Structural and Dynamic Basis for Assembly of the HIV-1 TAR-Tat Ribonucleoprotein Complex

by

Elizabeth A. Dethoff

A dissertation submitted in the partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2011

Doctoral Committee:
    Professor Hashim M. Al-Hashimi, Chair
    Professor Charles L. Brooks III
    Professor David R. Engelke
    Professor Ayyalusamy Ramamoorthy

2011

To my family.

Acknowledgements

I would first like to express my gratitude to my advisor, Hashim Al-Hashimi, for his expertise and guidance throughout my graduate career. Working with him has been a truly inspirational experience. Secondly, I need to thank all current and past Al-Hashimi lab members. Thanks to Dr. Qi Zhang and Dr. Alexandar Hansen for teaching me how to set up my first NMR experiments, and thanks to Dr. Anette Casiano for helping me synthesize my first RNA. I'm grateful for the many invaluable discussions I've had with Dr. Catherine Musselman, Dr. Maximilian Bailor, Evgenia Nikolova, and Dr. Jeetender Chugh. Thanks to Dr. Andrew Stelzer, Catherine Eichhorn, Jameson Bothe, and Jeremy Kratz, my officemates, for the many insightful conversations and for making grad school more enjoyable. I also thank the many collaborators I've worked with over the years, including Prof. Ioan Andricioaei and Aaron T. Frank. Last but not least, I am forever indebted to my family and fiancé Ryan for their unwavering encouragement and support.

Table of Contents

List of Figures

List of Tables

List of Appendices

Appendix

Abstract

   The transactivation response element (TAR), located at the 5' end of the HIV-1 genome, regulates the transcription elongation step of viral RNA. The TAR stem-loop binds the HIV viral transactivator protein (Tat), which enhances viral transcription by recruiting the positive transcription elongation factor b (P-TEFb). The TAR-Tat interaction and formation of the TAR/Tat/P-TEFb ribonucleoprotein (RNP) complex remains poorly understood from a structural and dynamic standpoint. To better understand the formation of this complex, we have studied the structural and dynamic features of free TAR to elucidate motions in both the bulge and apical loop that may be important for adaptive recognition of protein targets and, ultimately, for anti-viral therapeutic design.

   We used a combination of nuclear magnetic resonance (NMR), molecular dynamics (MD), and mutagenesis to characterize the structural dynamics of free TAR. Residual dipolar couplings (RDCs), motionally-decoupled ^{13}C relaxation (R_1, R_2), and ^{13}C $R_{1\rho}$ relaxation dispersion were used in combination to characterize global and local motions occurring on fast (ps to ns) and slow (μs to ms) timescales. These results were compared to an MD simulation of TAR. Our results show that the apical loop and bulge act as independent dynamical centers, with the apical loop undergoing complex dynamics on the microsecond to picosecond timescale that may be important for adaptive recognition. We have characterized a lowly-populated structure of the loop that is otherwise "invisible" to traditional NMR techniques, representing the first-ever characterization of an "invisible" state of an RNA loop by NMR. Overall, our results establish a broadly useful approach for characterizing the dynamics of RNA loops and shed light on the relationship between dynamics and biological function for the TAR element in the HIV genome.

Chapter 1

Introduction

1.1 RNA Structure and Dynamics

1.1.1 Role of RNA Dynamics in Biology

Francis Crick first proposed his "sequence hypothesis" in 1958, which articulated that DNA and/or RNA code for proteins and that the resulting protein sequence is dictated solely by the sequence of nucleic acid bases [1]. In the years that followed, it was generally accepted that DNA acts solely as a repository of genetic information and RNA functions as a passive carrier of this information. Proteins were mostly thought to play all catalytic and regulatory roles in a cell. One of the first pieces of evidence that disputed this pathway of information, now termed the "central dogma of biology," was the observation that RNA can code for DNA, a phenomenon known as reverse transcription [2,3]. Around this same time the discovery of non-coding RNAs (ncRNAs) such as tRNA and rRNA demonstrated that RNA has functions outside of serving as a template for protein production. Since then, the catalogue of RNA biological roles has exploded to include RNA modification [4], transcription and translation regulation [5,6], and gene silencing (Table 1.1) [7]. For example, many riboswitches bind a target ligand to regulate the transcription of genes or translation of proteins involved in the production or degradation of the ligand [8]. 7SK RNA is a small nuclear RNA (snRNA) that down-regulates RNA polymerase II activity by binding and sequestering the positive transcription elongation factor b (P-TEFb) and the transcriptional regulator HEXIM1 [9]. RNA can even control gene expression at the level of chromosome activation; the X chromosome in female somatic cells is inactivated by *Xist*, a member of the long non-coding RNAs (lncRNAs) [10]. It is likely that as we further investigate the functions of DNA, RNA, and proteins and their relationships to one other, we will continue to reshape our understanding of the flow and regulation of genetic information.

Type	Abbreviation	Function
7SK RNA	7SK	inhibits transcription
antisense RNA	aRNA	gene regulation
guide RNA	gRNA	mRNA editing
long non-coding RNA	lncRNA	various
messenger RNA	mRNA	codes for protein
microRNA	miRNA	gene regulation
Piwi-interacting RNA	piRNA	gene silencing
ribosomal RNA	rRNA	translation
Rnase P		tRNA maturation
Rnase MRP		rRNA maturation; DNA replication
signal recognition particle RNA	SRP RNA	directs protein traffic
small interfering RNA	siRNA	gene regulation
small nuclear RNA	snRNA	RNA splicing; various
small nucleolar RNA	snoRNA	RNA modification
telomerase RNA		telomere synthesis
transfer RNA	tRNA	translation
transfer-messenger RNA	tmRNA	tags proteins for proteolysis
Y RNA		DNA replication; cell proliferation

Table 1.1: Major classes and examples of functional RNAs. Adapted from references [11] and [12].

What is particularly amazing about the ability of RNA to carry out such a diverse range of functions is that it can do so despite being made up of only four nucleic acid building blocks. Growing studies suggest this functional complexity is derived in large part from RNA's ability to undergo dramatic conformation changes on binding ligands, proteins, or sensing changes in the cellular environment [13-19]. One quintessential example of the relationship between conformational change and biological function are riboswitches, which are *cis*-acting elements located at the 5' untranslated region (UTR) and sense changes in the cellular environment, such as temperature, metabolites, and metal ions, to adaptively change gene expression through complex structural rearrangements.

Most riboswitches consist of two domains: a highly conserved aptamer domain that binds a ligand at high specificity and an expression platform. Ligand binding induces a structural rearrangement in the aptamer domain that is transduced to the expression platform, altering its secondary structure. This change in the expression platform either
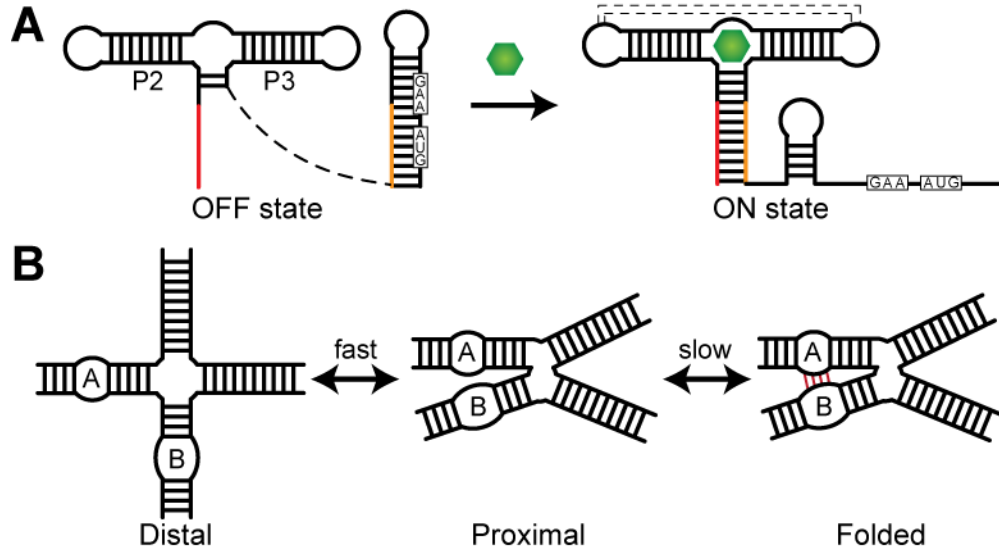
Figure 1.1: Role of conformational dynamics in RNA function. (A) Structural rearrangement of *add* adenosine-sensing riboswitch on binding adenosine. The Shine–Dalgarno sequence and start codon are boxed. (B) Equilibrium of three states of the hairpin ribozyme.

creates a transcription termination helix or sequesters the Shine-Dalgarno sequence, inhibiting translation of downstream genes involved in the biosynthetic pathway of the ligand. Conversely, the *add* adenine-sensing riboswitch activates expression at the level of translation in the presence of ligand (Figure 1.1). The aptamer domain consists of a three-way junction, where the apical loops of helix P2 and P3 form tertiary interactions necessary for ligand binding [20]. In the free form, these long-range tertiary loop-loop interactions are partially formed, which is thought to preform the binding pocket for efficient ligand recognition. Binding of adenosine results in a conformational change in the junction and in the distant loop-loop region of the aptamer domain, making the Shine–Dalgarno sequence and the AUG codon in the expression platform accessible to the ribosome which allows for the translation of downstream proteins [21].

Another class of functional RNAs that commonly undergo large structural rearrangements to carry out biological functions are ribozymes, a group of protein-independent RNA enzymes. For example, the hairpin ribozyme consists of four stems joined by a four-way junction and catalyzes sequence-specific cleavage and ligation of RNA (Figure 1.1) [22]. At equilibrium there exists three structural states: a distal, proximal, and folded state [23]. In the catalytically active folded position, helix A and B

loops are in close proximity and form tertiary interactions. The dynamic nature of the four-way junction allows for the fast interconversion of the distal and proximal state, repeatedly bringing the two loops within close proximity to promote formation of the folded state and the active site.

RNA often undergoes conformational changes during ribonucleoprotein (RNP) assembly. One example is the transactivation response element (TAR) of HIV-1, which undergoes conformational changes around the bulge that translate to reorientation of helices on binding to peptides that mimic its cognate protein Tat [24-27]. Similarly, the U1A protein, a component of the U1 small nuclear RNP complex (snRNP) [28], induces a conformational change in 3'-UTR RNA from a largely unstructured to a well-folded RNA [29]. The protein binding site is an unstructured, dynamic internal loop that becomes structured on binding U1A [30].

RNA dynamics are not limited to large interhelical motions that reorient helices to change the three-dimensional structure of the RNA. Local dynamic processes include base pair opening [31], repuckering of the sugar moiety [32], and base flipping [33]. To fully understand the relationship between dynamics and biological function, we must advance beyond solving static three-dimensional structures to fully characterize RNA dynamics at atomic resolution.

1.1.2 Structure and Dynamics of Apical Loops

Although RNA can form intricate three-dimensional structures to carry out biological functions as described in the previous section, RNA is largely considered to be a modular biomolecule, consisting of a series of structural elements strung together. The entire molecule is then tied together by long-range, intramolecular interactions, including key motifs such as kissing apical loops, tetraloop-tetraloop receptor interactions, pseudoknots, and base triples. Common RNA structural elements include A-form helices, apical loops, and internal loops or bulges. Apical loops cap the end of a double helix and serve both structural and functional roles. They are often sites of tertiary structure stabilization [8], protein interaction [34,35], and even bind small molecules [36].

For example, in the P4-P6 domain of the *Tetrahymena* group I intron RNA, a long-range interaction between a GAAA tetraloop in the P5abc extension and tetraloop

receptor of the P456 helical stack brings these two structural domains into proper tertiary orientation, allowing for a remarkably sharp bend between the two domains [37]. This interaction contributes significantly to the stability of the three-dimensional structure [38]. A second RNA that contains a biologically functional apical loop is the preQ$_1$ riboswitch. To date, the preQ$_1$ riboswitch has the smallest known ligand-binding domain of all known riboswtiches and its decaloop is the binding site of preqQ$_1$, a precursor of queuosine [36]. The small molecule is located in a deep pocket in the loop, forming interactions with two residues in the loop and one from the 3' ssRNA [39]. The ligand is covered by a base quartet formed by three residues in the loop and a fourth from a tail.

The most common and well-characterized class of apical loops is the tetraloop, the hairpin loop most frequently found in ribosomal RNAs in a variety of species [40]. The two most common consensus tetraloop sequences are GNRA and UNCG, where N is any nucleotide and R is any purine (A/G) (Figure 1.2). Both loops form unusually stable conformations, as revealed by computational and thermodynamic studies [41,42]. As such, the UUCG tetraloop is commonly used in NMR structural studies due to its high stability and favorable spectroscopic qualities [43]. The GNRA tetraloop is a widespread motif with diverse functionality ranging from protein recognition [44,45] to long-range RNA tertiary interactions[46], and can even be a target of enzymatic cleavage [47,48]. Structures of both loops have been determined by both NMR [49-52] and X-ray crystallography [53,54]. In the UUCG loop, the G (G34) is in the *syn* conformation. The noncanonical U-G base pair between U31 and G34 is stabilized by both base-base and base-sugar interactions [55]. Spin relaxation studies have shown that this unusual base pair is as stable as helical base pairs [56,57]. The GAGA tetraloop (G$_{14}$AGA$_{17}$) is stabilized by a sheared *anti-anti* G-A base pair and a heterogeneous network of up to five hydrogen bonds between G14 and A17 [51]. All residues are in the *anti* conformation. G14 is stacked on the 5' side of the loop and continuous stacking from A15 to A17 also contributes to the stability of the tetraloop (Figure 1.2) [51,52].

A recently recognized apical loop motif is the pseudo-triloop. This loop is formed when the first and fifth nucleotide in a hexanucleotide loop (hexaloop) form a *trans*-loop base pair, thus creating a triloop from positions two through four and a single nucleotide bulge at the sixth position (Figure 1.2). Secondary-structure prediction programs often
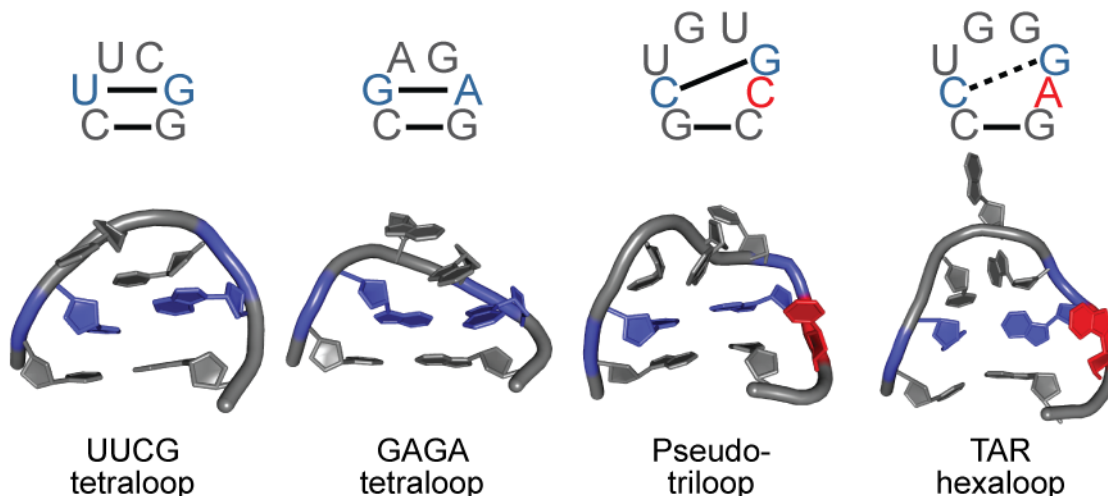
Figure 1.2: Example apical loops. (A) Crystal structure of UUCG loop and NMR-solved structures of (B) GAGA loop of HIV-1 [52], (C) Pseudo-triloop of HBV [63], and (D) and hexaloop of HIV-1 TAR [64].

predict a hexaloop over the pseudo-triloop motif, often overlooking this motif. Examples of this loop motif include the Iron Responsive Element (IRE) [58], the encapsidation signal of the Hepatitis B virus [59], the promoter of the brome mosaic virus [60], and the P3 domain of RNase MRP RNA [61]. It was recently proposed that the apical loop of the transactivation response element (TAR) of HIV-1 resembles a pseudo tri-loop, with a cross-loop base pair between the first (C30) and fifth (G34) nucleotide (Figure 1.2) [62]. Interestingly, many of the example pseudo-triloops listed above are involved in protein interactions. For example, the AGU pseudo-triloop of IRE makes 10 contacts to the iron regulatory protein 1 (IRP1) and is necessary for efficient protein binding [34]. Thus, this class of loops may serve as a general protein interaction motif.

Despite the obvious roles of apical loops in RNA:RNA and RNA:protein interactions and the fact that RNA often carries out functions through conformational rearrangement, very few studies have sought to characterize the dynamics of apical loops. Of the few studies that exist, most focus on the more thermodynamically stable tetraloop class [32,65-67]. This is likely due to the increased complexity of larger penta- and hexaloops. This thesis seeks to outline a general strategy for studying the dynamics of larger apical loops on the picosecond to millisecond timescale using a combination of NMR and mutagenesis. The RNA of interest is the hexaloop of TAR, which plays multiple roles in the life cycle of the human immunodeficiency 1 (HIV-1) virus.

1.2 The Transactivation Element of HIV-1

<u>1.2.1 Transcription Elongation of HIV-1</u>

HIV-1 Background

HIV/AIDS is one of the world's most urgent public health challenges, with more than 27 million deaths since its discovery in 1981 and approximately 33 million infections worldwide, two-thirds of which are in sub-Saharan Africa (UNAIDS/WHO AIDS). In 2009 alone, approximately 2.6 million individuals were newly infected with HIV and about 2 million individuals died from AIDS-related illness (UNAIDS/WHO AIDS). The HIV virus infects cells of the human immune system, impairing its ability to fight disease. Infection eventually results in the deterioration of the immune system, leaving the body susceptible to opportunistic infections and diseases.

With the annual number of new HIV infections in decline, the rate of growth of the epidemic appears to be slowing. However, the number of people living with HIV has continued to grow, in part due to the life-prolonging effects of antiretroviral therapy (ART) that has lead to a dramatic reduction in the morbidity and mortality related to the condition worldwide. As of 2009, 26 anti-HIV drugs were approved by the FDA [68]. Taken individually, these drugs are relatively effective in reducing viral replication, but HIV's high replication and mutation rate can create drug-resistant strains within days. Thus, anti-HIV drugs are usually taken in combination of two or more, with each drug targeting a different step in the replication cycle in the hopes of curbing resistance. Currently available drugs target only four out of more than 15 viral proteins and thus interfere with only a limited number of the major steps of the HIV life cycle [68].

Unfortunately, current drug treatments are not effective for all individuals due to drug resistance and very serious adverse health effects from taking the treatment long-term [69]. The latter is a major cause for medication non-adherence, which increases the likelihood of the development of resistance and is a particular problem in underdeveloped countries that often have access only to cheaper, more toxic medications. To better suppress virus replication and curb drug resistance, we must widen the range of targeted HIV replication steps by developing new therapeutic agents directed against new targets. Potentially outstanding targets include RNA structures essential for HIV replication, especially those steps that have no druggable protein target.
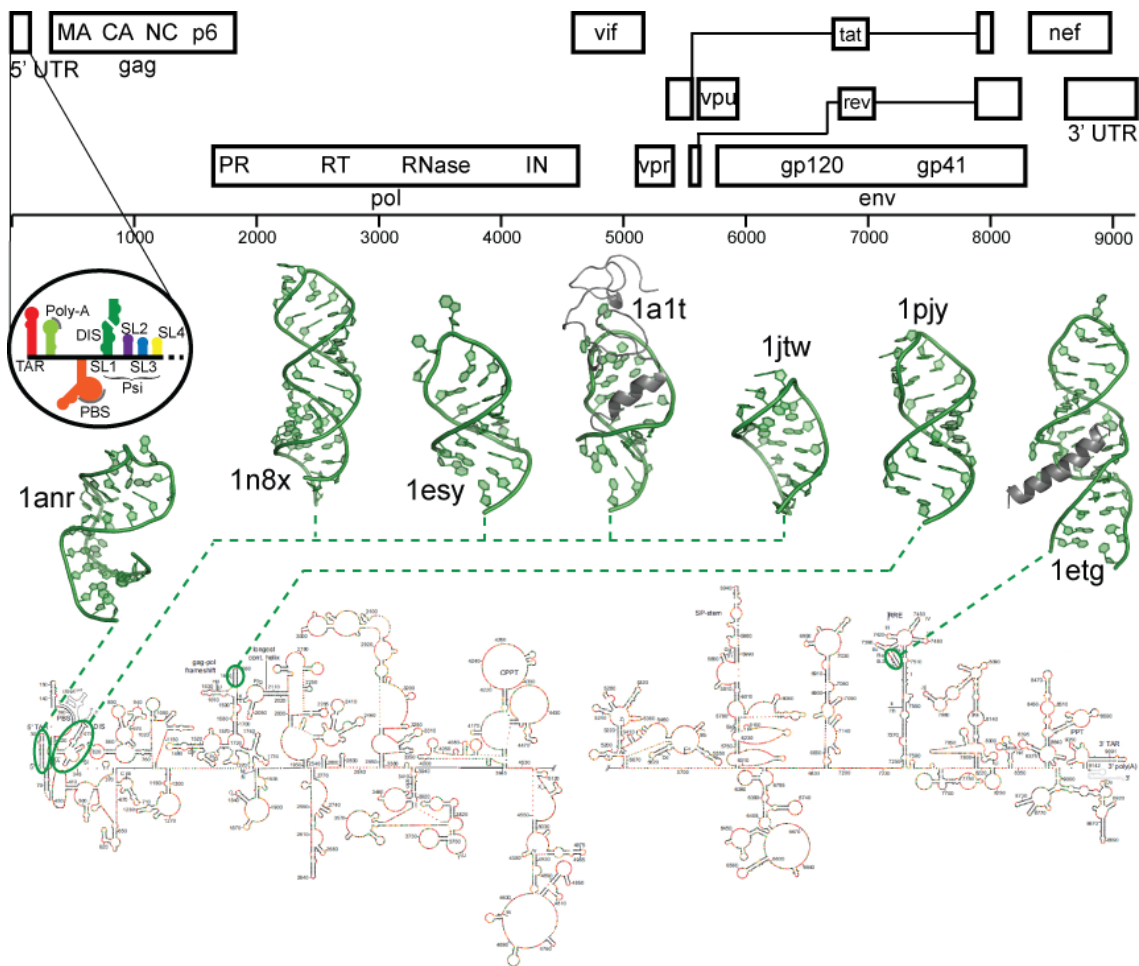
Figure 1.3: Genome of HIV-1 [70]. In green are structures of biologically important RNA elements (from left to right: TAR, SL1, SL2, SL3, SL4, frameshift inducing stem-loop, and RRE). PDB IDs are included next to each structure. Dashed lines and circles indicate location in the HIV-1 genome. Adapted by permission from Macmillan Publishers Ltd: *Nature* [71], copyright (2009).

