Structural Studies of Viral RNAs and Peptide Targeting

Yi-Ting Sun

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Gabriele Varani, Chair

Ashleigh Theberge

Joshua Vaughan

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Structural Studies of Viral RNAs and Peptide Targeting

Yi-Ting Sun

Chair of the Supervisory Committee:

Gabriele Varani

Department of Chemistry

RNA structures are involved in many biological processes and the progression of human disease, making them potential targets for therapeutic development. The first chapter introduces topics relevant to targeting RNA structures and to the subsequent chapters of this thesis: disease-associated RNA classes, the characteristics of RNA-binding ligands of different sizes (e.g., small molecules, peptides, engineered proteins), and NMR-based methods for RNA structure determination.

Chapter 2 describes the strategy and result of my research on designing cyclic β-hairpin peptidomimetics targeting pre-microRNA-20b (pre-miR-20b). Small cyclic peptidomimetics were rationally designed based on the sequence of β2-β3 hairpin of Rbfox2 protein, that recognize the terminal loop of precursor miR-20b. I identified a peptide with low µ-molar affinity for the miR-20b precursor and specifically targets the apical loop of pre-miR-20b. This work demonstrates that it is possible to mimic RNA-binding proteins with a minimal structurally pre-organized peptide, which

provide a starting point for designing or evolving small peptide mimetics of RNA-binding proteins.

Chapter 3 describes the results of structural analysis of a 70 nucleotides stem-loop RNA structure (called SLA) from Dengue virus serotype 1 (DENV1), which functions as the promoter for viral replication. NMR structure of a monomeric DENV1 SLA is assembled to high-resolution from independently folded structural elements, and SAXS modeling is used for independent validation. Both NMR structure determination and SAXS modeling result in an L-shape of RNA structure. It is very likely that the three-dimensional structure of SLA is conserved among flaviviruses because the sequence is highly conserved among different flavivirus genomes. This work establishes RNA structural features involved in Dengue replication and provides a foundation for the discovery of new antiviral drugs that target this essential replicative step.

ACKNOWLEDGEMENTS

First, I am very thankful to my supervisor Prof. Dr. Gaberiele Varani for his guidance and support with full encouragement and enthusiasm, and patience during my PhD study.

I would like to recognize Dr. Matt Shortridge for his guidance during my first year in the group. Thanks for the supporting of my peptide project, teaching me techniques around the lab including peptide and RNA synthesis, NMR experiments and structure calculations. I want to thank Dr. Shrikant Sharma for guiding me through peak assignments and NMR analysis of RNA molecules.

I would also like to thank all the members in the group. Thanks to Tom Pavelitz for the support on gel binding assays and being a great lab manager. Thanks to Dr. Greg Olsen and Rajan Paranji for NMR support. Thanks to Tommy, Bhawna and Emma for discussion their kind company. It is their kind help and support that have made my study and life in Seattle a wonderful time.

Finally, I want to express my gratitude to my family. Without their tremendous understanding and financial support, it would be impossible for me to complete my PhD degree.

TABLE OF CONTENTS

·Chapter 1

In the past few decades, RNA molecules have received increased attention as targets for drug discovery efforts, because RNA has been revealed to play critical roles in many human diseases. RNA molecules are involved in gene regulation processes such as transcription, splicing, mRNA transport and translation, and therefore they are also strictly regulated by other RNAs or RNA-binding proteins transcriptionally and post-transcriptionally. RNAs and RNA-protein interactions regulate viral replication (e.g. HIV TAR and Dengue SLA) which infect human cells, and mis-regulation of non coding RNAs in human cells can lead to diseases such as cancer. In this first chapter, I introduce topics relevant to the subsequent Chapters 2 and 3 are introduced: introductions to different kinds of RNA which play important role in human disease; the characteristics of RNA-binding ligands (e.g. small molecules, peptides, proteins), and the NMR-based methods for structure determination used in the chapters that follow.

Section 1: Non-Coding Regions of mRNAs and Non-Coding RNAs Play Important Roles in Gene Expression and Viral Replication

The analysis of the human genome (Lander et al., 2001; Venter et al., 2001) has revealed that less than 2% of the human genome is translated into proteins (Warner et al., 2018). In fact, most of the genome is expressed at the RNA level and involved in the regulation of gene expression (Breaker & Joyce, 2014; Castro-Oropeza et al., 2018; Di Leva et al., 2014; Esteller, 2011; Hill & Tran, 2021; Pereira-Castro & Moreira, 2021; Statello et al., 2021). This can occur at the transcriptional level, promoting or inhibiting the transcription of a gene, or the post-transcriptional level, controlling the fate of the transcribed RNA molecules. RNA elements that are involved in gene regulation include the well-known transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), as well as untranslated regions (UTRs) in mRNA, but also new players such as non-coding RNAs (ncRNAs), such as microRNAs, siRNAs and piRNAs, and large non-coding RNAs (lncRNAs). These RNAs can form complex secondary or higher order structures which participate in their function. Furthermore, a large number of ncRNAs and non coding regions of mRNA are found to be disease-associated (Castro-Oropeza et al., 2018; Di Leva et al., 2014; Esteller, 2011; Hill & Tran, 2021; Statello et al., 2021; Winkle et al., 2021). Many viral RNA elements, often within untranslated regions, were also found to be critical for viral replication and translation (Chen et al., 2018; Mazeaud et al., 2018; Neupane et al., 2020; Rasekhian et al., 2021; Stevens et al., 2006). Because of the growing interest in ncRNAs and untranslated regions of protein-coding RNAs (UTRs) are viewed as potential therapeutic targets. Various RNA molecules have been intensely studied and potential drugs have been discovered in the past decades (Chavali et al., 2019; Cooper et al., 2009; Falese et al., 2021; Hegde et al., 2021; Li & Chen, 2013; Ling et al., 2013; Matsui & Corey, 2017; Winkle et al., 2021), but only few have reached the clinical stage and questions exist concerning their modality (Cheung et al., 2018; Mullard, 2020; Ratni et al., 2018; Ratni et al., 2021; Singh et al., 2020; Zogg et al., 2022).

Section 1.1: UTRs Play Essential Roles in Post-Transcriptional and Translational Regulation of mRNAs

Untranslated regions (UTRs) of mRNA play critical roles in post-transcriptional regulation of gene expression. Many examples of cis-acting regulatory elements in 5'-UTR were found to enhance or suppress translation of the downstream protein coding genes, while 3'-UTR often regulates gene expression by influencing mRNA stability (Rasekhian et al., 2021; Wilkie et al., 2003). For instance, AU-rich elements in 3'-UTR, which can facilitate the shortening of poly(A) tail, are known to be responsible for translational repression and mRNA destabilization (Chen & Shyu, 1995; Kontoyiannis et al., 1999; Mayr, 2019; Otsuka et al., 2019). 3'-UTRs also contain binding sites for trans-acting regulatory elements such as microRNAs or RNA binding proteins which can induce degradation or stabilize a mRNA (Matoulkova et al., 2012).

Riboswitches are regulatory elements found usually in 5'-UTRs that control translation rates by directly binding small-molecule metabolites (Breaker, 2012; Coppins et al., 2007; Richards & Belasco, 2021; Roth & Breaker, 2009). Most known riboswitches occur in bacteria, but functional riboswitches have also been discovered in eukaryotic systems (Broft et al., 2021; Cheah et al., 2007; Dann et al., 2007; Nahvi et al., 2004; Sudarsan et al., 2003). The core of the riboswitch mechanism is a conformational change induced by ligand binding which promotes or suppresses the formation of critical RNA structures and, for example, mask or unmask ribosome-binding sites. Riboswitches provide proof of concept that RNA-ligand interactions can regulate gene expression and directly validate RNA as a promising target for drug discovery (Howe et al., 2015; Howe et al., 2016).

The canonical translation initiation of eukaryotic mRNA occurs through a cap-dependent mechanism, where the protein complex eIF4F (eukaryotic initiation factor 4F) which contains a RNA helicase and recognizes the mRNA 5'-m^7GTP cap through eIF4E, unwinds the mRNA structure and allow the small (40S) ribosomal subunit to land near the translation start site (Bhat et al., 2015; Jackson et al., 2010; Lopez-Ulloa et al., 2022). An alternative cap-independent mechanism for translation initiation was discovered in flaviviruses such as hepatitis C virus, where a complex structure within the 5'-UTR acts as an internal ribosome entry site (IRES) to recruit the 40S ribosomal subunit and promote translation initiation independent of the cap (Martinez-Salas et al., 2018). Various RNA viruses with uncapped genome depend on IRES elements to initiate viral protein synthesis

(Lee et al., 2017; Lopez-Ulloa et al., 2022; Pelletier & Sonenberg, 1988). IRES elements often adopt complex structures, which serve as the anchoring site for the ribosome through RNA–protein interactions which in many cases includes the use of canonical eIFs and RNA-binding proteins (RBPs) in unique non canonical ways (Jang & Wimmer, 1990; Khan et al., 2014; Lee et al., 2017; Lopez-Ulloa et al., 2022; Lozano & Martinez-Salas, 2015).

Section 1.2: Viral UTRs are Critical for Replication

Viral UTRs contains essential structural elements required not only for translation but also for viral replication. The HIV (human immunodeficiency virus) trans-activation response element RNA (TAR) is a classic example. Transcription of the integrated HIV-1 RNA in infected cells is activated by the complex of the viral Tat protein and TAR RNA, which is found at the 5 'end of nascent viral transcripts (Peterlin & Price, 2006; Stevens et al., 2006). Through an interaction with TAR, Tat protein and the transcription elongation factor-b (P-TEFb) are recruited to the elongating RNA polymerase II (RNAP II). The CDK9 kinase component of p-TEFb phosphorylates the C-terminal domain (CTD) of RNAP II and the repressive factors NELF and DSIF, enhancing transcription of the viral RNA (Figure 1.1). The recruitment of Tat and p-TEFb to TAR is required for reactivation of the integrated proviral genome in latently infected cells (Peterlin & Price, 2006; Stevens et al., 2006). Therefore, TAR has been an attractive target for antiviral development, and numerous small molecule and peptidic inhibitors of the Tat-TAR interaction have been presented during the past two decades (Alanazi et al., 2021; Baba, 2006; Gallego & Varani, 2001; Murchie et al., 2004; Zeiger et al., 2014).

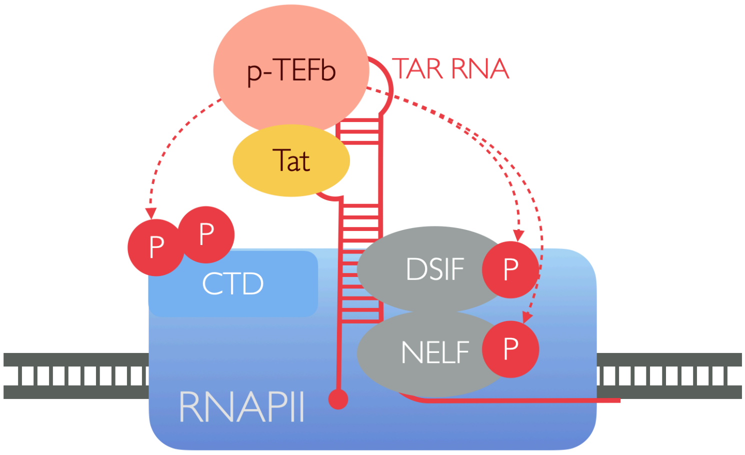


Figure 1.1. Schematic illustration of the activation of transcription of the integrated HIV-1 RNA in infected cells by the complex of the viral Tat protein, TAR RNA, and human cyclin T1, the cdk9 kinase, integrated into the p-TEFb complex. The complex is recruited to the polymerase RNAP II just downstream of the start of transcription, where the enzyme is stalled by the stable stem-loop. Phosphorylation by CDK9 kinase within p-TEFb inactivates NELF and DSI, and activates the CTD of RNAP II to relieve transcriptional blocks and enhance elongation of the viral RNA.

Another example of UTR structures regulating viral replication are found in Dengue virus, a flavivirus. Dengue virus (DENV) is a single-stranded positive sense RNA virus with a genome of about 11 kb flanked by 5 ’and 3 ’UTRs (Colavita et al., 2020; Dang et al., 2020; Kamau et al., 2019; Pascalis et al., 2020; Selisko et al., 2014). Studies have identified cis-acting RNA elements in both UTRs which are essential for translation and RNA synthesis, including the highly-conserved terminal stem-loop within the 3’UTR and stem-loop A (SLA) within the 5’UTR (Gebhard et al., 2011; Ng et al., 2017). The removal of either SL or SLA leads to a significant reduction of virus replication (Alvarez, De Lella Ezcurra, et al., 2005; Filomatori et al., 2006; Yu et al., 2008). The terminal SL appears to be important for viral

RNA synthesis, and is required for efficient translation as well (Alvarez, De Lella Ezcurra, et al., 2005; Ng et al., 2017). The SLA within the 5'UTR has been proposed to be the promoter for the viral RNA polymerase NS5 and a direct interaction between them has been shown to be critical for RNA synthesis (Choi, 2021; Filomatori et al., 2006; Gebhard et al., 2011). The structure and significance of the SLA is further discussed in Chapter 3.

DENV and other flavivirus genomes contain inverted complementary sequences at the ends of the RNA, mediating long-range RNA-RNA interactions which induce the cyclization of the genome (Gebhard et al., 2011; Ng et al., 2017). Two pairs of complementary regions in DENV, 5'-3'CS and 5'-3'UAR, are responsible for these long-range RNA-RNA interactions. The 5'UAR is located within the 5 'stem-loop B (SLB), and the 3'UAR is found at the bottom of 3'SL and the 3 'small hairpin (sHP). Visualization of DENV genome cyclization was demonstrated by atomic force microscopy (AFM) (Filomatori et al., 2006).

The current model for the initiation of DENV RNA replication starts with opening of 5'SLB, 3'sHP and 3'SL helices, allowing the complementary regions to form RNA duplexes involving 5'-3'CS and 5'-3-UAR to cyclize the viral genome. The genome cyclization positions the 5 'and 3'UTRs in close proximity, relocating the SLA-bound NS5 to the SL at the end of the 3'UTR to initiate (-)-strand synthesis (Figure 1.2) (Alvarez, Lodeiro, et al., 2005; Filomatori et al., 2011; Filomatori et al., 2006; Hodge et al., 2016; Mazeaud et al., 2018; van den Elsen et al., 2021; Villordo et al., 2010).

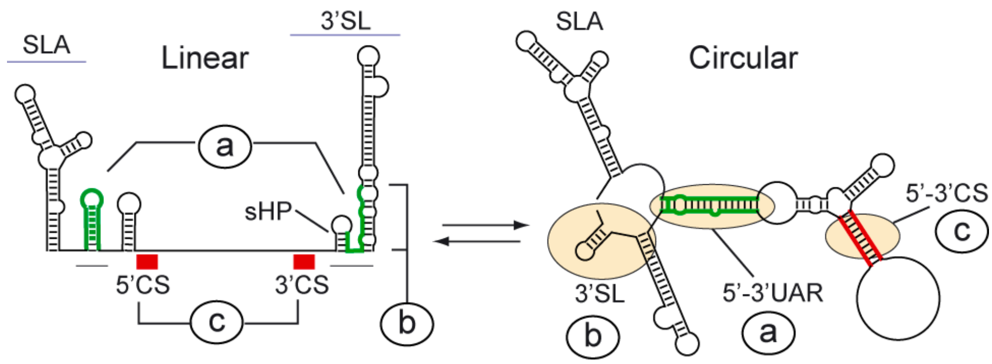


Figure 1.2. Schematic representation of the current model for the initiation of DENV RNA replication (Gebhard et al., 2011). This involves conformational changes in the RNA structure during the transition between the linear and circular forms of the DENV genome. Predicted changes of conserved RNA structures upon 5'-3' end hybridization are indicated: (a) the SLB and the sHP of the 3'SL structures, as observed in the linear conformation, open to form an extended duplex helix in the circular form; (b) the large stem of the 3'SL observed in the linear form opens, releasing the last nucleotides of the genome in the circular form; (c) the complementary sequences 5' and 3' CS form a duplex in the circular form through RNA-RNA interactions.

Section 1.3: Long Non-Coding RNAs

Long non-coding RNAs (lncRNAs) are non-protein coding RNA molecules that contains more than 200 nucleotides (Derrien et al., 2012; Statello et al., 2021). Some lncRNAs have been found to possess conserved structures between different species (Necsulea et al., 2014), and through a series of lncRNAs knock-out experiments, lncRNAs have demonstrated to perform critical roles in cellular physiology (Sauvageau et al., 2013). Many lncRNAs are currently believed to participate in the recruitment of transcription factors or chromatin-modifying complexes to chromatin, or otherwise inhibit the function of protein complexes (DiStefano, 2018; Mirzaei et al., 2022; Statello et al., 2021; Wapinski & Chang, 2011). One example is the well-known lncRNA HOTAIR (HOX transcript antisense intergenic RNA) which is a regulator of epidermal tissue development, and particularly abundantly expressed in human peripheral tissues (Cantile et al., 2021; Chen et al., 2021; Rinn et al., 2007; Schorderet & Duboule, 2011). HOTAIR is a 2 kb spliced and polyadenylated transcript which interacts with chromatin-remodeling enzymes and silences the genes of numerous tumor and metastasis suppressors (Gupta et al., 2010; Mozdarani et al., 2020; Rinn et al., 2007). Therefore, over-expression of HOTAIR can lead to tumor development and metastasis (Gupta et al., 2010; Kim et al., 2013; Mozdarani et al., 2020). Despite the growing interest in lncRNAs, the mechanism by which they exert their molecular functions remains largely to be studied, as does their structure.

Section 1.4: MicroRNAs

MicroRNAs (miRNAs) are small non-coding RNAs of 21-23 nucleotides that play key roles in post-transcriptional regulation of gene expression by binding to target mRNAs and inhibiting translation or inducing mRNA degradation (Bartel, 2009; Fabian et al., 2010; Hill & Tran, 2021). MiRNAs primarily base-pair to the 3'-untranslated region (UTR) of the mRNA by incomplete complementary base pairing. Each miRNA can regulate multiple genes and each gene can be regulated by multiple miRNAs (Fabian et al., 2010; Hill & Tran, 2021). As a result, even a small change in the level of a miRNA can combinatorially cause large phenotypic effects trough regulation of a large number of mRNAs.

MiRNAs are involved in all fundamental cellular processes (Baek et al., 2008; Bartel, 2009; Fabian et al., 2010). Currently, more than 2000 miRNAs are recognized in humans (Friedlander et al., 2014; Kozomara & Griffiths-Jones, 2014), targeting at least 60% of human protein-coding genes (Friedman et al., 2009). Mis-regulation of the expression of miRNAs is associated with many human diseases as well (Croce, 2009; Di Leva et al., 2014; Hill & Tran, 2021; Im & Kenny, 2012).

MiRNAs are regulated both transcriptionally and post-transcriptionally (Ha & Kim, 2014; Krol et al., 2010; Li & Yu, 2021). The generation of miRNAs starts from the primary miRNA (pri-miRNA) which is a long precursor RNA generally transcribed by RNA polymerase II (Figure 1.3). The Pri-miRNA is cleaved by the Drosha/DGCR8 microprocessor complex in the nucleus to produce the precursor miRNA (pre-miRNA), a 60-70-nucleotides-long stem-loop embedding the mature miRNA. The pre-miRNA is then cleaved into a ~22 nt miRNA:miRNA* duplex by the RNAse III enzyme Dicer in the cytoplasm. The mature miRNA is subsequently loaded onto the RNA-induced silencing complex (RISC) to guide the repression of target mRNAs (Ha & Kim, 2014; Krol et al., 2010). It has been found that the apical portion of the stem-loop miRNA precursors is often involved in the regulation of the maturation process, and several proteins have been found to regulate miRNA processing by binding to this apical loop region (Castilla-Llorente et al., 2013; Choudhury & Michlewski, 2012; Ha & Kim, 2014; Li & Yu, 2021).

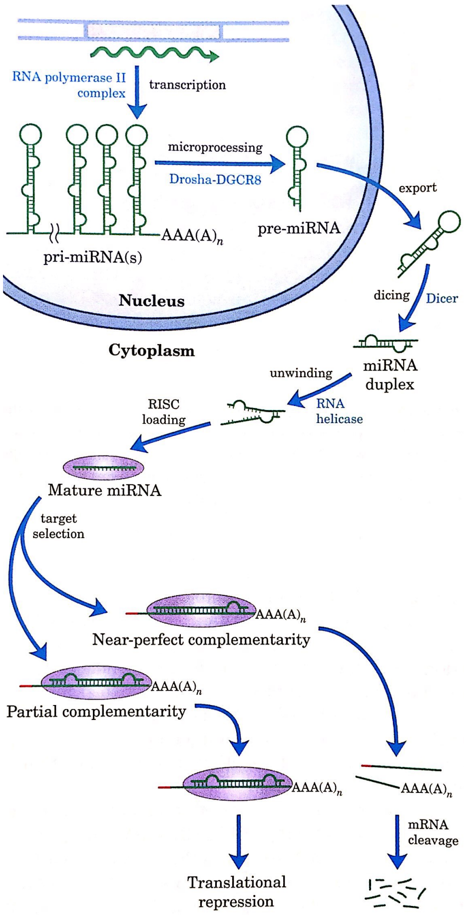


Figure 1.3. Synthesis and processing of miRNAs. Pri-miRNAs are processed by the Drosha/DGCR8 complex in the nucleus to generate pre-miRNAs, and then by Dicer-TRBP in the cytoplasm to generate the mature miRNAs. MiRNAs are loaded onto the RNA-induced silencing complex (RISC), and then bind to a target mRNA to suppress its translation or induce its degradation (Nelson et al., 2017).

Section 1.5: RNAs in Diseases

Many RNAs have been reported to be related to human diseases and thus received attention as potential targets for drug discovery. Diseases can be caused by mis-regulation of non-coding RNAs in cells, or the infection of virus particle.

For virus-induced diseases, viral UTRs are potential drug targets because of their significance in viral replications, which are introduced in Section 1.2. Inhibiting the RNA function and RNA-protein interactions can inhibit the multiplication of viruses and possibly lead to the cure of the disease. Many efforts have been dedicated to develop drug molecules targeting viral RNA motifs such as HIV TAR (Alanazi et al., 2021; Baba, 2006; Gallego & Varani, 2001; Murchie et al., 2004; Zeiger et al., 2014), HIV RRE (Chapman et al., 2002; Prado et al., 2016; Shuck-Lee et al., 2008) and HCV IRES (Carnevali et al., 2010; Dibrov et al., 2014; Ding et al., 2014; Rynearson et al., 2014; Zhou et al., 2013).

Cancer can be caused by mis-regulation of non-coding RNAs in human cells such as lncRNAs and miRNAs, which are introduced in Section 1.3 and 1.4. One famous example is miR-21, which had been identified as a cancer-promoting “oncomiR". Mir-21 targets and inhibits tumor suppressor genes including Phosphatase and Tensin Homolog (PTEN) and Programmed Cell Death 4 (PDCD4) (Gaur et al., 2011; Meng et al., 2007; Zhu et al., 2007; Zhu et al., 2008). It has been shown that miR-21 is commonly upregulated in tumor cells(Volinia et al., 2006). Studies have implicated the role of miR-21 in tumor pathogenesis and during all other stages of tumorigenesis (Chan et al., 2005; Feng & Tsao, 2016; Zhao et al., 2015). The development of anticancer therapeutics and drug targeting miR-21 has been carried out by different groups (Bose et al., 2012; Fu et al., 2021; Gumireddy et al., 2008; Javanmard et al., 2020; Naro et al., 2015; Watashi et al., 2010; Wickramasinghe et al., 2009).

The development of RNA-binding molecules in different sizes is subsequently introduced in the next section.

Section 2: RNA-binding Molecules Come in Different Sizes

With the growing awareness of RNA as a potential therapeutic target, different kinds of RNA-targeting molecules have been proposed. The size of these molecules varies from small drug-like molecules to proteins. In this section, the chemistry of different types of RNA-binding molecules is introduced.

Section 2.1: Oligonucleotides and their analogues

The potential usefulness of RNA-targeting drugs has been proven by preclinical studies and clinical trials using antisense oligonucleotides which bind to specific sequences through Watson-Crick base pairing (Bennett, 2019; Di Fusco et al., 2019). Two major therapeutic approaches in this field are antisense oligonucleotides that inhibit mRNA translation and oligonucleotides which function through the RNA interference (RNAi) pathway (Chery, 2016; Gagliardi & Ashizawa, 2021; Hung & Slotkin, 2021; Kole et al., 2012; Tian et al., 2021). Although these large and highly charged oligonucleotide molecules inevitably present stability, distribution and delivery challenges (Moreno & Pego, 2014), they are nevertheless considered important tools for gene-specific therapeutics that target the RNAs of disease-related genes.

Antisense oligonucleotides are sequences which can bind to specific target RNAs, such as regulatory element in UTRs, and inhibit it by inducing RNA degradation or by steric blocking access of other RNAs and proteins (Chan et al., 2006; Gagliardi & Ashizawa, 2021). The mechanism of RNAi induced post-transcriptional gene silencing involves two steps (Agrawal et al., 2003; Tian et al., 2021): First, a larger double-stranded RNA is cleaved into small interfering RNA (siRNA) by certain RNA nucleases. Second, these siRNAs join the multi-nuclease complex RISC and degrade single-stranded mRNAs containing complementary sequences.

Section 2.2: Small Molecules

Small molecules that affect RNA function have been identified in nature such as in RNA riboswitches, which are described in Section 1.1, and ribosomal RNA binding small molecule antibiotics produced by bacterial species (McCown et al., 2017; Wilson, 2014). These examples provide encouraging precedence for the development of RNA targeting small molecule therapeutics. A number of RNA binding small molecules have also been discovered through different drug discovery strategies, including high-throughput screening (Abulwerdi et al., 2019; Borgelt et al., 2021; Davila-Calderon et al., 2020; Hermann, 2016; Howe et al., 2015; Palacino et al., 2015; Prado et al., 2016; Rizvi et al., 2020; Seth et al., 2005; Sztuba-Solinska et al., 2014), structure-based rational design (Barros et al., 2016; Joly et al., 2014; Luu et al., 2016; Nguyen et al., 2015; Rzuczek et al., 2017), fragment-based approaches (Binas et al., 2021; Harner et al., 2013; Lee et al., 2014; Lundquist et al., 2021; Tam et al., 2019) and computational modeling (Park et al., 2011; Stelzer et al., 2011; Tanida & Matsuura, 2020). However, the discovery and development of small molecules to inhibit specific RNAs and RNA-protein interactions remains challenging, in part because of the large surface interface buried in the complex, because of the inherent flexibility of RNA (Aboul-ela, 2010; Falese et al., 2021) and of its high negative charge density.

High-throughput screening contributed by the development of automatic device and rapid assay methods is a widely used method in drug discovery (Zeng et al., 2020). Potential binders can be rapidly identified by screening large libraries against the target molecule even without knowing the detailed structure of a target or target-ligand complex. Binding signals can be detected by analytical methods such as fluorescence spectroscopy or mass spectrometry (Eydoux et al., 2021; Rizvi et al., 2018; Zeng et al., 2020). The Automated Ligand Identification System (ALIS), which was developed by Merck, is a label-free affinity-selection mass spectrometry (AS-MS) platform for high-throughput screening of small molecules. By coupling fast (<20 s) size exclusion chromatography to separate free ligands from target–ligand complexes to LC-MS ligand identification, this platform can screen up to 500,000 compounds a day with minimal and label-free sample consumption (Rizvi et al., 2018; Rizvi & Nickbarg, 2019). Overall, high-throughput screening has the advantages of automatic detection against large libraries at a rapid pace and require low sample amounts.

Structure-based rational design of small molecules relies on a detailed understanding of the target RNA structure, which can be revealed by structural biology approaches such as NMR spectroscopy or X-ray crystallography. Three-dimensional structures of RNAs and/or RNA-ligand complexes with atomic resolution can provide insight into possible "hot-spots" to initiate drug discovery efforts. For example, a rational design based on electrostatic potential analysis of the TAR RNA structure resulted in a series of TAR-binding small molecules and the discovery of a low-nM binder with Kd <100 nM (Davis et al., 2004; Murchie et al., 2004); the conjugation of modified nucleobases, which can recognize specific RNA base pairs, and positively charged amino acids, which provide favorable electrostatic interactions, also resulted in TAR-binding small molecules with mid-nM Kd values (Joly et al., 2014), but not drug-like characteristics.

Fragment-based drug design is an approach to develop potent small-molecule compounds from fragments with weak binding activity to a target (μM-mM range), low complexity in chemical structures and low molecular weight (<300 Da) (Congreve et al., 2003; Erlanson et al., 2016; Kirsch et al., 2019; Lundquist et al., 2021). After the identification of initial fragment hits, improvement of binding affinity and optimization can be achieved by fragment linking, merging or growing strategies, which are key approaches for fragment-based drug design (Diethelm-Varela, 2021; Erlanson et al., 2016). Compared to high-throughput screening, fragment-based screening has advantages such as lower experimental cost, more diverse hits, and exhibiting multiple ways to develop novel compounds (Hall et al., 2014; Hann et al., 2001; Kirsch et al., 2019; Osborne et al., 2020). In order to detect weak binders, fragment-based screening techniques require a detection method with high sensitivity, and thus nuclear magnetic resonance (NMR) spectroscopy and x-ray crystallography are more suitable methods than fluorescence (Chilingaryan et al., 2012; Diethelm-Varela, 2021; Harner et al., 2013; Kirsch et al., 2019; Moumne et al., 2012; Schiebel et al., 2016). In addition to its ability to detect a wide variety of binding affinities, NMR can also provide experimental restraints for the determination of target-fragment binding modes and thus has become one of the most frequently used method in fragment-based screening(Diethelm-Varela, 2021; Harner et al., 2013; Moumne et al., 2012).

Section 2.3: RNA Binding Proteins

RNA binding proteins (RBPs) are proteins that are involved in gene regulation by interacting with RNA to form ribonucleoprotein particles (RNPs), which govern the fate of their target RNAs, by regulating transcription, splicing, modification, localization, translation and decay (Gebauer et al., 2021; Gerstberger et al., 2014; Lukong et al., 2008; Mitchell & Parker, 2014), and more. The RBPs do so by binding to specific sequences or structures in RNA, often via structurally well-defined RNA-binding domains such as the RNA recognition motif (RRM), K-homology (KH) domain, dsRBDs and RNA-binding zinc-finger (Agrawal et al., 2019; Clery et al., 2008; Font & Mackay, 2010; Hentze et al., 2018; Lunde et al., 2007; Olejniczak et al., 2022; Valverde et al., 2008). Moreover, the discovery and investigation of lncRNAs, many of which interact with transcription factors or chromatin-remodeling enzymes, provide a notion that RBP function can also be regulated by RNAs (Cech & Steitz, 2014; Hentze et al., 2018). More than 1500 RBPs have been identified in the human proteome through modern large-scale quantitative methods (Gebauer et al., 2021; Gerstberger et al., 2014; Hentze et al., 2018). However, despite the large number of identified RBPs, our understanding of their constituent properties and interaction with RNA is still limited, compared for example to transcription factors.

The RNA recognition motif (RRM) is the most abundant RNA-binding motif in eukaryotes and is known to be involved in many gene regulation steps (Clery et al., 2008; Maris et al., 2005; Venter et al., 2001). RRMs are 80-100 amino acids long, with $\beta1\alpha1\beta2\beta3\alpha2\beta4$ topology that forms a four-stranded antiparallel β-sheet packed against two α-helices (Clery et al., 2008; Maris et al., 2005). Despite the similarity in structure, RRMs are able to recognize a wide variety of RNAs that have different sequence and length (Afroz et al., 2015; Burd & Dreyfuss, 1994; Collins et al., 2017; Gebhardt et al., 2015; Jonas et al., 2020), most commonly through the β-sheet surface, at the center of which are strands β1 and β3 with conserved signature sequences (Clery et al., 2008; Maris et al., 2005). Non-canonical interactions through the α-helices (Clery et al., 2013; Tintaru et al., 2007), protein loop (Auweter et al., 2006; Dominguez et al., 2010; Gebhardt et al., 2015; Skrisovska et al., 2007), and N- or C- terminal extensions (Auweter et al., 2006; Dominguez et al., 2010; Skrisovska et al., 2007) are also found in RRM-RNA complexes, to boost affinity or specificity of individual domains and even to redirect certain RRMs to bind to proteins.

Several RBPs regulate miRNA processing by binding at or near the apical region of the stem-loop precursors species, which is common to both pre- and pri- miRNAs (Michlewski et al., 2008; Newman et al., 2008). One example is provided by the highly conserved RRMs of Rbfox proteins (Rbfox2 RRM), which inhibits the maturation of miR-20b and miR-107 by binding to the terminal loop of the precursor miRNAs (Chen et al., 2016). The RRM of Rbfox proteins targets RNA specifically through the conserved (U)GCAUG element, a well-known cis-element for the regulation of alternative splicing (Auweter et al., 2006; Jin et al., 2003) which is found in the apical loop of pre-miR-20b and pre-miR-107 (Chen et al., 2016).

By binding to the apical loop, the Rbfox2 RRM inhibits processing of the two pri-miRNAs both *in vitro* and *in vivo*, and thus increases expression of the downstream tumor suppressor PTEN (Phosphatase and Tensin Homolog) (Chen et al., 2016; Zhou et al., 2014; Zhu et al., 2014).

Section 2.4: Peptide Mimetics

In addition to the intrinsic interest in protein engineering and evolution of RNA-binding protein, peptides provide much more advantageous starting points than proteins to discover inhibitors of RNA functions (Craik et al., 2013; Sachdeva et al., 2016; Sakamoto et al., 2017; Walker & Varani, 2019). Peptides possess a bigger interacting surface than small molecules, which improves specificity and binding activity, but are smaller than proteins and cheaper to manufacture and deliver (Craik et al., 2013; Jing & Jin, 2020). Compared to linear peptides, cyclic peptides often have better binding affinity, specificity and stability due to conformational rigidity and are less susceptible to proteolysis within cells and organisms (Gongora-Benitez et al., 2014; Jing & Jin, 2020).

Different approaches have been used to discover RNA binding peptides. High-throughput screening technologies, which enable the rapid evaluation of interactions between RNA and peptide on a large scale, is one of the methods (Fernandes et al., 2009; Hamzeh-Mivehroud et al., 2013; Henninot et al., 2018; Hilpert et al., 2009). A success of quantitative peptide microarray development was demonstrated by immobilizing hydrazide-linked peptides onto a epoxide-derivatized glass slide (Pai et al., 2012). RNA-peptide bindings were probed by RNA 5'-end fluorophore-labeling. Using this method, the group was able to profile interactions of six hairpin RNAs (IRES of domain IV of HCV, HIV RRE, HIV TAR, IRE, and thymidylate synthase mRNA) with 111 peptides, containing systematic amino acid replacements or deletions. Another success in discovery of RNA binding ligands occurred by using peptoid microarrays (Chirayil et al., 2009). Cysteine-terminated peptoids were printed on a N-(p-maleimidophenyl)isocynate pretreated surface through thiol-maleimide click chemistry. By screening peptoid microarrays, a peptoid binding to pre-miR21 with a Kd of 1.9 μM was identified. However, a disadvantage of microarray technologies is the potential for false negatives due to the masking of the peptide by the surface attachment.

Structure-based peptidomimetics were designed through a strategy where a protein-bound structure provides a rational to design a minimal peptide that mimics the activity of the full protein. The protein binding site structure and its interactions with the target molecule provides a starting point of the peptide design. Design of peptidomimetics binding to the HIV TAR represents a successful example of this strategy (Athanassiou et al., 2004; Athanassiou et al., 2007; Davidson et al., 2009; Davidson et al., 2011). Inspired by the structure of the TAR recognition Tat Arginine Rich Motif (ARM) in the BIV (bovine immunodeficiency virus) protein, which interacts with the major groove at the apical stem-loop junction of BIV TAR through a β-hairpin, peptide mimetics were made by grafting the protein sequence onto a β-hairpin inducing cyclic peptide scaffold. This led to the discovery of peptide BIV2 which binds to BIV TAR with a dissociation constant (Kd) of 0.15 μM (Athanassiou et al., 2004). Optimization of its binding to HIV TAR was carried out by screening small-to-medium sized peptide libraries, resulting eventually in the discovery of JB-181, a 30 pM binding ligand for HIV TAR (Shortridge et al., 2018). The Varani Group has also successfully targeted cancer related pre-miRNAs by using cyclic β-hairpin peptides, which will be described in Chapter 2.

A "semi-design" strategy using evolution and selection for RNA-binding peptide discovery was recently proposed (Belashov et al., 2018). Laboratory evolution is used to alter putative RNA-binding amino acids within a known protein. Thus, an RRM protein library was created by performing saturation mutagenesis in RRM1 and β2-β3 loop of U1A protein, then subjected to yeast-display screening. Desired TAR-binding proteins (TBPs) were identified by flow cytometry. Variant TBPs which recognize TAR through evolved amino acids in β2-β3 loop and possess nano-molar binding affinity were discovered. To find the minimal RNA binding module, short peptide mimics were then designed, which were stabilized by a side-chain disulfide linkage and comprised the lab-evolved loop sequence. The resulting peptides retained low-μM TAR binding affinity and demonstrated the ability to inhibit Tat-dependent transcription in cells.

Section 3: Structural analysis by NMR Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool to study biomolecule structure and is largely used in Varani group to investigate structures of proteins, peptides, small molecules and RNAs, and of their complexes.

The use of 2D NMR in structure determination of proteins was first demonstrated by the group of Kurt Wüthrich, and in 1985, the first structure of a small globular protein established using NMR spectroscopy was published (Williamson et al., 1985). With new technological developments, the credibility of NMR as structural tool for biomolecules has strengthened over the past few decades: the higher magnetic fields which are reached using new superconducting material improve spectra resolution; the use of cryoprobes where transmitter and receiver coils are maintained at low temperature to improve signal-to-noise ratio; and the development of multi-dimensional NMR using isotope-labeled samples (Marion, 2013). In addition, NMR can provide information on dynamics over time scales which is very difficult to investigate by crystallography. In the field of drug discovery, chemical shift mapping provides information on which part of the biomolecule is interacting with the ligand and NMR is a very powerful tool for screening or hit optimization (Diethelm-Varela, 2021; Harner et al., 2013; Moumne et al., 2012).

This section is an introduction to peptide and RNA NMR spectroscopy; additional experimental details are provided in Chapter 2 and Chapter 3.

Section 3.1: Total Correlation Spectroscopy (TOCSY)

In NMR spectroscopy, J coupling (spin-spin coupling) arises from indirect interactions between two nuclear spins, mediated by interactions between the nuclear spin and local electronic currents. The value of J-coupling is affected by the bond distance and bond angle, and thus can provide information about stereochemistry of a molecule, as well as conformation.

Total correlation spectroscopy (TOCSY) allows the observation of scalar correlations between all protons within a spin system. In TOCSY experiments, the evolution of the chemical shift is suppressed by an intense radio-frequency pulse along a selected direction, allowing magnetizations to remain aligned with the field, which is called the spin-lock field. The two most common pulse sequences used to generate the spin-lock field are DIPSI (Shaka et al., 1988) and MLEV (Bax & Davis, 1985). This enables the spin system to emerge only under the J-coupling and allows magnetization to be transferred as long as protons are coupled. For a long spin-lock time such as 80ms, magnetization can be transferred through up to 5 or 6 bonds, generating cross-peaks between all protons in the same spin system, unless being interrupted by small or zero proton-proton couplings along the spin system.

Section 3.2: Nuclear Overhauser Effect Spectroscopy (NOESY)

The Overhauser effect was first reported by Albert Overhauser as a polarization of nuclear spins in metals on saturation of electrons to which they are coupled (Overhauser, 1953). The Nuclear Overhauser effect (NOE) is a similar effect observed between nuclei, when nearby nuclear spins are saturated; then, the nuclear spin polarization of one nucleus can transfer to another nucleus within 5 Å through cross-relaxation, resulting in changes in the intensity of its NMR signal. The NOE effect is strongly dependent upon the distance between nuclei, and thus it is a key effect for obtaining three-dimensional structures of molecules in solution by NMR spectroscopy.

Nuclear Overhauser Effect Spectroscopy (NOESY) is an NMR spectroscopic method that identifies NOE interactions in molecules which can be used to measure internuclear distances. The cross peaks observed in NOESY spectra identify protons that are close to each other in space, providing spatial information, and is distinct from TOCSY which relies on J-couplings to provide spin-spin correlations.

Section 3.3: Peptide and Protein NMR Spectroscopy

NMR spectroscopy is a powerful tool to study proteins and peptides structure and dynamics in solution (Bax & Clore, 2019; Hu et al., 2021; Williamson, 1993; Wuthrich, 1990). A large number of NMR structures have been reported in the past several decades. The main constituents of biomolecules, e.g. hydrogen, carbon, nitrogen, have at least one NMR-active isotope. Thus, isotope-labeling can be used to obtain additional structural information for even complex structures.

NMR can provide evidence for the presence of stable secondary structure in peptides in aqueous solution, even for flexible peptides undergoing rapid interconversion between multiple low-energy conformations (Mielke & Krishnan, 2009; Wuthrich et al., 1984). In polypeptide structures, residues are separated by amide bonds, so correlations between HN, Hα and side chain protons can be observed in TOCSY spectra for each amino acid. Because the 20 amino acids have distinct side chain structures, each amino acid has a characteristic TOCSY pattern which allows them to be distinguished from each other in NMR spectra. For example, Gly is the only amino acid with 2 Hα protons, and the Hα chemical shifts are relatively upfield compared to other amino acids because of its lack of a side chain. Side chain protons in non-polar amino acid side chains, such as Ala and Leu, are usually upfield shifted; in contrast, side chain protons in polar and charged amino acids, such as Thr and Asp, are more downfield shifted because of the presence of electron-withdrawing group. Side chain HN in Arg and Lys (~7.3 ppm) and their backbone HN (~8.3 ppm) shows the same TOCSY patterns to side chain protons, making them easy to be recognized.

Characteristic NOE patterns and chemical shifts were identified for different secondary structures such as α-helix, β-sheet and β-turns (Mielke & Krishnan, 2009; Wishart et al., 1991; Wuthrich et al., 1984). For example, strong sequential iHN-(i+1)HN NOEs and iHα-(i+1)HN NOEs are used to establish sequential connectivity in α-helix and β-strand, respectively. Characteristic iHα-(i+4)HN and iHα-(i+3)Hβ NOEs are only found in α-helix because of its spiral structure; cross-strand HN-HN NOEs can be found in β-hairpins or β-sheets. Chemical shifts of Hα can also provide an additional validation for peptide and protein secondary structure. An α-helix often presents more upfield-shifted Hα chemical shifts compared to random-coil structures, while β-strand presents more downfield-shifted Hα chemical shifts (Mielke & Krishnan, 2009; Wishart et al., 1991).

Section 3.4: RNA and NMR Spectroscopy

Because of the increasing emphasis on the role of RNA role in gene expression and regulation, more efforts have been put into the investigation of its structure and function. High resolution X-ray crystal structures of various RNAs have been published, such as tRNA (Byrne et al., 2010; Chan et al., 2020), rRNAs (Ban et al., 2000; Vicens & Westhof, 2003; Yusupov et al., 2001), riboswitches (Edwards & Ferre-D'Amare, 2006; Garst et al., 2008; Lu et al., 2008) and disease-related viral RNAs (Hood et al., 2019; Ippolito & Steitz, 1998; Lee et al., 2021). However, the often-elongated structure and intrinsic flexibility of RNA makes crystallization often challenging or sometimes results in poor resolution at binding interfaces. Solution-state NMR spectroscopy is a suitable tool to study the structure and dynamics of flexible RNA molecules in solution with easier sample preparation requirements than X-ray crystallography (Barnwal et al., 2017; Marusic et al., 2019; Varani et al., 1996). However, NMR of RNA is also challenging because of severe spectral overlap resulting from the smaller chemical dispersion in nucleic acids. Compared to proteins which are composed of 20 different amino acids, RNAs are composed of only four different nucleosides, making unambiguous resonance assignments much more difficult to achieve (Furtig et al., 2003). In addition, RNA molecules often form extended structures, resulting in slower overall tumbling times and shorter transverse relaxation times (T2), causing increased linewidth which worsen the spectral overlap and prevent many advanced experiments from working as designed. These challenges become severe for RNAs that contain more than 40-50 nucleotides. Nevertheless, introducing isotope labeling can effectively reduce the complexity of NMR spectra and allow unambiguous assignments (Furtig et al., 2003; Varani et al., 1996).

Three-dimensional structural information essential for structure determination can be obtained through NMR experiments. Detailed explanation for the assignment process is covered in (Varani et al., 1996; Varani & Tinoco, 1991). Distance information between protons separated by less than 5 Å can be obtained from nuclear Overhauser effect spectroscopy (NOESY), and torsion angles from scalar couplings.

The assignments of aromatic and sugar protons from double-stranded helical regions are generally easier compared to single-stranded loops or mismatched base pair. The NOE patterns for A-form helical RNA structures is well-established. NOE interactions within Watson-Crick base pairs can be identified in H_2O NOESY spectra: strong NOEs are observable from Ade H2 to Ura imino protons, and Gua NH1 to Cyt amino protons. Sequential intrastrand connectivity can be established by aromatic-to-sugar or aromatic-to-aromatic proton NOEs. Aromatic protons within each base (H6/8) exhibit weak NOEs with intra-residue H1’and H2’and the preceding (i-1) H1’protons, while the intra-residue H6/8-H3’NOEs and sequential iH6/8-(i-1)H2’NOEs are often strong. Sequential aromatic-to-aromatic NOEs (H5 to H5 and H6/H8 to H6/H8) are observable at long mixing times. These patterns can be used if intrastrand base stacking is present, regardless of the presence or absence of base pairing.

The RNA sugar conformation and dynamics can be investigated through TOCSY. For 3'-endo sugar pucker, the H1'-H2 'coupling is very small, so generally no correlation between H1 'and other sugar protons is observable in TOCSY. The 2'-endo conformation has strong H1'-H2 'couplings, medium H2'-H3 'couplings and very weak H3'-H4 'couplings. Thus, strong H1'-H2', medium H1'-H3 'and weak/unobservable H1'-H4 'peaks are observed in TOCSY spectra for 2'-endo sugar pucker. For a mixture of 3 'and 2'-endo conformations, which could be a sign of dynamics, peaks from H1 'to H2', H3 'and H4 'are all observable.

Distance and torsion angle constraints determined from experimental NMR data are then used in simulated annealing protocols for structure determination, which is introduced in the next section. Compared to the structure determination of proteins of similar size, RNA structures often have a lower density of distance constraints due to their elongated helical structures, which results in a lower precision of the structure (Allain & Varani, 1997).

Section 3.5: Structure Determination

Xplor-NIH is a popular software package for biomolecular structure determination from NMR experimental data sources combined with known geometry and stereochemistry. Structure determination for polypeptides and RNA molecules can be achieved by seeking the minimum of a target energy function containing terms for experimental restraints, covalent geometry and non-bonded contacts using simulated annealing protocols and torsion angle dynamics. The energy function is a sum of terms for experimental, chemical and knowledge-based restraints. Experimental restraints include NOE-derived distance restraints and dihedral angle restraints obtained from chemical shift or scalar coupling data. Knowledge-based terms restrain the structure based on statistical trends observed in structural databases.

The interactive process of structure determination begins once a majority of NMR-derived conformational constraints are established confidently from the analysis of NOESY spectra. Distance restraints are derived from NOE assignments in 2D and 3D NOESY spectra which are sorted into strong (2.5 ±0.7 Å), medium (3.5 ±1.2 Å) and weak (4.5 ±1.5 Å) bins based on peak intensities, relative to fixed distances, for example, aromatic H5-H6 = 2.5 Å in Uri and Cyt for RNA, and aromatic Hδ-Hε = 2.5 Å in Phe and Try for polypeptide.

For RNA structure determination, base-pair planarity and hydrogen-bonding restraints are used for unambiguously established base pairs as identified by 2D NOESY involving NH base protons. Hydrogen bond, planarity and dihedral restraints are included for base-paired nucleotides surrounded by base pairs conforming to A-form helical structures, as established from the pattern of NOE cross-peaks (Varani et al., 1996; Varani & Tinoco, 1991). Similarly, hydrogen-bonding restraints are also used in peptide structure determination for secondary structure such as α-helix or β-sheet identified by characteristic NOE patterns and chemical shifts, as introduced in Section 3.3.

Conformational constraints, including distance, torsion angle and planarity constraints, are input into the NIH-XPLOR package for biomolecular structure determination (Schwieters et al., 2003). Compiled experimental restraints are then used in a simulated annealing procedure, initially undergoing high-temperature (2,500K to 298K) torsional angle dynamics, where incremental decreases in temperature are generated by progressively introducing Van der Waals terms and increasing force constants for angles, dihedral angles, NOEs, and the Van der Waals repulsive term. After the final cooling step, the RNA undergoes two sequential refinement steps, first in torsional angle space then in Cartesian space. Large number of structures are generated (200-1000 models) in the initial calculations to produce a reliable ensemble of converged structures. Restrain violations, if they occur, are then reviewed and reevaluation of spectral assignments and constraint binning are carried out. The lowest total energy structure calculated by the Xplor energy function is further refined by an extended simulated annealing calculation against the constraints listed above to generate 200-400 structures. The 10 structures with the lowest total energy, based on the Xplor energy function, were chosen as the representative structures to calculate RMSD, and structure quality analysis is conducted using servers such as MolProbity (Williams et al., 2018). Scripts and commands used for structure calculation are reported in the Appendix.

Section 4: References

Aboul-ela F. (2010). Strategies for the design of RNA-binding small molecules. *Future Med. Chem.*, *2*(1), 93-119. https://doi.org/10.4155/fmc.09.149

Abulwerdi F. A., Xu W., Ageeli A. A., Yonkunas M. J., Arun G., Nam H., Schneekloth J. S., Dayie T. K., Spector D., Baird N., et al. (2019). Selective Small-Molecule Targeting of a Triple Helix Encoded by the Long Noncoding RNA, MALAT1. *ACS Chem. Biol.*, *14*(2), 223-235. https://doi.org/10.1021/acschembio.8b00807

Afroz T., Cienikova Z., Clery A., & Allain F. H. (2015). One, Two, Three, Four! How Multiple RRMs Read the Genome Sequence. *Methods Enzymol.*, *558*, 235-278. https://doi.org/10.1016/bs.mie.2015.01.015

Agrawal N., Dasaradhi P. V., Mohmmed A., Malhotra P., Bhatnagar R. K., & Mukherjee S. K. (2003). RNA interference: biology, mechanism, and applications. *Microbiol. Mol. Biol. Rev.*, *67*(4), 657-685. https://doi.org/10.1128/MMBR.67.4.657-685.2003

Agrawal S., Kuo P. H., Chu L. Y., Golzarroshan B., Jain M., & Yuan H. S. (2019). RNA recognition motifs of disease-linked RNA-binding proteins contribute to amyloid formation. *Sci. Rep.*, *9*(1), 6171. https://doi.org/10.1038/s41598-019-42367-8

Alanazi A., Ivanov A., Kumari N., Lin X., Wang S., Kovalskyy D., & Nekhai S. (2021). Targeting Tat-TAR RNA Interaction for HIV-1 Inhibition. *Viruses*, *13*(10). https://doi.org/10.3390/v13102004

Allain F. H. T., & Varani G. (1997). How accurately and precisely can RNA structure be determined by NMR? *J. Mol. Biol.*, *267*(2), 338-351. https://doi.org/10.1006/jmbi.1996.0855

Alvarez D. E., De Lella Ezcurra A. L., Fucito S., & Gamarnik A. V. (2005). Role of RNA structures present at the 3'UTR of dengue virus on translation, RNA synthesis, and viral replication. *Virology*, *339*(2), 200-212. https://doi.org/10.1016/j.virol.2005.06.009

Alvarez D. E., Lodeiro M. F., Luduena S. J., Pietrasanta L. I., & Gamarnik A. V. (2005). Long-range RNA-RNA interactions circularize the dengue virus genome. *J. Virol.*, *79*(11), 6631-6643. https://doi.org/10.1128/JVI.79.11.6631-6643.2005

Athanassiou Z., Dias R. L., Moehle K., Dobson N., Varani G., & Robinson J. A. (2004). Structural mimicry of retroviral tat proteins by constrained beta-hairpin peptidomimetics: ligands with high affinity and selectivity for viral TAR RNA regulatory elements. *J. Am. Chem. Soc.*, *126*(22), 6906-6913. https://doi.org/10.1021/ja0497680

Athanassiou Z., Patora K., Dias R. L., Moehle K., Robinson J. A., & Varani G. (2007). Structure-guided peptidomimetic design leads to nanomolar beta-hairpin inhibitors of the Tat-TAR interaction of bovine immunodeficiency virus. *Biochemistry*, *46*(3), 741-751. https://doi.org/10.1021/bi0619371

Auweter S. D., Fasan R., Reymond L., Underwood J. G., Black D. L., Pitsch S., & Allain F. H. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.*, *25*(1), 163-173. https://doi.org/10.1038/sj.emboj.7600918

Baba M. (2006). Recent status of HIV-1 gene expression inhibitors. *Antiviral Res.*, *71*(2-3), 301-306. https://doi.org/10.1016/j.antiviral.2006.01.002

Baek D., Villen J., Shin C., Camargo F. D., Gygi S. P., & Bartel D. P. (2008). The impact of microRNAs on protein output. *Nature*, *455*(7209), 64-71. https://doi.org/10.1038/nature07242

Ban N., Nissen P., Hansen J., Moore P. B., & Steitz T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, *289*(5481), 905-920. https://doi.org/10.1126/science.289.5481.905

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. https://doi.org/10.1016/j.abb.2017.06.003

Barros S. A., Yoon I., & Chenoweth D. M. (2016). Modulation of the E. coli rpoH Temperature Sensor with Triptycene-Based Small Molecules. *Angew. Chem. Int. Ed.*, *55*(29), 8258-8261. https://doi.org/10.1002/anie.201601626

Bartel D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, *136*(2), 215-233. https://doi.org/10.1016/j.cell.2009.01.002

Bax A., & Clore G. M. (2019). Protein NMR: Boundless opportunities. *J. Magn. Reson.*, *306*, 187-191. https://doi.org/10.1016/j.jmr.2019.07.037

Bax A., & Davis D. G. (1985). Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *J. Magn. Reson.*, *65*(2), 355-360. https://doi.org/10.1016/0022-2364(85)90018-6

Belashov I. A., Crawford D. W., Cavender C. E., Dai P., Beardslee P. C., Mathews D. H., Pentelute B. L., McNaughton B. R., & 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. https://doi.org/10.1093/nar/gky529

Bennett C. F. (2019). Therapeutic Antisense Oligonucleotides Are Coming of Age. *Annu. Rev. Med.*, *70*, 307-321. https://doi.org/10.1146/annurev-med-041217-010829

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. https://doi.org/10.1038/nrd4505

Binas O., de Jesus V., Landgraf T., Volklein A. E., Martins J., Hymon D., Bains J. K., Berg H., Biedenbander T., Furtig B., et al. (2021). F-19 NMR-Based Fragment Screening for 14 Different Biologically Active RNAs and 10 DNA and Protein Counter-Screens. *Chembiochem*, *22*(2), 423-433. https://doi.org/10.1002/cbic.202000476

Borgelt L., Li F., Hommen P., Lampe P., Hwang J. M., Goebel G. L., Sievers S., & Wu P. (2021). Trisubstituted Pyrrolinones as Small-Molecule Inhibitors Disrupting the Protein-RNA Interaction of LIN28 and Let-7. *ACS Med. Chem. Lett.*, *12*(6), 893-898. https://doi.org/10.1021/acsmedchemlett.0c00546

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.*, *51*(4), 1019-1023. https://doi.org/10.1002/anie.201106455

Breaker R. R. (2012). Riboswitches and the RNA World. *Cold Spring Harb. Perspect. Biol.*, *4*(2). https://doi.org/10.1101/cshperspect.a003566

Breaker R. R., & Joyce G. F. (2014). The expanding view of RNA and DNA function. *Chem. Biol.*, *21*(9), 1059-1065. https://doi.org/10.1016/j.chembiol.2014.07.008

Broft P., Dzatko S., Krafcikova M., Wacker A., Hansel-Hertsch R., Dotsch V., Trantirek L., & Schwalbe H. (2021). In-Cell NMR Spectroscopy of Functional Riboswitch Aptamers in Eukaryotic Cells. *Angew. Chem. Int. Ed.*, *60*(2), 865-872. https://doi.org/10.1002/anie.202007184

Burd C. G., & Dreyfuss G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science*, *265*(5172), 615-621. https://www.ncbi.nlm.nih.gov/pubmed/8036511

Byrne R. T., Konevega A. L., Rodnina M. V., & Antson A. A. (2010). The crystal structure of unmodified tRNA(Phe) from Escherichia coli. *Nucleic Acids Res.*, *38*(12), 4154-4162. https://doi.org/10.1093/nar/gkq133

Cantile M., Di Bonito M., De Bellis M. T., & Botti G. (2021). Functional Interaction among lncRNA HOTAIR and MicroRNAs in Cancer and Other Human Diseases. *Cancers*, *13*(3). https://doi.org/10.3390/cancers13030570

Carnevali M., Parsons J., Wyles D. L., & Hermann T. (2010). A modular approach to synthetic RNA binders of the hepatitis C virus internal ribosome entry site. *Chembiochem*, *11*(10), 1364-1367. https://doi.org/10.1002/cbic.201000177

Castilla-Llorente V., Nicastro G., & Ramos A. (2013). Terminal loop-mediated regulation of miRNA biogenesis: selectivity and mechanisms. *Biochem. Soc. Trans.*, *41*(4), 861-865. https://doi.org/10.1042/BST20130058

Castro-Oropeza R., Melendez-Zajgla J., Maldonado V., & Vazquez-Santillan K. (2018). The emerging role of lncRNAs in the regulation of cancer stem cells. *Cell. Oncol.*, *41*(6), 585-603. https://doi.org/10.1007/s13402-018-0406-4

Cech T. R., & Steitz J. A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*, *157*(1), 77-94. https://doi.org/10.1016/j.cell.2014.03.008

