H6/H8 = 3.5Å). Restraints for overlapped regions were confirmed using truncated version of TL1 (TL-BOT and TL-TOP) in addition to corroboration by 3D ^{13}C-NOESY−HSQC spectra. Hydrogen bond, planarity and dihedral restraints were

included for base-paired nucleotides that were surrounded by base-pairs confirming to A-form helicity.

5.6 Structure Calculations

Xplor-NIH was used for all structure calculations (Schwieters et al, 2003). Compiled experimental restraints are initially used in a simulated annealing procedure starting from randomized coordinates and initially undergo high temperature (3500K to 298K) torsional angle dynamics where incremental decreases in temperature are represented by progressively introducing van der Waals terms and raising the force constants for angles, dihedral angles, NOEs and van der Waals repulsions. Following the final cooling step, the molecule underwent two sequential refinement steps in torsional angle space, then in Cartesian space. The lowest energy structure with minimum violations of distance (>0.5A) or torsion angle (>5°) restraints were used for in a follow-up refinement step incorporating a statistical base-base position potential for base-paired nucleotides. The 10 structure TL1 ensemble was calculated from 200 independent calculations using restraints from NOESY experiments and predicted values for A-type helical base-pairs (backbone and ribose dihedral angles, hydrogen-bond and planarity distance restraints). The 10-structure SL1 ensemble was calculated from 200 independent calculations using all NOE restraints from TL1 for residues 93-113 and 129-148 based on <0.1 ppm chemical shift overlap between the two spectra, NOEs from deuterated SL1 D2O NOESY spectra, and predicted values for A-type helical base-pairs (backbone and ribose dihedral angles, hydrogen-bond and planarity distance restraints).

5.7 Molecular Dynamic Simulations

A coarse grain molecular dynamics simulation software, simRNA, was used to computationally model SL1-SL2 (Boniecki, 2016). These calculations were performed in parallel, using the replica exchange Monte Carlo method, on the NMRBox server (Maciejewski et al., 2017). The secondary structure restraints were based on H1 and H3 imino resonance assignments from constructs SL1, TL1*-SL1 and TL1*-TL2. The total amount of time to run these calculations was approximately 323 CPU hours, using 3 CPUs per structure and 15 models per CPU. After obtaining the general topology of each model using simRNA, the models were further refined using a full atom minimization in Rosetta RNA De Novo (Das et al, 2010).

5.8 Computational Modeling

Structural analysis and RMSD calculations were done using a combination of PyMol and VMD (Schrödinger LLC, 2010; Humphrey et al, 1996). SAXS models were fitted using Chimera and positions were validated based on best global superposition relative to the structure coordinates (Pettersen et al, 2004; Cantara et al, 2017). The SL1-SL2 model bound to eIF3 was initially built by recognizing eIF3a and eIF3c subunits from previously predicted binding sites for HCV-IRES IIIabc domain (Sun et al, 2013; Hashem et al, 2013; Erzberger et al, 2014) and inspection of eIF3 subunits with basic patches, indicative of high nucleic acid binding affinity based on global electronegativity mapping using PyMol. Previous models of IRES elements bound to eIF3 locate the RNA within a cleft formed by the eIF3a and eIF3c subunits, which is also rich is basic sidechains. Positioning of SL1-SL2 was first attempted using the Chimera docking tool followed by modifications in PyMol for realistic positioning of the solvent exposed portion of SL1-SL2

Section 6: References

1. Andaya, A., Villa, N., Jia, W., Fraser, C., & Leary, J. (2014). Phosphorylation stoichiometries of human eukaryotic initiation factors. *Int J Mol Sci*, *15*(7), 11523-11538.
2. Anger, A. M., Armache, J. P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D. N., & Beckmann, R. (2013). Structures of the human and Drosophila 80S ribosome. *Nature*, *497*(7447), 80-85.
3. Barnwal, R. P., Loh, E., Godin, K. S., Yip, J., Lavender, H., Tang, C. M., & Varani, G. (2016). Structure and mechanism of a molecular rheostat, an RNA thermometer that modulates immune evasion by Neisseria meningitidis. *Nucleic Acids Res*, *44*(19), 9426-9437.
4. Bhat, M., Robichaud, N., Hulea, L., Sonenberg, N., Pelletier, J., & Topisirovic, I. (2015). Targeting the translation machinery in cancer. *Nat Rev Drug Discov*, *14*(4), 261-278.
5. Blau, L., Knirsh, R., Ben-Dror, I., Oren, S., Kuphal, S., Hau, P., ... & Vardimon, L. (2012). Aberrant expression of c-Jun in glioblastoma by internal ribosome entry site (IRES)-mediated translational activation. *Natl Acad Sci U S A*, *109*(42), E2875-E2884.
6. Boniecki, M. J., Lach, G., Dawson, W. K., Tomala, K., Lukasz, P., Soltysinski, T., ... & Bujnicki, J. M. (2015). SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res*, *44*(7), e63-e63.
7. Brown, E. A., Zhang, H., Ping, L. H., & Lemon, S. M. (1992). Secondary structure of the 5′ nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res*, *20*(19), 5041-5045.

8. Buratti, E., Tisminetzky, S., Zotti, M., & Baralle, F. E. (1998). Functional analysis of the interaction between HCV 5′ UTR and putative subunits of eukaryotic translation initiation factor eIF3. *Nucleic Acids Res*, *26*(13), 3179-3187.
9. Cantara, W. A., Olson, E. D., & Musier-Forsyth, K. (2017). Analysis of RNA structure using small-angle X-ray scattering. *Methods*, *113*, 46-55.
10. Chard, L. S., Kaku, Y., Jones, B., Nayak, A., & Belsham, G. J. (2006). Functional analyses of RNA structures shared between the internal ribosome entry sites of hepatitis C virus and the picornavirus porcine teschovirus 1 Talfan. *J Virol*, *80*(3), 1271-1279.
11. Collier, A. J., Gallego, J., Klinck, R., Cole, P. T., Harris, S. J., Harrison, G. P., ... & Walker, S. (2002). A conserved RNA structure within the HCV IRES eIF3-binding site. *Nat Struct Mol Biol*, *9*(5), 375-380.
12. Conte, M. R., Grüne, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S., & Curry, S. (2000). Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J*, *19*(12), 3132-3141.
13. Darty, K., Denise, A., & Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, *25*(15), 1974-1975.
14. Das, R., Karanicolas, J., & Baker, D. (2010). Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods*, *7*(4), 291-294.
15. de la Parra, C., Ernlund, A., Alard, A., Ruggles, K., Ueberheide, B., & Schneider, R. J. (2018). A widespread alternate form of cap-dependent mRNA translation initiation. *Nat Commun*, *9*(1), 3068.

16. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. D. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, *6*(3), 277-293.

17. Dibrov, S. M., Parsons, J., Carnevali, M., Zhou, S., Rynearson, K. D., Ding, K., ... & Hermann, T. (2013). Hepatitis C Virus Translation Inhibitors Targeting the Internal Ribosomal Entry Site: Miniperspective. *J Med Chem*, *57*(5), 1694-1707.

18. Erzberger, J. P., Stengel, F., Pellarin, R., Zhang, S., Schaefer, T., Aylett, C. H., ... & Ban, N. (2014). Molecular architecture of the 40S· eIF1· eIF3 translation initiation complex. *Cell*, *158*(5), 1123-1135.

19. Fraser, C. S., & Doudna, J. A. (2007). Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat Rev Microbiol*, *5*(1), 29-38.

20. Gallego, J., & Varani, G. (2002). The hepatitis C virus internal ribosome-entry site: a new target for antiviral research. *Biochem Soc Trans*, *30*(2), 140-145

21. Hashem, Y., Des Georges, A., Dhote, V., Langlois, R., Liao, H. Y., Grassucci, R. A., ... & Frank, J. (2013). Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. *Nature*, *503*(7477), 539-543.

22. Hellen, C. U., Pestova, T. V., Litterst, M., & Wimmer, E. (1994). The cellular polypeptide p57 (pyrimidine tract-binding protein) binds to multiple sites in the poliovirus 5'nontranslated region. *J Virol*, *68*(2), 941-950.

23. Hellen, C. U., & Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev*, *15*(13), 1593-1612.

24. Holcik, M., Sonenberg, N., & Korneluk, R. G. (2000). Internal ribosome initiation of translation and the control of cell death. *Trends Genet*, *16*(10), 469-473.

25. Hsieh, A. C., Liu, Y., Edlind, M. P., Ingolia, N. T., Janes, M. R., Sher, A., ... & Wang, S. (2012). The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature*, *485*(7396), 55-61.

26. Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph Model*, *14*(1), 33-38.

27. Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*, *11*(2), 113-127.

28. Jang, S. K., & Wimmer, E. (1990). Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein. *Genes Dev*, *4*(9), 1560-1572.

29. Johannes, G., Carter, M. S., Eisen, M. B., Brown, P. O., & Sarnow, P. (1999). Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. *Proc Natl Acad Sci U S A*, *96*(23), 13118-13123.

30. Johnson, R., Spiegelman, B., Hanahan, D., & Wisdom, R. (1996). Cellular transformation and malignancy induced by ras require c-jun. *Mol Cell Biol*, *16*(8), 4504-4511.

31. Khan, D., Bhat, P., & Das, S. (2014). HCV-like IRESs sequester eIF3: advantage virus. *Trends Microbiol*, *22*(2), 57-58.

32. Kieft, J. S., Zhou, K., Grech, A., Jubin, R., & Doudna, J. A. (2002). Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation. *Nat Struct Mol Biol*, *9*(5), 370-374.

33. Kieft, J. S., Zhou, K., Jubin, R., & Doudna, J. A. (2001). Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA*, *7*(2), 194-206.

34. Klinck, R., Westhof, E., Walker, S., Afshar, M., Collier, A., & Aboul-Ela, F. (2000). A potential RNA drug target in the hepatitis C virus internal ribosomal entry site. *RNA*, *6*(10), 1423-1431.

35. Lee, A. S., Kranzusch, P. J., & Cate, J. H. (2015). eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature*, *522*(7554), 111-114.

36. Lee, A. S., Kranzusch, P. J., Doudna, J. A., & Cate, J. H. (2016). eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature*, *536*(7614), 96-99.

37. Lee, W., Tonelli, M., & Markley, J. L. (2014). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, *31*(8), 1325-1327.

38. Lescoute, A., & Westhof, E. (2006). Topology of three-way junctions in folded RNAs. *RNA*, *12*(1), 83-93.

39. Lipsitz, R. S., & Tjandra, N. (2004). Residual dipolar couplings in NMR structure analysis. *Annu. Rev. Biophys. Biomol. Struct.*, *33*, 387-413.

40. Lu, X. J., & Olson, W. K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*, *31*(17), 5108-5121.

41. Luz, N., & Beck, E. W. (1991). Interaction of a cellular 57-kilodalton protein with the internal translation initiation site of foot-and-mouth disease virus. *J Virol*, *65*(12), 6486-6494.

42. Maciejewski, M. W., Schuyler, A. D., Gryk, M. R., Moraru, I. I., Romero, P. R., Ulrich, E. L., ... & Hoch, J. C. (2017). NMRbox: a resource for biomolecular NMR computation. *Biophys J*, *112*(8), 1529-1534.

43. Mailliot, J., & Martin, F. (2018). Viral internal ribosomal entry sites: four classes for one goal. *Wiley Interdiscip Rev: RNA*, *9*(2), e1458.

44. Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., & Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A*, *101*(19), 7287-7292.

45. Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J., & Hellen, C. U. (1998). A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initation of hepatitis C and classical swine fever virus RNAs. *Genes Dev*, *12*(1), 67-83.

46. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*, *25*(13), 1605-1612.

47. Quade, N., Boehringer, D., Leibundgut, M., Van Den Heuvel, J., & Ban, N. (2015). Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. *Nat Commun*, *6*, 7646.

48. Querol-Audi, J., Sun, C., Vogan, J. M., Smith, M. D., Gu, Y., Cate, J. H., & Nogales, E. (2013). Architecture of human translation initiation factor 3. *Structure*, *21*(6), 920-928.

49. Schrodinger, L. L. C. (2010). The PyMOL molecular graphics system. *Version*, *1*(5), 0.

50. Schwieters, C. D., & Clore, G. M. (2014). Using small angle solution scattering data in Xplor-NIH structure calculations. *Prog Nucl Magn Reson Spectrosc*, *80*, 1-11.

51. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *J Magn Reson Imaging*, *160*(1), 65-73.

52. Silvera, D., Formenti, S. C., & Schneider, R. J. (2010). Translational control in cancer. *Nature Rev Can*, *10*(4), 254-266.

53. Sizova, D. V., Kolupaeva, V. G., Pestova, T. V., Shatsky, I. N., & Hellen, C. U. (1998). 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*, *72*(6), 4775-4782.

54. Stumpf, C. R., & Ruggero, D. (2011). The cancerous translation apparatus. *Curr Opin Genet Dev*, *21*(4), 474-483.

55. Sun, C., Querol-Audí, J., Mortimer, S. A., Arias-Palomo, E., Doudna, J. A., Nogales, E., & Cate, J. H. (2013). Two RNA-binding motifs in eIF3 direct HCV IRES-dependent translation. *Nucleic Acids Res*, *41*(15), 7512-7521.

56. Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J*, *76*(6), 2879-2886.

57. Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr*, *25*(4), 495-503.

58. Terenin, I. M., Smirnova, V. V., Andreev, D. E., Dmitriev, S. E., & Shatsky, I. N. (2017). A researcher's guide to the galaxy of IRESs. *Cell Mol Life Sci*, *74*(8), 1431-1455.

59. Truitt, M. L., & Ruggero, D. (2016). New frontiers in translational control of the cancer genome. *Nat Rev Can*, *16*(5), 288-304.

60. Valášek, L. S., Zeman, J., Wagner, S., Beznosková, P., Pavlíková, Z., Mohammad, M. P., ... & Gunišová, S. (2017). Embraced by eIF3: structural and functional insights into the roles of eIF3 across the translation cycle. *Nucleic Acids Res*, *45*(19), 10948-10968.

61. Varani, G., Aboul-ela, F., & Allain, F. H. T. (1996). NMR investigation of RNA structure. *Prog Nucl Magn Reson Spectrosc*, *29*(1-2), 51-127.

62. Varani, G., Cheong, C., & Tinoco Jr, I. (1991). Structure of an unusually stable RNA hairpin. *Biochemistry*, *30*(13), 3280-3289.

63. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., ... & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics*, *59*(4), 687-696.

64. Wilkinson, K. A., Merino, E. J., & Weeks, K. M. (2006). Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc*, *1*(3), 1610-1616.

65. Wisdom, R., Johnson, R. S., & Moore, C. (1999). c-Jun regulates cell cycle progression and apoptosis by distinct mechanisms. *EMBO J*, *18*(1), 188-197.

66. Wüthrich, K. (1986). NMR with proteins and nucleic acids. *Europhysics News*, *17*(1), 11-13.

67. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, *31*(13), 3406-3415.

Table 4.1. NMR and Structure Statistics for TL1 RNA

NMR experimental restraints	
Total number of restraints	1380
Average no. of restraints per hydrogen atom	2.6
Average no. of restraints per nucleotide	28.1
Average no. of NOE-derived restraints per nucleotide	19.9
NOE-derived restraints	976
Intra-residue	517
Inter-residue	459
Sequential |i-j| = 1	361
Non-sequential |i-j| > 1	98
Dihedral restraints	280
Hydrogen-bonding restraints	88
Planarity restraints	36
Structure Analysis	
NOE violations (>0.5Å)	1
Dihedral angle violations (>5°)	0
Clashed Contacts (>0.5Å)	2
Bond Angle Outliers (>5°)	3
Heavy atom RMSD from mean structure, Å	
All RNA heavy atoms	1.29
All RNA backbone	1.32
Lower stem (93-103, 139-148)	0.57
Middle stem, upper internal loop (104-109, 132-136)	0.67
Upper stem (110-131)	0.99
Upper internal loop and proximal base-pairs (108-113, 129-134)	1.23
cUUCGg tetraloop (1-6)	0.28

Table 4.2. NMR and Structure Statistics for SL1 RNA

NMR experimental restraints	
Total number of restraints	1302
Average no. of restraints per hydrogen atom	2.2
Average no. of restraints per nucleotide	23.3
Average no. of NOE-derived restraints per nucleotide	16.4
NOE-derived restraints	916
Intra-residue	473
Inter-residue	443
Sequential |i-j| = 1	352
Non-sequential |i-j| > 1	91
Dihedral restraints	264
Hydrogen-bonding restraints	86
Planarity restraints	36
Structure Analysis	
NOE violations (>0.5Å)	1
Dihedral angle violations (>5°)	0
Clashed Contacts (>0.5Å)	1
Bond Angle Outliers (>5°)	2
Heavy atom RMSD from mean structure, Å	
All RNA heavy atoms	2.01
All RNA backbone	1.98
Stem (93-114, 128-148)	1.29
Apical loop (115-127)	2.24