RNA structures of the HIV-1 genome that could serve as potential drug targets are depicted in Figure 1.3. The majority of HIV-1 regulatory RNAs that have been identified thus far typically reside in the 5' and 3' untranslated regions, because until recently, ~85% of the HIV-1 genome structure was undefined. Recent SHAPE analysis of the entire HIV-1 genome has revealed the secondary structure of the uncharacterized regions, which may aid in the discovery of new druggable RNA structural elements. Current known RNA elements include the Rev responsive element (RRE), which can be targeted to disrupt nuclear export of unspliced viral RNA. The dimerization initiation site (DIS) can be targeted to block genome dimerization, packaging and viral maturation. The

transactivation response element (TAR), which will be the focus of this thesis, is a target for inhibiting viral transcription (Figure 1.3 and 1.4).

Transactivation of the HIV Genome

HIV is a retrovirus, and as such its genome is stored in infectious virions as an RNA dimer that is reverse transcribed by the viral protein reverse transcriptase upon infection of the host cell. Integrase, another viral protein, then inserts the newly made viral DNA into the human genome. Transcription of HIV is now dependent on both viral and host cell transcription factors (for review see reference [72]). The first step in transcription initiation is formation of the preinitiation complex. In this step, *cis*-acting elements in the viral DNA sequence recruit a number of host cell factors, including the TATA binding protein (TBP) that binds the TATA box located upstream from the transcription start site. After a cascade of additional factors join the complex, RNA Polymerase II (RNAPII) is recruited by the Transcription Factor for polymerase II H (TFIIH). To clear the promoter, TFIIH unwinds the DNA, opens the transcription bubble, and phosphorylates serine at position 5 in the C-terminal domain (CTD) of RNAPII. The polymerase is now free to proceed down the DNA and transcribe ~60nt of the 5' untranslated region (UTR). However, RNAPII has not yet entered productive elongation, for its processivity is still controlled by both negative and positive transcription factors. Two negative transcription elongation factor (NEFs) bind the complex: negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF), causing RNAPII to pause or stall [73,74]. At this point, the recruitment of a positive elongation factor to alleviate pausing and activate transcription is necessary; otherwise, the complex dissociates, resulting in abortive transcription.

Interestingly, it was known as early as the mid 1980s that the long terminal repeat (LTR) of the HIV-1 viral genome contains a *trans*-acting element that stimulates gene expression of the host cell by more than two orders of magnitude compared to an uninfected cell [75]. This was proposed to account for the virulent nature of the virus. This "transactivation" was proposed to be caused by a then unknown trans-activating factor (Tat) and has since been mapped on the HIV genome as an 86 amino acid (aa) protein (Figure 1.4C) [76]. At the time, proteins were only known to control gene

expression through interaction with the DNA promoter region, so it came as a surprise when the Tat binding site was mapped downstream of the initiation site, indicating that the transactivation-responsive region (TAR) is an RNA element [77-79]. Soon after this discovery, the sequence of TAR was also mapped and discovered to be a stable stem-loop structure at the 5' end of the leader region (Figure 1.4B) [80-84].

Early studies of TAR and Tat binding in the early 1990s showed that Tat binds the trinucleotide bulge but that the loop sequence was unnecessary [85,86]. This disagreed with previous studies that showed that the loop is a requirement for transactivation [80,81]. Thus, it was proposed that Tat functions by recruiting a third factor that binds the TAR loop and alleviates the pausing of RNAPII [87]. In 1995, Tat was shown to bind a nuclear serine/theonine kinase complex called the Tat-associated kinase (TAK) that phosphorylates the CTD domain of RNAPII [88]. TAK was later identified as the positive elongation factor b (P-TEFb), a protein complex consisting of Cyclin T1 (CycT1) and cyclin-dependent kinase 9 (Cdk9) (Figure 1.4A) [89,90]. Indeed, gel electrophoresis experiments confirmed that the loop sequence is required for CycT1 but not Tat binding and the data suggested that the binding of Tat and CycT1 to TAR is cooperative [91].

Once localized near the RNAPII complex through its interaction with TAR, Cdk9 of P-TEFb phosphorylates the NELF-e subunit of NELF, causing it to leave the RNAPII
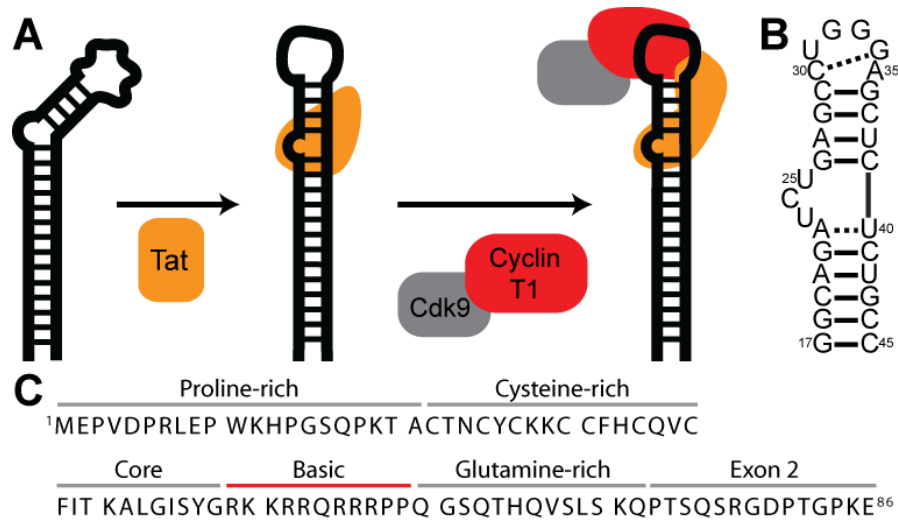
Figure 1.4: Transactivation of the HIV genome by TAR. (A) Traditional model of the transactivation step of the HIV genome. (B) Secondary structure of TAR. (C) Sequence of Tat from HIV-1 strain HXB2.

complex [92] and phosphorylates the spt5 subunit of DSIF, turning it into a positive elongation factor [93,94]. Finally, Cdk9 phosphorylates the CTD of RNP II, facilitating the transition from transcription initiation to elongation and allowing for formation of full-length HIV transcripts [95,96].

1.2.2 Transactivation RNP Complex

Proteins of the RNP Complex: Tat, Cyclin T1, and Cdk9

Tat is an 86-101 residue protein (9-11kDa) that is essential for transcription of viral mRNA (Figure 1.4C) [97-99]. The 99-101 aa Tat is the predominant form found in all clinical isolates, except subtype D. The most studied Tat form is from subtype B and is a truncated form at 86 aa long [97]. Tat is encoded by two exons and studies have shown that *in vitro*, the first exon (1–72aa) is sufficient for transcription but the second exon (14-34 aa in length, depending on the isolate) is necessary to function *in vivo* in other roles, such as activation of NF-kB [100], cellular uptake [101], induction of apoptosis [102,103], and binding the human translation elongation factor 1-delta [104].

The first exon of Tat can be divided into six regions [97,105]. The first region (1-21aa) is proline rich, and the second (22-37aa) contains seven well-conserved cysteines. The third region (38-48aa) contains a conserved [38]Phe and the sequence [43]LGISYG. The fourth, basic region (49-59aa) contains the well-conserved [49]RKKRRQRRRP, which is involved in many of Tat's functions. The fifth region (60-72aa) is glutamine-rich. The last region is encoded by the second exon.

All subtypes except C contain seven cysteines (22, 25, 27, 30, 31, 34, 37) in the cysteine-rich region, and there are usually no other cysteines found in the rest of the protein. Modification of these cysteines drastically changes the CD spectrum of Tat [106], and it has been suggested that these cysteines must bind zinc for Tat to function properly [107]. The seven cysteines are easily oxidized [108], which complicates recombinant expression and folding and results in a loss of biological activity [109].

Multiple post-translational modifications regulate Tat activity. For example, acetylation of lys50 by p300/CbP has been shown to dissociate Tat from the TAR complex [110-113]. Conversely, acetylation of lys28 enhances Tat binding to P-TEFb [110,114]. It has been suggested that this latter acetylation may inhibit HIV-1 splicing,

which allows for the synthesis of full-length transcripts needed for production of infectious virions [115].

To date, there have been no X-ray crystallography structural studies of full length HIV-1 Tat, though there are multiple NMR studies of different Tat variants. The first study examined an 86 aa HIV-1, Zaire 2 isolate variant under reducing conditions (1mM DTT) [116], which has been shown to inactivate Tat [117]. An 86- aa Tat Bru variant was studied at pH 4.5 and it was found that the basic and cysteine-rich regions which are functionally important are exposed to solvent [118]. Similarly, a study of Tat Mal from subtype D also showed that these two regions are exposed to solvent [119]. The only Tat study of a non-truncated form reported a structure similar to that of the Bru and Mal variants, with the C-terminus packed between the basic and cysteine-rich regions [120]. All three structures show that the core region consists of part of the N-terminus. Conversely, a study of the first exon of Tat (1-72aa) suggests that Tat is a naturally unfolded protein [121]. However, Tat folding has been confirmed by vaccine studies that raise antibodies against Tat epitopes [122-124], so it seems unlikely that Tat is completely unfolded. Instead, this NMR study suggests that the second exon of Tat may be important in the folding of the protein. The study was also complicated by the fact that the protein contained a His-tag and a 20aa T7 epitope.

The other protein partner of TAR is P-TEFb, which facilitates the transition from abortive to productive transcription elongation by RNAPII and is thus involved in the expression of many mRNAs [125]. Activity of P-TEFb is regulated by controlling its equilibrium between two major states: a small active form and a larger inactive form that includes 7SK RNA and hexamethylene bis-acetamide inducible protein 1 (HEXIM1) or 2 (HEXIM2) [9,126]. The distribution between these two states must be carefully regulated to ensure that P-TEFb does not haphazardly activate genes. Interestingly, HIV has evolved a mechanism to commandeer control of P-TEFb by evolving both an RNA element and a protein that mimic 7SK and HEXIM1: TAR and Tat. Tat shares an arginine-rich motif with HEXIM1 and so it can bind directly to 7SK, thus competing with HEXIM1 for P-TEFb. In this way, Tat extricates P-TEFb from the inactive RNP complex so that it can be used to activate the HIV genome transcription [127,128].

12

The human CycT1 (CycT1) is 726aa in length and contains a cyclin box repeat (31-250aa). The N-terminal cyclin box repeat is involved in binding and activating Cdk9. Although Cdk9 can form a complex with Cyclin T1, Cyclin T2, and Cyclin K, only its complex with CycT1 can activate transcription [129,130]. In the presence of Tat, CycT1 binds to TAR and Tat via its Tat-TAR recognition motif (TRM, 251-271aa) located at the carboxy-terminal end of the cyclin domain [131]. The C-terminal region of CycT1 (273-726aa) is not necessary for complex formation, and thus many studies use a truncated form of CycT1 [107,132]. The third protein of the transactivation RNP complex is Cdk9, a cyclin-dependent kinase. Tat and CycT1 bind TAR in the absence of Cdk9, so the latter is often excluded from binding studies.

Although no structure of the entire transactivation RNP complex exists, a crystal structure of the Tat:P-TEFb complex was recently published [133]. The structure shows that Tat has contacts to the T-loop of Cdk9 and makes numerous contacts to CycT1. Tat forms a zinc-mediated bridge with Cys261 of CycT1. Of significance is that Tat induces significant conformational changes in both P-TEFb subunits and that its own conformation is also altered so that it is in an extended conformation.

TAR:Tat:P-TEFb RNP Complex Formation

The exact order of binding of Tat and CycT1 to TAR is unknown (Figure 1.4A). CycT1 does not bind TAR in the absence of Tat and although Tat can bind TAR in the absence of CycT1, the interaction is inefficient [134]. Thus, it has been suggested that the two proteins likely bind TAR at the same time. This is supported by the fact that CycT1 and Tat binding is cooperative [91,134].

Tat binds to TAR via its arginine-rich motif (ARM, 49-60aa) [88,91]. R52 is necessary for TAR recognition [135], and this single residue is commonly used as a Tat mimic although it has a much lower affinity for TAR than full-length Tat [136]. Without CycT1, Tat binds weakly to all sides of the TAR loop, and in its presence K50 of Tat interacts with the G34 side of the TAR loop [35]. Residues 252-260 of CycT1 interact with the U31 side of the TAR loop. This agrees with a binding study that suggests there are three binding sites on TAR RNA [134].

Tat binds to CycT1 through its N-terminal activation domain (1-48aa), increasing the affinity of both Tat and Cyclin for TAR [137]. It has been suggested that CycT1 enhances the TAR:Tat interaction by inducing a significant conformation change in Tat, which is supported by the recent crystal structure of the Tat:P-TEFb complex [133]. The Tat:CycT1 interaction requires zinc and Cys261 in CycT1, suggesting it forms an intermolecular zinc-finger with Tat [107] which is also observed in the crystal structure [133]. As mentioned above, acetylation of Lys28 slightly enhances the interaction of Tat and CycT1, and Asn257 in the TRM of CycT1 is necessary for recognition of this acetylation [114].

1.2.3 The Transactivation Element of HIV-1

Structure and Dynamics of TAR

The transactivation response element (TAR) RNA stem-loop is located at the 5' end of the UTR of the HIV-1 genome and plays essential roles in various viral replication steps (discussed below) (Figure 1.4B) [78,138]. TAR contains a highly conserved trinucleotide bulge and hexanucleotide apical loop that serve as binding sites for Tat [83,84,139] and CycT1[35,91,134] during transcription elongation, respectively. TAR was originally assigned to the nucleotides +1 to +80 of the viral UTR, but subsequent deletion studies have determined that the minimal TAR sequence necessary for transactivation can be limited to +19 to +42 nucleotides [140]. Additional key features needed for interaction with Tat and P-TEFb have been established by a number of chemical probing, mutagenesis, gel shift assays, and peptide binding studies. Tat interacts with U23, while C24 and U25 generally act as spacers, for the identity of these two nucleotides are not critical [84,86]. The base pairs above and below the bulge are also necessary [83,84], and Tat makes contacts with phosphates at positions 21, 22, and 40 [141,142].

TAR undergoes conformational changes around the bulge which translate to reorientation of helices on binding to Tat-derived peptides [24-27], divalent ions [143-145], and small molecules [146-149]. In free TAR, U23 is stacked on A22. Upon binding Tat or a Tat mimic, U23 forms a base triple with A27-U38 [24,135]. Although initially

14

challenged [64], additional NMR and molecular dynamics studies support its existence [26,150,151].

It has been suggested that the flexibility at the bulge junction has an important role in adaptive recognition [64,151-155]. Recently, Al-Hashimi and coworkers have shown that unbound TAR samples inter-helical motions and local fluctuations at and around the bulge that are similar to conformational changes that take place on ligand recognition, suggesting that ligands likely select conformers from a dynamical ensemble of TAR structures via "tertiary capture" [153,154]. However, little attention has been given to the apical loop dynamics and their relation to biological function despite its crucial role in the transactivation of the viral genome (Figure 1.4B).

TAR Apical Loop Structure

In 1992, the first structure of the HIV-1 TAR element bound to argininamide was solved by NMR to reveal that the apical loop is largely heterogeneous in structure [24]. In the NMR structural studies since then, few NOEs have been observed between loop residues, indicating that this region of TAR is likely not stacked and samples multiple conformations [17,26,156]. However, early experiments using RNase and chemical modifiers as probes suggested that the loop contains some secondary or tertiary structure [157]. Residues C30, G33, and G34 are relatively inaccessible, while U31, G32, and A35 are accessible to solvent [157,158]. The unique chemical shifts of A35 suggest it is looped out into solution [158,159]. These results led to the proposal that the apical loop is stabilized by a cross-loop base pair between C30 and G34, with A35 flipped outside of the loop as depicted in Figure 1.5 [62,134,150,157,158].

A mutational and chemical probing study by Kulinski and coworkers show that a single mutation at C30 or G34 that disrupts the proposed cross-loop base pair significantly changes the chemical probing reactivity profile at these sites and that a double mutation partially restores the wild-type profile [62]. To probe the effects of these mutations on transactivation activity, Kulinski et. al. used a LTR-luciferase assay and found that again, single mutations of these residues significantly decreases Tat response but a double mutation partially restores activity. The fact that there is only partial restoration of activity suggests that the TAR loop has an additional role outside of
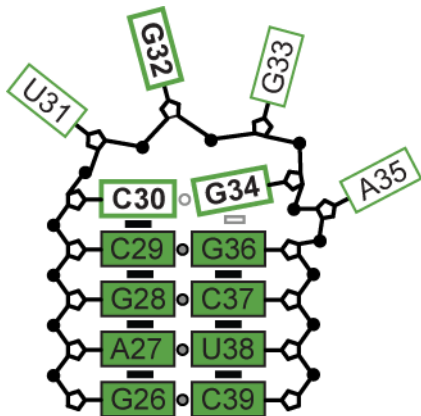
Figure 1.5: Schematic diagram of the proposed wtTAR structure.

providing stability to the upper helix. Indeed, this proposal is supported by a photo-crosslinking assay by Richter et. al., which determined that residues 252-260 of CycT1 interact with the U31 side of the TAR loop, while Tat residue K50 interacts with the G34 side of the loop [35]. Additional mutational studies show that the identity of U31, G33, and A35 are irrelevant, resulting in only a small loss of activity [134] and mutations of G32 or G34 reduce activity by half. Specifically, O6 and N7 of G34 are necessary for TAR:Tat:CycT1 complex formation. Also, A35 is needed as a spacer in the loop, possibly to distort the RNA backbone to give accessibility to the major groove of G34. The same study shows that the carboxyl group of G32 may also make a contact with the protein complex.

Considering how dynamic the apical loop is and how important it is for transactivation, it is surprising how few studies have focused on characterizing the structure and dynamics of the loop of free and bound TAR [160-162]. In a recent study by Bardaro et. al., ^{13}C NMR relaxation experiments were used to study changes in the motions of HIV-1 TAR in the presence of argininamide (K_d= 1mM) and two nanomolar-binding linear and cyclic peptide mimics of the Tat protein [161]. In free TAR, the authors found that the bulge and loop are highly dynamic. In the loop, every ribose (except C30) and every base (except C30 and U31, which were unresolved) exhibits fast motions at the ps–ns timescale. Slow motions are also observed in the ribose of every loop residue except G32 and G33. Of the bases that can be resolved, slow motions also exist in G34 and A35. On binding argininamide and the linear peptide, all resolved bases become less flexible on the ns-ps timescale except A35 in the Arg complex, which retains its flexibility. The ribose of C30 and U31 retain their slow motions, but the slow motions of A35 are lost. These results suggest that the linear peptide may interact with the loop and argininamide may have a second binding site near the loop. Alternatively, if these molecules do not interact with the loop, the changes they induce in the dynamics and structure of the bulge may be transmitted to the loop.

<u>1.2.4 Additional Roles of TAR in HIV Lifecycle</u>

TAR has a number of roles in the HIV lifecycle in addition to its namesake, including translation [163], the strand transfer step of reverse transcription [164,165], dimerization of the HIV-1 genome [166,167] and packaging of viral RNA into virions [168]. It has been suggested that TAR performs a role in RNA interference as well [169-171]. However, it has also been suggested by Das and coworkers that TAR has no essential function in HIV-1 other than transcription activation [172]. A few of these roles will be discussed in detail below.

Translational Control: PKR

Protein kinase R (PKR) is a double-stranded RNA-dependent serine-threonine kinase and plays an important role in antiviral defense (for review see reference [163]). PKR contains a Ser/Thr kinase domain and two N-terminal tandem copies of a dsRNA binding domain (dsRBD). It is translated in a latent form and is activated by autophosphorylation upon binding dsRNA, a common signal of viral invasion [163]. One PKR protein binds a minimum of 15 base pairs of cellular, viral, or synthetic dsRNA through its two dsRBDs, but a total of 30 base pairs is necessary to activate PKR [173-175]. The protein is referred to as a translational inhibitor because it phosphorylates the alpha unit of the translation elongation initiation factor 2 (eIF2α) in response to stress signals. Once phosphorylated, eIF2α stops translation of cellular and viral mRNAs, effectively inhibiting virus replication at the level of initiation.

Although the PKR and TAR relationship has been studied extensively, few details are known about the exact mechanism of binding. In fact, while most studies report activation of PKR by TAR [176-179], it has recently been suggested that TAR can either inhibit or activate PKR depending on its concentration [180]. The ability of high concentrations of TAR to inhibit PKR has been attributed to "dilution" of the protein on separate TAR molecules, but Heinicke et. al. propose that the formation of TAR dimers at different RNA and salt concentrations is the cause for this trend. To date, the exact mechanism of PKR inactivation/activation *in vivo* is unknown.

Kim et. al., studied the complex of TAR and the two dsRBDs and found that dsRBD1-dsRBD2 has two binding sites (one high affinity, one weak) and binds at a 1:1 stoichiometry [181]. Unfortunately, Kim and coworkers used imino chemical shift mapping and were thus limited in studying changes on binding in helical regions only. They show that the bulge and loop are important for binding, but do not specifically show that the protein binds the loop.

Reverse Transcription: Minus Strand-transfer Step

The genomic information of HIV-1 is stored in two single-stranded RNAs in virions. The reverse transcription of this RNA into double-stranded DNA involves multiple steps, including two intramolecular templates switches, or strand transfers. The first template switching is the minus-strand strong-stop transfer, which involves the formation of an RNA-DNA duplex. The newly transcribed minus-strand strong-stop DNA (ss-cDNA) is translocated to the 3'-end of the viral RNA and the complimentary repeat (R) regions base pair. The 3' R RNA sequence contains two secondary structures: the transactivation response element (TAR) and part of the poly(A) hairpin. Similarly, the ss-cDNA is predicted to form cTAR and cpoly(A) hairpins. These secondary structures have been proposed play a role in the transfer pathway [182].