Chan C. W., Badong D., Rajan R., & Mondragon A. (2020). Crystal structures of an unmodified bacterial tRNA reveal intrinsic structural flexibility and plasticity as general properties of unbound tRNAs. *RNA*, *26*(3), 278-289. https://doi.org/10.1261/rna.073478.119

Chan J. A., Krichevsky A. M., & Kosik K. S. (2005). MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res.*, *65*(14), 6029-6033. https://doi.org/10.1158/0008-5472.CAN-05-0137

Chan J. H., Lim S., & Wong W. S. (2006). Antisense oligonucleotides: from design to therapeutic application. *Clin. Exp. Pharmacol. Physiol.*, *33*(5-6), 533-540. https://doi.org/10.1111/j.1440-1681.2006.04403.x

Chapman R. L., Stanley T. B., Hazen R., & Garvey E. P. (2002). Small molecule modulators of HIV Rev/Rev response element interaction identified by random screening. *Antiviral Res.*, *54*(3), 149-162. https://doi.org/10.1016/s0166-3542(01)00222-4

Chavali S. S., Bonn-Breach R., & Wedekind J. E. (2019). Face-time with TAR: Portraits of an HIV-1 RNA with diverse modes of effector recognition relevant for drug discovery. *J. Biol. Chem.*, *294*(24), 9326-9341. https://doi.org/10.1074/jbc.REV119.006860

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-U497. https://doi.org/10.1038/nature05769

Chen C. Y. A., & Shyu A. B. (1995). Au-Rich Elements - Characterization and Importance in Messenger-Rna Degradation. *Trends Biochem. Sci.*, *20*(11), 465-470. https://doi.org/10.1016/S0968-0004(00)89102-1

Chen J. H., Zhang R. H., Lin S. L., Li P. F., Lan J. J., Song S. S., Gao J. M., Wang Y., Xie Z. J., Li F. C., et al. (2018). The Functional Role of the 3' Untranslated Region and Poly(A) Tail of Duck Hepatitis A Virus Type 1 in Viral Replication and Regulation of IRES-Mediated Translation. *Front. Microbiol.*, *9*, 2250. https://doi.org/10.3389/fmicb.2018.02250

Chen L., Qian X., Wang Z., & Zhou X. (2021). The HOTAIR lncRNA: A remarkable oncogenic promoter in human cancer metastasis. *Oncol. Lett.*, *21*(4), 302. https://doi.org/10.3892/ol.2021.12563

Chen Y., Zubovic L., Yang F., Godin K., Pavelitz T., Castellanos J., Macchi P., & Varani G. (2016). Rbfox proteins regulate microRNA biogenesis by sequence-specific binding to their precursors and target downstream Dicer. *Nucleic Acids Res.*, *44*(9), 4381-4395. https://doi.org/10.1093/nar/gkw177

Chery J. (2016). RNA therapeutics: RNAi and antisense mechanisms and clinical applications. *Postdoc J.*, *4*(7), 35-50. https://doi.org/10.14304/surya.jpr.v4n7.5

Cheung A. K., Hurley B., Kerrigan R., Shu L., Chin D. N., Shen Y., O'Brien G., Sung M. J., Hou Y., Axford J., et al. (2018). Discovery of Small Molecule Splicing Modulators of Survival Motor Neuron-2 (SMN2) for the Treatment of Spinal Muscular Atrophy (SMA). *J. Med. Chem.*, *61*(24), 11021-11036. https://doi.org/10.1021/acs.jmedchem.8b01291

Chilingaryan Z., Yin Z., & Oakley A. J. (2012). Fragment-based screening by protein crystallography: successes and pitfalls. *Int. J. Mol. Sci.*, *13*(10), 12857-12879. https://doi.org/10.3390/ijms131012857

Chirayil S., Chirayil R., & Luebke K. J. (2009). Discovering ligands for a microRNA precursor with peptoid microarrays. *Nucleic Acids Res.*, *37*(16), 5486-5497. https://doi.org/10.1093/nar/gkp549

Choi K. H. (2021). The Role of the Stem-Loop A RNA Promoter in Flavivirus Replication. *Viruses*, *13*(6). https://doi.org/10.3390/v13061107

Choudhury N. R., & Michlewski G. (2012). Terminal loop-mediated control of microRNA biogenesis. *Biochem. Soc. Trans.*, *40*(4), 789-793. https://doi.org/10.1042/BST20120053

Clery A., Blatter M., & Allain F. H. (2008). RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, *18*(3), 290-298. https://doi.org/10.1016/j.sbi.2008.04.002

Clery A., Sinha R., Anczukow O., Corrionero A., Moursy A., Daubner G. M., Valcarcel J., Krainer A. R., & Allain F. H. (2013). Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(30), E2802-2811. https://doi.org/10.1073/pnas.1303445110

Colavita F., Vairo F., Carletti F., Boccardo C., Ferraro F., Iaiani G., Al Moghazi S., Galardo G., Lalle E., Selvaggi C., et al. (2020). Full-length genome sequence of a dengue serotype 1 virus isolate from a traveler returning from Democratic Republic of Congo to Italy, July 2019. *Int. J. Infect. Dis.*, *92*, 46-48. https://doi.org/10.1016/j.ijid.2019.12.023

Collins K. M., Kainov Y. A., Christodolou E., Ray D., Morris Q., Hughes T., Taylor I. A., Makeyev E. V., & Ramos A. (2017). An RRM-ZnF RNA recognition module targets RBM10 to exonic sequences to promote exon exclusion. *Nucleic Acids Res.*, *45*(11), 6761-6774. https://doi.org/10.1093/nar/gkx225

Congreve M., Carr R., Murray C., & Jhoti H. (2003). A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today*, *8*(19), 876-877. https://doi.org/10.1016/s1359-6446(03)02831-9

Cooper T. A., Wan L., & Dreyfuss G. (2009). RNA and disease. *Cell*, *136*(4), 777-793. https://doi.org/10.1016/j.cell.2009.02.011

Coppins R. L., Hall K. B., & Groisman E. A. (2007). The intricate world of riboswitches. *Curr. Opin. Microbiol.*, *10*(2), 176-181. https://doi.org/10.1016/j.mib.2007.03.006

Craik D. J., Fairlie D. P., Liras S., & Price D. (2013). The future of peptide-based drugs. *Chem. Biol. Drug Des.*, *81*(1), 136-147. https://doi.org/10.1111/cbdd.12055

Croce C. M. (2009). Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, *10*(10), 704-714. https://doi.org/10.1038/nrg2634

Dang T. T., Pham M. H., Bui H. V., & Van Le D. (2020). Whole genome sequencing and genetic variations in several dengue virus type 1 strains from unusual dengue epidemic of 2017 in Vietnam. *Virol. J.*, *17*(1), 7. https://doi.org/10.1186/s12985-020-1280-z

Dann C. E., Wakeman C. A., Sieling C. L., Baker S. C., Irnov I., & Winkler W. C. (2007). Structure and mechanism of a metal-sensing regulatory RNA. *Cell*, *130*(5), 878-892. https://doi.org/10.1016/j.cell.2007.06.051

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. https://doi.org/10.1073/pnas.0900629106

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. https://doi.org/10.1093/nar/gkq713

Davila-Calderon J., Patwardhan N. N., Chiu L. Y., Sugarman A., Cai Z. G., Penutmutchu S. R., Li M. L., Brewer G., Hargrove A. E., & Tolbert B. S. (2020). IRES-targeting small molecule inhibits enterovirus 71 replication via allosteric stabilization of a ternary complex. *Nat. Commun.*, *11*(1). https://doi.org/10.1038/s41467-020-18594-3

Davis B., Afshar M., Varani G., Murchie A. I., Karn J., Lentzen G., Drysdale M., Bower J., Potter A. J., Starkey I. D., et al. (2004). Rational design of inhibitors of HIV-1 TAR

RNA through the stabilisation of electrostatic "hot spots". *J. Mol. Biol.*, *336*(2), 343-356. https://doi.org/10.1016/j.jmb.2003.12.046

Derrien T., Johnson R., Bussotti G., Tanzer A., Djebali S., Tilgner H., Guernec G., Martin D., Merkel A., Knowles D. G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.*, *22*(9), 1775-1789. https://doi.org/10.1101/gr.132159.111

Di Fusco D., Dinallo V., Marafini I., Figliuzzi M. M., Romano B., & Monteleone G. (2019). Antisense Oligonucleotide: Basic Concepts and Therapeutic Application in Inflammatory Bowel Disease. *Front. Pharmacol.*, *10*, 305. https://doi.org/10.3389/fphar.2019.00305

Di Leva G., Garofalo M., & Croce C. M. (2014). MicroRNAs in cancer. *Annu. Rev. Pathol.*, *9*, 287-314. https://doi.org/10.1146/annurev-pathol-012513-104715

Dibrov S. M., Parsons J., Carnevali M., Zhou S., Rynearson K. D., Ding K., Garcia Sega E., Brunn N. D., Boerneke M. A., Castaldi M. P., et al. (2014). Hepatitis C virus translation inhibitors targeting the internal ribosomal entry site. *J. Med. Chem.*, *57*(5), 1694-1707. https://doi.org/10.1021/jm401312n

Diethelm-Varela B. (2021). Using NMR Spectroscopy in the Fragment-Based Drug Discovery of Small-Molecule Anticancer Targeted Therapies. *Chemmedchem*, *16*(5), 725-742. https://doi.org/10.1002/cmdc.202000756

Ding K., Wang A., Boerneke M. A., Dibrov S. M., & Hermann T. (2014). Aryl-substituted aminobenzimidazoles targeting the hepatitis C virus internal ribosome entry site. *Bioorg. Med. Chem. Lett.*, *24*(14), 3113-3117. https://doi.org/10.1016/j.bmcl.2014.05.009

DiStefano J. K. (2018). The Emerging Role of Long Noncoding RNAs in Human Disease. *Methods Mol. Biol.*, *1706*, 91-110. https://doi.org/10.1007/978-1-4939-7471-9_6

Dominguez C., Fisette J. F., Chabot B., & Allain F. H. (2010). Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.*, *17*(7), 853-861. https://doi.org/10.1038/nsmb.1814

Edwards T. E., & Ferre-D'Amare A. R. (2006). Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure*, *14*(9), 1459-1468. https://doi.org/10.1016/j.str.2006.07.008

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. https://doi.org/10.1038/nrd.2016.109

Esteller M. (2011). Non-coding RNAs in human disease. *Nat. Rev. Genet.*, *12*(12), 861-874. https://doi.org/10.1038/nrg3074

Eydoux C., Fattorini V., Shannon A., Le T. T., Didier B., Canard B., & Guillemot J. C. (2021). A fluorescence-based high throughput-screening assay for the SARS-CoV RNA synthesis complex. *J. Virol. Methods*, *288*, 114013. https://doi.org/10.1016/j.jviromet.2020.114013

Fabian M. R., Sonenberg N., & Filipowicz W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.*, *79*, 351-379. https://doi.org/10.1146/annurev-biochem-060308-103103

Falese J. P., Donlic A., & Hargrove A. E. (2021). Targeting RNA with small molecules: from fundamental principles towards the clinic. *Chem. Soc. Rev.*, *50*(4), 2224-2243. https://doi.org/10.1039/d0cs01261k

Feng Y. H., & Tsao C. J. (2016). Emerging role of microRNA-21 in cancer. *Biomed. Rep.*, *5*(4), 395-402. https://doi.org/10.3892/br.2016.747

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. https://doi.org/10.1016/j.tibtech.2009.02.009

Filomatori C. V., Iglesias N. G., Villordo S. M., Alvarez D. E., & Gamarnik A. V. (2011). RNA sequences and structures required for the recruitment and activity of the dengue virus polymerase. *J. Biol. Chem.*, *286*(9), 6929-6939. https://doi.org/10.1074/jbc.M110.162289

Filomatori C. V., Lodeiro M. F., Alvarez D. E., Samsa M. M., Pietrasanta L., & Gamarnik A. V. (2006). A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev.*, *20*(16), 2238-2249. https://doi.org/10.1101/gad.1444206

Font J., & Mackay J. P. (2010). Beyond DNA: zinc finger domains as RNA-binding modules. *Methods Mol. Biol.*, *649*, 479-491. https://doi.org/10.1007/978-1-60761-753-2_29

Friedlander M. R., Lizano E., Houben A. J., Bezdan D., Banez-Coronel M., Kudla G., Mateu-Huertas E., Kagerbauer B., Gonzalez J., Chen K. C., et al. (2014). Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biol.*, *15*(4), R57. https://doi.org/10.1186/gb-2014-15-4-r57

Friedman R. C., Farh K. K., Burge C. B., & Bartel D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, *19*(1), 92-105. https://doi.org/10.1101/gr.082701.108

Fu Z. W., Wang L., Li S. J., Chen F., Au-Yeung K. K. W., & Shi C. (2021). MicroRNA as an Important Target for Anticancer Drug Development. *Front. Pharmacol.*, *12*. https://doi.org/10.3389/fphar.2021.736323

Furtig B., Richter C., Wohnert J., & Schwalbe H. (2003). NMR spectroscopy of RNA. *Chembiochem*, *4*(10), 936-962. https://doi.org/10.1002/cbic.200300700

Gagliardi M., & Ashizawa A. T. (2021). The Challenges and Strategies of Antisense Oligonucleotide Drug Delivery. *Biomedicines*, *9*(4). https://doi.org/10.3390/biomedicines9040433

Gallego J., & Varani G. (2001). Targeting RNA with small-molecule drugs: therapeutic promise and chemical challenges. *Acc. Chem. Res.*, *34*(10), 836-843. https://www.ncbi.nlm.nih.gov/pubmed/11601968

Garst A. D., Heroux A., Rambo R. P., & Batey R. T. (2008). Crystal structure of the lysine riboswitch regulatory mRNA element. *J. Biol. Chem.*, *283*(33), 22347-22351. https://doi.org/10.1074/jbc.C800120200

Gaur A. B., Holbeck S. L., Colburn N. H., & Israel M. A. (2011). Downregulation of Pdcd4 by mir-21 facilitates glioblastoma proliferation in vivo. *Neuro-Oncol.*, *13*(6), 580-590. https://doi.org/10.1093/neuonc/nor033

Gebauer F., Schwarzl T., Valcarcel J., & Hentze M. W. (2021). RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.*, *22*(3), 185-198. https://doi.org/10.1038/s41576-020-00302-y

Gebhard L. G., Filomatori C. V., & Gamarnik A. V. (2011). Functional RNA elements in the dengue virus genome. *Viruses*, *3*(9), 1739-1756. https://doi.org/10.3390/v3091739

Gebhardt A., Habjan M., Benda C., Meiler A., Haas D. A., Hein M. Y., Mann A., Mann M., Habermann B., & Pichlmair A. (2015). mRNA export through an additional cap-binding complex consisting of NCBP1 and NCBP3. *Nat. Commun.*, *6*, 8192. https://doi.org/10.1038/ncomms9192

Gerstberger S., Hafner M., & Tuschl T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.*, *15*(12), 829-845. https://doi.org/10.1038/nrg3813

Gongora-Benitez M., Tulla-Puche J., & Albericio F. (2014). Multifaceted roles of disulfide bonds. Peptides as therapeutics. *Chem. Rev.*, *114*(2), 901-926. https://doi.org/10.1021/cr400031z

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.*, *47*(39), 7482-7484. https://doi.org/10.1002/anie.200801555

Gupta R. A., Shah N., Wang K. C., Kim J., Horlings H. M., Wong D. J., Tsai M. C., Hung T., Argani P., Rinn J. L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, *464*(7291), 1071-1076. https://doi.org/10.1038/nature08975

Ha M., & Kim V. N. (2014). Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.*, *15*(8), 509-524. https://doi.org/10.1038/nrm3838

Hall R. J., Mortenson P. N., & Murray C. W. (2014). Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.*, *116*(2-3), 82-91. https://doi.org/10.1016/j.pbiomolbio.2014.09.007

Hamzeh-Mivehroud M., Alizadeh A. A., Morris M. B., Church W. B., & Dastmalchi S. (2013). Phage display as a technology delivering on the promise of peptide drug discovery. *Drug Discov. Today*, *18*(23-24), 1144-1157. https://doi.org/10.1016/j.drudis.2013.09.001

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. https://doi.org/10.1021/ci000403i

Harner M. J., Frank A. O., & Fesik S. W. (2013). Fragment-based drug discovery using NMR spectroscopy. *J. Biomol. NMR*, *56*(2), 65-75. https://doi.org/10.1007/s10858-013-9740-z

Hegde S., Tang Z., Zhao J., & Wang J. (2021). Inhibition of SARS-CoV-2 by Targeting Conserved Viral RNA Structures and Sequences. *Front. Chem.*, *9*, 802766. https://doi.org/10.3389/fchem.2021.802766

Henninot A., Collins J. C., & Nuss J. M. (2018). The Current State of Peptide Drug Discovery: Back to the Future? *J. Med. Chem.*, *61*(4), 1382-1414. https://doi.org/10.1021/acs.jmedchem.7b00318

Hentze M. W., Castello A., Schwarzl T., & Preiss T. (2018). A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.*, *19*(5), 327-341. https://doi.org/10.1038/nrm.2017.130

Hermann T. (2016). Small molecules targeting viral RNA. *Wiley Interdiscip. Rev. RNA*, *7*(6), 726-743. https://doi.org/10.1002/wrna.1373

Hill M., & Tran N. (2021). miRNA interplay: mechanisms and consequences in cancer. *Dis. Model Mech.*, *14*(4). https://doi.org/10.1242/dmm.047662

Hilpert K., Elliott M., Jenssen H., Kindrachuk J., Fjell C. D., Korner J., Winkler D. F., Weaver L. L., Henklein P., Ulrich A. S., et al. (2009). Screening and characterization of surface-tethered cationic peptides for antimicrobial activity. *Chem. Biol.*, *16*(1), 58-69. https://doi.org/10.1016/j.chembiol.2008.11.006

Hodge K., Tunghirun C., Kamkaew M., Limjindaporn T., Yenchitsomanus P. T., & Chimnaronk S. (2016). Identification of a Conserved RNA-dependent RNA Polymerase (RdRp)-RNA Interface Required for Flaviviral Replication. *J. Biol. Chem.*, *291*(33), 17437-17449. https://doi.org/10.1074/jbc.M116.724013

Hood I. V., Gordon J. M., Bou-Nader C., Henderson F. E., Bahmanjah S., & Zhang J. (2019). Crystal structure of an adenovirus virus-associated RNA. *Nat. Commun.*, *10*(1), 2871. https://doi.org/10.1038/s41467-019-10752-6

Howe J. A., Wang H., Fischmann T. O., Balibar C. J., Xiao L., Galgoci A. M., Malinverni J. C., Mayhood T., Villafania A., Nahvi A., et al. (2015). Selective small-molecule inhibition of an RNA structural element. *Nature*, *526*(7575), 672-677. https://doi.org/10.1038/nature15542

Howe J. A., Xiao L., Fischmann T. O., Wang H., Tang H., Villafania A., Zhang R., Barbieri C. M., & Roemer T. (2016). Atomic resolution mechanistic studies of ribocil: A highly selective unnatural ligand mimic of the E. coli FMN riboswitch. *RNA Biol.*, *13*(10), 946-954. https://doi.org/10.1080/15476286.2016.1216304

Hu Y., Cheng K., He L., Zhang X., Jiang B., Jiang L., Li C., Wang G., Yang Y., & Liu M. (2021). NMR-Based Methods for Protein Analysis. *Anal. Chem.*, *93*(4), 1866-1879. https://doi.org/10.1021/acs.analchem.0c03830

Hung Y. H., & Slotkin R. K. (2021). The initiation of RNA interference (RNAi) in plants. *Curr. Opin. Plant Biol.*, *61*, 102014. https://doi.org/10.1016/j.pbi.2021.102014

Im H. I., & Kenny P. J. (2012). MicroRNAs in neuronal function and dysfunction. *Trends Neurosci.*, *35*(5), 325-334. https://doi.org/10.1016/j.tins.2012.01.004

Ippolito J. A., & Steitz T. A. (1998). A 1.3-A resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation. *Proc. Natl. Acad. Sci. U. S. A.*, *95*(17), 9819-9824. https://doi.org/10.1073/pnas.95.17.9819

Jackson R. J., Hellen C. U. T., & Pestova T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.*, *11*(2), 113-127. https://doi.org/10.1038/nrm2838

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. https://doi.org/10.1101/gad.4.9.1560

Javanmard S. H., Vaseghi G., Ghasemi A., Rafiee L., Ferns G. A., Esfahani H. N., & Nedaeinia R. (2020). Therapeutic inhibition of microRNA-21 (miR-21) using locked-nucleic acid (LNA)-anti-miR and its effects on the biological behaviors of melanoma cancer cells in preclinical studies. *Cancer Cell Int.*, *20*(1). https://doi.org/10.1186/s12935-020-01394-6

Jin Y., Suzuki H., Maegawa S., Endo H., Sugano S., Hashimoto K., Yasuda K., & Inoue K. (2003). A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J.*, *22*(4), 905-912. https://doi.org/10.1093/emboj/cdg089

Jing X., & Jin K. (2020). A gold mine for drug discovery: Strategies to develop cyclic peptides into therapies. *Med. Res. Rev.*, *40*(2), 753-810. https://doi.org/10.1002/med.21639

Joly J. P., Mata G., Eldin P., Briant L., Fontaine-Vive F., Duca M., & Benhida R. (2014). Artificial nucleobase-amino acid conjugates: a new class of TAR RNA binding agents. *Chemistry*, *20*(7), 2071-2079. https://doi.org/10.1002/chem.201303664

Jonas K., Calin G. A., & Pichler M. (2020). RNA-Binding Proteins as Important Regulators of Long Non-Coding RNAs in Cancer. *Int. J. Mol. Sci.*, *21*(8). https://doi.org/10.3390/ijms21082969

Kamau E., Agoti C. N., Ngoi J. M., de Laurent Z. R., Gitonga J., Cotten M., Phan M. V. T., Nokes D. J., Delwart E., Sanders E., et al. (2019). Complete Genome Sequences of

Dengue Virus Type 2 Strains from Kilifi, Kenya. *Microbiol. Resour. Announc.*, *8*(4). https://doi.org/10.1128/MRA.01566-18

Khan D., Bhat P., & Das S. (2014). HCV-like IRESs sequester eIF3: advantage virus. *Trends Microbiol.*, *22*(2), 57-58. https://doi.org/10.1016/j.tim.2013.12.009

Kim K., Jutooru I., Chadalapaka G., Johnson G., Frank J., Burghardt R., Kim S., & Safe S. (2013). HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer. *Oncogene*, *32*(13), 1616-1625. https://doi.org/10.1038/onc.2012.193

Kirsch P., Hartman A. M., Hirsch A. K. H., & Empting M. (2019). Concepts and Core Principles of Fragment-Based Drug Design. *Molecules*, *24*(23). https://doi.org/10.3390/molecules24234309

Kole R., Krainer A. R., & Altman S. (2012). RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat. Rev. Drug Discov.*, *11*(2), 125-140. https://doi.org/10.1038/nrd3625

Kontoyiannis D., Pasparakis M., Pizarro T. T., Cominelli F., & Kollias G. (1999). Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: Implications for joint and gut-associated immunopathologies. *Immunity*, *10*(3), 387-398. https://doi.org/10.1016/S1074-7613(00)80038-2

Kozomara A., & Griffiths-Jones S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, *42*(Database issue), D68-73. https://doi.org/10.1093/nar/gkt1181

Krol J., Loedige I., & Filipowicz W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, *11*(9), 597-610. https://doi.org/10.1038/nrg2843

Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860-921. https://doi.org/10.1038/35057062

Lee E., Bujalowski P. J., Teramoto T., Gottipati K., Scott S. D., Padmanabhan R., & Choi K. H. (2021). Structures of flavivirus RNA promoters suggest two binding modes with NS5 polymerase. *Nat. Commun.*, *12*(1), 2530. https://doi.org/10.1038/s41467-021-22846-1

Lee K. M., Chen C. J., & Shih S. R. (2017). Regulation Mechanisms of Viral IRES-Driven Translation. *Trends Microbiol.*, *25*(7), 547-562. https://doi.org/10.1016/j.tim.2017.01.010

Lee M. K., Bottini A., Kim M., Bardaro M. F., Jr., Zhang Z., Pellecchia M., Choi B. S., & Varani G. (2014). A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication. *Chem. Commun.*, *50*(3), 368-370. https://doi.org/10.1039/c3cc46973e

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. https://doi.org/10.1016/j.biocel.2013.05.030

Li M., & Yu B. (2021). Recent advances in the regulation of plant miRNA biogenesis. *RNA Biol.*, *18*(12), 2087-2096. https://doi.org/10.1080/15476286.2021.1899491

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. https://doi.org/10.1038/nrd4140

Lopez-Ulloa B., Fuentes Y., Pizarro-Ortega M. S., & Lopez-Lastra M. (2022). RNA-Binding Proteins as Regulators of Internal Initiation of Viral mRNA Translation. *Viruses*, *14*(2). https://doi.org/10.3390/v14020188

Lozano G., & Martinez-Salas E. (2015). Structural insights into viral IRES-dependent translation mechanisms. *Curr. Opin. Virol.*, *12*, 113-120. https://doi.org/10.1016/j.coviro.2015.04.008

Lu C., Smith A. M., Fuchs R. T., Ding F., Rajashankar K., Henkin T. M., & Ke A. (2008). Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nat. Struct. Mol. Biol.*, *15*(10), 1076-1083. https://doi.org/10.1038/nsmb.1494

Lukong K. E., Chang K. W., Khandjian E. W., & Richard S. (2008). RNA-binding proteins in human genetic disease. *Trends Genet.*, *24*(8), 416-425. https://doi.org/10.1016/j.tig.2008.05.004

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. https://doi.org/10.1038/nrm2178

Lundquist K. P., Panchal V., Gotfredsen C. H., Brenk R., & Clausen M. H. (2021). Fragment-Based Drug Discovery for RNA Targets. *Chemmedchem*, *16*(17), 2588-2603. https://doi.org/10.1002/cmdc.202100324

Luu L. M., Nguyen L., Peng S. H., Lee J., Lee H. Y., Wong C. H., Hergenrother P. J., Chan H. Y. E., & Zimmerman S. C. (2016). A Potent Inhibitor of Protein Sequestration by Expanded Triplet (CUG) Repeats that Shows Phenotypic Improvements in a Drosophila Model of Myotonic Dystrophy. *Chemmedchem*, *11*(13), 1428-1435. https://doi.org/10.1002/cmdc.201600081

Marion D. (2013). An introduction to biological NMR spectroscopy. *Mol. Cell. Proteomics*, *12*(11), 3006-3025. https://doi.org/10.1074/mcp.O113.030239

Maris C., Dominguez C., & Allain F. H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, *272*(9), 2118-2131. https://doi.org/10.1111/j.1742-4658.2005.04653.x

Martinez-Salas E., Francisco-Velilla R., Fernandez-Chamorro J., & Embarek A. M. (2018). Insights into Structural and Mechanistic Features of Viral IRES Elements. *Front. Microbiol.*, *8*. https://doi.org/10.3389/fmicb.2017.02629

Marusic M., Schlagnitweit J., & Petzold K. (2019). RNA Dynamics by NMR Spectroscopy. *Chembiochem*, *20*(21), 2685-2710. https://doi.org/10.1002/cbic.201900072

Matoulkova E., Michalova E., Vojtesek B., & Hrstka R. (2012). The role of the 3 ' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol.*, *9*(5), 563-576. https://doi.org/10.4161/rna.20231

Matsui M., & Corey D. R. (2017). Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.*, *16*(3), 167-179. https://doi.org/10.1038/nrd.2016.117

Mayr C. (2019). What Are 3 ' UTRs Doing? *Cold Spring Harb. Perspect. Biol.*, *11*(10). https://doi.org/10.1101/cshperspect.a034728

Mazeaud C., Freppel W., & Chatel-Chaix L. (2018). The Multiples Fates of the Flavivirus RNA Genome During Pathogenesis. *Front. Genet.*, *9*, 595. https://doi.org/10.3389/fgene.2018.00595

McCown P. J., Corbino K. A., Stav S., Sherlock M. E., & Breaker R. R. (2017). Riboswitch diversity and distribution. *RNA*, *23*(7), 995-1011. https://doi.org/10.1261/rna.061234.117

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. https://doi.org/10.1053/j.gastro.2007.05.022

Michlewski G., Guil S., Semple C. A., & Caceres J. F. (2008). Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol. Cell*, *32*(3), 383-393. https://doi.org/10.1016/j.molcel.2008.10.013

Mielke S. P., & Krishnan V. V. (2009). Characterization of protein secondary structure from NMR chemical shifts. *Prog. Nucl. Magn. Reson. Spectrosc.*, *54*(3-4), 141-165. https://doi.org/10.1016/j.pnmrs.2008.06.002

Mirzaei S., Paskeh M. D. A., Hashemi F., Zabolian A., Hashemi M., Entezari M., Tabari T., Ashrafizadeh M., Raee P., Aghamiri S., et al. (2022). Long non-coding RNAs as new players in bladder cancer: Lessons from pre-clinical and clinical studies. *Life Sciences*, *288*. https://doi.org/10.1016/j.lfs.2021.119948

Mitchell S. F., & Parker R. (2014). Principles and properties of eukaryotic mRNPs. *Mol. Cell*, *54*(4), 547-558. https://doi.org/10.1016/j.molcel.2014.04.033

Moreno P. M., & Pego A. P. (2014). Therapeutic antisense oligonucleotides against cancer: hurdling to the clinic. *Front. Chem.*, *2*, 87. https://doi.org/10.3389/fchem.2014.00087

Moumne R., Catala M., Larue V., Micouin L., & Tisne C. (2012). Fragment-based design of small RNA binders: promising developments and contribution of NMR. *Biochimie*, *94*(7), 1607-1619. https://doi.org/10.1016/j.biochi.2012.02.002

Mozdarani H., Ezzatizadeh V., & Rahbar Parvaneh R. (2020). The emerging role of the long non-coding RNA HOTAIR in breast cancer development and treatment. *J. Transl. Med.*, *18*(1), 152. https://doi.org/10.1186/s12967-020-02320-0

Mullard A. (2020). FDA approves RNA-targeting small molecule. *Nat. Rev. Drug Discov.*, *19*(10), 659. https://doi.org/10.1038/d41573-020-00158-1

Murchie A. I., Davis B., Isel C., Afshar M., Drysdale M. J., Bower J., Potter A. J., Starkey I. D., Swarbrick T. M., Mirza S., et al. (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. https://doi.org/10.1016/j.jmb.2003.12.028

Nahvi A., Barrick J. E., & Breaker R. R. (2004). Coenzyme B-12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.*, *32*(1), 143-150. https://doi.org/10.1093/nar/gkh167

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. https://doi.org/10.1016/j.bmcl.2015.07.016

Necsulea A., Soumillon M., Warnefors M., Liechti A., Daish T., Zeller U., Baker J. C., Grutzner F., & Kaessmann H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, *505*(7485), 635-640. https://doi.org/10.1038/nature12943

Nelson D. L., Cox M. M., & Lehninger A. L. (2017). *Lehninger principles of biochemistry* (7th edition; international edition. ed.). W.H. Freeman/Macmillan Learning.

Neupane R., Pisareva V. P., Rodriguez C. F., Pisarev A. V., & Fernandez I. S. (2020). A complex IRES at the 5'-UTR of a viral mRNA assembles a functional 48S complex via an uAUG intermediate. *Elife*, *9*. https://doi.org/10.7554/eLife.54575

Newman M. A., Thomson J. M., & Hammond S. M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA*, *14*(8), 1539-1549. https://doi.org/10.1261/rna.1155108

Ng W. C., Soto-Acosta R., Bradrick S. S., Garcia-Blanco M. A., & Ooi E. E. (2017). The 5' and 3' Untranslated Regions of the Flaviviral Genome. *Viruses*, *9*(6). https://doi.org/10.3390/v9060137

Nguyen L., Luu L. M., Peng S. H., Serrano J. F., Chan H. Y. E., & Zimmerman S. C. (2015). Rationally Designed Small Molecules That Target Both the DNA and RNA Causing Myotonic Dystrophy Type 1. *J. Am. Chem. Soc.*, *137*(44), 14180-14189. https://doi.org/10.1021/jacs.5b09266

Olejniczak M., Jiang X., Basczok M. M., & Storz G. (2022). KH domain proteins: Another family of bacterial RNA matchmakers? *Mol. Microbiol.*, *117*(1), 10-19. https://doi.org/10.1111/mmi.14842

Osborne J., Panova S., Rapti M., Urushima T., & Jhoti H. (2020). Fragments: where are we now? *Biochem. Soc. Trans.*, *48*(1), 271-280. https://doi.org/10.1042/BST20190694

Otsuka H., Fukao A., Funakami Y., Duncan K. E., & Fujiwara T. (2019). Emerging Evidence of Translational Control by AU-Rich Element-Binding Proteins. *Front. Genet.*, *10*, 332. https://doi.org/10.3389/fgene.2019.00332

Overhauser A. W. (1953). Polarization of Nuclei in Metals. *Phys. Rev.*, *92*(2), 411-415. https://doi.org/10.1103/PhysRev.92.411

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. https://doi.org/10.1021/ja309760g

Palacino J., Swalley S. E., Song C., Cheung A. K., Shu L., Zhang X., Van Hoosear M., Shin Y., Chin D. N., Keller C. G., et al. (2015). SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.*, *11*(7), 511-517. https://doi.org/10.1038/nchembio.1837

Park S. J., Kim Y. G., & Park H. J. (2011). Identification of RNA Pseudoknot-Binding Ligand That Inhibits the-1 Ribosomal Frameshifting of SARS-Coronavirus by

Structure-Based Virtual Screening. *J. Am. Chem. Soc.*, *133*(26), 10094-10100. https://doi.org/10.1021/ja1098325

Pascalis H., Biscornet L., Toty C., Hafsia S., Roche M., Despres P., Atyame Nten C., Bibi J., Louange M., Gedeon J., et al. (2020). Complete Genome Sequences of Dengue Virus Type 2 Epidemic Strains from Reunion Island and the Seychelles. *Microbiol. Resour. Announc.*, *9*(4). https://doi.org/10.1128/MRA.01443-19

Pelletier J., & Sonenberg N. (1988). Internal Initiation of Translation of Eukaryotic Messenger-Rna Directed by a Sequence Derived from Poliovirus Rna. *Nature*, *334*(6180), 320-325. https://doi.org/10.1038/334320a0

Pereira-Castro I., & Moreira A. (2021). On the function and relevance of alternative 3'-UTRs in gene expression regulation. *Wiley Interdiscip. Rev. RNA*, *12*(5), e1653. https://doi.org/10.1002/wrna.1653

Peterlin B. M., & Price D. H. (2006). Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell*, *23*(3), 297-305. https://doi.org/10.1016/j.molcel.2006.06.014

Prado S., Beltran M., Coiras M., Bedoya L. M., Alcami J., & Gallego J. (2016). Bioavailable inhibitors of HIV-1 RNA biogenesis identified through a Rev-based screen. *Biochem. Pharmacol.*, *107*, 14-28. https://doi.org/10.1016/j.bcp.2016.02.007

Rasekhian M., Roohvand F., Habtemariam S., Marzbany M., & Kazemimanesh M. (2021). The Role of 3'UTR of RNA Viruses on mRNA Stability and Translation Enhancement. *Mini Rev. Med. Chem.*, *21*(16), 2389-2398. https://doi.org/10.2174/1389557521666210217092305

Ratni H., Ebeling M., Baird J., Bendels S., Bylund J., Chen K. S., Denk N., Feng Z., Green L., Guerard M., et al. (2018). Discovery of Risdiplam, a Selective Survival of Motor Neuron-2 (SMN2) Gene Splicing Modifier for the Treatment of Spinal Muscular

Atrophy (SMA). *J. Med. Chem.*, *61*(15), 6501-6517. https://doi.org/10.1021/acs.jmedchem.8b00741

Ratni H., Scalco R. S., & Stephan A. H. (2021). Risdiplam, the First Approved Small Molecule Splicing Modifier Drug as a Blueprint for Future Transformative Medicines. *ACS Med. Chem. Lett.*, *12*(6), 874-877. https://doi.org/10.1021/acsmedchemlett.0c00659

Richards J., & Belasco J. G. (2021). Riboswitch control of bacterial RNA stability. *Mol. Microbiol.*, *116*(2), 361-365. https://doi.org/10.1111/mmi.14723

Rinn J. L., Kertesz M., Wang J. K., Squazzo S. L., Xu X., Brugmann S. A., Goodnough L. H., Helms J. A., Farnham P. J., Segal E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, *129*(7), 1311-1323. https://doi.org/10.1016/j.cell.2007.05.022

Rizvi N. F., Howe J. A., Nahvi A., Klein D. J., Fischmann T. O., Kim H. Y., McCoy M. A., Walker S. S., Hruza A., Richards M. P., et al. (2018). Discovery of Selective RNA-Binding Small Molecules by Affinity-Selection Mass Spectrometry. *ACS Chem. Biol.*, *13*(3), 820-831. https://doi.org/10.1021/acschembio.7b01013

Rizvi N. F., & Nickbarg E. B. (2019). RNA-ALIS: Methodology for screening soluble RNAs as small molecule targets using ALIS affinity-selection mass spectrometry. *Methods*, *167*, 28-38. https://doi.org/10.1016/j.ymeth.2019.04.024

Rizvi N. F., Santa Maria J. P., Jr., Nahvi A., Klappenbach J., Klein D. J., Curran P. J., Richards M. P., Chamberlin C., Saradjian P., Burchard J., et al. (2020). Targeting RNA with Small Molecules: Identification of Selective, RNA-Binding Small Molecules Occupying Drug-Like Chemical Space. *SLAS Discov.*, *25*(4), 384-396. https://doi.org/10.1177/2472555219885373

Roth A., & Breaker R. R. (2009). The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annu. Rev. Biochem.*, *78*, 305-334. https://doi.org/10.1146/annurev.biochem.78.070507.135656

Rynearson K. D., Charrette B., Gabriel C., Moreno J., Boerneke M. A., Dibrov S. M., & Hermann T. (2014). 2-Aminobenzoxazole ligands of the hepatitis C virus internal ribosome entry site. *Bioorg. Med. Chem. Lett.*, *24*(15), 3521-3525. https://doi.org/10.1016/j.bmcl.2014.05.088

Rzuczek S. G., Colgan L. A., Nakai Y., Cameron M. D., Furling D., Yasuda R., & Disney M. D. (2017). Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat. Chem. Biol.*, *13*(2), 188-193. https://doi.org/10.1038/Nchembio.2251

Sachdeva S., Lobo S., & Goswami T. (2016). What is the future of noninvasive routes for protein- and peptide-based drugs? *Ther. Deliv.*, *7*(6), 355-357. https://doi.org/10.4155/tde-2016-0031

Sakamoto K., Otake K., & Umemoto T. (2017). Discovery of peptidic miR-21 processing inhibitor by mirror image phage display: A novel method to generate RNA binding D-peptides. *Bioorg. Med. Chem. Lett.*, *27*(4), 826-828. https://doi.org/10.1016/j.bmcl.2017.01.023

Sauvageau M., Goff L. A., Lodato S., Bonev B., Groff A. F., Gerhardinger C., Sanchez-Gomez D. B., Hacisuleyman E., Li E., Spence M., et al. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife*, *2*, e01749. https://doi.org/10.7554/eLife.01749

Schiebel J., Krimmer S. G., Rower K., Knorlein A., Wang X., Park A. Y., Stieler M., Ehrmann F. R., Fu K., Radeva N., et al. (2016). High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure*, *24*(8), 1398-1409. https://doi.org/10.1016/j.str.2016.06.010

Schorderet P., & Duboule D. (2011). Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet.*, *7*(5), e1002071. https://doi.org/10.1371/journal.pgen.1002071

Schwieters C. D., Kuszewski J. J., Tjandra N., & Clore G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, *160*(1), 65-73. https://www.ncbi.nlm.nih.gov/pubmed/12565051

Selisko B., Wang C., Harris E., & Canard B. (2014). Regulation of Flavivirus RNA synthesis and replication. *Curr. Opin. Virol.*, *9*, 74-83. https://doi.org/10.1016/j.coviro.2014.09.011

Seth P. P., Miyaji A., Jefferson E. A., Sannes-Lowery K. A., Osgood S. A., Propp S. S., Ranken R., Massire C., Sampath R., Ecker D. J., et al. (2005). SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. *J. Med. Chem.*, *48*(23), 7099-7102. https://doi.org/10.1021/jm050815o

Shaka A. J., Lee C. J., & Pines A. (1988). Iterative Schemes for Bilinear Operators - Application to Spin Decoupling. *J. Magn. Reson.*, *77*(2), 274-293. https://doi.org/10.1016/0022-2364(88)90178-3

Shortridge M. D., Wille P. T., Jones A. N., Davidson A., Bogdanovic J., Arts E., Karn J., Robinson J. A., & Varani G. (2018). An ultra-high affinity ligand of HIV-1 TAR reveals the RNA structure recognized by P-TEFb. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gky1197

Shuck-Lee D., Chen F. F., Willard R., Raman S., Ptak R., Hammarskjold M. L., & Rekosh D. (2008). Heterocyclic compounds that inhibit Rev-RRE function and human immunodeficiency virus type 1 replication. *Antimicrob. Agents Chemother.*, *52*(9), 3169-3179. https://doi.org/10.1128/Aac.00274-08

Singh R. N., Ottesen E. W., & Singh N. N. (2020). The First Orally Deliverable Small Molecule for the Treatment of Spinal Muscular Atrophy. *Neurosci. Insights*, *15*, 2633105520973985. https://doi.org/10.1177/2633105520973985

Skrisovska L., Bourgeois C. F., Stefl R., Grellscheid S. N., Kister L., Wenter P., Elliott D. J., Stevenin J., & Allain F. H. (2007). The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep.*, *8*(4), 372-379. https://doi.org/10.1038/sj.embor.7400910

Statello L., Guo C. J., Chen L. L., & Huarte M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.*, *22*(2), 96-118. https://doi.org/10.1038/s41580-020-00315-9

Stelzer A. C., Frank A. T., Kratz J. D., Swanson M. D., Gonzalez-Hernandez M. J., Lee J., Andricioaei I., Markovitz D. M., & Al-Hashimi H. M. (2011). Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.*, *7*(8), 553-559. https://doi.org/10.1038/Nchembio.596

Stevens M., De Clercq E., & Balzarini J. (2006). The regulation of HIV-1 transcription: molecular targets for chemotherapeutic intervention. *Med. Res. Rev.*, *26*(5), 595-625. https://doi.org/10.1002/med.20081

Sudarsan N., Barrick J. E., & Breaker R. R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA*, *9*(6), 644-647. https://doi.org/10.1261/rna.5090103

Sztuba-Solinska J., Shenoy S. R., Gareiss P., Krumpe L. R. H., Le Grice S. F. J., O'Keefe B. R., & Schneekloth J. S. (2014). Identification of Biologically Active, HIV TAR RNA-Binding Small Molecules Using Small Molecule Microarrays. *J. Am. Chem. Soc.*, *136*(23), 8402-8410. https://doi.org/10.1021/ja502754f

Tam B., Sherf D., Cohen S., Eisdorfer S. A., Perez M., Soffer A., Vilenchik D., Akabayov S. R., Wagner G., & Akabayov B. (2019). Discovery of small-molecule inhibitors targeting the ribosomal peptidyl transferase center (PTC) of M. tuberculosis. *Chem. Sci.*, *10*(38), 8764-8767. https://doi.org/10.1039/c9sc02520k

Tanida Y., & Matsuura A. (2020). Alchemical free energy calculations via metadynamics: Application to the theophylline-RNA aptamer complex. *J. Comput. Chem.*, *41*(20), 1804-1819. https://doi.org/10.1002/jcc.26221

Tian Z. L., Liang G. H., Cui K. L., Liang Y. Y., Wang Q., Lv S. Y., Cheng X. X., & Zhang L. (2021). Insight Into the Prospects for RNAi Therapy of Cancer. *Front. Pharmacol.*, *12*. https://doi.org/10.3389/fphar.2021.644718

Tintaru A. M., Hautbergue G. M., Hounslow A. M., Hung M. L., Lian L. Y., Craven C. J., & Wilson S. A. (2007). Structural and functional analysis of RNA and TAP binding to SF2/ASF. *EMBO Rep.*, *8*(8), 756-762. https://doi.org/10.1038/sj.embor.7401031

Valverde R., Edwards L., & Regan L. (2008). Structure and function of KH domains. *FEBS J.*, *275*(11), 2712-2726. https://doi.org/10.1111/j.1742-4658.2008.06411.x

van den Elsen K., Quek J. P., & Luo D. H. (2021). Molecular Insights into the Flavivirus Replication Complex. *Viruses*, *13*(6). https://doi.org/10.3390/v13060956

Varani G., Aboulela F., & Allain F. H. T. (1996). NMR investigation of RNA structure. *Prog. Nucl. Mag. Res. Sp.*, *29*, 51-127. https://doi.org/Doi 10.1016/0079-6565(96)01028-X

Varani G., & Tinoco I., Jr. (1991). RNA structure and NMR spectroscopy. *Q. Rev. Biophys.*, *24*(4), 479-532. https://doi.org/10.1017/s0033583500003875

Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G., Smith H. O., Yandell M., Evans C. A., Holt R. A., et al. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304-1351. https://doi.org/10.1126/science.1058040

Vicens Q., & Westhof E. (2003). Molecular recognition of aminoglycoside antibiotics by ribosomal RNA and resistance enzymes: an analysis of x-ray crystal structures. *Biopolymers*, *70*(1), 42-57. https://doi.org/10.1002/bip.10414

Villordo S. M., Alvarez D. E., & Gamarnik A. V. (2010). A balance between circular and linear forms of the dengue virus genome is crucial for viral replication. *RNA*, *16*(12), 2325-2335. https://doi.org/10.1261/rna.2120410

Volinia S., Calin G. A., Liu C. G., Ambs S., Cimmino A., Petrocca F., Visone R., Iorio M., Roldo C., Ferracin M., et al. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. U. S. A.*, *103*(7), 2257-2261. https://doi.org/10.1073/pnas.0510565103

Walker M. J., & Varani G. (2019). Design of RNA-targeting macrocyclic peptides. *Methods Enzymol.*, *623*, 339-372. https://doi.org/10.1016/bs.mie.2019.04.029

Wapinski O., & Chang H. Y. (2011). Long noncoding RNAs and human disease. *Trends Cell Biol.*, *21*(6), 354-361. https://doi.org/10.1016/j.tcb.2011.04.001

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. https://doi.org/10.1038/nrd.2018.93

Watashi K., Yeung M. L., Starost M. F., Hosmane R. S., & Jeang K. T. (2010). Identification of Small Molecules That Suppress MicroRNA Function and Reverse Tumorigenesis. *J. Biol. Chem.*, *285*(32), 24707-24716. https://doi.org/10.1074/jbc.M109.062976

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. https://doi.org/10.1093/nar/gkp117

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. https://doi.org/10.1016/S0968-0004(03)00051-3

Williams C. J., Headd J. J., Moriarty N. W., Prisant M. G., Videau L. L., Deis L. N., Verma V., Keedy D. A., Hintze B. J., Chen V. B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*, *27*(1), 293-315. https://doi.org/10.1002/pro.3330

Williamson M. P. (1993). Peptide structure determination by NMR. *Methods Mol. Biol.*, *17*, 69-85. https://doi.org/10.1385/0-89603-215-9:69

Williamson M. P., Havel T. F., & Wuthrich K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.*, *182*(2), 295-315. https://doi.org/10.1016/0022-2836(85)90347-x

Wilson D. N. (2014). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.*, *12*(1), 35-48. https://doi.org/10.1038/nrmicro3155

Winkle M., El-Daly S. M., Fabbri M., & Calin G. A. (2021). Noncoding RNA therapeutics - challenges and potential solutions. *Nat. Rev. Drug Discov.*, *20*(8), 629-651. https://doi.org/10.1038/s41573-021-00219-z

Wishart D. S., Sykes B. D., & Richards F. M. (1991). Simple techniques for the quantification of protein secondary structure by 1H NMR spectroscopy. *FEBS Lett.*, *293*(1-2), 72-80. https://www.ncbi.nlm.nih.gov/pubmed/1959674

Wuthrich K. (1990). Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.*, *265*(36), 22059-22062. https://www.ncbi.nlm.nih.gov/pubmed/2266107

Wuthrich K., Billeter M., & Braun W. (1984). Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J. Mol. Biol.*, *180*(3), 715-740. https://doi.org/10.1016/0022-2836(84)90034-2

Yu L., Nomaguchi M., Padmanabhan R., & Markoff L. (2008). Specific requirements for elements of the 5' and 3' terminal regions in flavivirus RNA synthesis and viral replication. *Virology*, *374*(1), 170-185. https://doi.org/10.1016/j.virol.2007.12.035

Yusupov M. M., Yusupova G. Z., Baucom A., Lieberman K., Earnest T. N., Cate J. H., & Noller H. F. (2001). Crystal structure of the ribosome at 5.5 A resolution. *Science*, *292*(5518), 883-896. https://doi.org/10.1126/science.1060089

Zeiger M., Stark S., Kalden E., Ackermann B., Ferner J., Scheffer U., Shoja-Bazargani F., Erdel V., Schwalbe H., & Gobel M. W. (2014). Fragment based search for small molecule inhibitors of HIV-1 Tat-TAR. *Bioorg. Med. Chem. Lett.*, *24*(24), 5576-5580. https://doi.org/10.1016/j.bmcl.2014.11.004

Zeng W., Guo L., Xu S., Chen J., & Zhou J. (2020). High-Throughput Screening Technology in Industrial Biotechnology. *Trends Biotechnol.*, *38*(8), 888-906. https://doi.org/10.1016/j.tibtech.2020.01.001

Zhao Y., Soh T. S., Zheng J., Chan K. W., Phoo W. W., Lee C. C., Tay M. Y., Swaminathan K., Cornvik T. C., Lim S. P., et al. (2015). A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog.*, *11*(3), e1004682. https://doi.org/10.1371/journal.ppat.1004682

Zhou S., Rynearson K. D., Ding K., Brunn N. D., & Hermann T. (2013). Screening for inhibitors of the hepatitis C virus internal ribosome entry site RNA. *Bioorg. Med. Chem.*, *21*(20), 6139-6144. https://doi.org/10.1016/j.bmc.2013.03.054

Zhou W., Shi G., Zhang Q., Wu Q., Li B., & Zhang Z. (2014). MicroRNA-20b promotes cell growth of breast cancer cells partly via targeting phosphatase and tensin homologue (PTEN). *Cell Biosci.*, *4*(1), 62. https://doi.org/10.1186/2045-3701-4-62

Zhu J., Chen L., Zou L., Yang P., Wu R., Mao Y., Zhou H., Li R., Wang K., Wang W., et al. (2014). MiR-20b, -21, and -130b inhibit PTEN expression resulting in B7-H1 over-expression in advanced colorectal cancer. *Hum. Immunol.*, *75*(4), 348-353. https://doi.org/10.1016/j.humimm.2014.01.006

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. https://doi.org/10.1074/jbc.M611393200

Zhu S., Wu H., Wu F., Nie D., Sheng S., & Mo Y. Y. (2008). MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res.*, *18*(3), 350-359. https://doi.org/10.1038/cr.2008.24

Zogg H., Singh R., & Ro S. (2022). Current Advances in RNA Therapeutics for Human Diseases. *Int. J. Mol. Sci.*, *23*(5). https://doi.org/10.3390/ijms23052736

Chapter 2

A small cyclic β-hairpin peptide mimics the Rbfox2 RRM and binds to the precursor miRNA 20b

Abstract

The RRM (RNA-recognition motif), the most abundant RNA-binding motif in eukaryotes, is a well-structured domain of about 90 amino acids, yet the β2-β3 hairpin corresponding to strands 2 and 3 of the β-sheet, and the intervening loop makes essential interactions with RNA in many RRM complexes. I designed a series of small cyclic peptides mimics of the β2-β3 hairpin of Rbfox2 protein, that recognize the terminal loop of precursor miR-20b, to investigate whether I could mimic the full RNA-binding protein with a minimal structurally pre-organized peptide. Within a small library of 7 cyclic peptides, I identified a peptide with low μ-molar affinity for the miR-20b precursor. NMR titration data suggest that this peptide specifically targets the apical loop of pre-miR-20b. My work shows that it is possible to mimic RNA-binding proteins with designed stable peptides, which provide a starting point for designing or evolving small peptide mimetics of RRM proteins.

Section 2.1: Introduction

The RRM is the most abundant RNA-binding motif in eukaryotes (Venter et al., 2001). RRM-containing proteins are involved in all aspects of post-transcriptional gene regulation, including pre-mRNA processing, mRNA export, and stability (Maris et al., 2005). RRMs are typically 80–100 amino acids long and fold to form a four-stranded antiparallel β-sheet, packed against two α-helices, with a split αβ topology (Clery et al., 2008; Maris et al., 2005). Despite the similarity in structure, RRMs are able to bind to a wide variety of RNAs that have different sequences and lengths, in some cases with high affinity and specificity, although this is most often found in proteins containing multiple domains (Afroz et al., 2015; Burd & Dreyfuss, 1994). The versatile sequence specificity provides RRM-containing proteins with diverse biological functions by endowing them with the ability to target many cellular RNAs (Clery et al., 2008; Maris et al., 2005). Typically, RRM proteins recognize RNA through the β-sheet surface, at the center of which are strands β1 and β3, which contain the conserved signature sequences that define the domain (Clery et al., 2008; Maris et al., 2005). However, noncanonical interactions through the α-helices (Clery et al., 2013; Tintaru et al., 2007), protein loop (Auweter et al., 2006; Dominguez et al., 2010; Skrisovska et al., 2007), and N- or C- terminal extensions (Clery et al., 2011; Oberstrass et al., 2005; Tsuda et al., 2009) are also found in RRM-RNA complexes, to boost affinity or specificity of individual domains and even to redirect RRMs to bind to proteins.

Among RNA with activity regulated by RRM proteins are miRNAs (Ha & Kim, 2014; Krol et al., 2010), which are small noncoding RNAs of 21–23 nucleotides that play key regulatory roles in protein expression by binding target mRNAs and inhibiting translation or inducing mRNA degradation (Bartel, 2009; Fabian et al., 2010). The primary miRNA (pri-miRNA), which is the long precursor of miRNA transcribed by RNA polymerases, is cleaved by the Drosha/DGCR8 microprocessor complex in the nucleus to release the precursor miRNA (pre-miRNA), a 60-70 nt stem-loop embedding the mature miRNA. The pre-miRNA is then cleaved into about 22 nt miRNA:miRNA* duplexes by the RNAseIII enzyme Dicer in the cytoplasm, from which the mature miRNA is loaded onto the RNA-induced silencing complex (RISC) to guide the repression of target mRNAs. Several RRM proteins regulate miRNA processing by binding at or near the apical region of the stem-loop precursors species, which is common to both pre- and pri-miRNAs (Michlewski et al., 2008; Newman et al., 2008), including the highly conserved Rbfox2 RRM, which inhibits the maturation of miR-20b and miR-107 by binding to the terminal loop of the precursor miRNAs (Chen et al., 2016).

Rbfox proteins are tissue-specific alternative splicing factors (Auweter et al., 2006; Jin et al., 2003; Kuroyanagi, 2009), and include three family members in mammals with nearly identical N-terminal RRMs: Rbfox1(A2BP1), Rbfox2(RBM9), and Rbfox3(NeuN) (Kuroyanagi, 2009). The RRM of Rbfox proteins targets RNA specifically through the conserved (U)GCAUG element, a well-known *cis* element for the regulation of alternative splicing (Auweter et al., 2006; Jin et al., 2003), which is also found in the apical loop of the two miRNA precursors targeted by Rbfox2 (Chen et al., 2016). Through this interaction, the Rbfox2 RRM inhibits the processing of the two pri-miRNAs in vitro and in mammalian cells (Chen et al., 2016). The downregulation of miR-20b by Rbfox2 increases the expression of downstream targets (Zhou et al., 2014), such as the tumor suppressor phosphatase and tensin homologue (PTEN) (Chen et al., 2016). Conversely, upregulation of miR-107 by knocking-down Rbfox2 decreases the expression of Dicer protein (Chen et al., 2016); a downstream target of miR-107 (Li et al., 2015; Martello et al., 2010; Ristori et al., 2015).

The structure of the complex of Rbfox2 RRM and pre-miR-20b was determined by nuclear magnetic resonance (NMR) spectroscopy (Figure 2.1) (Chen et al., 2016). In the complex, three base pairs in the apical stem immediately below the loop open up, drastically enlarging the size of the terminal loop. The β2β3 hairpin inserts into the opened-up terminal loop of pre-miR-20b, anchoring the RNA loop-helix junction against the protein surface. Inspired by this structure, I designed a small set of short cyclic peptide mimics of the β2β3 and investigated whether it was possible to mimic the core of recognition elements of RRM proteins with a minimal peptide and target pre-miR-20b.

In addition to the intrinsic interest in protein engineering and evolution of RNA-binding protein, which might have emerged from much smaller β-hairpins by subsequent duplication and fusion, peptides provide much more advantageous starting points to discover miRNA inhibitors (Craik et al., 2013), an attractive prospect since mis-expression of miRNAs is observed in many human diseases (Croce, 2009; Ha & Kim, 2014; Im & Kenny, 2012), often as a result of post-transcriptional mis-regulation of their maturation. Peptides potentially combine the specificity and binding activity of proteins, but are smaller in size and cheaper to manufacture and deliver (Craik et al., 2013). A peptidic miR-21 processing inhibitor was discovered and reported in (Sakamoto et al., 2017). Compared to linear peptides, cyclic peptides have better binding affinity, specificity and stability due to conformational rigidity and are less susceptible to proteolysis within cells and organisms (Gongora-Benitez et al., 2014).

Cyclic peptides mimic of Arginine-rich domain proteins that potently disrupt the HIV TAR RNA-Tat protein interactions have been discovered in the Varani group (Athanassiou et al., 2004; Athanassiou et al., 2007; Davidson et al., 2009; Lalonde et al., 2011), and a peptide mimic of an RRM stabilized by a perfluoroaryl side chain linkage was discovered through evolution and selection (Belashov et al., 2018). I wished to use a similar design strategy with the Rbfox RRM2, but unlike the Tat protein, which forms a short β-hairpin that can immediately be mimicked by cyclic peptides, RRM proteins such as Rbfox2 are much larger and have complex structures. By mimicking the β2β3 hairpin of Rbfox2 based on the structure of Figure 2.1, I discovered a cyclic peptide mimetic of this protein which selectively binds to pre-miR-20b with a dissociation constant (Kd) ~50 μM, and opens the way for the optimization of this new peptide by rational design and genetic selection.