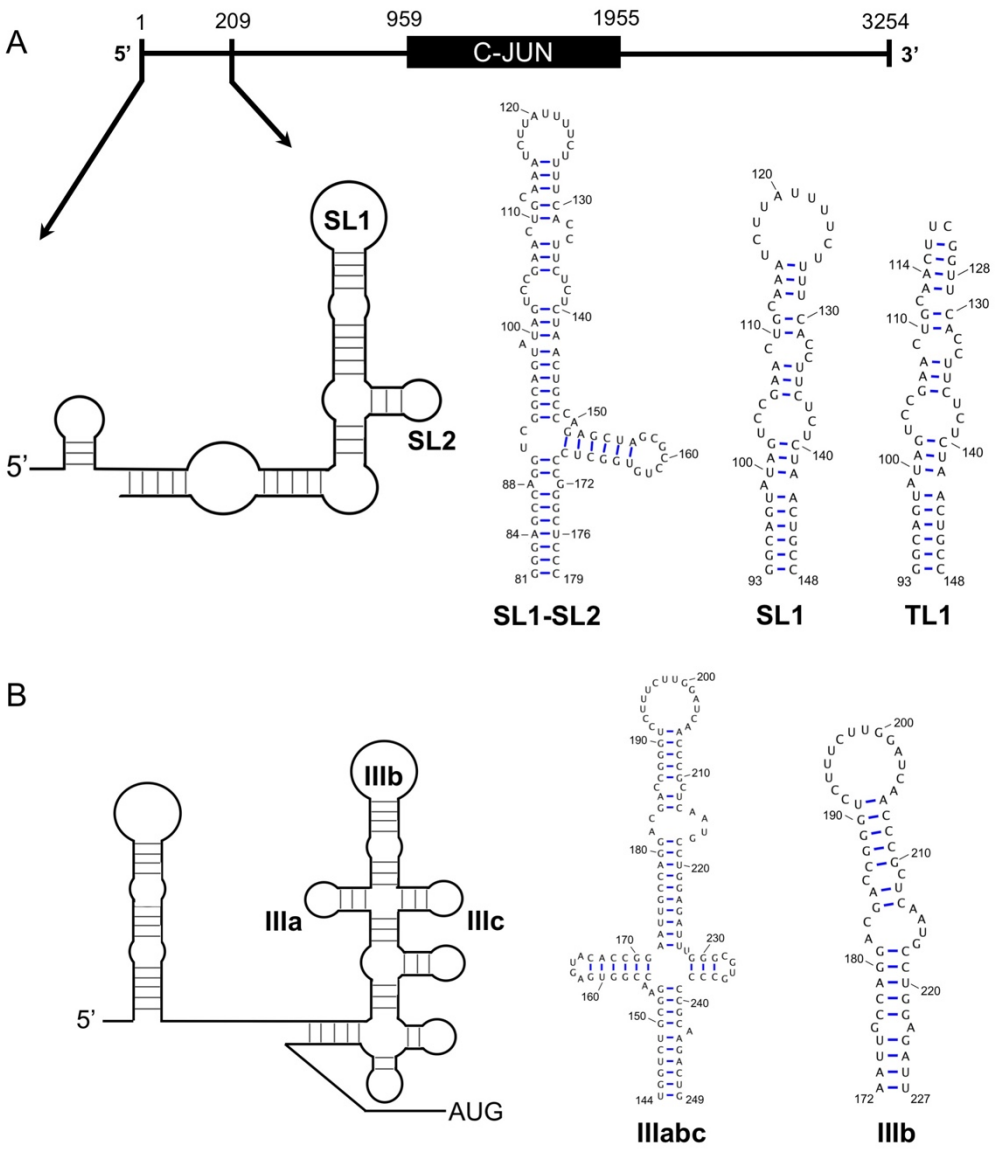


Figure 4.1. (**A**) Representation of the c-JUN transcript showing the position of the stem-loop responsible for specific translational initiation within the 5'-UTR (93-148, SL1) that cross-links to eIF3. Numbering is based on cap analysis gene expression (CAGE) data deposited in FANTOM. The proposed secondary structure of TSS +1-209 is depicted and the location of SL1 and its proximal stem-loop (SL2) are labeled. Sequence and secondary structure of SL1, along with the TL1 construct used for structure determination. (**B**) Representation of the HCV IRES secondary structure with eIF3 subdomains IIIa, IIIb, and IIIc labeled. Secondary structure and sequence of subdomain IIIb (172-227) is represented as well.

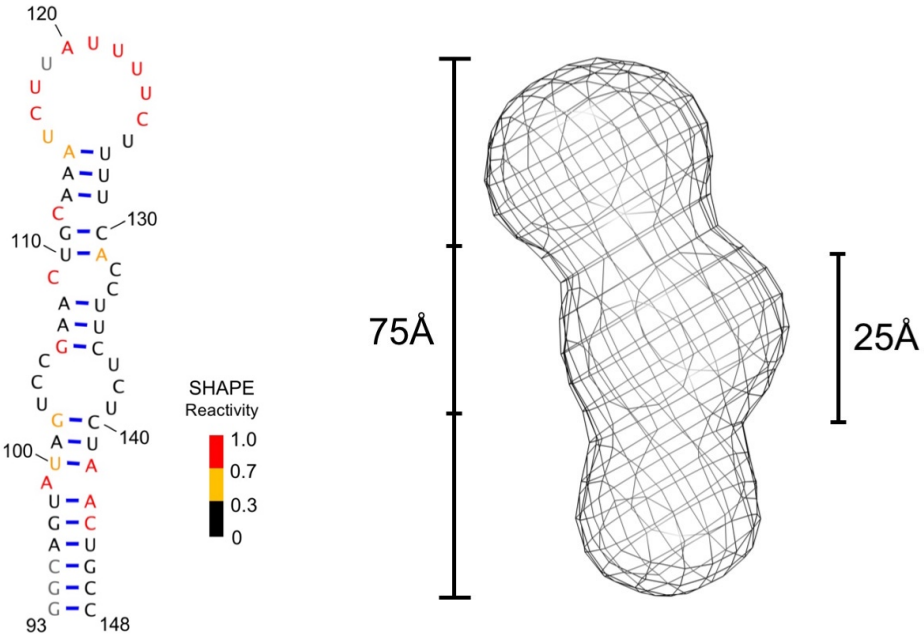


Figure 4.2. (**left**) Secondary structure of SL1 with color-coded SHAPE reactivities mapped to each nucleotide. Reactivities are normalized based on the RNAStructure software (Mathews et al, 2004). No data could be acquired for nucleotides represented in grey. (**right**) Refined SAXS envelope for SL1 generated by ATSAS (Svergun et al, 1999; Volkov et al, 2003).

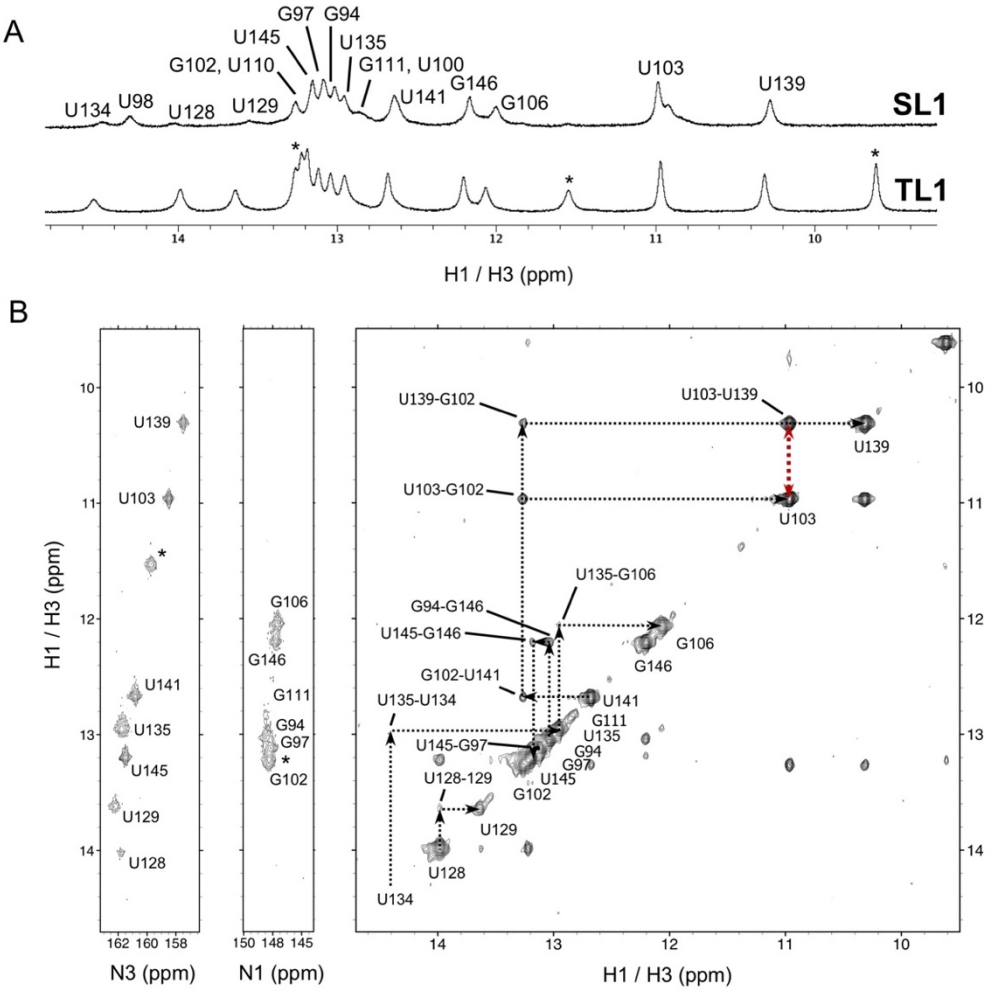


Figure 4.3. (**A**) 1D ^{1}H NMR spectra for the SL1 and TL1 RNAs show very similar spectra, indicative of very similar structures; (*) indicates resonances from the UUCG tetraloop. (**B**) (right) H_2O NOESY spectrum for TL1 with black dotted lines showing sequential NOE correlations involving imino protons. (left) ^{15}N HSQC with N3 and N1 where assignments are aligned with H_2O NOESY cross-peaks, confirming resonance assignments. The red dotted line represents the strong U103-H3 to U139-H3 NOE, supportive of the U103-U139 base-pair.

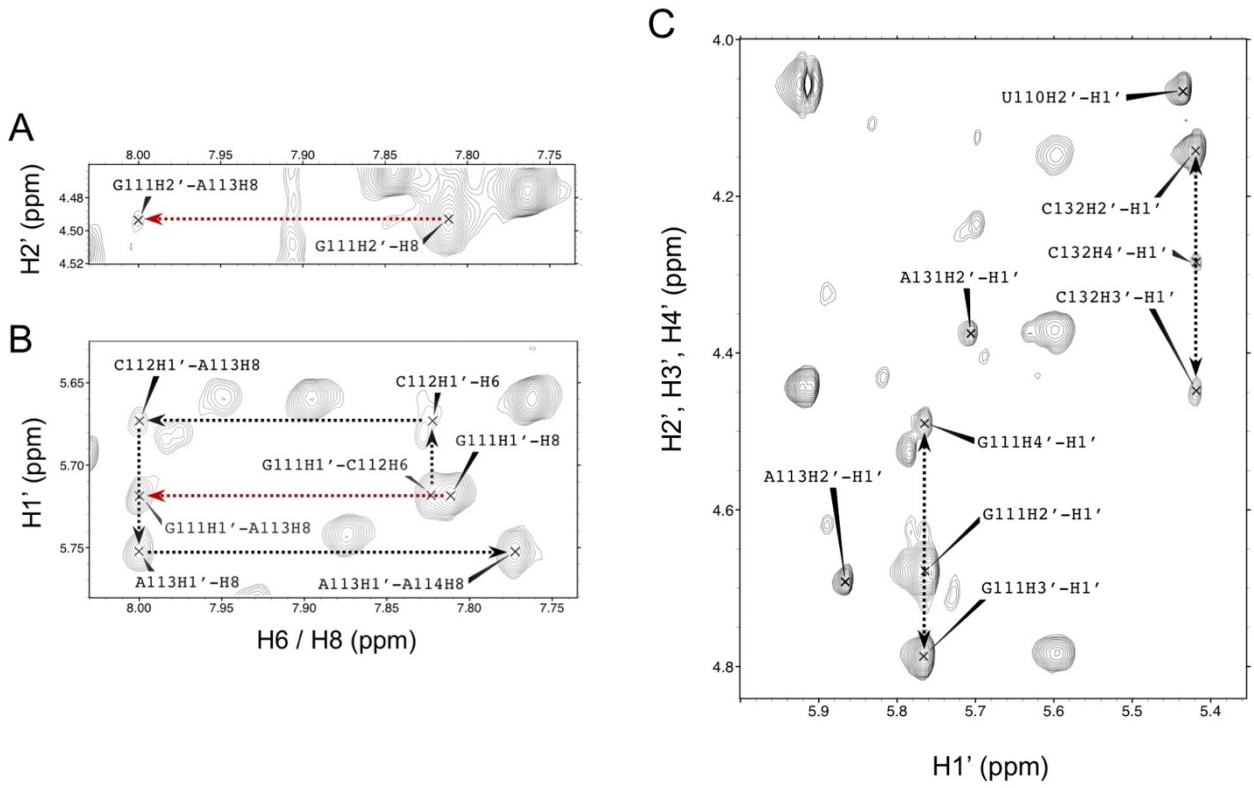


Figure 4.4. Regions of the D_2O NOESY spectrum of TL1 exhibiting atypical NOE patterns for the upper internal loop region. Black dotted lines show sequential NOE interactions and deviations from standard helical patterns are represented as red dotted lines. (**A**) An atypical NOE cross-peak between G111-H2' to A113-H8. (**B**) The H1' to H6/H8 helical "walk" from G111 to A114. The red dotted line represents an atypical NOE interaction between G111-H1' and A113-H8. (**C**) D_2O TOCSY spectrum of TL1 TOP. Peaks between H1' to H2', H3', and H4' for G111 and C132 suggest deviation from canonical 3'-endo conformation to a mixed 2'-endo/3'-endo pattern.

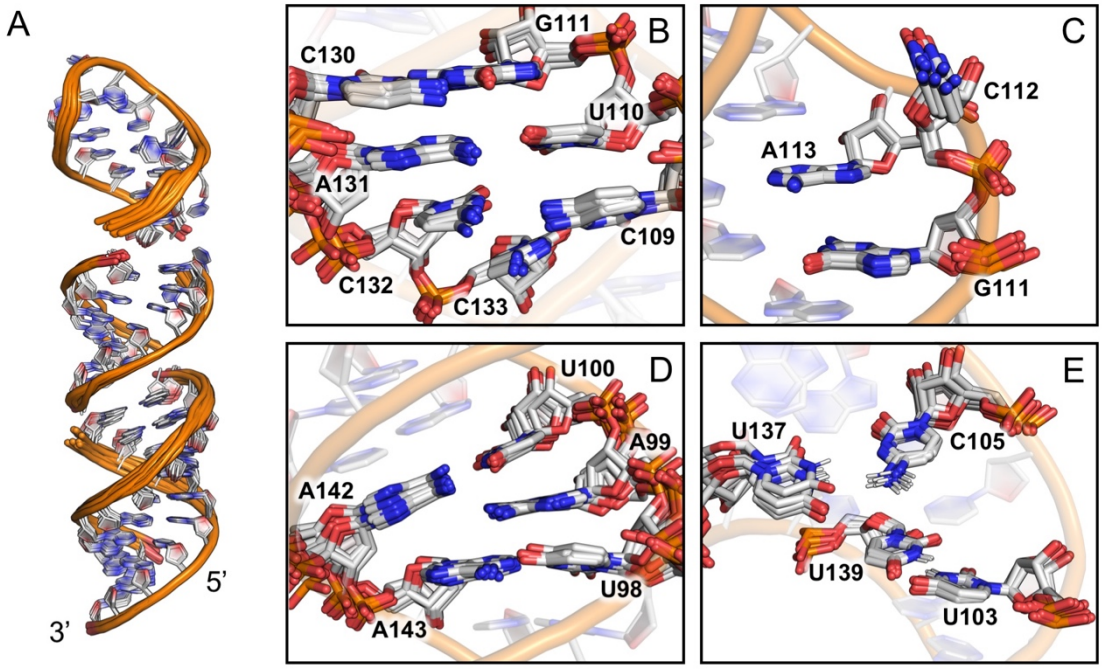


Figure 4.5. Structure of TL1 (white, PDB: 6MXQ) (**A**) Local superposition for: (lower) 93-103, 139-148; (middle) 104-109, 132-138; and (top) 110-131 sections of the structure. (**B**) The unpaired nucleotides C109, C132, and C133 of the upper internal loop. (**C**) C112 bulges out and is displaced from the base-stack between G111 and A113. (**D**) the unpaired A99 is inserted into the base-stack between U98 and U100. (**E**) U103-U139 base-pair and long range C105-U137 electrostatic interactions. B-E shown from sectional alignments of the ensemble.

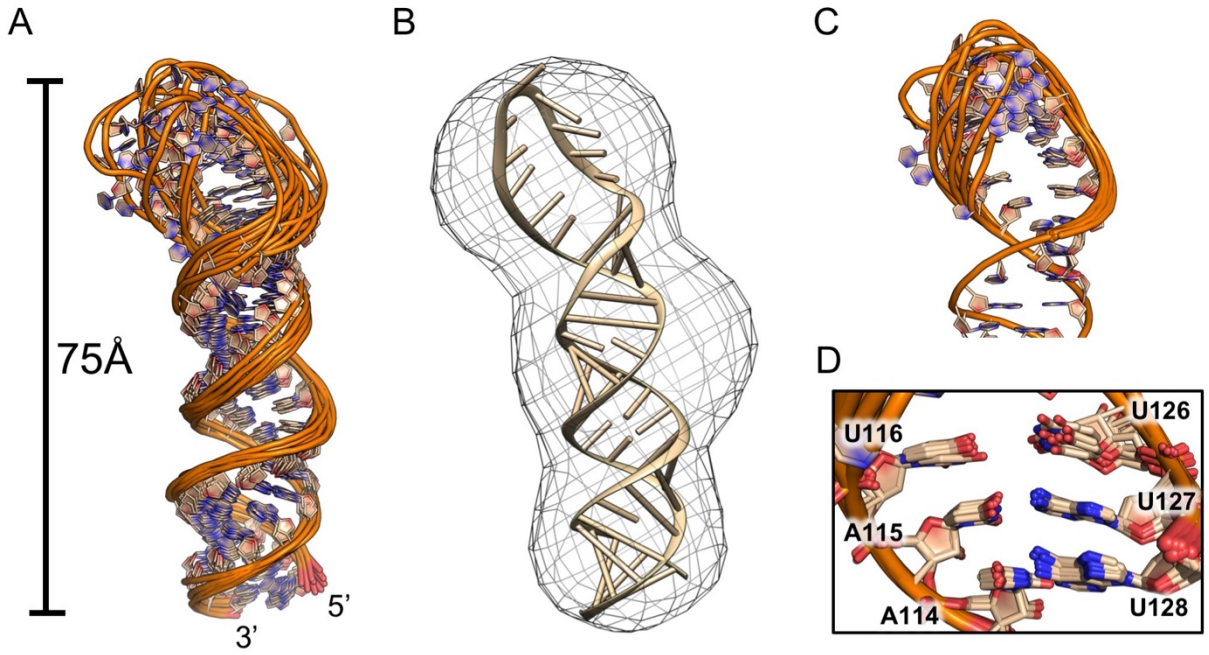


Figure 4.6. Structure of the SL1 apical loop (wheat, PDB: 6NOA). (**A**) Ensemble of 10 converged structures, superposed over all heavy atoms. (**B**) Lowest energy model for the SL1 apical loop fit into the SAXS envelope. (**C**) (top) Ensemble of 10 models superimposed with respect to the apical loop residues. (bottom) A115-U127 base-pairing with U116 and U126 stacking located at the stem-apical loop interface.