The nucleocapsid (NC) protein facilitates the annealing of the RNA-DNA R sequences [183-185]. Although many studies have investigated the binding of NC to cTAR and TAR [164,186], the mechanism is still poorly understood. Multiple pathways of the annealing process have been proposed. One such mechanism involves a kissing duplex, which contains a loop-loop interaction between the apical loops of TAR RNA and cTAR DNA [165]. In a second mechanism, the annealing depends on nucleation through the 3'/5' termini, resulting in a "zipper" intermediate [187]. Further complicating the matter, single-molecule FRET studies show that there may be multiple pathways for the DNA and RNA to anneal [188]. However, many of these studies were likely complicated by the fact that "mini" TAR and cTAR constructs were used. More recently, Vo et. al. studied full-length TAR, cTAR, and NC and found that the structure of the intermediate is affected by the concentration of NC [189]. In the absence of NC, annealing involves fast formation of an unstable loop-loop "kissing" intermediate

followed by a slow conversion to the extended duplex. Under saturating NC conditions, the mechanism is dominated by the "zipper" pathway.

The NC:TAR complex has been studied by chemical and enzymatic probing and gel retardation [164]. When NC is added to TAR, the three G residues in the loop are protected (Figure 1.5). NC binding does not change the three-dimensional structure of TAR and likely does not facilitate the formation of dimers. It has been proposed that the NC zinc fingers interact with the guanine residues of the loop and the basic residues of NC interact with the upper stem.

Translational Control: TRBP

Cells have many lines of defense against viral infection, including the relatively newly discovered RNA interference (RNAi) pathway. In this mechanism, dsRNA originating from viruses are recognized by Dicer, a cellular protein, which cleaves the RNA into small duplexes. These small RNA duplexes then serve as a template for sequence-specific degradation. The human TAR-binding protein (TRBP) was first discovered due to its high affinity for TAR [190] and has since been identified as a binding partner to Dicer and a player in the RNAi pathway [191]. In fact, a lack of TRBP results in a loss of RNA silencing function [192]. By sequestering TRBP from Dicer, TAR is capable of reducing RNA silencing [193]. In this way, TAR can subvert the cell's natural defenses against infection.

Similar to PKR, TRBP is a double-stranded RNA (dsRNA)-binding protein. It has three binding domains, of which two have high affinities for dsRNA while the third domain binds to Dicer. Circular dichromism studies using a 24 amino acid peptide from the dsRBD of TRBP show that the peptide binds to TAR with a similar affinity as the full-length protein and binds at a 3:1 ratio [194]. This study by Erard and coworkers also suggests that the TRBP peptide induces a conformation change in TAR and maps the higher-affinity binding site to the upper stem and loop. Modeling and mutation studies suggest that the peptide likely interacts with G26 above the bulge and G33 of the loop (Figure 1.5).

The apical loop of HIV-1 TAR interacts with a number of very different proteins. It is possible that the structure heterogeneity of the loop enables it to interact with such a

wide breadth of partners. Thus, it is the aim of this thesis to characterize the dynamics of the apical loop of TAR over relevant biological timescales using techniques outlined in section 1.3 in hopes of gaining insight into the relationship between dynamics and function.

1.3 NMR Methods to Characterize Nucleic Acid Dynamics

1.3.1 The Study of RNA by NMR

Since its inception, the field of nuclear magnetic resonance (NMR) spectroscopy has advanced tremendously beyond the simple one-dimensional spectra of water and paraffin wax measured by Bloch and Purcell in the 1940s [195,196]. NMR has since become an indispensible tool for the determination of nucleic acid structure and dynamics, as it is the only experimental technique that offers atomic-level resolution of biomolecules in solution. A key feature of NMR that makes it such an elegant and precise tool is the chemical shift, a phenomenon first recognized in the early 1950s [197]. Chemical shifts originate from the intrinsic frequency of a given nucleus. The frequency signal of a nucleus depends on the magnetic field strength, so theoretically one would expect that every proton in an NMR sample would have the same frequency. However, local perturbations in the magnetic field caused by nearby nuclei create deviations or shifts in the frequency of every nuclei. Thus, chemical shifts differentiate magnetically nonequivalent nuclei and provide information on the local environment.

In the 1960s, the study of biomolecules by NMR was time-consuming and plagued by spectral overlap. Three of the biggest NMR developments towards studying RNA (outside of general instrumental development) include the invention of fourier-transform (FT) NMR, the development of multi-dimensional experiments, and isotopic labeling schemes. Prior to the invention of FT NMR, nuclei were excited by slowly sweeping the entire frequency range of interest. The invention of FT NMR by Ernst and Anderson in 1966 allowed for the excitation of nuclei over a range of frequencies and for the disentanglement of induction signals from multiple nuclei [198]. FT NMR significantly decreased the time needed to record NMR experiments and increased the signal-to-noise ratio. The invention of two-dimensional (2D) NMR in the 1970s reduced spectral overlap [199,200], allowing for the full assignment of a single protein [201] and

the determination of the three-dimensional structure of a protein [202]. Finally, $^{13}C/^{15}N$ isotopic labeling of carbons and nitrogens in RNA resulted in an explosion of multidimensional NMR techniques that would otherwise have been too time-consuming using natural-abundance [203,204], enabling the determination of the three-dimensional structure of an RNA [205].

In addition to structure determination, NMR can be used to determine the local and global dynamics of a biomolecule on the picosecond to second timescale (discussed in detail below). For example, ^{13}C and ^{15}N relaxation experiments are commonly used to probe fast (nanosecond-picosecond) dynamics [206], while relaxation dispersion experiments are used to measure slow (microsecond to millisecond) dynamics [207-209]. Residual dipolar couplings (RDCs) bridge these two timescales by reporting on sub-millisecond timescales [210,211]. These measurements also provide long-range structural information, which is particularly useful in the study of nucleic acids.

It should be mentioned that there is a large discrepancy between the number of nucleic acid-only and protein-only NMR structures reported in the PDB database (~420 and ~7700 to date, respectively). Also, most reported nucleic acid structures are relatively small compared to proteins. This is likely due to the fact that the measurement of NMR parameters for nucleic acids is complicated by carbon-carbon scalar couplings, narrow dispersion of chemical shifts, small proton density, and complexities arising from dynamics. However, recent advances in the field of nucleic acid NMR, including alternate labeling schemes [212], combination with other techniques such as SAXS and molecular dynamics [213,214], and novel experimental developments [215-219] will hopefully lessen the gap between the study of proteins and nucleic acids by NMR in the years to come.

1.3.2 RNA Resonance Assignment

Before the structure and dynamics of an RNA can be characterized, the identity of each chemical shift or resonance in a spectrum must be determined. For example, in a 2D spectrum of the aromatic carbons, each peak represents a different C2H2, C6H6, or C8H8 bond vector and must be assigned to the appropriate nucleotide in the RNA. To do this, individual nuclei are connected to other nuclei nearby via either through-space or

through-bond experiments. In this way, a network of distance relationships is created which is then used to assign peaks (for reviews see references [215,220]).

Through-space assignment strategies rely on the nuclear Overhauser effects (NOEs) that occur between protons that are less than 5Å apart. Dipole-dipole coupling between two proximal protons results in an NOE cross peak, where the intensity of this cross peak is related to the distance between the two protons. NOEs can be measured using an NOESY experiment, which provides both intra- and internucleotide connectivities to create what is referred to as an NOE walk [221]. In a nonexchangeable NOESY, an H8 or H6 proton will have an NOE peak to its own H1' and to the H1' of the 3'-adjacent residue. In this way, one can "walk" along the RNA in the 5' to 3' direction (see Figure 1.6A for an example NOE walk). Additional intranucleotide connections (e.g. H5 to H6 of cytosines and uridines) and internuclear connections are also observed. Of particular usefulness are H2 protons, which have a cross peak to the H1' of the 3'-adjacent residue, to the opposite strand in the 5' direction, and to the H3 imino proton of the uracil to which it is base paired. An exchangeable NOESY can be used to not only assign imino resonances but to also determine the base pairing pattern of the RNA, because non-base paired imino proton resonances exchange rapidly with water and are not observed. Resonance overlap can lead to assignment ambiguities; therefore heteronuclear 3D NOESY-HSQC experiments are commonly used to increase the resolution of NOE assignments. Generally, NOE walks are only useful in A-form regions, because loops and bulges often lack close internucleotide proton interactions and dynamics can result in alternate conformations that give rise to unconventional NOEs.

NOESY data can be supplemented by information from through-bond assignment experiments that utilize scalar couplings, such as the HCN triple resonance experiment used for isotopically labeled samples (Figure 1.6B) [222]. In this experiment, the C6H6/C6H8 of the nucleobase is connected to the ribose C1'H1' via their shared N1/9, providing unambiguous assignment connections. It also clearly identifies these aromatic and sugar resonances as belonging to either Cs, Us, or As/Gs based on their very different N1/N9 chemical shifts (~147-171, 140-149, and 161-174 ppm, respectively).

Other through-bond experiments that provide intranucleotide assignments include the HCCH-COSY, which can be used to correlate adenosine H2 to H8 protons [223,224], and the HCCNH-TOCSY, which correlates base paired exchangeable imino protons
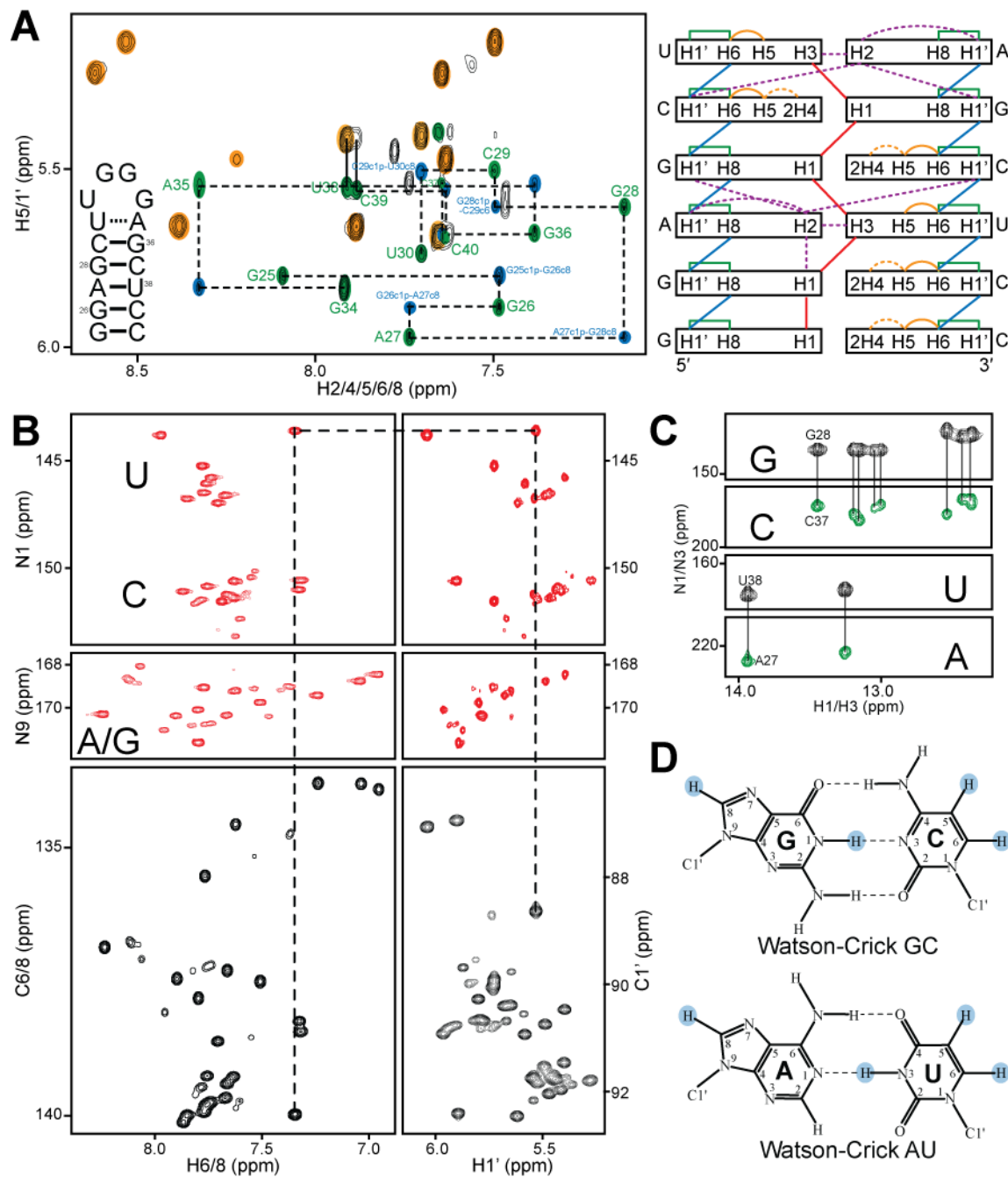
Figure 1.6: NMR assignment strategies. (A) Example nonexchangeable NOESY walk showing intra-(green) and internucleotide (blue) NOEs (*left*) and a schematic of common NOE connections (*right*). (B) 2D HCN connecting C8H8/C6H6 to C1'H1'. (C) HNN-COSY depicting hydrogen bonds. (D) Watson Crick base pairs. Aromatic protons of interest are highlighted in blue.

(H1/H3) to the H8/H6 of the nucleobase [225]. Finally, it is useful to verify the presence of base pairing outside of the exchangeable NOESY experiment by an HNN-COSY experiment (Figure 1.6C). This experiment uses the two-bond J_{NN} couplings between the N3/N1 of the donor uridine/guanosine and the N1/N3 of acceptor adenosine/cytidine (~7Hz) to correlate the donor and acceptor and identify base pairing partners [226,227].

1.3.3 Residual Dipolar Couplings

Until recently, the determination of three-dimensional structures of biomolecules was limited to the use of short-range NOE distance restraints ($\geq$ 5Å) and local angular constraints determined from scalar couplings. While these short-range parameters are usually sufficient for protein structure determination, which are more likely to contain interdomain NOEs, the propagation of error along an extended nucleic acid helix can result in significant errors in the overall helical shape and the global orientation of one helix relative to another. Also, compared to proteins, RNA contains fewer protons, and its limited chemical shift dispersion can result in significant spectral overlap, culminating in fewer probes for structure determination. Residual dipolar couplings (RDCs) help to overcome this issue, as they provide information on the orientation of individual bond vectors relative to the magnetic field [228,229]. RDCs are commonly used as a penalty function in the structure determination of RNAs or can be used to orient individual RNA fragments relative to one another.

RDC Theory

Dipolar couplings arise from the through-space coupling between the dipoles of two spins and are dependent upon the angle between the vector connecting the two nuclei and the external magnetic field (Figure 1.7). For nuclei I and S, the time-averaged Hamiltonian is written as

$$H_{DD}^{hetero} = -\left(\frac{\mu_o}{4\pi}\right)\left(\frac{\gamma_I\gamma_S h^2}{4\pi^2 r_{IS}^3}\right)\langle 3cos^2\theta - 1\rangle \vec{I}_z \cdot \vec{S}_z \qquad (1.1)$$

where γ_I and γ_S are the gyromagnetic ratios of nuclei *I* and *S*, μ_o is the permeability in a vacuum, *h* is Plank's constant, r_{IS} is the distance between the two nuclei, and θ is the angle between the vector connecting the two nuclei and the principle direction of the

external magnetic field ($\vec{B}_o$). The angular brackets around $\langle 3cos^2\theta - 1 \rangle$ denote both a time and ensemble averaging. For two directly bonded spin ½ nuclei, i and j, this equation simplifies to

$$H_{DD}^{hetero} = -\left(\frac{\mu_o \gamma_i \gamma_j h^2}{8\pi^3 r_{ij}^3}\right)\langle \frac{3cos^2\theta - 1}{2} \rangle \tag{1.2}$$

In solution, RNA tumbles isotropically and samples all possible orientations in three-dimensional space. When averaged over the entire ensemble of RNAs and over time, the dipolar coupling of a single bond vector averages out to zero. In this case, the only coupling between the two spins is the scalar coupling. However, if the RNA was somehow held static relative to the magnetic field, the dipolar coupling would become measureable. The coupling between a carbon and proton spin can be on the order of kilohertz, which is too large to measure by NMR. Thus, instead of completely aligning every RNA molecule in an NMR sample, we wish to attain a *partial* alignment in which there is a slight bias in the distribution of molecular orientations to measure a *residual* dipolar coupling (RDC). Through-space dipolar couplings (D) modulate the effective average field at a given nucleus and result in a splitting of the peak, similar to the effect of through-bond scalar couplings (J). RDCs are determined by measuring the splitting of a resonance under isotropic conditions (J) and under partially aligned conditions (J+D). By subtracting the splitting without alignment from the splitting with partial alignment, D is obtained (Figure 1.7).

To measure RDCs, the optimal level of alignment must be induced [230]. If too few molecules are aligned (~1 in 10,000), the RDC value will be too small to measure accurately. Conversely, if too many molecules are aligned (~1 in 100), the RDC value will be very large, rendering NMR spectra intractable. The ideal degree of alignment is when ~1 in 1,000 RNA molecules are aligned. Ordering media is most commonly used to induce this orientational bias, and varying the concentration of the media will change the average degree of alignment. In nucleic acid studies, the most commonly used alignment media is filamentous bacteriophage (Pf1) [231,232]. Phage is a rod-like molecule, estimated to be ~20,000Å in length and ~60Å in diameter and undergoes a phase transition from isotropic to ordered when placed in a magnetic field (Figure 1.7A) [233]. The negative charge of phage ensures there will be little attractive interactions between
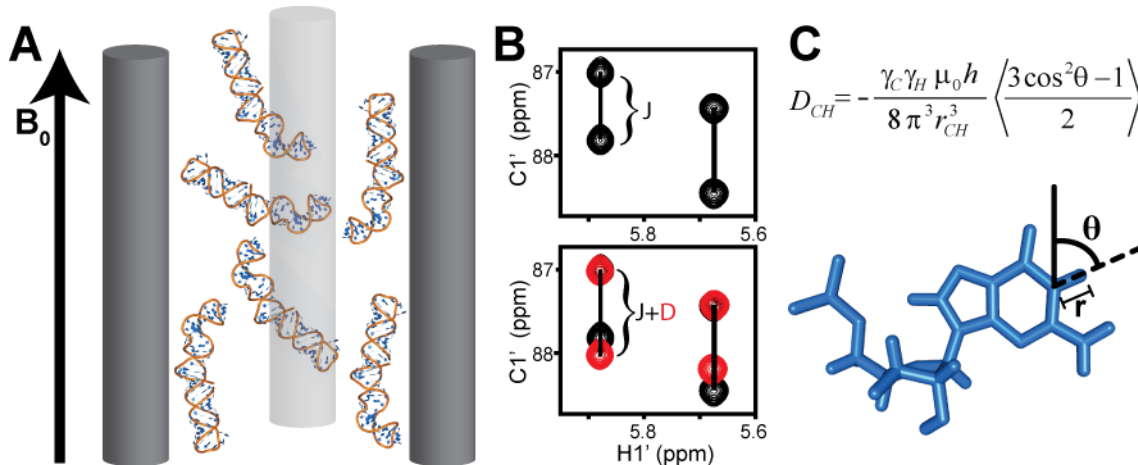
Figure 1.7: Measurements of RDCs. (A) Partial alignment of RNA using phage ordering media. (B) Measurement of RDCs as a contribution to splitting of resonances. (C) Dipolar coupling between a carbon and hydrogen spin depends on the average angle (θ) of the internuclear bond vector relative to the magnetic field.

the ordering media and nucleic acid. Phage is also tolerant to high salt conditions and a wide range of temperatures. For a small RNA of 30nt, a phage concentration of ~20 mg/mL will sterically align ~1 in 1000 molecules.

Due to the uniform distribution of charge in nucleic acids, RNA alignment in neutral or charged ordering media is largely governed by sterics and thus the overall molecular shape [234,235]. Effectively, the alignment media serves as a large wall and the RNA molecules located close to the wall can only sample certain conformations depending on their global shape. For example, the long axis of a helix of seven base pairs is along the helical axis and thus would be oriented with the helical axis parallel to the ordering media to avoid steric clashes. Increasing the length of the helix would increase the anisotropy of the shape of the RNA and thus increase the asisotropy of alignment.

RDC Analysis

As mentioned at the beginning of this section, RDCs are used primarily in two ways. The first technique uses RDCs as penalty functions in simulated annealing protocols [236,237]. In this strategy, RDCs and other NMR-measured parameters, such as NOEs and dihedral angles, are combined to determine the local structure of RNA fragments. This requires the back-calculation of RDCs throughout the annealing simulation. This calculation is not as straightforward as that for NOEs because it requires

knowing the two order tensor parameters S_{zz} and η (discussed below). In certain cases, these values can be estimated *a priori* [238].

The second type of RDC approach involves determining the orientation of molecular fragments relative to the magnetic field and, thus, to each other. This strategy first requires the division of the RNA into separate fragments of known structure, in this case being A-form helices. The validity of this step is supported by a statistical survey of RNA X-ray structures (solved with <3Å resolution) by Al-Hashimi and coworkers which shows that the local conformation of non-terminal Watson-Crick base pairs can be modeled *a priori* using a standard idealized A-form helix geometry [239]. This study also outlines an approach for taking into account structural noise in the A-form geometry and error in RDC measurements in the determination of the fragment alignment.

Secondly, this approach requires the determination of the RDC-derived order tensor that describes the symmetry and degree of alignment of a molecular fragment relative to the magnetic field [240,241]. The time-averaged angular term in equation (1.2) can be rewritten in terms of the time-independent orientation of the internuclear bond vector relative to an arbitrary frame (ϕ) and the five order tensor elements (S_{ij}):

$$\left\langle \frac{3cos^2\theta - 1}{2} \right\rangle = \sum_{ij=xyz} S_{ij} cos\phi_i cos\phi_j \tag{1.3}$$

$$S_{ij} = \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{pmatrix} \tag{1.4}$$

where ϕ_n is the average angular position of the bond vector and the *n*th axis of the arbitrary frame, and S_{ij} is the order tensor that describes the alignment of the molecule relative to the magnetic field. The frame in which the order matrix is diagonal is referred to as the principal axis system (PAS) of the order tensor. The tensor is a symmetric ($S_{ij}=S_{ji}$), traceless ($S_{xx}+S_{yy}+S_{zz}=0$), 3 × 3 matrix and therefore has five independent parameters. Theoretically, if a structure is assumed *a priori*, the five order tensor elements can be determined by the measurement of five or more independent RDCs.