Section 2.2: Result

Section 2.2.1: Peptide Design

Cyclic β-hairpin peptides were designed to mimic the β2β3 hairpin in the Rbfox2 RRM structure (Figure 2.1B). This part of the protein forms a β-hairpin in all RRMs, and amino acids from it, including both the strands and intervening loop, are often involved in sequence-specific RNA recognition. For example, the specificity of U1A and U2B'' proteins can be switched by swapping β2 among the two proteins (Price et al., 1998). Given its role in sequence-specific RNA recognition, we sought to mimic these hairpins with synthetic peptides, cyclized to form a stable β-hairpin with a sequence and structure matching those of the β2β3 hairpin of Rbfox2. We reasoned that the peptide would retain some of the activity of the complete protein, and that full activity could later be restored by further cycles of design and/or selection.

To provide proof of principle that this goal can be achieved, seven cyclic peptides were designed (Table 2.1) varying in length (14, 16 and 18 residues), as was originally done when mimicking the BIV Tat protein (Athanassiou et al., 2004). It was later found that despite the 14-amino-acid-long cyclic peptide would be more stable, the 16-amino-acid-long cyclic peptides mimic the protein structure best, as described herein. The ^{D}Pro-^{L}Pro type IIβ '-turn was introduced into all sequences to promote β-hairpin formation (Athanassiou et al., 2004). The β2β3 loop in Rbfox2 RRM is composed of Asn151, Glu152, Arg153 and Gly154. We transplanted this loop and part of the β2 and β3 strands onto the type IIβ '-turn to mimic the RNA interacting segment observed in Rbfox2-RNA structure (Figure 2.2). Direct transplantation of residue 148-159 and 146-161 from Rbfox2 RRM generate cyclic peptides CPfox1 and CPfox3, respectively. The rest of the peptides are designed rationally. The basic side chain of residue 156 provides a salt bridge interaction to the RNA backbone (Maris et al., 2005); therefore, Lys or Arg were used in position 11 in CPfox peptides. Aromatic residues on strand β3 (Phe158 and Phe160) which often interact with the RNA sugar in RRM/RNA interactions (Clery et al., 2008; Maris et al., 2005) were left unchanged in all mimetic peptides (position 13 and 15). Gly residues on strand 2 were substituted with Arg in CPfox2 and CPfox4-7 (positon 12 and/or 14) to increase β-sheet propensity and solubility and to rigidify the structure. Position 2 in CPfox5-7 was changed into Lys to increase the basicity of peptide. for the same reason, position 5 in CPfox6 was changed into Arg. Finally, in CPfox7, the Glu in the center of the RNA interacting loop (position 7) was mutated to Gln to remove the acidic side chain while retaining side chain size and hydrophilicity.

Section 2.2.2: Structure of the Peptides

The ability of peptides to bind the target RNA is highly affected by their structure. Folded peptides which pre-organize the interacting groups often give better binding affinity and specificity than unfolded peptides. Thus, we investigated the structure of each peptide mimetic in aqueous solution by NMR spectroscopy to confirm their rigidity. 1D and 2D NMR spectra, including 1H NOESY, TOCSY and natural abundant HSQC spectra, were obtained for all but peptide CPfox3, for which 2D spectra could not be recorded because of its low solubility in water.

The 1D and 2D ^{1}H NMR spectra of CPfox1 show that the peptide is unfolded (Figure 2.3 and 2.4), as revealed by the narrow diversion of HN peaks and multiple conformations are present in exchange, as revealed by the broad peak shape. In contrast, the spectra of CPfox2 demonstrate formation of a highly stable β-hairpin structure (Figure 2.3 and 2.5). The HN peaks are widely dispersed, and most of the $^{3}J_{HN\text{-}H\alpha}$ constants for residues within the putative β-strands are >9.0 Hz, as expected for a β-sheet structure. In addition, many intra-strand and cross-strand NOE signals expected for β-sheet structures are observed in the NOESY spectrum of CPfox2. For CPfox3, the 1D 1H NMR spectra show a moderate dispersion of

HN peaks, suggesting a less stable β-hairpin structure compared to CPfox2 (Figure 2.3). Unfortunately, no 2D NMR information could be collected for CPfox3 because of its low solubility. The 1D and 2D NMR spectra show that 16-mers CPfox4-7 have β-sheet structure (Figure 2.3 and 2.6-2.9). The dispersed HN peaks and β-sheet NOE signals found in 2D NOESY spectra of these peptides indicate the formation of β-hairpin. However, the observation of fewer cross-strand NOEs and moderate dispersion of HN peaks, compared to CPfox2, suggests that the β-hairpins are moderately structured or perhaps experiencing conformational exchange. This is not inconsistent with their length, less favorable than the 14-residue sequences of CPfox2 for β-hairpin formation.

Altogether, the NMR studies demonstrate that CPfox1 and CPfox2 are fully unfolded and folded peptides, respectively, while the rest of the peptide mimetics fold into moderately stable β-hairpins.

Structure calculations were performed for CPfox2 and CPfox4-7 using NOE-derived restraints (see Section 2.4.4). The results are shown in Figure 2.10 and summarized in Table 2.2. The 10 lowest structures converge well for each peptide, with no NOE restraint violation greater than 0.2 Å and high structural definition. The backbone RMSD values of the 5 calculated peptide structures are all between 1.0 and 1.4 Å (Table 2.2).

The calculations confirm that the 14-mer CPfox2 is a highly structured antiparallel β-hairpin with type-II and type-IIβ '-turns at each end of the hairpin (Figure 2.10A). However, the hydrogen-bonding residues observed in the structure of this peptide differ from those observed in the Rbfox2 RRM protein, suggesting that CPfox2 is a very poor structural mimic of β2β3 in Rbfox2. Namely, residue Ile1, Phe3, Glu5, Ser8, Arg10 and Arg12 are hydrogen-bonded across the β-strand, while the corresponding residues in Rbfox2 RRM are not. As a consequence, the side chains in the peptide point in the opposite direction compared to the corresponding residue in the Rbfox2 RRM (Figure 2.11A). We encountered a similar problem with BIV Tat and were unable to simply transplant the BIV ARM sequence onto a ^{D}Pro-^{L}Pro template, but had instead to changes the sequence to mimic the peptide structure observed in the complex with RNA (Athanassiou et al., 2004).

The structure calculation of CPfox4 revealed a less structured β-hairpin compared to CPfox2 (Figure 2.10B), with a longer and more flexible loop. However, this 16-mer peptide presents the same hydrogen-bonding residues and side chain geometry as Rbfox2 RRM (Figure 2.11B). The structures calculated for the other 16-mer peptides also showed the side chain geometry observed in the Rbfox2-RNA structure, suggesting that 16-mer peptides can mimic the β2β3 hairpin of the Rbfox2 RRM, while 14-mer peptides cannot because of intrinsic structural preferences induced by the ^{D}Pro-^{L}Pro turn. Among all 16-mers, both CPfox4 and CPfox7 possess a flat round loop (Figure 2.10B and E), while the loops of CPfox5 and CPfox6 are more structured (Figure 2.10C and D).

Section 2.2.3: Peptide Binding to Pre-miR-20b

Electrophoretic mobility shift assays (EMSA) were used to test the interaction between the peptide mimetics and pre-miR-20b RNA. Among the designs, none of the 14- or 18-mer peptides (CPfox1-3) showed any detectable binding (Figure 2.12) at concentrations of 200-500 μM. This is not in retrospect surprising, once the structures of the peptides are analyzed, since these are poor structural mimetic of the β2β3 hairpin of Rbfox2, with different cross-strand hydrogen bonding pattern and therefore different displays of amino acid side chains to the RNA. CPfox5 and CPfox6 bind to RNA, but the interactions appear to be non-specific, since it can be significantly reduced when binding assays are carried out in the presence of 250-fold excess of competitor tRNA (Figure 2.12 and 2.13). However, peptidomimetics CPfox4 and CPfox7 show binding to pre-miR-20b in the presence of competitor tRNA (Figure 2.12), albeit with a relatively weak apparent Kd of ~50 μM for CPfox4 and weaker for CPfox7. However, EMSA measurements are not accurate at this Kd because of the fast exchange between the bound and free forms of the RNA on the gels; as a result, the binding band of CPfox7/pre-miR-20b was broadly distributed in the lane (Figure 2.14). Precipitations were also observed in both CPfox4/pre-miR-20b and CPfox7/pre-miR-20b mixture when the peptide concentration is above 100 μM. It is therefore very likely that the Kd's are underestimated.

Section 2.2.4: NMR studies of peptide-RNA interaction

The interaction between pre-miR-20b and CPfox4 and CPfox7 were further investigated by NMR titration experiments. 1D ^{1}H NMR spectra of pre-miR-20b RNA with increasing amounts of CPfox4 or CPfox7 were recorded (Figure 2.15 and 2.16). Some peaks shifted and broadened when the concentration of peptide increased, confirming binding as observed by EMSAs, while some of the peaks remained unchanged during the titration, indicating that the interaction occurs at a specific site on the RNA.

2D TOCSY spectra of free RNA and the CPfox4/RNA complex (at 2.5:1 ratio) were also recorded (Figure 2.17). The peak corresponding to the C30 H5-H6 resonances in pre-miR-20b, a nucleotide in the RNA apical loop, shifts significantly after the addition of CPfox4 peptide. The H5-H6 resonance peaks of U25, U26 and U36, which are immediately below the apical loop (Figure 2.1), shift as well. The result suggests a direct interaction between the peptide and this RNA apical loop.

On the other hand, only weak changes in chemical shifts were found in the TOCSY spectrum of CPfox7/RNA (2.5:1 ratio) mixture (Figure 2.18). Direct interactions between pre-miR-20b apical loop and CPfox7 were not observed by NMR, though this might be due to the more limited solubility and weak affinity of this peptide.

Section 2.3: Discussion

RRM protein recognize RNAs through a variety of binding modes (Afroz et al., 2015; Clery et al., 2008), but most use the large relatively featureless β-sheet surface; thus, it is difficult to mimic RRM/RNA binding with small molecules for the purpose of therapeutic RNA targeting. However, peptide analogues that mimic key protein structural elements can provide specific and potent therapeutic leads; for example, α -helical regions of DNA-binding proteins can be mimicked with stapled peptides (Frank et al., 2014), whereby the stapling stabilizes the α-helical structure that is necessary for recognition. Since the β2β3 hairpin and its intervening loop often provides critical contacts in RRM-RNA complexes, we reasoned that it might be possible to use cyclic β-hairpin peptides to mimic the binding activity of the much larger complete domain. We used the Rbfox2 RRM and its interaction with the pre-miR-20b RNA as a starting point for design, in part because this protein binds to a structured RNA, which would make it more likely to observe direct binding compared to single stranded RNA targets recognized by other RRMs, including Rbfox2 binding to splicing signals.

Based on the structure of Rbfox2 RRM/pre-miR-20b complex, I designed peptide mimetics of β2β3 hairpin, tested their folding with NMR and measured binding affinities by EMSA gel shift assays. The first synthetic peptide mimetic of the β2β3 hairpin in Rbfox2 RRM was a straightforward transplant of residues Ile148-Gly159 of Rbfox2 onto a hairpin-inducing ^{D}Pro-^{L}Pro template (Table 2.1). By pre-structuring the mimetic peptide into a conformation close to that observed the bound complex with the templating amino acids, I expected to favor binding by reducing entropic losses associated with peptide folding from an unstructured state (Athanassiou et al., 2004). However, the resulting design, CPfox1, does not adopt a β-hairpin conformation, most likely because of the high Gly-content of the sequence, which introduces excessive conformational flexibility.

In order to improve the stabilities of the design, two different strategies were adopted. The first was replacing Gly residues with Arg, which were introduced to improve the β-sheet propensity and its solubility and potentially provide improved binding to RNA. The resulting mimetic CPfox2 forms a stable β-hairpin structures in aqueous solution (Figure 2.10A), but failed to bind to RNA. Structural analysis revealed that the side chains in this peptide point in a direction opposite to that observed in the Rbfox2 RRM (Figure 2.11A). This occurs because the (^{D}Pro-^{L}Pro) template forms a type-II β-turn, which dictates the conformation of the rest of the peptide, making residue 1, 3, 5, 8,10 and 12 form cross-strand backbone hydrogen bonds. The corresponding residues in Rbfox2 RRM are non-hydrogen-bonding residues instead. When we elongated the β-strand by four amino acids, to provide more cross-strand hydrogen bonds, the resulting CPfox3 peptide is better structured than CPfox1, but has low solubility in water and could not be fully analyzed by NMR.

This situation was also encountered when we first transplanted the sequence of BIV Tat onto the template to mimic the BIV Tat-TAR complex (Athanassiou et al., 2004). In that case, it was necessary to mimic the structure of the peptide, not just its sequence, to observe any binding. Based on the structure of CPfox2, we synthesized 16-mer peptide mimetics rather than 14-mers to restore the hydrogen bonding pattern observed in the complex. It is well-known that the length of cyclic peptide influences the stability of β-hairpin structures (Gibbs et al., 1998). Cyclic peptides with 4n+2 residues form more stable β-hairpins compared to those with 4n residues. Thus, it is not surprising that CPfox4-7, 16-mer cyclic peptides, are all less structured than the 14-mer CPfox2 from which they originate. However, by increasing the peptide length by 2 residues and placing the turn in phase with the correct residues, the side chains of the 16-mer peptides adopt conformations consistent with the β2β3 hairpin within Rbfox2 RRM (Figure 2.11B). Mid-μM binding to the pre-miR-20b RNA was observed for two of these peptides (CPfox4 and CPfox7), although their Kd couldn't be accurately measured by EMSA due to the fast exchange of the complex and limited solubility under high peptide concentration (>100μM).

CPfox4 with a flat and flexible loop, corresponding to the β2β3 loop in the Rbfox2 RRM, binds to pre-miR-20b best among all peptidomimetics. Flexibility can help improve the RNA binding affinity by allowing direct interactions through conformational adaptation (Leulliot & Varani, 2001). The flexibility in the peptide loop can thus help improve its binding to RNA, particularly in the initial phases of design when binding is expected to be weak and interactions sub-optimal. Consistent with this hypothesis, CPfox5 and CPfox6, which possess a more structured loop than CPfox4, only bind to RNA non-specifically (binding activity is abrogated in the presence of excess tRNA). Thus, the structure and binding results support the hypothesis that flexibility in the binding loop facilitates binding, yet a pre-structured β-hairpin is required for the interaction to occur, presenting each amino acid in the correct orientation with respect to the RNA stem-loop.

When the binding between pre-miR-20b and CPfox4 was examined by NMR, gradual peak shifts were observed in 1D ^{1}H NMR spectra upon addition of the peptides, suggesting that the bound form of pre-miR-20b is in fast exchange with the free form, consistent with the mid-μM affinity observed by EMSA. In addition, the TOCSY spectra showed that the peaks of four nucleotides located in the apical loop shifts significantly after the addition of CPfox4 (Figure 2.17). The peak shift of C30 indicates a direct interaction between the RNA apical loop and the peptide also observed in the structure of Rbfox2 RRM bound to pre-miR-20b (Figure 2.19).

Although CPfox4 is basic, the interaction with RNA is not simply the result of non-specific electrostatics. First of all, the binding affinity of CPfox7 is weaker than CPfox4, although CPfox7 is more positively charged; second, CPfox4 cannot be competed away by tRNA. To further confirm the specificity of CPfox4 binding to pre-miR-20b, an EMSA experiment with CPfox4 and HIV TAR was conducted and no binding was detected (Figure 2.20), despite the RNA's similar size, secondary structure and loop size. Binding between CPfox4 and mutant pre-miR-20b (C30A), and between mutant CPfox4 (cyclo-EIIFNE**A**GSKRFRFpP) and pre-miR-20b were also examined. By mutating the Cytosine in RNA or Arginine in peptide involved in the interacting interface (Figure 2.19), the CPfox4 and pre-miR-20b binding affinity is erased (Figure 2.21).

The results presented here demonstrate the successful mimicking of a full RNA-binding protein domain with a small cyclic peptide, and adds a second approach, beyond direct evolution, to generate small peptide mimics of relatively large RRM proteins (Belashov et al., 2018). We created a specific, although weak binder for the pre-miR-20b apical loop by grafting the sequence of the β2β3 loop of RBfox2 RRM onto a β-hairpin promoting template, and ensuring through the design that the correct amino acid side chains are exposed to the RNA. In the Rbfox2 RRM and in many such proteins, residues within β1, the β1α1 loop and the C-terminal loop also interact with RNA, creating a large interface (Figure 2.1). Thus, it is not surprising that affinity is decreased 3 orders of magnitude, since we have discarded half of the interacting functional groups in the interface and disrupted a stable protein fold. Clearly, this peptide represents a starting not a final point for the successful design of more potent and specific peptides, as we did successfully with the BIV and HIV TAR RNAs, by using structure-based peptide design (Athanassiou et al., 2004; Athanassiou et al., 2007; Davidson et al., 2009) or evolution/selection methods (Belashov et al., 2018). This work also showed the possibility of structure-based design strategy of RNA-binding ligands. By understanding the crucial interactions between a protein and RNA, we can mimic the binding interface with properly designed peptides. Due to the more powerful X-ray crystallography and NMR instrumentation, an increasing number of protein-RNA structures are available in the Protein Data Bank (PDB) Thus, more potential RNA-binding peptides can be discovered with a better understanding of protein-RNA interactions. The discovery of CPfox peptides might provide leads for the development of miRNA maturation inhibitors, and proof of principle that even complex interfaces as observed between RRMs and RNA can be mimicked by much smaller peptide structures.

Section 2.4: Materials and Methods

Section 2.4.1: Peptide Synthesis

Peptides were synthesized by solid phase peptide synthesis using Fmoc chemistry (Athanassiou et al., 2004; Fields & Noble, 1990). Resin and protected amino acids were purchased from AappTec and P3 Biosystems. Arg(Pbf)-preloaded 2-chlorotrityl resin was used for the synthesis. Linear peptide synthesis was performed on a Liberty BlueTM automated microwave peptide synthesizer with DMF as the solvent. Piperidine (20% in DMF) was used for deprotection, and DIC/Oxyma for coupling. The linear precursor was cleaved from the resin with 5% acetic acid in DCM for 2 hours. Cyclization was performed with 6 equivalents of PyAOP/HOAt and 2,4,6-trimethylpyridine in DMF solution, and progress was monitored by HPLC. The peptide was then treated with TFA/triisoproprylsilane/H2O (95:2.5:2.5) solution for 2 hours for side-chain deprotection.

Cyclic peptides were purified by reverse phase HPLC with a semi-preparative C18 column. The gradients were from 7.5% to 100% isopropanol in H2O with 0.1% TFA and 5% to 100% Acetonitrile in H_2O with 0.1% TFA. Each peptide showed a single peak (>95% purity) in HPLC. The identity of all peptides was confirmed with Bruker Esquire Ion Trap mass spectrometry. HPLC and mass spectrometry data are shown in Figure 2.22-2.28.

Section 2.4.2: RNA Preparation and Binding Assays

Pre-miR-20b RNA (G19 to C41) was prepared by *in vitro* transcription with in-house purified T7 RNA polymerase with a commercially synthetic oligonucleotide template (IDT) (Beckert & Masquida, 2011; Milligan et al., 1987). The RNA oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis (PAGE), electroeluted and concentrated by ethanol precipitation.

In order to prepare labeled RNA for binding assays, purified RNA was 5'-dephosphorylated with Calf Intestinal protein (CIP) then re-phosphorylated with [γ-^{32}P]-ATP and T4 polynucleotide kinase. RNA was annealed and snap cooled before use. Peptide and RNA were incubated in buffer containing 12.5 mM Tris-HCl (pH 7.4), 12.5 mM KCl, 250-fold excess tRNA, and 0.25% Triton X-100 for 30 minutes. The samples were loaded onto 12% native polyacrylamide gel in 0.5X TBE buffer and electrophoresed at 300 V and 4 W for 2 hours. Dried gels were exposed to a phosphor imaging plate and scanned by GE Typhoon FLA 9000 gel imager.

Section 2.4.3: NMR Binding Studies

NMR titration experiments were performed on Bruker DRX 500 MHz or Avance III 600 MHz spectrometer, equipped with cryocooled probes. Free RNA samples were prepared at a concentration of 200 μM in 10 mM phosphate buffer (pH 6.5, H_2O:D_2O= 9:1). The 1D 1H NMR spectra at different ratios of peptide to RNA (0:1, 0.25:1, 0.5:1, 1:1, 1.5:1, 2:1, 2.5:1) were collected. The 2D TOCSY spectra were collected for free RNA and peptide/RNA (2.5:1) sample. All spectra were collected at 5 °C unless specified.

Section 2.4.4: NMR Spectroscopy and Structure Determination

Peptide samples at a concentration of 2-5 mg/mL were prepared in H_2O:D_2O (9:1) 10 mM phosphate buffer at pH 6.5. 1D and 2D NMR spectra were collected on Bruker DRX 500 MHz or Avance III 600 MHz spectrometer equipped with cryo-cooled probes. 2D NMR data (TOCSY, NOESY, HSQC) were processed with TOPSPIN (Bruker) and analyzed in NMRFAM-SPARKY (Lee et al., 2015). Two mixing time (100 ms and 300 ms) were used for 2D ^{1}H NOESY spectra to assist quantitative evaluation of internuclear distances.

Peptide structures were calculated with Xplor-NIH using torsional angle dynamics and simulated annealing from a starting peptide template generated with randomized torsion angles (Schwieters et al., 2003). Interproton distance restraints were derived from NOE peaks in 2D ^{1}H NOESY spectra and sorted into strong (2.5 ± 0.7 Å), medium (3.5 ± 1.2 Å) and weak (4.5 ± 1.5 Å) bins based on relative peak intensities. For each peptide, 200 structures were calculated and 10 structures with the lowest energy were selected for analysis from the population of converged structures.

Section 2.5: References

Afroz T., Cienikova Z., Clery A., & Allain F. H. (2015). One, Two, Three, Four! How Multiple RRMs Read the Genome Sequence. *Methods Enzymol.*, *558*, 235-278. https://doi.org/10.1016/bs.mie.2015.01.015

Athanassiou Z., Dias R. L., Moehle K., Dobson N., Varani G., & Robinson J. A. (2004). Structural mimicry of retroviral tat proteins by constrained beta-hairpin peptidomimetics: ligands with high affinity and selectivity for viral TAR RNA regulatory elements. *J. Am. Chem. Soc.*, *126*(22), 6906-6913. https://doi.org/10.1021/ja0497680

Athanassiou Z., Patora K., Dias R. L., Moehle K., Robinson J. A., & Varani G. (2007). Structure-guided peptidomimetic design leads to nanomolar beta-hairpin inhibitors of the Tat-TAR interaction of bovine immunodeficiency virus. *Biochemistry*, *46*(3), 741-751. https://doi.org/10.1021/bi0619371

Auweter S. D., Fasan R., Reymond L., Underwood J. G., Black D. L., Pitsch S., & Allain F. H. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.*, *25*(1), 163-173. https://doi.org/10.1038/sj.emboj.7600918

Bartel D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, *136*(2), 215-233. https://doi.org/10.1016/j.cell.2009.01.002

Beckert B., & Masquida B. (2011). Synthesis of RNA by in vitro transcription. *Methods Mol. Biol.*, *703*, 29-41. https://doi.org/10.1007/978-1-59745-248-9_3

Belashov I. A., Crawford D. W., Cavender C. E., Dai P., Beardslee P. C., Mathews D. H., Pentelute B. L., McNaughton B. R., & 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. https://doi.org/10.1093/nar/gky529

Burd C. G., & Dreyfuss G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science*, *265*(5172), 615-621. https://www.ncbi.nlm.nih.gov/pubmed/8036511

Chen Y., Zubovic L., Yang F., Godin K., Pavelitz T., Castellanos J., Macchi P., & Varani G. (2016). Rbfox proteins regulate microRNA biogenesis by sequence-specific binding to their precursors and target downstream Dicer. *Nucleic Acids Res.*, *44*(9), 4381-4395. https://doi.org/10.1093/nar/gkw177

Clery A., Blatter M., & Allain F. H. (2008). RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, *18*(3), 290-298. https://doi.org/10.1016/j.sbi.2008.04.002

Clery A., Jayne S., Benderska N., Dominguez C., Stamm S., & Allain F. H. (2011). Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-beta1. *Nat. Struct. Mol. Biol.*, *18*(4), 443-450. https://doi.org/10.1038/nsmb.2001

Clery A., Sinha R., Anczukow O., Corrionero A., Moursy A., Daubner G. M., Valcarcel J., Krainer A. R., & Allain F. H. (2013). Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(30), E2802-2811. https://doi.org/10.1073/pnas.1303445110

Craik D. J., Fairlie D. P., Liras S., & Price D. (2013). The future of peptide-based drugs. *Chem. Biol. Drug Des.*, *81*(1), 136-147. https://doi.org/10.1111/cbdd.12055

Croce C. M. (2009). Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, *10*(10), 704-714. https://doi.org/10.1038/nrg2634

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. https://doi.org/10.1073/pnas.0900629106

Dominguez C., Fisette J. F., Chabot B., & Allain F. H. (2010). Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.*, *17*(7), 853-861. https://doi.org/10.1038/nsmb.1814

Fabian M. R., Sonenberg N., & Filipowicz W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.*, *79*, 351-379. https://doi.org/10.1146/annurev-biochem-060308-103103

Fields G. B., & Noble R. L. (1990). Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int. J. Pept. Protein Res.*, *35*(3), 161-214. https://www.ncbi.nlm.nih.gov/pubmed/2191922

Frank A. O., Vangamudi B., Feldkamp M. D., Souza-Fagundes E. M., Luzwick J. W., Cortez D., Olejniczak E. T., Waterson A. G., Rossanese O. W., Chazin W. J., et al. (2014). Discovery of a potent stapled helix peptide that binds to the 70N domain of replication protein A. *J. Med. Chem.*, *57*(6), 2455-2461. https://doi.org/10.1021/jm401730y

Gibbs A. C., Kondejewski L. H., Gronwald W., Nip A. M., Hodges R. S., Sykes B. D., & Wishart D. S. (1998). Unusual beta-sheet periodicity in small cyclic peptides. *Nat. Struct. Biol.*, *5*(4), 284-288. https://www.ncbi.nlm.nih.gov/pubmed/9546219

Gongora-Benitez M., Tulla-Puche J., & Albericio F. (2014). Multifaceted roles of disulfide bonds. Peptides as therapeutics. *Chem. Rev.*, *114*(2), 901-926. https://doi.org/10.1021/cr400031z

Ha M., & Kim V. N. (2014). Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.*, *15*(8), 509-524. https://doi.org/10.1038/nrm3838

Im H. I., & Kenny P. J. (2012). MicroRNAs in neuronal function and dysfunction. *Trends Neurosci.*, *35*(5), 325-334. https://doi.org/10.1016/j.tins.2012.01.004

Jin Y., Suzuki H., Maegawa S., Endo H., Sugano S., Hashimoto K., Yasuda K., & Inoue K. (2003). A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J.*, *22*(4), 905-912. https://doi.org/10.1093/emboj/cdg089

Krol J., Loedige I., & Filipowicz W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, *11*(9), 597-610. https://doi.org/10.1038/nrg2843

Kuroyanagi H. (2009). Fox-1 family of RNA-binding proteins. *Cell. Mol. Life Sci.*, *66*(24), 3895-3907. https://doi.org/10.1007/s00018-009-0120-5

Lalonde M. S., Lobritz M. A., Ratcliff A., Chamanian M., Athanassiou Z., Tyagi M., Wong J., Robinson J. A., Karn J., Varani G., et al. (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. https://doi.org/10.1371/journal.ppat.1002038

Lee W., Tonelli M., & Markley J. L. (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, *31*(8), 1325-1327. https://doi.org/10.1093/bioinformatics/btu830

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. https://doi.org/10.1021/bi010680y

Li Y., Mao L., Gao Y., Baral S., Zhou Y., & Hu B. (2015). MicroRNA-107 contributes to post-stroke angiogenesis by targeting Dicer-1. *Sci. Rep.*, *5*, 13316. https://doi.org/10.1038/srep13316

Maris C., Dominguez C., & Allain F. H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, *272*(9), 2118-2131. https://doi.org/10.1111/j.1742-4658.2005.04653.x

Martello G., Rosato A., Ferrari F., Manfrin A., Cordenonsi M., Dupont S., Enzo E., Guzzardo V., Rondina M., Spruce T., et al. (2010). A MicroRNA targeting dicer for metastasis control. *Cell*, *141*(7), 1195-1207. https://doi.org/10.1016/j.cell.2010.05.017

Michlewski G., Guil S., Semple C. A., & Caceres J. F. (2008). Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol. Cell*, *32*(3), 383-393. https://doi.org/10.1016/j.molcel.2008.10.013

Milligan J. F., Groebe D. R., Witherell G. W., & Uhlenbeck O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, *15*(21), 8783-8798. https://www.ncbi.nlm.nih.gov/pubmed/3684574

Newman M. A., Thomson J. M., & Hammond S. M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA*, *14*(8), 1539-1549. https://doi.org/10.1261/rna.1155108

Oberstrass F. C., Auweter S. D., Erat M., Hargous Y., Henning A., Wenter P., Reymond L., Amir-Ahmady B., Pitsch S., Black D. L., et al. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, *309*(5743), 2054-2057. https://doi.org/10.1126/science.1114066

Price S. R., Evans P. R., & Nagai K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature*, *394*(6694), 645-650. https://doi.org/10.1038/29234

Ristori E., Lopez-Ramirez M. A., Narayanan A., Hill-Teran G., Moro A., Calvo C. F., Thomas J. L., & Nicoli S. (2015). A Dicer-miR-107 Interaction Regulates Biogenesis of Specific miRNAs Crucial for Neurogenesis. *Dev. Cell*, *32*(5), 546-560. https://doi.org/10.1016/j.devcel.2014.12.013

Sakamoto K., Otake K., & Umemoto T. (2017). Discovery of peptidic miR-21 processing inhibitor by mirror image phage display: A novel method to generate RNA binding D-peptides. *Bioorg. Med. Chem. Lett.*, *27*(4), 826-828. https://doi.org/10.1016/j.bmcl.2017.01.023

Schwieters C. D., Kuszewski J. J., Tjandra N., & Clore G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, *160*(1), 65-73. https://www.ncbi.nlm.nih.gov/pubmed/12565051

Skrisovska L., Bourgeois C. F., Stefl R., Grellscheid S. N., Kister L., Wenter P., Elliott D. J., Stevenin J., & Allain F. H. (2007). The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep.*, *8*(4), 372-379. https://doi.org/10.1038/sj.embor.7400910

Tintaru A. M., Hautbergue G. M., Hounslow A. M., Hung M. L., Lian L. Y., Craven C. J., & Wilson S. A. (2007). Structural and functional analysis of RNA and TAP binding to SF2/ASF. *EMBO Rep.*, *8*(8), 756-762. https://doi.org/10.1038/sj.embor.7401031

Tsuda K., Kuwasako K., Takahashi M., Someya T., Inoue M., Terada T., Kobayashi N., Shirouzu M., Kigawa T., Tanaka A., et al. (2009). Structural basis for the sequence-specific RNA-recognition mechanism of human CUG-BP1 RRM3. *Nucleic Acids Res.*, *37*(15), 5151-5166. https://doi.org/10.1093/nar/gkp546

Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G., Smith H. O., Yandell M., Evans C. A., Holt R. A., et al. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304-1351. https://doi.org/10.1126/science.1058040

Zhou W., Shi G., Zhang Q., Wu Q., Li B., & Zhang Z. (2014). MicroRNA-20b promotes cell growth of breast cancer cells partly via targeting phosphatase and tensin homologue (PTEN). *Cell Biosci.*, *4*(1), 62. https://doi.org/10.1186/2045-3701-4-62

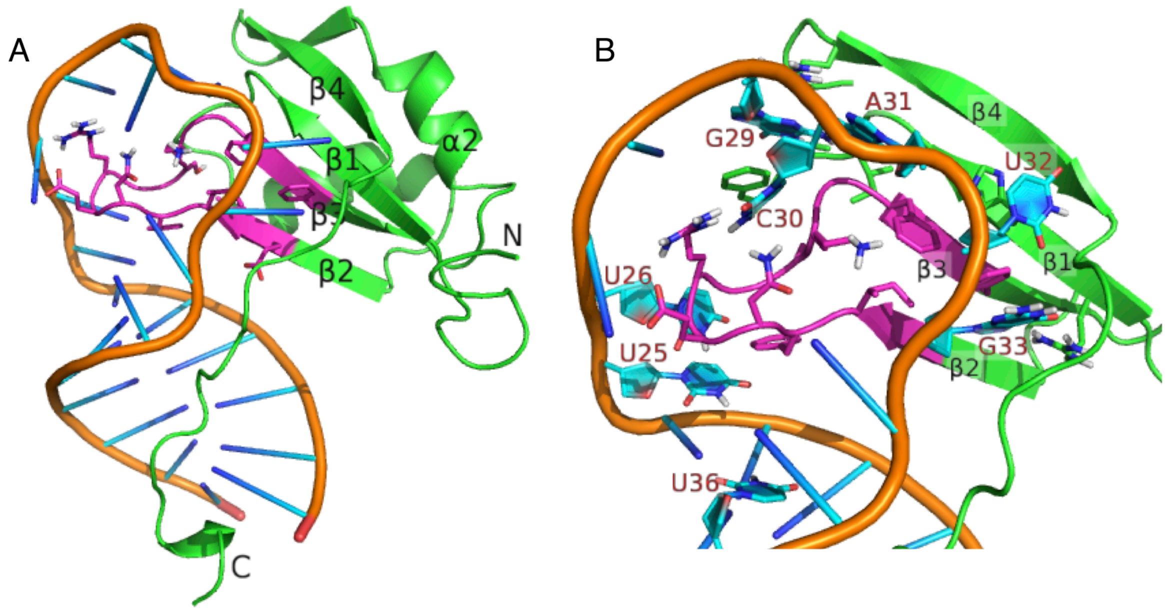


Figure 2.1. (A) The NMR structure of the complex of Rbfox2 RRM protein and pre-miR-20b RNA. (B) A closer look into the Rbfox2 RRM and pre-miR-20b interface (Chen et al., 2016). (PDB: 2N82) The β2β3 loop mimicked in our designed cyclic peptides is colored in magenta in both images.

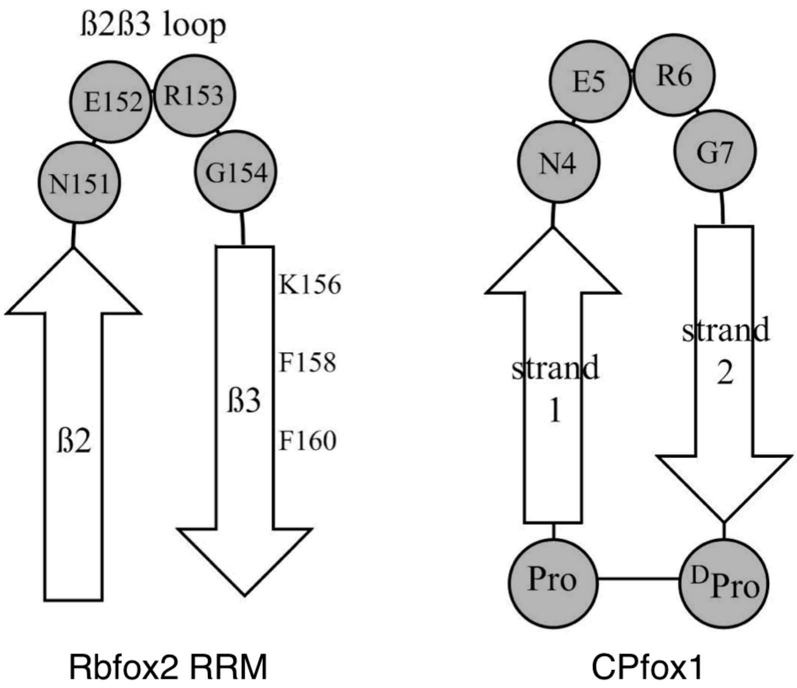


Figure 2.2. Illustration of CPfox1 cyclic peptide design. Residues 148-159 of the Rbfox2 RRM, including the β2β3 hairpin, are transplanted onto the DPro-LPro template to generate a cyclic peptides structure that presents the same amino acids in the apical loop. Subsequent designs are based on this simple transplant.

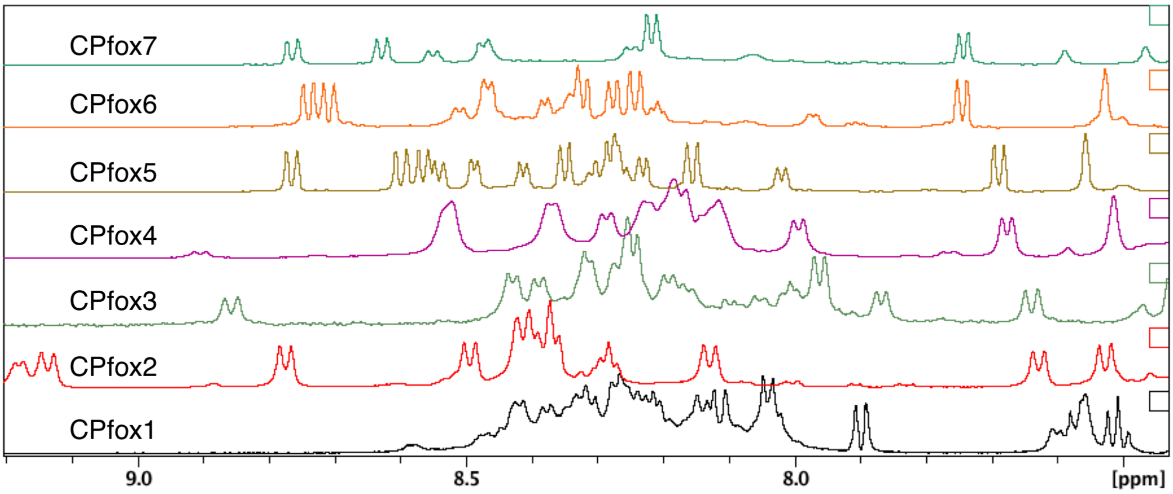


Figure 2.3. Overlay of 1D NMR spectra of cyclic peptide mimics. The narrow diversion of HN peaks and broad peak shape of CPfox1 show that the peptide is unfolded; the widely dispersed HN peaks of CPfox2 suggest that the peptide forms highly structured β-hairpin; the moderate dispersion of HN peaks of CPfox3-7 suggest that the peptides form less stable β-hairpin structure compared to CPfox2.

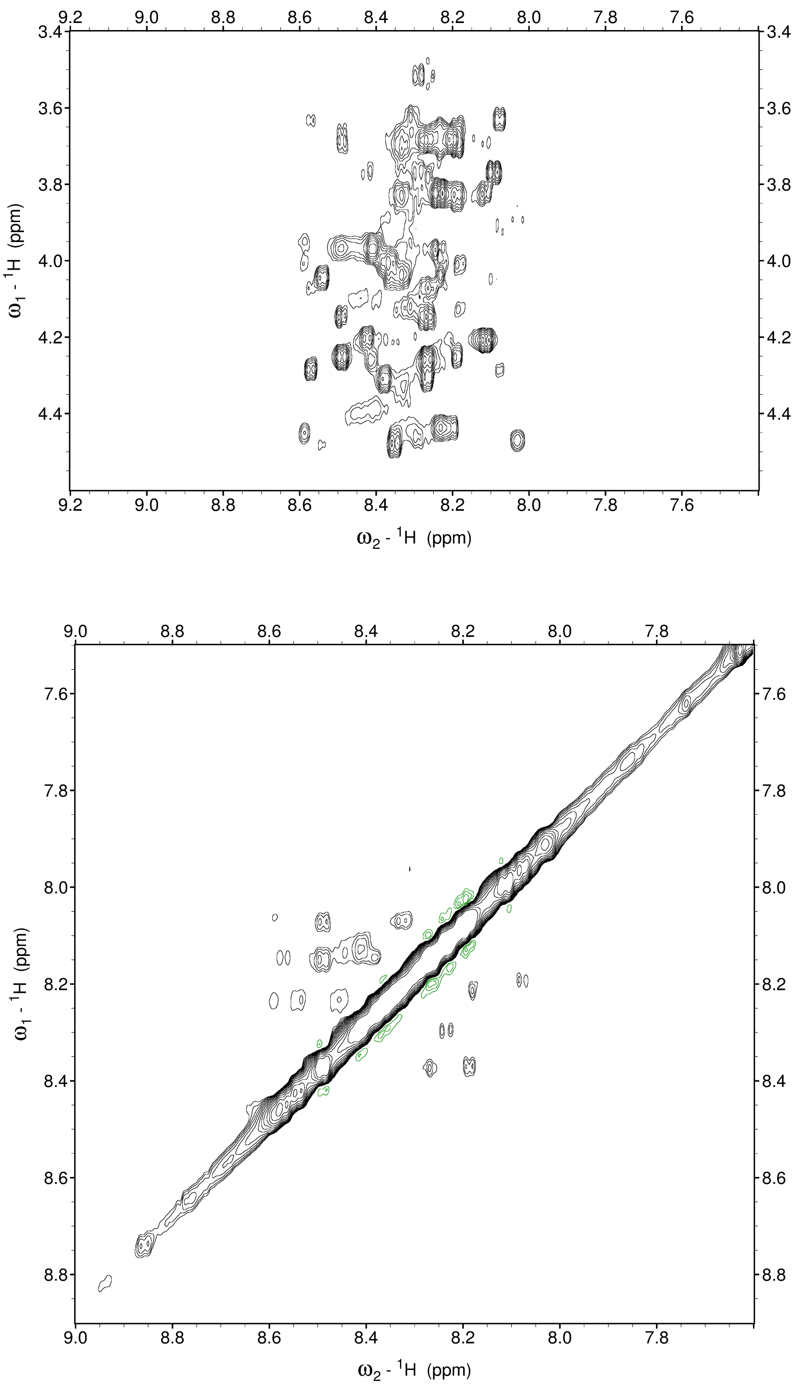


Figure 2.4. HN-Hα and HN-HN regions of the 2D NOESY spectra for CPfox1. Main chain Hα-HN and HN-HN NOE assignments are shown.

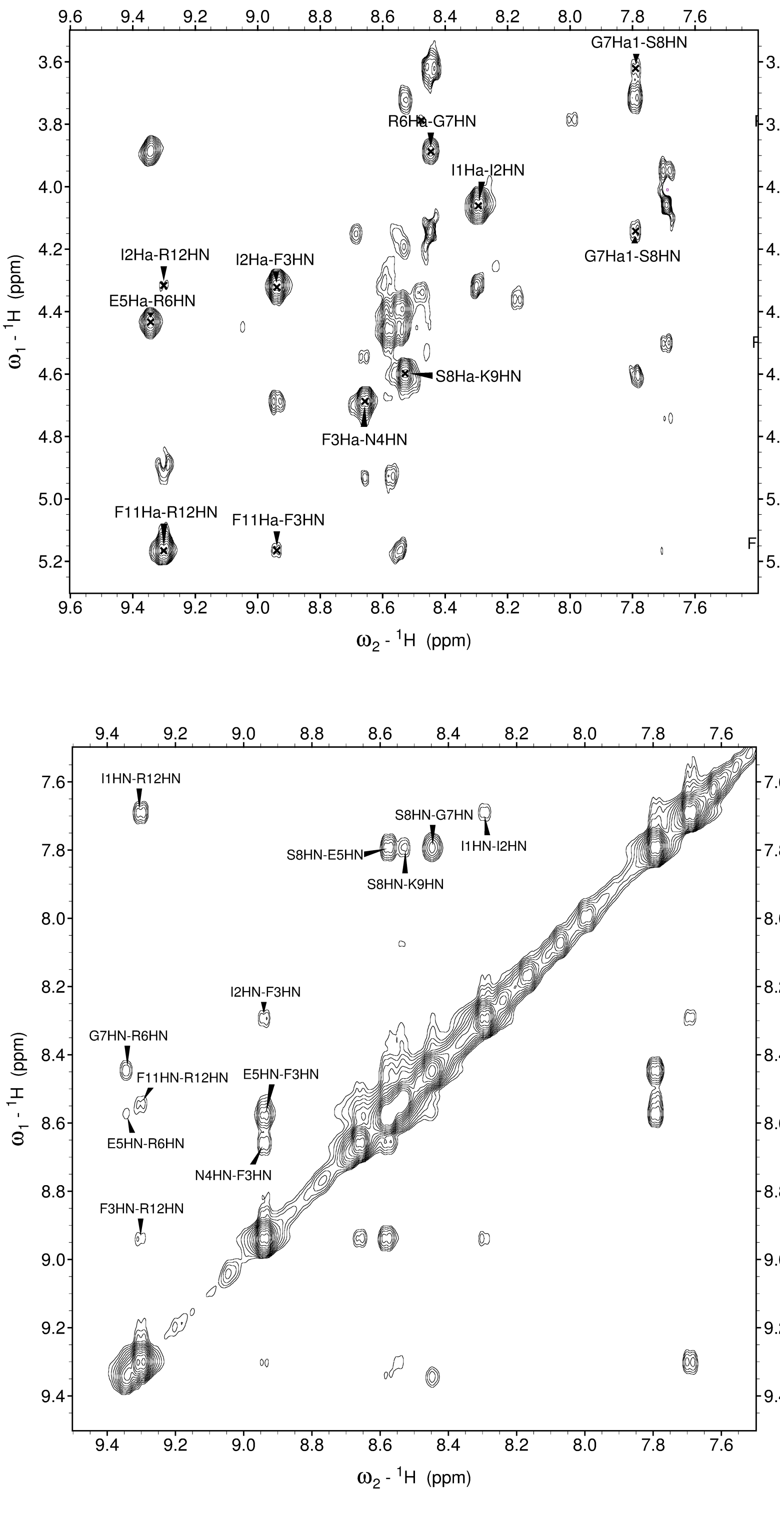


Figure 2.5. HN-Hα and HN-HN regions of the 2D NOESY spectra for CPfox2. Main chain Hα-HN and HN-HN NOE assignments are shown.

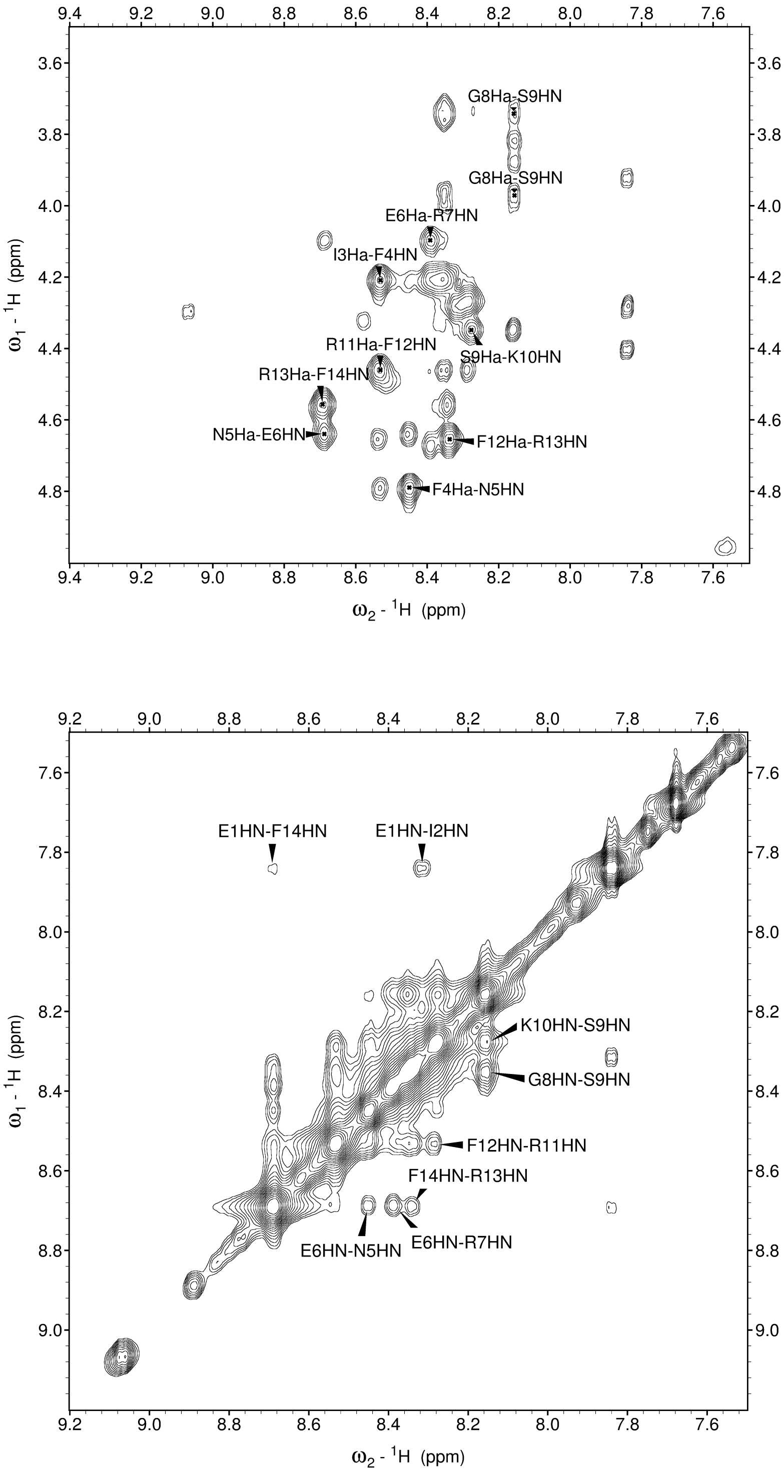


Figure 2.6. HN-Hα and HN-HN regions of the 2D NOESY spectra for CPfox4. Main chain Hα-HN and HN-HN NOE assignments are shown.

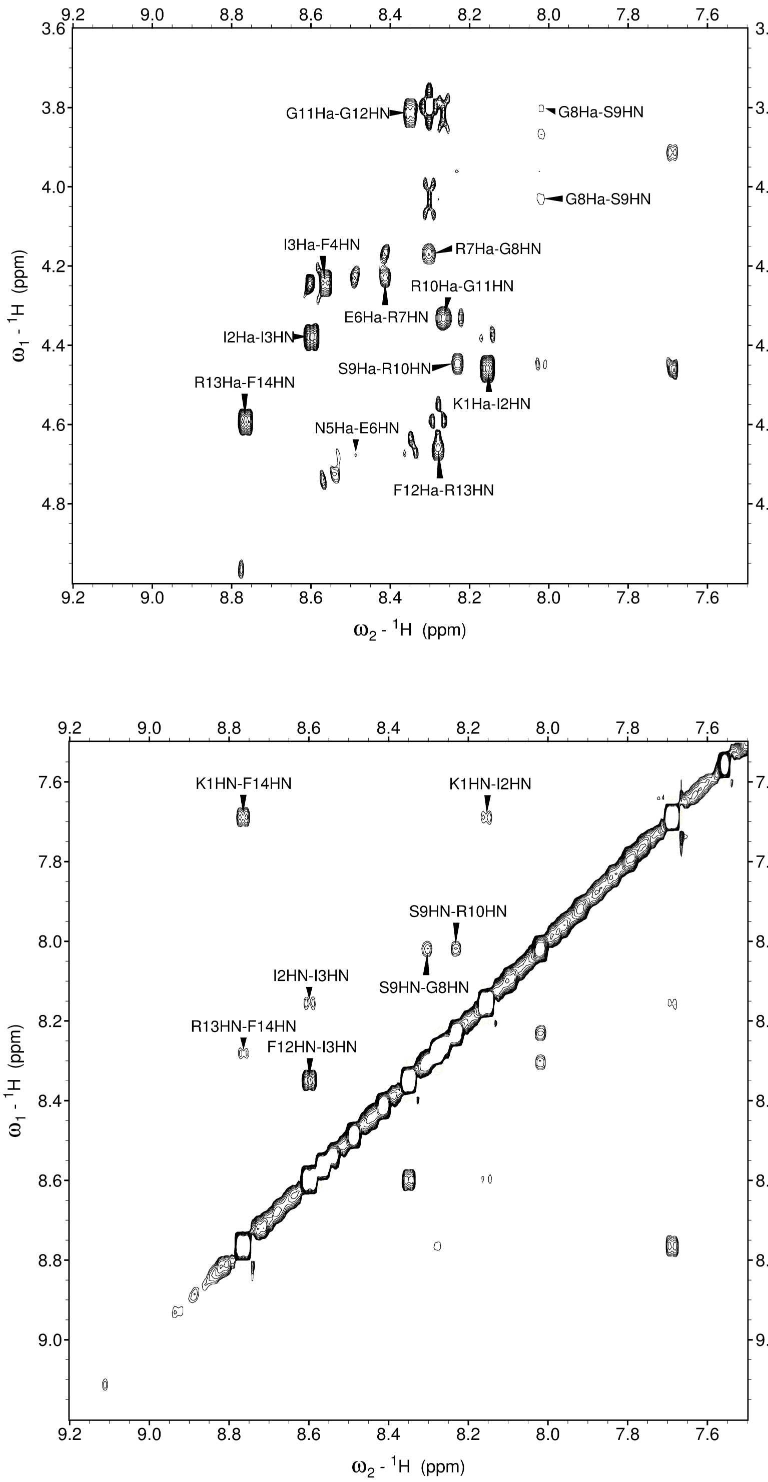


Figure 2.7. HN-Hα and HN-HN regions of the 2D NOESY spectra for CPfox5. Main chain Hα-HN and HN-HN NOE assignments are shown.

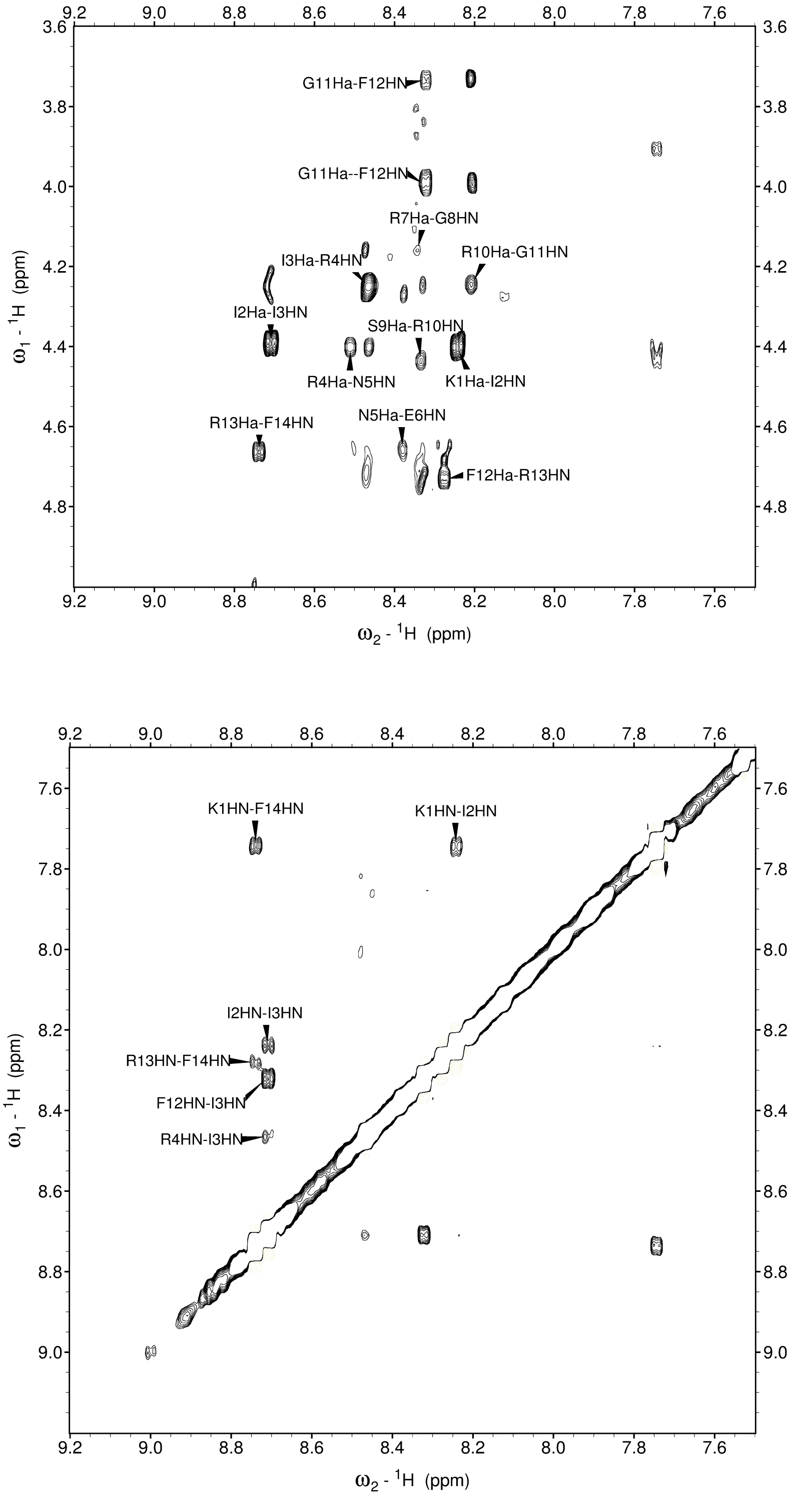


Figure 2.8. HN-Hα and HN-HN regions of the 2D NOESY spectra for CPfox6. Main chain Hα-HN and HN-HN NOE assignments are shown.