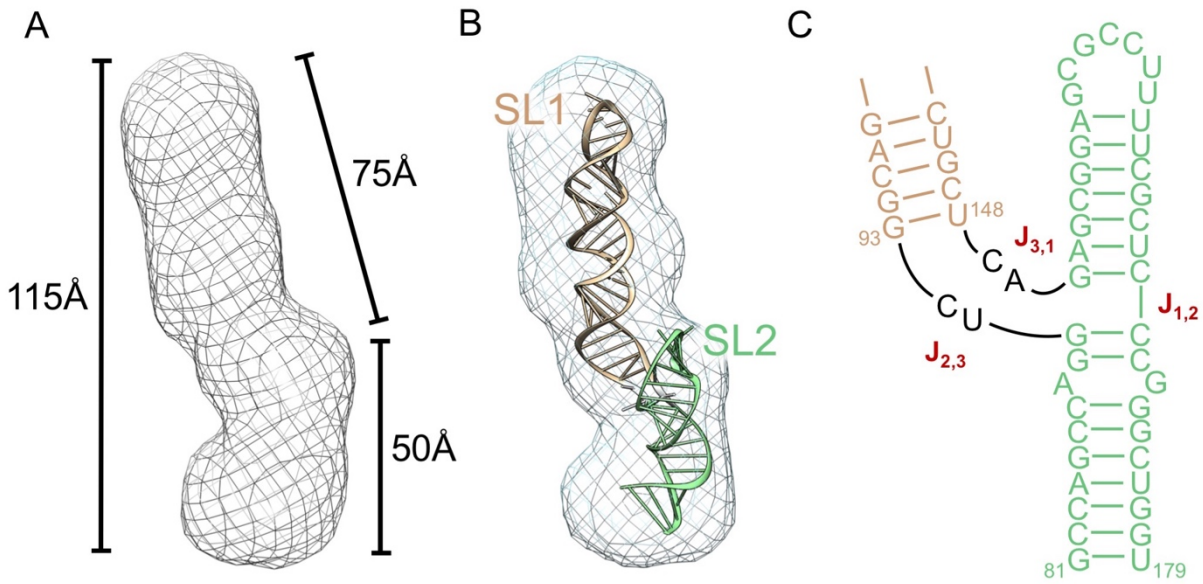


Figure 4.7. (A) SAXS envelope for SL1-SL2 generated by the ATSAS software package. (**B**) Model of SL1-SL2 generated by simRNA and fit into the refined SAXS envelope for SL1-SL2. SL1 (wheat), SL2 and 5'-3' helix (light green). Secondary structure restraints for simRNA calculations were based on the imino NMR peak assignments. SL1 and SL2 annotation are based on the comparison with the SL1 SAXS envelope. (**C**) Secondary structure of SL1-SL2 represented with SL2 and the 5'-3' stem coaxially stacked. Helical junction enumerations are based on (Lescoute et al, 2006).

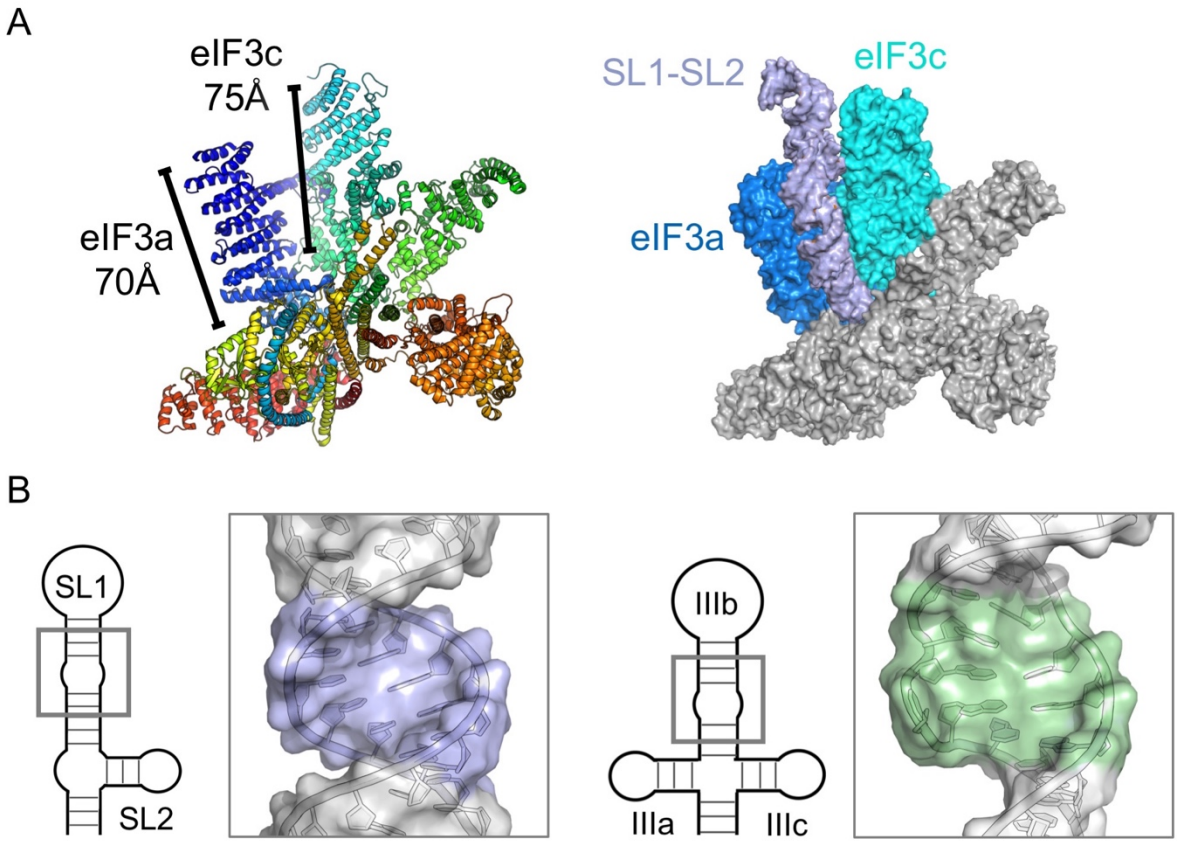
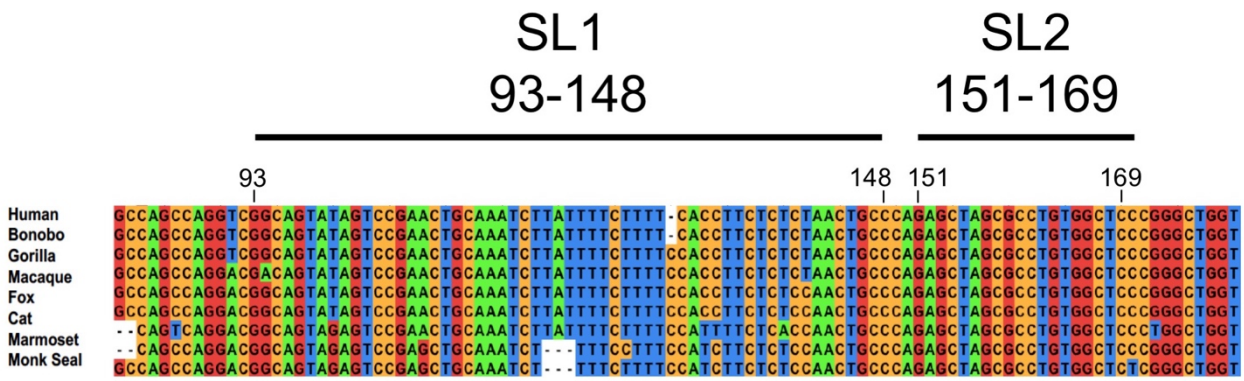
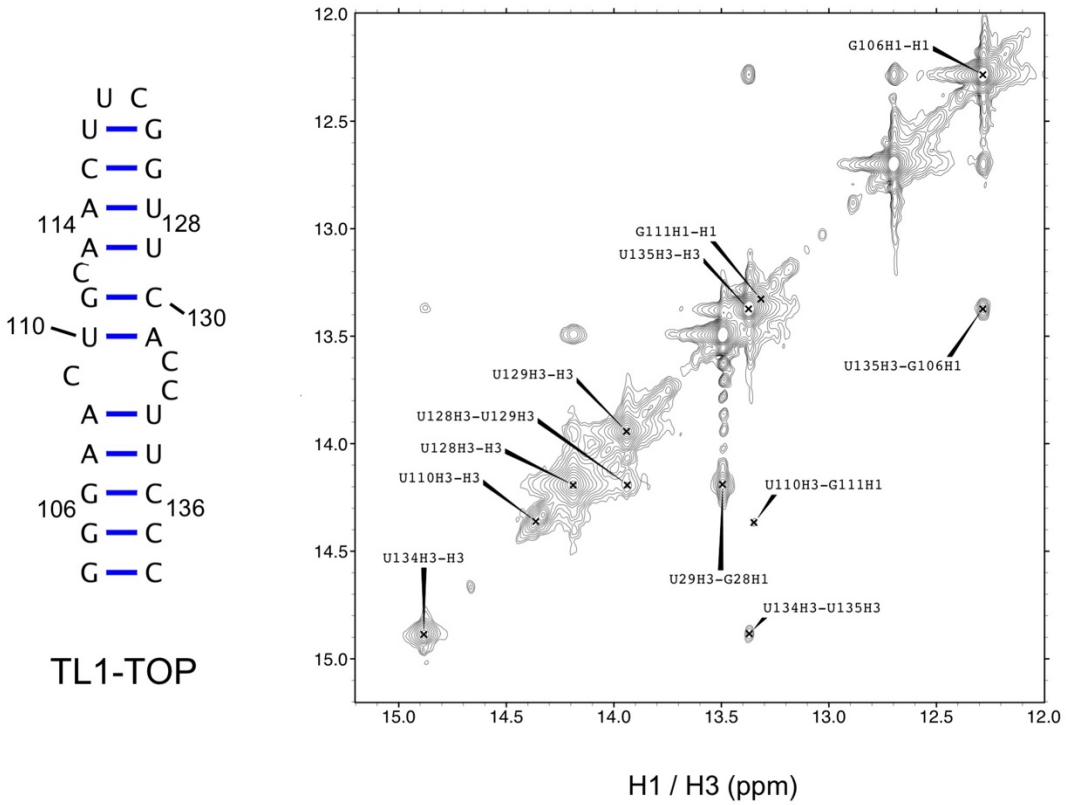


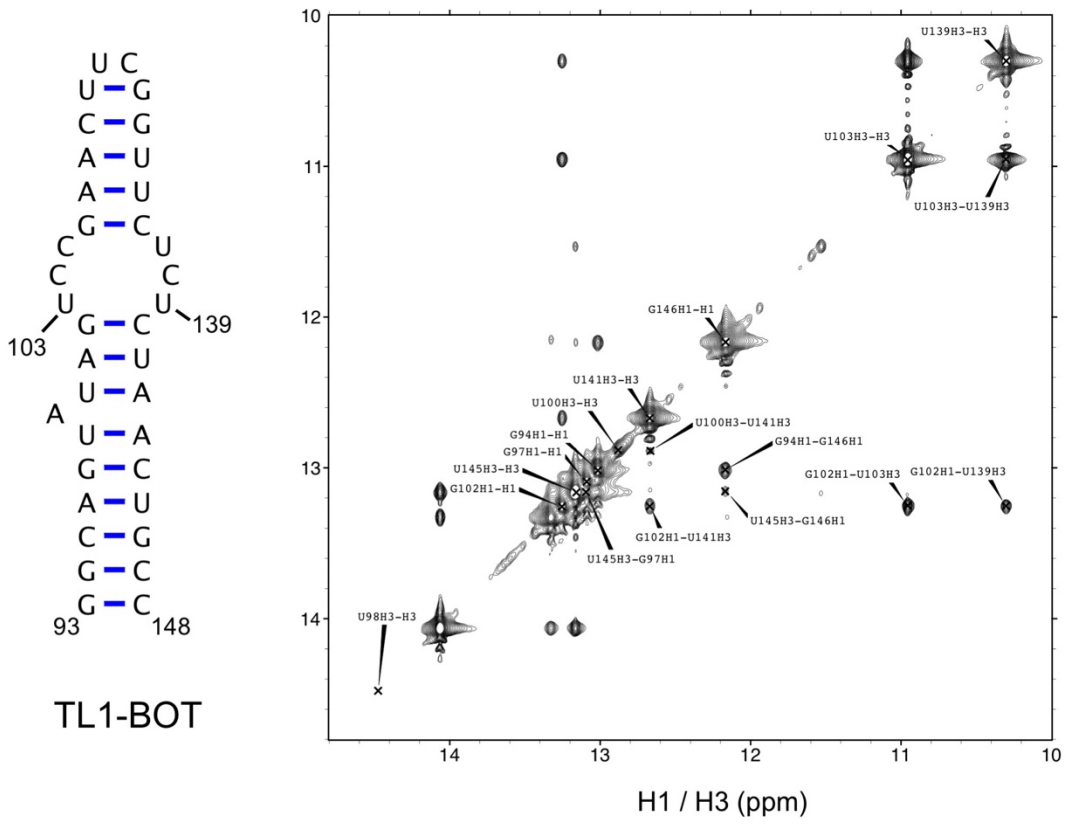
Figure 4.8. (**A**) EIF3 recognizes the HCV IRES IIIabc in a highly basic cleft located between eIF3a (blue) and eIF3c (teal); the dimensions of SL1-SL2 fit within this cleft similarly to previously modeled HCV-like IRES IIIb subdomains (Sun et al, 2013; Hashem et al, 2013; Erzberger et al, 2014). (**B**) Comparison of the internal loops and enlarged major grooves associated with eIF3 recognition: (left) SL1, (right) HCV IRES IIIb. Unpaired nucleotides are in light green, and base-paired nucleotides in light blue.



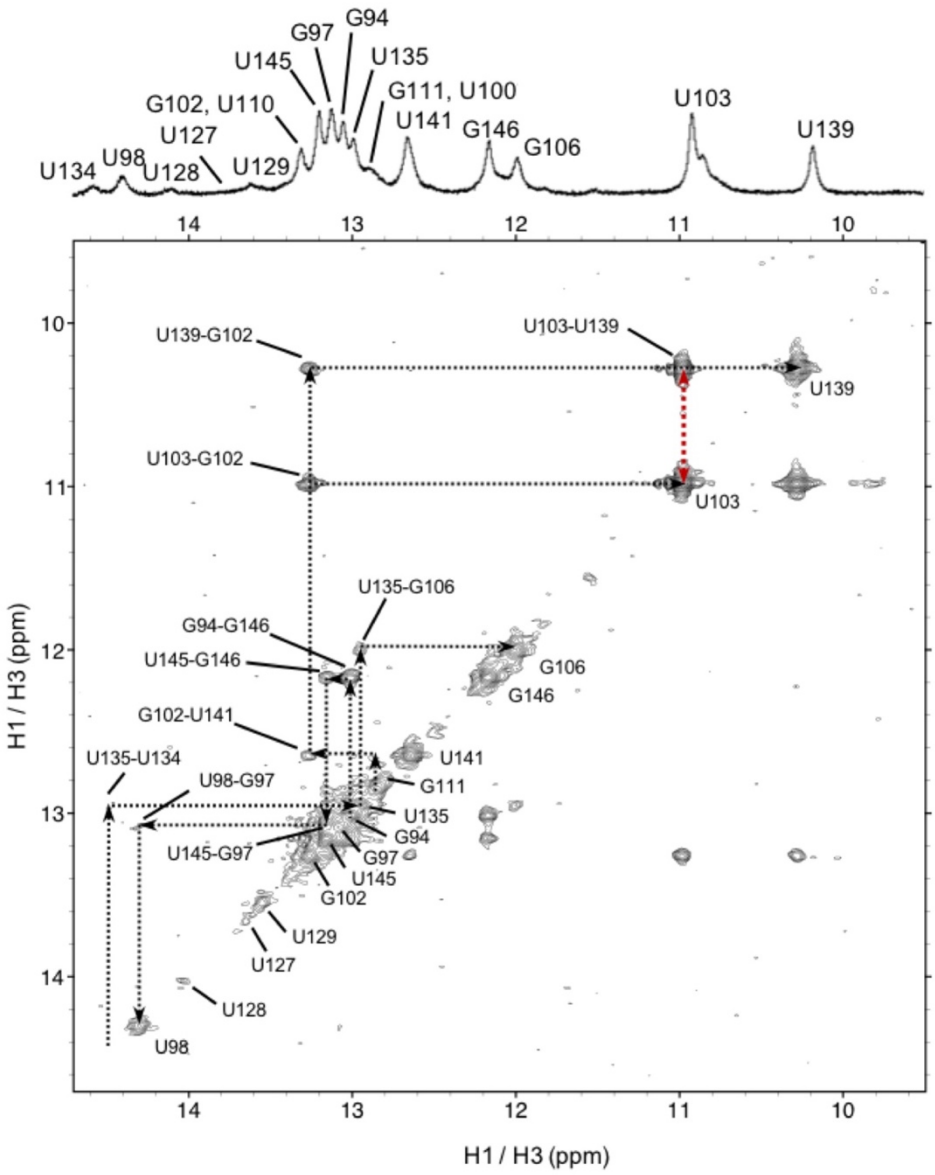
Supplementary Figure S4.1. Clustal Omega multiple sequence alignment for human c-JUN mRNA from NM_002228. Nucleotides 81-179 are displayed and listed based on similarity to the human sequence. Species shown are those with almost identical sequence alignment and highly conserved secondary structure when predicted by the Mfold web server.