The average orientation of the helical fragment relative to the magnetic field is defined by three angular parameters (S_{xx}, S_{yy}, S_{zz}) and is commonly represented by three Euler angles (α, β, γ). Two principal order parameters define the degree $\left(\vartheta = \right.$

27
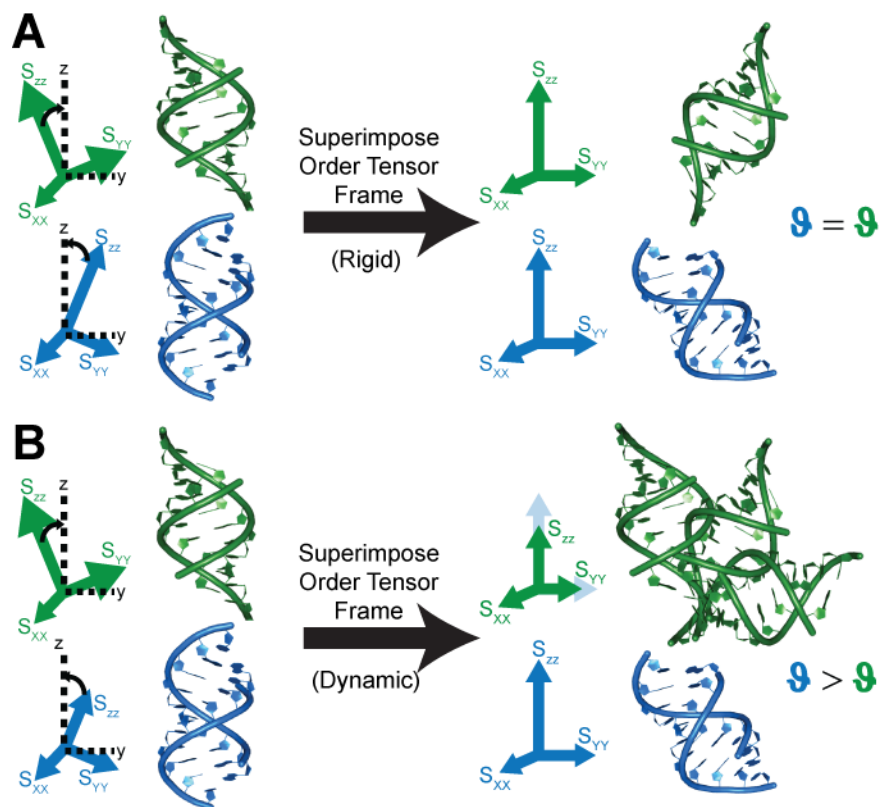
Figure 1.8: Order tensor analysis of RDCs. The orientation of each helix is individually determined. (A) By superimposing the two order tensor frames, the average orientation of the two helices to one another can be ascertained. (B) Comparison of the degree of order of each helix reports on interhelical motions.

$$\sqrt{\frac{2}{3}\left(S_{xx}^2 + S_{yy}^2 + S_{zz}^2\right)}\,\Big)\quad \text{and}\quad \text{asymmetry}\quad \left(\eta = \frac{S_{xx}-S_{yy}}{S_{zz}}\right) \text{of alignment} \quad [242].\quad \text{The}$$

generalized degree of order (ϑ) is sensitive to the dynamics of the molecular fragment originating from changes in overall alignment and internal motions.

Finally, after the order tensor of each helical fragment is determined and each fragment is rotated into its PAS, the relative orientation of the two fragments to one another can be determined by superimposing their order tensor frames (Figure 1.8) [241,242]. This step assumes that the fragments share a common view of the external magnetic field, which is a fair assumption if they are a part of the same molecule. The directions of the principal axes are unknown, resulting in a degeneracy of 4^{n-1} when assembling n fragments. For RNA helices connected by a junction, three of the four possible inter-helical orientations can usually be omitted because they result in

disallowed junction distances. This degeneracy can also be overcome by measuring RDCs under different alignment conditions to achieve independent RDC sets [243,244].

By comparing the two principal order tensor parameters (ϑ and η) for each helix, not only structural but dynamic information can be determined [242] (Figure 1.8). For example, if two helices connected by a bulge are completely rigid with respect to one another, they should feel the same degree and asymmetry of alignment: ϑ and η values for each helix would be equal. However, if one helix is dynamic relative to the other, the degree of alignment felt by the dynamic helix will be attenuated and proportional to the amplitude of the interhelical motions. A measure of this attenuation is represented by the internal degree of order ($\vartheta_{int} = \vartheta_i/\vartheta_j$; $\vartheta_i < \vartheta_j$), which reports on interhelical dynamics occurring on the sub-millisecond timescale. The ϑ_{int} value ranges from 0 for maximum interhelical motion to 1 for complete interhelical rigidity. The asymmetry parameter provides insight into the asymmetry of interhelical motions. For example, suppose an RNA has a $\vartheta_{int} < 1$ and the dynamic helix has a larger η value than the helix that dominates alignment. This suggests interhelical dynamics are present that are not isotropic but rather have a specific directionality. However, RDC interpretation is complicated when both fragments are of equal size. When a single domain does not dominate alignment, interhelical dynamics may go undetected due to motional coupling. Al-Hashimi and coworkers have developed techniques to overcome the issue of motional couplings [153,154] and this issue is discussed in-depth in Chapters 3 and 4 [245].

1.3.4 Spin Relaxation

The vector sum of all of the nuclear magnets in an NMR sample is referred to as the net magnetization. At thermal equilibrium, the net magnetization is oriented along the $+z$ direction, parallel to the external magnetic field. Upon application of a radiofrequency pulse, the magnetization is tipped away from the z axis toward the transverse plane. After the pulse, the net magnetization returns to its equilibrium position. The rate of this spin relaxation back to equilibrium is of particular importance for various reasons. It determines how long of a delay to use between NMR data acquisitions, determines the linewidth of NMR resonances, and is the basis for the nuclear Overhauser effect (NOE), which is commonly used in biomolecule structure determination. Finally, the relaxation

back to equilibrium is ultimately caused by dynamic motions, so measurement of these rates gives us detailed information about molecular motions at atomic-level resolution.

Relaxation Theory

The rate at which magnetization relaxes along the *z* direction is called longitudinal relaxation (R_1), while the rate at which magnetization relaxes due to loss of coherence along the transverse plane is called transverse relaxation (R_2). Rates R_1 and R_2 describe the auto-relaxation of individual nuclei, while the NOE describes the cross-relaxation rate of dipolar-coupled spins. All three parameters are measured for individual nuclei and report on the dynamics experienced by that particular nucleus on the ps-ns timescale, including both overall and local motions. For spin ½ nuclei, such as ^{1}H, ^{15}N, and ^{13}C, conformational dynamics on timescales faster than overall tumbling affect these relaxation rates by stochastically modulating the CSA and dipole-dipole Hamiltonians. The dipole-dipole interaction originates from the modulation of the nuclear magnetic flux between the spins of two nuclei as the internuclear vector rotates relative to the magnetic field. Similarly, the CSA interaction is a result of shielding variations experienced by a nucleus due to a fluctuation in its local magnetic field as a result of reorientation relative to the magnetic field. If a molecule is completely rigid, the CSA and dipole-dipole interactions will be modulated similarly across the entire molecule due to overall tumbling, and every nucleus will experience similar relaxation rates. Thus, if different relaxation rates are observed for particular nuclei in a molecule, they can be generally attributed to local flexibility. Assuming no chemical exchange and that the CSA tensor is axially symmetric with the principal axis collinear with the bond vector, the following equations represent the relaxation constants in a system of spins *IS* [246,247]:

$$R_1 = \frac{d^2}{4}\left[J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)\right] + c^2 J(\omega_S) \tag{1.5}$$

$$R_2 = \frac{d^2}{8}\left[4J(0) + J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I) + 6J(\omega_I + \omega_S)\right] +$$

$$\frac{c^2}{6}\left[4J(0) + 3J(\omega_s)\right] + R_{ex} \tag{1.6}$$

$$NOE = 1 + \frac{d^2}{4R_1}\frac{\gamma_S}{\gamma_N}\left[6J(\omega_I + \omega_S) - J(\omega_I - \omega_S)\right] \tag{1.7}$$

$$d = \frac{\mu_0 h \gamma_I \gamma_S}{8\pi^2 \langle r_{IS}^3 \rangle}, c = \frac{\Delta\sigma\omega_S}{\sqrt{3}} \tag{1.8}$$

where μ_o is the permeability of a vacuum, h is Planck's constant, r_{IS} is the IS bond length, and $\Delta\sigma$ is the CSA of the S spin. γ_I and γ_S are the gyromagnetic ratios and ω_I and ω_S are the Larmor frequency of I and S, respectively. Here, $J(\omega)$ is the spectral density function. The above equations are valid for a ^{1}H-^{15}N spin system, where the CSA tensor is axially symmetric. Solving these equations requires knowledge of the values of r_{IS} and $\Delta\sigma$.

Relaxation Analysis

Measured relaxation rates are generally interpreted through density mapping to determine $J(\omega)$ [248,249] or by using a program such as ModelFree to fit relaxation data to a "model-free" functional form of $J(\omega)$ that contains a limited number of free parameters [250]. The former approach allows for the complete determination of the correlation function, C(t), which describes both the overall and internal motions of the IS bond vector of interest:

$$J(\omega) = 2 \int_0^\infty C(t) \cos(\omega t)\, dt \tag{1.9}$$

The ModelFree approach developed by Lipari and Szabo assumes that internal and overall motions are uncorrelated, allowing the spectral density function to be divided into two terms [251]:

$$J(\omega) = 2 \int_0^\infty C_I(t) C_O(t) \cos(\omega t)\, dt \tag{1.10}$$

Here, C_I and C_O are the correlation functions describing internal and overall motions, respectively. With the separation of internal and overall motions, the internal motions are now measured relative to the molecular frame as opposed to the laboratory frame. Also, if isotropic tumbling can be assumed, the correlation function of overall motion can be simplified to the following:

$$C_O(t) = \frac{1}{5} e^{-6\tau_m} \tag{1.12}$$

where τ_m is the overall correlation time. This approximation holds true for globular proteins and small RNAs (~30 nt) that can be modeled on the rotation of a sphere. However, RNAs with elongated helices are better modeled as the diffusion of a rod.

Under the ModelFree formalism, C_I will have a value of 1 at time 0, and at infinite time, C_I is equal to the square of the generalized order parameter (S^2), which is

dependent on the range of angular motion sampled by the bond vector. The internal correlation can be rewritten as [251]:

$$C_I(t) = S^e + (1 - S^2)e^{-t/\tau} \tag{1.13}$$

where τ represents the timescale associated with the angular reorientation. Thus, a rigid bond vector (relative to the molecular frame) would have an internal correlation function of 1 for all times ($0 \rightarrow \infty$). Conversely, a dynamic bond vector will decay to S^2 at infinite time. S^2 ranges from 0 to 1 for completely isotropic and rigid motions, respectively.

With this simplification, it is possible to fit measured spin relaxation data to the function and solve for the parameters describing the amplitude and timescale of motion for individual bond vectors. Clore and coworkers later extended this approach by introducing two correlation times (fast and slow) for situations where a single correlation time is insufficient [252].

Both analyses have their limitations. The spectral density formalism is very time-consuming, may over interpret relaxation data, and can be insensitive to certain motions. However, the assumptions underlying the more simplistic ModelFree formalism do not hold true for all biomolecules. Most pertinent to the study of RNA is the "decoupling approximation" which requires that internal and overall motions are not correlated. This approximation is commonly violated by RNAs with interhelical dynamics (discussed below), limiting the utility of ModelFree in the analysis of RNA dynamics.

[13]C Relaxation Measurements in Nucleic Acids

In the study of proteins, ^{15}N relaxation parameters are commonly measured to probe internal dynamics on the ps-ns timescale. This approach is not as useful for the study of dynamics in nucleic acids because 1) there are fewer nitrogen nuclei to measure overall and 2) of the measurable nitrogens, most are located in helical regions. In dynamic apical loops and bulges, the imino hydrogens exchange readily with water and are not observable in NMR spectra. Thus, while nitrogen relaxation measurements are useful for revealing interhelical dynamics, they fail to probe the most dynamic regions of an RNA. To fully characterize the dynamics of RNA, the relaxation rates of a more abundant nuclei, such as ^{13}C, must be measured. However, the measurement of carbon relaxation in RNA presents additional complications.

32

Compared to proteins, RNA has a much smaller spectral dispersion, particularly in the sugar region, resulting in peak overlap. Additionally, unlike imino groups in proteins, RNA carbons are not isolated spin systems. With the exception of guanine C8 and adenosine C2, the measurement and interpretation of RNA carbon relaxation data is complicated by the presence of C-C interactions and asymmetric CSA tensors [253]. Finally, analysis of relaxation data of RNA is further complicated by the motional coupling of internal and overall motions. If large interhelical motions of an RNA occur on a timescale similar to that of overall rotational diffusion, the approximation is violated and makes the disentanglement of the collection of motions intractable. To overcome the latter complication, Al-Hashimi and coworkers devised an elongation strategy wherein one helix of a two-domain RNA is elongated by twenty-two base pairs that, through strategic

Figure 1.9: Decoupling of global and internal motions by helix elongation. (A) Global motions lead to a coupled change in the rotational diffusion tensor. (B) Elongation of one domain orients S_{zz} (and D_{zz}) parallel to the long axis, effectively decoupling motional modes.

isotopic labeling, are NMR-invisible (Figure 1.9) [153]. The elongated stem dominates overall tumbling, making the diffusion tensor less sensitive to collective motions of the smaller domain. Incidentally, the increased size of the RNA also decreases the tumbling rate, which widens the timescale sensitivity of spin relaxation measurements. Anchoring alignment of an RNA to a single helical axis also simplifies the computation of NMR parameters from molecular dynamics (MD) trajectories in such as way that they can be directly compared with experimental NMR data [254,255].

For RNA, longitudinal (R_1) and rotating frame ($R_{1\rho}$) relaxation rates are usually measured for C2, C5, C6, C8, and C1' carbons. Relaxation rates are determined by fitting peak intensities to a mono-exponential decay. Measured $R_{1\rho}$ values have R_1 contributions due to off-resonance effects and must be corrected:

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta \qquad (1.14)$$

where θ=arctan(ω_{SL}/Ω) is the effective tilt angle of the spin lock field from the static field, ω_{SL} is the amplitude of the spin lock field (in Hz) , and Ω is the resonance offset from the spin lock carrier frequency (in Hz).

1.3.5 Chemical Exchange

Spin relaxation measurements are very useful when studying the dynamics of a biomolecule on the ps-ns timescale, but slower dynamic processes (μs-s) will go undetected. Dynamic processes occurring at these timescales include base pair opening [31], ligand binding [256], repuckering of the sugar moiety [32], and base flipping [33]. These slow motions increase the relaxation rate of transverse magnetization (R_2) due to changes in the local magnetic environment through a relaxation mechanism known as chemical exchange. To probe dynamics at these slower timescales, line shape analysis, *zz*-exchange techniques, Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion, and $R_{1\rho}$ relaxation dispersion experiments are commonly used (for reviews see references [207-209]).

In CPMG experiments, a train of 180° pulses are used to refocus magnetization in the transverse plane. Monitoring the effective R_2 rate as a function of the delay between the 180° pulses reveals information about the chemical exchange kinetics. CPMG experiments are commonly used to measure slow motions in proteins but its application in the study of RNA dynamics is complicated due to artifacts introduced by scalar couplings between adjacent carbons. To circumvent this obstacle, alternative labeling schemes that remove these couplings have been created. For example, Johnson and coworkers introduced a labeling scheme in which only C2' and C4' are isotopically labeled, eliminating $^1J_{CC}$ effects in the sugar moiety [32]. A labeling scheme that uses a 2'-^{13}C-methoxy uridine was developed by Kreutz and coworkers and used to study the folding of a multi-stable RNA by CPMG [257]. However, site-specific labeling often

requires the preparation of multiple samples and special labeling techniques are not readily commercially available.

$R_{1\rho}$ experiments lock the magnetization along the direction of an effective field by application of a radiofrequency pulse, or spin lock. These experiments are usually limited to studying exchange processes on the microsecond timescale, because traditional R_{1p} experiments are limited by the lower bound of the spin lock (~1kHz). Below this field strength, conventional decoupling techniques fail to suppress scalar coupling evolution and off-resonance effects [258,259]. Al-Hashimi and coworkers recently developed a ^{13}C R_{1p} experiment that extends the measureable timescale of motions to milliseconds by using weak radiofrequency fields [219]. The experiment utilizes selective Hartman-Hahn polarization transfers to excite specific spins of interest, thus avoiding carbon-carbon scalar coupling and collects data in a 1D manner. Care must be taken to avoid Hartman-Hahn matching conditions by choosing appropriate experimental parameters, such as the spin lock power and offset.

^{13}C R_{1p} Relaxation Dispersion Theory

The relaxation dispersion experiment monitors chemical exchange, a phenomenon that occurs when a chemical shift exchanges between chemically-distinct states, as depicted for a simple 2-state model in Figure 1.10A and by the following:

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} = B \qquad (1.15)$$

where k_1 and k_{-1} are the rate constants for the forward and reverse reaction, respectively. Exchange between two states increases the relaxation rate of transverse magnetization, so measured $R_{1\rho}$ values are actually a combination of both R_2 and R_{ex} relaxation rates:

$$R_{1\rho} = R_1 cos^2\theta + (R_2 + R_{ex})sin^2\theta \qquad (1.16)$$

$$\theta = \arctan\left(\frac{\omega_1}{\Omega}\right) \qquad (1.17)$$

Here, R_2 is relaxation due to normal relaxation mechanisms (dipole-dipole, CSA, etc.), while R_{ex} is relaxation due to chemical exchange; ω_1 is the spin lock power and Ω is the resonance offset from the spin lock carrier frequency. Monitoring the $R_{1\rho}$ rate as a function of the strength and offset of the spin lock reveals the rate of exchange.
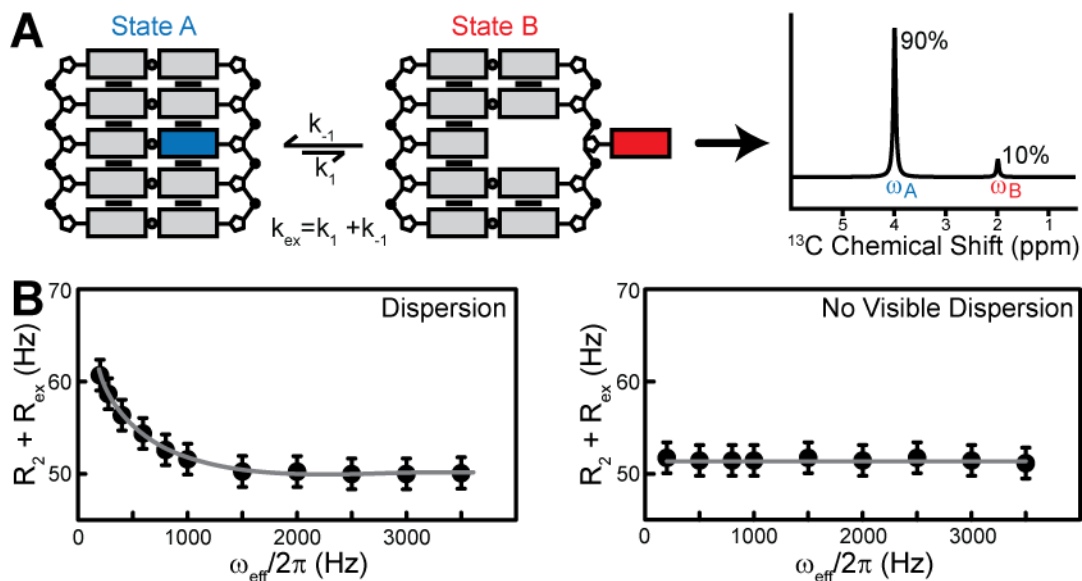
Figure 1.10: Measurement of chemical exchange. (A) Exchange between two chemically distinct states A and B. (B) Example on-resonance profiles for regimes in which there is dispersion (*left*) and in which there is no visible dispersion (*right*).

The exchange rate in equation 1.16 (R_{ex}) can be described by the Bloch-McConnell equations, and simplifications to the solutions of these equations have been derived for specific exchange conditions, such as asymmetric or fast exchange [260]. When exchange is sufficiently fast ($k_{ex} \gg \Delta\omega$), the following approximation holds true:

$$R_{ex} = sin^2\theta \frac{\phi_{ex}k_{ex}}{k_{ex}^2 + \omega_{eff}^2}$$

$$\omega_{eff} = \sqrt{(\omega_1^2 + \Omega^2)} \ , \qquad \phi_{ex} = p_a p_b \Delta\omega^2 \qquad\qquad (1.17)$$

$$\Delta\omega = \Omega_b - \Omega_a \ , \qquad \Omega = p_a\Omega_a + p_b\Omega_b$$

In the first equation, k_{ex} is the rate of the exchange process ($k_{ex} = k_1 + k_{-1}$) and ω_{eff} is the effective spin lock power at Ω, the population-weighted resonance offset from the spin lock frequency (the observed resonance). It is generally assumed that $p_a > p_b$, where p_a and p_b are the populations of state a and b, respectively, and $p_a + p_b = 1$. The chemical shift difference between the two states is represented by $\Delta\omega$. At the fast exchange limit, R_{1p} is not dependent on the individual parameters p_a, p_b, and $\Delta\omega$ but on their combination (ϕ_{ex}). In this regime, p_a, p_b, and $\Delta\omega$ cannot be determined without additional experimental information.