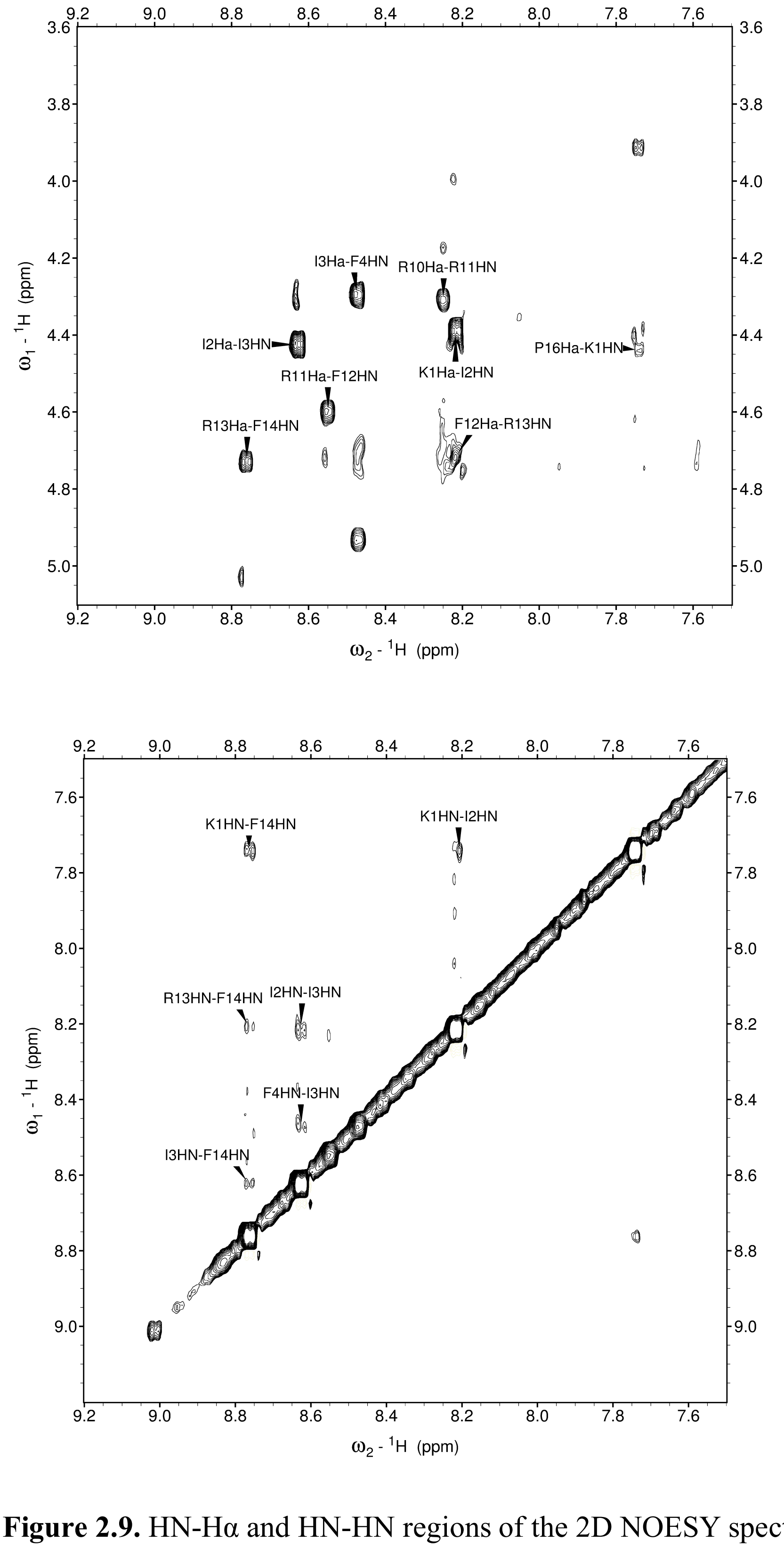


Figure 2.9. HN-Hα and HN-HN regions of the 2D NOESY spectra for CPfox7. Main chain Hα-HN and HN-HN NOE assignments are shown.

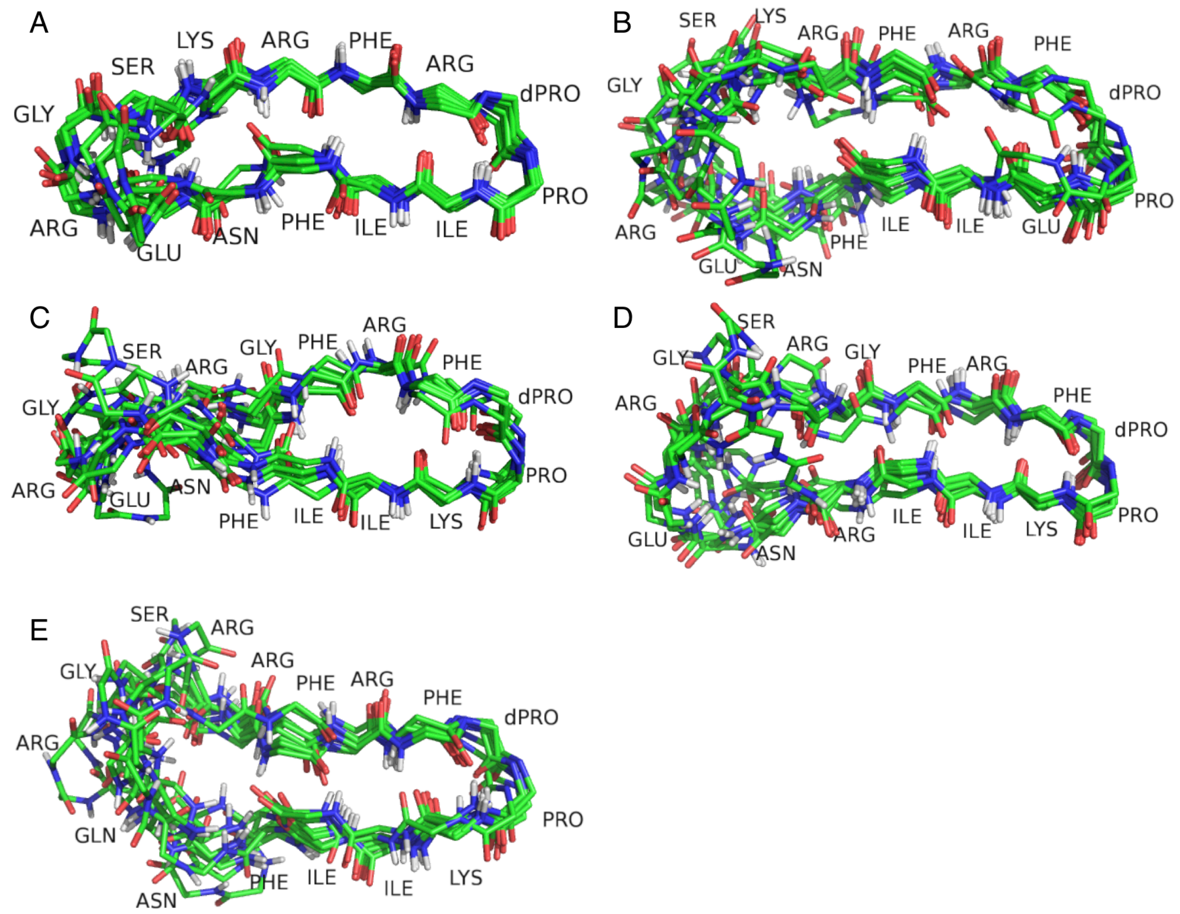


Figure 2.10. Superposition of the 10 NMR structures of lowest energy for (A) CPfox2, (B) CPfox4, (C) CPfox5, (D) CPfox6 and (E) CPfox7. In each image, the β2β3 loop mimic is on the left, and the ^{D}Pro-^{L}Pro template on the right.

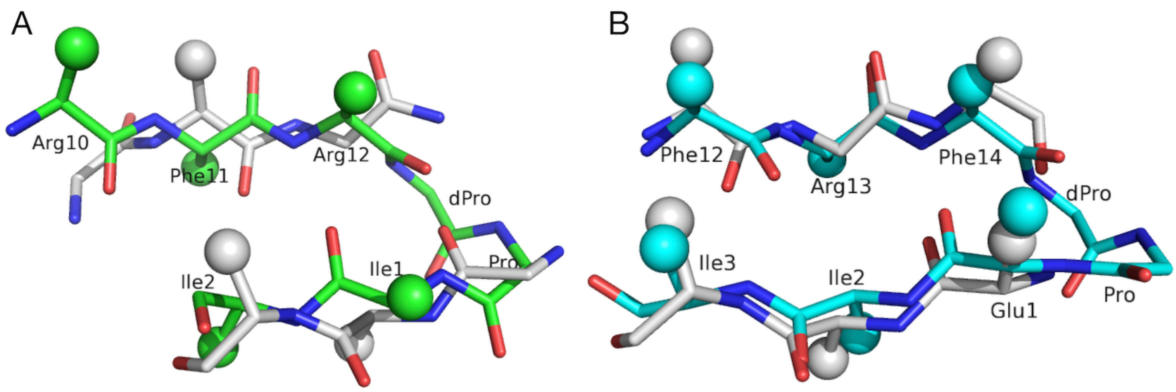


Figure 2.11. Superposition of (A) the NMR structure of CPfox2 (lowest energy, green) and Rbfox2 RRM β2β3 (grey) and (B) the structure of CPfox4 (lowest energy, cyan) and Rbfox2 RRM β2β3 (grey). Amino acid side chains are represented as spheres.

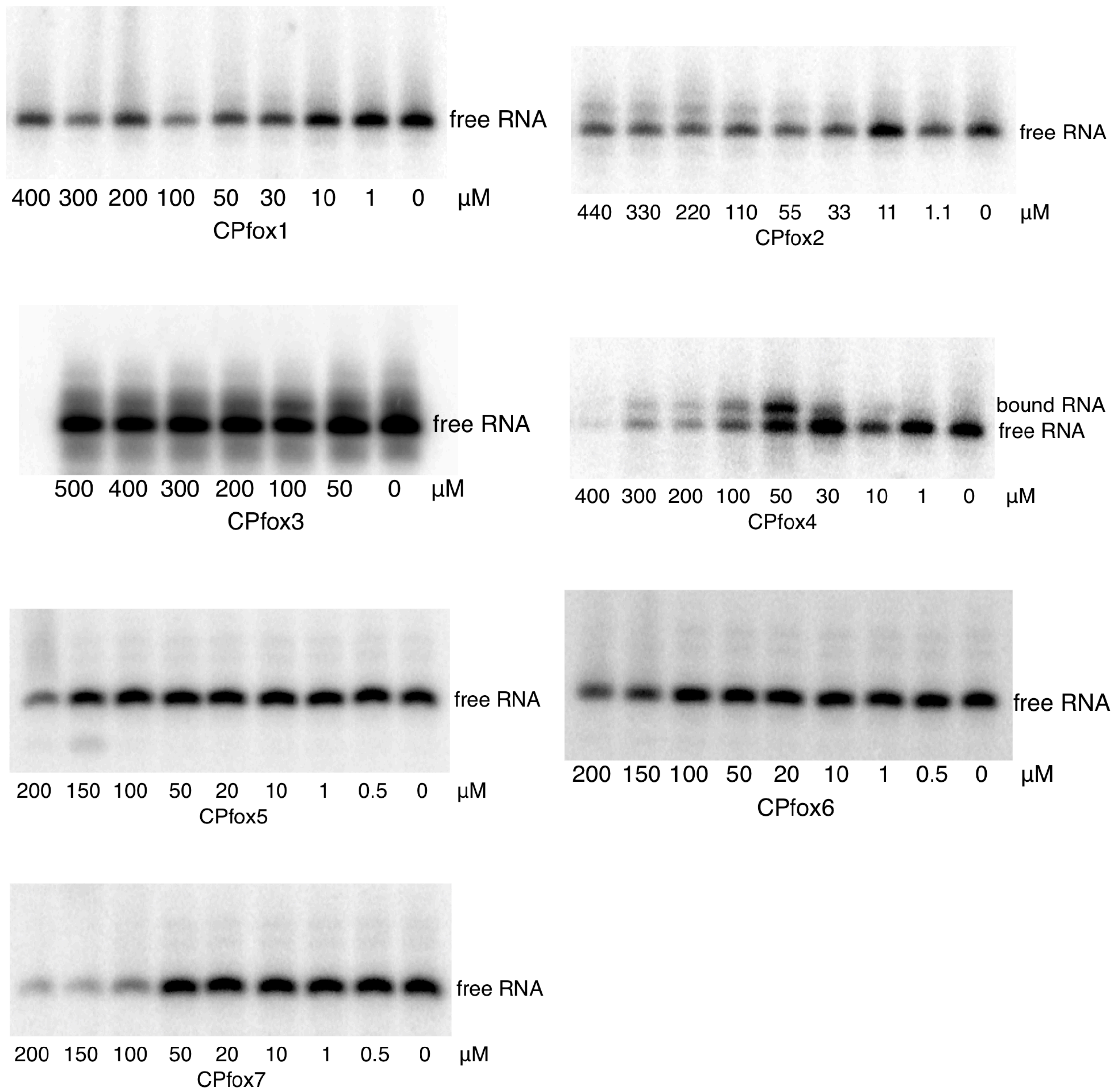


Figure 2.12. EMSA of Rbfox2 peptide mimetics binding to pre-miR-20b. Various concentration of peptide and 30 nM RNA were incubated in Tris buffer (pH7.4) containing 250-fold excess tRNA, and 0.25% Triton X-100 for 30 minutes. The samples were loaded onto 12% native polyacrylamide gel in 0.5X TBE buffer and electrophoresed at 300 V and 4 W for 2 hours. The decreases of band intensity observed above 100 μM are because of the limited solubility the peptides.

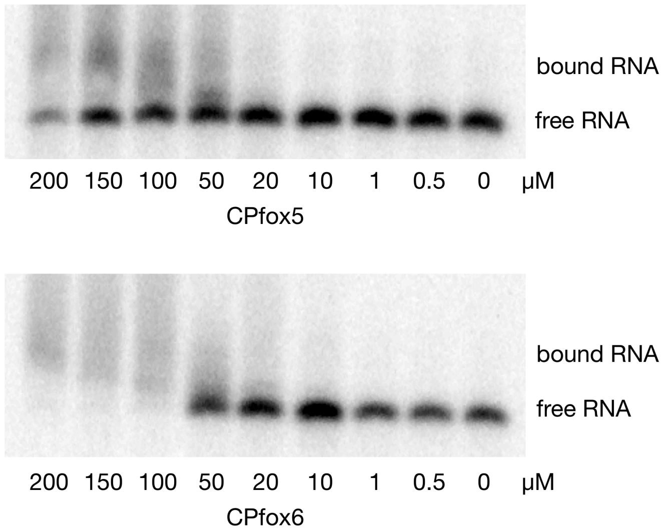


Figure 2.13. EMSA of CPfox5 and CPfox6 binding to pre-miR-20b in the absence of competition tRNA. Various concentration of peptide and 30 nM RNA were incubated in Tris buffer (pH7.4), and 0.25% Triton X-100 for 30 minutes. The samples were loaded onto 12% native polyacrylamide gel in 0.5X TBE buffer and electrophoresed at 300 V and 4 W for 2 hours.

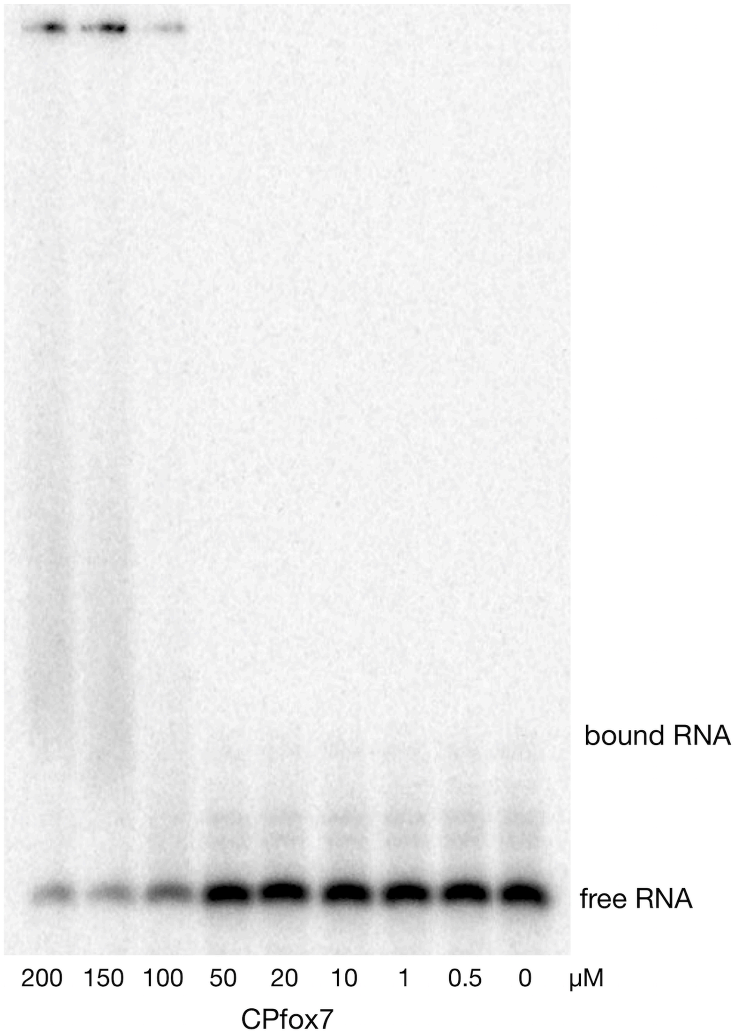


Figure 2.14. EMSA of CPfox7 binding to pre-miR-20b. Various concentration of peptide and 30 nM RNA were incubated in Tris buffer (pH7.4) containing 250-fold excess tRNA, and 0.25% Triton X-100 for 30 minutes. The samples were loaded onto 12% native polyacrylamide gel in 0.5X TBE buffer and electrophoresed at 300 V and 4 W for 2 hours. The broadly distributed binding band of CPfox7/pre-miR-20b implies the fast exchange between the bound and free forms of the RNA on the gel.

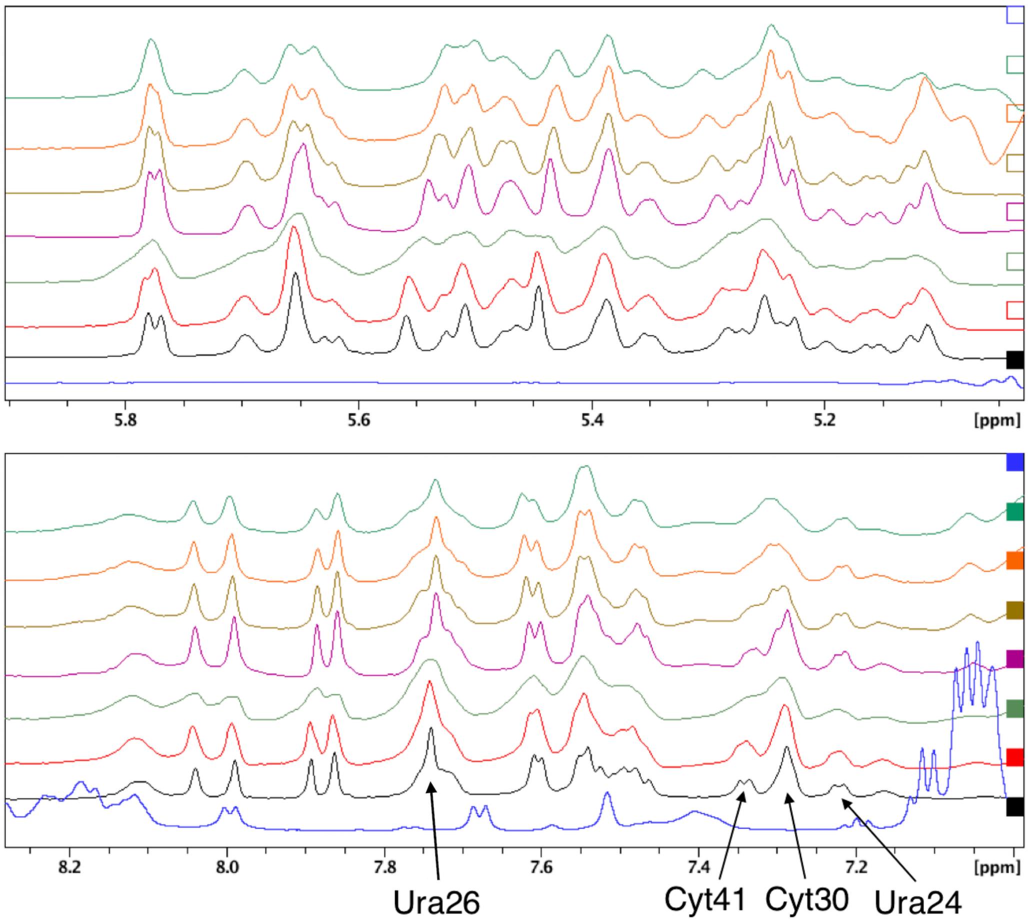


Figure 2.15. 1D ^{1}H NMR titration pre-miR-20b RNA with CPfox4. Peptide and RNA ratios are 0:1 (black), 0.25:1 (red), 0.5:1 (dark green), 1:1 (purple), 1.5:1 (brown), 2:1 (orange), 2.5:1 (green). The spectrum of free CPfox4 is shown in blue.

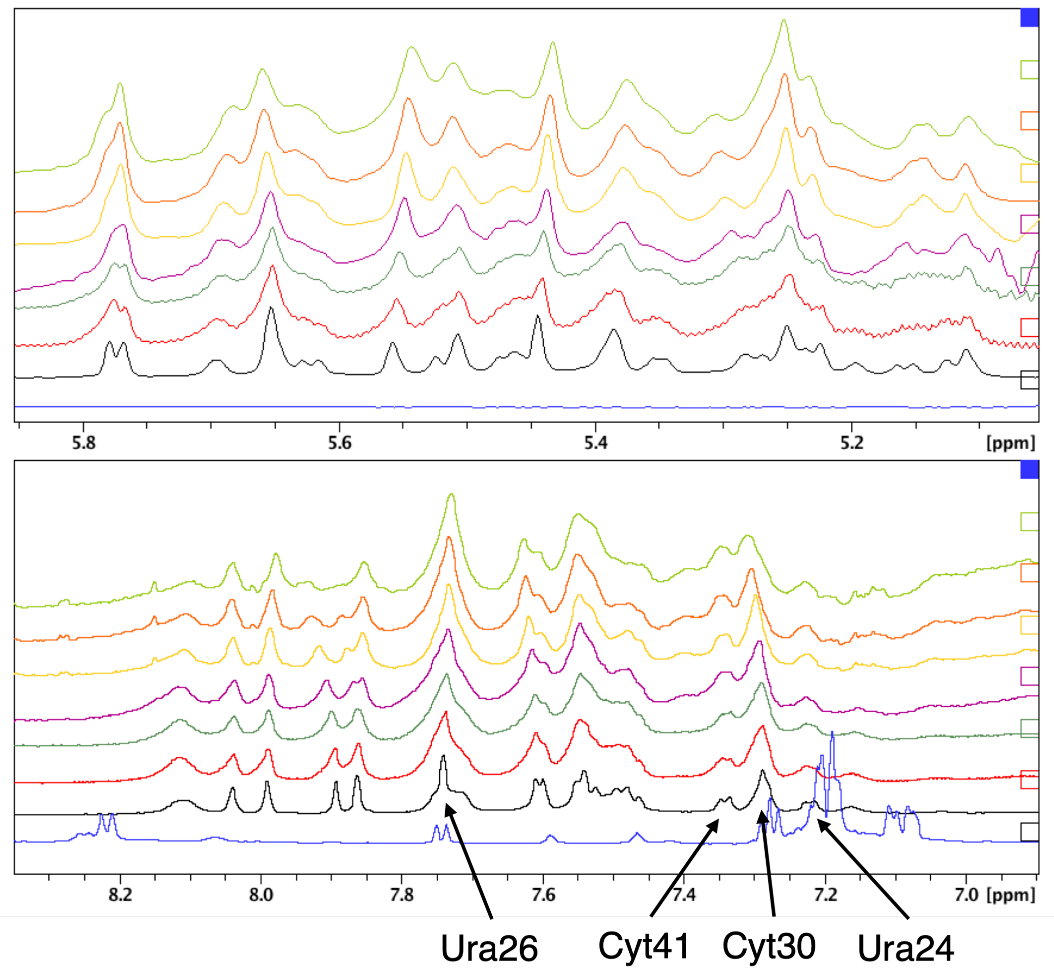


Figure 2.16. 1D ^{1}H NMR titration pre-miR-20b RNA with CPfox7. Peptide and RNA ratios are 0:1 (black), 0.25:1 (red), 0.5:1 (dark green), 1:1 (purple), 1.5:1 (yellow), 2:1 (orange), 2.5:1 (green). The spectrum of free CPfox7 is shown in blue.

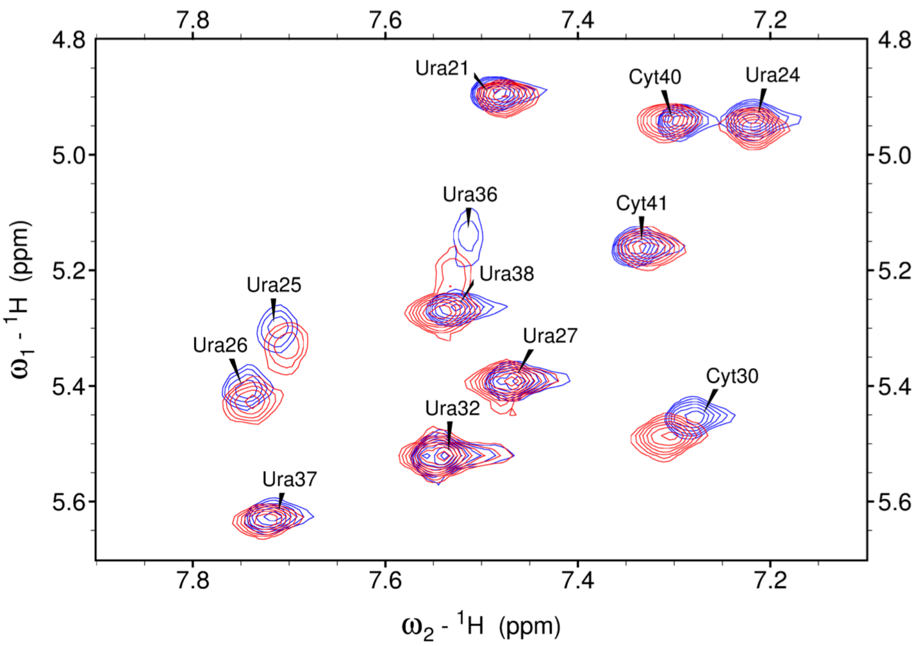


Figure 2.17. 2D TOCSY spectra of free pre-miR-20b (blue) and CPfox4/pre-miR-20b (2.5:1) mixture (red). Peaks of U25, U26, C30 and U36 shift after the addition of CPfox4.

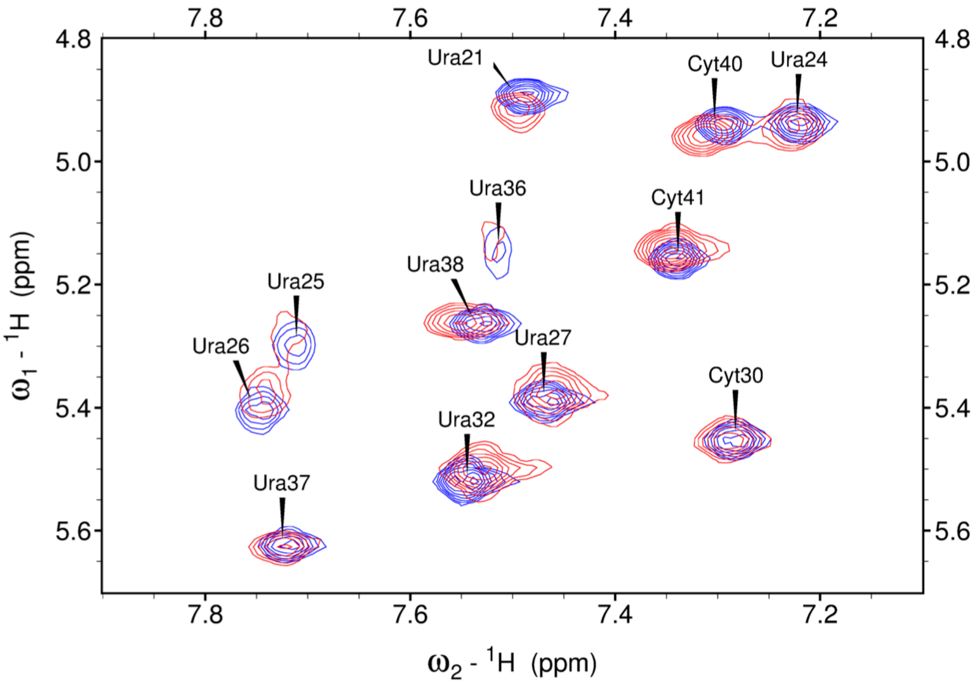


Figure 2.18. 2D TOCSY spectra of free pre-miR-20b (blue) and CPfox7/pre-miR-20b (2.5:1) mixture (red).

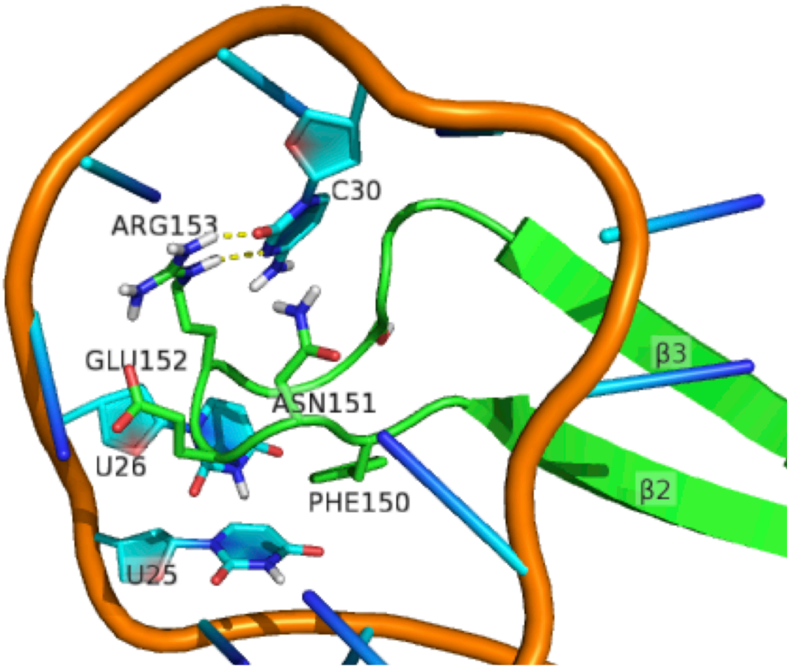


Figure 2.19. The interface between Rbfox2 RRM β2β3 loop and pre-miR-20b has C30 in the center, in direct contact with the amino acid side chains of the protein (PDB:2N82).

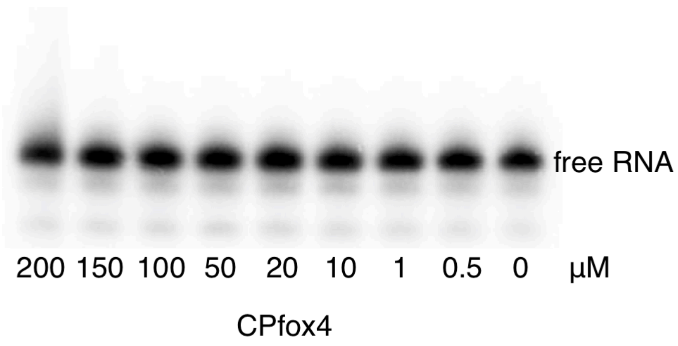


Figure 2.20. EMSA of CPfox4 binding to HIV TAR RNA. Various concentration of peptide and 30 nM RNA were incubated in Tris buffer (pH7.4) containing 250-fold excess tRNA, and 0.25% Triton X-100 for 30 minutes. The samples were loaded onto 12% native polyacrylamide gel in 0.5X TBE buffer and electrophoresed at 300 V and 4 W for 2 hours.

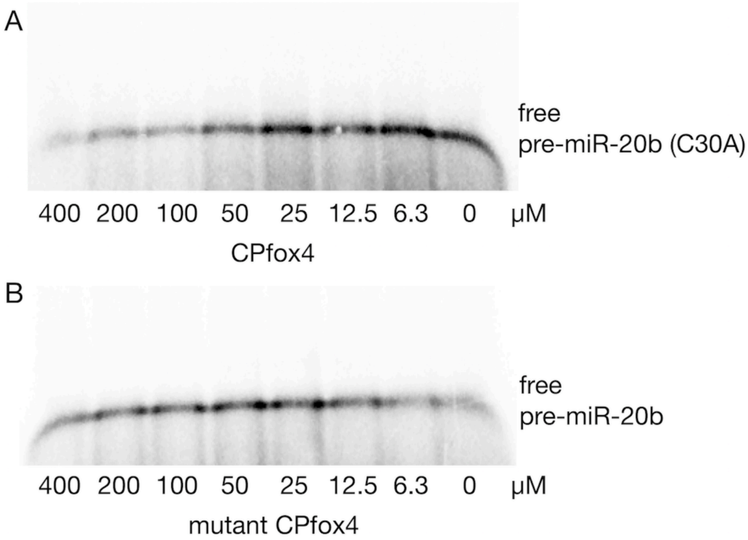
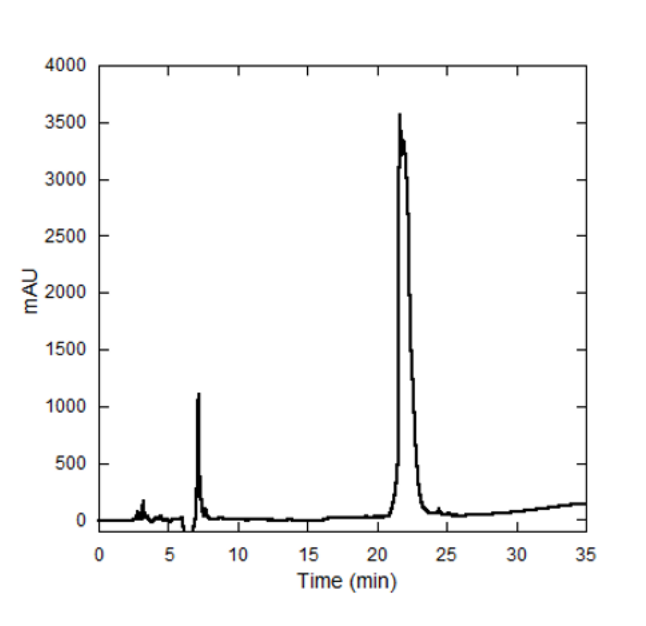


Figure 2.21. EMSA of (A) CPfox4 binding to mutant pre-miR-20b (C30A) and (B) mutant CPfox4 (cyclo-EIIFNE**A**GSKRFRFpP) binding to pre-miR-20b. Various concentration of peptide and 10 nM RNA were incubated in HEPES buffer (pH7.2) containing 250-fold excess tRNA, and 0.25% Triton X-100 for 30 minutes. The samples were loaded onto 12% native polyacrylamide gel in 0.5X TBE buffer and electrophoresed at 300 V and 4 W for 2 hours.



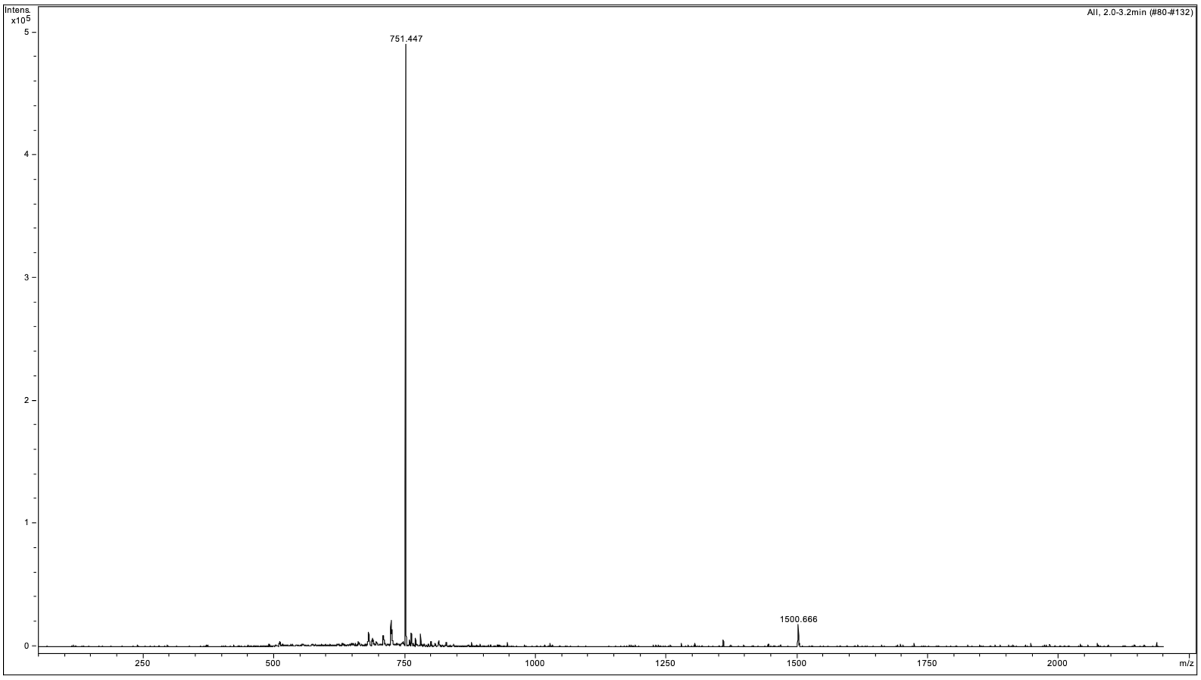
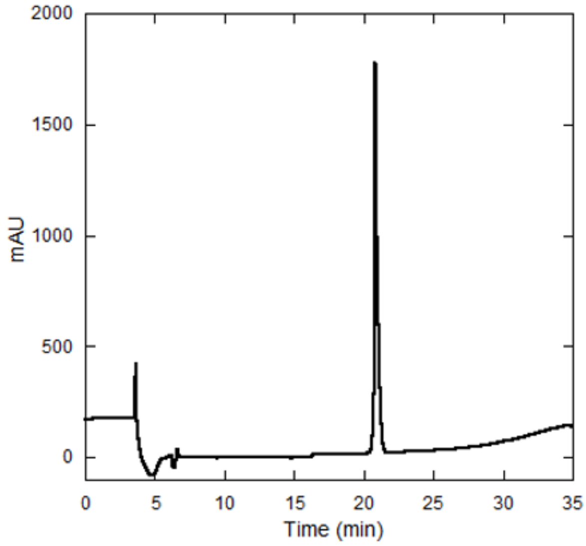


Figure 2.22. Chromatogram and mass spectrometry (ESI-ion trap) of CPfox1. The exact mass of CPfox1 is1499.79. $[MH^{+}]$= 1500.67, $[MH_2^{2+}]$= 751.45 were observed.



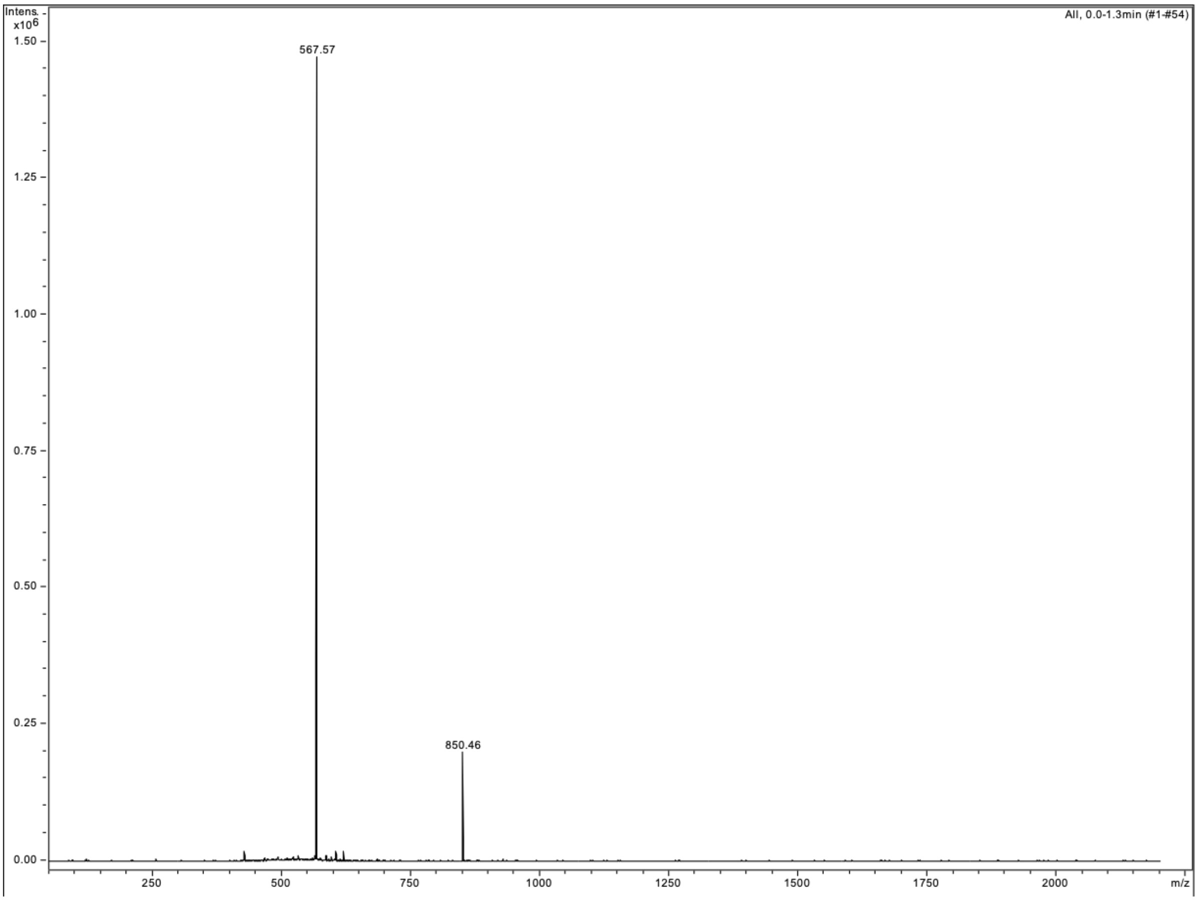


Figure 2.23. Chromatogram and mass spectrometry (ESI-ion trap) of CPfox2. The exact mass of CPfox2 is 1697.95. $[MH_2^{2+}]$= 850.46, $[MH_3^{3+}]$= 567.57 were observed.

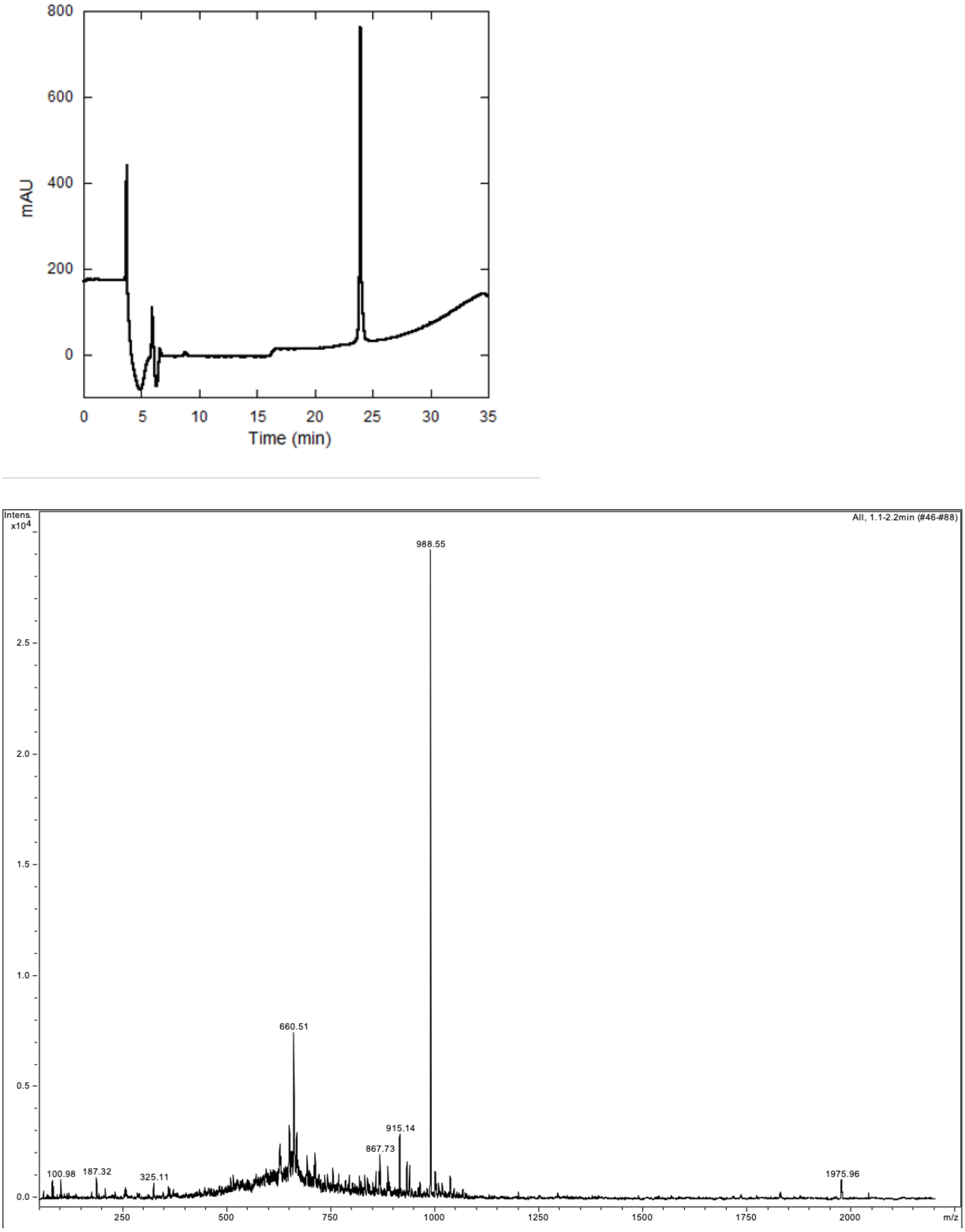
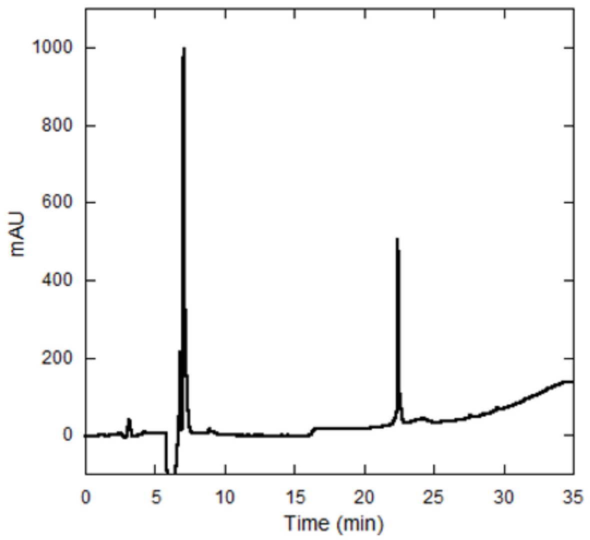


Figure 2.24. Chromatogram and mass spectrometry (ESI-ion trap) of CPfox3. The exact mass of CPfox3 is 1974.04. [MH^+]= 1975.96, [MH_2^{2+}]= 988.55, [MH_3^{3+}]= 660.51 were observed.



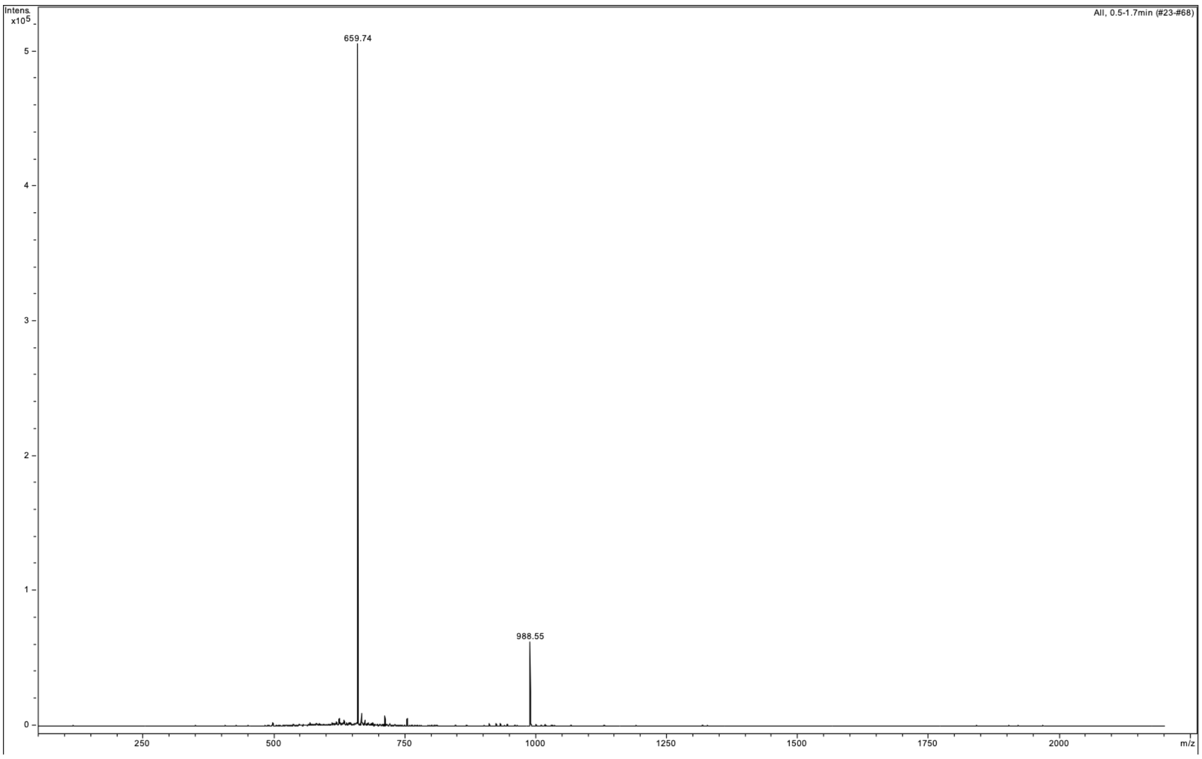


Figure 2.25. Chromatogram and mass spectrometry (ESI-ion trap) of CPfox4. The exact mass of CPfox4 is 1974.06. [MH_2^{2+}]= 988.55, [MH_3^{3+}]= 659.74 were observed.

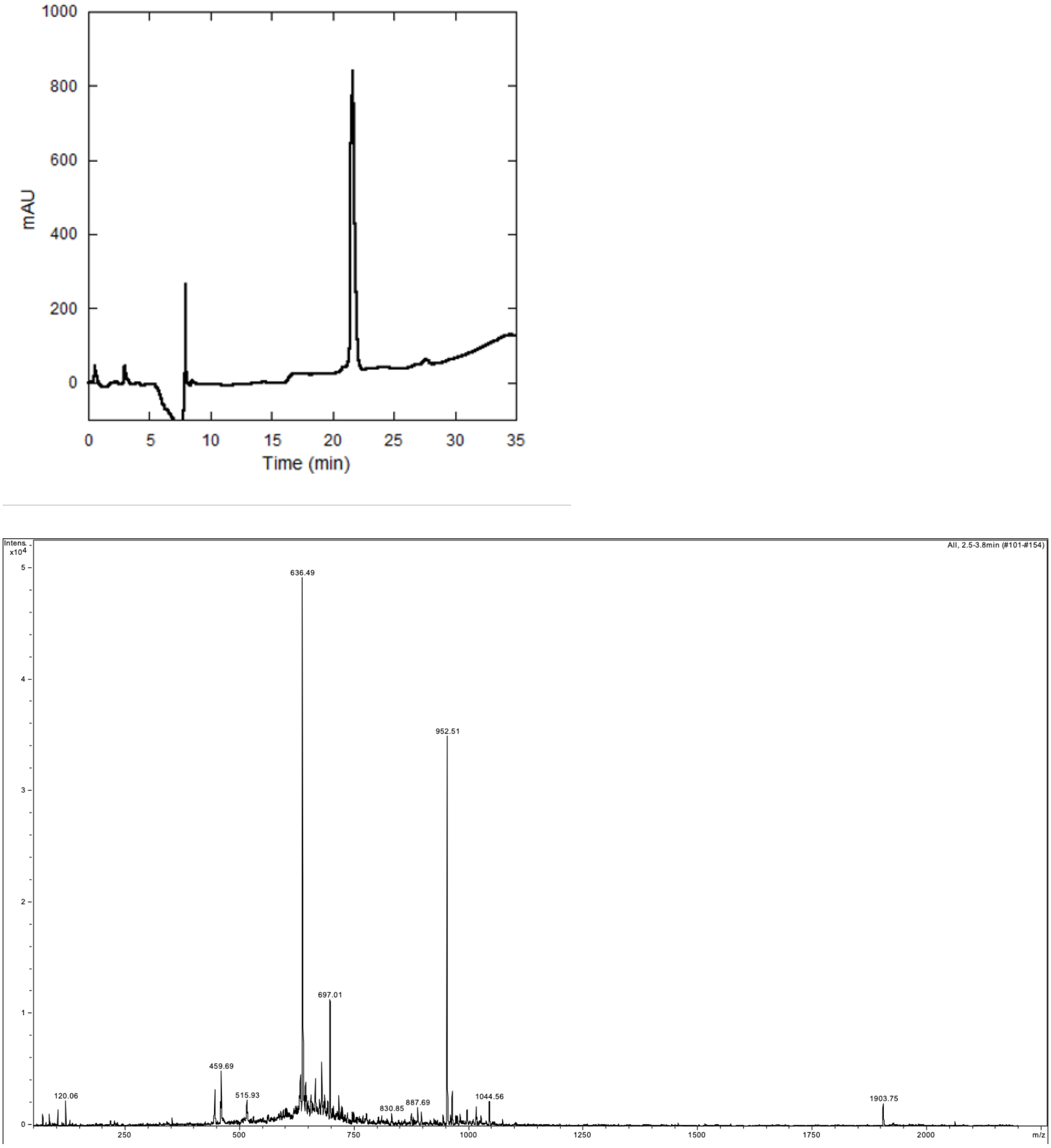
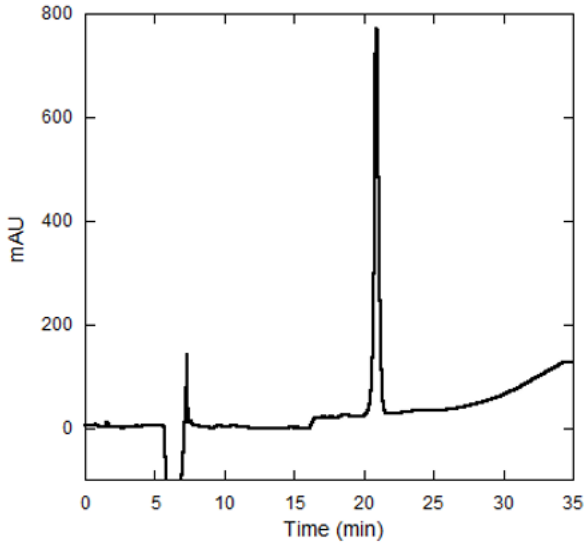


Figure 2.26. Chromatogram and mass spectrometry (ESI-ion trap) of CPfox5. The exact mass of CPfox5 is 1902.04. [MH^+]= 1903.75, [MH_2^{2+}]= 952.51, [MH_3^{3+}]= 636.49 were observed.



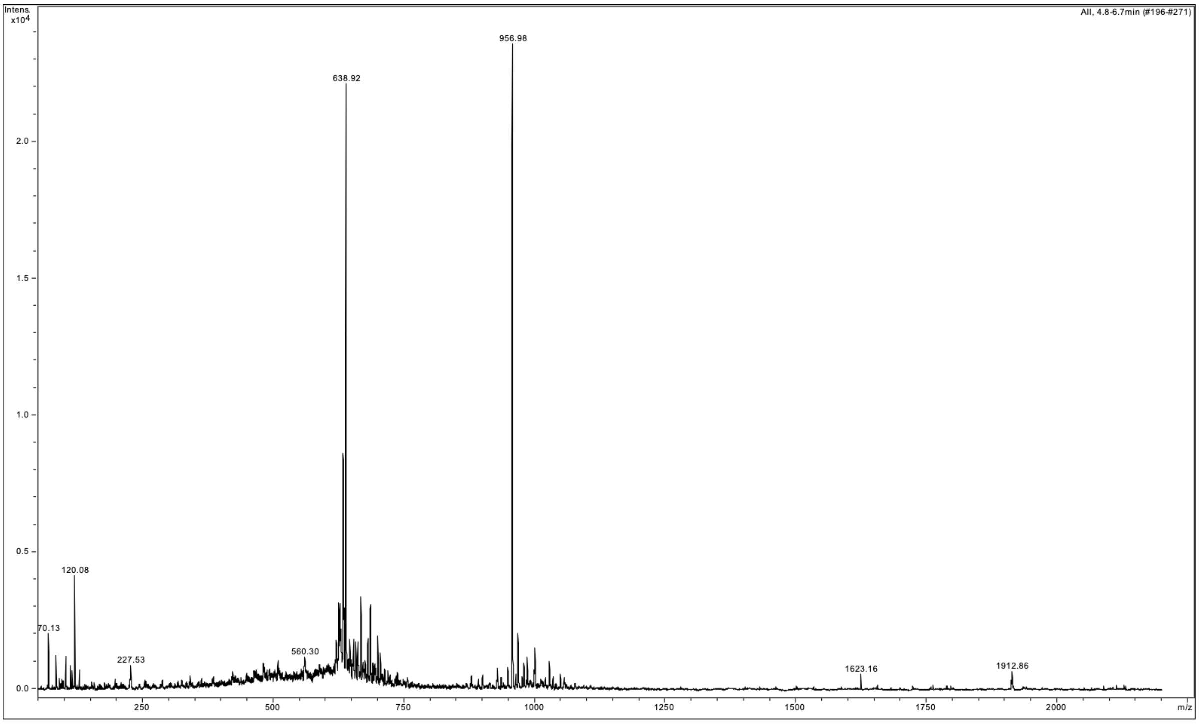
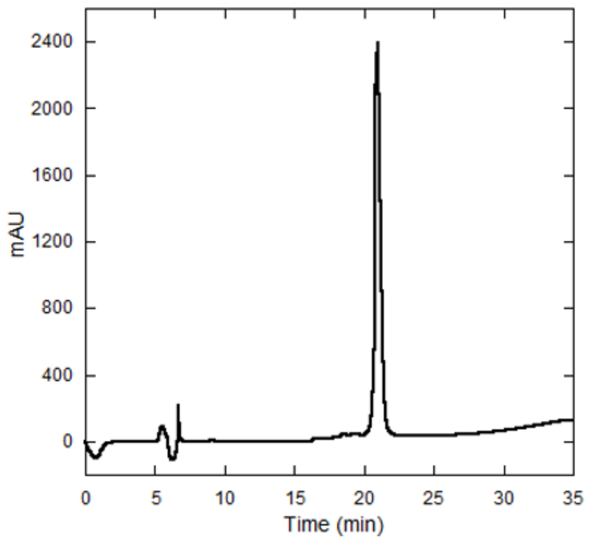


Figure 2.27. Chromatogram and mass spectrometry (ESI-ion trap) of CPfox6. The exact mass of CPfox6 is 1911.07. $[MH^{+}]$= 1912.86, $[MH_2^{2+}]$= 956.98, $[MH_3^{3+}]$= 638.92 were observed.