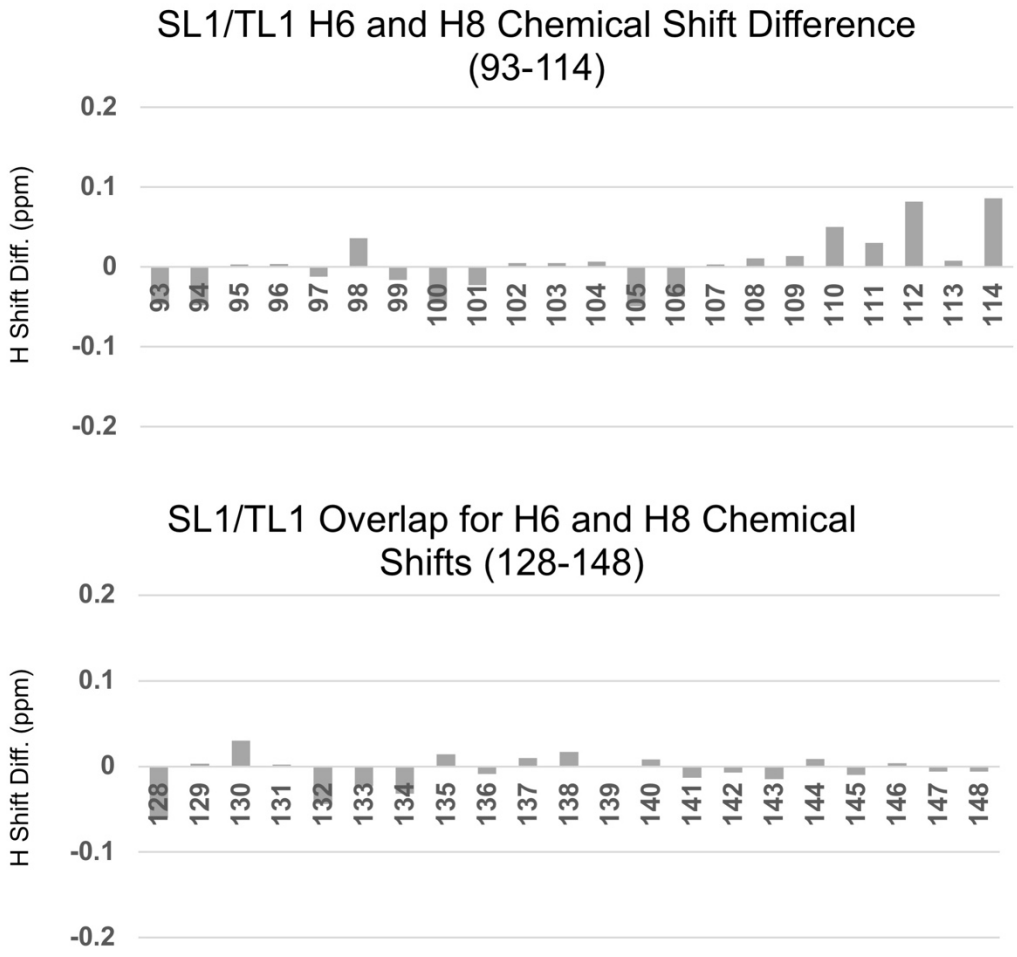
Supplementary Figure S4.2. (left) TL1-TOP sequence and secondary structure. (right) Assigned H_2O NOESY spectrum revealing base-paired imino protons corresponding to G106 to A114, and U128 to C136.



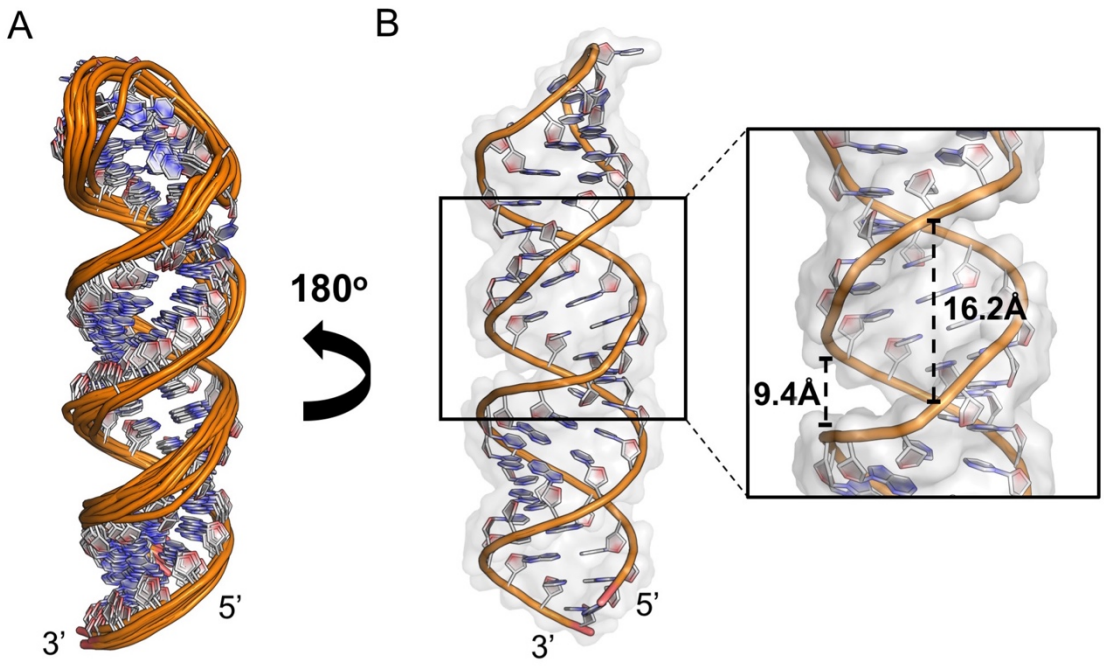
Supplementary Figure S4.3. (left) TL1-BOT sequence and secondary structure. (right) Assigned H_2O NOESY spectrum revealing for base-paired imino protons corresponding to G93 to U103, and U139 to C148.



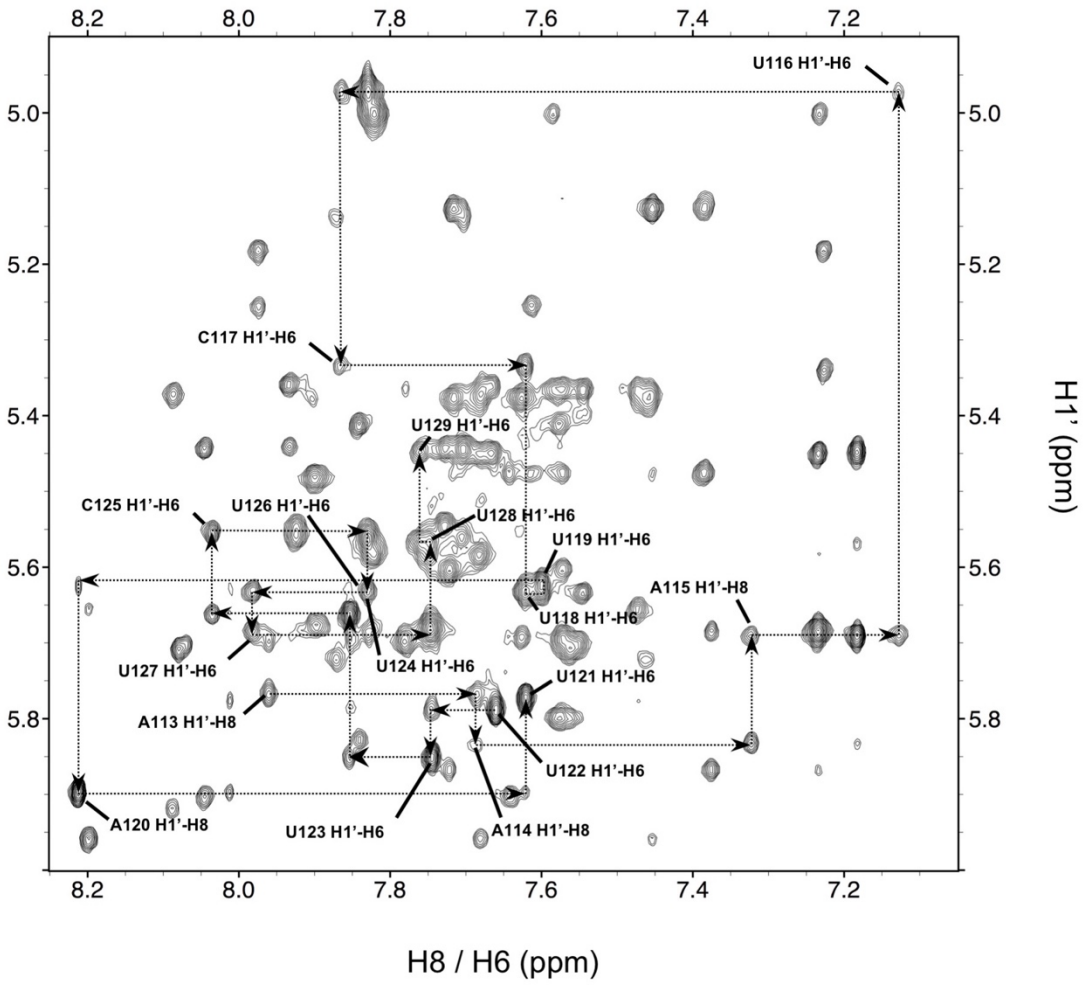
Supplementary Figure S4.4. H_2O NOESY spectrum of SL1 RNA, with black dotted lines showing sequential NOE interactions involving imino protons. The red dotted line represents the strong U103-H3 to U139-H3 NOE supportive of the U103-U139 base-pair.



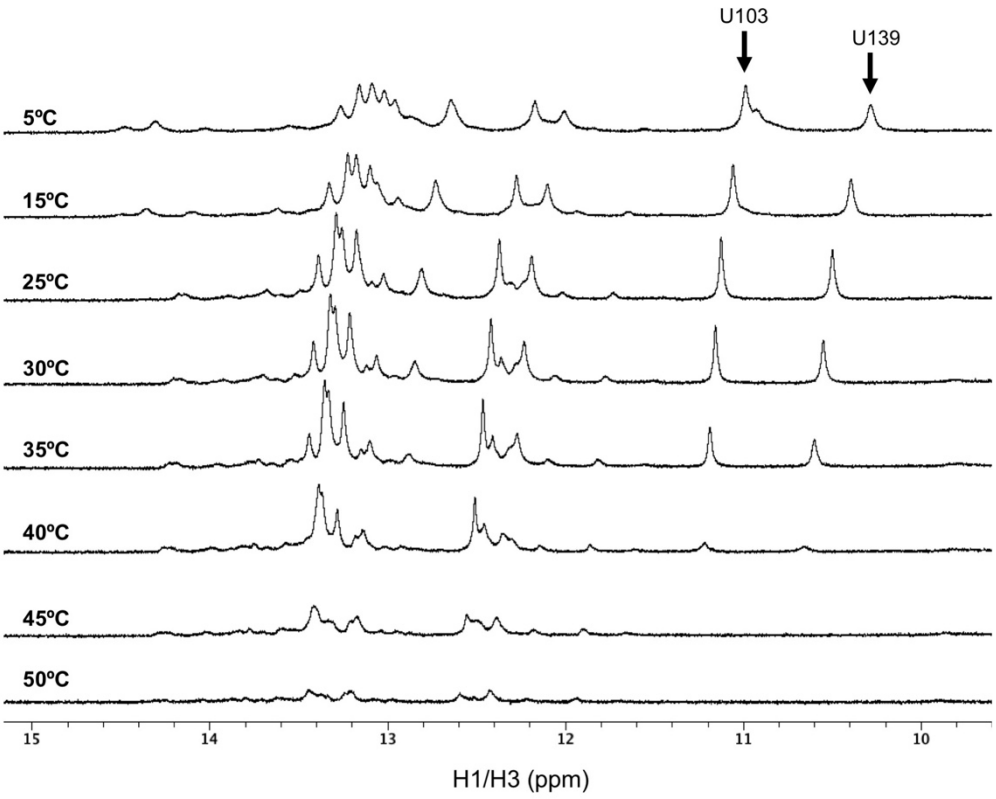
Supplementary Figure S4.5. Comparison of H6 and H8 chemical shifts values for SL1 and TL1 based on comparison of D_2O NOESY spectra.



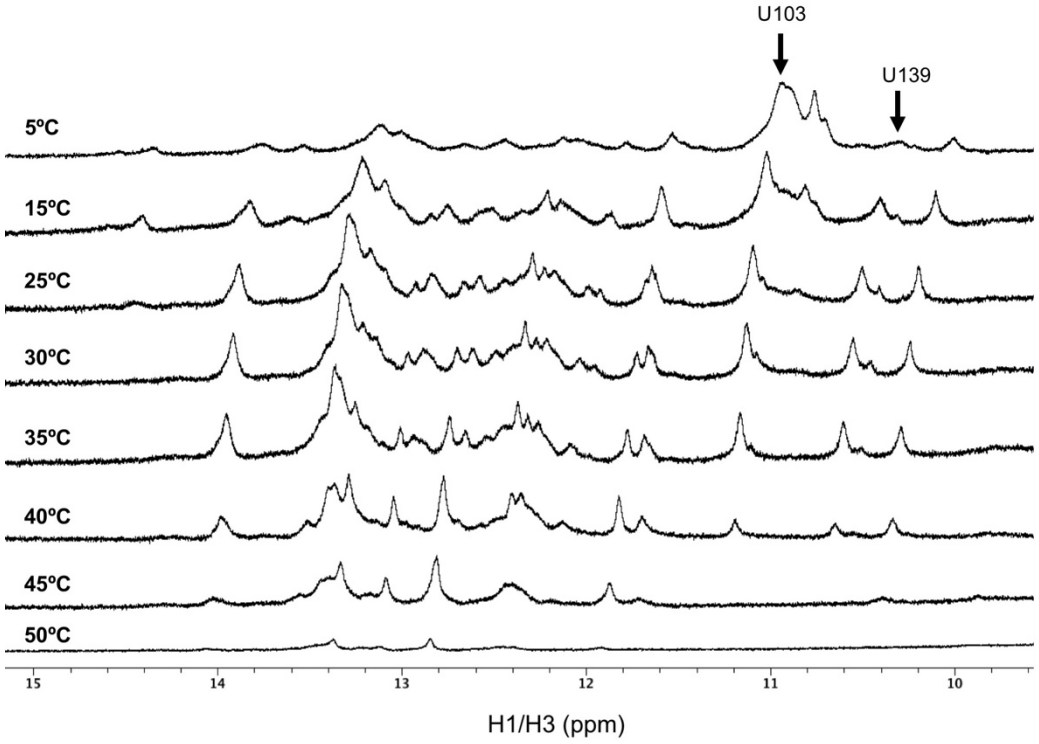
Supplementary Figure S4.6. (**A**) Global superposition of the 10 lowest energy structures from the ensemble of converged structure for the TL1 stem-loop. (**B**) Narrowest and widest major groove widths among converged structures are measured from A101-P to A134-P and C104 to A131, respectively.



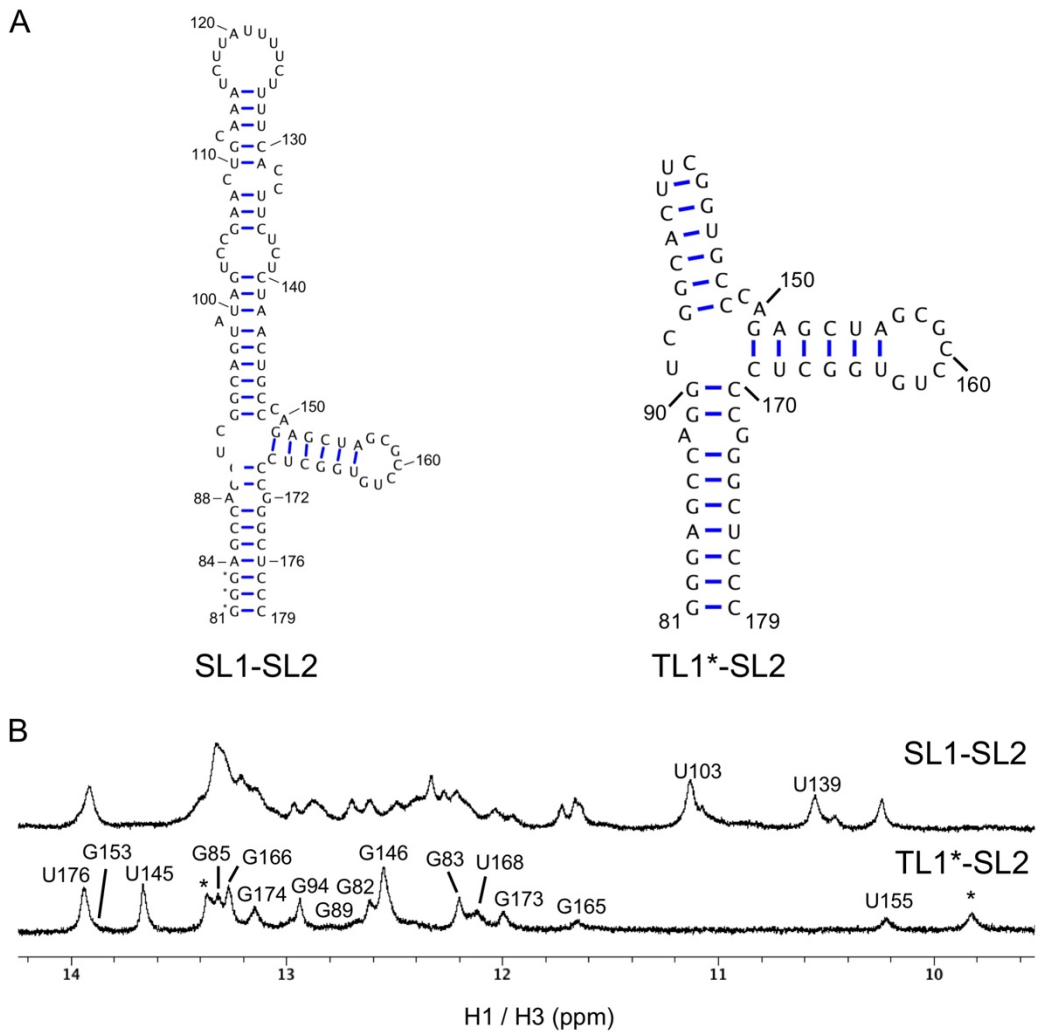
Supplementary Figure S4.7. D_2O NOESY spectrum recorded for SL1 using a highly deuterated sample were deuterium was replaced for protons C5, C2', C3', C4', C5', and C5'' for each nucleotide. Black dotted lines represent sequential H1' to H6/H8 "helical walk" NOE connectivities for apical loop nucleotides A113 to U129.