A commonly used approximation to the solution of the Bloch-McConnell equations under the influence of a spin lock field derived by Trott and Palmer is used when the populations of the two states are highly skewed or asymmetric ($p_a \gg p_b$) [261]. Here, the lowly populated state is assumed to have little effect on the population-weighted resonance frequency (Ω), so Ω is assumed to be equal to Ω_a, the chemical shift of the highly populated state a. Similarly, $\omega_{eff} \sim \omega_a$, the effective spin lock at Ω_a. This estimation is valid for all timescales. Thus, the relaxation rate constant becomes:

$$R_{ex} = sin^2\theta \frac{p_a p_b \Delta\omega^2 k_{ex}}{\omega_a^2 \omega_b^2 / \omega_{eff}^2 + k_{ex}^2} \approx sin^2\theta \frac{p_a p_b \Delta\omega^2 k_{ex}}{(\Delta\omega + \Omega)^2 + \omega_1^2 + k_{ex}^2} \tag{1.18}$$

The most accurate approximation of the relaxation rate uses Laguerre's method of polynomial root finding to derive the following equation [260]:

$$R_{ex} = sin^2\theta \frac{p_a p_b \Delta\omega^2 k_{ex}}{\omega_a^2 \omega_b^2 / \omega_{eff}^2 + k_{ex}^2 - sin^2\theta p_a p_b \Delta\omega^2 \left(1 + \frac{2k_{ex}^2(p_a\omega_a^2 + p_b\omega_b^2)}{\omega_a^2 \omega_b^2 + \omega_e^2 k_{ex}^2}\right)} \tag{1.19}$$

Although equation 1.19 appears to be quite complex, there are only three unknowns for which to solve for: p_a, k_{ex}, and $\Delta\omega$. Additional equations have been derived for other exchange regimes, including those with more than two states. To determine which equation best describes a specific exchange process, a model comparison of the goodness-of-fit must be completed using a model criterion such as the F-test or Akaike method.

Data Measurement and Analysis

Before $R_{1\rho}$ experiments are completed, it is necessary to calculate Hartman-Hahn transfers to avoid unwanted magnetization transfers to neighboring carbons. These calculations require knowledge of the chemical shifts of the carbon of interest and any scalar-coupled carbons. The maximum efficiency of a transfer between scalar-coupled spins I and S is given by the following [262]:

$$A_{HAHA} = \left(1 + \left(\frac{\omega_{eff,I} - \omega_{eff,S}}{J_{IS}(1 + \cos(\theta_I - \theta_S))}\right)^2\right)^{-1} \tag{1.20}$$

Here, $w_{eff,x}$ is the effective spin lock at spin x, J_{IS} is the scalar coupling constant of spins I and S, and θ_x is the tip angle of the magnetization of spin x with respect to the external magnetic field.

If the measured $R_{1\rho}$ rate changes as a function of the strength and offset of the spin lock, this is called relaxation dispersion and is an indication that there is chemical exchange (Figure 1.10B). However, it should be noted that the absence of dispersion does not unequivocally prove there is no chemical exchange. In fact, there is a limited window in which exchange is measureable using the $R_{1\rho}$ dispersion experiments developed by Al-Hashimi et. al. [219]. If the chemical shift difference between the two states is too small or the exchange rate is too slow or too fast, relaxation dispersion may not be observed.
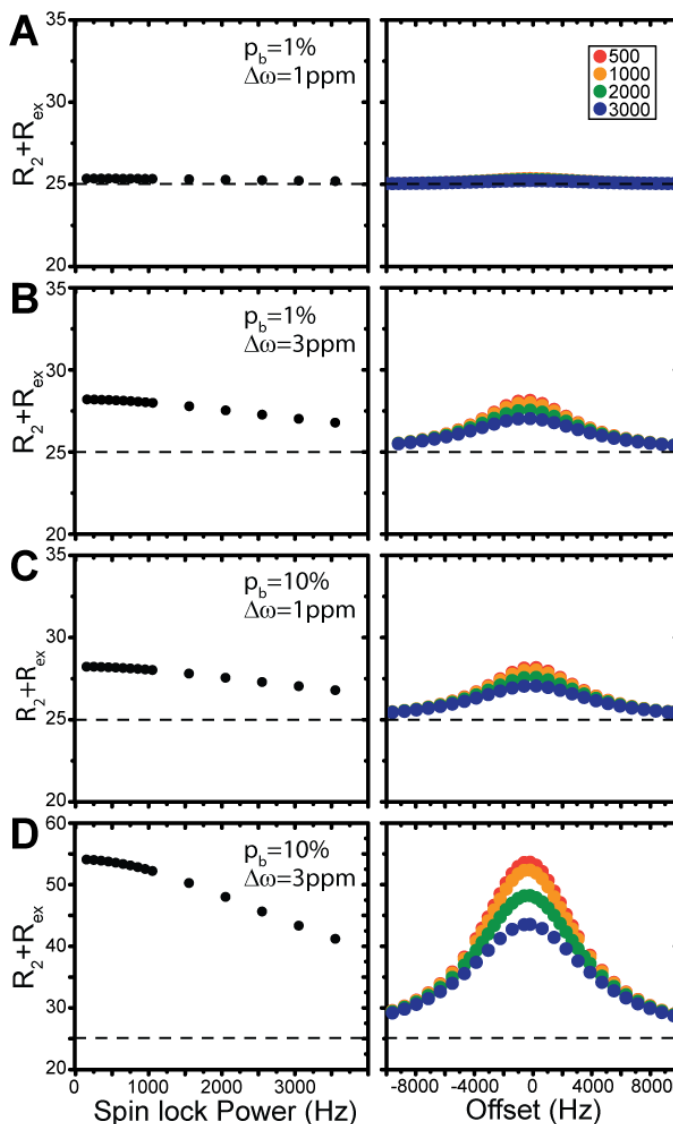
Figure 1.11: Simulated dispersion data near the fast exchange regime. (A-D) Simulated $R_2 + R_{ex}$ is plotted as a function of spin lock power (in Hz) on the left and as a function of offset and spin lock power (in Hz) on the right. Two values for population (10%, 1%) and chemical shift difference (1ppm, 3ppm) were used. See inset for key.

This is demonstrated in simulations of R_{2eff} (R_2+R_{ex}) data near the fast exchange limit (Figure 1.11). In these simulations, $k_{ex} = 25000$ Hz (40μs), $R_2 = 25$Hz, and combinations of different populations (p_b=1%, 10%) and chemical shift differences (1ppm, 3ppm) are used. Simulations show that in the fast exchange regime, exchange will likely go undetected if the population is less than 1% and the chemical shift difference is less than 1ppm, while the easiest exchange regime to detect and measure accurately would possess a large difference in the population and chemical shift of the two states (Figure 1.11A and D). The final two regimes in Figure 1.11B and C, would require the measurement of a small change in R_{2eff} (about 3Hz), which is on the same order of magnitude of R_{2eff} measurement error, making accurate measurement of the exchange parameters difficult. Finally, regime B and C have very similar ϕ_{ex} values, and the plots show why determination of the individual parameters of ϕ_{ex} is usually impossible at the fast exchange regime.

If the populations of the dynamic process can be accurately determined, the difference in the Gibbs free energy between exchanging states can be calculated. For the two-state exchange showed in equation 1.15:

$$K_{eq} = \frac{[B]}{[A]} = \frac{p_b}{p_b}, \quad \Delta G = -RT \ln(K_{eq}) \tag{1.21}$$

where ΔG is the difference in the Gibbs free energy, T is the temperature at which the chemical exchange parameters were measured, and R is the gas constant.

The objective of this dissertation is to use a combination of the NMR techniques described above, molecular dynamics (MD), and mutagenesis to characterize the structural dynamics of free TAR. RDCs, motionally-decoupled ^{13}C relaxation (R_1, R_2), and [13]C $R_{1\rho}$ relaxation dispersion are used in combination to characterize global and local motions occurring on fast (ps to ns) and slow (μs to ms) timescales. The [13]C $R_{1\rho}$ relaxation dispersion experiment is used to characterized a lowly-populated structure of the loop that is otherwise "invisible" to traditional NMR techniques, representing the first-ever characterization of an "invisible" conformation of an RNA apical loop by NMR. Overall, this dissertation establishes a broadly useful approach for characterizing the dynamics of RNA loops and sheds light on the relationship between dynamics and biological function for the TAR element in the HIV genome.

1.4 References

[1] F.H. Crick, On protein synthesis. Symp Soc Exp Biol 12 (1958) 138-63.

[2] Baltimor.D, Rna-Dependent Synthesis of DNA by Virions of Mouse Leukemia Virus. Cold Spring Harbor Symposia on Quantitative Biology 35 (1970) 843-&.

[3] H.M. Temin, and S. Mizutani, Viral Rna-Dependent DNA Polymerase - Rna-Dependent DNA Polymerase in Virions of Rous Sarcoma Virus. Nature 226 (1970) 1211-&.

[4] W.A. Decatur, and M.J. Fournier, RNA-guided nucleotide modification of ribosomal and other RNAs. Journal of Biological Chemistry 278 (2003) 695-698.

[5] A. Serganov, and D.J. Patel, Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nat Rev Genet 8 (2007) 776-90.

[6] O. Bensaude, and G. Diribarne, 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. Rna Biology 6 (2009) 122-128.

[7] A. Wutz, M. Leeb, and P.A. Steffen, X chromosome inactivation sparked by non-coding RNAs. Rna Biology 6 (2009) 94-99.

[8] R.K. Montange, and R.T. Batey, Riboswitches: emerging themes in RNA structure and function. Annu Rev Biophys 37 (2008) 117-33.

[9] J.H.N. Yik, R.C. Chen, R. Nishimura, J.L. Jennings, A.J. Link, and Q. Zhou, Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. Molecular Cell 12 (2003) 971-982.

[10] B. Panning, J. Dausman, and R. Jaenisch, X chromosome inactivation is mediated by Xist RNA stabilization. Cell 90 (1997) 907-916.

[11] G. Storz, An expanding universe of noncoding RNAs. Science 296 (2002) 1260-3.

[12] A.F. Bompfunewerer, C. Flamm, C. Fried, G. Fritzsch, I.L. Hofacker, J. Lehmann, K. Missal, A. Mosig, B. Muller, S.J. Prohaska, B.M.R. Stadler, P.F. Stadler, A. Tanzer, S. Washietl, and C. Witwer, Evolutionary patterns of non-coding RNAs. Theory in Biosciences 123 (2005) 301-369.

[13] J.P. Simorre, P. Legault, A.B. Hangar, P. Michiels, and A. Pardi, A conformational change in the catalytic core of the hammerhead ribozyme upon cleavage of an RNA substrate. Biochemistry 36 (1997) 518-525.

[14] X.W. Zhuang, H. Kim, M.J.B. Pereira, H.P. Babcock, N.G. Walter, and S. Chu, Correlating structural dynamics and function in single ribozyme molecules. Science 296 (2002) 1473-1476.

[15] M.V. Rodnina, T. Daviter, K. Gromadski, and W. Wintermeyer, Structural dynamics of ribosomal RNA during decoding on the ribosome. Biochimie 84 (2002) 745-754.

[16] H.F. Noller, and A. Baucom, Structure of the 70 S ribosome: implications for movement. Biochemical Society Transactions 30 (2002) 1159-1161.

[17] F. Aboul-ela, J. Karn, and G. Varani, The Structure of the Human-Immunodeficiency-Virus Type-1 Tar RNA Reveals Principles of RNA Recognition By Tat Protein. Journal of Molecular Biology 253 (1995) 313-332.

[18] J. Noeske, H. Schwalbe, and J. Wohnert, Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain. Nucleic Acids Res 35 (2007) 5262-73.

[19] O.M. Ottink, S.M. Rampersad, M. Tessari, G.J. Zaman, H.A. Heus, and S.S. Wijmenga, Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined induced fit mechanism. Rna 13 (2007) 2202-12.

[20] M. Mandal, B. Boese, J.E. Barrick, W.C. Winkler, and R.R. Breaker, Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. Cell 113 (2003) 577-86.

[21] R. Rieder, K. Lang, D. Graber, and R. Micura, Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. Chembiochem 8 (2007) 896-902.

[22] P.B. Rupert, and A.R. Ferre-D'Amare, Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. Nature 410 (2001) 780-786.

[23] E. Tan, T.J. Wilson, M.K. Nahas, R.M. Clegg, D.M.J. Lilley, and T. Ha, A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. Proceedings of the National Academy of Sciences of the United States of America 100 (2003) 9308-9313.

[24] J.D. Puglisi, R. Tan, B.J. Calnan, A.D. Frankel, and J.R. Williamson, Conformation of the TAR RNA-arginine complex by NMR spectroscopy. Science 257 (1992) 76-80.

[25] J.R. Williamson, Induced fit in RNA-protein recognition. Nature Structural Biology 7 (2000) 834-837.

[26] K.S. Long, and D.M. Crothers, Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. Biochemistry 38 (1999) 10059-10069.

[27] M. Zacharias, and P.J. Hagerman, The Bend in RNA Created By the Transactivation Response Element Bulge of Human-Immunodeficiency-Virus Is Straightened By Arginine and By Tat-Derived Peptide. Proceedings of the National Academy of Sciences of the United States of America 92 (1995) 6052-6056.

[28] W.C. Boelens, E.J.R. Jansen, W.J. Vanvenrooij, R. Stripecke, I.W. Mattaj, and S.I. Gunderson, The Human U1 Snrnp-Specific U1a Protein Inhibits Polyadenylation of Its Own Premessenger Rna. Cell 72 (1993) 881-892.

[29] K.B. Hall, Interaction of Rna Hairpins with the Human U1a N-Terminal Rna-Binding Domain. Biochemistry 33 (1994) 10076-10088.

[30] Z. Shajani, G. Drobny, and G. Varani, Binding of U1A Protein Changes RNA Dynamics As Observed by (13)C NMR Relaxation Studies(,). Biochemistry 46 (2007) 5875-83.

[31] J.H. Lee, and A. Pardi, Thermodynamics and kinetics for base-pair opening in the P1 duplex of the Tetrahymena group I ribozyme. Nucleic Acids Research 35 (2007) 2965-2974.

[32] J.E. Johnson, and C.G. Hoogstraten, Extensive Backbone Dynamics in the GCAA RNA Tetraloop Analyzed Using C-13 NMR Spin Relaxation and Specific Isotope Labeling. Journal of the American Chemical Society 130 (2008) 16757-16769.

[33] H. Blad, N.J. Reiter, F. Abildgaard, J.L. Markley, and S.E. Butcher, Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. Journal of Molecular Biology 353 (2005) 540-555.

[34] J.B. Goforth, S.A. Anderson, C.P. Nizzi, and R.S. Eisenstein, Multiple determinants within iron-responsive elements dictate iron regulatory protein binding and regulatory hierarchy. Rna-a Publication of the Rna Society 16 (2010) 154-169.

[35] S. Richter, Y.H. Ping, and T.M. Rana, TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. Proc Natl Acad Sci U S A 99 (2002) 7928-33.

[36] A. Roth, W.C. Winkler, E.E. Regulski, B.W.K. Lee, J. Lim, I. Jona, J.E. Barrick, A. Ritwik, J.N. Kim, R. Welz, D. Iwata-Reuyl, and R.R. Breaker, A riboswitch selective for the queuosine precursor preQ(1) contains an unusually small aptamer domain. Nature Structural & Molecular Biology 14 (2007) 308-317.

[37] J.H. Cate, A.R. Gooding, E. Podell, K.H. Zhou, B.L. Golden, C.E. Kundrot, T.R. Cech, and J.A. Doudna, Crystal structure of a group I ribozyme domain: Principles of RNA packing. Science 273 (1996) 1678-1685.

[38] F.L. Murphy, and T.R. Cech, Gaaa Tetraloop and Conserved Bulge Stabilize Tertiary Structure of a Group-I Intron Domain. Journal of Molecular Biology 236 (1994) 49-63.

[39] R.C. Spitale, A.T. Torelli, J. Krucinska, V. Bandarian, and J.E. Wedekind, The Structural Basis for Recognition of the PreQ(0) Metabolite by an Unusually Small Riboswitch Aptamer Domain. Journal of Biological Chemistry 284 (2009) 11012-11016.

[40] C.R. Woese, S. Winker, and R.R. Gutell, Architecture of Ribosomal-Rna - Constraints on the Sequence of Tetra-Loops. Proceedings of the National Academy of Sciences of the United States of America 87 (1990) 8467-8471.

[41] J. Srinivasan, J. Miller, P.A. Kollman, and D.A. Case, Continuum solvent studies of the stability of RNA hairpin loops and helices. Journal of Biomolecular Structure & Dynamics 16 (1998) 671-+.

[42] N. Spackova, and J. Sponer, Molecular dynamics simulations of sarcin-ricin rRNA motif. Nucleic Acids Research 34 (2006) 697-708.

[43] M. Molinaro, and I. Tinoco, Use of Ultra-Stable Uncg Tetraloop Hairpins to Fold Rna Structures - Thermodynamic and Spectroscopic Applications. Nucleic Acids Research 23 (1995) 3056-3063.

[44] C.C. Correll, I.G. Wool, and A. Munishkin, The two faces of the Escherichia coli 23 S rRNA sarcin/ricin domain: the structure at 1.11 A resolution. J Mol Biol 292 (1999) 275-87.

[45] A. Munishkin, and I.G. Wool, The ribosome-in-pieces: binding of elongation factor EF-G to oligoribonucleotides that mimic the sarcin/ricin and thiostrepton domains of 23S ribosomal RNA. Proc Natl Acad Sci U S A 94 (1997) 12280-4.

[46] C.D. Downey, J.L. Fiore, C.D. Stoddard, J.H. Hodak, D.J. Nesbitt, and A. Pardi, Metal ion dependence, thermodynamics, and kinetics for intramolecular docking of a GAAA tetraloop and receptor connected by a flexible linker. Biochemistry 45 (2006) 3664-73.

[47] Y. Endo, and I.G. Wool, The site of action of alpha-sarcin on eukaryotic ribosomes. The sequence at the alpha-sarcin cleavage site in 28 S ribosomal ribonucleic acid. J Biol Chem 257 (1982) 9054-60.

[48] T.P. Hausner, J. Atmadja, and K.H. Nierhaus, Evidence that the G2661 region of 23S rRNA is located at the ribosomal binding sites of both elongation factors. Biochimie 69 (1987) 911-23.

[49] C.J. Cheong, G. Varani, and I. Tinoco, Solution Structure of an Unusually Stable Rna Hairpin, 5'ggac(Uucg)Gucc. Nature 346 (1990) 680-682.

[50] G. Varani, C.J. Cheong, and I. Tinoco, Structure of an Unusually Stable Rna Hairpin. Biochemistry 30 (1991) 3280-3289.

[51] F.M. Jucker, H.A. Heus, P.F. Yip, E.H.M. Moors, and A. Pardi, A network of heterogeneous hydrogen bonds in GNRA tetraloops. Journal of Molecular Biology 264 (1996) 968-980.

[52] D.J. Kerwood, M.J. Cavaluzzi, and P.N. Borer, Structure of SL4 RNA from the HIV-1 packaging signal. Biochemistry 40 (2001) 14518-14529.

[53] E. Ennifar, A. Nikulin, S. Tishchenko, A. Serganov, N. Nevskaya, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov, and P. Dumas, The crystal structure of UUCG tetraloop. Journal of Molecular Biology 304 (2000) 35-42.

[54] C.C. Correll, and K. Swinger, Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 angstrom resolution. Rna-a Publication of the Rna Society 9 (2003) 355-363.

[55] F.H.T. Allain, and G. Varani, Structure of the P1 Helix from Group-I Self-Splicing Introns. Journal of Molecular Biology 250 (1995) 333-353.

[56] M. Akke, R. Fiala, F. Jiang, D. Patel, and A.G. Palmer, Base dynamics in a UUCG tetraloop RNA hairpin characterized by N-15 spin relaxation: Correlations with structure and stability. RNA 3 (1997) 702-709.

[57] P. Vallurupalli, and L.E. Kay, A suite of 2H NMR spin relaxation experiments for the measurement of RNA dynamics. J Am Chem Soc 127 (2005) 6893-901.

[58] L.G. Laing, and K.B. Hall, A model of the iron responsive element RNA hairpin loop structure determined from NMR and thermodynamic data. Biochemistry 35 (1996) 13586-13596.

[59] S. Flodell, J. Schleucher, J. Cromsigt, H. Ippel, K. Kidd-Ljunggren, and S. Wijmenga, The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. Nucleic Acids Research 30 (2002) 4803-4811.

[60] P.C.J. Haasnoot, F.T. Brederode, R.C.L. Olsthoorn, and J.F. Bol, A conserved hairpin structure in Alfamovirus and Bromovirus subgenomic promoters is required for efficient RNA synthesis in vitro. Rna-a Publication of the Rna Society 6 (2000) 708-716.

[61] T.J.M. Welting, S. Mattijssen, F.M.A. Peters, N.L. van Doorn, L. Dekkers, W.J. van Venrooij, H.A. Heus, L. Bonafe, and G.J.M. Pruijn, Cartilage-hair hypoplasia-associated mutations in the RNase MRP P3 domain affect RNA folding and ribonucleoprotein assembly. Biochimica Et Biophysica Acta-Molecular Cell Research 1783 (2008) 455-466.

[62] T. Kulinski, M. Olejniczak, H. Huthoff, L. Bielecki, K. Pachulska-Wieczorek, A.T. Das, B. Berkhout, and R.W. Adamiak, The apical loop of the HIV-1 TAR RNA hairpin is stabilized by a cross-loop base pair. J Biol Chem 278 (2003) 38892-901.

[63] S. Flodell, M. Petersen, F. Girard, J. Zdunek, K. Kidd-Ljunggren, J. Schleucher, and S. Wijmenga, Solution structure of the apical stem-loop of the human hepatitis B virus encapsidation signal. Nucleic Acids Res 34 (2006) 4449-57.

[64] F. Aboul-ela, J. Karn, and G. Varani, Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. Nucleic Acids Research 24 (1996) 3974-3981.

[65] N.V. Eldho, and K.T. Dayie, Internal bulge and tetraloop of the catalytic domain 5 of a group II intron ribozyme are flexible: Implications for catalysis. Journal of Molecular Biology 365 (2007) 930-944.

[66] L. Trantirek, E. Caha, P. Kaderavek, and R. Fiala, NMR (13)C-relaxation study of base and sugar dynamics in GCAA RNA hairpin tetraloop. J Biomol Struct Dyn 25 (2007) 243-52.

[67] E. Duchardt, and H. Schwalbe, Residue Specific Ribose and Nucleobase Dynamics of the cUUCGg RNA Tetraloop Motif by NMR 13C Relaxation. Journal of Biomolecular Nmr 32 (2005) 295-308.

[68] S.N. Richter, I. Frasson, and G. Palu, Strategies for inhibiting function of HIV-1 accessory proteins: a necessary route to AIDS therapy? Curr Med Chem 16 (2009) 267-86.

[69] B. Roca, Adverse drug reactions to antiretroviral medication. Front Biosci 14 (2009) 1785-92.

[70] C. Kuiken, B. Korber, and R.W. Shafer, HIV sequence databases. AIDS Rev 5 (2003) 52-61.

[71] J.M. Watts, K.K. Dang, R.J. Gorelick, C.W. Leonard, J.W. Bess, Jr., R. Swanstrom, C.L. Burch, and K.M. Weeks, Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460 (2009) 711-6.