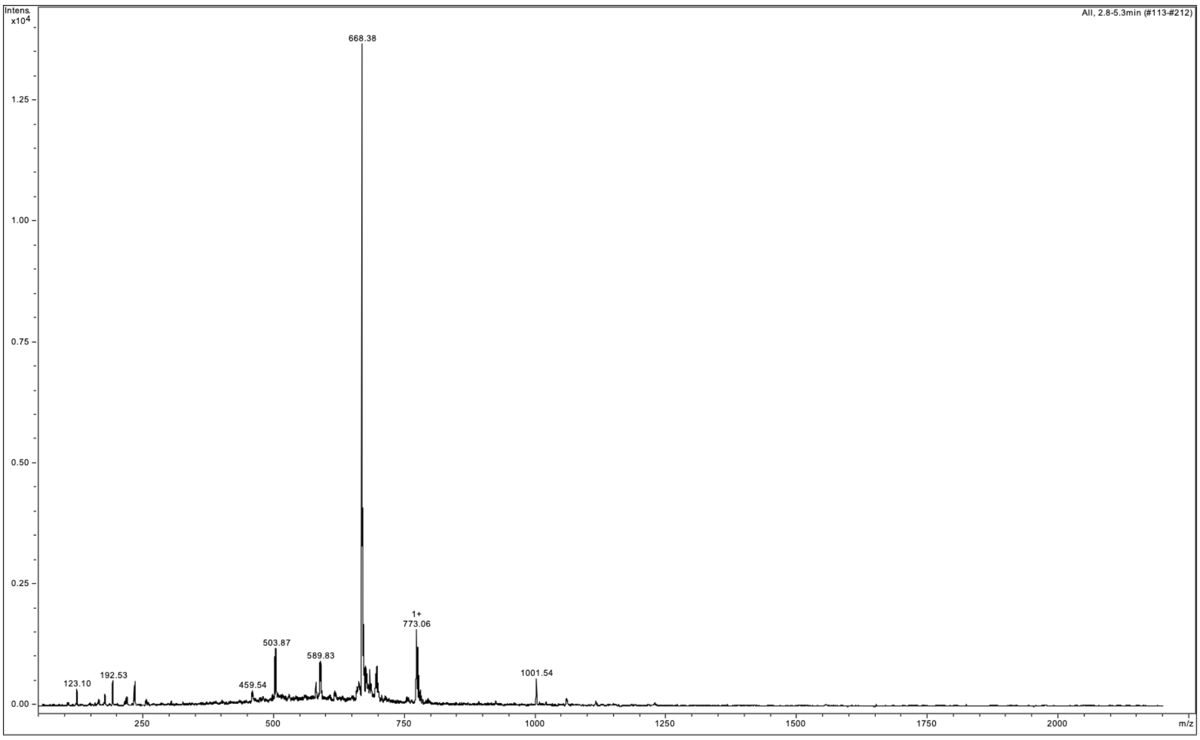


Figure 2.28. Chromatogram and mass spectrometry (ESI-ion trap) of CPfox7. The exact mass of CPfox7 is 2000.13. $[MH_2^{2+}]$= 1001.54, $[MH_3^{3+}]$= 668.83 were observed.

Table 2.1. Amino acid sequence of Rbfox2 RRM (residue 146-161) and all 7 CPfox mimetic peptides (a),(b),(c)

Peptide/Protein	Position																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Rbfox2(146-161)	V	E	I	I	F	**N**	**E**	**R**	**G**	S	K	G	F	G	F	V		
CPfox1			I	I	F	**N**	**E**	**R**	**G**	S	K	G	F	G	p	P		
CPfox2			I	I	F	**N**	**E**	**R**	**G**	S	K	R	F	R	p	P		
CPfox3	V	E	I	I	F	**N**	**E**	**R**	**G**	S	K	G	F	G	F	V	p	P
CPfox4		E	I	I	F	**N**	**E**	**R**	**G**	S	K	R	F	R	F	p	P	
CPfox5		K	I	I	F	**N**	**E**	**R**	**G**	S	R	G	F	R	F	p	P	
CPfox6		K	I	I	R	**N**	**E**	**R**	**G**	S	R	G	F	R	F	p	P	
CPfox7		K	I	I	F	**N**	**Q**	**R**	**G**	S	R	R	F	R	F	p	P	

(a) All peptides are cyclized following synthesis, as described in the methods section.
(b) The lower-case p represents ^{D}Pro.
(c) Residues corresponding to the β2β3 loop in the Rbfox2 RRM are underlined.

Table 2.2. Experimental restraints and structural statistics for the 10 lowest energy structures of CPfox mimetic peptides.

	CPfox2	CPfox4	CPfox5	CPfox6	CPfox7
No. of NOE-derived restraints	167	131	220	183	158
intraresidue	85	70	119	113	94
sequential	51	43	63	43	36
medium range ($2 \leq$ i-j ≤ 5)	24	7	27	18	19
long range (i-j ≤ 6)	7	11	11	9	9
No. of NOE violation					
number (> 0.2 Å)	0	0	0	0	0
maximum (Å)	0	0	0	0	0
mean RMSD value (Å)					
all heavy atoms	1.89 ±0.46	2.30 ±0.40	2.08 ±0.60	2.26 ±0.85	2.16 ±0.18
backbone atoms	1.00 ±0.34	1.39 ±0.21	1.06 ±0.38	1.29 ±0.49	1.29 ±0.22

Chapter 3

Structure of the Dengue Virus RNA Promoter

Abstract

Dengue virus, a single-stranded positive sense RNA virus, is the most prevalent mosquito-borne pathogen in the world. Like all RNA viruses, it uses conserved structural elements within its genome to control essential replicative steps. A 70 nucleotides stem-loop RNA structure (called SLA) found at the 5'-end of the genome of all flaviviruses, functions as the promoter for viral replication. This highly conserved structure interacts with the viral polymerase NS5 to initiate RNA synthesis. Here I report the NMR structure of a monomeric SLA from Dengue virus serotype 1, assembled to high-resolution from independently folded structural elements. The DENV1 SLA has an L-shape structure, where the top and side helices are coaxially-stacked and the bottom helix is roughly perpendicular to them. Because the sequence is highly conserved among different flavivirus genomes, it is very likely that the three-dimensional fold and local structure of SLA are also conserved among flaviviruses and required for efficient replication. This work provides structural insight into the Dengue promoter and provides the foundation for the discovery of new antiviral drugs that target this essential replicative step.

Section 3.1: Introduction

Flaviviruses such as dengue (DENV), West Nile (WNV), yellow fever (YFV) and Zika virus (ZIKV), cause severe human diseases. Among them, Dengue fever is the most prevalent mosquito-borne viral disease in humans (Clyde et al., 2006; Gubler, 2006). It is estimated that up to 400 million people become infected with it each year (Murray et al., 2013). Any of the four dengue serotypes (DENV1 to DENV4) can produce clinical symptoms ranging from a flu-like syndrome to the severe and even fatal Dengue hemorrhagic fever (Clyde et al., 2006; Gubler, 2006). Despite the significant impact of dengue infection on human health, effective vaccines are not yet available after > 70 years of efforts (Dyer, 2017), and small molecule treatment is only beginning to show promise (Lin et al., 2017; Raut et al., 2015; Saleem et al., 2019).

Dengue, like all flaviviruses, is a single-stranded positive sense RNA virus with a genome of about 11 kb, with a 5'-type I cap but without a polyadenylated tail (Colavita et al., 2020; Dang et al., 2020; Kamau et al., 2019; Pascalis et al., 2020; Selisko et al., 2014). The viral genome encodes a long polyprotein which is subsequently processed by both host and viral proteases to generate 10 mature viral proteins. Three structural proteins, Capsid (C), Envelop (E), and prM, assemble new viral particles, while seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) are responsible for viral RNA replication (Garcia-Blanco et al., 2016; Klema et al., 2015; Mazeaud et al., 2018; Neufeldt et al., 2018; Ng et al., 2017). The NS5 protein consists of an N-terminal methyltransferase (MTase) domain which is involved in capping, and a C-terminal RNA-dependent RNA polymerase (RdRp) domain which is responsible for RNA synthesis (Ferrer-Orta et al., 2006; Neufeldt et al., 2018; Yap et al., 2007; Zhao et al., 2015).

The coding sequence is flanked by 5'- and 3'-UTRs which contains highly conserved cis-acting RNA elements that are important for translation of viral proteins, RNA synthesis, encapsidation, and genome dimerization (Gebhard et al., 2011; Ng et al., 2017). In particular, the highly conserved stem-loop A (SLA) in the 5'-UTR, close to the 5'-end of the genome, functions as a 'promoter 'for the NS5 protein to initiate minus-strand synthesis (Choi, 2021; Filomatori et al., 2011; Gebhard et al., 2011). The interaction between SLA and NS5 is critical for RNA synthesis and was confirmed to be direct by biochemical studies; removal of SLA and mutation of certain nucleotides drastically decreases RNA replication (Filomatori et al., 2006; Lodeiro et al., 2009). Consistent with its essential functional role, the sequence and secondary structure of the SLA are highly conserved (Figure 3.1).

The current model for DENV RNA synthesis involves cyclization of the genome, mediated by long-range RNA-RNA interactions through inverted complementary sequences near the opposite ends of the RNA, which position the 5'- and 3'-UTRs in close proximity (Alvarez, Lodeiro, et al., 2005; Filomatori et al., 2011; Filomatori et al., 2006; Villordo et al., 2010). This interaction allows the SLA-bound NS5 to transfer to the 3'-stem loop (3'-SL) at the end of the 3'-UTR and initiate minus-strand synthesis (Hodge et al., 2016; Mazeaud et al., 2018). Removal of the SLA or 3'-SL leads to a significant reduction of viral replication (Alvarez, De Lella Ezcurra, et al., 2005; Filomatori et al., 2006; Yu et al., 2008), suggesting that a small molecule targeting these RNAs could have anti-viral activity.

In all flaviviruses, the SLA forms a Y-shaped secondary structure, consisting of a bottom helix with a conserved U-rich bulge, a short side stem-loop and an apical stem-loop containing an internal loop (Figure 3.1) (Dethoff et al., 2018; Gritsun & Gould, 2007). A crystal structure of DENV2 SLA was determined using a chimeric RNA, wherein the DENV2 SLA was inserted into the anticodon loop of a human tRNA (Lee et al., 2021). It revealed a large L-shape dimer, where the side stem-loop of each monomer is base paired with another molecule in the crystal. It is unclear whether dimerization is an artifact of crystallization, or a relevant functional state, perhaps transient, of the virus.

Here I report the NMR structure of the monomeric DENV1 SLA. Because of the severe peak overlap caused by the large size of the RNA, we use a divide-and-conquer approach to assemble a high-resolution structure from three independently folded structural elements corresponding to two of the three stem-loops, and the three-way junction (Figure 3.2), which were then assembled to generate the full SLA RNA through NMR analysis of the complete element (Barnwal et al., 2016). The SLA has an L-shape structure, consistent with the crystal structure, where the top and side stem-loops are coaxially-stacked and the bottom helix is roughly perpendicular to it. I only observe formation of a dimeric conformation at near mM NMR concentrations, above 0.5 mM; suggesting that if a dimer does indeed form in the cell, it is not because of its thermodynamic stability. My structure will facilitate understanding of the mechanism of dengue virus replication and provides the foundation for the discovery of new antiviral drugs, which is already under way in the group.

Section 3.2: Results

Section 3.2.1: Construct design

The spectra of the complete SLA RNA are characterized by extensive spectral overlap and broad lines because of its relatively large size, 70 nts, causing difficulties in obtaining unambiguous peak assignment and, especially, collecting a large number of constraints for structure determination (Figure 3.3). Thus, we adopted the strategy of reconstructing the complete structure by analyzing smaller independently folded structural elements (divide-and-conquer) (Barnwal et al., 2016). This can be done because RNA structure is modular, and its folding hierarchical; in other words, secondary structural elements generally fold independently outside of the context of the complete RNA (Tinoco & Bustamante, 1999). This is very different from proteins, where secondary and tertiary structure folding generally cannot be separated. Nevertheless, the validity of this approach has to be experimentally verified in each case, which I do as described below.

I divided the DENV1 SLA into three structural segments that overlap to generate the complete promoter: the bottom helix (DenvBS), the top stem-loop (DenvTSL) and the three-way junction (DenvSLAsh) (Figure 3.2 and 3.4).

A few additional G-C base pairs were added at the end of the sequences to improve *in vitro* transcription and stabilize the local secondary structure, if needed. The addition of tetraloops stabilizes RNA structure and, especially, improves linewidth with multiple RNAs we have studied in the past (Barnwal et al., 2016; Sharma & Varani, 2020; Walker et al., 2020). Thus, a UUCG tetraloop was added to the top of DenvBS, while UUCG and GAAA tetraloop were added to the top and side stemloop of DenvSLAsh, respectively. For the complete SLA, we also incorporated UUCG and GAAA tetraloop into the top and side stem-loop of DENV1 SLA, respectively, to generate a stabilized SLA which we named DenvSLATL (Figure 3.2-3.4), and to avoid dimer formation through the side stem-loop, as described in Section 3.2.4. The top terminal loop is important site for SLA function (Filomatori et al., 2011; Filomatori et al., 2006; Lodeiro et al., 2009); thus the top loop of DenvTSL was retained as wild type.

Section 3.2.2: The structure of the individual domains recapitulates what is observed in the full SLA

The divide-and-conquer approach is only warranted if the structure of individual domains faithfully recapitulates what is observed in the complete RNA. For each of the smaller structural segments, imino proton peaks of base-paired residues predicted from the secondary structure were observed in H_2O NOESYs, except for fast exchanging imino protons for unpaired nucleotides and the base-pairs at the end of helical stretches (Figure 3.5). Assignments of these imino protons allowed me to establish and verify the predicted secondary structures for each domain.

High quality NMR spectra could be collected for a tetraloop-stabilized SLA model that avoid dimerization through the side stem-loop (see below, DenvSLATL) and peak assignment was facilitated by the assignments of each segment. Reassuringly, I find that each NH in the smaller segments has similar chemical environment to its corresponding NH in the complete RNA, thus presenting similar chemical shifts and verifying that the structure observed in the complete SLA is retained in the individual fragments. Namely, overlaying the NOESY NMR spectra of the segments on the DenvSLATL spectra revealed high similarities in chemical shifts and NOE patterns (Figure 3.6), which confirms the secondary structure of the fragments coincide with what is seen, for the corresponding domain, in the complete RNA.

NOESY spectra in D_2O were then collected to assign non-exchangeable protons. Extensive overlap in the sugar proton region was relieved by deuteration of H3', H4', H5', H5 "and H5 protons (Figure 3.7-3.9), even if this approach prevented assignments of most of those protons. ^{15}N- and ^{13}C-labeled samples were prepared for DenvTSL to record ^{1}H-^{15}N HSQC and ^{1}H-^{13}C HSQC to distinguish ambiguous peaks in NOESY spectra, as well. Through deuteration, I was able to observe characteristic sequential NOEs in helices and assign non-exchangeable protons for each of the separate Denv structural segments (Figure 3.7-3.9).

The existence of a single dominant and monomeric conformation was confirmed in each case by the number of Ura and Cyt H5-H6 peaks in TOCSY spectra, which was in all cases consistent with the sequence (Figure 3.10). Formation of base-paired double helices was confirmed by cross-strand NOEs involving imino resonances, as well as sequential NOEs involving both exchangeable and non-exchangeable protons (Figure 3.7-3.9). Thus, DenvBS and DenvTSL were confirmed to form stem loop structures as expected; the monomeric/dimeric state of the three-way junction is discussed below.

Section 3.2.3: NMR Structures of the individual structural elements of DENV1 SLA RNA

Once assignments were completed, NOE distance constraints were systematically tabulated for structure determination, and the constrain list was refined by multiple rounds of structure calculations. For each of the three structural segments, structure calculations were performed independently using distance and torsion angle restraints derived from NMR experiments, as summarized in the Tables 3.1-3.3. Once the segments were completed, restraints from corresponding nucleotides in each segment were also added to the restraint table for structure calculation of the complete monomeric SLA (see Section 3.2.5).

RNA structures were calculated with Xplor-NIH using torsion angle dynamics and simulated annealing starting from randomized coordinates against the restraint table (Schwieters et al., 2003). A total of 1,000 independent calculations were executed and the lowest energy structure was further refined by an extended simulated annealing calculation against the experimental restraints. 400 structures were eventually generated, and the 10 lowest scoring structures were selected for presentation for each RNA molecule. The resulting structures converge well, with no NOE restraint violation greater than 0.5 Å. Representative structures for DenvSLA segments were presented with PyMol (Figure 3.11) and the overall heavy atom RMSDs are listed in Tables 3.1-3.3 as well.

As predicted, DenvBS and DenvTSL form double helical stems interrupted by internal loops. In both cases, no direct evidence of non-canonical base-pair formation in the internal loops was found, but sequential aromatic NOE correlations (H6/H8-H6/H8) in those regions were observed (Figure 3.7 and 3.8), together with other sugar-base NOEs, implying the retention of helical stacking throughout the internal loop. Many of the unpaired nucleotides are therefore stacked inside the helices, and potentially base paired, but we lack the information to impose such constraints, perhaps because any non-canonical base pairs are only transiently formed.

The results provide insight into the local structure in these regions. In the U-rich internal loop of DenvBS, Ura12 and Ura30 are coplanar and Ura29 stacks between Gua28 and Ura30 (Figure 3.12A). Nucleotides in the internal loop of DenvTSL (Cyt7 to Ade9 and Gua20 to Ade22) also stack within the helical stem, retaining coaxial stacking (Figure 3.12B). Although the terminal loop of DenvTSL varies somewhat within the 10 calculated structures, the topology is clearly established. The Gua15 base points toward the major groove, while Ade14 and Ura16 points outward, in the direction of the solvent, and could conceivably provide direct interactions with the RdRP or other accessory proteins (Figure

3.12C). The significance of each region in Dengue SLA had been validated by biological studies and is discussed further in Section 3.3.

Section 3.2.4: Structure of the 3-way junction

The key structural element in SLA is the 3-way junction which organizes the complete element; because a dimer is observed in the DENV2 crystal structure, it remains unclear what the three-dimensional organization of the full SLA would be. In order to investigate the three-way junction in DENV1 SLA with the necessary resolution, we prepared an RNA containing the sequence of the three-way junction and side stem-loop, but with shortened top and bottom helical stems to reduce spectra overlap. Furthermore, two RNAs were synthesized, one containing the wild-type side stem loop (SLAshCUUG) and the other containing a GAAA-stabilized side stem loop (DenvSLAsh) (Figure 3.2 and 3.4). This was done because I observed extra base pairs in the NMR spectra of SLAshCUUG. The signal intensity for those extra base pairs is concentration dependent, which indicates a monomer-dimer equilibrium at the 0.5-1.5 mM concentration of our NMR experiments (Figure 3.13A). By substituting CUUG with GAAA tetraloop, dimerization was eliminated (Figure 3.13B).

In the NOESY spectra of SLAshCUUG, I did not observe NH corresponding to base paired nucleotides for the side stem-loop or cross-strand peaks involving the AH2s. Nevertheless, sequential H1'-H6/H8 and H2'-H6/H8 were observed for the helical sections of the side stem-loop, which indicate stacking and perhaps unstable base pairing. I presume that with a less stabilizing loop sequence, the side stem-loop in monomeric SLAshCUUG forms an unstable short helix, allowing the dimer to form through RNA-RNA interactions in the side stem-loop at the mM concentrations of NMR experiments (Figure 3.13C). Considering the low copy number of viral RNAs in cell, it is very unlikely that the dimer is the thermodynamically favored SLA structure and therefore I collected NMR spectra of the monomeric DenvSLAsh for structural analysis. However, I cannot exclude the possibility that dimer formation occurs at intermediate steps of viral replication; this is further discussed below.

Base-pair imino NOEs and characteristic sequential NOE patterns for helical stem nucleotides were observed in NOESY spectra for the three double helical stems (Figure 3.5C and 3.9), validating the secondary structure prediction. For nucleotides Gua7, Gua8 and Ade9, only sequential NOEs were observed, but no other inter-stranded or inter-nucleotide NOEs were identified. NOE interactions corresponding to Gua23H8-Ade24H8, Ura33H6-Ade34H8 and Ade34-Cyt10H6 were observed in NOESY spectra as well (Figure 3.14A), consistent with continuous stacking between the top and side stem-loops, and suggestive of coaxial stacking of the two helices.

The calculated structure of DenvSLAsh demonstrates formation of an L-shape three-way junction, where the top and side stem loops in stack coaxially, and the bottom helix is roughly perpendicular to the coaxial stack (Figure 3.11). Ade34 in the three-way junction stacks within the coaxial helix, between Cyt10 in the top stem and Ura33 in the side stem. Gua7, Gua8 and Ade9 are instead single stranded, giving the three-way junction some flexibility, and thus allowing the angle of the bottom helix to wiggle relative to the top coaxial stack (Figure 3.14B).

The topology of the three-way junction structure was independently validated by SAXS analysis (Figure 3.14C and 3.15). Both SLAshCUUG and DenvSLAsh exhibit L-shape SAXS envelops which agree very well with the L-shape NMR structure and provide independent validation. They also demonstrate that the structure is fully monomeric below 0.1 mM concentration.

Section 3.2.5: Structure of the Denv1 promoter

By examining each of the independently folded secondary structure elements separately, I was able to establish the local structures for the two stem-loops and 3-way junction and side stem-loop that form Denv1 SLA. Spectral and structural information collected from these spectra were used to assist peak assignments and structure determination for the complete SLA structure.

A single conformation of DenvSLATL was confirmed by the number of Ura and Cyt H5-H6 peaks in TOCSY spectra (Figure 3.16). Overlay of the NOESY and TOCSY spectra of DenvSLATL with those of its structural segments revealed very similar chemical shifts and NOE patterns (Figure 3.6 and 3.16) which allowed me to safely transfer the much larger number and more confidently assigned restraints obtained from the individual segments. Structure calculations were then performed for DenvSLATL using distance and torsion angle restraints derived from experimental NMR data of DenvSLATL and each of the structural segments, as was done for each of the elements in the structure. The 10 lowest structures converge with no NOE restraint violation greater than 0.5 Å, and then independent SAXS analysis were performed to validate the structure. Importantly, the topology of the structure of DenvSLATL was independently validated by SAXS analysis, and the SAXS envelope agree very well with the L-shape NMR structure (Figure 3.17). The structural statistics are summarized in Table 3.4.

The calculated structure of DenvSLATL reveals an L-shape RNA, where the top and side stem loop are coaxial and the bottom stem roughly perpendicular to it (Figure 3.17A) , as per the smaller three-way junction. The arrangement of nucleotides in the three-way junction is as observed in DenvSLAsh: Ade56 stacks between the base-pairs of Cyt22-Gua45 and Ade46-Ura66, which are at the end of the top and side stem-loop, respectively, while nucleotides in the longer internal junction loop (Gua19 to Ade21) are single stranded, with fewer NOE interactions, and thus the orientation of the bottom stem is less well defined. The U-rich bulge in the bottom stem is important for SLA function. In my NMR structures, Ura12 and Ura64 are co-planar and stack within the helix. However, two conformations of Ura63 were observed in the calculated structures: one with the base within the helix and another with the base pointing outward (Figure 3.17B). Eight out of the lowest 10 structures have Ura63 stacked within the helix and two structures have Ura63 pointing outward, suggesting conformational flexibility.

Section 3.3: Discussion

The Denv SLA structure functions as an RNA 'promoter 'for the viral RNA-dependent RNA polymerase enzyme NS5 and is essential viral replication. NS5 is recruited to the viral RNA through this element, which has high conservation in both sequence and secondary structure across flaviviruses (Choi, 2021; Filomatori et al., 2006; Gebhard et al., 2011). The presence of a U-rich bulge in the bottom helix is essential for SLA function, and the high conservation of the 3-way junction supports a functional role for the overall 3D shape of the RNA, which is determined by the topology of the junction. I observe a rigid conformation for the three-way junction, with co-axial stacking of the top and side stem-loops, to generate a well-defined L-shape structure, with the bottom helix emanating at a nearly 90 angle, but with some flexibility because of the single stranded nucleotides linking the bottom and top helices (Figure 3.17A).

Flexibility could be the result of insufficient experimental information (e.g. due to spectral overlap) or intrinsic dynamic. One way to look at dynamic without recording relaxation data, which would be very challenging for an RNA this size, is to examine the conformation of the sugar pucker, as deduced from the TOCSY spectra. For 3'-endo sugar pucker, the H1'-H2 'coupling is very small and generally no peak is observable, especially for an RNA this size. The 2'-endo conformation has strong H1'-H2', medium H1'-H3 'peaks and very weak/unobservable H1'-H4 'peaks. For a mixture of 3 'and 2'-endo conformations, which is a sign of dynamics, peaks from H1 'to H2', H3 'and H4 'are observed. According to the TOCSY spectra (Figure 3.19), Ura29 in the U-rich bulge occupies the 2'-endo conformation; Ade9 and Gua8 in the three-way junction are 2'-endo as well. This is interesting, since the 2'-endo conformation extends the phosphate-phosphate distance. Unfortunately, I cannot establish the conformation of Ade34, in the loop between the side and bottom stems, because of spectral overlap around 4.4 ppm; this is an important nucleotide to determine the orientation of the lower helix with respect to the rest of the structure. The Cytosine in the top loop is 2'-endo. Only the sugar of the Adenine that follows occupies a mixture of 2 'and 3'-endo conformations, which implies a dynamic structure for this nucleotide. It is possible that conformational flexibility plays a role in SLA function, since RNA-binding proteins often exploit induced fit (Leulliot & Varani, 2001), but this remains to be investigated.

In the structure, Ura63 in the U-rich bulge occupies 2 conformations; the base can point inward or outward, but the two conformations do not affect the rest of the RNA structure (Figure 3.17B); in the recently reported crystal structure of a dimeric Denv2 SLA, the same Uracil points outward (Lee et al., 2021). The fact that DENV3 has a single U bulge in the bottom stem and mutational studies of the U-UU bulge imply that Ura63 is critical for SLA promoter function, but the two remaining nucleotides are not. Deletion or mutations of the U-rich bulge largely impair viral RNA replication *in vivo*, but the effect on the binding affinity to RdRp and its *in vitro* activity is insignificant (Filomatori et al., 2011; Lodeiro et al., 2009). This suggests that the U-rich bulge interacts with another protein which is important for viral replication in infected cells.

The apical loop is important for SLA function and highly conserved; the CAG(X)U sequence is found in all four Dengue serotypes (Figure 3.1), and mutations impair both viral replication *in vivo* and RdRp activity *in vitro* (Filomatori et al., 2011; Filomatori et al., 2006; Lodeiro et al., 2009). Interestingly, mutations in the terminal loop do not significantly affect RdRp binding, suggesting the terminal loop might play an important role in post-binding steps to promote polymerase activity (Filomatori et al., 2011). In the structure, I observe that Gua32 points towards the major groove and faces the other two loop nucleotides Ade31 and Ura33, which point outward (Figure 3.12C). The orientation of the loop nucleotides could be important for RdRp activity. The top stem, however, is less conserved among dengue serotypes. I*n vivo* biological assays have shown that the presence of an internal loop in the top stem has no significant effect on SLA function either (Lodeiro et al., 2009), yet shortening the helix significantly decreases viral replication and RdRp binding (Filomatori et al., 2011; Lodeiro et al., 2009). This implies that, although the helical stem is unlikely to provide direct interactions with the NS5 protein, it might be important to space and orient the terminal loop relative to the three-way junction and the U-rich bulge.

The three-way junction is likely to provide a structural framework to orient the SLA and different domains of NS5, giving it an L-shaped conformation confirmed by the SAXS analysis. The top and side stem-loop are coaxially stacked, as expected, since no single-stranded nucleotide is found at the junction between them (Lescoute & Westhof, 2006). The stacking of Ade56 in the linker between side and bottom stem constrains the direction of the bottom stem to be roughly perpendicular to the coaxially stacked helix.

The recently reported crystal structure of Denv2 SLA also presents a L-shape structure, but this is generated by dimerization in the crystal created by the unpaired side loop which forms 'kissing loop 'interactions (Lee et al., 2021). In the crystal structure, the side loop was open and engaged in loop-loop interactions. We observe two conformations for a construct containing the wild-type loop sequence, consistent with monomer-dimer equilibrium. Replacement of the wild-type loop with GAAA stabilizes the monomer and allowed me to establish the structure of a monomeric SLA RNA. It remains to be seen whether SLA dimerization is of functional relevance. Because I observe increased dimer peak intensity at the NMR concentrations, 0.5-1.5 mM, and no dimer below 0.5 mM, I presume that under cellular conditions, the SLA RNA will be entirely monomeric. However, this conclusion does not preclude the possibility that the dimeric structure is transiently present as a result of protein binding or dimerization of the genome, and in fact it could be an elegant way for the virus to regulate its promoter activity.

I expect the 3D structure of the three-way junction to be conserved in flaviviruses (Figure 3.1 and 3.20), because no nucleotides are predicted to be present in the junction between top and side stem in all flavivirus sequences, and the linker between side and bottom stem is always short, while the longest linker occurs invariably between bottom and top helices. It follows that the shape of the SLA and the orientation of the top and bottom stems are likely important for SLA to be recognized by the viral polymerase.

In summary, I have established the 3D structure of the thermodynamically favored, monomeric form of the SLA promoter from Dengue virus serotype 1; which is very likely to be representative of all dengue serotypes, and indeed it is very likely all flaviviruses will share the same global structural arrangement. In addition to providing a framework for interpreting biochemical data and NS5 activity, this structure also lays the groundwork to identify small molecule inhibitors that target the highly conserved 3-way junction, which are being actively pursued in the group.

Section 3.4: Materials and Methods

Section 3.4.1: RNA Preparation

All RNA molecules were synthesized by in vitro transcription with *in house* purified T7 RNA polymerase using synthetic DNA oligonucleotide templates (purchased from Integrated DNA Technologies) and standard methods (Milligan et al., 1987). Partially deuterated RNA molecules (deuteration of H5, H3', H4', H5 'and H5 "protons) were synthesized using selective deuterated NTPs (from Cambridge Isotopes). ^{15}N- and ^{13}C-labeled samples were synthesized using isotope-labeled NTPs (from Cambridge Isotopes). The sequence and secondary structures of all the RNAs, as verified by NMR, are shown in Figure 3.4.

The RNA oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis (PAGE), electroeluted and concentrated by ethanol precipitation (Gubser & Varani, 1996). After extensive dialysis into 10 mM potassium phosphate buffer (pH 6.5), the RNAs were annealed by heating briefly to 90 °C followed by snap cooling in an ice bath. Final RNA concentrations used for NMR studies were 0.6~1.2 mM. For experiments studying non-exchangeable protons, samples were lyophilized to dryness and dissolved into D_2O. Samples used to study exchangeable protons were dissolved in H_2O:D_2O (9:1).

Section 3.4.2: NMR Spectroscopy

All NMR spectra were collected at 25℃ on Avance III 600 MHz, AVANCE III 700MHz or AVANCE III 800MHz spectrometer equipped with cryogenic probes (600 and 800). The 1D ^1H spectra were recorded using the excitation sculpting pulse sequence. 2D total correlation spectroscopy (TOCSY) spectra were recorded with mixing times of 80 ms. The exchangeable and non-exchangeable 2D NOESY spectra were recorded with various mixing times (100, 200 and 300 ms) to assist spectral assignments and quantitative evaluation of internuclear distances by comparison with peak intensities for pair of protons with fixed distances. Spectra for selectively deuterated samples were collected in the same manner. 2D ^{1}H-^{15}N and ^{1}H-^{13}C HSQC spectra were recorded on isotope labeled samples, if needed to confirm assignments. All NMR data were processed with TOPSPIN (Bruker) and analyzed in NMRFAM-SPARKY (Lee et al., 2015). Assignments of RNA spectra were guided by predicted RNA chemical shift values and based on well-established double-helical sequential NOE patterns (Varani et al., 1996; Varani & Tinoco, 1991).

Section 3.4.3: Experimental Restraints and Structure Determination

Interproton distance restraints were derived from NOE cross-peaks in 2D 1H NOESY spectra and sorted into strong (2.5 ±0.7 Å), medium (3.5 ±1.2 Å) and weak (4.5 ±1.5 Å) bins based on peak intensities, relative to fixed distances (e.g. H5-H6 = 2.5 Å, H3'-H6/H8 = 3.5 Å). Base-pair planarity and hydrogen-bonding restraints were used for unambiguously established base pairs as identified by 2D NOESY involving NH protons. Hydrogen bond, planarity and dihedral restraints were included for base-paired nucleotides that were surrounded by base pairs confirming to A-form helical structures, as established from the pattern of NOE cross-peaks (Varani et al., 1996; Varani & Tinoco, 1991).

It has been often found within the group that flexible loops capping stem-loops lead to loss of spectral quality, most likely due to non-specific aggregation or multimerization at NMR concentrations (Barnwal et al., 2016; Varani et al., 1991); this was the case for this RNA as well, with the added complication of dimerization through kissing loop interactions which created multiple conformations (Figure 3.3). Thus, UUCG and GAAA tetraloops were used to replace the dynamic apical loop or added to the end of the bottom helix to improve spectra quality (Banas et al., 2010; Jucker & Pardi, 1995; Varani et al., 1991).

Experimental constraints for structure calculation of the complete DenvSLATL were compiled by dividing the RNA into three segments: DenvBS, DenvTSL and DenvSLAsh (Figure 3.2), corresponding to the bottom helix, apical stem-loop and three-way junction and side stem-loop. Overlay of 2D ^{1}H-^{1}H NOESY spectra from the segments, with a spectrum of the complete Dengue 1 SLA model (called DenvSLATL), showed strong similarities in the chemical shifts, which allowed the transfer of segments assignments to DenvSLATL for assignments for structure calculations (Figure 3.5 and 3.6). Restraints derived from DenvBS spanned nucleotides G1-U18 and A57-A70; those obtained from DenvTSL spanned nucleotides C22-G45; DenvSLAsh contributed information around the three-way junction, namely A15-A25 and U42-U60; obviously, there is overlap which allowed us to further verify that the divide-and-conquer approach was warranted. The NMR experimental constraints are summarized in Tables 3.1-3.4.

RNA structures were calculated with Xplor-NIH using torsion angle dynamics and simulated annealing from a single extended RNA starting template (Schwieters et al., 2003). Compiled experimental restraints were used in a simulated annealing procedure, initially undergoing high-temperature (2,500K to 298K) torsional angle dynamics, where incremental decreases in temperature were generated by progressively introducing Van der Waals terms and increasing force constants for angles, dihedral angles, NOEs, and the Van der Waals repulsive term. After the final cooling step, the RNA underwent two sequential refinement steps, first in torsional angle space then in Cartesian space. A total of 1,000 independent calculations were executed for each RNA, and the lowest total energy structure calculated by the Xplor energy function was further refined by an extended simulated annealing calculation against the constraints listed above to generate 400 structures. The 10 structures with the lowest total energy, based on the Xplor energy function, were chosen as the representative structures to calculate RMSD. The structure quality analysis was conducted using MolProbity

(Williams et al., 2018). Representative structures of DenvBS, DenvTSL, DenvSLAsh and DenvSLATL were deposited in Protein Data Bank with accession codes: 7K4L, 7UME, 7UMD and 7UMC, respectively.

Section 3.4.4: Small-angle X-ray Scattering (SAXS)

RNA samples for SAXS were prepared similarly to samples made for NMR spectroscopy, but at different concentrations (1-5 mg/mL), dissolved in 20 mM Tris, 100 mM NaCl and 0.1 mM EDTA (pH 6.5). SAXS experiments were recorded on an *in-house* state-of-art SAXS instrument (BioSAXS-2000) at Argonne National Laboratory. The data were processed using RAW and particle distance distribution function P(r) plots were calculated using GNOM (Svergun, 1992) and used for low resolution *ab initio* shape reconstruction with DAMMIN (Svergun, 1999). Processed SAXS data curves of DenvSLAsh, SLAshCUUG and DenvSLATL are shown in Figures 3.21-3.23. A total of 20 models were generated with DAMMIN using the ATSAS online server (https://www.embl-hamburg.de/biosaxs/) (Manalastas-Cantos et al., 2021). The representative model was selected with a suite of software tools (DAMSEL, SAMSUP, DAMAVER and DAMFILT) for comparison and fitting to the NMR structure (Volkov & Svergun, 2003).

DAMSEL compares models, finds the most probable model and identifies outliers; DAMSUP aligns all models with the most probable model; DAMAVER averages these aligned models and computes a probability map; and DAMFILT filters the average model at a given default cut-off volume, which is the expected volume of the generated PDB file. DAMFILT removes loosely defined and lower occupancy atoms and generates a most probable compact model.

Section 3.5: References

Alvarez D. E., De Lella Ezcurra A. L., Fucito S., & Gamarnik A. V. (2005). Role of RNA structures present at the 3'UTR of dengue virus on translation, RNA synthesis, and viral replication. *Virology*, *339*(2), 200-212. https://doi.org/10.1016/j.virol.2005.06.009

Alvarez D. E., Lodeiro M. F., Luduena S. J., Pietrasanta L. I., & Gamarnik A. V. (2005). Long-range RNA-RNA interactions circularize the dengue virus genome. *J. Virol.*, *79*(11), 6631-6643. https://doi.org/10.1128/JVI.79.11.6631-6643.2005

Banas P., Hollas D., Zgarbova M., Jurecka P., Orozco M., Cheatham T. E., 3rd, Sponer J., & Otyepka M. (2010). Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.*, *6*(12), 3836-3849. https://doi.org/10.1021/ct100481h

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. https://doi.org/10.1093/nar/gkw584

Choi K. H. (2021). The Role of the Stem-Loop A RNA Promoter in Flavivirus Replication. *Viruses*, *13*(6). https://doi.org/10.3390/v13061107

Clyde K., Kyle J. L., & Harris E. (2006). Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. *J. Virol.*, *80*(23), 11418-11431. https://doi.org/10.1128/JVI.01257-06

Colavita F., Vairo F., Carletti F., Boccardo C., Ferraro F., Iaiani G., Al Moghazi S., Galardo G., Lalle E., Selvaggi C., et al. (2020). Full-length genome sequence of a dengue serotype 1 virus isolate from a traveler returning from Democratic Republic of Congo to Italy, July 2019. *Int. J. Infect. Dis.*, *92*, 46-48. https://doi.org/10.1016/j.ijid.2019.12.023

Dang T. T., Pham M. H., Bui H. V., & Van Le D. (2020). Whole genome sequencing and genetic variations in several dengue virus type 1 strains from unusual dengue epidemic of 2017 in Vietnam. *Virol. J.*, *17*(1), 7. https://doi.org/10.1186/s12985-020-1280-z

Dethoff E. A., Boerneke M. A., Gokhale N. S., Muhire B. M., Martin D. P., Sacco M. T., McFadden M. J., Weinstein J. B., Messer W. B., Horner S. M., et al. (2018). Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci. U. S. A.*, *115*(45), 11513-11518. https://doi.org/10.1073/pnas.1716689115

Dyer O. (2017). Philippines halts dengue immunisation campaign owing to safety risk. *BMJ*, *359*, j5759. https://doi.org/10.1136/bmj.j5759

Ferrer-Orta C., Arias A., Escarmis C., & Verdaguer N. (2006). A comparison of viral RNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.*, *16*(1), 27-34. https://doi.org/10.1016/j.sbi.2005.12.002

Filomatori C. V., Iglesias N. G., Villordo S. M., Alvarez D. E., & Gamarnik A. V. (2011). RNA sequences and structures required for the recruitment and activity of the dengue virus polymerase. *J. Biol. Chem.*, *286*(9), 6929-6939. https://doi.org/10.1074/jbc.M110.162289

Filomatori C. V., Lodeiro M. F., Alvarez D. E., Samsa M. M., Pietrasanta L., & Gamarnik A. V. (2006). A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev.*, *20*(16), 2238-2249. https://doi.org/10.1101/gad.1444206

Garcia-Blanco M. A., Vasudevan S. G., Bradrick S. S., & Nicchitta C. (2016). Flavivirus RNA transactions from viral entry to genome replication. *Antiviral Res.*, *134*, 244-249. https://doi.org/10.1016/j.antiviral.2016.09.010

Gebhard L. G., Filomatori C. V., & Gamarnik A. V. (2011). Functional RNA elements in the dengue virus genome. *Viruses*, *3*(9), 1739-1756. https://doi.org/10.3390/v3091739

Gritsun T. S., & Gould E. A. (2007). Origin and evolution of flavivirus 5'UTRs and panhandles: trans-terminal duplications? *Virology*, *366*(1), 8-15. https://doi.org/10.1016/j.virol.2007.04.011

Gubler D. J. (2006). Dengue/dengue haemorrhagic fever: history and current status. *Novartis Foundation Symposia*, *277*, 3-16; discussion 16-22, 71-13, 251-253. https://doi.org/10.1002/0470058005.ch2

Gubser C. C., & Varani G. (1996). Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3'-untranslated region and interaction with the U1A protein. *Biochemistry*, *35*(7), 2253-2267. https://doi.org/10.1021/bi952319f

Hodge K., Tunghirun C., Kamkaew M., Limjindaporn T., Yenchitsomanus P. T., & Chimnaronk S. (2016). Identification of a Conserved RNA-dependent RNA Polymerase (RdRp)-RNA Interface Required for Flaviviral Replication. *J. Biol. Chem.*, *291*(33), 17437-17449. https://doi.org/10.1074/jbc.M116.724013

Jucker F. M., & Pardi A. (1995). GNRA tetraloops make a U-turn. *RNA*, *1*(2), 219-222. https://www.ncbi.nlm.nih.gov/pubmed/7585251

Kamau E., Agoti C. N., Ngoi J. M., de Laurent Z. R., Gitonga J., Cotten M., Phan M. V. T., Nokes D. J., Delwart E., Sanders E., et al. (2019). Complete Genome Sequences of Dengue Virus Type 2 Strains from Kilifi, Kenya. *Microbiol. Resour. Announc.*, *8*(4). https://doi.org/10.1128/MRA.01566-18

Klema V. J., Padmanabhan R., & Choi K. H. (2015). Flaviviral Replication Complex: Coordination between RNA Synthesis and 5'-RNA Capping. *Viruses*, *7*(8), 4640-4656. https://doi.org/10.3390/v7082837

Lee E., Bujalowski P. J., Teramoto T., Gottipati K., Scott S. D., Padmanabhan R., & Choi K. H. (2021). Structures of flavivirus RNA promoters suggest two binding modes with NS5 polymerase. *Nat. Commun.*, *12*(1), 2530. https://doi.org/10.1038/s41467-021-22846-1

Lee W., Tonelli M., & Markley J. L. (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, *31*(8), 1325-1327. https://doi.org/10.1093/bioinformatics/btu830

Lescoute A., & Westhof E. (2006). Topology of three-way junctions in folded RNAs. *RNA*, *12*(1), 83-93. https://doi.org/10.1261/rna.2208106

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. https://doi.org/10.1021/bi010680y

Lin K. H., Ali A., Rusere L., Soumana D. I., Kurt Yilmaz N., & Schiffer C. A. (2017). Dengue Virus NS2B/NS3 Protease Inhibitors Exploiting the Prime Side. *J. Virol.*, *91*(10). https://doi.org/10.1128/JVI.00045-17

Lodeiro M. F., Filomatori C. V., & Gamarnik A. V. (2009). Structural and functional studies of the promoter element for dengue virus RNA replication. *J. Virol.*, *83*(2), 993-1008. https://doi.org/10.1128/JVI.01647-08

Manalastas-Cantos K., Konarev P. V., Hajizadeh N. R., Kikhney A. G., Petoukhov M. V., Molodenskiy D. S., Panjkovich A., Mertens H. D. T., Gruzinov A., Borges C., et al. (2021). ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.*, *54*(Pt 1), 343-355. https://doi.org/10.1107/S1600576720013412

Mazeaud C., Freppel W., & Chatel-Chaix L. (2018). The Multiples Fates of the Flavivirus RNA Genome During Pathogenesis. *Front. Genet.*, *9*, 595. https://doi.org/10.3389/fgene.2018.00595

Milligan J. F., Groebe D. R., Witherell G. W., & Uhlenbeck O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, *15*(21), 8783-8798. https://www.ncbi.nlm.nih.gov/pubmed/3684574

Murray N. E., Quam M. B., & Wilder-Smith A. (2013). Epidemiology of dengue: past, present and future prospects. *Clin. Epidemiol.*, *5*, 299-309. https://doi.org/10.2147/CLEP.S34440

Neufeldt C. J., Cortese M., Acosta E. G., & Bartenschlager R. (2018). Rewiring cellular networks by members of the Flaviviridae family. *Nat. Rev. Microbiol.*, *16*(3), 125-142. https://doi.org/10.1038/nrmicro.2017.170

Ng W. C., Soto-Acosta R., Bradrick S. S., Garcia-Blanco M. A., & Ooi E. E. (2017). The 5' and 3' Untranslated Regions of the Flaviviral Genome. *Viruses*, *9*(6). https://doi.org/10.3390/v9060137

Pascalis H., Biscornet L., Toty C., Hafsia S., Roche M., Despres P., Atyame Nten C., Bibi J., Louange M., Gedeon J., et al. (2020). Complete Genome Sequences of Dengue Virus Type 2 Epidemic Strains from Reunion Island and the Seychelles. *Microbiol. Resour. Announc.*, *9*(4). https://doi.org/10.1128/MRA.01443-19

Raut R., Beesetti H., Tyagi P., Khanna I., Jain S. K., Jeankumar V. U., Yogeeswari P., Sriram D., & Swaminathan S. (2015). A small molecule inhibitor of dengue virus type 2 protease inhibits the replication of all four dengue virus serotypes in cell culture. *Virol. J.*, *12*, 16. https://doi.org/10.1186/s12985-015-0248-x

Saleem H. N., Batool F., Mansoor H. J., Shahzad-ul-Hussan S., & Saeed M. (2019). Inhibition of Dengue Virus Protease by Eugeniin, Isobiflorin, and Biflorin Isolated from the Flower Buds of Syzygium aromaticum (Cloves). *Acs Omega*, *4*(1), 1525-1533. https://doi.org/10.1021/acsomega.8b02861

Schwieters C. D., Kuszewski J. J., Tjandra N., & Clore G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, *160*(1), 65-73. https://www.ncbi.nlm.nih.gov/pubmed/12565051

Selisko B., Wang C., Harris E., & Canard B. (2014). Regulation of Flavivirus RNA synthesis and replication. *Curr. Opin. Virol.*, *9*, 74-83. https://doi.org/10.1016/j.coviro.2014.09.011

Sharma S., & Varani G. (2020). NMR structure of Dengue West Nile viruses stem-loop B: A key cis-acting element for flavivirus replication. *Biochem. Biophys. Res. Commun.*, *531*(4), 522-527. https://doi.org/10.1016/j.bbrc.2020.07.115

Svergun D. I. (1992). Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J. Appl. Crystallogr.*, *25*, 495-503. https://doi.org/Doi 10.1107/S0021889892001663

Svergun D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.*, *76*(6), 2879-2886. https://doi.org/Doi 10.1016/S0006-3495(99)77443-6

Tinoco I., Jr., & Bustamante C. (1999). How RNA folds. *J. Mol. Biol.*, *293*(2), 271-281. https://doi.org/10.1006/jmbi.1999.3001

Varani G., Aboulela F., & Allain F. H. T. (1996). NMR investigation of RNA structure. *Prog. Nucl. Mag. Res. Sp.*, *29*, 51-127. https://doi.org/Doi 10.1016/0079-6565(96)01028-X

Varani G., Cheong C., & Tinoco I., Jr. (1991). Structure of an unusually stable RNA hairpin. *Biochemistry*, *30*(13), 3280-3289. https://doi.org/10.1021/bi00227a016

Varani G., & Tinoco I., Jr. (1991). RNA structure and NMR spectroscopy. *Q. Rev. Biophys.*, *24*(4), 479-532. https://doi.org/10.1017/s0033583500003875

Villordo S. M., Alvarez D. E., & Gamarnik A. V. (2010). A balance between circular and linear forms of the dengue virus genome is crucial for viral replication. *RNA*, *16*(12), 2325-2335. https://doi.org/10.1261/rna.2120410

Volkov V. V., & Svergun D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, *36*, 860-864. https://doi.org/10.1107/S0021889803000268

Walker M. J., Shortridge M. D., Albin D. D., Cominsky L. Y., & Varani G. (2020). Structure of the RNA Specialized Translation Initiation Element that Recruits eIF3 to the 5'-UTR of c-Jun. *J. Mol. Biol.*, *432*(7), 1841-1855. https://doi.org/10.1016/j.jmb.2020.01.001

Williams C. J., Headd J. J., Moriarty N. W., Prisant M. G., Videau L. L., Deis L. N., Verma V., Keedy D. A., Hintze B. J., Chen V. B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.*, *27*(1), 293-315. https://doi.org/10.1002/pro.3330

Yap T. L., Xu T., Chen Y. L., Malet H., Egloff M. P., Canard B., Vasudevan S. G., & Lescar J. (2007). Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J. Virol.*, *81*(9), 4753-4765. https://doi.org/10.1128/JVI.02283-06

Yu L., Nomaguchi M., Padmanabhan R., & Markoff L. (2008). Specific requirements for elements of the 5' and 3' terminal regions in flavivirus RNA synthesis and viral replication. *Virology*, *374*(1), 170-185. https://doi.org/10.1016/j.virol.2007.12.035

Zhao Y., Soh T. S., Zheng J., Chan K. W., Phoo W. W., Lee C. C., Tay M. Y., Swaminathan K., Cornvik T. C., Lim S. P., et al. (2015). A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog.*, *11*(3), e1004682. https://doi.org/10.1371/journal.ppat.1004682

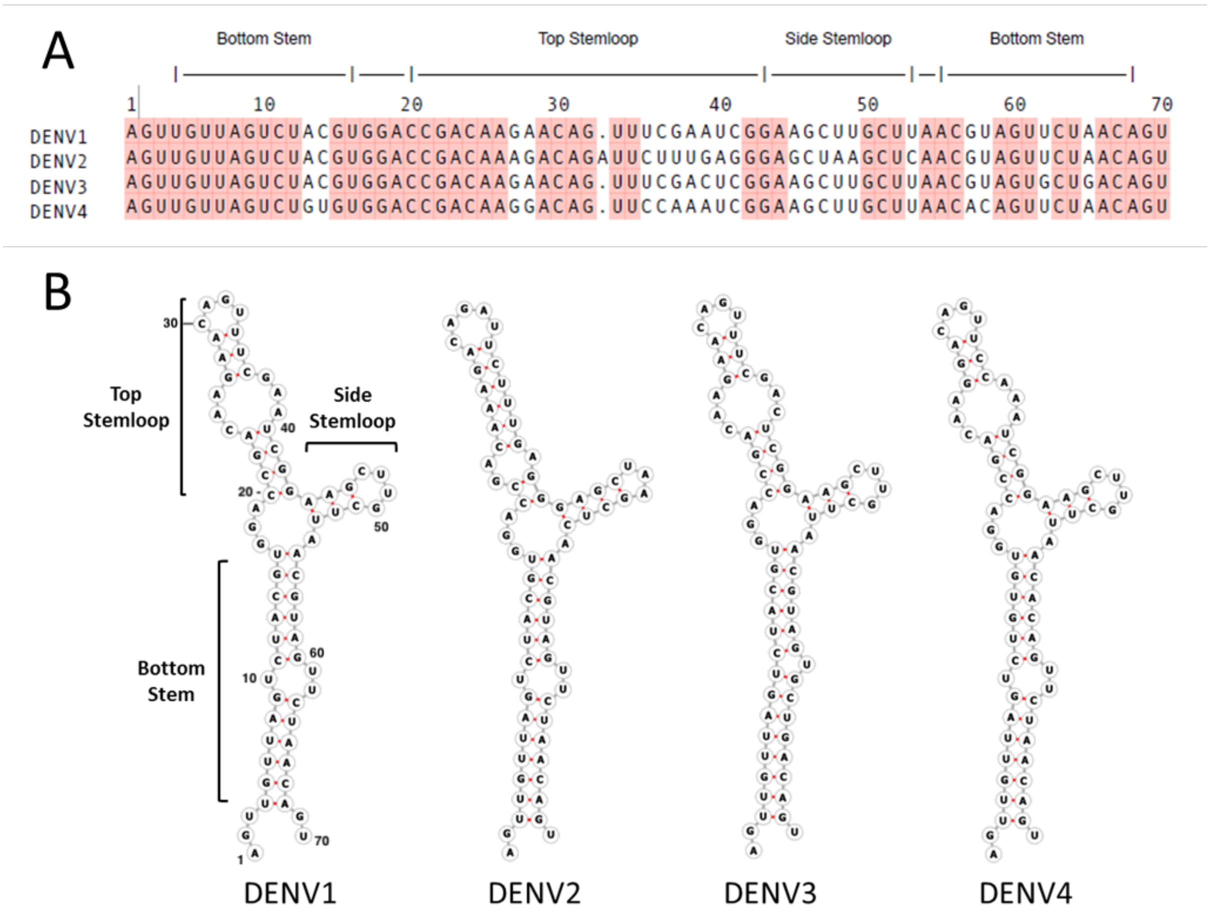


Figure 3.1. The sequence and secondary structure of the SLA promoter element that recruits the NS5 protein to initiate minus-strand RNA synthesis is very highly conserved among the four Dengue virus serotypes (DENV1-4). (A) Alignment of the sequences of the first 70 nucleotides of DENV1, DENV2, DENV3, and DENV4. The regions corresponding to the predicted secondary structure elements (bottom helix, top and side stem-loops) are indicated at the top. Conserved nucleotides are shaded. (B) Predicted RNA secondary structure of the SLA for the four DENV serotypes, as confirmed by SHAPE (Dethoff et al., 2018).

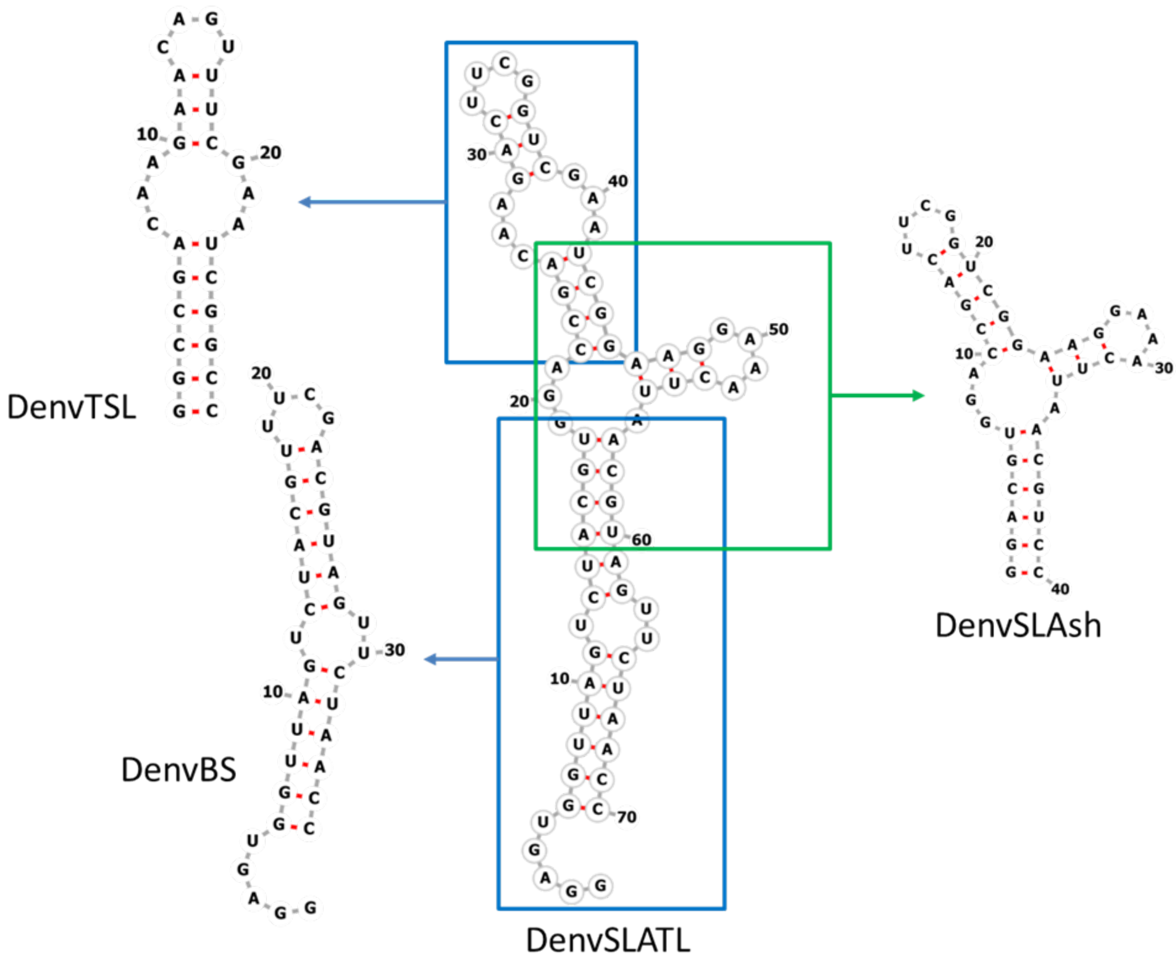


Figure 3.2. The sequences and secondary structures of each RNA segment studied in this work. Three smaller constructs, corresponding to independently folded secondary structure domains, were prepared to facilitate structure determination: DenvBS represents the bottom helix; DenvTSL the top stem-loop and DenvSLAsh the three-way junction. Secondary structures for each segment were predicted using the UNAFold web server (http://www.unafold.org) (Markham & Zuker, 2008; Zuker, 2003) and verified by the NMR assignments.