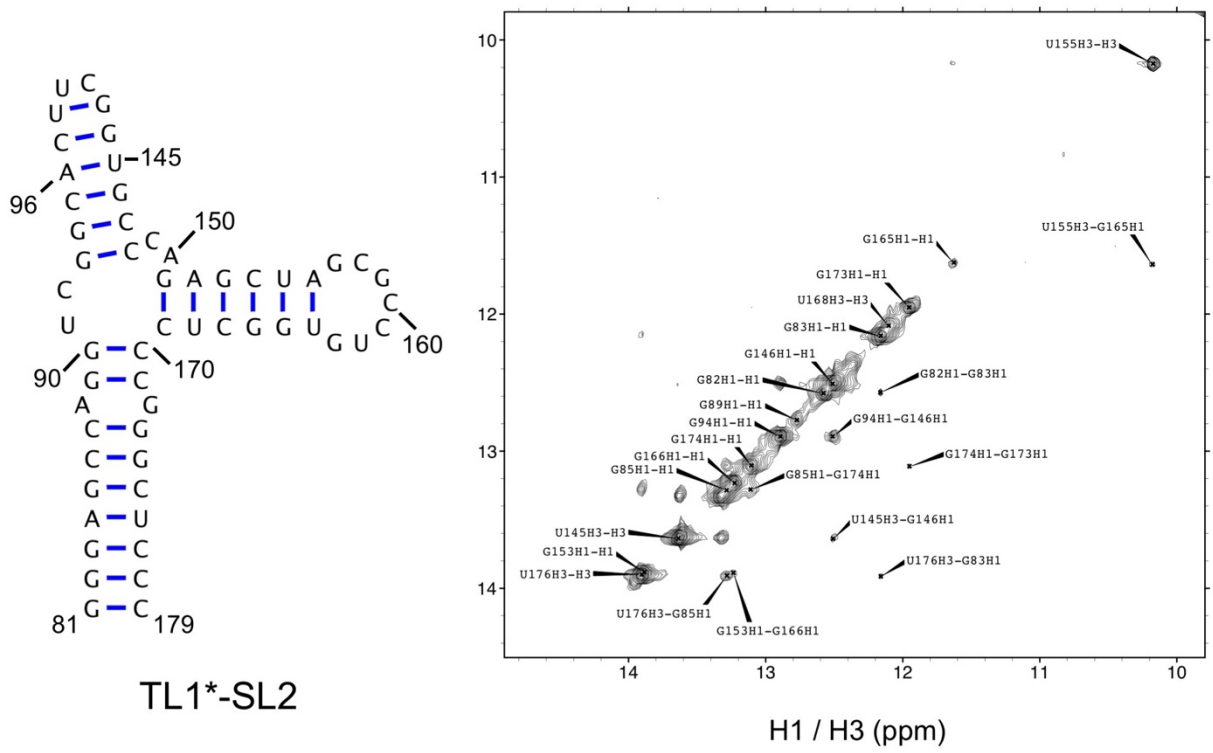
Supplementary Figure S4.8. 1D ^{1}H spectra for SL1 where U103 and U139 are tracked for spectra collected between 5°C and 50°C in 5°C increments. Stable formation of the U103-U139 base-pair is observed until 45°C.



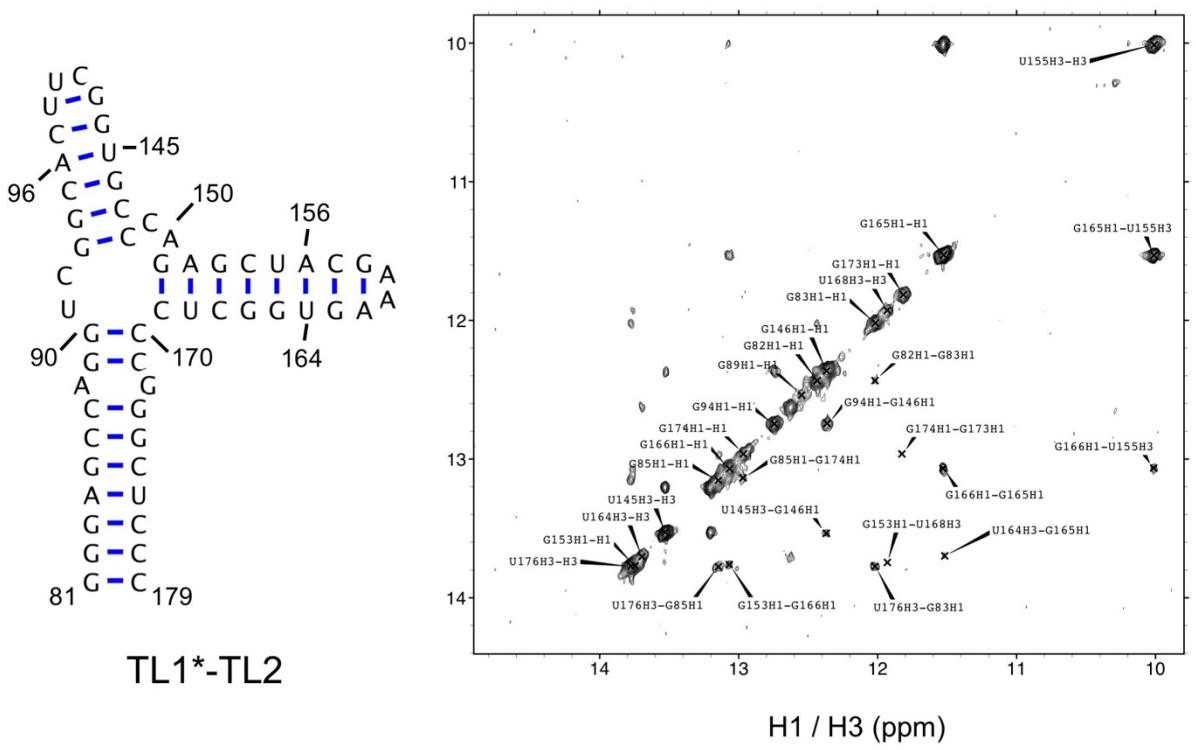
Supplementary Figure S4.9. 1D ^{1}H spectra for SL1-SL2 where U103 and U139 are tracked for spectra collected between 5°C and 50°C in 5°C increments. Stable formation of the U103-U139 base-pair is observed until 45°C.



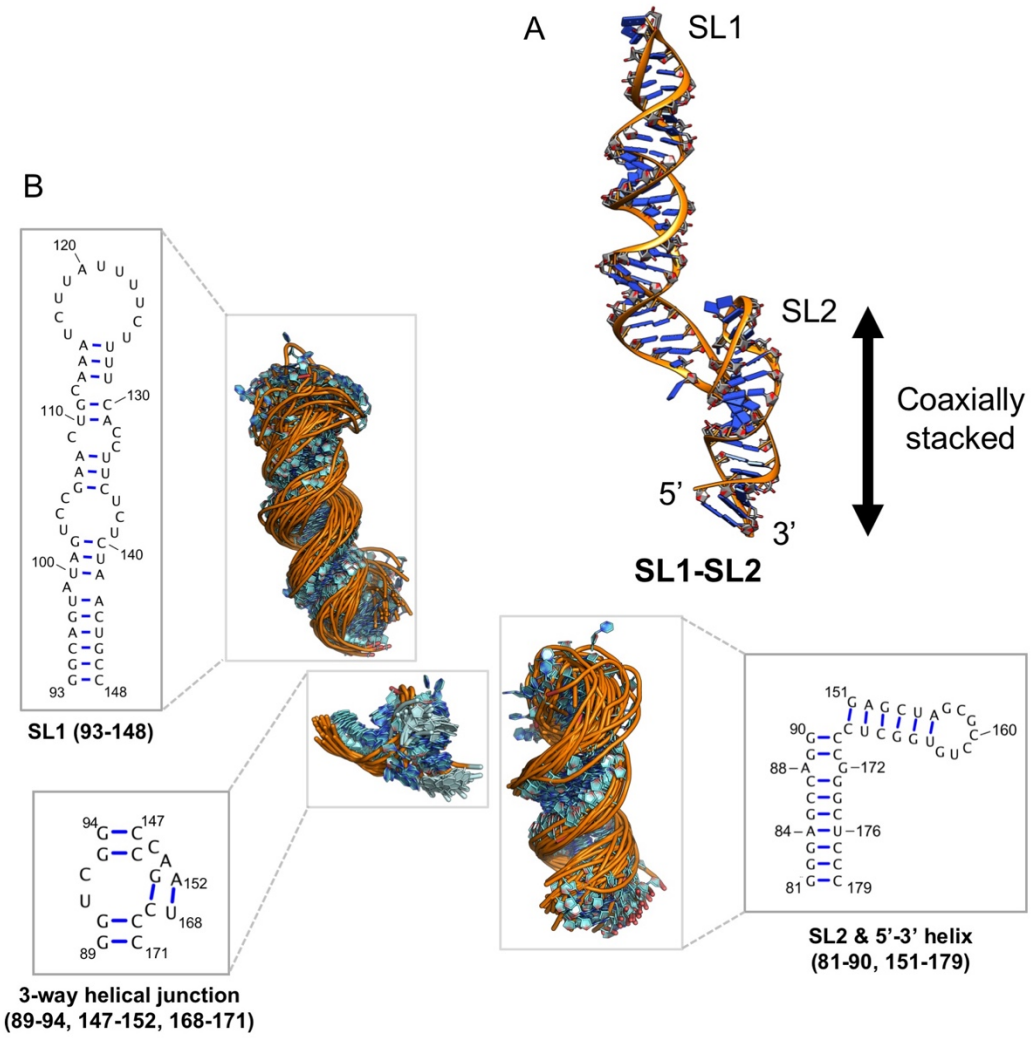
Supplementary Figure S4.10. (**A**) Representation of the secondary structure and sequence of SL1-SL2 and TL1*-SL2 construct. TL1*-SL2 was designed to have SL1 truncated and capped by a tetraloop to facilitate H_2O NOESY peak assignments. (**B**) 1D ^1H spectra for SL1-SL2 compared to the truncated TL1*-SL2. (*) indicates resonances originating in the cUUCGc tetraloop.



Supplementary Figure S4.11. (left) Representation of the TL1*-SL2 sequence and secondary structure, where SL1 is truncated at the A96-U145 base-pair and capped with a cUUCGg tetraloop to facilitate resonance assignments. (right) Assigned H_2O NOESY spectrum for the same RNA.



Supplementary Figure S4.12. (left) Representation of the TL1*-TL2 sequence and secondary structure, where SL1 is truncated at the A96-U145 base-pair and capped with a cUUCGg tetraloop to facilitate resonance assignments. SL2 is capped with a cGAAAg tetraloop facilitate resonance assignments. (right) Assigned H_2O NOESY spectrum for the same RNA.



Supplementary Figure S4.13. (**A**) Lowest energy model calculated from simRNA (Boniecki, 2016) and refined by Rosetta RNA De Novo energy minimization (Das et al, 2010) for SL1-SL2. (**B**) 24 simRNA-generated models sectionally aligned to the Rosetta RNA De Novo energy minimization model. Secondary structures represented for each individual section: SL1 (93-148; RMSD = 4.07 Å), SL2 & 5'-3' helix (81-90, 151-179; RMSD = 4.22 Å), 3-way helical junction (89-94, 147-152, 168-171; RMSD = 2.56 Å).

APPENDIX A. RESONANCE ASSIGNMENTS

A1. Resonance Assignments for c-JUN TL1 G93-C148

```
#####################################################################
#      Chemical Shift Ambiguity Index Value Definitions
#
# The values other than 1 are used for those atoms with different
# chemical shifts that cannot be assigned to stereospecific atoms
# or to specific residues or chains.
#
#  Index Value          Definition
#
#     1          Unique (including isolated methyl protons,
#                   geminal atoms, and geminal methyl
#                   groups with identical chemical shifts)
#                   (e.g. ILE HD11, HD12, HD13 protons)
#     2          Ambiguity of geminal atoms or geminal methyl
#                   proton groups (e.g. ASP HB2 and HB3
#                   protons, LEU CD1 and CD2 carbons, or
#                   LEU HD11, HD12, HD13 and HD21, HD22,
#                   HD23 methyl protons)
#     3          Aromatic atoms on opposite sides of
#                   symmetrical rings (e.g. TYR HE1 and HE2
#                   protons)
#     4          Intraresidue ambiguities (e.g. LYS HG and
#                   HD protons or TRP HZ2 and HZ3 protons)
#     5          Interresidue ambiguities (LYS 12 vs. LYS 27)
#     6          Intermolecular ambiguities (e.g. ASP 31 CA
#                   in monomer 1 and ASP 31 CA in monomer 2
#                   of an asymmetrical homodimer, duplex
#                   DNA assignments, or other assignments
#                   that may apply to atoms in one or more
#                   molecule in the molecular assembly)
#     9          Ambiguous, specific ambiguity not defined
#
#####################################################################
loop_
   _Atom_chem_shift.ID
   _Atom_chem_shift.Comp_index_ID
   _Atom_chem_shift.Comp_ID
   _Atom_chem_shift.Atom_ID
   _Atom_chem_shift.Atom_type
   _Atom_chem_shift.Atom_isotope_number
   _Atom_chem_shift.Val

1          1          C          H41        H          1          8.240
2          1          C          H42        H          1          6.386
3          1          C          H1'        H          1          5.317
4          1          C          H2'        H          1          4.322
5          1          C          H3'        H          1          4.069
6          1          C          H4'        H          1          4.373
7          1          C          H5         H          1          5.062
8          1          C          H5'        H          1          4.367
9          1          C          H5"        H          1          3.959
10         1          C          H6         H          1          7.145
11         1          C          C1'        C          13         93.758
12         1          C          C2'        C          13         75.575
13         1          C          C3'        C          13         72.176
14         1          C          C4'        C          13         79.606
15         1          C          C5         C          13         97.220
16         1          C          C5'        C          13         65.138
17         1          C          C6         C          13         140.285
18         2          U          H3         H          1          11.549
```

19	2	U	H1'	H	1	5.604
20	2	U	H2'	H	1	3.732
21	2	U	H3'	H	1	4.476
22	2	U	H4'	H	1	4.323
23	2	U	H5	H	1	5.581
24	2	U	H5'	H	1	4.403
25	2	U	H5"	H	1	4.059
26	2	U	H6	H	1	7.702
27	2	U	C1'	C	13	94.343
28	2	U	C2'	C	13	75.961
29	2	U	C3'	C	13	73.135
30	2	U	C4'	C	13	82.313
31	2	U	C5	C	13	104.959
32	2	U	C5'	C	13	65.435
33	2	U	C6	C	13	140.698
34	2	U	N3	N	15	159.627
35	3	U	H1'	H	1	6.054
36	3	U	H2'	H	1	4.616
37	3	U	H3'	H	1	3.976
38	3	U	H4'	H	1	4.431
39	3	U	H5	H	1	5.815
40	3	U	H5'	H	1	4.185
41	3	U	H5"	H	1	4.056
42	3	U	H6	H	1	7.974
43	3	U	C1'	C	13	89.308
44	3	U	C2'	C	13	74.500
45	3	U	C3'	C	13	77.749
46	3	U	C4'	C	13	87.020
47	3	U	C5	C	13	105.430
48	3	U	C5'	C	13	67.780
49	3	U	C6	C	13	144.615
50	4	C	H41	H	1	8.048
51	4	C	H42	H	1	6.431
52	4	C	H1'	H	1	5.908
53	4	C	H2'	H	1	4.041
54	4	C	H3'	H	1	4.436
55	4	C	H4'	H	1	3.751
56	4	C	H5	H	1	6.074
57	4	C	H5'	H	1	3.563
58	4	C	H5"	H	1	2.671
59	4	C	H6	H	1	7.650
60	4	C	C1'	C	13	89.114
61	4	C	C2'	C	13	77.689
62	4	C	C3'	C	13	80.300
63	4	C	C4'	C	13	84.507
64	4	C	C5	C	13	98.455
65	4	C	C5'	C	13	67.300
66	4	C	C6	C	13	142.827
67	5	G	H1	H	1	9.616
68	5	G	H1'	H	1	5.917
69	5	G	H2'	H	1	4.775
70	5	G	H3'	H	1	5.589
71	5	G	H4'	H	1	4.365
72	5	G	H5'	H	1	4.382
73	5	G	H5"	H	1	4.136
74	5	G	H8	H	1	7.815
75	5	G	C1'	C	13	94.443
76	5	G	C2'	C	13	76.853
77	5	G	C3'	C	13	75.920
78	5	G	C4'	C	13	83.200
79	5	G	C5'	C	13	69.050
80	5	G	C8	C	13	142.866
81	6	G	H1	H	1	13.225
82	6	G	H21	H	1	8.543
83	6	G	H22	H	1	5.355
84	6	G	H1'	H	1	4.471
85	6	G	H2'	H	1	4.383
86	6	G	H3'	H	1	4.182
87	6	G	H4'	H	1	4.342
88	6	G	H5'	H	1	4.436