[72] M. Stevens, E. De Clercq, and J. Balzarini, The regulation of HIV-1 transcription: molecular targets for chemotherapeutic intervention. Med Res Rev 26 (2006) 595-625.

[73] Y. Yamaguchi, T. Takagi, T. Wada, K. Yano, A. Furuya, S. Sugimoto, J. Hasegawa, and H. Handa, NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell 97 (1999) 41-51.

[74] T. Wada, T. Takagi, Y. Yamaguchi, A. Ferdous, T. Imai, S. Hirose, S. Sugimoto, K. Yano, G.A. Hartzog, F. Winston, S. Buratowski, and H. Handa, DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev 12 (1998) 343-56.

[75] J. Sodroski, C. Rosen, F. Wongstaal, S.Z. Salahuddin, M. Popovic, S. Arya, R.C. Gallo, and W.A. Haseltine, Trans-Acting Transcriptional Regulation of Human T-Cell Leukemia-Virus Type-Iii Long Terminal Repeat. Science 227 (1985) 171-173.

[76] J. Sodroski, R. Patarca, C. Rosen, F. Wongstaal, and W. Haseltine, Location of the Trans-Activating Region on the Genome of Human T-Cell Lymphotropic Virus Type-Iii. Science 229 (1985) 74-77.

[77] B.R. Cullen, Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. Cell 46 (1986) 973-982.

[78] M.A. Muesing, D.H. Smith, and D.J. Capon, Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. Cell 48 (1987) 691-701.

[79] A.P. Rice, and M.B. Mathews, Transcriptional but Not Translational Regulation of Hiv-1 by the Tat Gene-Product. Nature 332 (1988) 551-553.

[80] M.J. Selby, E.S. Bain, P.A. Luciw, and B.M. Peterlin, Structure, Sequence, and Position of the Stem Loop in Tar Determine Transcriptional Elongation by Tat through the Hiv-1 Long Terminal Repeat. Genes & Development 3 (1989) 547-558.

[81] S. Feng, and E.C. Holland, Hiv-1 Tat Trans-Activation Requires the Loop Sequence within Tar. Nature 334 (1988) 165-167.

[82] C. Dingwall, I. Ernberg, M.J. Gait, S.M. Green, S. Heaphy, J. Karn, A.D. Lowe, M. Singh, and M.A. Skinner, Hiv-1 Tat Protein Stimulates Transcription by Binding to a U-Rich Bulge in the Stem of the Tar Rna Structure. Embo Journal 9 (1990) 4145-4153.

[83] K.M. Weeks, and D.M. Crothers, Rna Recognition By Tat-Derived Peptides - Interaction in the Major Groove. Cell 66 (1991) 577-588.

[84] M.J. Churcher, C. Lamont, F. Hamy, C. Dingwall, S.M. Green, A.D. Lowe, P.J.G. Butler, M.J. Gait, and J. Karn, High-Affinity Binding of Tar Rna By the Human-Immunodeficiency- Virus Type-1 Tat Protein Requires Base-Pairs in the Rna Stem and Amino-Acid-Residues Flanking the Basic Region. Journal of Molecular Biology 230 (1993) 90-110.

[85] M.G. Cordingley, R.L. LaFemina, P.L. Callahan, J.H. Condra, V.V. Sardana, D.J. Graham, T.M. Nguyen, K. LeGrow, L. Gotlib, A.J. Schlabach, and et al., Sequence-specific interaction of Tat protein and Tat peptides with the transactivation-responsive sequence element of human immunodeficiency virus type 1 in vitro. Proc Natl Acad Sci U S A 87 (1990) 8985-9.

[86] M. Sumner-Smith, S. Roy, R. Barnett, L.S. Reid, R. Kuperman, U. Delling, and N. Sonenberg, Critical chemical features in trans-acting-responsive RNA are required for interaction with human immunodeficiency virus type 1 Tat protein. J Virol 65 (1991) 5196-202.

[87] C.H. Herrmann, and A.P. Rice, Specific Interaction of the Human-Immunodeficiency-Virus Tat Proteins with a Cellular Protein-Kinase. Virology 197 (1993) 601-608.

[88] C.H. Herrmann, and A.P. Rice, Lentivirus Tat Proteins Specifically Associate with a Cellular Protein-Kinase, Tak, That Hyperphosphorylates the Carboxyl-Terminal Domain of the Large Subunit of Rna-Polymerase .2. Candidate for a Tat Cofactor. Journal of Virology 69 (1995) 1612-1620.

[89] T. Pe'ery, Y. Ramanathan, Y.R. Zhu, J.M. Peng, N. Marshall, T. Marshall, B. Amendt, D.H. Price, and M.B. Mathews, HIV-1 TAT associates with a CTD kinase that is the human homolog of Drosophila P-TEFb. Faseb Journal 11 (1997) A1206-A1206.

[90] Y.R. Zhu, T. Peery, T.M. Peng, Y. Ramanathan, N. Marshall, T. Marshall, B. Amendt, M.B. Mathews, and D.H. Price, Transcription elongation factor P-TEFb is required for HIV-1 Tat transactivation in vitro. Genes & Development 11 (1997) 2622-2632.

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

[92] K. Fujinaga, D. Irwin, Y.H. Huang, R. Taube, T. Kurosu, and B.M. Peterlin, Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Molecular and Cellular Biology 24 (2004) 787-795.

[93] T. Wada, T. Takagi, Y. Yamaguchi, D. Watanabe, and H. Handa, Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. EMBO J 17 (1998) 7395-403.

[94] Y.H. Ping, and T.M. Rana, DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 Tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase

II and DSIF during transcription elongation. Journal of Biological Chemistry 276 (2001) 12951-12958.

[95] P.D. Bieniasz, T.A. Grdina, H.P. Bogerd, and B.R. Cullen, Recruitment of cyclin T1/P-TEFb to an HIV type I long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. Proceedings of the National Academy of Sciences of the United States of America 96 (1999) 7791-7796.

[96] B. Majello, G. Napolitano, A. Giordano, and L. Lania, Transcriptional regulation by targeted recruitment of cyclin-dependent CDK9 kinase in vivo. Oncogene 18 (1999) 4598-605.

[97] K.T. Jeang, H. Xiao, and E.K. Rich, Multifaceted activities of the HIV-1 transactivator of transcription, Tat. Journal of Biological Chemistry 274 (1999) 28837-28840.

[98] L. Ratner, W. Haseltine, R. Patarca, K.J. Livak, B. Starcich, S.F. Josephs, E.R. Doran, J.A. Rafalski, E.A. Whitehorn, K. Baumeister, L. Ivanoff, S.R. Petteway, M.L. Pearson, J.A. Lautenberger, T.S. Papas, J. Ghrayeb, N.T. Chang, R.C. Gallo, and F. Wongstaal, Complete Nucleotide-Sequence of the Aids Virus, Htlv-Iii. Nature 313 (1985) 277-284.

[99] B.R. Cullen, The Hiv-1 Tat Protein - an Rna Sequence-Specific Processivity Factor. Cell 63 (1990) 655-657.

[100] U. Mahlknecht, I. Dichamp, A. Varin, C. Van Lint, and G. Herbein, NF-kappaB-dependent control of HIV-1 transcription by the second coding exon of Tat in T cells. J Leukoc Biol 83 (2008) 718-27.

[101] M. Ma, and A. Nath, Molecular determinants for cellular uptake of Tat protein of human immunodeficiency virus type 1 in brain cells. J Virol 71 (1997) 2495-9.

[102] S.R. Bartz, and M. Emerman, Human immunodeficiency virus type 1 Tat induces apoptosis and increases sensitivity to apoptotic signals by up-regulating FLICE/caspase-8. J Virol 73 (1999) 1956-63.

[103] G.R. Campbell, J.D. Watkins, D. Esquieu, E. Pasquier, E.P. Loret, and S.A. Spector, The C terminus of HIV-1 Tat modulates the extent of CD178-mediated apoptosis of T cells. Journal of Biological Chemistry 280 (2005) 38376-38382.

[104] H. Xiao, C. Neuveut, M. Benkirane, and K.T. Jeang, Interaction of the second coding exon of Tat with human EF-1 delta delineates a mechanism for HIV-1-mediated shut-off of host mRNA translation. Biochem Biophys Res Commun 244 (1998) 384-9.

[105] M. Kuppuswamy, T. Subramanian, A. Srinivasan, and G. Chinnadurai, Multiple functional domains of Tat, the trans-activator of HIV-1, defined by mutational analysis. Nucleic Acids Res 17 (1989) 3551-61.

[106] J.M. Peloponese, Y. Collette, C. Gregoire, C. Bailly, D. Campese, E.F. Meurs, D. Olive, and E.P. Loret, Full peptide synthesis, purification, and characterization of six Tat variants - Differences observed between HIV-1 isolates from Africa and other continents. Journal of Biological Chemistry 274 (1999) 11473-11478.

[107] M.E. Garber, P. Wei, V.N. KewalRamani, T.P. Mayall, C.H. Herrmann, A.P. Rice, D.R. Littman, and K.A. Jones, The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. Genes & Development 12 (1998) 3512-3527.

[108] B. Ensoli, L. Buonaguro, G. Barillari, V. Fiorelli, R. Gendelman, R.A. Morgan, P. Wingfield, and R.C. Gallo, Release, Uptake, and Effects of Extracellular Human-Immunodeficiency-Virus Type-1 Tat Protein on Cell-Growth and Viral Transactivation. Journal of Virology 67 (1993) 277-287.

[109] R. Pierleoni, M. Menotta, A. Antonelli, C. Sfara, G. Serafini, S. Dominici, M.E. Laguardia, A. Salis, G. Damonte, L. Banci, M. Porcu, P. Monini, B. Ensoli, and M. Magnani, Effect of the redox state on HIV-1 tat protein multimerization and cell internalization and trafficking. Molecular and Cellular Biochemistry 345 (2010) 105-118.

[110] R.E. Kiernan, C. Vanhulle, L. Schiltz, E. Adam, H. Xiao, F. Maudoux, C. Calomme, A. Burny, Y. Nakatani, K.T. Jeang, M. Benkirane, and C. Van Lint, HIV-1 Tat transcriptional activity is regulated by acetylation. Embo Journal 18 (1999) 6106-6118.

[111] L.W. Deng, C. de la Fuente, P. Fu, L. Wang, R. Donnelly, J.D. Wade, P. Lambert, H. Li, C.G. Lee, and F. Kashanchi, Acetylation of HIV-1 Tat by CBP/P300 increases transcription of integrated HIV-1 genome and enhances binding to core histones. Virology 277 (2000) 278-295.

[112] K. Kaehlcke, A. Dorr, C. Hetzer-Egger, V. Kiermer, P. Henklein, M. Schnoelzer, E. Loret, P.A. Cole, E. Verdin, and M. Ott, Acetylation of Tat defines a CyclinT1-independent step in HIV transactivation. Molecular Cell 12 (2003) 167-176.

[113] M. Ott, A. Dorr, C. Hetzer-Egger, K. Kaehlcke, M. Schnolzer, P. Henklein, P. Cole, M.M. Zhou, and E. Verdin, Tat acetylation: a regulatory switch between early and late phases in HIV transcription elongation. Novartis Found Symp 259 (2004) 182-93; discussion 193-6, 223-5.

[114] I. D'Orso, and A.D. Frankel, Tat acetylation modulates assembly of a viral-host RNA-protein transcription complex. Proceedings of the National Academy of Sciences of the United States of America 106 (2009) 3101-3106.

[115] R. Berro, K. Kehn, C. de la Fuente, A. Pumfery, R. Adair, J. Wade, A.M. Colberg-Poley, J. Hiscott, and F. Kashanchi, Acetylated tat regulates human immunodeficiency virus type 1 splicing through its interaction with the splicing regulator p32. Journal of Virology 80 (2006) 3189-3204.

[116] P. Bayer, M. Kraft, A. Ejchart, M. Westendorp, R. Frank, and P. Rosch, Structural Studies of Hiv-1 Tat Protein. Journal of Molecular Biology 247 (1995) 529-535.

[117] S.E.C. Koken, A.E. Greijer, K. Verhoef, J. Vanwamel, A.G. Bukrinskaya, and B. Berkhout, Intracellular Analysis of in-Vitro Modified Hiv Tat Protein. Journal of Biological Chemistry 269 (1994) 8366-8375.

[118] J.M. Peloponese, C. Gregoire, S. Opi, D. Esquieu, J. Sturgis, E. Lebrun, E. Meurs, Y. Collette, D. Olive, A.M. Aubertin, M. Witvrow, C. Pannecouque, E. De Clercq, C. Bailly, J. Lebreton, and E.P. Loret, H-1-C-13 nuclear magnetic resonance assignment and structural characterization of HIV-1 Tat protein. Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences 323 (2000) 883-894.

[119] C. Gregoire, J.M. Peloponese, D. Esquieu, S. Opi, G. Campbell, M. Solomiac, E. Lebrun, J. Lebreton, and E.P. Loret, Homonuclear H-1-NMR Assignment and Structural Characterization of Human Immunodeficiency Virus Type 1 Tat Mal Protein. Biopolymers 62 (2001) 324-335.

[120] J.D. Watkins, G.R. Campbell, H. Halimi, and E.P. Loret, Homonuclear H-1 NMR and circular dichroism study of the HIV-1 Tat Eli variant. Retrovirology 5 (2008) -.

[121] S. Shojania, and J.D. O'Neil, HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1-72) by NMR spectroscopy. J Biol Chem 281 (2006) 8347-56.

[122] M.C. Re, M. Vignoli, G. Furlini, D. Gibellini, V. Colangeli, F. Vitone, and M. La Placa, Antibodies against full-length Tat protein and some low-molecular-weight Tat-peptides correlate with low or undetectable viral load in HIV-1 seropositive patients. J Clin Virol 21 (2001) 81-9.

[123] S. Butto, V. Fiorelli, A. Tripiciano, M.J. Ruiz-Alvarez, A. Scoglio, F. Ensoli, M. Ciccozzi, B. Collacchi, M. Sabbatucci, A. Cafaro, C.A. Guzman, A. Borsetti, A. Caputo, E. Vardas, M. Colvin, M. Lukwiya, G. Rezza, B. Ensoli, and T.M.S. Grp, Sequence conservation and antibody cross-recognition of clade B human immunodeficiency virus (HIV) type 1 Tat protein in HIV-1-infected Italians, Ugandans, and South Africans. Journal of Infectious Diseases 188 (2003) 1171-1180.

[124] G.R. Campbell, D. Senkaali, J. Watkins, D. Esquieu, S. Opi, D.L. Yirrell, P. Kaleebu, and E.P. Loret, Tat mutations in an African cohort that do not prevent transactivation but change its immunogenic properties. Vaccine 25 (2007) 8441-8447.

[125] B.M. Peterlin, and D.H. Price, Controlling the elongation phase of transcription with P-TEFb. Molecular Cell 23 (2006) 297-305.

[126] A.A. Michels, A. Fraldi, Q. Li, T.E. Adamson, F. Bonnet, V.T. Nguyen, S.C. Sedore, J.P. Price, D.H. Price, L. Lania, and O. Bensaude, Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. Embo Journal 23 (2004) 2608-2619.

[127] J.H.N. Yik, R.C. Chen, A.C. Pezda, C.S. Samford, and Q. Zhou, A human immunodeficiency virus type 1 tat-like arginine-rich RNA-binding domain is essential for HEXIM1 to inhibit RNA polymerase II transcription through 7SK snRNA-Mediated inactivation of P-TEFb. Molecular and Cellular Biology 24 (2004) 5094-5105.

[128] B.J. Krueger, K. Varzavand, J.J. Cooper, and D.H. Price, The Mechanism of Release of P-TEFb and HEXIM1 from the 7SK snRNP by Viral and Cellular Activators Includes a Conformational Change in 7SK. Plos One 5 (2010) -.

[129] J.M. Peng, Y.R. Zhu, J.T. Milton, and D.H. Price, Identification of multiple cyclin subunits of human P-TEFb. Genes & Development 12 (1998) 755-762.

[130] J. Wimmer, K. Fujinaga, R. Taube, T.P. Cujec, Y.R. Zhu, J.M. Peng, D.H. Price, and B.M. Peterlin, Interactions between Tat and TAR and human immunodeficiency virus replication are facilitated by human cyclin T1 but not cyclins T2a or T2b. Virology 255 (1999) 182-189.

[131] J. Karn, Tackling Tat. Journal of Molecular Biology 293 (1999) 235-254.

[132] D. Ivanov, Y.T. Kwak, E. Nee, J. Guo, L.F. Garcia-Martinez, and R.B. Gaynor, Cyclin T1 domains involved in complex formation with Tat and TAR RNA are critical for tat-activation. Journal of Molecular Biology 288 (1999) 41-56.

[133] T.H. Tahirov, N.D. Babayeva, K. Varzavand, J.J. Cooper, S.C. Sedore, and D.H. Price, Crystal structure of HIV-1 Tat complexed with human P-TEFb. Nature 465 (2010) 747-U2.

[134] S. Richter, H. Cao, and T.M. Rana, Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. Biochemistry 41 (2002) 6391-7.

[135] J.D. Puglisi, L. Chen, A.D. Frankel, and J.R. Williamson, Role of Rna Structure in Arginine Recognition of Tar Rna. Proceedings of the National Academy of Sciences of the United States of America 90 (1993) 3680-3684.

[136] C.A. Smith, V. Calabro, and A.D. Frankel, An RNA-Binding chameleon. Molecular Cell 6 (2000) 1067-1076.

[137] J. Zhang, N. Tamilarasu, S.W. Hwang, M.E. Garber, I. Huq, K.A. Jones, and T.M. Rana, HIV-1 TAR RNA enhances the interaction between Tat and cyclin T1. Journal of Biological Chemistry 275 (2000) 34314-34319.

[138] C.A. Rosen, J.G. Sodroski, and W.A. Haseltine, The Location of Cis-Acting Regulatory Sequences in the Human T-Cell Lymphotropic Virus Type-Iii (Htlv-Iii/Lav) Long Terminal Repeat. Cell 41 (1985) 813-823.

[139] B. Berkhout, and K.T. Jeang, Detailed Mutational Analysis of Tar Rna - Critical Spacing between the Bulge and Loop Recognition Domains. Nucleic Acids Research 19 (1991) 6169-6176.

[140] A. Jakobovits, D.H. Smith, E.B. Jakobovits, and D.J. Capon, A Discrete Element 3' of Human Immunodeficiency Virus-1 (Hiv-1) and Hiv-2 Messenger-Rna Initiation Sites Mediates Transcriptional Activation by an Hiv Trans Activator. Molecular and Cellular Biology 8 (1988) 2555-2561.

[141] C.E. Pritchard, J.A. Grasby, F. Hamy, A.M. Zacharek, M. Singh, J. Karn, and M.J. Gait, Methylphosphonate Mapping of Phosphate Contacts Critical for Rna Recognition by the

Human-Immunodeficiency-Virus Tat and Rev Proteins. Nucleic Acids Research 22 (1994) 2592-2600.

[142] F. Hamy, U. Asseline, J. Grasby, S. Iwai, C. Pritchard, G. Slim, P.J.G. Butler, J. Karn, and M.J. Gait, Hydrogen-Bonding Contacts in the Major Groove Are Required for Human-Immunodeficiency-Virus Type-1 Tat Protein Recognition of Tar Rna. Journal of Molecular Biology 230 (1993) 111-123.

[143] J.A. Ippolito, and T.A. Steitz, A 1.3-angstrom resolution crystal structure of the HIV-1 trans- activation response region RNA stem reveals a metal ion- dependent bulge conformation. Proceedings of the National Academy of Sciences of the United States of America 95 (1998) 9819-9824.

[144] T.E. Edwards, and S.T. Sigurdsson, EPR spectroscopic analysis of TAR RNA-metal ion interactions. Biochem Biophys Res Commun 303 (2003) 721-5.

[145] A. Casiano-Negroni, X. Sun, and H.M. Al-Hashimi, Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. Biochemistry 46 (2007) 6525-35.

[146] C. Faber, H. Sticht, K. Schweimer, and P. Rosch, Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. J Biol Chem 275 (2000) 20660-6.

[147] Z. Du, K.E. Lind, and T.L. James, Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. Chem Biol 9 (2002) 707-12.

[148] B. Davis, M. Afshar, G. Varani, A.I. Murchie, J. Karn, G. Lentzen, M. Drysdale, J. Bower, A.J. Potter, I.D. Starkey, T. Swarbrick, and F. Aboul-ela, Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". J Mol Biol 336 (2004) 343-56.

[149] A.I. Murchie, B. Davis, C. Isel, M. Afshar, M.J. Drysdale, J. Bower, A.J. Potter, I.D. Starkey, T.M. Swarbrick, S. Mirza, C.D. Prescott, P. Vaglio, F. Aboul-ela, and J. Karn, Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. J Mol Biol 336 (2004) 625-38.

[150] R. Nifosi, C.M. Reyes, and P.A. Kollman, Molecular Dynamics Studies of the HIV-1 TAR and its Complex with Argininamide. Nucleic Acids Research (2000) 4944-4955.

[151] H.M. Al-Hashimi, Y. Gosser, A. Gorin, W. Hu, A. Majumdar, and D.J. Patel, Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. J Mol Biol 315 (2002) 95-102.

[152] T.E. Edwards, and S. Th Sigurdsson, Electron paramagnetic resonance dynamic signatures of TAR RNA-small molecule complexes provide insight into RNA structure and recognition. Biochemistry 41 (2002) 14843-7.

[153] Q. Zhang, X. Sun, E.D. Watt, and H.M. Al-Hashimi, Resolving the motional modes that code for RNA adaptation. Science 311 (2006) 653-6.

[154] Q. Zhang, A.C. Stelzer, C.K. Fisher, and H.M. Al-Hashimi, Visualizing spatially correlated dynamics that directs RNA conformational transitions. Nature 450 (2007) 1263-7.

[155] G.L. Olsen, D.C. Echodu, Z. Shajani, M.F. Bardaro, Jr., G. Varani, and G.P. Drobny, Solid-State Deuterium NMR Studies Reveal mus-ns Motions in the HIV-1 Transactivation Response RNA Recognition Site. J Am Chem Soc (2008).

[156] J.A. Jaeger, J. SantaLucia, Jr., and I. Tinoco, Jr., Determination of RNA structure and thermodynamics. Annu Rev Biochem 62 (1993) 255-87.

[157] R.A. Colvin, and M.A. Garciablanco, Unusual Structure of the Human-Immunodeficiency-Virus Type-1 Transactivation Response Element. Journal of Virology 66 (1992) 930-935.