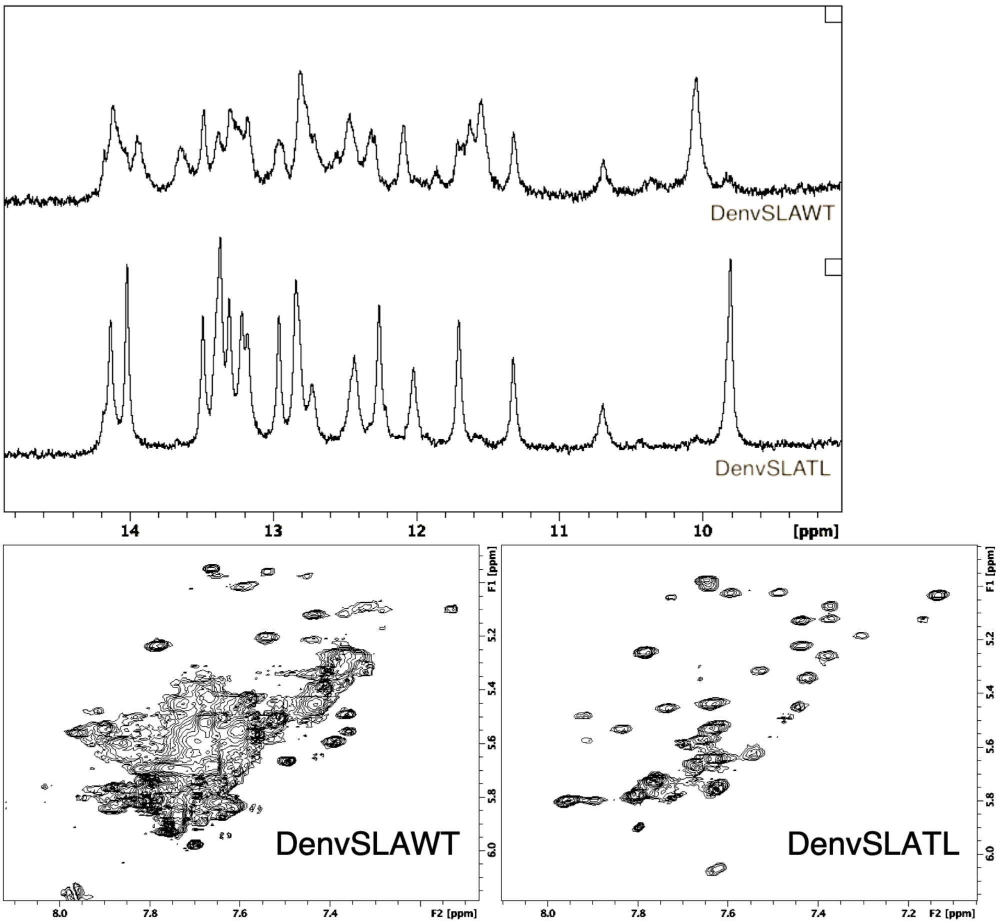
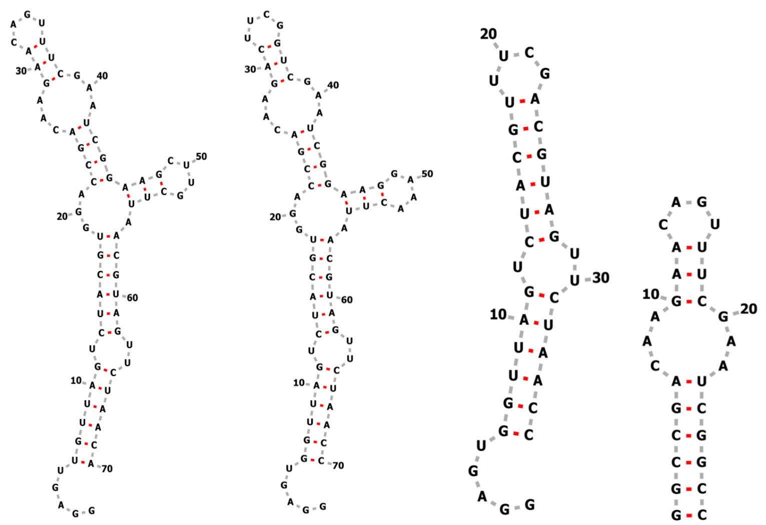


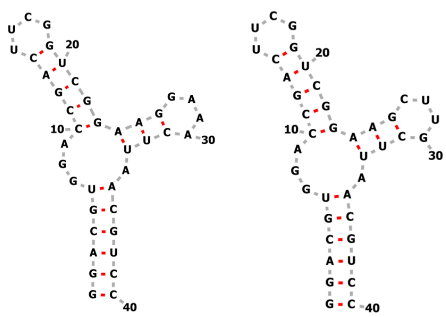
Figure 3.3. (A) Comparison of the imino region of the 1D 1H spectra of wild-type DENV1 SLA (called DenvSLAWT) and a UUCG-tetraloop-stabilized SLA promoter (called DenvSLATL). The addition of stabilizing tetraloops has been observed to improve linewidth with multiple RNAs we studied in the past, and was also observed in this case. The spectra of DenvSLAWT and DenvSLATL are similar, but the quality is much higher for the tetraloop-containing RNA with less overlaps and well-separated peaks. This is also in part caused by the partial dimerization of the wide-type RNA through the side stem-loop at the mM concentrations of NMR experiments, which would further increase the linewidth and the number of peaks in the spectrum. (B) Comparison of the H5-H6 region of the 2D TOCSY spectra of wild-type (DenvSLAWT) and tetraloop-stabilized SLA (DenvSLATL). While many of the peaks are in similar locations, the spectra for DenvSLAWT have much broader

linewidth resulting in severe peak overlaps, which would make peak assignments extremely difficult. The addition of tetraloops improves linewidth and allows identification of discrete peaks, as we have observed with multiple RNAs in the past (Barnwal et al., 2016; Sharma & Varani, 2020; Walker et al., 2020). This strategy is analogous to what is often done in crystallography, where crystallization modules (e.g. protein binding sites) are introduced to facilitate crystal packing and increase resolution.

Name	Length	Sequence
DenvSLAWT	70	GGAGU UGUUA GUCUA CGUGG ACCGA CAAGA ACAGU UUCGA AUCGG AAGCU UGCUU AACGU AGUUC UAACA
DenvSLATL	70	GGAGU GGUUA GUCUA CGUGG ACCGA CAAGA CUUCG GUCGA AUCGG AAGGA AACUU AACGU AGUUC UAACC
DenvBS	36	GGAGU GGUUA GUCUA CGUUU CGACG UAGUU CUAAC C
DenvTSL	28	GGCCG ACAAG AACAG UUUCG AAUCG GCC
DenvSLAsh	40	GGACG UGGAC CGACU UCGGU CGGAA GGAAA CUUAA CGUCC
SLAshCUUG	40	GGACG UGGAC CGACU UCGGU CGGAA GCUUG CUUAA CGUCC



DevnSLAWT DenvSLATL DenvBS DenvTSL



DenvSLAsh SLAshCUUG

Figure 3.4. Sequences and secondary structures of all RNAs studied in this work; all secondary structures were verified by NMR assignments of NH resonances.

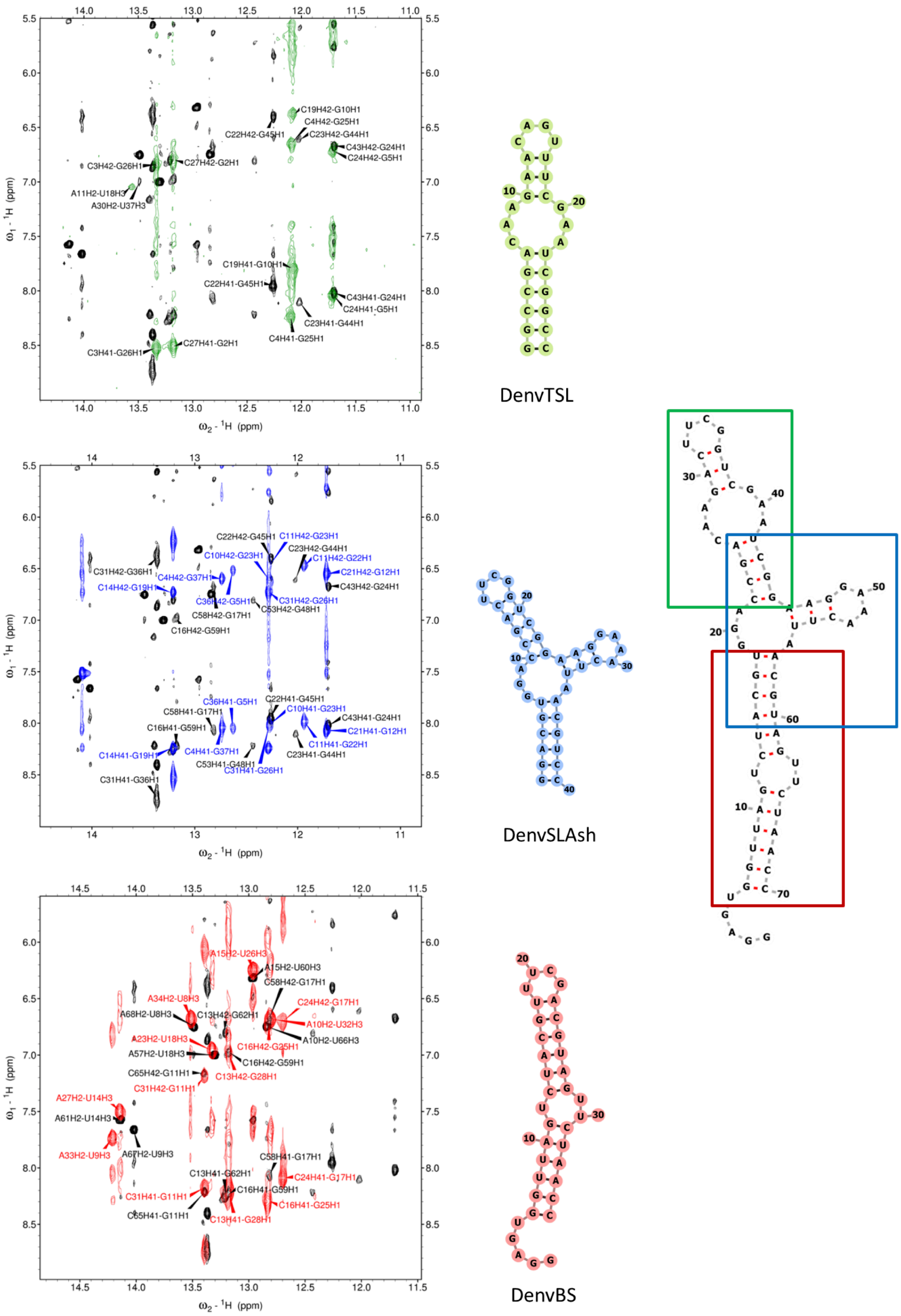


Figure 3.5. Overlay of the imino regions of NOESY spectra for DenvSLATL (black) to DenvBS (red), DenvSLAsh (blue), and DenvTSL (green). Assignments for each RNA are shown in the corresponding color. High similarity of chemical shifts and NOE patterns were

observed in each segment; thus, assignments for DenvSLATL were safely transferred from each corresponding segment.

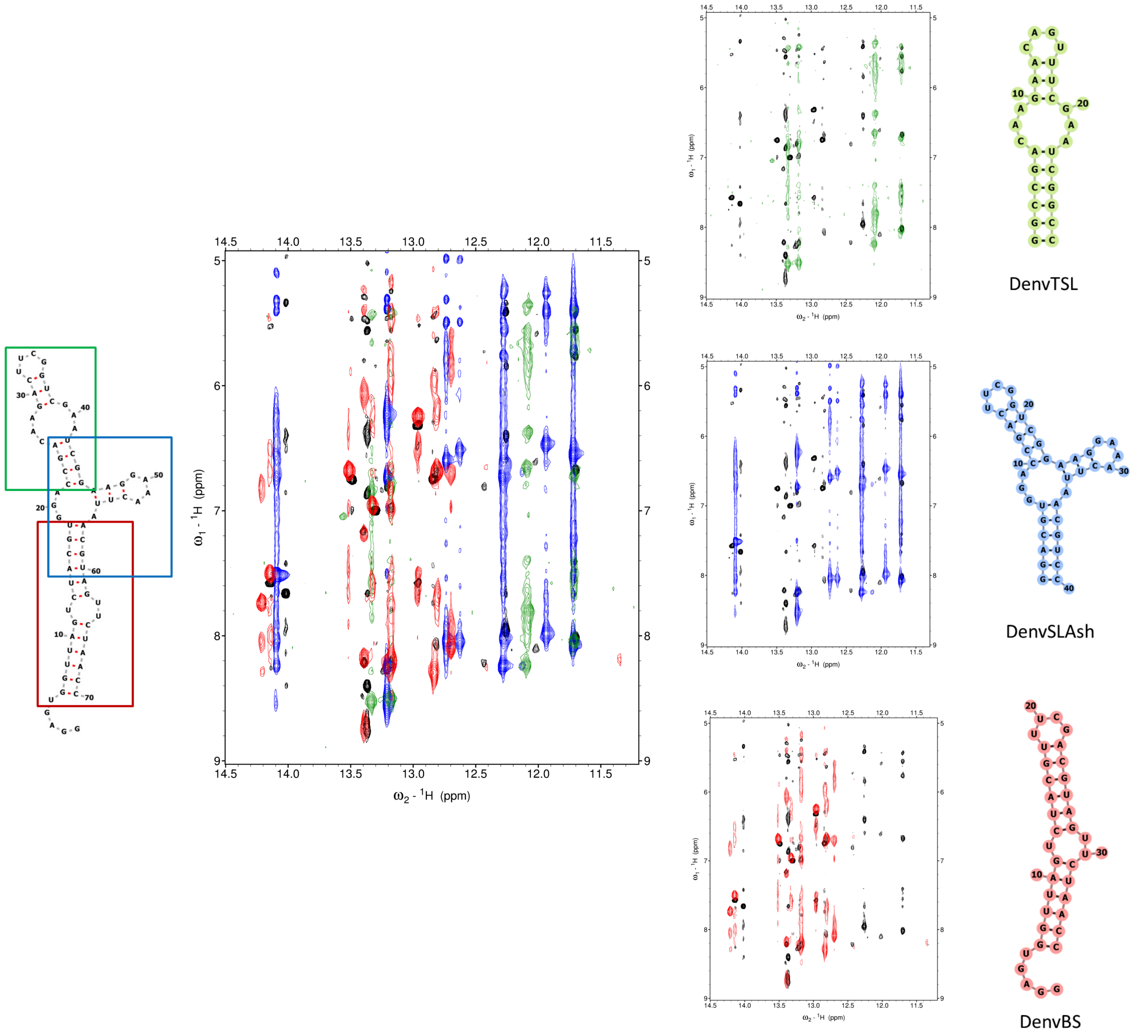


Figure 3.6. Overlay of the imino region of NOESY spectra for DenvSLATL (black) and the three smaller constructs which were prepared to facilitate structure determination. Black color corresponds to DenvSLATL, red to DenvBS, the bottom helix; green corresponds to DenvTSL, the top stem-loop, and blue corresponds to DenvSLAsh, the three-way junction. ^{1}H NOESY spectra were recorded in 10 mM potassium phosphate buffer (pH 6.5, 90% H2O/ 10% D2O) at 25°C. Despite small differences, high similarity of chemical shifts and NOE patterns were observed, allowing the transfer of peak assignments from each segment to the SLA model oligonucleotide (Figure 3.5).

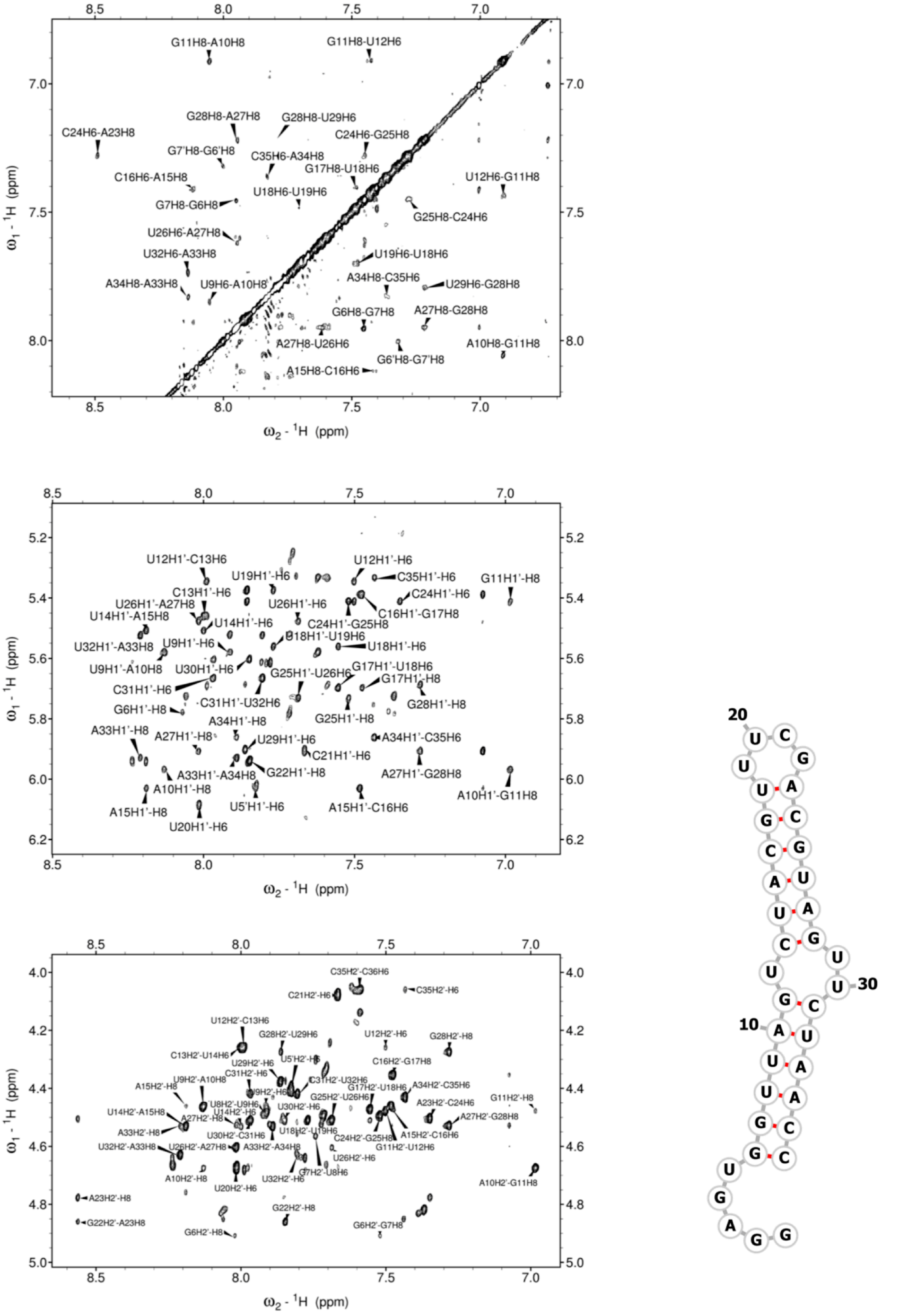


Figure 3.7. NOESY spectra of non-exchangeable proton region (H6/H8-H6/H8 and sugar proton-H6/H8) for DenvBS; the secondary structure is shown on the right. Deuteration of H5, H3', H4', H5', H5 ''protons relieved the extensive overlap in the sugar proton region.

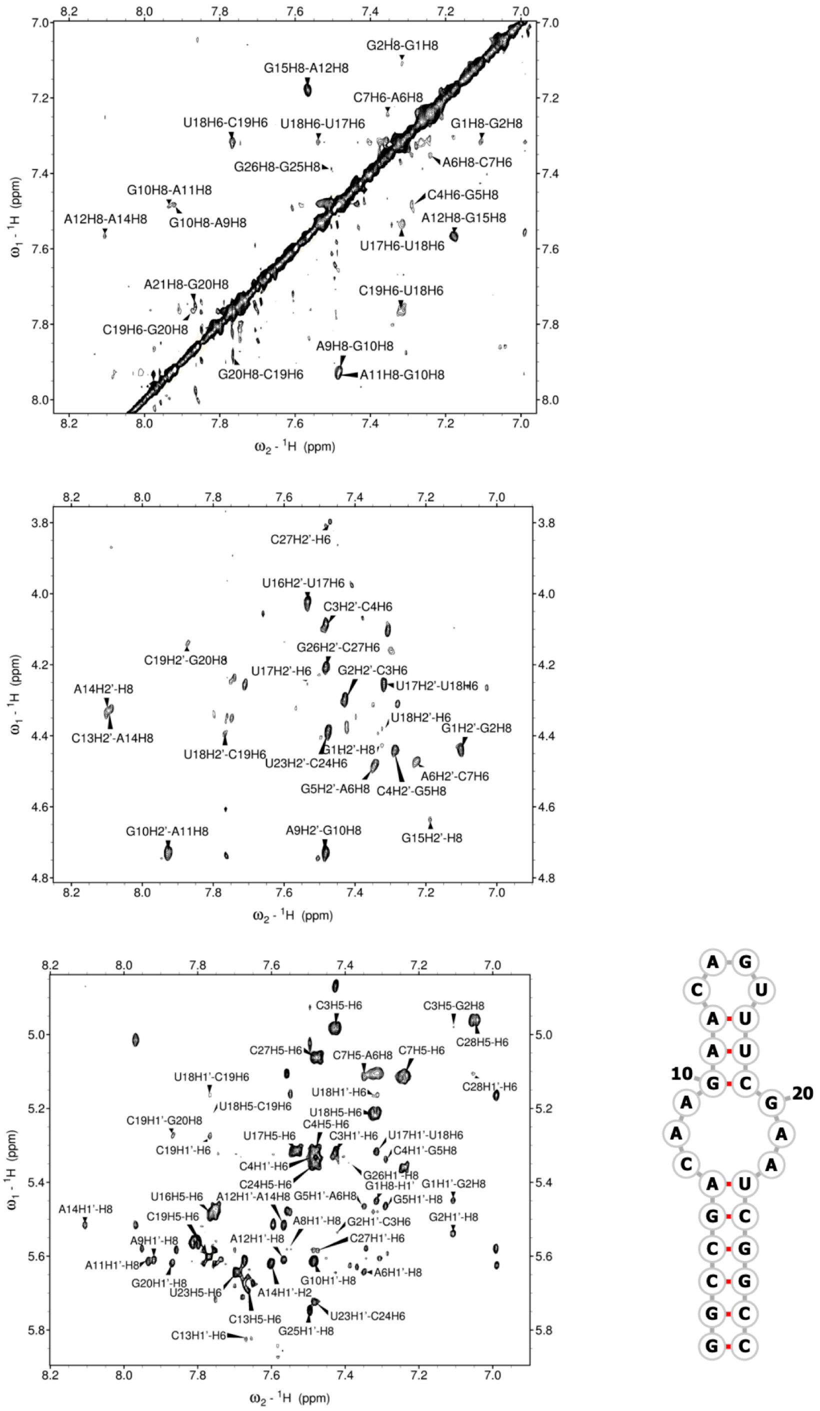


Figure 3.8. NOESY spectra of non-exchangeable proton region (H6/H8-H6/H8 and sugar proton-H6/H8) for DenvTSL; the secondary structure is shown on the right. Deuteration of H5, H3', H4', H5', H5 ''protons relieved the extensive overlap in the sugar proton region.

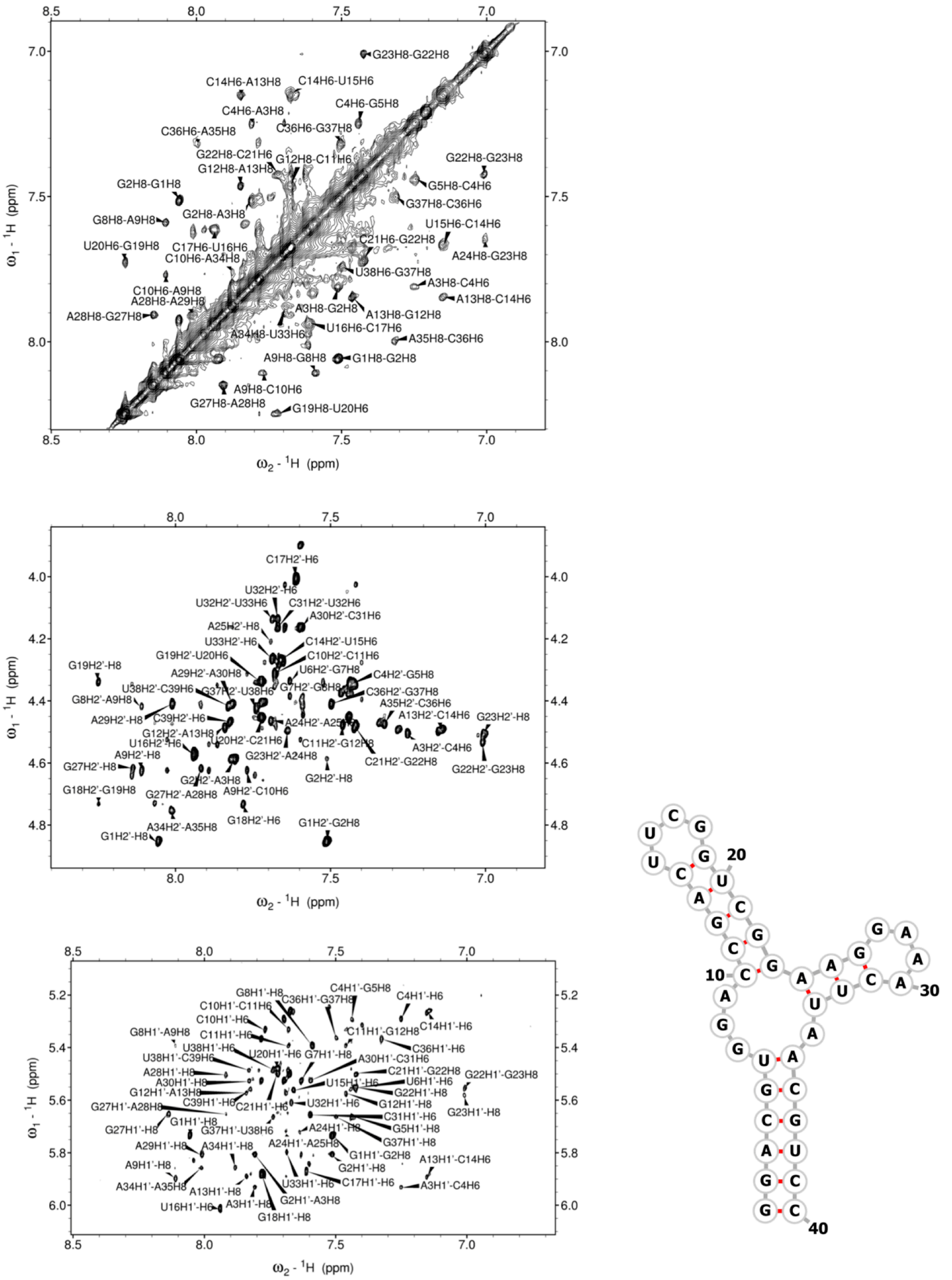


Figure 3.9. NOESY spectra of non-exchangeable proton region (H6/H8-H6/H8 and sugar proton-H6/H8) for DenvSLAsh; the secondary structure is shown on the right. Deuteration of H5, H3', H4', H5', H5 ''protons relieved the extensive overlap in the sugar proton region.

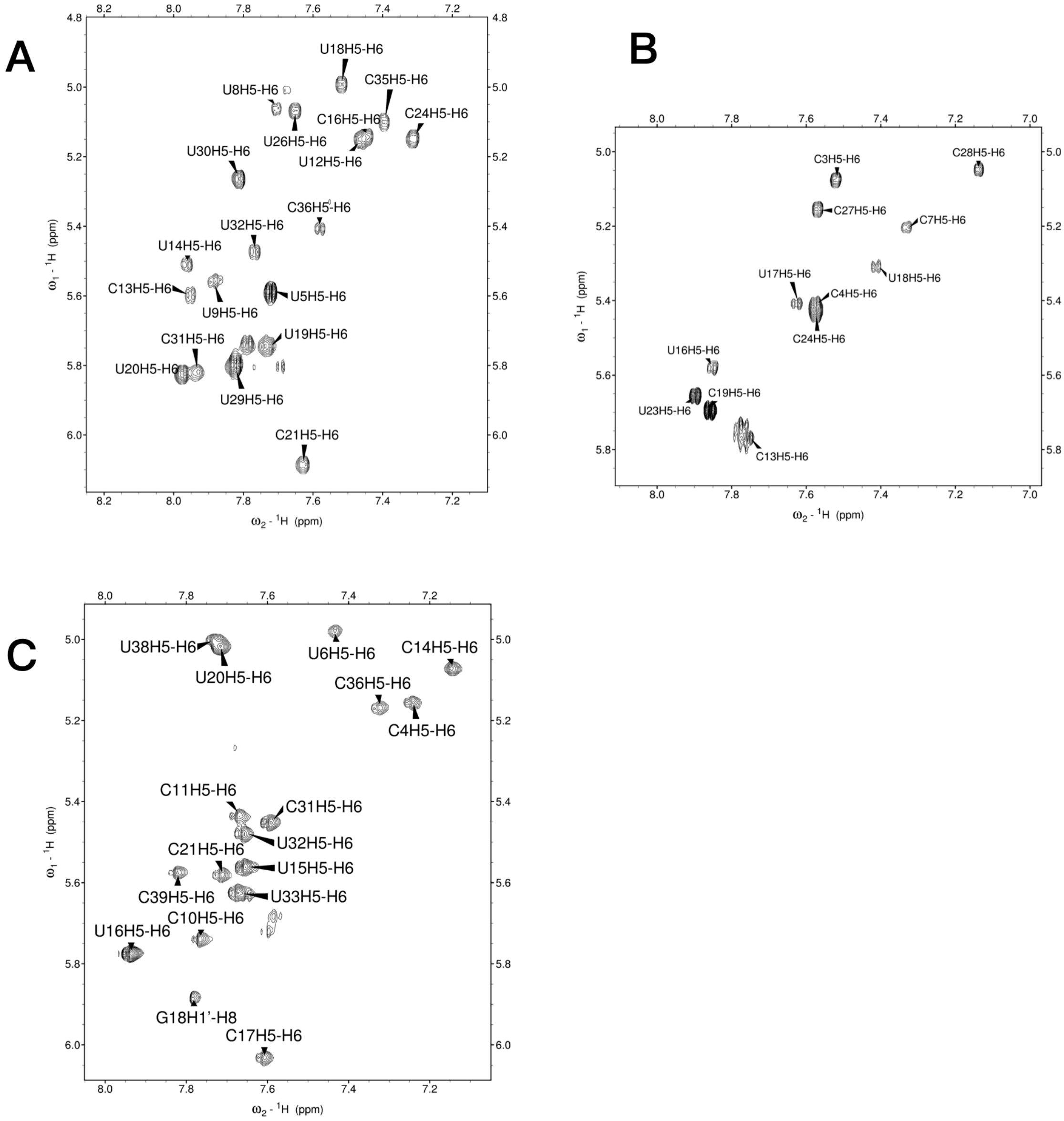


Figure 3.10. TOCSY spectra for the three independently folded structural segments: (A) DenvBS, (B) DenvTSL, and (C) DenvSLAsh. The existence of a single monomeric conformation for each RNA was confirmed from the number of Ura and Cyt H5-H6 peaks in TOCSY spectra, which was as predicted for each sequence.

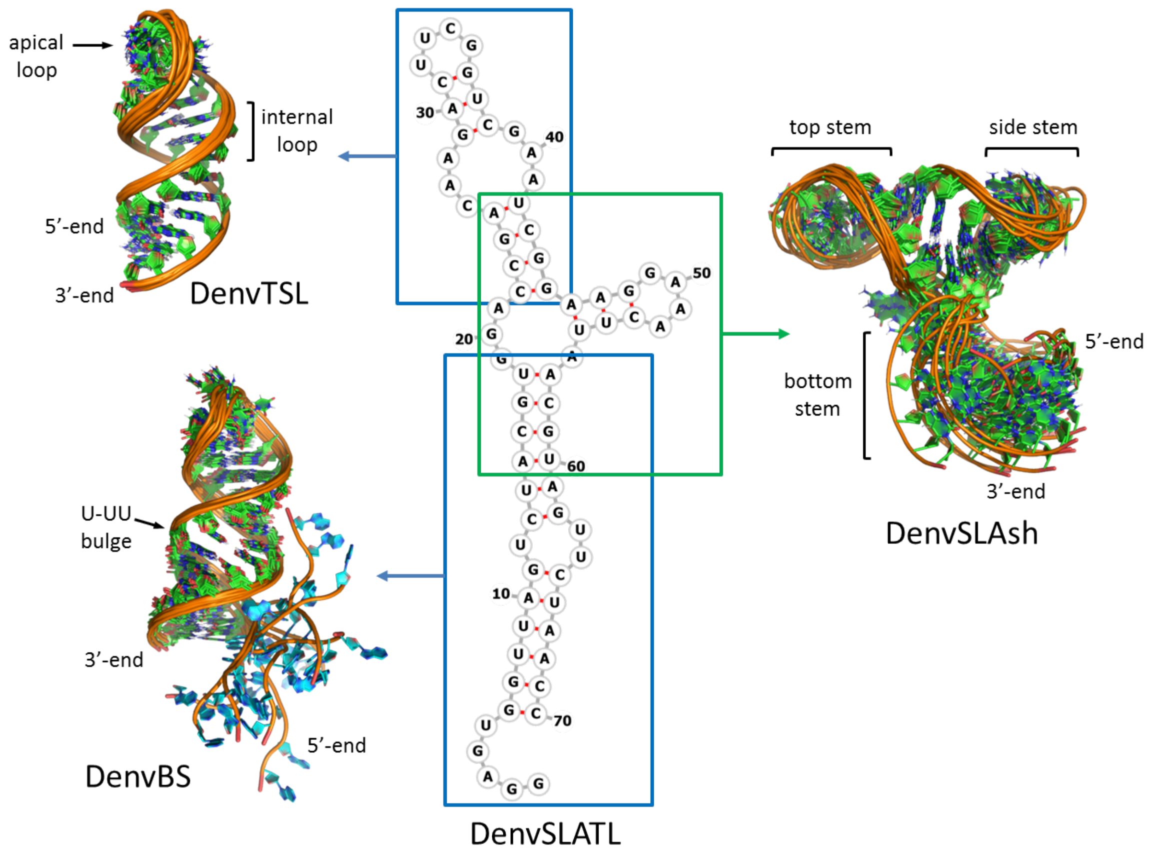


Figure 3.11. Superposition of the 10 NMR structures of lowest energy for each independently folded RNA segment studied in this work. DenvBS represents the bottom helix and forms a well-defined double helix with an unstructured 5'-tail (G1-U5, in cyan) and a U-rich bulge; DenvTSL represents the top stem-loop and forms a short helix interrupted by an internal loop, capped by a partially flexible loop; DenvSLAsh represents the three-way junction and forms a L-shape structure, where the top and side stem helices stack coaxially, and the bottom helix is roughly perpendicular to the coaxial stack; this last image was generated by superposing only the two stacked helices to emphasize the coaxial stack and partial flexibility at the 3-way junction.

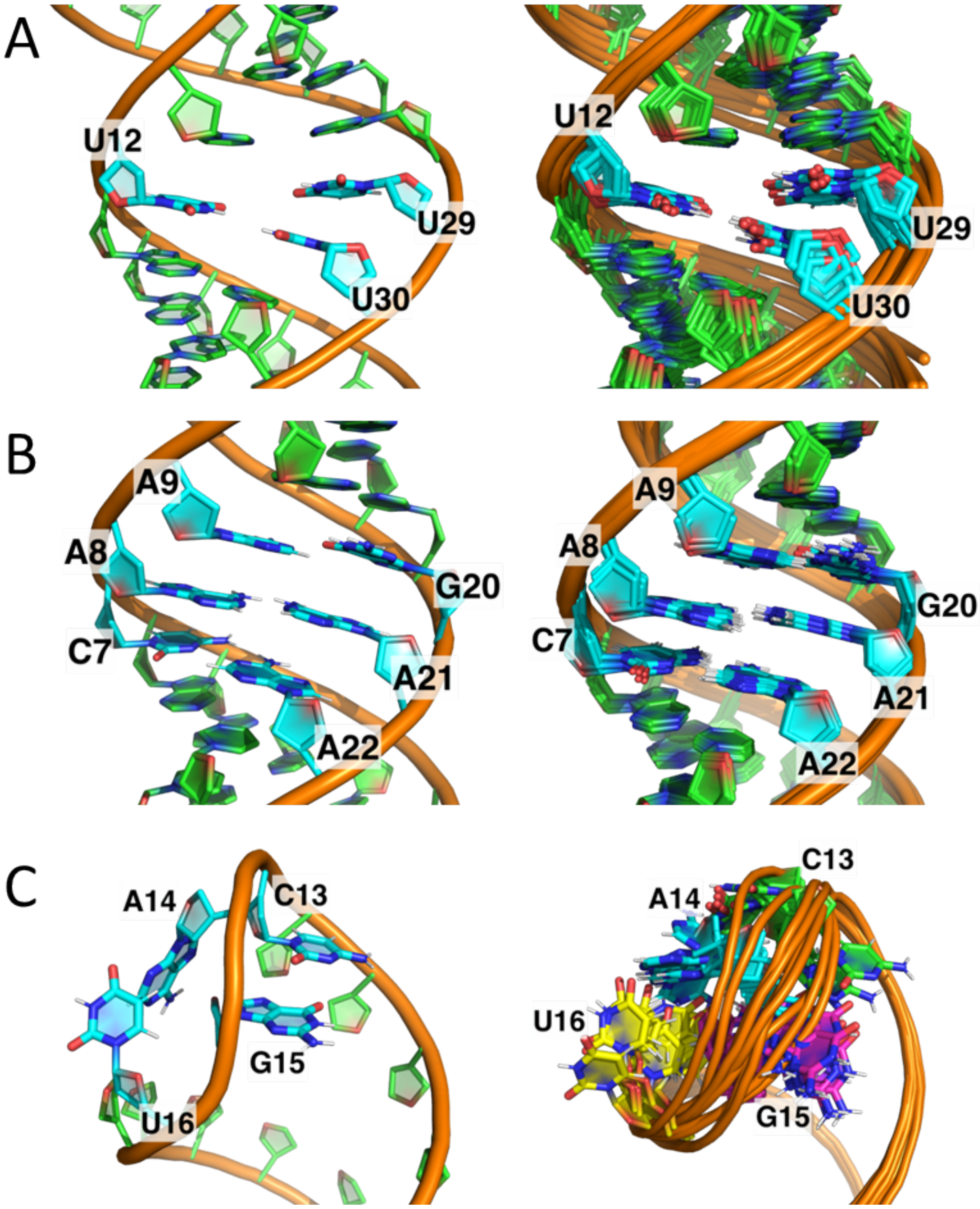


Figure 3.12. (A) Close-up view of the U-UU bulge in DenvBS, drawn from the lowest energy structure (left) and superposition of 10 representative NMR structures (right). Ura12 and Ura30 are coplanar, potentially forming an unstable base pair (no NH peak is visible in 2D spectra), while Ura29 stacks between Gua28 and Ura30, retaining continuous stacking. (B) A close-up view of the internal loop in DenvTS, taken from the lowest energy structure (left) and the superposition of 10 representative NMR structures (right). Nucleotides Cyt7-Ade9 and Gua20-Ade22 across the internal loop stack within the helix. Although no evidence

of non-canonical base-pairs formation in this region was observed in NMR spectra, uninterrupted sequential NOEs were observed (Figure 3.8). (C) A close-up view of the terminal loop in DenvTSL. from the lowest energy structure (left) and the superposition of 10 representative NMR structures (right). In this last image, nucleotides are colored differently for clarity (C13 in green, A14 in cyan, G15 in magenta, and U16 in yellow). In all 10 structures, Gua15 points toward the major groove while Ade14 and Ura16 point outward towards the solvent, while Cyt13 is poorly defined.

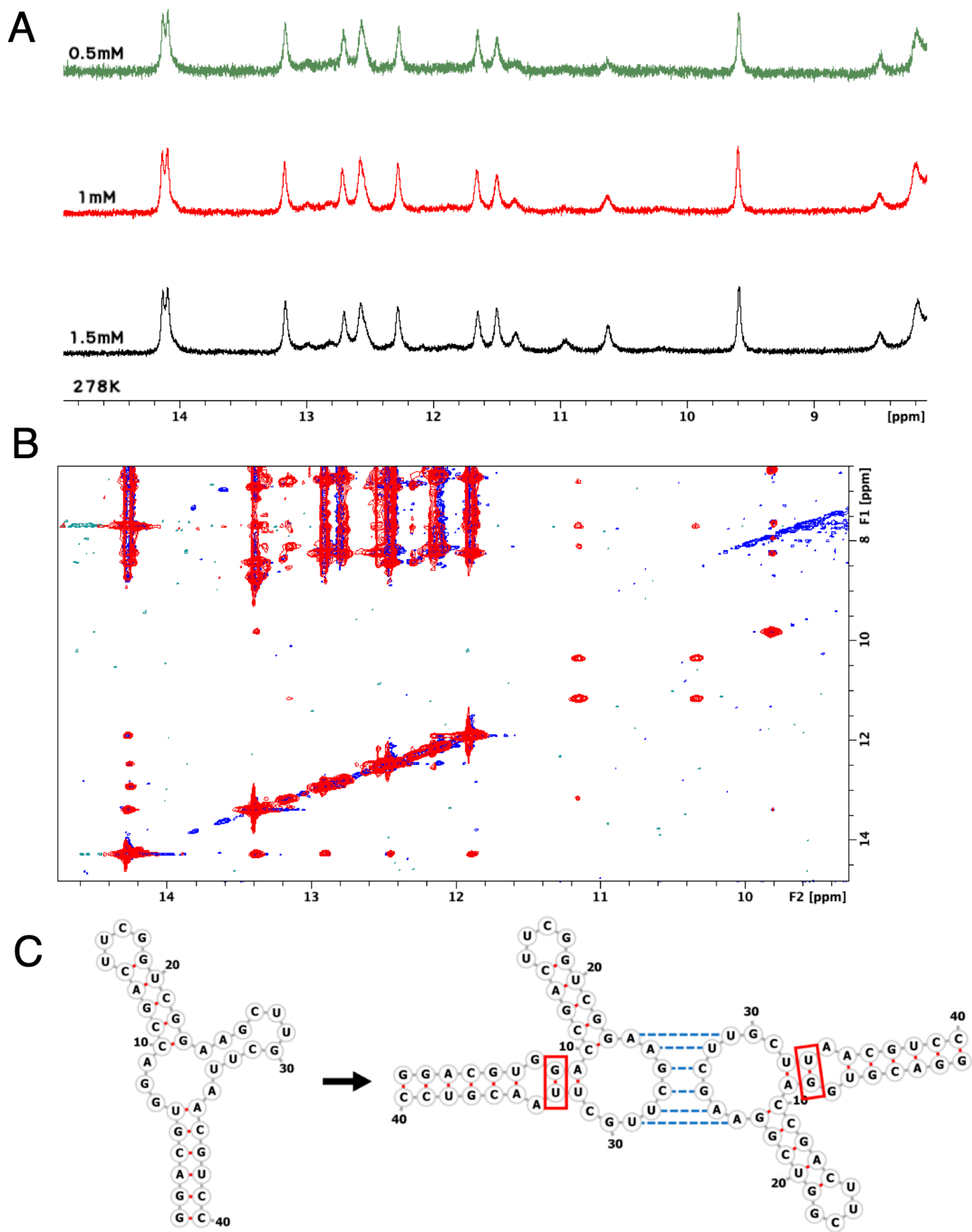


Figure 3.13. (A) 1D ^{1}H NMR spectra of SLAshCUUG recorded at 5°C at concentrations of 0.5 mM (green), 1 mM (red), and 1.5 mM (black). Concentration-dependent peaks are observed between 10-12 ppm and between 12 and 13 ppm, which imply formation of new base pairs as a result of dimerization, which of course becomes more favorable at higher RNA concentration. (B) Overlay of 2D ^{1}H NOESY spectra for SLAshCUUG (red) and DenvSLAsh (blue). For SLAshCUUG, extra base pair peaks are observed between 10-11 ppm, consistent with formation of a GU base pair. By substituting CUUG with a GAAA

tetraloop, the extra resonances are eliminated. Peaks corresponding to base pairs in the side stem were not observed in the spectra of SLAshCUUG, implying that the secondary structure of the side stem is likely to be unstable. (C) Secondary structures of the SLAshCUUG monomer and the presumed SLAshCUUG dimer. The secondary structure shown here corresponds to the dimer observed in the crystal structure (Lee et al., 2021) and contains a GU base pair.

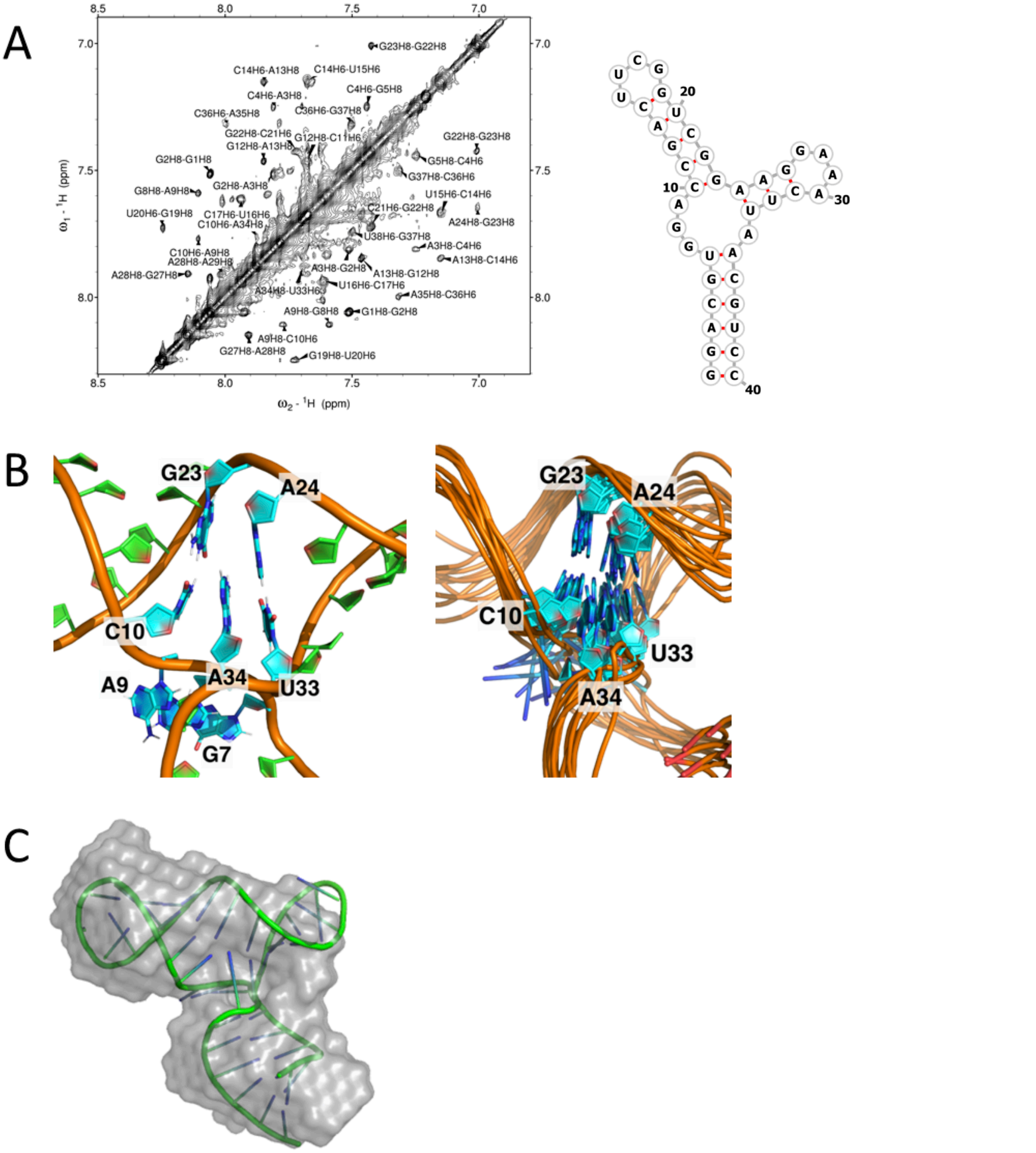


Figure 3.14. (A) NOESY spectrum of DenvSLAsh showing H6/H8-H6/H8 interactions; the secondary structure of DenvSLAsh is shown on the right. This spectrum was recorded at 25°C in 10 mM potassium phosphate buffer (pH 6.5, 90% H_2O/ 10% D_2O). The following NOE interactions: G23H8-A24H8, U33H6-A34H8 and A34-C10H6 within the three-way junction establish coaxial stacking of the top and side stem-loops. (B) A close-up view of the three-way junction in DenvSLAsh from the lowest energy structure (left) and superposition

of 10 representative NMR structure (right). Ade34 stacks within the coaxial helix, between Cyt10 at the end of top stem and Ura33 at the end of the side stem-loop. (C) The lowest energy structure of DenvSLAsh in shown in cartoon representation superposed on the SAXS model; the SAXS results were not used for NMR refinement, and therefore provide independent validation of the structure.

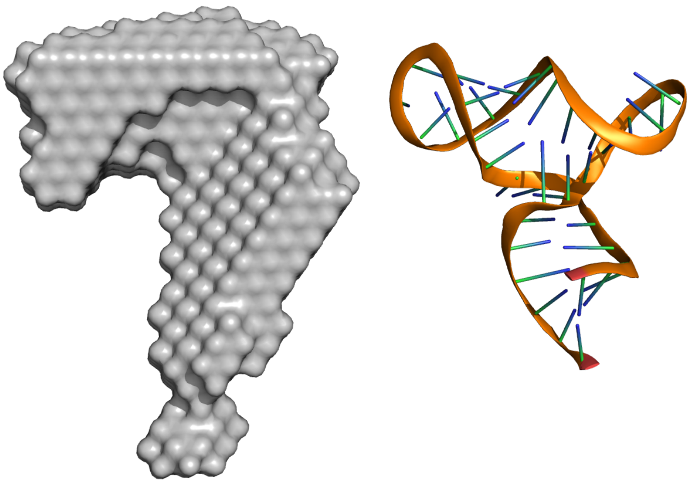


Figure 3.15. The SAXS model of SLAshCUUG demonstrates a fully monomeric structure at concentrations below those used for NMR. Data was collected using the same process described in Method at RNA concentrations of 0.1-2 mg/mL. The structure of DenvSLAsh is shown on the right for comparison.

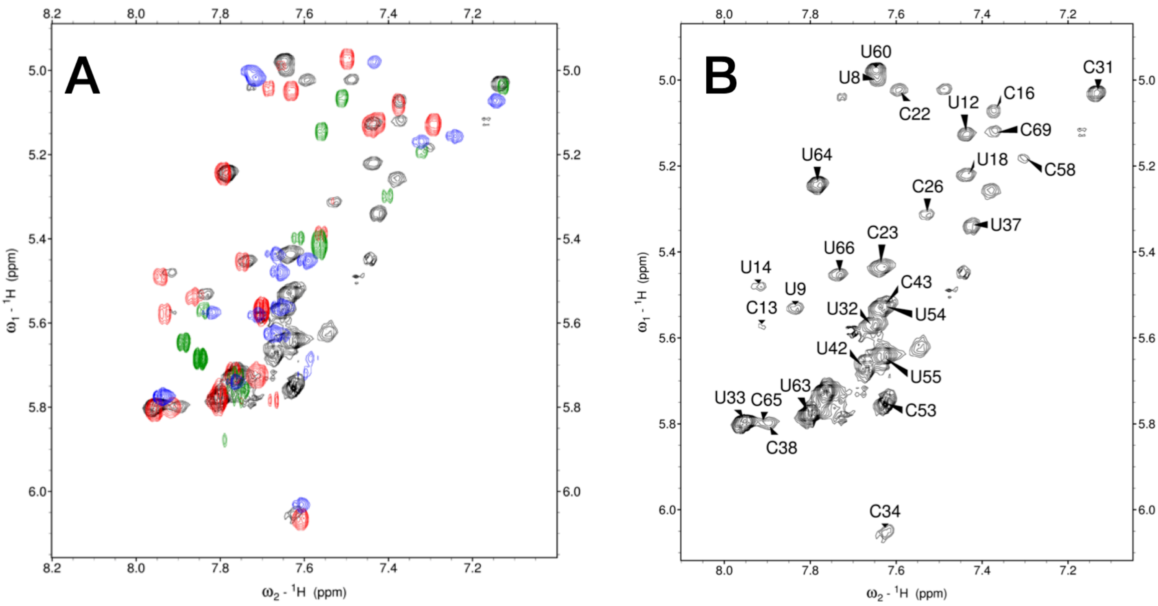


Figure 3.16. (A) Overlay of the TOCSY spectra for DenvSLATL (black), DenvBS (red), DenvTSL (green) and DenvSLAsh (blue). The similarity of chemical shifts allowed me to assign H5-H6 cross-peaks for DenvSLATL, as shown. (B) H5-H6 peak assignments for the TOCSY spectrum of the complete DenvSLATL. A single dominant monomeric conformation was confirmed from the number of Ura and Cyt H5-H6 peaks in TOCSY spectra, which was consistent with its sequence.

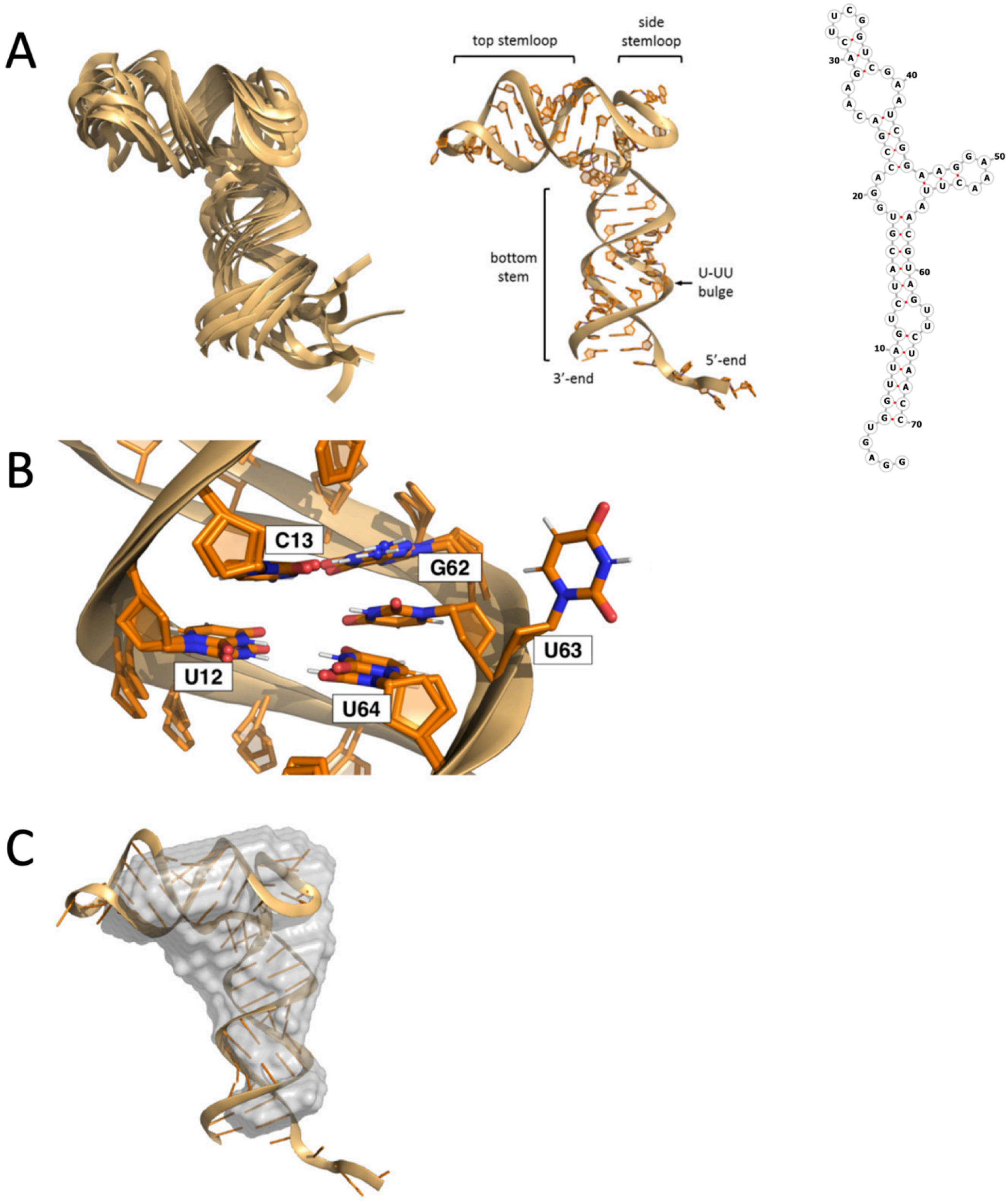


Figure 3.17. (A) Superposition of the 10 NMR structures of DenvSLATL (left), the structure with the lowest calculated energy (middle), and the secondary structure (right) are also shown. (B) A close-up view of the U-rich bulge of DenvSLATL. Two conformations of U-UU bulge were observed in the calculated NMR structures, as shown in the image. (C) The lowest energy structure of DenvSLATL in cartoon representation superposed on the SAXS envelope; the SAXS results were not used for refinement, and therefore provide independent validation of the NMR structure.