89	6	G	H5"	H	1	4.219
90	6	G	H8	H	1	8.276
91	6	G	C1'	C	13	93.646
92	6	G	C2'	C	13	75.239
93	6	G	C3'	C	13	74.609
94	6	G	C4'	C	13	81.011
95	6	G	C5'	C	13	69.678
96	6	G	C8	C	13	139.011
97	93	G	H1'	H	1	5.771
98	93	G	H2'	H	1	4.897
99	93	G	H3'	H	1	4.672
100	93	G	H4'	H	1	4.514
101	93	G	H5'	H	1	4.303
102	93	G	H5"	H	1	4.233
103	93	G	H8	H	1	8.094
104	93	G	C1'	C	13	91.473
105	93	G	C2'	C	13	75.022
106	93	G	C3'	C	13	75.258
107	93	G	C4'	C	13	82.457
108	93	G	C5'	C	13	67.201
109	93	G	C8	C	13	139.067
110	94	G	H1	H	1	13.038
111	94	G	H21	H	1	6.753
112	94	G	H22	H	1	5.430
113	94	G	H1'	H	1	5.859
114	94	G	H2'	H	1	4.509
115	94	G	H3'	H	1	4.464
116	94	G	H4'	H	1	4.455
117	94	G	H5'	H	1	4.245
118	94	G	H5"	H	1	4.164
119	94	G	H8	H	1	7.581
120	94	G	C1'	C	13	92.943
121	94	G	C2'	C	13	75.200
122	94	G	C3'	C	13	75.234
123	94	G	C4'	C	13	81.697
124	94	G	C5'	C	13	66.079
125	94	G	C8	C	13	136.660
126	94	G	N1	N	15	148.048
127	95	C	H41	H	1	8.138
128	95	C	H42	H	1	6.599
129	95	C	H1'	H	1	5.471
130	95	C	H2'	H	1	4.535
131	95	C	H3'	H	1	4.485
132	95	C	H4'	H	1	4.302
133	95	C	H5	H	1	5.213
134	95	C	H5'	H	1	4.264
135	95	C	H5"	H	1	4.073
136	95	C	H6	H	1	7.616
137	95	C	C1'	C	13	92.323
138	95	C	C2'	C	13	75.195
139	95	C	C3'	C	13	75.219
140	95	C	C4'	C	13	82.352
141	95	C	C5	C	13	97.236
142	95	C	C5'	C	13	67.408
143	95	C	C6	C	13	140.618
144	96	A	H61	H	1	7.466
145	96	A	H62	H	1	6.046
146	96	A	H1'	H	1	5.891
147	96	A	H2	H	1	6.984
148	96	A	H2'	H	1	4.585
149	96	A	H3'	H	1	4.598
150	96	A	H4'	H	1	4.351
151	96	A	H5'	H	1	4.453
152	96	A	H5"	H	1	4.110
153	96	A	H8	H	1	7.902
154	96	A	C1'	C	13	92.947
155	96	A	C2	C	13	152.622
156	96	A	C2'	C	13	75.658
157	96	A	C3'	C	13	74.796
158	96	A	C4'	C	13	82.724

159	96	A	C5'	C	13	65.589
160	96	A	C8	C	13	139.354
161	97	G	H1	H	1	13.120
162	97	G	H21	H	1	7.975
163	97	G	H22	H	1	5.732
164	97	G	H1'	H	1	5.416
165	97	G	H2'	H	1	4.372
166	97	G	H3'	H	1	4.301
167	97	G	H4'	H	1	4.384
168	97	G	H5'	H	1	4.397
169	97	G	H5"	H	1	3.986
170	97	G	H8	H	1	7.044
171	97	G	C1'	C	13	92.993
172	97	G	C2'	C	13	75.228
173	97	G	C3'	C	13	73.539
174	97	G	C4'	C	13	83.939
175	97	G	C5'	C	13	66.327
176	97	G	C8	C	13	135.931
177	97	G	N1	N	15	147.679
178	98	U	H1'	H	1	5.688
179	98	U	H2'	H	1	4.242
180	98	U	H3'	H	1	4.528
181	98	U	H4'	H	1	4.407
182	98	U	H5	H	1	5.053
183	98	U	H5'	H	1	4.062
184	98	U	H6	H	1	7.551
185	98	U	C1'	C	13	92.881
186	98	U	C2'	C	13	76.016
187	98	U	C3'	C	13	73.006
188	98	U	C4'	C	13	84.436
189	98	U	C5	C	13	103.476
190	98	U	C5'	C	13	65.846
191	98	U	C6	C	13	142.194
192	99	A	H1'	H	1	5.961
193	99	A	H2	H	1	7.603
194	99	A	H2'	H	1	4.646
195	99	A	H3'	H	1	4.570
196	99	A	H4'	H	1	4.386
197	99	A	H5'	H	1	4.122
198	99	A	H5"	H	1	4.310
199	99	A	H8	H	1	8.229
200	99	A	C1'	C	13	90.948
201	99	A	C2'	C	13	76.006
202	99	A	C3'	C	13	76.049
203	99	A	C4'	C	13	78.283
204	99	A	C5'	C	13	66.904
205	99	A	C8	C	13	141.589
206	100	U	H1'	H	1	5.461
207	100	U	H2'	H	1	4.418
208	100	U	H3'	H	1	4.484
209	100	U	H4'	H	1	4.338
210	100	U	H5	H	1	5.381
211	100	U	H5'	H	1	4.120
212	100	U	H5"	H	1	4.349
213	100	U	H6	H	1	7.504
214	100	U	C1'	C	13	93.452
215	100	U	C2'	C	13	75.192
216	100	U	C3'	C	13	73.156
217	100	U	C4'	C	13	82.358
218	100	U	C5	C	13	103.954
219	100	U	C5'	C	13	66.514
220	100	U	C6	C	13	141.931
221	101	A	H61	H	1	7.397
222	101	A	H62	H	1	6.091
223	101	A	H1'	H	1	5.987
224	101	A	H2	H	1	7.014
225	101	A	H2'	H	1	4.613
226	101	A	H3'	H	1	4.601
227	101	A	H4'	H	1	4.349
228	101	A	H5'	H	1	4.215

229	101	A	H8	H	1	8.133
230	101	A	C1'	C	13	92.179
231	101	A	C2	C	13	152.558
232	101	A	C2'	C	13	75.148
233	101	A	C3'	C	13	73.501
234	101	A	C4'	C	13	81.578
235	101	A	C5'	C	13	67.302
236	101	A	C8	C	13	140.253
237	102	G	H1	H	1	13.268
238	102	G	H21	H	1	8.529
239	102	G	H22	H	1	5.923
240	102	G	H1'	H	1	5.409
241	102	G	H2'	H	1	4.423
242	102	G	H3'	H	1	3.953
243	102	G	H4'	H	1	4.274
244	102	G	H5'	H	1	4.302
245	102	G	H5"	H	1	3.977
246	102	G	H8	H	1	6.924
247	102	G	C1'	C	13	92.669
248	102	G	C2'	C	13	76.210
249	102	G	C3'	C	13	73.284
250	102	G	C4'	C	13	84.685
251	102	G	C5'	C	13	67.009
252	102	G	C8	C	13	135.280
253	102	G	N1	N	15	147.854
254	103	U	H3	H	1	10.967
255	103	U	H1'	H	1	5.254
256	103	U	H2'	H	1	4.305
257	103	U	H3'	H	1	4.242
258	103	U	H4'	H	1	4.282
259	103	U	H5	H	1	5.064
260	103	U	H5'	H	1	4.367
261	103	U	H5"	H	1	4.344
262	103	U	H6	H	1	7.305
263	103	U	C1'	C	13	93.939
264	103	U	C2'	C	13	72.893
265	103	U	C3'	C	13	71.870
266	103	U	C4'	C	13	82.391
267	103	U	C5	C	13	103.629
268	103	U	C5'	C	13	63.199
269	103	U	C6	C	13	140.119
270	103	U	N3	N	15	158.071
271	104	C	H41	H	1	8.327
272	104	C	H42	H	1	7.314
273	104	C	H1'	H	1	5.335
274	104	C	H2'	H	1	3.999
275	104	C	H3'	H	1	4.397
276	104	C	H4'	H	1	4.359
277	104	C	H5	H	1	5.751
278	104	C	H5"	H	1	3.972
279	104	C	H6	H	1	8.038
280	104	C	C1'	C	13	94.368
281	104	C	C2'	C	13	75.418
282	104	C	C3'	C	13	72.197
283	104	C	C4'	C	13	83.369
284	104	C	C5	.	13	96.203
285	104	C	C5'	C	13	67.195
286	104	C	C6	C	13	142.607
287	105	C	H1'	H	1	5.536
288	105	C	H2'	H	1	4.146
289	105	C	H3'	H	1	4.206
290	105	C	H4'	H	1	4.220
291	105	C	H5	H	1	5.443
292	105	C	H5'	H	1	4.285
293	105	C	H5"	H	1	4.354
294	105	C	H6	H	1	7.651
295	105	C	C1'	C	13	92.784
296	105	C	C2'	C	13	75.799
297	105	C	C3'	C	13	75.996
298	105	C	C4'	C	13	81.414

299	105	C	C5	C	13	97.037
300	105	C	C6	C	13	142.001
301	106	G	H1	H	1	12.063
302	106	G	H21	H	1	5.856
303	106	G	H1'	H	1	5.635
304	106	G	H2'	H	1	4.648
305	106	G	H3'	H	1	4.507
306	106	G	H4'	H	1	4.430
307	106	G	H5'	H	1	4.001
308	106	G	H8	H	1	7.613
309	106	G	C1'	C	13	90.824
310	106	G	C2'	C	13	75.329
311	106	G	C3'	C	13	72.683
312	106	G	C4'	C	13	83.388
313	106	G	C5'	C	13	64.532
314	106	G	C8	C	13	137.302
315	106	G	N1	N	15	147.144
316	107	A	H61	H	1	7.692
317	107	A	H62	H	1	6.224
318	107	A	H1'	H	1	5.941
319	107	A	H2	H	1	7.307
320	107	A	H2'	H	1	4.749
321	107	A	H3'	H	1	4.531
322	107	A	H4'	H	1	4.429
323	107	A	H5'	H	1	4.058
324	107	A	H5"	H	1	4.483
325	107	A	H8	H	1	7.813
326	107	A	C1'	C	13	93.000
327	107	A	C2	C	13	153.492
328	107	A	C2'	C	13	75.288
329	107	A	C3'	C	13	72.405
330	107	A	C4'	C	13	82.193
331	107	A	C5'	C	13	64.788
332	107	A	C8	C	13	139.822
333	108	A	H62	H	1	6.546
334	108	A	H1'	H	1	5.745
335	108	A	H2	H	1	7.908
336	108	A	H2'	H	1	4.546
337	108	A	H3'	H	1	4.276
338	108	A	H4'	H	1	4.298
339	108	A	H5'	H	1	4.053
340	108	A	H5"	H	1	4.427
341	108	A	H8	H	1	7.450
342	108	A	C1'	C	13	93.068
343	108	A	C2	C	13	153.540
344	108	A	C2'	C	13	75.543
345	108	A	C3'	C	13	75.872
346	108	A	C4'	C	13	83.377
347	108	A	C5'	C	13	66.086
348	108	A	C8	C	13	138.665
349	109	C	H1'	H	1	5.067
350	109	C	H2'	H	1	4.054
351	109	C	H3'	H	1	4.234
352	109	C	H4'	H	1	4.293
353	109	C	H5'	H	1	4.391
354	109	C	H5	H	1	5.180
355	109	C	H6	H	1	7.314
356	109	C	C1'	C	13	93.981
357	109	C	C2'	C	13	75.169
358	109	C	C3'	C	13	73.168
359	109	C	C4'	C	13	81.732
360	109	C	C5	.	13	97.098
361	109	C	C5'	C	1	63.630
362	109	C	C6	C	13	140.961
363	110	U	H1'	H	1	5.436
364	110	U	H2'	H	1	4.153
365	110	U	H3'	H	1	4.218
366	110	U	H4'	H	1	4.126
367	110	U	H5'	H	1	3.964
368	110	U	H5	H	1	5.346

369	110	U	H6	H	1	7.649
370	110	U	C1'	C	13	92.754
371	110	U	C2'	C	13	75.547
372	110	U	C3'	C	13	75.185
373	110	U	C4'	C	13	81.977
374	110	U	C5	C	13	103.228
375	110	U	C5'	C	13	64.275
376	110	U	C6	C	13	141.713
377	111	G	H1'	H	1	5.770
378	111	G	H2'	H	1	4.775
379	111	G	H3'	H	1	4.617
380	111	G	H4'	H	1	4.476
381	111	G	H5'	H	1	4.198
382	111	G	H5"	H	1	4.336
383	111	G	H8	H	1	7.902
384	111	G	C1'	C	13	88.687
385	111	G	C2'	C	13	75.284
386	111	G	C3'	C	13	76.799
387	111	G	C4'	C	13	82.378
388	111	G	C5'	C	13	66.898
389	111	G	C8	C	13	138.632
390	111	G	N1	N	15	148.087
391	112	C	H1'	H	1	5.603
392	112	C	H2'	H	1	4.410
393	112	C	H3'	H	1	4.478
394	112	C	H4'	H	1	4.343
395	112	C	H5	H	1	5.988
396	112	C	H5'	H	1	4.193
397	112	C	H6	H	1	7.933
398	112	C	C1'	C	13	90.874
399	112	C	C2'	C	13	75.945
400	112	C	C3'	C	13	73.724
401	112	C	C4'	C	13	82.080
402	112	C	C5	C	13	98.626
403	112	C	C5'	C	13	66.874
404	112	C	C6	C	13	144.038
405	113	A	H1'	H	1	5.833
406	113	A	H2	H	1	7.016
407	113	A	H2'	H	1	4.526
408	113	A	H3'	H	1	4.363
409	113	A	H4'	H	1	4.588
410	113	A	H5'	H	1	4.168
411	113	A	H8	H	1	8.052
412	113	A	C1'	C	13	92.198
413	113	A	C2	C	13	152.596
414	113	A	C2'	C	13	75.252
415	113	A	C3'	C	13	74.581
416	113	A	C4'	C	13	82.007
417	113	A	C5'	C	13	67.194
418	113	A	C8	C	13	140.751
419	114	A	H61	H	1	8.034
420	114	A	H62	H	1	6.268
421	114	A	H1'	H	1	5.967
422	114	A	H2	H	1	7.776
423	114	A	H2'	H	1	4.536
424	114	A	H3'	H	1	4.482
425	114	A	H4'	H	1	4.478
426	114	A	H5'	H	1	4.371
427	114	A	H8	H	1	7.809
428	114	A	C1'	C	13	92.739
429	114	A	C2	C	13	154.737
430	114	A	C2'	C	13	74.546
431	114	A	C3'	C	13	72.387
432	114	A	C4'	C	13	82.098
433	114	A	C5'	C	13	67.088
434	114	A	C8	C	13	139.638
435	128	U	H3	H	1	13.987
436	128	U	H1'	H	1	5.527
437	128	U	H2'	H	1	4.534
438	128	U	H3'	H	1	4.401

439	128	U	H4'	H	1	4.355
440	128	U	H5	H	1	5.058
441	128	U	H5'	H	1	4.009
442	128	U	H5"	H	1	3.986
443	128	U	H6	H	1	7.733
444	128	U	C1'	C	13	92.763
445	128	U	C2'	C	13	72.380
446	128	U	C3'	C	13	72.150
447	128	U	C4'	C	13	82.082
448	128	U	C5	C	13	102.310
449	128	U	C5'	C	13	64.307
450	128	U	C6	C	13	141.858
451	128	U	N3	N	15	162.293
452	129	U	H3	H	1	13.632
453	129	U	H1'	H	1	5.537
454	129	U	H2'	H	1	4.453
455	129	U	H3'	H	1	4.426
456	129	U	H4'	H	1	4.370
457	129	U	H5	H	1	5.541
458	129	U	H5'	H	1	4.054
459	129	U	H6	H	1	7.814
460	129	U	C1'	C	13	93.879
461	129	U	C2'	C	13	75.277
462	129	U	C3'	C	13	72.510
463	129	U	C4'	C	13	82.120
464	129	U	C5	C	13	103.644
465	129	U	C5'	C	13	64.849
466	129	U	C6	C	13	142.960
467	129	U	N3	N	15	162.777
468	130	C	H41	H	1	7.335
469	130	C	H1'	H	1	5.216
470	130	C	H2'	H	1	4.271
471	130	C	H3'	H	1	4.331
472	130	C	H4'	H	1	4.602
473	130	C	H5'	H	1	4.071
474	130	C	H5	H	1	5.660
475	130	C	H6	H	1	7.778
476	130	C	C1'	C	13	93.802
477	130	C	C2'	C	13	75.306
478	130	C	C3'	C	13	72.651
479	130	C	C4'	C	13	81.064
480	130	C	C5	C	13	98.038
481	130	C	C5'	C	13	65.068
482	130	C	C6	C	13	141.056
483	131	A	H1'	H	1	5.753
484	131	A	H2	H	1	7.271
485	131	A	H2'	H	1	4.391
486	131	A	H3'	H	1	4.463
487	131	A	H4'	H	1	4.424
488	131	A	H5'	H	1	4.301
489	131	A	H5"	H	1	4.043
490	131	A	H8	H	1	7.908
491	131	A	C1'	C	13	92.255
492	131	A	C2	C	13	153.595
493	131	A	C2'	C	13	75.264
494	131	A	C3'	C	13	72.446
495	131	A	C4'	C	13	82.775
496	131	A	C5'	C	13	65.382
497	131	A	C8	C	13	139.981
498	132	C	H1'	H	1	5.374
499	132	C	H2'	H	1	4.122
500	132	C	H3'	H	1	4.405
501	132	C	H4'	H	1	4.259
502	132	C	H5'	H	1	4.028
503	132	C	H5	H	1	5.477
504	132	C	H6	H	1	7.448
505	132	C	C1'	C	13	92.749
506	132	C	C2'	C	13	75.915
507	132	C	C3'	C	13	74.212
508	132	C	C4'	C	13	83.326

509	132	C	C5	C	13	97.733
510	132	C	C5'	C	13	65.839
511	132	C	C6	C	13	142.305
512	133	C	H1'	H	1	5.545
513	133	C	H2'	H	1	4.320
514	133	C	H3'	H	1	4.372
515	133	C	H4'	H	1	4.356
516	133	C	H5	H	1	5.815
517	133	C	H5'	H	1	4.058
518	133	C	H5"	H	1	4.075
519	133	C	H6	H	1	7.997
520	133	C	C1'	C	13	94.297
521	133	C	C2'	C	13	75.389
522	133	C	C3'	C	13	72.692
523	133	C	C4'	C	13	82.746
524	133	C	C5	C	13	97.464
525	133	C	C5'	C	13	65.090
526	133	C	C6	C	13	144.358
527	134	U	H3	H	1	14.538
528	134	U	H1'	H	1	5.742
529	134	U	H2'	H	1	4.457
530	134	U	H3'	H	1	4.535
531	134	U	H4'	H	1	4.336
532	134	U	H5	H	1	5.684
533	134	U	H5'	H	1	4.180
534	134	U	H5"	H	1	4.068
535	134	U	H6	H	1	7.966
536	134	U	C1'	C	13	92.139
537	134	U	C2'	C	13	75.313
538	134	U	C3'	C	13	73.688
539	134	U	C4'	C	13	82.430
540	134	U	C5	C	13	103.758
541	134	U	C5'	C	13	65.154
542	134	U	C6	C	13	142.837
543	135	U	N3	N	15	162.154
544	135	U	H3	H	1	12.953
545	135	U	H1'	H	1	5.641
546	135	U	H2'	H	1	4.479
547	135	U	H3'	H	1	4.324
548	135	U	H4'	H	1	4.349
549	135	U	H5'	H	1	4.170
550	135	U	H5	H	1	5.567
551	135	U	H6	H	1	7.811
552	135	U	C1'	C	13	93.840
553	135	U	C2'	C	13	76.008
554	135	U	C3'	C	13	72.530
555	135	U	C4'	C	13	83.976
556	135	U	C5	C	13	103.605
557	135	U	C5'	C	13	65.658
558	135	U	C6	C	13	141.753
559	136	C	H41	H	1	7.929
560	136	C	H42	H	1	6.753
561	136	C	H1'	H	1	5.494
562	136	C	H2'	H	1	4.366
563	136	C	H3'	H	1	4.511
564	136	C	H4'	H	1	4.412
565	136	C	H5'	H	1	4.148
566	136	C	H5	H	1	5.572
567	136	C	H6	H	1	7.728
568	136	C	C1'	C	13	93.876
569	136	C	C2'	C	13	75.575
570	136	C	C3'	C	13	72.260
571	136	C	C4'	C	13	81.951
572	136	C	C5	C	13	97.561
573	136	C	C5'	C	13	65.450
574	136	C	C6	C	13	141.782
575	137	U	H1'	H	1	5.415
576	137	U	H2'	H	1	4.104
577	137	U	H3'	H	1	4.234
578	137	U	H4'	H	1	4.407