[158] A.D. Critchley, I. Haneef, D.J. Cousens, and P.G. Stockley, Modeling and solution structure probing of the HIV-1 TAR stem-loop. J Mol Graph 11 (1993) 92-7, 124.

[159] M.J. Michnicka, J.W. Harper, and G.C. King, Selective Isotopic Enrichment of Synthetic Rna - Application to the Hiv-1 Tar Element. Biochemistry 32 (1993) 395-400.

[160] C.M. Gherghe, Z. Shajani, K.A. Wilkinson, G. Varani, and K.M. Weeks, Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S-2) in RNA. Journal of the American Chemical Society 130 (2008) 12244-+.

[161] M.F. Bardaro, Z. Shajani, K. Patora-Komisarska, J.A. Robinson, and G. Varani, How binding of small molecule and peptide ligands to HIV-1 TAR alters the RNA motional landscape. Nucleic Acids Research 37 (2009) 1529-1540.

[162] J.R. Williamson, P. Vallurupalli, L. Scott, M. Hennig, and L.E. Kay, New RNA labeling methods offer dramatic sensitivity enhancements in H-2 NMR relaxation spectra. Journal of the American Chemical Society 128 (2006) 9346-9347.

[163] G. Clerzius, J.F. Gelinas, and A. Gatignol, Multiple levels of PKR inhibition during HIV-1 replication. Reviews in Medical Virology 21 (2011) 42-53.

[164] I. Kanevsky, F. Chaminade, D. Ficheux, A. Moumen, R. Gorelick, M. Negroni, J.L. Darlix, and P. Fosse, Specific interactions between HIV-1 nucleocapsid protein and the TAR element. J Mol Biol 348 (2005) 1059-77.

[165] B. Berkhout, N.L. Vastenhouw, B.I.F. Klasens, and H. Huthoff, Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. Rna-a Publication of the Rna Society 7 (2001) 1097-1114.

[166] E.S. Andersen, S.A. Contera, B. Knudsen, C.K. Damgaard, F. Besenbacher, and J. Kjems, Role of the trans-activation response element in dimerization of HIV-1 RNA. J Biol Chem 279 (2004) 22243-9.

[167] R. Song, J. Kafaie, and M. Laughrea, Role of the 5' TAR Stem-Loop and the U5-AUG Duplex in Dimerization of HIV-1 Genomic RNA. Biochemistry 47 (2008) 3283-93.

[168] C. Helga-Maria, M.L. Hammarskjold, and D. Rekosh, An intact TAR element and cytoplasmic localization are necessary for efficient packaging of human immunodeficiency virus type 1 genomic RNA. Journal of Virology 73 (1999) 4127-4135.

[169] Z. Klase, P. Kale, R. Winograd, M.V. Gupta, M. Heydarian, R. Berro, T. McCaffrey, and F. Kashanchi, HIV-1 TAR element is processed by Dicer to yield a viral micro-RNA involved in chromatin remodeling of the viral LTR. Bmc Molecular Biology 8 (2007).

[170] D.L. Ouellet, I. Plante, P. Landry, C. Barat, M.E. Janelle, L. Flamand, M.J. Tremblay, and P. Provost, Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. Nucleic Acids Research 36 (2008) 2353-2365.

[171] Z. Klase, R. Winograd, J. Davis, L. Carpio, R. Hildreth, M. Heydarian, S. Fu, T. McCaffrey, E. Meiri, M. Ayash-Rashkovsky, S. Gilad, Z. Bentwich, and F. Kashanchi, HIV-1 TAR miRNA protects against apoptosis by altering cellular gene expression. Retrovirology 6 (2009).

[172] A.T. Das, A. Harwig, M.M. Vrolijk, and B. Berkhout, The TAR hairpin of human immunodeficiency virus type 1 can be deleted when not required for Tat-mediated activation of transcription. Journal of Virology 81 (2007) 7742-7748.

[173] P.C. Bevilacqua, and T.R. Cech, Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. Biochemistry 35 (1996) 9983-9994.

[174] L. Manche, S.R. Green, C. Schmedt, and M.B. Mathews, Interactions between Double-Stranded-Rna Regulators and the Protein-Kinase Dai. Molecular and Cellular Biology 12 (1992) 5238-5248.

[175] J.W. Ucci, Y. Kobayashi, G. Choi, A.T. Alexandrescu, and J.L. Cole, Mechanism of interaction of the double-stranded RNA (dsRNA) binding domain of protein kinase R with short dsRNA sequences. Biochemistry 46 (2007) 55-65.

[176] S. Roy, M. Agy, A.G. Hovanessian, N. Sonenberg, and M.G. Katze, The Integrity of the Stem Structure of Human-Immunodeficiency-Virus Type-1 Tat-Responsive Sequence Rna

Is Required for Interaction with the Interferon-Induced 68,000-Mr Protein-Kinase. Journal of Virology 65 (1991) 632-640.

[177] R.K. Maitra, N.A.J. McMillan, S. Desai, J. McSwiggen, A.G. Hovanessian, G. Sen, B.R.G. Williams, and R.H. Silverman, HIV-1 TAR RNA HAS AN INTRINSIC ABILITY TO ACTIVATE INTERFERON-INDUCIBLE ENZYMES. Virology 204 (1994) 823-827.

[178] B.W. Carpick, V. Graziano, D. Schneider, R.K. Maitra, X. Lee, and B.R.G. Williams, Characterization of the solution complex between the interferon-induced, double-stranded RNA-activated protein kinase and HIV-I trans-activating region RNA. Journal of Biological Chemistry 272 (1997) 9510-9516.

[179] S.A. McKenna, D.A. Lindhout, C.E. Aitken, and J.D. Puglisi, Molecular insights into PKR activation by viral double-stranded RNA. Structure and Biophysics - New Technologies for Current Challenges in Biology Beyond 231 (2007) 99-110.

[180] L.A. Heinicke, C.J. Wong, J. Lary, S.R. Nallagatla, A. Diegelman-Parente, X. Zheng, J.L. Cole, and P.C. Bevilacqua, RNA Dimerization Promotes PKR Dimerization and Activation. Journal of Molecular Biology 390 (2009) 319-338.

[181] I. Kim, C.W. Liu, and J.D. Puglisi, Specific recognition of HIV TAR RNA by the dsRNA binding domains (dsRBD1-dsRBD2) of PKR. J Mol Biol 358 (2006) 430-42.

[182] A. Moumen, L. Polomack, B. Roques, H. Buc, and M. Negroni, The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. Nucleic Acids Research 29 (2001) 3814-3821.

[183] J.L. Darlix, A. Vincent, C. Gabus, H. Derocquigny, and B. Roques, Transactivation of the 5' to 3' Viral-DNA Strand Transfer by Nucleocapsid Protein during Reverse Transcription of Hiv-1 Rna. Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences 316 (1993) 763-771.

[184] J.H. Guo, L.E. Henderson, J. Bess, B. Kane, and J.G. Levin, Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. Journal of Virology 71 (1997) 5178-5188.

[185] M. LapadatTapolsky, C. Gabus, M. Rau, and J.L. Darlix, Possible roles of HIV-1 nucleocapsid protein in the specificity of proviral DNA synthesis and in its variability. Journal of Molecular Biology 268 (1997) 250-260.

[186] S. Bernacchi, S. Stoylov, E. Piemont, D. Ficheux, B.P. Roques, J.L. Darlix, and Y. Mely, HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. Journal of Molecular Biology 317 (2002) 385-399.

[187] J. Godet, H. de Rocquigny, C. Raja, N. Glasser, D. Ficheux, J.L. Darlix, and Y. Mely, During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. Journal of Molecular Biology 356 (2006) 1180-1192.

[188] H.W. Liu, G. Cosa, C.F. Landes, Y.N. Zeng, B.J. Kovaleski, D.G. Mullen, G. Barany, K. Musier-Forsyth, and P.F. Barbara, Single-molecule FRET studies of important intermediates in the nucleocapsid-protein-chaperoned minus-strand transfer step in HIV-1 reverse transcription. Biophysical Journal 89 (2005) 3470-3479.

[189] M.N. Vo, G. Barany, I. Rouzina, and K. Musier-Forsyth, HIV-1 Nucleocapsid Protein Switches the Pathway of Transactivation Response Element RNA/DNA Annealing from Loop-Loop "Kissing" to "Zipper". Journal of Molecular Biology 386 (2009) 789-801.

[190] A. Gatignol, A. Bucklerwhite, B. Berkhout, and K.T. Jeang, Characterization of a Human Tar Rna-Binding Protein That Activates the Hiv-1 Ltr. Science 251 (1991) 1597-1600.

[191] A.D. Haase, L. Jaskiewicz, H.D. Zhang, S. Laine, R. Sack, A. Gatignol, and W. Filipowicz, TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. Embo Reports 6 (2005) 961-967.

[192] T.P. Chendrimada, R.I. Gregory, E. Kumaraswamy, J. Norman, N. Cooch, K. Nishikura, and R. Shiekhattar, TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature 436 (2005) 740-744.

[193] K.T. Jeang, Y. Bennasser, and M.L. Yeung, HIV-1 TAR RNA subverts RNA interference in transfected cells through sequestration of TAR RNA-binding protein, TRBP. Journal of Biological Chemistry 281 (2006) 27674-27678.

[194] M. Erard, D.G. Barker, F. Amalric, K.T. Jeang, and A. Gatignol, An Arg/Lys-rich core peptide mimics TRBP binding to the HIV-1 TAR RNA upper-stem/loop. J Mol Biol 279 (1998) 1085-99.

[195] F. Bloch, Nuclear induction. Phys. Rev. 70 (1946) 460-474.

[196] E.M. Purcell, H.C. Torrey, and R.V. Pound, Resonance Absorption by Nuclear Magnetic Moments in a Solid. Physical Review 69 (1946) 37-38.

[197] J.T. Arnold, S.S. Dharmatti, and M.E. Packard, Chemical Effects on Nuclear Induction Signals from Organic Compounds. Journal of Chemical Physics 19 (1951) 507-507.

[198] R.R. Ernst, and W.A. Anderson, Application of Fourier Transform Spectroscopy to Magnetic Resonance. Review of Scientific Instruments 37 (1966) 93-+.

[199] J. Jeener, Ampere Summer School, Basko Polje, Yugoslavia, 1971.

[200] W.P. Aue, E. Bartholdi, and R.R. Ernst, 2-Dimensional Spectroscopy - Application to Nuclear Magnetic- Resonance. Journal of Chemical Physics 64 (1976) 2229-2246.

[201] K. Wüthrich, G. Wider, G. Wagner, and W. Braun, Journal of Molecular Biology 155 (1982) 311-318.

[202] M.P. Williamson, T.F. Havel, and K. Wuthrich, Solution Conformation of Proteinase Inhibitor-Iia from Bull Seminal Plasma by H-1 Nuclear Magnetic-Resonance and Distance Geometry. Journal of Molecular Biology 182 (1985) 295-315.

[203] E.P. Nikonowicz, A. Sirr, P. Legault, F.M. Jucker, L.M. Baer, and A. Pardi, Preparation of C-13 and N-15 Labeled Rnas for Heteronuclear Multidimensional Nmr-Studies. Nucleic Acids Research 20 (1992) 4507-4513.

[204] R.T. Batey, M. Inada, E. Kujawinski, J.D. Puglisi, and J.R. Williamson, Preparation of Isotopically Labeled Ribonucleotides for Multidimensional Nmr-Spectroscopy of Rna. Nucleic Acids Research 20 (1992) 4515-4523.

[205] S.A. White, M. Nilges, A. Huang, A.T. Brunger, and P.B. Moore, Nmr Analysis of Helix-I from the 5s Rna of Escherichia-Coli. Biochemistry 31 (1992) 1610-1621.

[206] A.G. Palmer, 3rd, NMR characterization of the dynamics of biomacromolecules. Chem Rev 104 (2004) 3623-40.

[207] A.G. Palmer, 3rd, C.D. Kroenke, and J.P. Loria, Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. Methods Enzymol 339 (2001) 204-38.

[208] A.G. Palmer, 3rd, and F. Massi, Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. Chem Rev 106 (2006) 1700-19.

[209] A.K. Mittermaier, and L.E. Kay, Observing biological dynamics at atomic resolution using NMR. Trends in Biochemical Sciences 34 (2009) 601-611.

[210] J.H. Prestegard, H.M. Al-Hashimi, and J.R. Tolman, NMR structures of biomolecules using field oriented media and residual dipolar couplings. Quarterly Reviews of Biophysics 33 (2000) 371-424.

[211] J.R. Tolman, Dipolar couplings as a probe of molecular dynamics and structure in solution. current opinion in structural biology 11 (2001) 532-539.

[212] M.F. Summers, K. Lu, and Y. Miyazaki, Isotope labeling strategies for NMR studies of RNA. Journal of Biomolecular Nmr 46 (2010) 113-125.

[213] Y.X. Wang, X.B. Zuo, J.B. Wang, P. Yu, and S.E. Butcher, Rapid global structure determination of large RNA and RNA complexes using NMR and small-angle X-ray scattering. Methods 52 (2010) 180-191.

[214] A.C. Stelzer, A.T. Frank, M.H. Bailor, I. Andricioaei, and H.M. Al-Hashimi, Constructing atomic-resolution RNA structural ensembles using MD and motionally decoupled NMR RDCs. Methods 49 (2009) 167-73.

[215] B. Furtig, C. Richter, J. Wohnert, and H. Schwalbe, NMR spectroscopy of RNA. Chembiochem 4 (2003) 936-62.

[216] Z. Shajani, and G. Varani, NMR studies of dynamics in RNA and DNA by (13)C relaxation. Biopolymers 86 (2007) 348-59.

[217] Q. Zhang, and H.M. Al-Hashimi, Extending the NMR spatial resolution limit for RNA by motional couplings. Nat Methods 5 (2008) 243-5.

[218] M. Getz, X. Sun, A. Casiano-Negroni, Q. Zhang, and H.M. Al-Hashimi, Review NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. Biopolymers 86 (2007) 384-402.

[219] A.L. Hansen, E.N. Nikolova, A. Casiano-Negroni, and H.M. Al-Hashimi, Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. J Am Chem Soc 131 (2009) 3818-9.

[220] S.S. Wijmenga, and B.N.M. van Buuren, The use of NMR methods for conformational studies of nucleic acids. Progress in Nuclear Magnetic Resonance Spectroscopy 32 (1998) 287-387.

[221] K. Wüthrich, NMR of Proteins and Nucleic Acids, Wiley, New York, 1986.

[222] V. Sklenar, R.D. Peterson, M.R. Rejante, and J. Feigon, 2-Dimensional and 3-Dimensional Hcn Experiments for Correlating Base and Sugar Resonances in N-15,C-13-Labeled Rna Oligonucleotides. Journal of Biomolecular Nmr 3 (1993) 721-727.

[223] P. Legault, B.T. Farmer, L. Mueller, and A. Pardi, Through-Bond Correlation of Adenine Protons in a C-13-Labeled Ribozyme. Journal of the American Chemical Society 116 (1994) 2203-2204.

[224] J.P. Marino, J.H. Prestegard, and D.M. Crothers, Correlation of Adenine H2/H8 Resonances in Uniformly C-13 Labeled Rnas By 2d Hcch-Tocsy - a New Tool For H-1 Assignment. Journal of the American Chemical Society 116 (1994) 2205-2206.

[225] V. Sklenar, T. Dieckmann, S.E. Butcher, and J. Feigon, Through-bond correlation of imino and aromatic resonances in C-13-,N-15-labeled RNA via heteronuclear TOCSY. Journal of Biomolecular Nmr 7 (1996) 83-87.

[226] A.J. Dingley, and S. Grzesiek, Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (2)J(NN) couplings. Journal of the American Chemical Society 120 (1998) 8293-8297.

[227] K. Pervushin, A. Ono, C. Fernandez, T. Szyperski, M. Kainosho, and K. Wuthrich, NMR scaler couplings across Watson-Crick base pair hydrogen bonds in DNA observed by transverse relaxation optimized spectroscopy. Proceedings of the National Academy of Sciences of the United States of America 95 (1998) 14147-14151.

[228] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, and J.H. Prestegard, Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information For Structure Determination in Solution. Proceedings of the National Academy of Sciences of the United States of America 92 (1995) 9279-9283.

[229] N. Tjandra, and A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278 (1997) 1111-1114.

[230] N. Tjandra, Establishing a degree of order: obtaining high-resolution NMR structures from molecular alignment. Structure With Folding & Design 7 (1999) R205-R211.

[231] M.R. Hansen, L. Mueller, and A. Pardi, Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nature Structural Biology 5 (1998) 1065-1074.

[232] G.M. Clore, M.R. Starich, and A.M. Gronenborn, Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. Journal of the American Chemical Society 120 (1998) 10571-10572.

[233] M.R. Hansen, P. Hanson, and A. Pardi, Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. Methods in Enzymology 317 (2000) 220-240.

[234] B. Wu, M. Petersen, F. Girard, M. Tessari, and S.S. Wijmenga, Prediction of molecular alignment of nucleic acids in aligned media. Journal Of Biomolecular Nmr 35 (2006) 103-115.

[235] M. Zweckstetter, G. Hummer, and A. Bax, Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. Biophys J 86 (2004) 3444-60.

[236] C.A. Bewley, K.R. Gustafson, M.R. Boyd, D.G. Covell, A. Bax, G.M. Clore, and A.M. Gronenborn, Solution structure of cyanovirin-N, a potent HIV-inactivating protein. Nature Structural Biology 5 (1998) 571-578.

[237] G.M. Clore, M.R. Starich, C.A. Bewley, M.L. Cai, and J. Kuszewski, Impact of residual dipolar couplings on the accuracy of NMR structures determined from a minimal number of NOE restraints. Journal of the American Chemical Society 121 (1999) 6513-6514.

[238] G.M. Clore, A.M. Gronenborn, and A. Bar, A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. Journal of Magnetic Resonance 133 (1998) 216-221.

[239] C. Musselman, S.W. Pitt, K. Gulati, L.L. Foster, I. Andricioaei, and H.M. Al-Hashimi, Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. J Biomol NMR 36 (2006) 235-49.

[240] A. Saupe, Recent results in the field of liquid crystals. Angew. Chem., Int. Ed. Engl. 7 (1968) 97-112.

[241] J.A. Losonczi, M. Andrec, M.W.F. Fischer, and J.H. Prestegard, Order matrix analysis of residual dipolar couplings using singular value decomposition. Journal of Magnetic Resonance 138 (1999) 334-342.

[242] J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, and J.H. Prestegard, Structural and dynamic analysis of residual dipolar coupling data for proteins. Journal Of The American Chemical Society 123 (2001) 1416-1424.

[243] B.E. Ramirez, and A. Bax, Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. Journal of the American Chemical Society 120 (1998) 9106-9107.

[244] H.M. Al-Hashimi, H. Valafar, M. Terrell, E.R. Zartler, M.K. Eidsness, and J.H. Prestegard, Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. Journal of Magnetic Resonance 143 (2000) 402-406.

[245] E.A. Dethoff, A.L. Hansen, Q. Zhang, and H.M. Al-Hashimi, Variable helix elongation as a tool to modulate RNA alignment and motional couplings. J Magn Reson 202 (2010) 117-21.

[246] J. Cavanagh, W.J. Fairbrother, A.G.I. Palmer, and N.J. Skelton, Protein NMR Spectroscopy, Academic Press, San Diego, CA, 1996.

[247] A. Abragam, Principles of Nuclear Magnetism, Clarendon Press, Oxford, 1961.

[248] J.W. Peng, and G. Wagner, Mapping of Spectral Density-Functions Using Heteronuclear Nmr Relaxation Measurements. Journal of Magnetic Resonance 98 (1992) 308-332.

[249] J.W. Peng, and G. Wagner, Mapping of the Spectral Densities of N-H Bond Motions in Eglin-C Using Heteronuclear Relaxation Experiments. Biochemistry 31 (1992) 8571-8586.

[250] A.M. Mandel, M. Akke, and A.G. Palmer, Backbone Dynamics of Escherichia-Coli Ribonuclease Hi - Correlations With Structure and Function in an Active Enzyme. Journal of Molecular Biology 246 (1995) 144-163.

[251] G. Lipari, and A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. Journal of the American Chemical Society 104 (1982) 4546-4559.

[252] G.M. Clore, A. Szabo, A. Bax, L.E. Kay, P.C. Driscoll, and A.M. Gronenborn, Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. J . Am. Chem. Soc 112 (1990) 4989-4991.

[253] A.L. Hansen, and H.M. Al-Hashimi, Dynamics of large elongated RNA by NMR carbon relaxation. J Am Chem Soc 129 (2007) 16072-82.

[254] C. Musselman, H.M. Al-Hashimi, and I. Andricioaei, iRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. Biophys J 93 (2007) 411-22.

[255] A.T. Frank, A.C. Stelzer, H.M. Al-Hashimi, and I. Andricioaei, Constructing RNA Dynamical Ensembles by Combining MD and Motionally Decoupled NMR RDCs: New insights into RNA Dynamics and Adaptive Ligand Recognition. Nucleic Acids Res 37 (2009) 3670-9.

[256] M.P. Latham, G.R. Zimmermann, and A. Pardi, NMR Chemical Exchange as a Probe for Ligand-Binding Kinetics in a Theophylline-Binding RNA Aptamer. Journal of the American Chemical Society 131 (2009) 5052-+.

[257] K. Kloiber, R. Spitzer, M. Tollinger, R. Konrat, and C. Kreutz, Probing RNA dynamics via longitudinal exchange and CPMG relaxation dispersion NMR spectroscopy using a sensitive 13C-methyl label. Nucleic Acids Res.

[258] D.M. Korzhnev, V.Y. Orekhov, and L.E. Kay, Off-resonance R(1rho) NMR studies of exchange dynamics in proteins with low spin-lock fields: an application to a Fyn SH3 domain. J Am Chem Soc 127 (2005) 713-21.

[259] F. Massi, E. Johnson, C. Wang, M. Rance, and A.G. Palmer, 3rd, NMR R1 rho rotating-frame relaxation with weak radio frequency fields. J Am Chem Soc 126 (2004) 2247-56.

[260] V.Z. Miloushev, and A.G. Palmer, R-1p relaxation for two-site chemical exchange: General approximations and some exact solutions. Journal of Magnetic Resonance 177 (2005) 221-227.

[261] O. Trott, and A.G. Palmer, R-1 rho relaxation outside of the fast-exchange limit. Journal of Magnetic Resonance 154 (2002) 157-160.