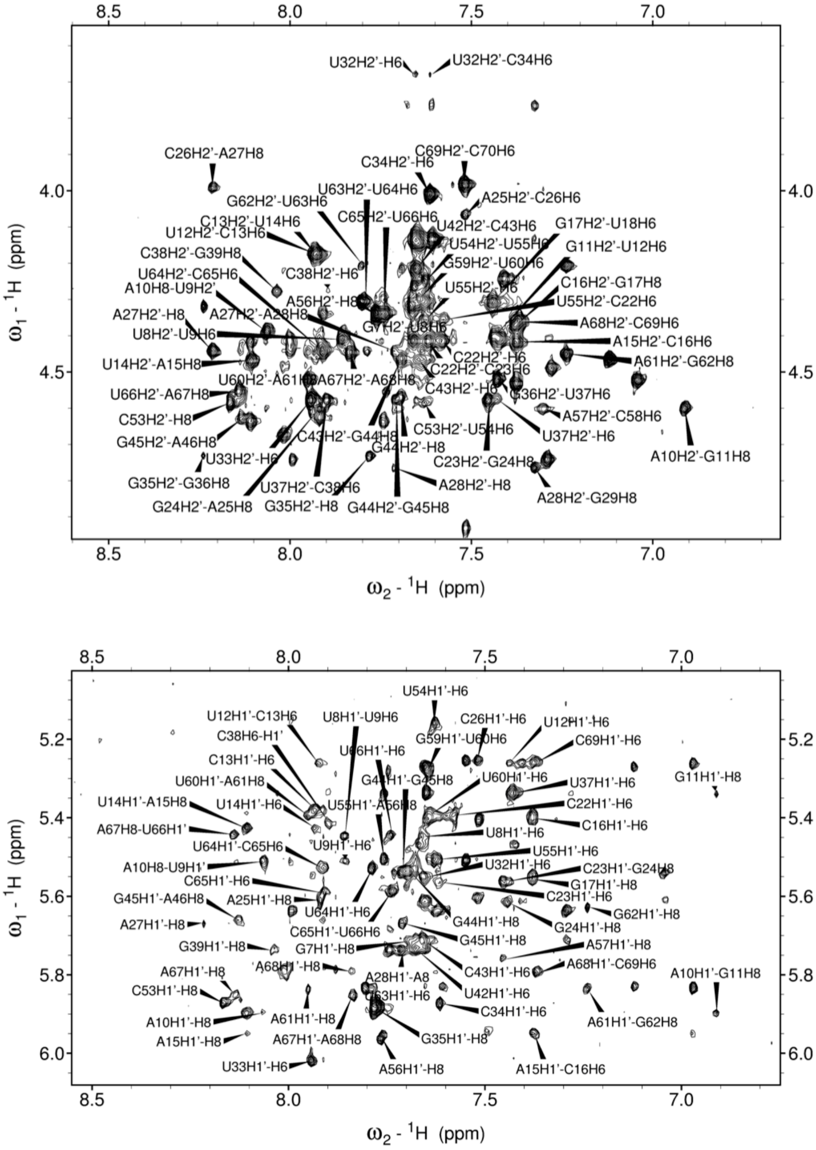


Figure 3.18. NOESY spectra of the H1'-H6/H8 and H2'-H6/H8 region for DenvSLATL. Selective deuteration of H5, H3', H4', H5', H5 ''protons was applied to relieve the extensive overlap in the sugar proton region, leading to interpretable and fully assignable spectra.

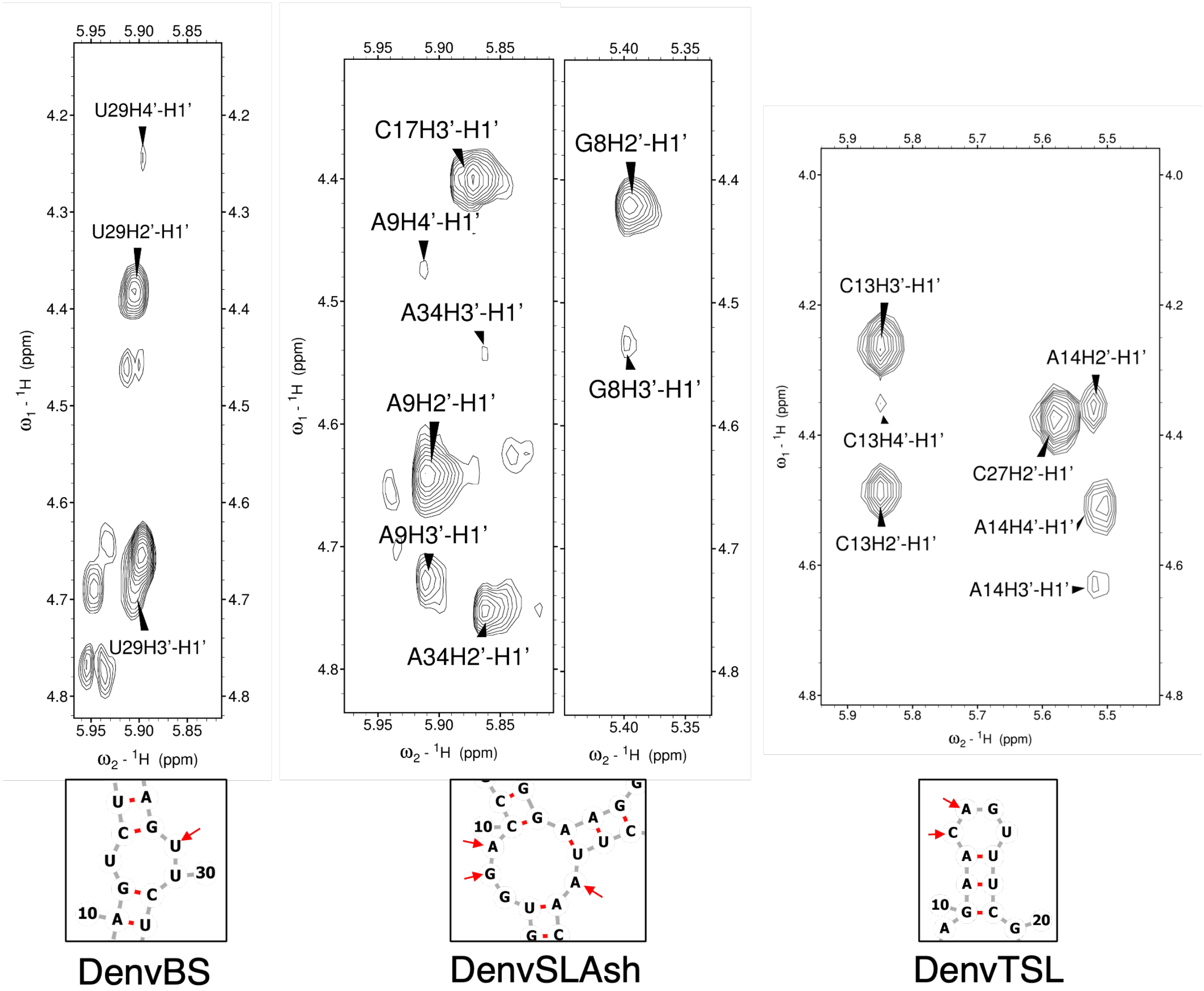


Figure 3.19. Sugar proton region of TOCSY spectra of DenvBS, DenvSLAsh and DenvTSL. For 3'-endo sugar pucker, no peak is observable because the scalar couplings are small; 2'-endo conformation has strong H1'-H2', medium H1'-H3 'peaks and very weak/unobservable H1'-H4 'peaks. For a mixture of 3 'and 2'-endo conformations, peaks from H1 'to H2', H3 ' and H4 'are observed. According to the TOCSY spectra, Ura29 within the U-bulge, Ade9 and Gua8 in the three-way junction and the Cytosine in the upper loop have 2'-endo conformations. The conformation of Ade34 cannot be established because of spectral overlap around 4.4 ppm. Only the Ade in the apical loop occupies a mixture of 2 'and 3'-endo conformations, which implies a dynamic structure for this nucleotide.

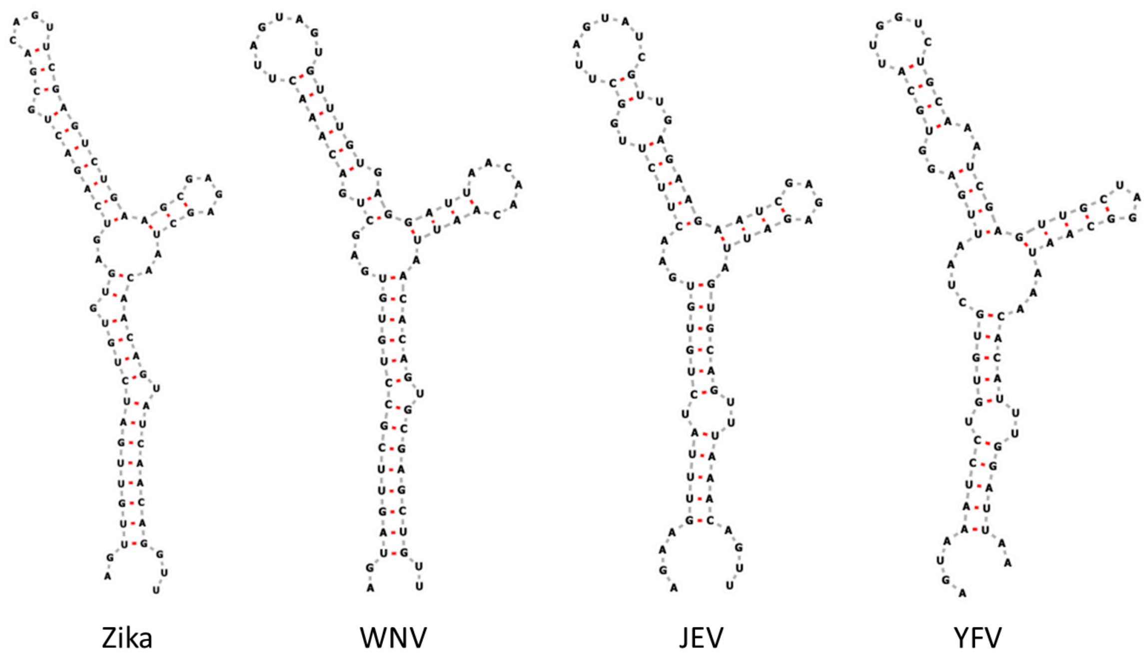


Figure 3.20. Predicted SLA secondary structures, produced by Vienna fold, for different flaviviruses (Lodeiro et al., 2009; Markham & Zuker, 2008; Zuker, 2003). Results are shown for Zika virus (NC_012532, nucleotide 1-106), West Nile virus (NC_001563, nucleotide 1-96), Japanese encephalitis virus (NC_001437, nucleotide 1-95), and Yellow fever virus (NC_002031, nucleotide 1-118).

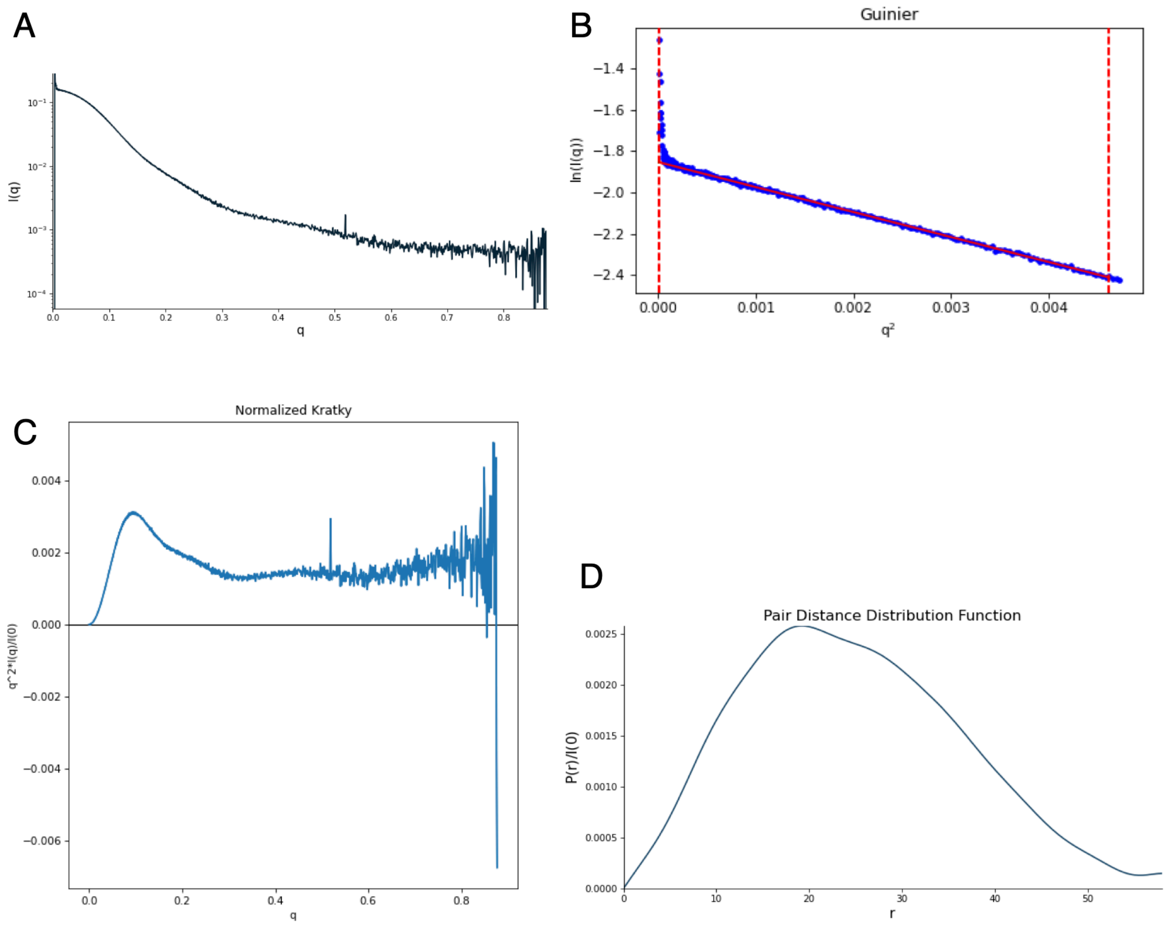


Figure 3.21. Processed SAXS data of DenvSLAsh: (A) scattering curve, (B) Guinier approximation curve; (C) normalized Kratky plot; and (D) particle distance distribution function plot. The radius of gyration Rg and the scattering intensity at zero angle I(0) obtained by Guinier approximation are 19.0 Å and 0.16 cm^{-1}, respectively. The particle distance distribution function P(r) plots was calculated using GNOM (Svergun, 1992).

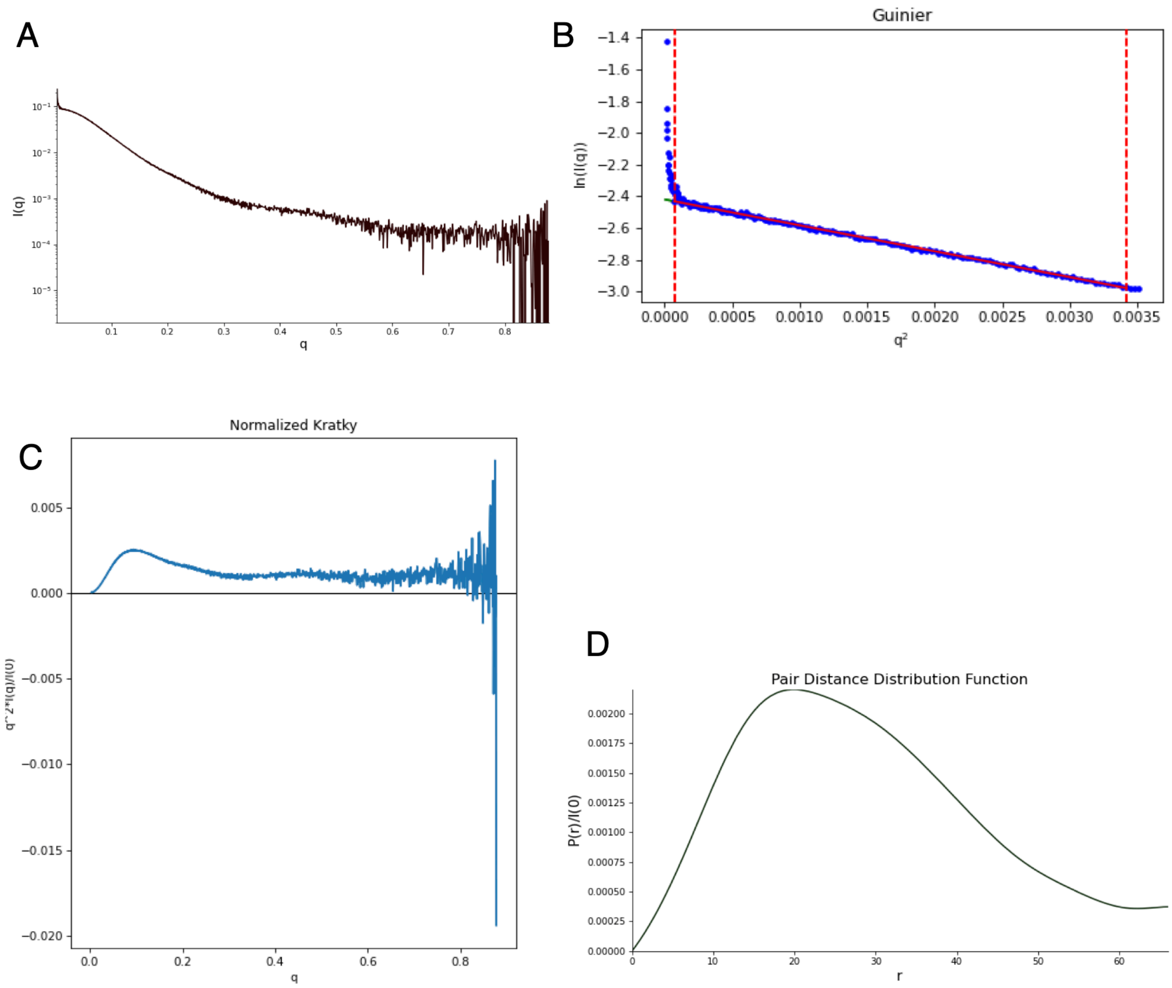


Figure 3.22. Processed SAXS data of SLAshCUUG: (A) scattering curve; (B) Guinier approximation curve; (C) normalized Kratky plot; and (D) particle distance distribution function plot. The radius of gyration Rg and the scattering intensity at zero angle I(0) obtained by Guinier approximation are 22.1 Å and 0.09 cm^{-1}, respectively. The particle distance distribution function P(r) plots was calculated using GNOM (Svergun, 1992).

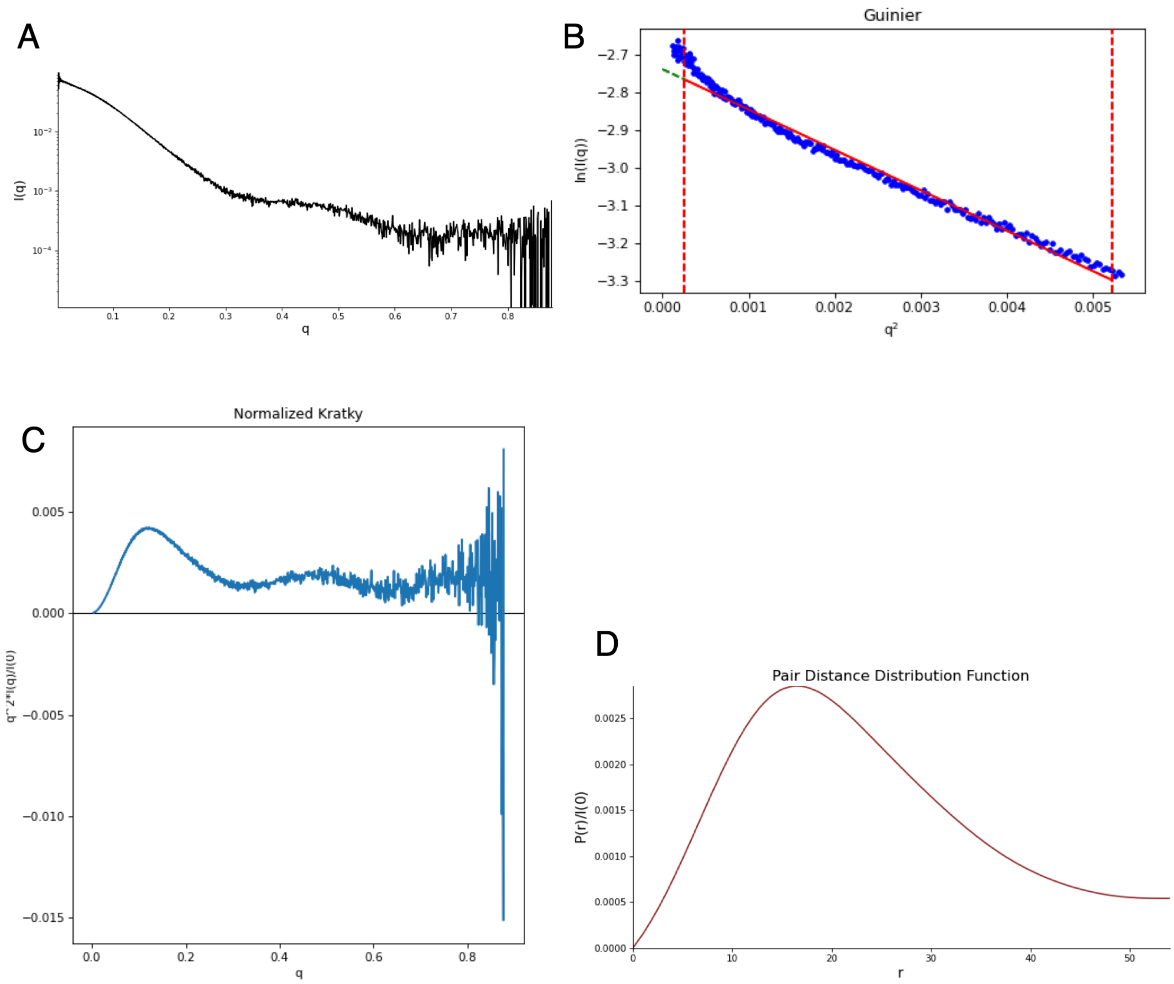


Figure 3.23. Processed SAXS data of DenvSLATL: (A) scattering curve; (B) Guinier approximation curve; (C) normalized Kratky plot; and (D) particle distance distribution function plot. The radius of gyration Rg and the scattering intensity at zero angle I(0) obtained by Guinier approximation are 18.0 Å and 0.06 cm^{-1}, respectively. The particle distance distribution function P(r) plots was calculated using GNOM (Svergun, 1992).

Table 3.1. NMR experimental restraints and structural statistics for the 10 lowest energy structures of DenvBS (bottom helix)

	DenvBS
NMR Peak Assignment (excluding the unstructured 5'-tail, nucleotides 1-5)	
All H1′, H2′, H5, H6, H8, AdeH2, 87.5% of H3', 59.4% of H4' and 18.8% of H5'/H5'' were assigned. Extensive overlap in the sugar proton region was relieved by deuteration of H3', H4', H5', H5'' and H5 protons to assign H1' and H2' unambiguously, but this prevented assignments of many of those sugar protons. 63.2% of imino protons and 85.7% of Cyt amino protons were assigned.	
NMR Experimental Restraints	
Total number of restraints	539
Total NOE Restraints	299
Intra-residue	162
Inter-residue	137
Sequential $\|i-j\| = 1$	116
Non-sequential $\|i-j\| > 1$	21
Hydrogen bond restraints	10
Dihedral Angle Restraints[(a)]	216
Planarity Restraints	24
Structure Analysis	
NOE violations >0.5 Å	0
Torsion angle violations >5º	0
Deviation from Idealized Geometry	
RMS Deviation of Bond lengths (Å)	0.00
RMS Deviation of Bond angles (º)	0.63
Heavy Atom RMSD to the Mean Structure (Å)	
All RNA heavy atoms (6-36, excluding the unstructured 5'-tail)	1.02
All RNA backbone (6-36, excluding the unstructured 5'-tail)	1.00
RNA upper stem heavy atoms (13-28)	0.72
RNA upper stem backbone (13-28)	0.60
RNA lower stem heavy atoms (6-12, 29-36)	0.50
RNA lower stem backbone (6-12, 29-36)	0.53

(a) Generic dihedral restraints were applied to A-form helical regions confirmed by NOESY spectra.

Table 3.2. NMR experimental restraints and structural statistics for the 10 lowest energy structures of DenvTSL (apical stem-loop)

	DenvTSL
NMR Peak Assignment	
All H1′, H2′, H5, H6, H8, and 25% of AdeH2[(a)], 92.9% of H3’, 67.9% of H4 ’and 37.5% of H5'/H5’’ were assigned. Extensive overlap in the sugar proton region was relieved by deuteration of H3’, H4’, H5’, H5 ’’and H5 protons to assign H1 ’and H2 ’unambiguously, but this prevented assignment of many of those deuterated sugar protons. 58.3% of imino protons and 75% of Cyt amino protons were assigned.	
NMR Experimental Restraints	
Total number of restraints	475
Total NOE Restraints	251
Intra-residue	135
Inter-residue	116
Sequential |i-j| = 1	92
Non-sequential |i-j| > 1	24
Hydrogen bond restraints	9
Dihedral Angle Restraints[(b)]	206
Planarity Restraints	18
Structure Analysis	
NOE violations >0.5 Å	0
Torsion angle violations >5º	0
Deviation from Idealized Geometry	
RMS Deviation of Bond lengths (Å)	0.00
RMS Deviation of Bond angles (º)	0.64
Heavy Atom RMSD to the Mean Structure (Å)	
All RNA heavy atoms (1-28)	0.85
All RNA backbone (1-28)	0.75
RNA upper stem heavy atoms (7-22)	1.03
RNA upper stem backbone (7-22)	0.84

(a) H2 of Adenines located in single-stranded loops were not assigned.
(b) Generic dihedral restraints were applied to A-form helical regions confirmed by NOESY spectra.

Table 3.3. NMR experimental restraints and structural statistics for the 10 lowest energy structures of DenvSLAsh (three-way junction)

	DenvSLAsh
NMR Peak Assignment	
90% of H1′, 95% of H2′, 94.1% of H5, 97.5% of H6/H8, 47.5% of H3', 27.5% of H4', 16.3% of H5'/H5 ''and 40% of AdeH2[(a)] were assigned. Extensive overlap in the sugar proton region was relieved by deuteration of H3', H4', H5', H5'' and H5 protons to assign H1' and H2' unambiguously, but this prevented assignments of many of those deuterated sugar protons. 65% of imino protons and 80% of Cyt amino protons were assigned.	
NMR Experimental Restraints	
Total number of restraints	840
Total NOE Restraints	562
Intra-residue	262
Inter-residue	300
Sequential |i-j| = 1	187
Non-sequential |i-j| > 1	113
Hydrogen bond restraints	41
Dihedral Angle Restraints[(b)]	250
Planarity Restraints	28
Structure Analysis	
NOE violations >0.5 Å	0
Torsion angle violations >5º	0
Deviation from Idealized Geometry	
RMS Deviation of Bond lengths (Å)	0.00
RMS Deviation of Bond angles (º)	0.68
Heavy Atom RMSD to the Mean Structure (Å)	
All RNA heavy atoms (1-40)	2.75
All RNA backbone (1-40)	2.67
RNA top and side stem heavy atoms (10-33)	2.07
RNA top and side stem backbone (10-33)	1.88
RNA bottom stem heavy atoms (1-9, 34-40)	1.76
RNA bottom stem backbone (1-9, 34-40)	1.11

(a) H2 of Adenines located in single-stranded loop were not assigned.
(b) Generic dihedral restraints were applied to A-form helical regions confirmed by NOESY spectra.

Table 3.4. NMR experimental restraints and structural statistics for the 10 lowest energy structures of DenvSLATL (full-length monomeric SLA)

	DenvSLATL
NMR Experimental Restraints[(a)]	
Total number of restraints	1277
Total NOE Restraints	843
Intra-residue	402
Inter-residue	441
Sequential |i-j| = 1	303
Non-sequential |i-j| > 1	138
Hydrogen bond restraints	68
Dihedral Angle Restraints	384
Planarity Restraints	50
Structure Analysis	
NOE violations >0.5 Å	0
Torsion angle violations >5º	0
Deviation from Idealized Geometry	
RMS Deviation of Bond lengths (Å)	0.00
RMS Deviation of Bond angles (º)	0.66
Heavy Atom RMSD to the Mean Structure (Å)	
All RNA heavy atoms (6-70)	3.98[(b)]
All RNA backbone (6-70)	4.08[(b)]
Top stemloop heavy atoms (22-45, corresponding to DenvTSL)	1.33
Top stemloop backbone (22-45, corresponding to DenvTSL)	1.26
Bottom stem heavy atoms (6-18, 57-70, corresponding to DenvBS)	1.02
Bottom stem backbone (6-18, 57-70, corresponding to DenvBS)	1.00
Three-way junction heavy atoms (15-25, 42-60, corresponding to DenvSLAsh)	2.68
Three-way junction backbone (15-25, 42-60, corresponding to DenvSLAsh)	2.68

(a) Experimental restraints were derived from structural segments DenvBS, DenvTSL and DenvSLAsh (as described in Materials and Methods).
(b) The RMSD is calculated without the unstructured single-stranded 5'-tail (residues 1-5).

Appendix A.

Scripts and Commands for Peptide Structure Calculation

A1. Scripts

A1.1. peptide.seq

Write the sequence with 3-letter code of amino acids. For example:

```
GLY SER LYS ARG PHE ARG PRO PRO ILE ILE PHE ASN GLU ARG
```

A1.2. generate_cyclic_psf.inp

```
rtf @TOPPAR:protein.top end

evaluate ($kbbang = 500.0)
evaluate ($kbbimp = 500.0)

parameter @TOPPAR:protein.par end

segment
  name = "PeptideName"
  SETUP=TRUE
  chain
    LINK PEPP    HEAD - *     TAIL + PRO      END  { LINK to PRO }
    LINK PEPT    HEAD - *     TAIL + *       END
    sequence
      @peptide.seq
    end
  end
end
vector do ( resid = encode ( decode ( resid ) + 0) ) (segid "PeptideName")
vector do (segid = "PeptideName") (segid "PeptideName")

                          {*Make cyclic peptide.*}
   patch PEPT reference=-=( resid 14 ) reference=+=( resid 1 )
   end

                      {*Change residue 7 into L-form.*}
   patch ltod
     reference=nil=( resid 7)
   end

write structure output = "PeptideName_generate_psf2.psf" end
write coor output = "PeptideName_generate_psf2.pdb" end

stop
```

A1.3. generate_template_cmplx_ani.inp

```
remarks  file  nmr/generate_template.inp
remarks  Generates a "template" coordinate set.  This produces
```

```
remarks  an arbitrary extended conformation with ideal geometry.
remarks  Author: Axel T. Brunger

{====>}
structure @PeptideName_generate_psf2.psf
        end                     {*Read structure file.*}

parameter
{====>}
  @TOPPAR:protein.par                    {*Read parameters.*}
end

!{ Set force constants for S-S bond lengths and angles to zero  }
!parameter
!   bonds ( name SG ) ( name SG ) 0. 1.
!   angle ( name CB ) ( name SG ) ( name SG ) 0. 1
!end

vector ident (x) ( all )
vector do (x=x/10.) ( all )
vector do (y=random(0.5) ) ( all )
vector do (z=random(0.5) ) ( all )

vector do (fbeta=50) (all)            {*Friction coefficient, in 1/ps.*}
vector do (mass=100) (all)                 {*Heavy masses, in amus.*}

parameter
  nbonds
    cutnb=5.5 rcon=20. nbxmod=-2 repel=0.9  wmin=0.1 tolerance=1.
    rexp=2 irexp=2 inhibit=0.25
  end
end

flags exclude * include bond angle vdw end

minimize powell nstep=5000  nprint=10 end

flags include impr end

minimize powell nstep=5000 nprint=10 end

dynamics  verlet
  nstep=5000  timestep=0.001 iasvel=maxwell  firsttemp= 300.
  tcoupling = true  tbath = 300.   nprint=100  iprfrq=0
end

parameter
  nbonds
    rcon=2. nbxmod=-3 repel=0.75
  end
end

minimize powell nstep=5000 nprint=25 end

dynamics  verlet
  nstep=300  timestep=0.005 iasvel=maxwell  firsttemp= 300.
  tcoupling = true  tbath = 300.   nprint=100  iprfrq=0
```

```
end

flags exclude vdw elec end
vector do (mass=1.) ( name h* )
hbuild selection=( name h* ) phistep=360 end
flags include vdw elec end

minimize powell nstep=5000 nprint=50 end
                                        {*Write coordinates.*}
remarks produced by nmr/generate_template.inp
write coordinates output=PeptideName_only.pdb end

stop
```

A1.4. peptide_anneal.inp

```
!
! slow cooling protocol in torsion angle space. Uses only vDw
! NOE constraints with optional coupling, carbon, diheadral and radius of gyration.
!
! CDS 5/2/00
!

eval ($numStructs = 204)       !total number of structures to calculate
eval ($randomSeed = 753)     !random seed
eval ($use_graphics = false) ! VMD-xplor graphics are not used.
eval ($maxResid = 56)        !used for properly configuring TA dynamics

!
! get parallel info, which computer and how many processes.
!
cpyth "from os import environ as env"
cpyth "xplor.command('eval ($proc_num=%s)'  % env['XPLOR_PROCESS']      )"
cpyth "xplor.command('eval ($num_procs=%s)' % env['XPLOR_NUM_PROCESSES'])"
eval ($num_procs=min($num_procs,$numStructs))
if ( $proc_num >= $num_procs ) then  stop end if
eval ($firstStruct = int($proc_num * $numStructs / $num_procs))
eval ($lastStruct  = int(($proc_num+1)  * $numStructs / $num_procs))

!
! get parallel info
!
!cpyth "from os import environ as env"
!cpyth "xplor.command('eval ($proc_num=%s)'  % env['XPLOR_PROCESS']      )"
!cpyth "xplor.command('eval ($num_procs=%s)' % env['XPLOR_NUM_PROCESSES'])"

!
! read in the PSF and initial PDB files
!
parameter @TOPPAR:protein.par end
structure         @PeptideName_generate_psf2.psf
 end

!rgyr
!collapse
```

```
! scale 1.0
!radius of gyration
!ordered region of protein, constant=(#res^0.38)*2.2
!assign (resid 1:8) 50.0 4.8
!end

noe
{====>}
  nres=3000          {*Estimate greater than the actual number of NOEs.*}
  class all
{====>}

  @PeptideName_noe.tbl                    {*Read NOE distance ranges.*}

!   class hb
  @hbond.tbl

end

!restraints dihed
!   reset
!   nass = 5000
!   {====>}
!   @PeptideName_angles.tbl                  {*Read dihedral angle restraints.*}
!end

!couplings
!  nres 400
!  potential harmonic
!  class phi
!  degen 1
!  force 1.0
!  coefficients 6.98 -1.38 1.72 -60.0
!{====>}
!  @jna_coup.tbl   {* regular (non-Glycine) hnha couplings *}
!end

!carbon
!   phistep 180
!   psistep 180
!   nres = 200
!   class all
!   force 0.5
!   potential harmonic
!   @/home/mds8575/PROGRAMS/XPLOR/xplor-nih-2.36/databases/c13shifts/rcoil_c13.tbl
!   @/home/mds8575/PROGRAMS/XPLOR/xplor-nih-
2.36/databases/c13shifts/expected_edited_c13.tbl
!   @carbon.tbl
!end


!
! starting coords
!

coor @PeptideName_only.pdb
```

```
! Fix dipolar coupling tensor axis
!  this applies for the powell minimization- but not for the torsion angle
!  dynamics
!constraints fix (resname "ANI") end

flags exclude * include bond angle impr end
mini powell nstep 1000 nprint 100 end
  {* Minimize with only the covalent constraints. *}

coor copy end

!
! annealing settings
!

eval ($init_t  = 3500.01)
eval ($final_t = 100)

eval ($cool_steps = 25000)
eval ($tempstep = 25)

eval ($ncycle_vdw = 100)

eval ($ncycle = ($init_t-$final_t)/$tempstep)
eval ($nstep = int($cool_steps/$ncycle))

eval ($ini_rad  = 0.4)
eval ($fin_rad  = 0.80)
eval ($radfact = ($fin_rad/$ini_rad)^(1/$ncycle))

eval ($ini_con = 0.004)
eval ($fin_con = 4.0)
eval ($k_vdwfact = ($fin_con/$ini_con)^(1/$ncycle))

eval ($ini_ang = 0.4)
eval ($fin_ang = 1.0)
eval ($ang_fac = ($fin_ang/$ini_ang)^(1/$ncycle))

eval ($ini_imp = 0.4)     ! was 0.1
eval ($fin_imp = 1.0)
eval ($imp_fac = ($fin_imp/$ini_imp)^(1/$ncycle))

eval ($hitemp_noe = 20.0) ! was 150
eval ($ini_noe = 2.0)   ! was 2
eval ($fin_noe = 30.0)
eval ($noe_fac = ($fin_noe/$ini_noe)^(1/$ncycle))

eval ($ini_timestep = 0.010)    !reduced from 0.015
eval ($fin_timestep = 0.003)
eval ($timestep_fac = ($fin_timestep/$ini_timestep)^(1/$ncycle))

!scaling factors for the vdw loop
eval ($radfact_vdw = ($fin_rad/$ini_rad)^(1/$ncycle_vdw))
eval ($k_vdwfact_vdw = ($fin_con/$ini_con)^(1/$ncycle_vdw))
eval ($ang_fac_vdw = ($fin_ang/$ini_ang)^(1/$ncycle_vdw))
eval ($imp_fac_vdw = ($fin_imp/$ini_imp)^(1/$ncycle_vdw))
```

```
vector do (mass = 100.0) (all)
vector do (fbeta = 10.0) (all)

set echo on message on end

!
! set up the torsion angle dynamics topology stuff
! to eliminate degrees of freedom in the aromatics

dynamics internal

  set echo on message on end
  cloop=false
  auto torsion
  maxe 10000
end

constraints
   interaction (not resname ANI) (not resname ANI)
      weights * 1 angl $ini_ang impr $ini_imp  {* Scale covalent constraints. *}
   end
end

parameter                   {*Parameters for the repulsive energy term.*}
   nbonds
      repel=0.5               {*Initial value for effective atom radius  *}
                         {*--modified later.*}
      rexp=2 irexp=2 rcon=1.
      nbxmod=-2               {*Initial value for nbxmod--modified later.*}
      wmin=0.01
      cutnb=4.5 ctonnb=2.99 ctofnb=3.
      tolerance=0.5
   end
end

!restraints dihedral
!     scale=5.                       {*Initial weight--modified later.*}
!end

eval ($count = $firstStruct)
while ($count < $lastStruct) loop structure
    !
    ! Be sure different structures start with different seeds
    !
    eval ($seed = $randomSeed+$count)
    set seed $seed end

                if ( $use_graphics = true ) then
                  ps
                    defi $count
                      bonds (not name h)
                    end
                  end
                end if

      !
      ! reset the force constants and call the energy terms
```

```
!

eval ($bath  = $init_t)
eval ($radius = $ini_rad)
eval ($k_vdw = $ini_con)
eval ($k_ang = $ini_ang)
eval ($k_imp = $ini_imp)
eval ($knoe  = $ini_noe)
eval ($timestep = $ini_timestep)

flags
  exclude * include bonds angl impr noe
end

!
! re-init the coords
!

coor swap end
coor copy end {* Save the first structure to copy to use as a reference *}

!
! set some high-temp force constants
!

noe
 ceiling=1000
 averaging  * sum
 potential  * square
 sqconstant * 1.0
 sqexponent * 2
 sqoffset   * 0.0
 soexponent * 1
 asymptote  * 1.0
 rswitch * 0.5
 rswitch * 3
 scale      * $hitemp_noe

 averaging  hb sum
 potential  hb biharmonic
 sqconstant hb 1.0
 sqexponent hb 2
 scale      hb $hitemp_noe

end


parameters
  nbonds
    atom
    nbxmod 4  {* Can use 4 here, due to internal coordinate dynamics. *}
    wmin 0.01
    cutnb 4.5
    tolerance 0.5
    rexp 2
    irex 2
    repel $radius
```

```
    rcon $k_vdw
  end
end

vector do (vx = maxwell(3500)) (all)
vector do (vy = maxwell(3500)) (all)
vector do (vz = maxwell(3500)) (all)
{* Set initial velocities to fit a temperature of 3500K. *}
{* High temperature to promote convergence.            *}

!
! high temp dynamics
!

evaluate ($tol = $bath/1000)
dynamics internal
  nstep 0
        endt 20
  timestep $timestep
  tbath $bath
        response 20
  nprint 100
        etol $tol
end

         if ( $use_graphics = true ) then
          ps
            append $count
              bonds (not name h)
            end
          end
         end if

flags exclude * include bond angle impr vdw noe end

eval ($bath  = $init_t)

evaluate ($i_vdw = 0)
while ($i_vdw < $ncycle_vdw) loop vdw    {* iterate changes in van der  *}
                                         {* Waals interaction. Scale and*}
                                         {* radius.                  *}
  evaluate ($i_vdw=$i_vdw+1)
  {* Min function is used to keep variables within ini_variable and   *}
  {* fin_variable.                                              *}
  eval ($k_vdw = min($fin_con,$k_vdw*$k_vdwfact_vdw))
  eval ($radius = min($fin_rad,$radius*$radfact_vdw))
  eval ($k_ang = min($fin_ang,$k_ang*$ang_fac_vdw))
  eval ($k_imp = min ($fin_imp,$k_imp*$imp_fac_vdw))

  parameter
    nbonds
      rcon $k_vdw
      repel $radius
    end
  end
  constraints
    interaction (all) (all)
```

```
      weights * 1 angl $k_ang impr $k_imp
    end
  end

  !
  ! vdw dynamics {* dynamics with Van der Waals repulsion turned on. *}
  !

  dynamics internal
    timestep $timestep
          endt 1
    tbath $bath
    nprint $nstep
    ntrfr 10
  end

       if ( $use_graphics = true ) then
          ps
            append $count
                  bonds (not name h)
            end
          end
       end if
end loop vdw

!
! cooling
!
restraints dihedral
   scale=200.   end {* increase dihedral term  *}

evaluate ($i_cool = 0)

while ($i_cool < $ncycle) loop cool
  evaluate ($i_cool=$i_cool+1)

  eval ($bath = $bath  - $tempstep)
  eval ($knoe = $knoe*$noe_fac)    {* Scale during cooling. *}
  eval ($timestep = $timestep*$timestep_fac)
  eval ($k_vdw = min($fin_con,$k_vdw*$k_vdwfact))
  eval ($radius = min($fin_rad,$radius*$radfact))
  eval ($k_ang = min($fin_ang,$k_ang*$ang_fac))
  eval ($k_imp = min($fin_imp,$k_imp*$imp_fac))

  noe scale * $knoe end
  parameter
    nbonds
      rcon $k_vdw
      repel $radius
    end
  end
  constraints
   interaction (not resname ANI) (not resname ANI)
    weights * 1 angl $k_ang impr $k_imp {* Scale impropers and angles *}
   end
  end
```

```
  !
  ! cooling dynamics
  !

  evaluate ($tol = $bath/1000)
  dynamics internal
        endt 2
    tbath $bath
    nprint $nstep
    ntrfr 10
  end

       if ( $use_graphics = true ) then
         ps
           append $count
                bonds (not name h)
           end
         end
       end if
end loop cool

!
! final minimization - with fixed tensor axis
!

! constraints fix (resname "ANI") end
mini powell nstep 1500 nprint 50 end

       if ( $use_graphics = true ) then
        ps
          append $count
            bonds (not name h)
          end
        end
       end if

!
! recenter
!

coor orient end

!
! analysis
! evaluates RMS deviations and violations

print threshold 0.5 noe
eval ($rms_noe = $result)
eval ($violations_noe = $violations)

print thres 0.05 bonds
eval ($rms_bonds = $result)

print thres 5. angles
eval ($rms_angles = $result)

print thres 5. impropers
```

```
    eval ($rms_impropers = $result)

   ! print thres 5. cdih
   ! eval  ($rms_cdih = $result)
   ! eval  ($violations_cdih=$violations)

   ! carbon print threshold = 1.5 end
   ! evaluate ($rms_ashift = $rmsca)
   ! evaluate ($rms_bshift = $rmscb)
   ! evaluate ($viol_shift = $violations)

   ! couplings print threshold 1.0 class phi end
   ! evaluate ($rms_coup = $result)
   ! evaluate ($viols_coup = $violations)

    remarks     overall bonds angles improper vdw noe
    remarks energies: $ener $bond $angl $impr $vdw $noe
    remarks ===============================================================
    remarks          bonds angles impropers noe
    remarks  $rms_bonds $rms_angles $rms_impropers $rms_noe
    remarks ===============================================================
    remarks              noe
    remarks violations :    $violations_noe
    remarks ===============================================================

    eval ($suffix = ".pdb")
    evaluate ($file = "PeptideName_anneal_" + encode($count) + $suffix)
    write coor output= $file end

!     {* Test for convergence. *}
!     if ($violations_noe < 4) then
!       if ($violations_cdih < 2) then
!         eval ($suffix = ".pdb_converged")  {* Change the suffix if converged. *}
!              evaluate ($file = "complex_anneal_" + encode($count) + $suffix)
!              write coor output= $file end
!       end if
!     end if


    eval ($count = $count + 1)
  end loop structure
stop
```

A1.5. xplor_energy.sh

```
#!/bin/csh
set PPATH='/Users/PATH_TO_STRUCTURE_DIRECTORY'

## Sort structures in the order of their energy. ##
grep "REMARK energies:" $PPATH/*pdb >> $PPATH/energy.tmp

sort -k3n energy.tmp >> energy.out

rm energy.tmp
```

```
## Sort structures in the order of violations. ##
grep "REMARK violations.:" $PPATH/*pdb >> $PPATH/energy.tmp

sort -k4 energy.tmp >> energy.out

rm energy.tmp
```

A1.6. prtvio.inp

```
evaluate ($19 = 0.4)
set echo=true mess=on display=prtvio.out print=prtvio.out end
structure @PeptideName_generate_psf2.psf end
noe reset nres=20000
class all @PeptideName_noe.tbl
potential all soft sqconst all 1.0 sqexponent all 2 aver all sum
end
for $2 in
(
"PeptideName_anneal_1.pdb"
"PeptideName_anneal_2.pdb"
"PeptideName_anneal_3.pdb"
"PeptideName_anneal_4.pdb"
"PeptideName_anneal_5.pdb"
"PeptideName_anneal_6.pdb"
"PeptideName_anneal_7.pdb"
"PeptideName_anneal_8.pdb"
"PeptideName_anneal_9.pdb"
"PeptideName_anneal_10.pdb"
)

loop test
coor @@$2
display FILE= $2
print threshold $19 noe
end loop test
stop
```

A2. Restraint Table Format

NOEs are binned into 3 categories: strong (2.5 ± 0.7 Å), medium (3.5 ±1.2 Å) and weak (4.5 ±1.5 Å). Distance restraints for atoms are depicted in a format as following:

```
assign (residue 1 and name HA) (residue 1 and name HN) 2.5 0.7 0.7
assign (residue 2 and name HB) (residue 2 and name HN) 3.5 1.2 1.2
assign (residue 3 and name HG) (residue 3 and name HN) 4.5 1.5 1.5
```

Two atoms are selected and their distance restraint is written after that. The first number is the restrained distance between the two atoms. The second number is the minus range of distance, and the third number is the plus range.
Naming for every atom in each amino acid can be found on the website of Biological Magnetic Resonance Data Bank (BMRB) (https://bmrb.io/ref_info/).

A3. Commands and Steps

1. Generate a dummy cyclic peptide based on sequence.

```
/PATH/xplor <generate_cyclic_psf.inp> cyclic.out
```

2. Generate an arbitrary extended conformation of peptide with ideal geometry as the starting structure for simulated annealing in the next step.

```
/PATH/xplor <generate_template_complx_ani.inp> temp.out
```

3. Do structure calculations using simulated annealing process. Distribute parallel works to multiple CPUs, if exist, using *-smp #*, where # is the number of CPUs.

```
/PATH/xplor -smp # -o anneal.out peptide_anneal.inp
```

4. Find out 10 structures with the lowest energy.

```
sh xplor_energy.sh
```

5. Check if NOE violations occur in the 10 lowest energy structures. If violations occur, re-inspect the spectra and NOE table to find out mistakes or incorrect assignments and fix them.

```
/PATH/xplor <prtvio.inp> prtvio.out
```

6. Repeat step 3-5 until no NOE violation occurs. The 10 structures with the lowest total energy are chosen as the representative structures. Examine the structure quality by MolProbity server.

Appendix B.

Scripts and Commands for RNA Structure Calculation

B1. Scripts

B1.1. generate_aptamer_psf.py

```
# Generate PSF for aptamer RNA from the sequence.

# Aptamer sequence.
# RNA name= RNA_NAME
seq = '''GUA ADE URI CYT GUA ADE URI CYT GUA ADE URI CYT’''

# Make desired nucleic acid topology and parameter files the default.
import protocol
protocol.topology['nucleic'] = 'nucleic-3.1.top'
protocol.parameters['nucleic'] = 'nucleic-3.1.par' # needed because seqToPSF
                                   # below initialzes params.
                                   # as side effect
# Generate PSF file.
import psfGen
psfGen.seqToPSF(seq, seqType='rna', psfFilename='RNA_NAME.psf')
```

B1.2. fold.py

```
#
# Protocol for folding an input extended RNA structure in torsion angle space
# (with optional final Cartesian minimization).
#
# This script comprises the following stages:
#  (1) High temperature (sampling) stage, with small force constants for the
#      different energy terms and limited van der Waals-like repulsions.
#  (2) Simulated annealing stage, where force constants are increased.
#  (3) Minimization in torsion angle space.
#  (4) Minimization in Cartesian space (optional).
#
# The experimental restraints consist of distances (from NOEs and hydrogen
# bonds) and torsion angles (e.g., from J-couplings).  Additionally, planarity
# restraints are implemented to prevent undue buckling across Watson-Crick
# pairs.
#
# The force field covalent energy terms for bond lengh and angles, and improper
# dihedral angles are supplemented with a statistical potential for torsion
# angles, which reduces backbone conformational outliers.  This is the so-called
# RNA-ff1 force field, implemented here as originally described in:
# Bermejo et al., Improving NMR Structures of RNA, Structure (2016) 24: 806-815.
#
# This script is intended to serve as a reference for RNA structure calculation,
# allowing the easy removal/addition of experimental energy terms.
#
# Authors: Guillermo A. Bermejo and Charles D. Schwieters.
#
```

```
#
# Total number of structures.
#
nstructures = 1000

#
# Base name for output PDB files.
# ---
# This string must contain the "STRUCTURE" literal to be replaced by the
# structure number in the PDB filename.  The (optional) "SCRIPT" literal is
# replaced by the name of this file (or stdin if redirected using <).
#
outfilename = 'SCRIPT_STRUCTURE.pdb'

#
# Set random seed.
#
import protocol
protocol.initRandomSeed(3421)   # by specific seed number

#
# Read PSF file(s).
#
protocol.initStruct(['RNA_NAME.psf'])

#
# Load paramaters.
# ---
# Read covalent and nonbonded parameters from parameter file(s).
# Note that only covalent parameters for bond lengths and angles, and impropers
# dihedrals are used in this script.  The torsion angle parameters, if any, are
# ommited because they are provided by a statistical potential below.
#
protocol.initParams(['nucleic-3.1.par'])

#
# Generate extended conformation with satisfied covalent geometry.
#
protocol.genExtendedStructure(sel=0, verbose=0, maxFixupIters=500)

#protocol.loadPDB("cmplx.tmplt.pdb")
#inputStructures="cmplx.tmplt.pdb"

#
# Create a PotList() to contain the energy terms that will be active
# during structure calculations.
#
from potList import PotList
potList = PotList()

#
# Lists highTempParams and rampedParams will hold simulationTools.StaticRamp and
# simulationTools.MultRamp objects to handle parameter changes between the high
# termperature and annealing stage, and within the annealing stage (e.g., ramped
# force constants).
#
from simulationTools import StaticRamp, MultRamp, InitialParams, AnnealIVM
```

```
highTempParams = []
rampedParams = []

# Below, the entire setup of each energy term (including their treatment during
# the high temperature and annealing stages) is performed in self-contained
# sections, so that removal of a term or addition of a new one can be done
# simply by commenting out or adding the corresponding section, respcetively.

#
# Set up distance restraint potential (e.g., from NOEs).

import noePotTools
noe = PotList('noe')
for (name, scale, table) in [('all', 1, 'RNA_noe.tbl'),

                   # add entries for more restraint tables here
                   ]:
   pot = noePotTools.create_NOEPot(name, table)
   # pot.setPotType("soft") # if you suspect there are bad NOEs
   pot.setScale(scale)
   noe.append(pot)
potList.append(noe)
rampedParams.append(MultRamp(2, 30, "noe.setScale( VALUE )"))

#
# Set up torsion angle restraint potential (e.g., from J-couplings).
#
import xplorPot
dihedralTables = ['RNA_dihedralSL.tbl' ]
protocol.initDihedrals(dihedralTables)
potList.append(xplorPot.XplorPot('CDIH'))
highTempParams.append(StaticRamp("potList['CDIH'].setScale(10)"))
rampedParams.append(StaticRamp("potList['CDIH'].setScale(200)"))

#
# Set up potential for base-pair planarity restraints.
#
protocol.initPlanarity('RNA_planeSL.inp')
potList.append(xplorPot.XplorPot('PLAN'))
# (The setup of this term remains unchanged throughout; no need to involve
# highTempParams and/or rampedParams.)

#
# Set up statistical torsion angle potential (torsionDB).
#
import torsionDBPotTools
torsiondb = torsionDBPotTools.create_TorsionDBPot(name='torsiondb',system='rna')
potList.append(torsiondb)
rampedParams.append(MultRamp(0.5, 4, "torsiondb.setScale(VALUE)"))

#
# setup parameters for atom-atom repulsive term. (van der Waals-like term)
#
from repelPotTools import create_RepelPot,initRepel
repel = create_RepelPot('repel')
potList.append(repel)
```

```
rampedParams.append( StaticRamp("initRepel(repel,use14=False)") )
rampedParams.append( MultRamp(.004,4,  "repel.setScale( VALUE)") )
# nonbonded interaction only between CA atoms
highTempParams.append( StaticRamp("""initRepel(repel,
                                  use14=True,
                                  scale=0.004,
                                  repel=1.2,
                                  moveTol=45,
                                  interactingAtoms="name C1'"
                                  )""") )

# Selected 1-4 interactions.
import torsionDBPotTools
repel14 = torsionDBPotTools.create_Terminal14Pot('repel14')
potList.append(repel14)
highTempParams.append(StaticRamp("repel14.setScale(0)"))
rampedParams.append(MultRamp(0.004, 4, "repel14.setScale(VALUE)"))

#
# Set up bond length potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potList.append(xplorPot.XplorPot('BOND'))
# (The setup of this term remains unchanged throughout; no need to involve
# highTempParams and/or rampedParams.)

#
# Set up bond angle potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potList.append(xplorPot.XplorPot('ANGL'))
rampedParams.append(MultRamp(0.4, 1.0, "potList['ANGL'].setScale(VALUE)"))

#
# Set up improper dihedral angle potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potList.append(xplorPot.XplorPot('IMPR'))
rampedParams.append(MultRamp(0.1, 1.0, "potList['IMPR'].setScale(VALUE)"))

#
# Done with energy terms.

#
# Give atoms uniform weights, except for anisotropy axes (if any).
#
protocol.massSetup()

#
# Set up IVM object(s).
#

# IVM object for torsion-angle dynamics/minimization.
import ivm
dyn = ivm.IVM()

# Make ligand (resid 34) rigid.
```

```
#dyn.group('''resid 34 and (name n1 or name c1 or name c2 or name o2
#    or name n3 or name c3 or name c4 or name n9 or name c8
#    or name n7 or name c5 or name c6 or name o6 or name h7)''')

# Argument flexRiboseRing below is a string that selects residues whose ribose
# rings will have all endocyclic angles flexible.  In general, all ribose rings
# should be selected.  In this example, the non-RNA ligand (residue 34) has to
# be excluded.
#protocol.torsionTopology(dyn, flexRiboseRing='resid 1:33 and resid 35:50')

protocol.torsionTopology(dyn, flexRiboseRing='all')

## Optional IVM object for final Cartesian minimization.
minc = ivm.IVM()
protocol.cartesianTopology(minc)

#
# Temperature set up.
#
temp_ini = 3500.0   # initial temperature
temp_fin = 25.0     # final temperature

def calcOneStructure(loopInfo):
  """Calculate a structure.

  """
  # Generate initial structure by randomizing torsion angles.
  import monteCarlo
  monteCarlo.randomizeTorsions(dyn)

  # Randomize position of ligand.
  # (In the case where there is no ligand, comment out the next two lines.)
  #import atomAction
  #atomAction.randomizeDomainPos('resid 1:14', deltaPos=45)

  # Set torsion angles from restraints.
  # (They start satisfied, allowing the shortening of high temp dynamics.)
  #import torsionTools
  #torsionTools.setTorsionsFromTable(dihedralTables)

  # Fix up covalent geometry.
  # (The torsion restraints may include ring torsions and distort geometry.)
  while True:
    try:
      protocol.fixupCovalentGeom(maxIters=100, useVDW=1)
      break
    except protocol.CovalentViolation:
      pass

  #
  # High Temperature Dynamics Stage.
  #

  # Initialize parameters for high temperature dynamics.
  InitialParams(rampedParams)
  InitialParams(highTempParams) # purposedly overides some
                                # setups in rampedParams
```

```
    # Set up IVM object and run.
    protocol.initDynamics(dyn,
                 potList=potList,
                 bathTemp=temp_ini,
                 initVelocities=1,
                 finalTime=15,   # run for finalTime or
                 numSteps=15001, # numSteps * 0.001, whichever is less
                 printInterval=100)

    dyn.setETolerance(temp_ini/100) # used to find stepsize (default: temp/1000)

    dyn.run()

    #
    # Simulated Annealing Stage.
    #

    # Initialize parameters for annealing.
    InitialParams(rampedParams)

    # Set up IVM object for annealing.
    protocol.initDynamics(dyn,
                 potList=potList,
                 finalTime=0.2,  # run for finalTime or
                 numSteps=201,   # numSteps * 0.001, whichever is less
                 printInterval=100)

    # Set up cooling loop and run.
    AnnealIVM(initTemp=temp_ini,
         finalTemp=temp_fin,
         tempStep=12.5,
         ivm=dyn,
         rampedParams=rampedParams).run()

    #
    # Torsion angle minimization.
    #
    protocol.initMinimize(dyn,
                 potList=potList,
                 printInterval=50)
    dyn.run()

##   #
##   # Cartesian minimization (optional).
##   #
    protocol.initMinimize(minc,
                 potList=potList,
                 dEPred=10)
    minc.run()

from simulationTools import StructureLoop
StructureLoop(numStructures=nstructures,
        pdbTemplate=outfilename,
        doWriteStructures=True,
        structLoopAction=calcOneStructure,
        # Arguments for generating structure statistics:
```

```
        genViolationStats=True,
        averageSortPots=[potList['noe']], # terms for structure sorting
                # potList['CDIH']],
        averageTopFraction=0.1,  # top fraction of structs. to report on
        averagePotList=potList,  # terms analyzed
        averageFitSel='not (name H* or resname ANI)', # selection to fit
        ).run()                          # to average struct.
                                 # and report precision
```