579	137	U	H5'	H	1	4.441
580	137	U	H5	H	1	5.398
581	137	U	H6	H	1	7.733
582	137	U	C1'	C	13	93.649
583	137	U	C2'	C	13	75.591
584	137	U	C3'	C	13	72.187
585	137	U	C4'	C	13	81.927
586	137	U	C5	C	13	103.954
587	137	U	C5'	C	13	66.406
588	137	U	C6	C	13	142.728
589	138	C	H41	H	1	7.804
590	138	C	H42	H	1	7.038
591	138	C	H1'	H	1	5.495
592	138	C	H2'	H	1	4.174
593	138	C	H3'	H	1	4.421
594	138	C	H4'	H	1	4.349
595	138	C	H5	H	1	5.747
596	138	C	H5'	H	1	4.177
597	138	C	H6	H	1	8.017
598	138	C	C1'	C	13	93.791
599	138	C	C2'	C	13	75.048
600	138	C	C3'	C	13	75.343
601	138	C	C4'	C	13	81.528
602	138	C	C5	C	13	97.544
603	138	C	C5'	C	13	65.530
604	138	C	C6	C	13	144.067
605	139	U	H3	H	1	10.314
606	139	U	H1'	H	1	5.610
607	139	U	H2'	H	1	4.302
608	139	U	H3'	H	1	4.493
609	139	U	H4'	H	1	4.076
610	139	U	H5	H	1	5.461
611	139	U	H5'	H	1	4.062
612	139	U	H6	H	1	7.763
613	139	U	C1'	C	13	93.969
614	139	U	C2'	C	13	75.419
615	139	U	C3'	C	13	73.118
616	139	U	C4'	C	13	80.279
617	139	U	C5	C	13	102.683
618	139	U	C5'	C	13	64.680
619	139	U	C6	C	13	143.007
620	139	U	N3	N	15	156.855
621	140	C	H41	H	1	8.003
622	140	C	H42	H	1	7.098
623	140	C	H1'	H	1	5.628
624	140	C	H2'	H	1	4.388
625	140	C	H3'	H	1	4.290
626	140	C	H4'	H	1	4.316
627	140	C	H5	H	1	5.846
628	140	C	H5'	H	1	4.076
629	140	C	H5"	H	1	4.002
630	140	C	H6	H	1	7.975
631	140	C	C1'	C	13	94.102
632	140	C	C2'	C	13	74.984
633	140	C	C3'	C	13	75.155
634	140	C	C4'	C	13	83.307
635	140	C	C5	C	13	98.081
636	140	C	C5'	C	13	64.969
637	140	C	C6	C	13	141.985
638	141	U	H3	H	1	12.681
639	141	U	H1'	H	1	5.506
640	141	U	H2'	H	1	4.570
641	141	U	H3'	H	1	4.514
642	141	U	H4'	H	1	4.287
643	141	U	H5	H	1	5.439
644	141	U	H5'	H	1	4.233
645	141	U	H6	H	1	7.767
646	141	U	C1'	C	13	93.911
647	141	U	C2'	C	13	75.269
648	141	U	C3'	C	13	75.246

649	141	U	C4'	C	13	83.055
650	141	U	C5	C	13	103.901
651	141	U	C5'	C	13	66.710
652	141	U	C6	C	13	141.755
653	141	U	N3	N	15	161.027
654	142	A	H1'	H	1	5.973
655	142	A	H2	H	1	6.817
656	142	A	H2'	H	1	4.506
657	142	A	H3'	H	1	4.502
658	142	A	H4'	H	1	4.412
659	142	A	H5'	H	1	4.173
660	142	A	H5"	H	1	4.198
661	142	A	H8	H	1	8.107
662	142	A	C1'	C	13	92.932
663	142	A	C2	C	13	152.733
664	142	A	C2'	C	13	74.080
665	142	A	C3'	C	13	74.613
666	142	A	C4'	C	13	82.133
667	142	A	C5'	C	13	65.905
668	142	A	C8	C	13	139.938
669	143	A	H1'	H	1	5.570
670	143	A	H2	H	1	7.448
671	143	A	H2'	H	1	4.458
672	143	A	H3'	H	1	4.446
673	143	A	H4'	H	1	4.466
674	143	A	H5'	H	1	4.002
675	143	A	H5"	H	1	4.232
676	143	A	H8	H	1	7.728
677	143	A	C1'	C	13	92.662
678	143	A	C2	C	13	153.803
679	143	A	C2'	C	13	77.362
680	143	A	C3'	C	13	72.106
681	143	A	C4'	C	13	83.122
682	143	A	C5'	C	13	64.337
683	143	A	C8	C	13	139.132
684	144	C	H41	H	1	7.964
685	144	C	H42	H	1	6.747
686	144	C	H1'	H	1	5.185
687	144	C	H2'	H	1	4.083
688	144	C	H3'	H	1	4.300
689	144	C	H4'	H	1	4.301
690	144	C	H5	H	1	5.169
691	144	C	H5'	H	1	3.997
692	144	C	H6	H	1	7.402
693	144	C	C1'	C	13	93.618
694	144	C	C2'	C	13	75.310
695	144	C	C3'	C	13	72.055
696	144	C	C4'	C	13	81.789
697	144	C	C5	C	13	96.882
698	144	C	C5'	C	13	66.136
699	144	C	C6	C	13	140.581
700	145	U	H3	H	1	13.187
701	145	U	H1'	H	1	5.445
702	145	U	H2'	H	1	4.531
703	145	U	H3'	H	1	4.429
704	145	U	H4'	H	1	4.357
705	145	U	H5	H	1	5.256
706	145	U	H5'	H	1	4.037
707	145	U	H5"	H	1	4.470
708	145	U	H6	H	1	7.759
709	145	U	C1'	C	13	93.438
710	145	U	C2'	C	13	75.136
711	145	U	C3'	C	13	72.169
712	145	U	C4'	C	13	82.068
713	145	U	C5	C	13	103.559
714	145	U	C5'	C	13	64.593
715	145	U	C6	C	13	141.561
716	145	U	N3	N	15	161.882
717	146	G	H1	H	1	12.204
718	146	G	H21	H	1	7.880

719	146	G	H22	H	1	5.636
720	146	G	H1'	H	1	5.744
721	146	G	H2'	H	1	4.432
722	146	G	H3'	H	1	4.489
723	146	G	H4'	H	1	4.387
724	146	G	H5'	H	1	4.021
725	146	G	H5"	H	1	4.172
726	146	G	H8	H	1	7.688
727	146	G	C1'	C	13	92.732
728	146	G	C2'	C	13	75.170
729	146	G	C3'	C	13	73.110
730	146	G	C4'	C	13	82.067
731	146	G	C5'	C	13	65.380
732	146	G	C8	C	13	136.275
733	146	G	N1	N	15	147.260
734	147	C	H41	H	1	8.233
735	147	C	H42	H	1	6.628
736	147	C	H1'	H	1	5.428
737	147	C	H2'	H	1	4.167
738	147	C	H3'	H	1	4.386
739	147	C	H4'	H	1	4.335
740	147	C	H5	H	1	5.151
741	147	C	H5'	H	1	4.426
742	147	C	H5"	H	1	3.996
743	147	C	H6	H	1	7.619
744	147	C	C1'	C	13	93.944
745	147	C	C2'	C	13	75.586
746	147	C	C3'	C	13	75.561
747	147	C	C4'	C	13	82.006
748	147	C	C5	C	13	96.941
749	147	C	C5'	C	13	65.128
750	147	C	C6	C	13	140.972
751	148	C	H1'	H	1	5.701
752	148	C	H2'	H	1	3.947
753	148	C	H3'	H	1	4.115
754	148	C	H4'	H	1	3.979
755	148	C	H5	H	1	5.450
756	148	C	H5'	H	1	4.200
757	148	C	H5"	H	1	3.974
758	148	C	H6	H	1	7.609
759	148	C	C1'	C	13	92.843
760	148	C	C2'	C	13	77.586
761	148	C	C3'	C	13	69.720
762	148	C	C4'	C	13	82.275
763	148	C	C5	C	13	98.066
764	148	C	C5'	C	13	67.730
765	148	C	C6	C	13	141.783

A2. Resonance Assignments for c-JUN SL1 G115-C127

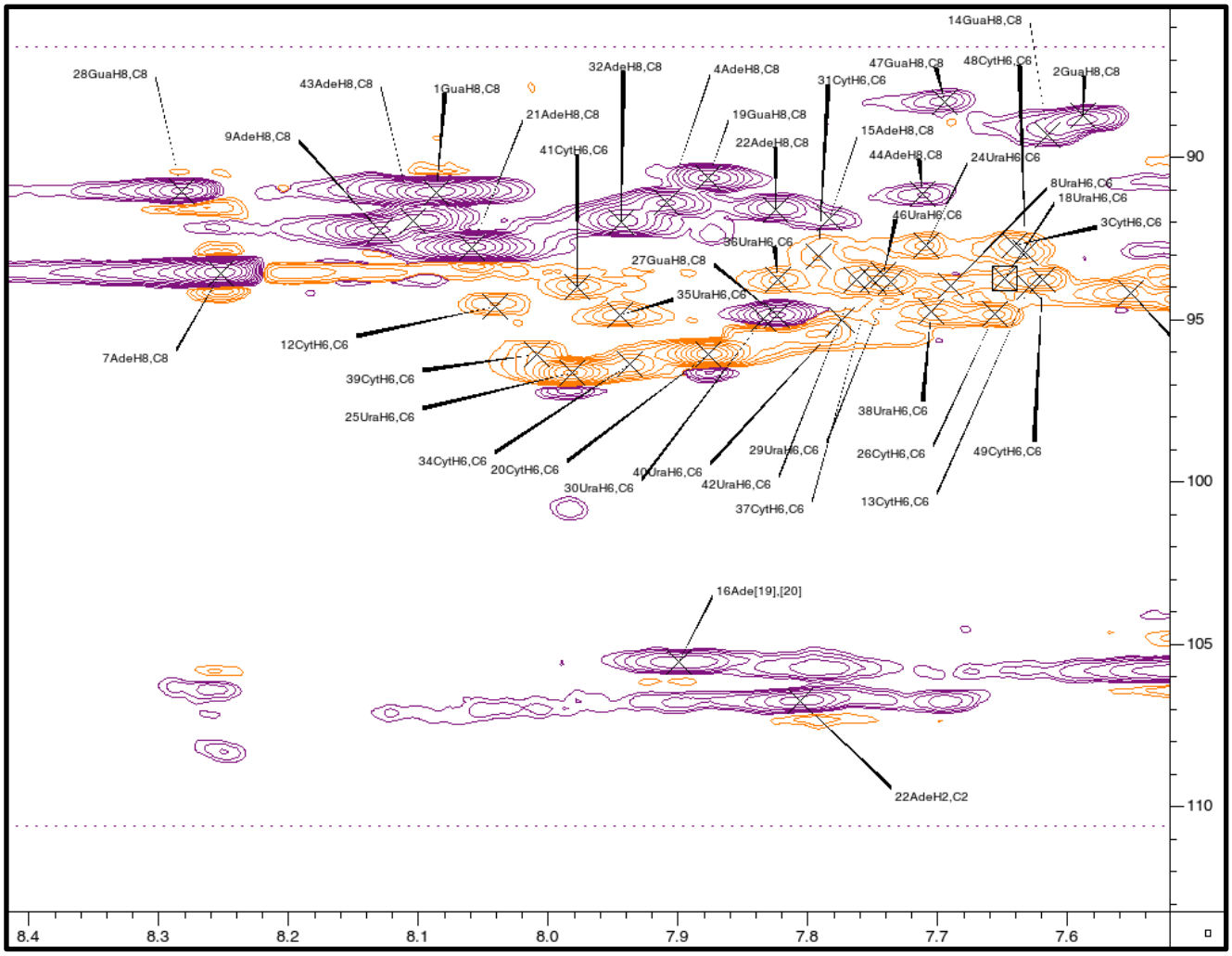
```
#####################################################################
#      Chemical Shift Ambiguity Index Value Definitions
#
# The values other than 1 are used for those atoms with different
# chemical shifts that cannot be assigned to stereospecific atoms
# or to specific residues or chains.
#
#  Index Value          Definition
#
#     1          Unique (including isolated methyl protons,
#                   geminal atoms, and geminal methyl
#                   groups with identical chemical shifts)
#                   (e.g. ILE HD11, HD12, HD13 protons)
#     2          Ambiguity of geminal atoms or geminal methyl
#                   proton groups (e.g. ASP HB2 and HB3
#                   protons, LEU CD1 and CD2 carbons, or
#                   LEU HD11, HD12, HD13 and HD21, HD22,
#                   HD23 methyl protons)
#     3          Aromatic atoms on opposite sides of
#                   symmetrical rings (e.g. TYR HE1 and HE2
#                   protons)                        #
#     4          Intraresidue ambiguities (e.g. LYS HG and
#                   HD protons or TRP HZ2 and HZ3 protons)
#     5          Interresidue ambiguities (LYS 12 vs. LYS 27)
#     6          Intermolecular ambiguities (e.g. ASP 31 CA
#                   in monomer 1 and ASP 31 CA in monomer 2
#                   of an asymmetrical homodimer, duplex
#                   DNA assignments, or other assignments
#                   that may apply to atoms in one or more
#                   molecule in the molecular assembly)
#     9          Ambiguous, specific ambiguity not defined
#
#####################################################################
loop_
   _Atom_chem_shift.ID
   _Atom_chem_shift.Comp_index_ID
   _Atom_chem_shift.Comp_ID
   _Atom_chem_shift.Atom_ID
   _Atom_chem_shift.Atom_type
   _Atom_chem_shift.Atom_isotope_number
   _Atom_chem_shift.Val

339         115         A         H1'         H         1         5.746
340         115         A         H2          H         1         7.902
341         115         A         H2'         H         1         4.503
342         115         A         H8          H         1         7.396
343         116         U         H1'         H         1         5.027
344         116         U         H2'         H         1         4.211
345         116         U         H6          H         1         7.200
346         117         C         H1'         H         1         5.389
347         117         C         H2'         H         1         4.132
348         117         C         H6          H         1         7.938
349         118         U         H1'         H         1         5.687
350         118         U         H2'         H         1         4.169
351         118         U         H6          H         1         7.694
352         119         U         H1'         H         1         5.676
353         119         U         H2'         H         1         4.180
354         119         U         H6          H         1         7.673
355         120         A         H1'         H         1         5.952
356         120         A         H2          H         1         8.085
357         120         A         H2'         H         1         4.724
358         120         A         H8          H         1         8.285
359         121         U         H1'         H         1         5.831
360         121         U         H2'         H         1         4.294
361         121         U         H6          H         1         7.694
```

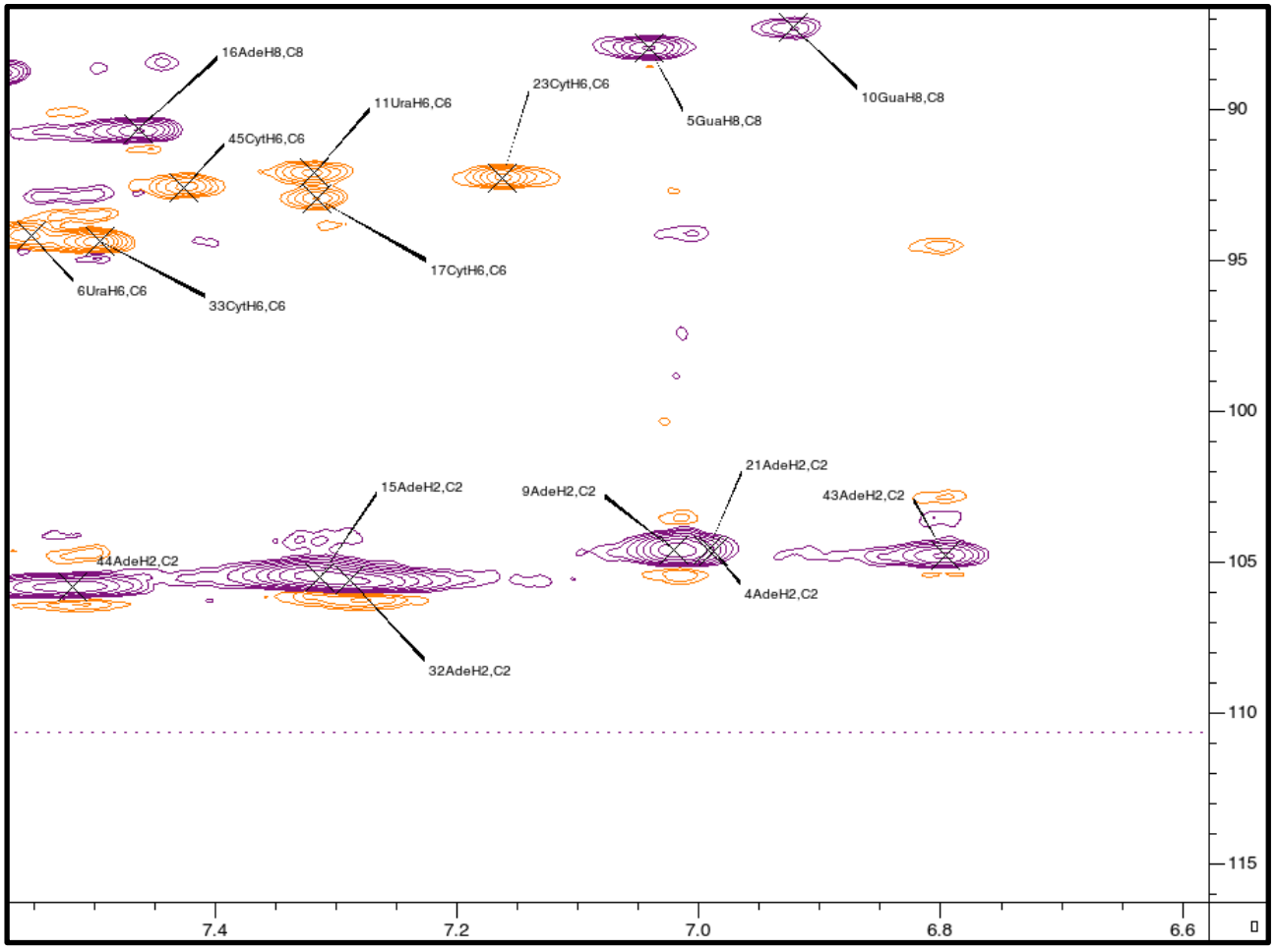
362	122	U	H1'	H	1	5.843
363	122	U	H2'	H	1	4.322
364	122	U	H6	H	1	7.734
365	123	U	H1'	H	1	5.906
366	123	U	H2'	H	1	4.404
367	123	U	H6	H	1	7.817
368	124	U	H1'	H	1	5.716
369	124	U	H2'	H	1	4.463
370	124	U	H6	H	1	7.926
371	125	C	H1'	H	1	5.608
372	125	C	H2'	H	1	4.311
373	125	C	H6	H	1	8.107
374	126	U	H1'	H	1	5.686
375	126	U	H2'	H	1	4.413
376	126	U	H6	H	1	7.903
377	127	U	H1'	H	1	5.740
378	127	U	H2'	H	1	4.472
379	127	U	H6	H	1	8.054

APPENDIX B ASSIGNED SPECTRA

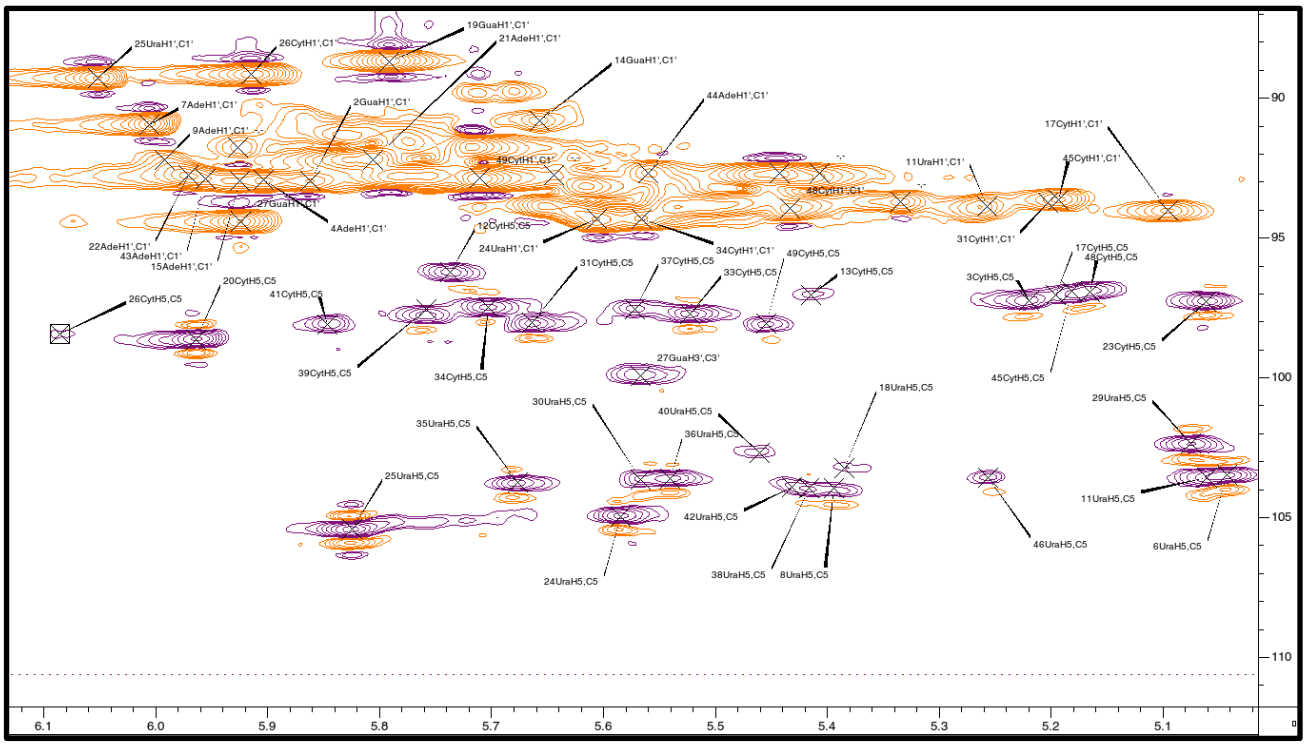
Appendix B1. ^{13}C Filtered 2D HSQC – TL1 H8, H6, H2 (Downfield) Assignments (^{1}H 8.4-7.5 ppm, ^{13}C 110-85 ppm)



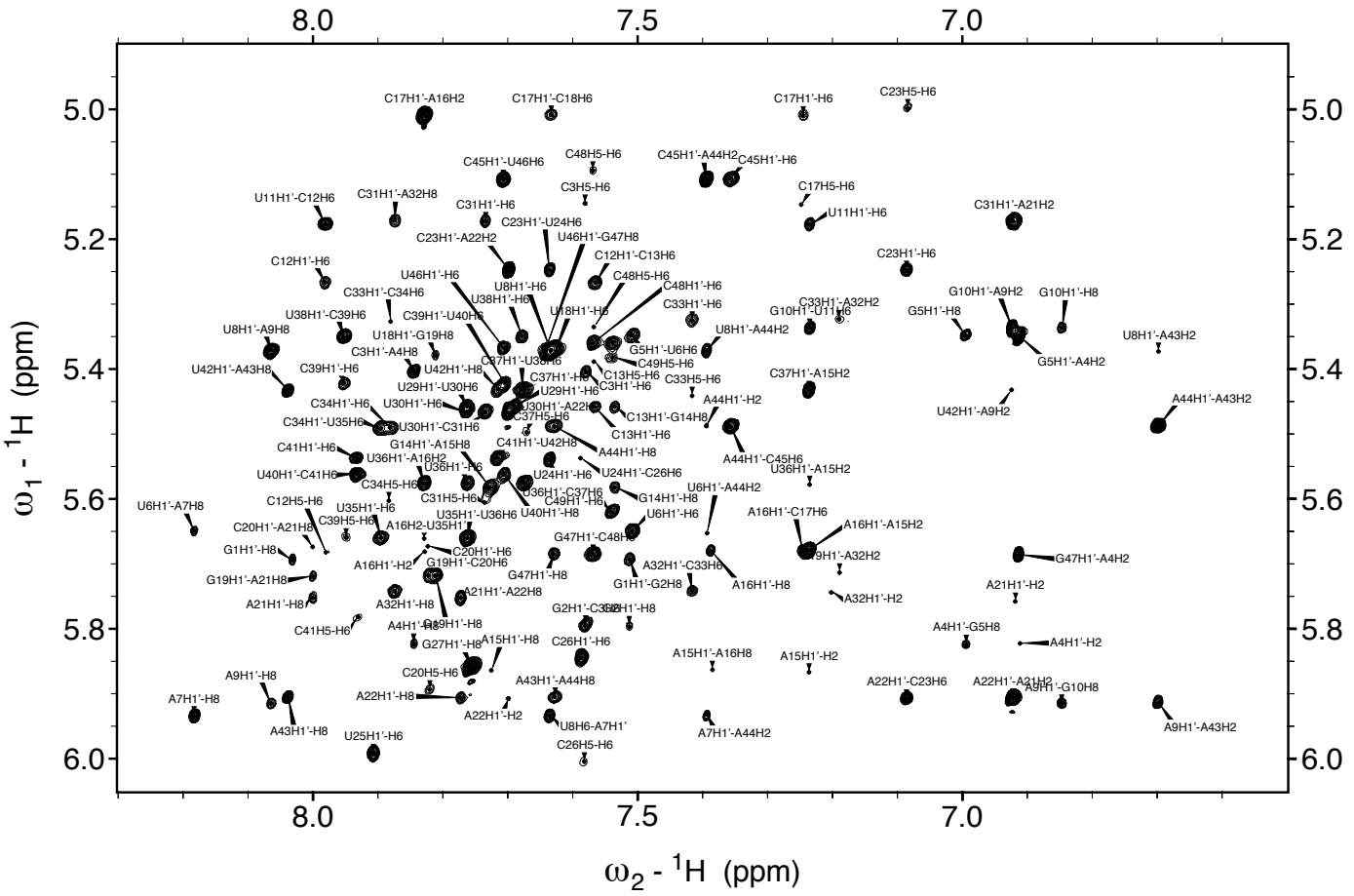
Appendix B2. ^{13}C Filtered 2D HSQC – TL1 H8, H6, H2 (Upfield) Assignments (^{1}H 7.5-6.6 ppm, ^{13}C 110-85 ppm)



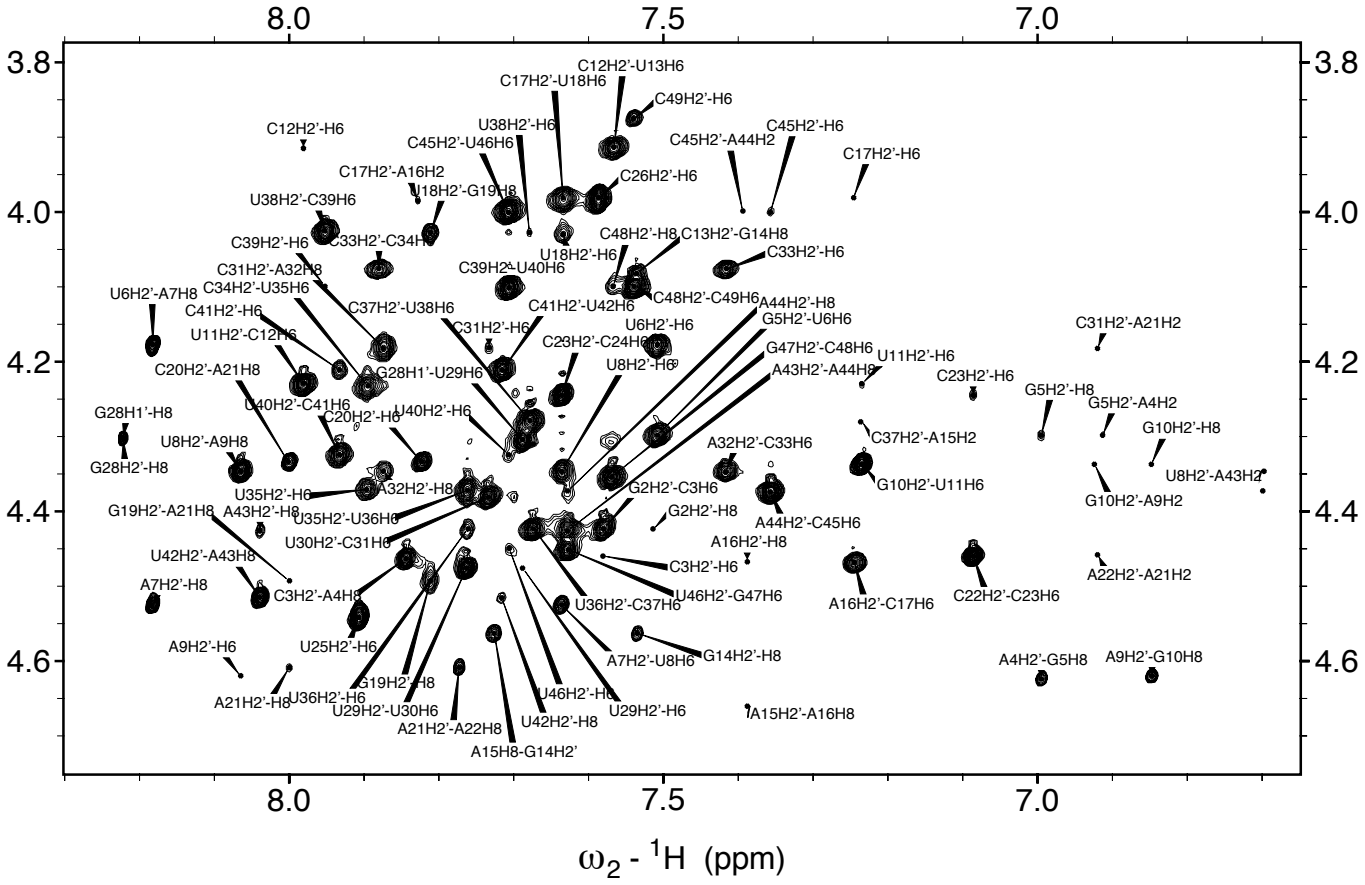
Appendix B3. ^{13}C Filtered 2D HSQC – TL1 H5 Assignments (^{1}H 6.1-5.0 ppm, ^{13}C 110-85 ppm)



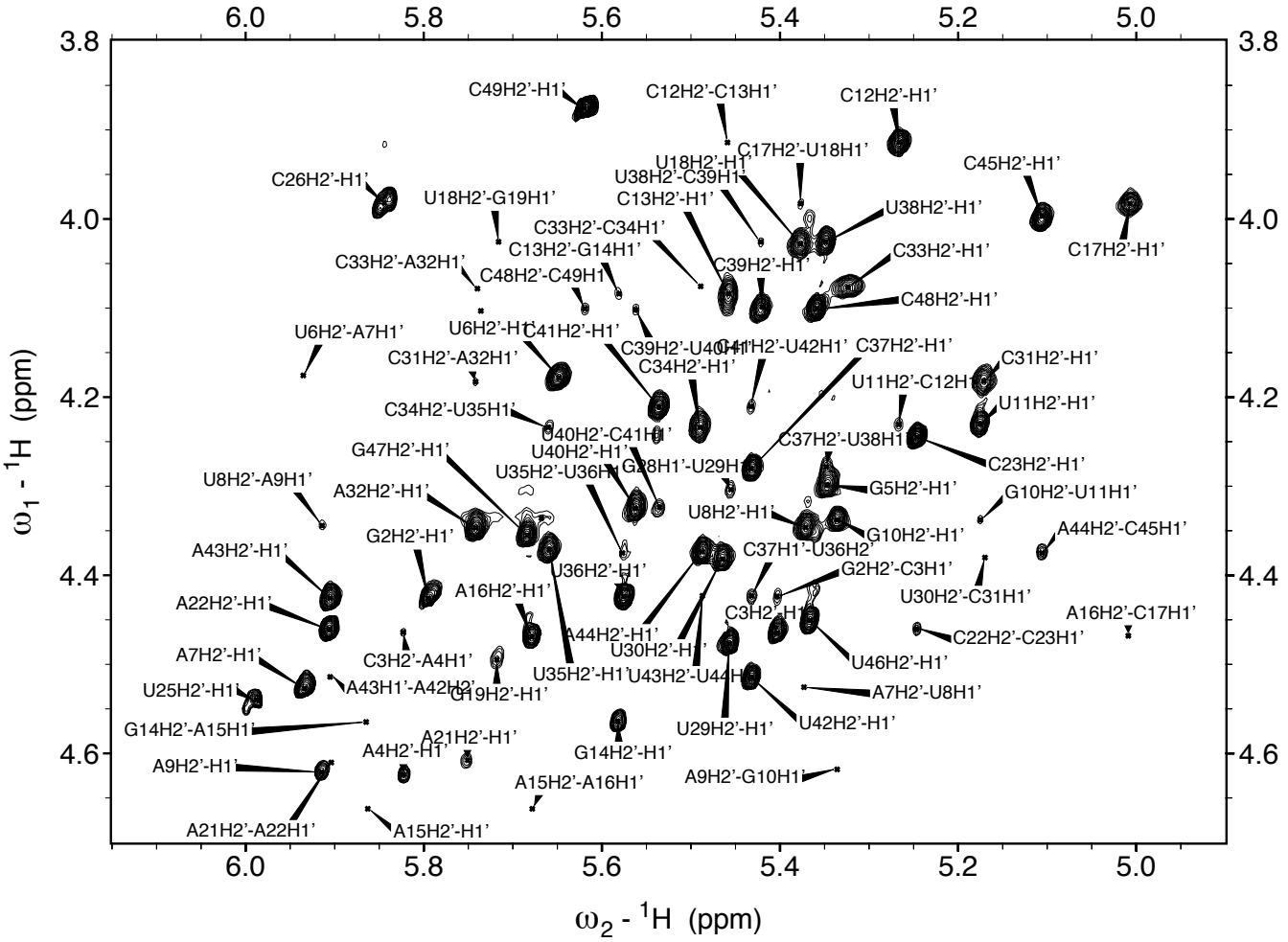
Appendix B4. D_2O 2D NOESY – Selectively Deuterated TL1 H2, H6, H8, H1' Assignments (^1H 8.4-6.8 ppm, ^1H 6.1-5.0 ppm)



Appendix B5. D_2O 2D NOESY – Selectively Deuterated TL1 H6, H8, H1', H2' Assignments (^1H 8.4-6.8 ppm, ^1H 5.0-3.9 ppm)



Appendix B6. D_2O 2D NOESY – Selectively Deuterated TL1 H1', H2', H3', H4', H5', H5'' Assignments (^1H 6.2-5.1 ppm, ^1H 5.0-3.9 ppm)



Appendix B7. D_2O 2D NOESY – TL1 H2, H6, H8, H1' Assignments (^1H 8.4-6.8 ppm, ^1H 6.1-5.0 ppm)