B1.3. refine.py

```
# Protocol refining an input folded RNA structure in torsion angle space (with
# optional final Cartesian minimization).
#
# This script comprises the following stages:
#  (1) High temperature (sampling) stage, with small force constants for the
#     different energy terms and limited van der Waals-like repulsions.
#  (2) Simulated annealing stage, where force constants are increased.
#  (3) Minimization in torsion angle space.
#  (4) Minimization in Cartesian space (optional).
#
# The experimental restraints consist of distances (from NOEs and hydrogen
# bonds), torsion angles (e.g., from J-couplings), and RDCs.  Additionally,
# planarity restraints are implemented to prevent undue buckling across Watson-
# Crick pairs.
#
# The force field covalent energy terms for bond lengh and angles, and improper
# dihedral angles are supplemented with statistical potentials for torsion
# angles (which reduces backbone conformational outliers) and a base-base
# positional preferences.  This is the so-called RNA-ff1 force field,
# implemented here as originally described in:
# Bermejo et al., Improving NMR Structures of RNA, Structure (2016) 24: 806-815.
#
# This script is intended to serve as a reference for RNA structure calculation,
# allowing the easy removal/addition of experimental energy terms.
#
# Authors: Guillermo A. Bermejo and Charles D. Schwieters.
#

#
# Total number of structures.
#
nstructures = 400

#
# Base name for output PDB files.
# ---
# This string must contain the "STRUCTURE" literal to be replaced by the
# structure number in the PDB filename.  The (optional) "SCRIPT" literal is
# replaced by the name of this file (or stdin if redirected using <).
#
outfilename = 'SCRIPT_STRUCTURE.pdb'

#
# Set random seed.
#
```

```
import protocol
protocol.initRandomSeed(3421)   # by specific seed number

#
# Read PSF file(s).
#
protocol.initStruct(['RNA_NAME.psf'])

#
# Load paramaters.
# ---
# Read covalent and nonbonded parameters from parameter file(s).
# Note that only covalent parameters for bond lengths and angles, and impropers
# dihedrals are used in this script.  The torsion angle parameters, if any, are
# ommited because they are provided by a statistical potential below.
#
protocol.initParams(['nucleic-3.1.par'])

#
# Read input structure.
#
import glob
infilename = glob.glob('fold_RNA_LOWEST.pdb')[0] # name of top-ranked struct. from fold.py
protocol.initCoords(infilename)

#protocol.loadPDB("fold_cjun_39.sa")

#
# Create a potList.PotList() to contain the energy terms that will be active
# during structure calculations.
#
from potList import PotList
potList = PotList()

#
# Lists highTempParams and rampedParams will hold simulationTools.StaticRamp and
# simulationTools.MultRamp objects to handle parameter changes between the high
# termperature and annealing stage, and within the annealing stage (e.g., ramped
# force constants).
#
from simulationTools import StaticRamp, MultRamp, InitialParams, AnnealIVM

highTempParams = []
rampedParams = []

# Below, the entire setup of each energy term (including their treatment during
# the high temperature and annealing stages) is performed in self-contained
# sections, so that removal of a term or addition of a new one can be done
# simply by commenting out or adding the corresponding section, respcetively.

#
# Set up RDC potential.
#

# List with RDC data.
# ---
# Each entry, a tuple, must contain strings for (in this order): arbitrary name
```

```
# for the alignment medium, arbitrary name for the experiment (e.g., "CH" for
# C-H RDCs), and the path of the corresponding restraint table.
#        medium  exp.  table
#rdcData = [('pf1', 'NH', 'RDC.tbl')
#         ]

# Orientation tensor(s).
#import varTensorTools
#tensors = {}
#for item in rdcData:
#    medium = item[0]
#    tensors[medium] = varTensorTools.create_VarTensor(medium)

# RDCs per se.
#import rdcPotTools
#rdcPotTools.correctGyromagneticSigns() # change sign of NH RDCs after scaling
#rdcs = PotList('rdc')
#for (medium, exp, table) in rdcData:
#    name = '%s_%s' % (exp, medium)
#    rdc = rdcPotTools.create_RDCPot(name, table, tensors[medium])
#    #rdcPotTools.scale_toCH(rdc)
#    rdcs.append(rdc)
#potList.append(rdcs)
#rampedParams.append(MultRamp(0.05,1.0, "rdcs.setScale(VALUE)"))

# Calculate initial tensor orientation and setup tensor calculation during
# simulated annealing.
#for tensor in tensors.keys():
#    highTempParams.append(StaticRamp(
#        "varTensorTools.calcTensor(tensors['%s'])" % medium))

#
# Set up distance restraint potential (e.g., from NOEs).
#
import noePotTools
noe = PotList('noe')
for (name, scale, table) in [('all', 1, 'RNA_noe.tbl')
                       # add entries for more restraint tables here
                       ]:
    pot = noePotTools.create_NOEPot(name, table)
    # pot.setPotType("soft") # if you suspect there are bad NOEs
    pot.setScale(scale)
    noe.append(pot)
potList.append(noe)
rampedParams.append(MultRamp(2, 30, "noe.setScale(VALUE)"))

#
# Set up torsion angle restraint potential (e.g., from J-couplings).
#
import xplorPot
dihedralTables = ['RNA_dihedralSL.tbl']
protocol.initDihedrals(dihedralTables)
potList.append(xplorPot.XplorPot('CDIH'))
highTempParams.append(StaticRamp("potList['CDIH'].setScale(10)"))
rampedParams.append(StaticRamp("potList['CDIH'].setScale(200)"))

#
```

```
# Set up potential for base-pair planarity restraints.
#
protocol.initPlanarity('RNA_planeSL.inp')
potList.append(xplorPot.XplorPot('PLAN'))
# (The setup of this term remains unchanged throughout; no need to involve
# highTempParams and/or rampedParams.)

#
# Set up statistical torsion angle potential (torsionDB).
#
import torsionDBPotTools
torsiondb = torsionDBPotTools.create_TorsionDBPot(name='torsiondb',system='rna')
potList.append(torsiondb)
rampedParams.append(MultRamp(0.5, 4, "torsiondb.setScale(VALUE)"))

#
# setup parameters for atom-atom repulsive term. (van der Waals-like term)
#
from repelPotTools import create_RepelPot,initRepel
repel = create_RepelPot('repel')
potList.append(repel)
rampedParams.append( StaticRamp("initRepel(repel,use14=False)") )
rampedParams.append( MultRamp(.004,4,  "repel.setScale( VALUE)") )
# nonbonded interaction only between CA atoms
highTempParams.append( StaticRamp("""initRepel(repel,
                              use14=True,
                              scale=0.004,
                              repel=1.2,
                              moveTol=45,
                              interactingAtoms="name C1'"
                              )""") )

# Selected 1-4 interactions.
import torsionDBPotTools
repel14 = torsionDBPotTools.create_Terminal14Pot('repel14')
potList.append(repel14)
highTempParams.append(StaticRamp("repel14.setScale(0)"))
rampedParams.append(MultRamp(0.004, 4, "repel14.setScale(VALUE)"))


#
# Set up bond length potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potList.append(xplorPot.XplorPot('BOND'))
# (The setup of this term remains unchanged throughout; no need to involve
# highTempParams and/or rampedParams.)

#
# Set up bond angle potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potList.append(xplorPot.XplorPot('ANGL'))
rampedParams.append(MultRamp(0.4, 1.0, "potList['ANGL'].setScale(VALUE)"))

#
# Set up improper dihedral angle potential.
```

```
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potList.append(xplorPot.XplorPot('IMPR'))
rampedParams.append(MultRamp(0.1, 1.0, "potList['IMPR'].setScale(VALUE)"))

#
# Set up statistical base-base positional potential.
# ---
# The selection argument in initOrie below should select all residues to be
# affected by the potential.
# In this particular example, the small-molecule ligand (residue 34) is ommited,
# along with residue 27, known from preliminary calculations to stick out of the
# structure.  In general, however, all RNA residues should be selected (e.g.,
# for an isolated RNA molecule use selection='all').
# Reference: Clore, GM & Kuszewski, J, (2003) J. Am. Chem. Soc. 125:1518-1525.
#
#protocol.initOrie(system='rna', selection='resid 1:23 or resid 35:56')
protocol.initOrie(system='rna', selection='all')
potList.append(xplorPot.XplorPot('ORIE'))
rampedParams.append(MultRamp(0.002,0.3,"xplor.command('orie scale VALUE end')"))

#
# Done with energy terms.

#
# Give atoms uniform weights, except for anisotropy axes (if any).
#
protocol.massSetup()

#
# Set up IVM object(s).
#

# IVM object for torsion-angle dynamics/minimization.
import ivm
dyn = ivm.IVM()

# Make ligand (resid 34) rigid.
#dyn.group('''resid 34 and (name n1 or name c1 or name c2 or name o2
 #    or name n3 or name c3 or name c4 or name n9 or name c8
 #    or name n7 or name c5 or name c6 or name o6 or name h7)''')

# Orientation tensor setup.
# ---
# If RDCs are not included in the calculations by commenting out their setup
# section above, this set up has to be commented out also.
#for tensor in tensors.values():
#    tensor.setFreedom("fixDa, fixRh")  # fix tensor Rh, Da, vary orientation
#    tensor.setFreedom("varyDa, varyRh")  # vary tensor Rh, Da, vary orientation

# Argument flexRiboseRing below is a string that selects residues whose ribose
# rings will have all endocyclic angles flexible.  In general, all ribose rings
# should be selected.  In this example, the non-RNA ligand (residue 34) has to
# be excluded.
#protocol.torsionTopology(dyn, flexRiboseRing='resid 1:33 and resid 35:56')
protocol.torsionTopology(dyn, flexRiboseRing='all')
```

```
## Optional IVM object for final Cartesian minimization.
minc = ivm.IVM()
##
##for tensor in tensors.values():
##    tensor.setFreedom("varyDa, varyRh") # allow all tensor parameters float here
##
protocol.cartesianTopology(minc)

#
# Temperature set up.
#
temp_ini = 3000.0   # initial temperature
temp_fin = 25.0     # final temperature

def calcOneStructure(loopInfo):
    """Calculate a structure.

    """
    #
    # High Temperature Dynamics Stage.
    #

    # Initialize parameters for high temperature dynamics.
    InitialParams(rampedParams)
    InitialParams(highTempParams) # purposedly overides some
                                  # setups in rampedParams

    # Set up IVM object and run.
    protocol.initDynamics(dyn,
                  potList=potList,
                  bathTemp=temp_ini,
                  initVelocities=1,
                  finalTime=15,   # run for finalTime or
                  numSteps=15001, # numSteps * 0.001, whichever is less
                  printInterval=100)

    dyn.setETolerance(temp_ini/100) # used to find stepsize (default: temp/1000)

    dyn.run()

    #
    # Simulated Annealing Stage.
    #

    # Initialize parameters for annealing.
    InitialParams(rampedParams)

    # Set up IVM object for annealing.
    protocol.initDynamics(dyn,
                  potList=potList,
                  finalTime=0.63,  # run for finalTime or
                  numSteps=631,   # numSteps * 0.001, whichever is less
                  printInterval=100)

    # Set up cooling loop and run.
    AnnealIVM(initTemp=temp_ini,
          finalTemp=temp_fin,
```

```
            tempStep=12.5,
            ivm=dyn,
            rampedParams=rampedParams).run()

    #
    # Torsion angle minimization.
    #
    protocol.initMinimize(dyn,
                  potList=potList,
                  printInterval=50)
    dyn.run()

    #
    # Cartesian minimization (optional).
    #
    protocol.initMinimize(minc,
                  potList=potList,
                  dEPred=10)
    minc.run()

from simulationTools import StructureLoop
StructureLoop(numStructures=nstructures,
            pdbFilesIn=infilename,
            pdbTemplate=outfilename,
            doWriteStructures=True,
            structLoopAction=calcOneStructure,
            # Arguments for generating structure statistics:
            genViolationStats=True,
            averageSortPots=[potList['noe'], # terms for structure sorting
                      potList['CDIH']],
                      #potList['rdc']],
            averageTopFraction=0.1,  # top fraction of structs. to report on
            averagePotList=potList,  # terms analyzed
            averageFitSel='not (name H* or resname ANI)',
            ).run()
```

B2. Restraint Table Format

B2.1. Distance Restraint

NOEs are binned into 3 categories: strong (2.5 ± 0.7 Å), medium (3.5 ±1.2 Å) and weak (4.5 ±1.5 Å). Distance restraints for atoms are depicted in a format as following, which is the same as the format mentioned in Appendix A2:

```
assign (residue 1 and name H5) (residue 1 and name H6) 2.5 0.7 0.7
assign (residue 2 and name H1’) (residue 2 and name H6) 3.5 1.2 1.2
assign (residue 3 and name H1’) (residue 3 and name H8) 4.5 1.5 1.5
```

Two atoms are selected and their distance restraint is written after that. The first number is the restrained distance between the two atoms. The second number is the minus range of distance, and the third number is the plus range.
Naming for every atom in each nucleic acid can be found on the website of Biological Magnetic Resonance Data Bank (BMRB) (https://bmrb.io/ref_info/).

B2.2. Dihedral Restraint

It is known that the sugar puckers of nucleic acids in a A-form helix have 3'-endo conformations. The conformation of sugar pucker can also be confirmed through TOCSY spectra (described in Chapter 1 Section 3.3).

Dihedral restraints are depicted in a format as following:

```
assign (resid 1 and name O4’) (resid 1 and name C1')
        (resid 1 and name C2') (resid 1 and name C3') 1.0 -20.0 10.0 2
```

Four atoms are selected to specify the dihedral angle.
The first number specifies the energy constant in kcal·mole^{-1}·rad^{-2}. The second number specifies the angle in degrees to which the dihedral angle is restrained. The third number specifies the range around the restrained angle. The last number specifies the exponent of the torsion angle energy function in Xplor-NIH (https://nmr.cit.nih.gov/xplor-nih/doc/current/xplor/cdih.html).

An example of dihedral restraints for a 3'-endo sugar pucker in an A-form helix :

```
!     GUA2
assign (resid 2 and name O4’) (resid 2 and name C1')
        (resid 2 and name C2') (resid 2 and name C3') 1.0 -20.0 10.0 2
assign (resid 2 and name C1’) (resid 2 and name C2')
        (resid 2 and name C3’) (resid 2 and name C4') 1.0  35.0 5.0 2
assign (resid 2 and name C5') (resid 2 and name C4')
        (resid 2 and name C3’) (resid 2 and name O3') 1.0 80.0 25.0 2
assign (resid 2 and name O4') (resid 2 and name C1')
        (resid 2 and name N9 ) (resid 2 and name C4 ) 1.0  -150.0 90.0 2
!backbone alpha
assign (resid 1 and name O3') (resid 2 and name P  )
        (resid 2 and name O5') (resid 2 and name C5') 1.0 -60.0 40.0 2
!backbone beta
assign (resid 2 and name P  ) (resid 2 and name O5')
        (resid 2 and name C5') (resid 2 and name C4') 1.0 180.0 50.0 2
!backbone gama
assign (resid 2 and name O5') (resid 2 and name C5')
        (resid 2 and name C4') (resid 2 and name C3') 1.0 60.0 40.0 2
!backbone epsilon
assign (resid 2 and name C4') (resid 2 and name C3')
        (resid 2 and name O3') (resid 3 and name P  ) 1.0 -160.0 50.0 2
!backbone zeta
assign (resid 2 and name C3') (resid 2 and name O3')
        (resid 3 and name P  ) (resid 3 and name O5') 1.0 -70.0 50.0 2
```

An example of dihedral restraints for a 2'-endo sugar pucker conformation:

```
!     GUA1
assign (resid 1 and name O4') (resid 1 and name C1')
        (resid 1 and name C2') (resid 1 and name C3') 1.0 35 10.0 2
assign (resid 1 and name C1') (resid 1 and name C2')
        (resid 1 and name C3’) (resid 1 and name C4') 1.0  -35.0 5.0 2
assign (resid 1 and name C5') (resid 1 and name C4')
        (resid 1 and name C3') (resid 1 and name O3') 1.0 145.0 20.0 2
assign (resid 1 and name O4') (resid 1 and name C1')
```

```
(resid 1 and name N9 ) (resid 1 and name C4 ) 1.0 -150.0 90.0 2
```

B2.3. Planarity Restraint
Planarity restraints are set for base pairs to ensure their hydrogen bonds and bases are coplanar.
An example of planarity restraints for canonical Watson-Crick base pairings:

```
! G6-C36 WC base pair
 group
  select= ((resid 6 and (name N1 or name C6 or name C2)) or (resid 36 and (name N3)))
  weight=80
 end
 group
  select= ((resid 6 and (name N1)) or (resid 36 and (name N3 or name C2 or name C4)))
  weight=80
 end

! A34-U8 WC base pair
group
  select= ((resid 34 and (name N1 or name C2 or name C6)) or  (resid 8 and (name N3)))
  weight=80
 end
 group
  select= ((resid 34 and (name N1)) or (resid 8 and (name N3 or name C2 or name C4)))
  weight=80
 end
```

B3. Commands and Steps

1. Generate the initial RNA aptamer .psf file.

```
/PATH/xplor -py generate_aptamer_psf.py -o aptamer.out
```

2. Do structure calculations using simulated annealing process. Distribute parallel works to multiple CPUs, if exist, using *-smp #*, where # is the number of CPUs.

```
/PATH/xplor -py fold.py -smp # -o fold.out
```

3. Check the output fold_RNA_##.pdb.stats file to find out restraint violations and 10% of all structures with the lowest energy. If violations occur, re-inspect the spectra and restraint tables to find out mistakes or incorrect assignments and fix them.

4. Repeat step 2 and 3 until no violation occurs. Pick the lowest energy structure for a further refinement by an extended simulated annealing calculation.

```
/PATH/xplor -py refine.py -smp # -o refine.out
```

5. Check the output refine_RNA_##.pdb.stats file to make sure no violation occurs and find out the 10 structures with the lowest total energy which are chosen as the representative structures. Examine the structure quality by MolProbity server.

Appendix C.

Resonance Assignments

C1. Resonance Assignments for DenvBS

```
loop_
_Entity_comp_index.ID
_Entity_comp_index.Comp_ID

1 GUA
2 GUA
3 ADE
4 GUA
5 URI
6 GUA
7 GUA
8 URI
9 URI
10 ADE
11 GUA
12 URI
13 CYT
14 URI
15 ADE
16 CYT
17 GUA
18 URI
19 URI
20 URI
21 CYT
22 GUA
23 ADE
24 CYT
25 GUA
26 URI
27 ADE
28 GUA
29 URI
30 URI
31 CYT
32 URI
33 ADE
34 ADE
35 CYT
36 CYT

stop_

loop_
_Atom_chem_shift.ID
_Atom_chem_shift.Comp_index_ID
_Atom_chem_shift.Seq_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	5	5	URI	H1'	H	1	5.742
2	5	5	URI	H2'	H	1	4.375
3	5	5	URI	H5	H	1	5.626
4	5	5	URI	H6	H	1	7.762
5	6	6	GUA	H1'	H	1	5.749
6	6	6	GUA	H2'	H	1	4.912
7	6	6	GUA	H8	H	1	8.022
8	7	7	GUA	H1'	H	1	5.829
9	7	7	GUA	H2'	H	1	4.571
10	7	7	GUA	H8	H	1	7.523
11	8	8	URI	H1'	H	1	5.548
12	8	8	URI	H2'	H	1	4.495
13	8	8	URI	H5	H	1	5.100
14	8	8	URI	H6	H	1	7.744
15	9	9	URI	H1'	H	1	5.579
16	9	9	URI	H2'	H	1	4.467
17	9	9	URI	H5	H	1	5.599
18	9	9	URI	H6	H	1	7.926
19	10	10	ADE	H1'	H	1	5.968
20	10	10	ADE	H2'	H	1	4.680
21	10	10	ADE	H8	H	1	8.125
22	11	11	GUA	H1'	H	1	5.410
23	11	11	GUA	H2'	H	1	4.485
24	11	11	GUA	H8	H	1	6.980
25	12	12	URI	H1'	H	1	5.344
26	12	12	URI	H2'	H	1	4.263
27	12	12	URI	H3'	H	1	4.378
28	12	12	URI	H4'	H	1	4.483
29	12	12	URI	H5	H	1	5.191
30	12	12	URI	H6	H	1	7.503
31	13	13	CYT	H1'	H	1	5.461
32	13	13	CYT	H2'	H	1	4.264
33	13	13	CYT	H5	H	1	5.635
34	13	13	CYT	H6	H	1	7.991
35	14	14	URI	H1'	H	1	5.509
36	14	14	URI	H2'	H	1	4.537
37	14	14	URI	H5	H	1	5.546
38	14	14	URI	H6	H	1	8.003
39	15	15	ADE	H1'	H	1	6.031
40	15	15	ADE	H2'	H	1	4.466
41	15	15	ADE	H8	H	1	8.189
42	16	16	CYT	H1'	H	1	5.388
43	16	16	CYT	H2'	H	1	4.358
44	16	16	CYT	H5	H	1	5.183
45	16	16	CYT	H6	H	1	7.482
46	17	17	GUA	H1'	H	1	5.699
47	17	17	GUA	H2'	H	1	4.477
48	17	17	GUA	H8	H	1	7.475
49	18	18	URI	H1'	H	1	5.561
50	18	18	URI	H2'	H	1	4.512
51	18	18	URI	H3'	H	1	4.316

52	18	18	URI	H5	H	1	5.033
53	18	18	URI	H6	H	1	7.556
54	19	19	URI	H1'	H	1	5.376
55	19	19	URI	H2'	H	1	3.781
56	19	19	URI	H3'	H	1	4.511
57	19	19	URI	H4'	H	1	4.405
58	19	19	URI	H5	H	1	5.785
59	19	19	URI	H5'	H	1	4.484
60	19	19	URI	H6	H	1	7.770
61	20	20	URI	H1'	H	1	6.086
62	20	20	URI	H2'	H	1	4.681
63	20	20	URI	H3'	H	1	4.016
64	20	20	URI	H4'	H	1	4.475
65	20	20	URI	H5	H	1	5.865
66	20	20	URI	H6	H	1	8.015
67	21	21	CYT	H1'	H	1	5.904
68	21	21	CYT	H2'	H	1	4.082
69	21	21	CYT	H3'	H	1	4.379
70	21	21	CYT	H4'	H	1	3.724
71	21	21	CYT	H5	H	1	6.124
72	21	21	CYT	H6	H	1	7.667
73	22	22	GUA	H1'	H	1	5.939
74	22	22	GUA	H2'	H	1	4.862
75	22	22	GUA	H3'	H	1	5.611
76	22	22	GUA	H4'	H	1	4.399
77	22	22	GUA	H8	H	1	7.848
78	23	23	ADE	H2'	H	1	4.510
79	23	23	ADE	H8	H	1	8.561
80	24	24	CYT	H1'	H	1	5.411
81	24	24	CYT	H2'	H	1	4.497
82	24	24	CYT	H5	H	1	5.187
83	24	24	CYT	H6	H	1	7.350
84	25	25	GUA	H1'	H	1	5.731
85	25	25	GUA	H2'	H	1	4.516
86	25	25	GUA	H8	H	1	7.520
87	26	26	URI	H1'	H	1	5.478
88	26	26	URI	H2'	H	1	4.608
89	26	26	URI	H5	H	1	5.107
90	26	26	URI	H6	H	1	7.689
91	27	27	ADE	H1'	H	1	5.907
92	27	27	ADE	H2'	H	1	4.534
93	27	27	ADE	H8	H	1	8.019
94	28	28	GUA	H1'	H	1	5.692
95	28	28	GUA	H2'	H	1	4.279
96	28	28	GUA	H8	H	1	7.288
97	29	29	URI	H1'	H	1	5.903
98	29	29	URI	H2'	H	1	4.381
99	29	29	URI	H5	H	1	5.843
100	29	29	URI	H6	H	1	7.863
101	30	30	URI	H1'	H	1	5.605
102	30	30	URI	H2'	H	1	4.519
103	30	30	URI	H3'	H	1	4.449
104	30	30	URI	H5	H	1	5.302
105	30	30	URI	H6	H	1	7.851
106	31	31	CYT	H1'	H	1	5.668
107	31	31	CYT	H2'	H	1	4.424
108	31	31	CYT	H5	H	1	5.857

109	31	31	CYT	H6	H	1	7.973
110	32	32	URI	H1'	H	1	5.523
111	32	32	URI	H2'	H	1	4.633
112	32	32	URI	H5	H	1	5.514
113	32	32	URI	H6	H	1	7.806
114	33	33	ADE	H1'	H	1	5.929
115	33	33	ADE	H2'	H	1	4.532
116	33	33	ADE	H8	H	1	8.209
117	34	34	ADE	H1'	H	1	5.863
118	34	34	ADE	H2'	H	1	4.430
119	34	34	ADE	H8	H	1	7.897
120	35	35	CYT	H1'	H	1	5.337
121	35	35	CYT	H2'	H	1	4.060
122	35	35	CYT	H5	H	1	5.139
123	35	35	CYT	H6	H	1	7.436
124	36	36	CYT	H1'	H	1	5.763
125	36	36	CYT	H5	H	1	5.445
126	36	36	CYT	H6	H	1	7.619
127	8	8	URI	H3	H	1	13.51
128	9	9	URI	H3	H	1	14.22
129	10	10	ADE	H2	H	1	6.683
130	11	11	GUA	H1	H	1	13.39
131	13	13	CYT	H41	H	1	8.208
132	13	13	CYT	H42	H	1	6.980
133	14	14	URI	H3	H	1	14.14
134	15	15	ADE	H2	H	1	6.252
135	16	16	CYT	H41	H	1	8.294
136	16	16	CYT	H42	H	1	6.713
137	17	17	GUA	H1	H	1	12.70
138	18	18	URI	H3	H	1	13.33
139	23	23	ADE	H2	H	1	6.958
140	24	24	CYT	H41	H	1	8.088
141	24	24	CYT	H42	H	1	6.706
142	25	25	GUA	H1	H	1	12.82
143	26	26	URI	H3	H	1	12.96
144	27	27	ADE	H2	H	1	7.503
145	28	28	GUA	H1	H	1	13.18
146	31	31	CYT	H41	H	1	8.180
147	31	31	CYT	H42	H	1	7.193
148	32	32	URI	H3	H	1	12.80
149	33	33	ADE	H2	H	1	7.738
150	34	34	ADE	H2	H	1	6.680

stop_

C2. Resonance Assignments for DenvTSL

```
 loop_
_Entity_comp_index.ID
_Entity_comp_index.Comp_ID

 1 GUA
 2 GUA
 3 CYT
 4 CYT
 5 GUA
 6 ADE
 7 CYT
 8 ADE
 9 ADE
10 GUA
11 ADE
12 ADE
13 CYT
14 ADE
15 GUA
16 URI
17 URI
18 URI
19 CYT
20 GUA
21 ADE
22 ADE
23 URI
24 CYT
25 GUA
26 GUA
27 CYT
28 CYT

 stop_

 loop_
_Atom_chem_shift.ID
_Atom_chem_shift.Comp_index_ID
_Atom_chem_shift.Seq_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	1	GUA	H1'	H	1	6.132
2	1	1	GUA	H2'	H	1	5.309
3	1	1	GUA	H8	H	1	6.364
4	2	2	GUA	H1'	H	1	5.688
5	2	2	GUA	H2'	H	1	4.450
6	2	2	GUA	H3'	H	1	4.416
7	2	2	GUA	H8	H	1	7.256
8	3	3	CYT	H1'	H	1	5.476
9	3	3	CYT	H2'	H	1	4.240
```

10	3	3	CYT	H3'	H	1	4.389
11	3	3	CYT	H5	H	1	5.132
12	3	3	CYT	H6	H	1	7.578
13	4	4	CYT	H1'	H	1	5.486
14	4	4	CYT	H2'	H	1	4.592
15	4	4	CYT	H5	H	1	5.466
16	4	4	CYT	H6	H	1	7.632
17	5	5	GUA	H1'	H	1	5.613
18	5	5	GUA	H2'	H	1	4.641
19	5	5	GUA	H3'	H	1	4.449
20	5	5	GUA	H8	H	1	7.441
21	6	6	ADE	H1'	H	1	5.793
22	6	6	ADE	H2'	H	1	4.628
23	6	6	ADE	H3'	H	1	4.422
24	6	6	ADE	H8	H	1	7.499
25	7	7	CYT	H1'	H	1	5.312
26	7	7	CYT	H2'	H	1	4.883
27	7	7	CYT	H5	H	1	5.264
28	7	7	CYT	H6	H	1	7.392
29	8	8	ADE	H1'	H	1	5.769
30	8	8	ADE	H2'	H	1	4.885
31	8	8	ADE	H3'	H	1	4.768
32	8	8	ADE	H8	H	1	7.752
33	9	9	ADE	H1'	H	1	5.761
34	9	9	ADE	H2'	H	1	4.885
35	9	9	ADE	H8	H	1	8.072
36	10	10	GUA	H1'	H	1	5.763
37	10	10	GUA	H2'	H	1	4.885
38	10	10	GUA	H3'	H	1	4.698
39	10	10	GUA	H8	H	1	7.770
40	11	11	ADE	H1'	H	1	5.761
41	11	11	ADE	H2'	H	1	4.488
42	11	11	ADE	H3'	H	1	4.698
43	11	11	ADE	H4'	H	1	4.189
44	11	11	ADE	H8	H	1	8.007
45	12	12	ADE	H1'	H	1	5.759
46	12	12	ADE	H2'	H	1	4.795
47	12	12	ADE	H8	H	1	7.716
48	13	13	CYT	H1'	H	1	5.973
49	13	13	CYT	H2'	H	1	4.489
50	13	13	CYT	H3'	H	1	4.400
51	13	13	CYT	H4'	H	1	4.190
52	13	13	CYT	H5	H	1	5.619
53	13	13	CYT	H5'	H	1	3.944
54	13	13	CYT	H6	H	1	7.767
55	14	14	ADE	H1'	H	1	5.666
56	14	14	ADE	H2	H	1	7.751
57	14	14	ADE	H2'	H	1	4.490
58	14	14	ADE	H3'	H	1	4.763
59	14	14	ADE	H4'	H	1	4.626
60	14	14	ADE	H8	H	1	8.256
61	15	15	GUA	H1'	H	1	3.768
62	15	15	GUA	H2'	H	1	4.796
63	15	15	GUA	H3'	H	1	4.225
64	15	15	GUA	H4'	H	1	4.261
65	15	15	GUA	H8	H	1	7.332
66	16	16	URI	H1'	H	1	5.164

67	16	16	URI	H2'	H	1	4.175
68	16	16	URI	H3'	H	1	4.311
69	16	16	URI	H4'	H	1	4.108
70	16	16	URI	H5	H	1	5.638
71	16	16	URI	H6	H	1	7.911
72	17	17	URI	H1'	H	1	5.467
73	17	17	URI	H2'	H	1	4.411
74	17	17	URI	H5	H	1	5.466
75	17	17	URI	H6	H	1	7.684
76	18	18	URI	H1'	H	1	5.313
77	18	18	URI	H2'	H	1	4.535
78	18	18	URI	H3'	H	1	4.424
79	18	18	URI	H5	H	1	5.365
80	18	18	URI	H6	H	1	7.468
81	19	19	CYT	H1'	H	1	5.425
82	19	19	CYT	H2'	H	1	4.293
83	19	19	CYT	H3'	H	1	4.435
84	19	19	CYT	H5	H	1	5.752
85	19	19	CYT	H6	H	1	7.917
86	20	20	GUA	H1'	H	1	5.768
87	20	20	GUA	H2'	H	1	4.836
88	20	20	GUA	H8	H	1	7.751
89	21	21	ADE	H1'	H	1	5.598
90	21	21	ADE	H2'	H	1	4.590
91	21	21	ADE	H8	H	1	7.894
92	22	22	ADE	H1'	H	1	5.688
93	22	22	ADE	H2'	H	1	4.414
94	22	22	ADE	H8	H	1	7.256
95	23	23	URI	H1'	H	1	5.873
96	23	23	URI	H2'	H	1	4.541
97	23	23	URI	H3'	H	1	4.514
98	23	23	URI	H5	H	1	5.791
99	23	23	URI	H6	H	1	7.840
100	24	24	CYT	H1'	H	1	5.510
101	24	24	CYT	H2'	H	1	4.362
102	24	24	CYT	H3'	H	1	4.396
103	24	24	CYT	H5	H	1	5.212
104	24	24	CYT	H6	H	1	7.630
105	25	25	GUA	H1'	H	1	5.896
106	25	25	GUA	H2'	H	1	4.428
107	25	25	GUA	H8	H	1	7.650
108	26	26	GUA	H1'	H	1	5.500
109	26	26	GUA	H2'	H	1	4.360
110	26	26	GUA	H8	H	1	7.537
111	27	27	CYT	H1'	H	1	5.733
112	27	27	CYT	H2'	H	1	3.965
113	27	27	CYT	H5	H	1	5.497
114	27	27	CYT	H6	H	1	7.631
115	28	28	CYT	H1'	H	1	5.733
116	28	28	CYT	H2'	H	1	3.966
117	28	28	CYT	H5	H	1	5.497
118	28	28	CYT	H6	H	1	7.630
119	2	2	GUA	H1	H	1	13.18
120	3	3	CYT	H41	H	1	8.528
121	3	3	CYT	H42	H	1	6.843
122	4	4	CYT	H41	H	1	8.235
123	4	4	CYT	H42	H	1	6.653

124	5	5	GUA	H1	H	1	11.72
125	6	6	ADE	H2	H	1	7.506
126	10	10	GUA	H1	H	1	12.08
127	11	11	ADE	H2	H	1	7.047
128	18	18	URI	H3	H	1	13.56
129	19	19	CYT	H41	H	1	7.791
130	19	19	CYT	H42	H	1	6.372
131	23	23	URI	H3	H	1	14.13
132	24	24	CYT	H41	H	1	8.034
133	24	24	CYT	H42	H	1	6.719
134	25	25	GUA	H1	H	1	12.10
135	26	26	GUA	H1	H	1	13.34
136	27	27	CYT	H41	H	1	8.509
137	27	27	CYT	H42	H	1	6.800

stop_

C3. Resonance Assignments for DenvSLAsh

```
loop_
   _Entity_comp_index.ID
   _Entity_comp_index.Comp_ID

     1 GUA
     2 GUA
     3 ADE
     4 CYT
     5 GUA
     6 URI
     7 GUA
     8 GUA
     9 ADE
    10 CYT
    11 CYT
    12 GUA
    13 ADE
    14 CYT
    15 URI
    16 URI
    17 CYT
    18 GUA
    19 GUA
    20 URI
    21 CYT
    22 GUA
    23 GUA
    24 ADE
    25 ADE
    26 GUA
    27 GUA
    28 ADE
    29 ADE
    30 ADE
    31 CYT
    32 URI
    33 URI
    34 ADE
    35 ADE
    36 CYT
    37 GUA
    38 URI
    39 CYT
    40 CYT

stop_

loop_
   _Atom_chem_shift.ID
   _Atom_chem_shift.Comp_index_ID
   _Atom_chem_shift.Seq_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	1	GUA	H1'	H	1	5.696
2	1	1	GUA	H2'	H	1	4.825
3	1	1	GUA	H3'	H	1	4.604
4	1	1	GUA	H4'	H	1	4.464
5	1	1	GUA	H5'	H	1	4.155
6	1	1	GUA	H5''	H	1	4.269
7	1	1	GUA	H8	H	1	8.026
8	2	2	GUA	H1	H	1	12.285
9	2	2	GUA	H1'	H	1	5.774
10	2	2	GUA	H2'	H	1	4.558
11	2	2	GUA	H3'	H	1	4.503
12	2	2	GUA	H5'	H	1	4.236
13	2	2	GUA	H8	H	1	7.479
14	3	3	ADE	H1'	H	1	5.890
15	3	3	ADE	H2	H	1	7.523
16	3	3	ADE	H2'	H	1	4.478
17	3	3	ADE	H5'	H	1	4.550
18	3	3	ADE	H8	H	1	7.769
19	4	4	CYT	H1'	H	1	5.247
20	4	4	CYT	H2'	H	1	4.321
21	4	4	CYT	H3'	H	1	4.292
22	4	4	CYT	H41	H	1	8.043
23	4	4	CYT	H42	H	1	6.598
24	4	4	CYT	H5	H	1	5.090
25	4	4	CYT	H6	H	1	7.203
26	5	5	GUA	H1	H	1	12.630
27	5	5	GUA	H1'	H	1	5.624
28	5	5	GUA	H2'	H	1	4.324
29	5	5	GUA	H8	H	1	7.393
30	6	6	URI	H1'	H	1	5.518
31	6	6	URI	H2'	H	1	4.309
32	6	6	URI	H5	H	1	4.941
33	6	6	URI	H6	H	1	7.393
34	7	7	GUA	H1'	H	1	5.503
35	7	7	GUA	H2'	H	1	4.367
36	7	7	GUA	H3'	H	1	4.554
37	7	7	GUA	H8	H	1	7.603
38	8	8	GUA	H1'	H	1	5.364
39	8	8	GUA	H2'	H	1	4.397
40	8	8	GUA	H3'	H	1	4.523
41	8	8	GUA	H4'	H	1	4.321
42	8	8	GUA	H8	H	1	7.560
43	9	9	ADE	H1'	H	1	5.872
44	9	9	ADE	H2'	H	1	4.598
45	9	9	ADE	H3'	H	1	4.730
46	9	9	ADE	H4'	H	1	4.469
47	9	9	ADE	H8	H	1	8.074
48	10	10	CYT	H1'	H	1	5.297
49	10	10	CYT	H2'	H	1	4.287
50	10	10	CYT	H3'	H	1	4.333
51	10	10	CYT	H4'	H	1	4.418
52	10	10	CYT	H41	H	1	8.019
53	10	10	CYT	H42	H	1	6.612
54	10	10	CYT	H5	H	1	5.658

55	10	10	CYT	H6	H	1	7.723
56	11	11	CYT	H1'	H	1	5.359
57	11	11	CYT	H2'	H	1	4.448
58	11	11	CYT	H3'	H	1	4.377
59	11	11	CYT	H41	H	1	7.983
60	11	11	CYT	H42	H	1	6.467
61	11	11	CYT	H5	H	1	5.359
62	11	11	CYT	H6	H	1	7.639
63	12	12	GUA	H1	H	1	11.722
64	12	12	GUA	H1'	H	1	5.550
65	12	12	GUA	H2'	H	1	4.457
66	12	12	GUA	H5'	H	1	4.043
67	12	12	GUA	H8	H	1	7.428
68	13	13	ADE	H1'	H	1	5.850
69	13	13	ADE	H2	H	1	7.503
70	13	13	ADE	H2'	H	1	4.469
71	13	13	ADE	H3'	H	1	4.407
72	13	13	ADE	H5'	H	1	4.549
73	13	13	ADE	H8	H	1	7.799
74	14	14	CYT	H1'	H	1	5.214
75	14	14	CYT	H2'	H	1	4.240
76	14	14	CYT	H3'	H	1	4.289
77	14	14	CYT	H4'	H	1	4.052
78	14	14	CYT	H41	H	1	8.242
79	14	14	CYT	H42	H	1	6.731
80	14	14	CYT	H5	H	1	4.989
81	14	14	CYT	H6	H	1	7.102
82	15	15	URI	H1'	H	1	5.533
83	15	15	URI	H2'	H	1	3.652
84	15	15	URI	H5	H	1	5.497
85	15	15	URI	H6	H	1	7.623
86	16	16	URI	H1'	H	1	5.977
87	16	16	URI	H2'	H	1	4.542
88	16	16	URI	H3'	H	1	3.932
89	16	16	URI	H4'	H	1	4.360
90	16	16	URI	H5	H	1	5.723
91	16	16	URI	H6	H	1	7.904
92	17	17	CYT	H1'	H	1	5.830
93	17	17	CYT	H2'	H	1	3.977
94	17	17	CYT	H3'	H	1	4.371
95	17	17	CYT	H4'	H	1	3.678
96	17	17	CYT	H5	H	1	5.973
97	17	17	CYT	H6	H	1	7.577
98	18	18	GUA	H1	H	1	9.636
99	18	18	GUA	H1'	H	1	5.852
100	18	18	GUA	H2'	H	1	4.700
101	18	18	GUA	H21	H	1	7.759
102	18	18	GUA	H22	H	1	6.396
103	18	18	GUA	H3'	H	1	5.516
104	18	18	GUA	H4'	H	1	4.329
105	18	18	GUA	H8	H	1	7.752
106	19	19	GUA	H1	H	1	13.213
107	19	19	GUA	H2'	H	1	4.310
108	19	19	GUA	H3'	H	1	4.173
109	19	19	GUA	H5'	H	1	4.413
110	19	19	GUA	H5''	H	1	4.305
111	19	19	GUA	H8	H	1	8.216

112	20	20	URI	H1'	H	1	5.429
113	20	20	URI	H2'	H	1	4.425
114	20	20	URI	H3	H	1	14.099
115	20	20	URI	H5	H	1	4.950
116	20	20	URI	H6	H	1	7.687
117	21	21	CYT	H1'	H	1	5.467
118	21	21	CYT	H2'	H	1	4.455
119	21	21	CYT	H41	H	1	8.067
120	21	21	CYT	H42	H	1	6.548
121	21	21	CYT	H5	H	1	5.507
122	21	21	CYT	H5'	H	1	4.148
123	21	21	CYT	H6	H	1	7.678
124	22	22	GUA	H1	H	1	11.939
125	22	22	GUA	H1'	H	1	5.523
126	22	22	GUA	H2'	H	1	4.508
127	22	22	GUA	H3'	H	1	4.372
128	22	22	GUA	H4'	H	1	4.017
129	22	22	GUA	H8	H	1	7.384
130	23	23	GUA	H1	H	1	12.278
131	23	23	GUA	H1'	H	1	5.555
132	23	23	GUA	H2'	H	1	4.474
133	23	23	GUA	H3'	H	1	4.338
134	23	23	GUA	H5'	H	1	3.965
135	23	23	GUA	H8	H	1	6.973
136	24	24	ADE	H1'	H	1	5.692
137	24	24	ADE	H2	H	1	7.166
138	24	24	ADE	H2'	H	1	4.450
139	24	24	ADE	H8	H	1	7.610
140	25	25	ADE	H1'	H	1	5.544
141	25	25	ADE	H2	H	1	6.791
142	25	25	ADE	H2'	H	1	4.179
143	25	25	ADE	H8	H	1	7.662
144	26	26	GUA	H1	H	1	12.239
145	27	27	GUA	H1'	H	1	5.617
146	27	27	GUA	H2'	H	1	4.596
147	27	27	GUA	H8	H	1	8.108
148	28	28	ADE	H1'	H	1	5.468
149	28	28	ADE	H2'	H	1	4.364
150	28	28	ADE	H5'	H	1	4.505
151	28	28	ADE	H8	H	1	7.880
152	29	29	ADE	H1'	H	1	5.779
153	29	29	ADE	H2'	H	1	4.380
154	29	29	ADE	H8	H	1	7.965
155	30	30	ADE	H1'	H	1	5.493
156	30	30	ADE	H2'	H	1	4.134
157	30	30	ADE	H4'	H	1	4.337
158	30	30	ADE	H8	H	1	7.793
159	31	31	CYT	H1'	H	1	5.618
160	31	31	CYT	H2'	H	1	4.132
161	31	31	CYT	H3'	H	1	4.056
162	31	31	CYT	H4'	H	1	3.902
163	31	31	CYT	H41	H	1	8.012
164	31	31	CYT	H42	H	1	6.737
165	31	31	CYT	H5	H	1	5.391
166	31	31	CYT	H6	H	1	7.559
167	32	32	URI	H1'	H	1	5.581
168	32	32	URI	H2'	H	1	4.137

169	32	32	URI	H3	H	1	13.432
170	32	32	URI	H5	H	1	5.444
171	32	32	URI	H6	H	1	7.635
172	33	33	URI	H1'	H	1	5.768
173	33	33	URI	H2'	H	1	4.236
174	33	33	URI	H3	H	1	13.644
175	33	33	URI	H5	H	1	5.587
176	33	33	URI	H6	H	1	7.656
177	34	34	ADE	H1'	H	1	5.827
178	34	34	ADE	H2'	H	1	4.725
179	34	34	ADE	H3'	H	1	4.533
180	34	34	ADE	H8	H	1	7.846
181	35	35	ADE	H2'	H	1	4.459
182	35	35	ADE	H8	H	1	7.972
183	36	36	CYT	H1'	H	1	5.323
184	36	36	CYT	H2'	H	1	4.384
185	36	36	CYT	H3'	H	1	4.290
186	36	36	CYT	H41	H	1	8.050
187	36	36	CYT	H42	H	1	6.517
188	36	36	CYT	H5	H	1	5.090
189	36	36	CYT	H6	H	1	7.271
190	37	37	GUA	H1	H	1	12.738
191	37	37	GUA	H1'	H	1	5.619
192	37	37	GUA	H2'	H	1	4.392
193	37	37	GUA	H5'	H	1	4.326
194	37	37	GUA	H5''	H	1	4.013
195	37	37	GUA	H8	H	1	7.458
196	38	38	URI	H1'	H	1	5.451
197	38	38	URI	H2'	H	1	4.386
198	38	38	URI	H3	H	1	14.092
199	38	38	URI	H5	H	1	4.958
200	38	38	URI	H6	H	1	7.709
201	39	39	CYT	H1'	H	1	5.535
202	39	39	CYT	H2'	H	1	4.438
203	39	39	CYT	H41	H	1	8.241
204	39	39	CYT	H42	H	1	6.727
205	39	39	CYT	H5	H	1	5.506
206	39	39	CYT	H6	H	1	7.782

stop_

C4. Resonance Assignments for DenvSLATL

```
loop_
  _Entity_comp_index.ID
  _Entity_comp_index.Comp_ID

   1 GUA
   2 GUA
   3 ADE
   4 GUA
   5 URI
   6 GUA
   7 GUA
   8 URI
   9 URI
  10 ADE
  11 GUA
  12 URI
  13 CYT
  14 URI
  15 ADE
  16 CYT
  17 GUA
  18 URI
  19 CYT
  20 URI
  21 URI
  22 CYT
  23 CYT
  24 GUA
  25 ADE
  26 CYT
  27 ADE
  28 ADE
  29 GUA
  30 ADE
  31 CYT
  32 URI
  33 URI
  34 CYT
  35 GUA
  36 GUA
  37 URI
  38 CYT
  39 GUA
  40 URI
  41 CYT
  42 URI
  43 CYT
  44 GUA
  45 GUA
  46 ADE
  47 GUA
  48 GUA
  49 ADE
  50 GUA
```

51 URI
52 URI
53 CYT
54 URI
55 URI
56 ADE
57 ADE
58 CYT
59 GUA
60 URI
61 ADE
62 GUA
63 URI
64 URI
65 CYT
66 URI
67 ADE
68 ADE
69 CYT
70 CYT

stop_

loop_
_Atom_chem_shift.ID
_Atom_chem_shift.Comp_index_ID
_Atom_chem_shift.Seq_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	7	7	GUA	H1	H	1	13.370
2	7	7	GUA	H1’	H	1	5.715
3	7	7	GUA	H2’	H	1	4.413
4	7	7	GUA	H8	H	1	7.690
5	8	8	URI	H1’	H	1	5.449
6	8	8	URI	H2’	H	1	4.410
7	8	8	URI	H3	H	1	13.485
8	8	8	URI	H6	H	1	7.661
9	9	9	URI	H1’	H	1	5.512
10	9	9	URI	H2’	H	1	4.388
11	9	9	URI	H3	H	1	14.025
12	9	9	URI	H6	H	1	7.855
13	10	10	ADE	H1’	H	1	5.899
14	10	10	ADE	H2	H	1	6.746
15	10	10	ADE	H2’	H	1	4.600
16	10	10	ADE	H8	H	1	8.066
17	11	11	GUA	H1	H	1	13.396
18	11	11	GUA	H1’	H	1	5.340
19	11	11	GUA	H2’	H	1	4.403
20	11	11	GUA	H8	H	1	6.910
21	12	12	URI	H1’	H	1	5.263
22	12	12	URI	H2’	H	1	4.175
23	12	12	URI	H6	H	1	7.436
24	13	13	CYT	H1’	H	1	5.380

25	13	13	CYT	H2’	H	1	4.174
26	13	13	CYT	H41	H	1	8.252
27	13	13	CYT	H42	H	1	6.800
28	13	13	CYT	H6	H	1	7.927
29	14	14	URI	H1’	H	1	5.427
30	14	14	URI	H2’	H	1	4.467
31	14	14	URI	H3	H	1	14.142
32	14	14	URI	H6	H	1	7.931
33	15	15	ADE	H1’	H	1	5.951
34	15	15	ADE	H2	H	1	6.320
35	15	15	ADE	H2’	H	1	4.418
36	15	15	ADE	H8	H	1	8.104
37	16	16	CYT	H1’	H	1	5.404
38	16	16	CYT	H2’	H	1	4.373
39	16	16	CYT	H41	H	1	8.225
40	16	16	CYT	H42	H	1	6.977
41	16	16	CYT	H6	H	1	7.376
42	17	17	GUA	H1	H	1	12.819
43	17	17	GUA	H1’	H	1	5.562
44	17	17	GUA	H2’	H	1	4.307
45	17	17	GUA	H8	H	1	7.377
46	18	18	URI	H3	H	1	13.305
47	18	18	URI	H6	H	1	7.442
48	22	22	CYT	H1’	H	1	5.394
49	22	22	CYT	H2’	H	1	4.415
50	22	22	CYT	H41	H	1	7.946
51	22	22	CYT	H42	H	1	6.400
52	22	22	CYT	H6	H	1	7.580
53	23	23	CYT	H1’	H	1	5.563
54	23	23	CYT	H2’	H	1	4.578
55	23	23	CYT	H41	H	1	8.101
56	23	23	CYT	H42	H	1	6.609
57	23	23	CYT	H6	H	1	7.621
58	24	24	GUA	H1	H	1	11.703
59	24	24	GUA	H1’	H	1	5.617
60	24	24	GUA	H2’	H	1	4.601
61	24	24	GUA	H8	H	1	7.445
62	25	25	ADE	H1’	H	1	5.609
63	25	25	ADE	H2’	H	1	4.067
64	25	25	ADE	H8	H	1	7.919
65	26	26	CYT	H1’	H	1	5.255
66	26	26	CYT	H2’	H	1	3.990
67	26	26	CYT	H6	H	1	7.519
68	27	27	ADE	H1’	H	1	5.668
69	27	27	ADE	H2’	H	1	4.442
70	27	27	ADE	H8	H	1	8.214
71	28	28	ADE	H1’	H	1	5.739
72	28	28	ADE	H2’	H	1	4.764
73	28	28	ADE	H8	H	1	7.713
74	29	29	GUA	H8	H	1	7.324
75	30	30	ADE	H2	H	1	7.001
76	31	31	CYT	H41	H	1	8.724
77	31	31	CYT	H42	H	1	6.383
78	32	32	URI	H1’	H	1	5.552
79	32	32	URI	H2’	H	1	3.679
80	32	32	URI	H6	H	1	7.653
81	33	33	URI	H1’	H	1	6.021

82	33	33	URI	H2’	H	1	4.578
83	33	33	URI	H6	H	1	7.941
84	34	34	CYT	H1’	H	1	5.874
85	34	34	CYT	H2’	H	1	4.010
86	34	34	CYT	H6	H	1	7.614
87	35	35	GUA	H1’	H	1	5.885
88	35	35	GUA	H2’	H	1	4.732
89	35	35	GUA	H8	H	1	7.780
90	36	36	GUA	H1	H	1	13.364
91	36	36	GUA	H1’	H	1	5.470
92	36	36	GUA	H2’	H	1	4.521
93	36	36	GUA	H8	H	1	8.239
94	37	37	URI	H1’	H	1	5.339
95	37	37	URI	H2’	H	1	4.577
96	37	37	URI	H3	H	1	13.488
97	37	37	URI	H6	H	1	7.428
98	38	38	CYT	H1’	H	1	5.416
99	38	38	CYT	H2’	H	1	4.277
100	38	38	CYT	H6	H	1	7.896
101	39	39	GUA	H1’	H	1	5.737
102	39	39	GUA	H2’	H	1	4.829
103	39	39	GUA	H8	H	1	8.039
104	42	42	URI	H1’	H	1	5.710
105	42	42	URI	H2’	H	1	4.282
106	42	42	URI	H6	H	1	7.674
107	43	43	CYT	H1’	H	1	5.714
108	43	43	CYT	H2’	H	1	4.464
109	43	43	CYT	H41	H	1	8.020
110	43	43	CYT	H42	H	1	6.668
111	43	43	CYT	H6	H	1	7.652
112	44	44	GUA	H1	H	1	12.020
113	44	44	GUA	H1’	H	1	5.537
114	44	44	GUA	H2’	H	1	4.573
115	44	44	GUA	H8	H	1	7.693
116	45	45	GUA	H1	H	1	12.261
117	45	45	GUA	H1’	H	1	5.665
118	45	45	GUA	H2’	H	1	4.628
119	45	45	GUA	H8	H	1	7.708
120	46	46	ADE	H8	H	1	8.131
121	48	48	GUA	H1	H	1	12.433
122	53	53	CYT	H1’	H	1	5.873
123	53	53	CYT	H2’	H	1	4.586
124	53	53	CYT	H41	H	1	8.219
125	53	53	CYT	H42	H	1	6.808
126	53	53	CYT	H6	H	1	8.163
127	54	54	URI	H1’	H	1	5.166
128	54	54	URI	H2’	H	1	4.253
129	54	54	URI	H6	H	1	7.629
130	55	55	URI	H1’	H	1	5.506
131	55	55	URI	H2’	H	1	4.357
132	55	55	URI	H6	H	1	7.626
133	56	56	ADE	H1’	H	1	5.959
134	56	56	ADE	H2’	H	1	4.340
135	56	56	ADE	H8	H	1	7.759
136	57	57	ADE	H1’	H	1	5.761
137	57	57	ADE	H2	H	1	7.000
138	57	57	ADE	H2’	H	1	4.603

139	57	57	ADE	H8	H	1	7.455
140	58	58	CYT	H41	H	1	8.069
141	58	58	CYT	H42	H	1	6.676
142	58	58	CYT	H6	H	1	7.301
143	59	59	GUA	H1	H	1	13.179
144	59	59	GUA	H1’	H	1	5.284
145	59	59	GUA	H2’	H	1	4.299
146	60	60	URI	H1’	H	1	5.397
147	60	60	URI	H2’	H	1	4.519
148	60	60	URI	H3	H	1	12.963
149	60	60	URI	H6	H	1	7.642
150	61	61	ADE	H1’	H	1	5.837
151	61	61	ADE	H2	H	1	7.570
152	61	61	ADE	H2’	H	1	4.454
153	61	61	ADE	H8	H	1	7.949
154	62	62	GUA	H1	H	1	13.210
155	62	62	GUA	H1’	H	1	5.629
156	62	62	GUA	H2’	H	1	4.206
157	62	62	GUA	H8	H	1	7.238
158	63	63	URI	H1’	H	1	5.835
159	63	63	URI	H2’	H	1	4.306
160	63	63	URI	H6	H	1	7.800
161	64	64	URI	H1’	H	1	5.529
162	64	64	URI	H2’	H	1	4.442
163	64	64	URI	H6	H	1	7.788
164	65	65	CYT	H1’	H	1	5.590
165	65	65	CYT	H2’	H	1	4.340
166	65	65	CYT	H41	H	1	8.219
167	65	65	CYT	H6	H	1	7.911
168	66	66	URI	H1’	H	1	5.441
169	66	66	URI	H2’	H	1	4.554
170	66	66	URI	H3	H	1	12.849
171	66	66	URI	H6	H	1	7.738
172	67	67	ADE	H1’	H	1	5.854
173	67	67	ADE	H2	H	1	7.664
174	67	67	ADE	H2’	H	1	4.445
175	67	67	ADE	H8	H	1	7.061
176	68	68	ADE	H1’	H	1	5.793
177	68	68	ADE	H2	H	1	6.744
178	68	68	ADE	H2’	H	1	4.360
179	68	68	ADE	H8	H	1	7.836
180	69	69	CYT	H1’	H	1	5.259
181	69	69	CYT	H2’	H	1	3.982
182	69	69	CYT	H41	H	1	8.400
183	69	69	CYT	H42	H	1	6.860
184	69	69	CYT	H6	H	1	7.370
185	70	70	CYT	H6	H	1	7.519

stop_