[262] A. Bax, and D.G. Davis, Practical Aspects of Two-Dimensional Transverse Noe Spectroscopy. J. Mag. Reson 63 (1985) 207-213.

Chapter 2

Characterizing dynamics in TAR and exploring long-range motional correlations by NMR, mutagenesis, and molecular dynamics

2.1 Introduction

The transactivation response element (TAR) RNA hairpin [1] (Figure 2.1) plays a diverse role in HIV viral replication and is a major target for developing anti-HIV therapies [2]. TAR has two primary binding sites: a trinucleotide bulge and hexanucleotide apical loop. The bulge binds the viral transactivator protein Tat [3] which stimulates transcription-elongation of HIV mRNA by recruiting the positive transcription elongation factor P-TEFb [4,5]. The TAR apical loop contacts P-TEFb via the cyclin T1 (CytT1) subunit [4-7] and is also believed to bind a number of other proteins, including the TAR RNA-binding protein (TRBP), which inhibits protein kinase R (PKR) [8,9] and increases RNA translation [2,10], the nucleocapsid protein (NC), which promotes minus-strand transfer during reverse transcription [11], and the eukaryotic translation initiation factor 2 (eIF2) which is believed to contact both the TAR apical loop and sub-apical lower stem [12]. The apical loop is also believed to aid dimerization of the HIV-1 genome by forming self-interactions in kissing complexes [13,14].

Most biophysical studies have so far focused on characterizing the conformational properties of the TAR bulge where the protein Tat binds. Studies have shown that recognition of TAR by Tat-derived peptides results in large changes in the relative orientation of helices that are coupled to local rearrangements in the bulge conformation [15-19]. Numerous studies have also established that the TAR bulge can adopt wide ranging conformations and thereby bind to a variety of targets including divalent ions [20,21] and chemically distinct small molecules that bind the HIV-1 TAR bulge and inhibit its interaction with Tat [22-25]. Recent studies show that inherent flexibility at the bulge junction plays an important role in adaptive recognition [26-32]. In particular, new domain-elongation NMR techniques allowed resolution of inter-helical motions and local

fluctuations in and around the bulge that mirror the structural rearrangements that take place upon recognition [28,29]. Thus, small molecules likely select conformers in the TAR dynamical ensemble by "tertiary capture" rather than induce new TAR conformations [28,29]. These NMR studies were focused on the TAR bulge and were conducted on a TAR mutant (TAR$_m$) in which the wild-type (wt) apical loop was replaced with a spectroscopically more favorable UUCG tetraloop [33] (Figure 2.1).

Few biophysical studies have examined the TAR apical loop which is also being targeted in the design of anti-HIV therapeutics [34-37]. No structures have been reported to date for the TAR apical loop when in complex with cognate protein targets. Available structures of TAR show significant differences in the apical loop conformation likely because it is highly flexible in solution. To date, no quantitative experiments have been preformed to characterize dynamics in the apical loop and its potential roles in adaptive recognition. Furthermore, studies provide evidence for long-range interactions between the apical loop and the bulge [38,39] that could make the conformational dynamics of wtTAR different from that of TAR$_m$ which has previously been studied by NMR. Based on sequence covariation, structure probing, and molecular modeling, Berkhout and co-workers proposed that apical loop residue U31 forms a U31•A22-U40 base-triple in unbound TAR involving the A22-U40 basepair in the lower stem [38]. Isotropic reorientational eigenmode dynamics (iRED) analysis of a 20 ns molecular dynamics (MD) simulation of wtTAR revealed that dynamics at the apical loop, particularly residue U31, is correlated by unknown mechanisms to dynamics at the bulge, particularly residues U23 and C24 [39]. Such long-range interactions may provide the basis for coupling recognition events at the different TAR binding sites and could account in part for the high cooperativity of Tat and CytT1 binding to these two different TAR sites [7]. Long range motional correlations are increasingly implicated in the functions of RNA [40,41] but are notoriously difficult to characterize based on experimental methods.

Figure 2.1: Secondary structures of wtTAR (black) and TAR$_m$ (blue).

Here, we report the NMR and MD characterization of the conformational dynamics of the wtTAR apical loop and use a mutagenesis approach to experimentally examine potential couplings it may have to the trinucleotide bulge. Our results show that the apical loop and bulge act as independent dynamical centers with the apical loop undergoeing complex dynamics at multiple timescales that are likely important for adaptive recognition. Our study establishes an NMR experimental framework for studying motional correlations in RNA and indentifies one approach for indirectly establishing motional correlations between distinct regions in an RNA.

2.2 Material and Methods

2.2.1 Preparation and Purification of Uniformly ^{13}C/^{15}N Labeled RNA

E-AU-wtTAR, E-GC-wtTAR, and wtTAR samples were prepared by *in vitro* transcription using T7 RNA polymerase (Takara Mirus Bio, Inc.), uniformly ^{13}C/^{15}N labeled nucleotide triphosphates (ISOTEC, Inc.), unlabeled nucleotide triphosphates (Sigma), and synthetic DNA templates (Integrated DNA Technologies, Inc.) containing the T7 promoter and sequence of interest. Short and elongated RNAs were purified using 20% and 15% (w/v) denaturing polyacrylamide gel electrophoresis, respectively, with 8M urea and TBE. The RNA was eluted from the gel in 20 mM Tris pH 8 buffer followed by ethanol precipitation. The RNA pellet was dissolved in water and exchanged into NMR buffer (15 mM sodium phosphate, 0.1 mM EDTA, and 25 mM NaCl at pH ~6.4) multiple times using a Centricon Ultracel YM-3 concentrator (Millipore Corp.). The elongated samples were annealed at 95°C for 5 min before being exchanged into NMR buffer. The final RNA concentration of the E-AU-wtTAR, E-GC-wtTAR, and wtTAR samples was 1.5 mM, 0.5 mM, and 0.7 mM, respectively. The aligned NMR sample of wtTAR was prepared by adding Pf1-phage (Asla Biotech Ltd.) in NMR buffer to wtTAR to yield a final Pf1-phage concentration of ~19 mg/mL and a final RNA concentration of 0.35mM. The addition of phage did not affect the average structure of RNA as judged from careful comparison of the chemical shifts in the absence and presence of phage (data not shown).

2.2.2 Resonance Assignments

All NMR experiments were performed at 298 K on an Avance Bruker 600 MHz spectrometer equipped with a triple-resonance cryogenic (5mm) probe and an Avance Bruker 900 MHz spectrometer equipped with a TCI cryoprobe. NMR spectra were analyzed using NMR Draw [42] and Sparky 3 [43]. The ^{1}H, ^{13}C, and ^{15}N resonances in wtTAR were assigned using standard homonuclear and heteronuclear, 2D and 3D experiments [44] on unlabeled and uniformly ^{13}C/^{15}N labeled wtTAR sample, respectively. The non-exchangeable NOE walk was uninterrupted from C19 to U23, G26 to C29, and G34 to C39 indicating a continuous helical conformation in these regions. The NOE walk was interrupted at bulge residue C24 and apical loop residue C30 and became convoluted at the terminal ends due to peak multiplicity and end-fraying effects. The NOESY data was supplemented with 3D HCN experiments [45] for intranucleotide H8/H6-N1/N9 H1' correlations and a 2D HCCNH experiment [46] for base C6/C5 to imino H1/H3 correlations in pyrimidines. All resonance intensities in E-wtTAR were measured as previously described for E-TAR$_m$ [29]. Values for various bond vector types were normalized to a minimum value of 0.1 independently for G-C and A-U residues, as previously described [29].

2.2.3 Measurement and Order Tensor Analysis of RDCs

2D ^{13}C–^{1}H (or ^{15}N-^{1}H) S^{3}E HSQC experiments [47,48] were used to measure one bond $^1D_{C6H6}$, $^1D_{C8H8}$, $^1D_{C5H5}$, $^1D_{C2H2}$, $^1D_{C1'H1'}$, and $^1D_{N1/3H1/3}$ RDCs in non-elongated wtTAR using ~19 mg/ml of Pf1 phage [49,50] as an ordering medium. RDC measurement error was estimated from duplicate experiments that yielded splittings along the ^{1}H dimension or ^{15}N/^{13}C dimension, as previously described [51,52]. The measured RDCs are listed in Table 2.1.

Residue (bond vector)	RDC (Hz)	Residue (bond vector)	RDC (Hz)	Residue (bond vector)	RDC (Hz)
G17 (C8H8)	*-4.9*	G26 (C8H8)	22.8	*A35 (C1'H1')*	*-9.1*
G18 (C8H8)	6.6	A27 (C1'H1')	-23.9	*A35 (C8H8)*	*3.2*
G18 (N1H1)	-5.9	A27 (C2H2)	21.6	G36 (C8H8)	27.4
A20 (C2H2)	16.7	A27 (C8H8)	22.2	G36 (N1H1)	-12

A20 (C8H8)	13.7	G28 (C8H8)	25	U38 (N3H3)	-9
G21 (C1'H1')	2.1	G28 (N1H1)	-11.8	C39 (C1'H1')	-1.9
G21 (C8H8)	9.1	C29 (C6H6)	26.4	C39 (C5H5)	26.4
G21 (N1H1)	-3.3	C29 (C1'H1')	-4.3	*U40 (C1'H1')*	*-1.9*
A22 (C1'H1')	*15.8*	*C30 (C1'H1')*	*7.9*	*U40 (C5H5)*	*22.7*
A22 (C2H2)	*5.9*	*C30 (C5H5)*	*20.5*	C41 (C5H5)	20.6
A22 (C8H8)	*5.8*	*U31 (C1'H1')*	*-10.3*	C41 (C6H6)	4.9
U23 (C1'H1')	*6.9*	*U31 (C6H6)*	*-4.9*	U42 (C5H5)	13.2
U23 (C5H5)	*-4.6*	*G32 (C1'H1')*	*4*	U42 (N3H3)	-4.4
U23 (C6H6)	*13.9*	*G32 (C8H8)*	*2.4*	G43 (C8H8)	12.9
C24 (C5H5)	*-1*	*G33 (C8H8)*	*15.5*	G43 (N1H1)	-6.9
C24 (C6H6)	*7.6*	*G33 (C1'H1')*	*0.9*	C44 (C5H5)	8.8
U25 (C5H5)	*-3.4*	*G34 (C1'H1')*	*6.9*	C45 (C1'H1')	*12.2*
U25 (C6H6)	*12.9*	*G34 (C8H8)*	*22.5*		

Table 2.1: RDCs (in Hz) of wtTAR measured at 900MHz. Values shown in italic were excluded from the order tensor analysis.

Domain	N	CN	RMSD (Hz)	R	η	ϑ x 10^{-3}	ϑ_{int}	θ	ξ
I	14	5.7	1.2	0.99	0.36 ± 0.1	0.65 ± 0.08	0.54	45	-41
II	13	3.0	1.7	0.99	0.10 ± 0.05	1.19 ± 0.05	± 0.07	± 7	± 50

Table 2.2. Order tensor analysis of RDCs measured in the wtTAR helical stems.

The RDCs measured in Watson-Crick basepairs in domains I and II were subjected to an order tensor analysis using idealized A-form helices as input coordinates [53]. Statistics for this analysis are summarized in Table 2.2. RDCs from the flexible residues G17-C45 and A22-U40 were excluded from analysis [53]. The helices were constructed using the Biopolymers module in Insight II (Molecular Simulations, Inc.), followed by a correction of the propeller twist angles from $+15^{o}$ to the idealized A-form value of -15° [53,54]. The measured RDCs and idealized-A-form helices were used to determine best-fit order tensors for each domain using singular value decomposition [55] implemented by the in-house written program RAMAH [56]. Order tensor errors due to "A-form structural noise" and RDC uncertainty were estimated using the program AFORM-RDC [53]. The overall RNA structure was assembled by rotating each domain into its principal axis system (PAS) of its best-fitted order tensor. The helices were then

assembled using the program Insight II (Molecular Simulations, Inc). Three of the four possible inter-helical orientations were omitted because they resulted in U40 (P)-C39 (O3') and A22 (O3')-G26 (P) distances greater than 1.59 Å and 28 Å, respectively. Inter-helical angles for all RNA structures were calculated using the in-house program Euler-RNA as previously described [57].

2.2.4 Measurement of Nucleobase Carbon Relaxation

The power dependence of $R_{1\rho}$ was measured for nucleobase C2, C6 and C8 carbons at 25°C using the pulse sequence shown in Figure 2.2. The same pulse sequence was used to measure R_1 by replacing the ^{13}C spinlock and adiabatic passages with a series of high power ^{1}H 180° pulses separated by 10 ms. A total of ten and six relaxation decays were measured for wtTAR and TAR$_m$, respectively, with spinlock strengths ranging from 1 to 5 kHz and at 3 different offsets (Table 2.3). The power dependence of the observed $R_{1\rho}$ was fitted using,

$$R_{1\rho} = R_1 \cos^2(\theta) + R_2^0 \sin^2(\theta) + \frac{\sin^2(\theta)\phi_{ex}k_{ex}}{\omega_{eff}^2 + k_{ex}^2}$$

(2.1)

where θ = arctan (ω_1/Ω) is the effective tilt angle of the magnetization; ω_1 is the spinlock power; Ω is the chemical shift offset of the resonance from the spinlock frequency; ω_{eff} = $(\omega_1^2 + \Omega^2)^{1/2}$ is the effective field strength; R_2^0 is the intrinsic transverse relaxation rate; ϕ_{ex} = $p_a p_b \Delta\omega^2$, where p_a and p_b are the populations of the exchanging states a and b; $\Delta\omega$ is the difference in chemical shift between the states (rad/s); and k_{ex} is the timescale of the motion [58]. To ascertain the presence of exchange in the dispersion curves, F-test statistics were used at the 99% confidence level between one (R_2) and three (R_2, f_{ex}, k_{ex}) parameter fits to the data. For all but the A35 C8 (p-value 3.86 x 10^{-4}) and G34 C8 (p-value 3.24 x 10^{-5}) resonances in wtTAR, which share a chemical exchange time of 33.4 ± 5.8 ms, no measurable exchange is observed on the microsecond time scale (see Appendix 1).

Using the R_2 values measured using the highest power and off-resonance spinlocks, the value of 2R_2-R_1 was computed providing an estimate for the amplitudes of ps-ns motions. 2R_2-R_1 is, to a good approximation, proportional to S^2·J (0), where S^2 is the order parameter describing motions occurring faster than overall tumbling of the
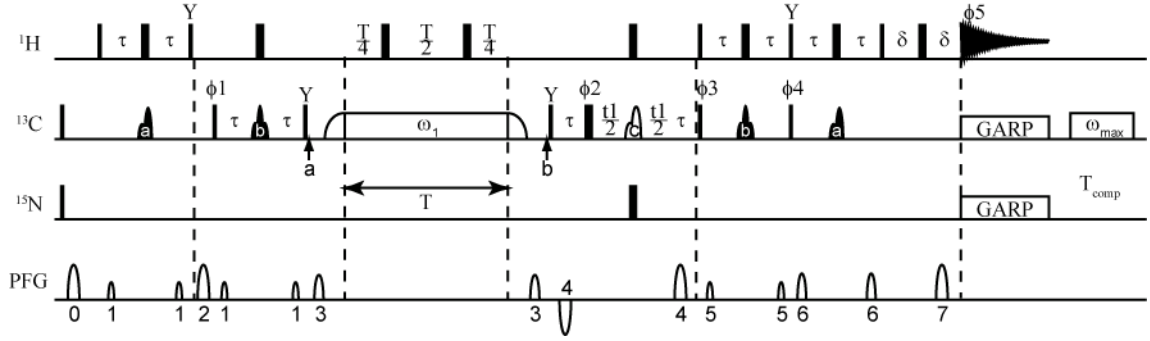
Figure 2.2: Pulse sequence for measurement of $R_{1\rho}$. Narrow and wide rectangles indicate 90° and 180° pulses, respectively. Selective pulses a and b are on resonance and c is off resonance. The a pulse selectively inverts carbon magnetization of interest while the b (c) pulses selectively refocus (invert) carbon magnetization to eliminate C-C scalar coupling evolution. The delays $\tau = \frac{1}{4} \times J_{CH} = 1.25$ ms and $\delta = 700$ μs. An optimized sampling strategy was used to minimize experiment time[60,61]. Pulse field gradients (PFG) are sine shaped with the following amplitude (G/cm)/durations (ms): G0 = 4.0/0.8, G1 = 1.1/0.25, G2 = 6.7/0.8, G3 = -7.2/0.8, G4 = 4.778/0.55, G5 = 1.3/0.25, G6 = -0.7/0.25, G7 = 9.5/0.55. Unless noted otherwise, all pulses have an x phase. The phase cycle is ϕ_1 = (x, -x), ϕ_2 = (x, x, y, y, -x, -x, -y, -y), ϕ_3 = 4 (x) 4 (-x), ϕ_4 = 4 (y) 4 (-y), and ϕ_5 = (x, -x, -x, x, -x, x, x, -x). GARP decoupling is applied on both ^{13}C and ^{15}N channels with strengths of 2.4 kHz and 1.0 kHz respectively. Axial peaks are shifted to the edge of the spectrum by incrementing the 90° ^{13}C pulse at point b and ϕ_5 by 180° for each t1 increment. Between points a and b in the experiment, the hydrogen carrier frequency is shifted to the center of the H8/H6 region while the carbon carrier can be shifted off-resonance to employ higher effective fields which suppress chemical exchange. A heat compensation element is applied after acquisition to account for the wide variation in RF irradiation from the use of different spinlock strengths. An off-resonance spinlock is applied at the highest field strength applied during the dispersion experiments for a time T_{comp} such that $T_{comp}\omega_{max}^2 + T\omega_1^2$ is a constant for all experiments and delay times[58]. For measuring R_1, the spinlock and adiabatic passages are replaced by a series of high power 180° pulses separated by 10 ms.

nucleic acid and J (0) is the spectral density function at zero field. The value of $2R_2$-R_1 is, to a good approximation, independent of the timescale of internal motions as well as the number of contributing relaxation mechanisms under the assumption that all relaxation mechanisms experience similar amplitude isotropic motions [59,60]. Relative S^2 (S_{rel}^2) values can therefore be estimated from the ratio of $2R_2$-R_1 measured for each residue to that of the largest value obtained in the well-structured A-form helices. Orientational dependencies of the ^{13}C relaxation rates for RNA the size of wtTAR are expected to only vary $2R_2$-R_1 by c.a. 10% [60].

	wtTAR		TAR$_m$	
Spin lock power (kHz)	^{13}C carrier frequency (ppm)		Spin lock power (kHz)	^{13}C carrier frequency (ppm)
5.0637	179.9, 165.0, 139.5		5.0637	134.2
3.6868	139.5		3.6868	134.2
3.161	139.5			
2.6132	139.5		2.6132	134.2
2.0503	139.5, 154.0		2.0503	134.2
1.5156	139.5, 154.0		1.5156	134.2
1.2367	139.5, 154.0			
0.9766	139.5, 154.0		0.9766	134.2

Table 2.3: Spin lock powers and carrier frequencies used in nucleobase C2, C6 and C8 $R_{1\rho}$ relaxation experiments

2.2.5 Molecular Dynamics Simulations

A 65 nanosecond all-atom molecular dynamics (MD) simulation of wtTAR was performed using the CHARMM force field (parameter set 27) [62]. Structure 3 of the family of free TAR NMR structures (PDB 1ANR) [32] was used as the starting coordinates, chosen as it yields the best agreement with previous residual dipolar couplings measurements [30]. TAR was neutralized using Na$^+$ counter ions and solvated in a 35 Å sphere of TIP3P water cite [63]. A stochastic boundary potential was used [64], allowing for more than 9 Å distance between the surface of the sphere and all RNA atoms. The solvated system was minimized and heated to 300 K while harmonically constraining the heavy atoms of the RNA with a force constant of 62 kcal/mol/ Å^2 for 100 ps, after which the constraints were removed. The system was then preliminarily equilibrated for 1 ns, and a production-run trajectory was followed up to 65 ns. A Nose-Hoover thermostat using a coupling constant of 50 ps^{-1} [65,66] was employed to maintain a constant temperature of 300 K throughout the simulation, with a 1 femtosecond time step. Details of the structural dynamics and correlations from this simulation are reported elsewhere [39,53].

2.3 Results and Discussion

2.3.1 NMR Characterization of wtTAR and TAR$_m$

If a long-range conformational coupling exists between the TAR apical loop and bulge, then replacement of the wt CUGGGA apical loop with a UUCG loop should impact the conformational properties of the bulge and therefore its NMR resonances. We therefore first compared 2D HSQC spectra of uniformly $^{13}C/^{15}N$ labeled wtTAR with those of the TAR$_m$ mutant (Figure 2.1). Representative spectra and the differences in chemical shifts are shown in Figure 2.3. Chemical shift differences were calculated using the following equation: $\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\alpha\Delta\delta_X)^2}$, where $\Delta\delta_H$ and $\Delta\delta_X$ are the changes in proton and carbon/nitrogen chemical shift and α is the ratio of the H and C/N gyromagnetic ratio. Expected differences are observed for domain II residues C29, G36, G28 and C37 that are near the apical loop. However, some differences are also observed at the bulge (U23) and neighboring residue A22 in domain I. The latter chemical shift differences are very similar in both amplitude and direction to the shifts induced by Na$^+$ ions and could possibly arise from small differences in sample preparation and having less Na$^+$ in wtTAR [52]. The difference observed for U38, which is involved in Tat recognition and which is flexible in unbound TAR [29], is likely due to the apical loop mutation. Finally, a greater deal of peak multiplicity is observed for terminal residues G17, G18 and C19 in wtTAR compared TAR$_m$ that most likely arises from presence of a combination of n and n+1 transcripts.

We examined if wtTAR binds the ligand argininamide (ARG), which is used as a mimic of Tat [15,67], in a manner similar to TAR$_m$. The ARG-induced chemical shift perturbations were virtually identical for wtTAR and TAR$_m$ (Figure 2.4), indicating that the apical loop does not impact ARG binding at the bulge. Nevertheless, ARG binding induced small perturbations at the apical loop at a distance from the bulge binding site. The possible source of these perturbations is discussed in a subsequent section.

We did not obtain any evidence for the proposed U31•A22-U40 base triple [38] in unbound wtTAR (Figure 2.5). We were unable to observe the U31 imino proton to detect a cross peak between H5 of U31 and the N7 of A22 using the *trans*-hydrogen bond J$_{NN}$ H5 (C5)NN NMR experiment which employs the non-exchangeable H5 [68]. Thus, the U31•A22-U40 base tripe either does not form or only forms transiently in solution.
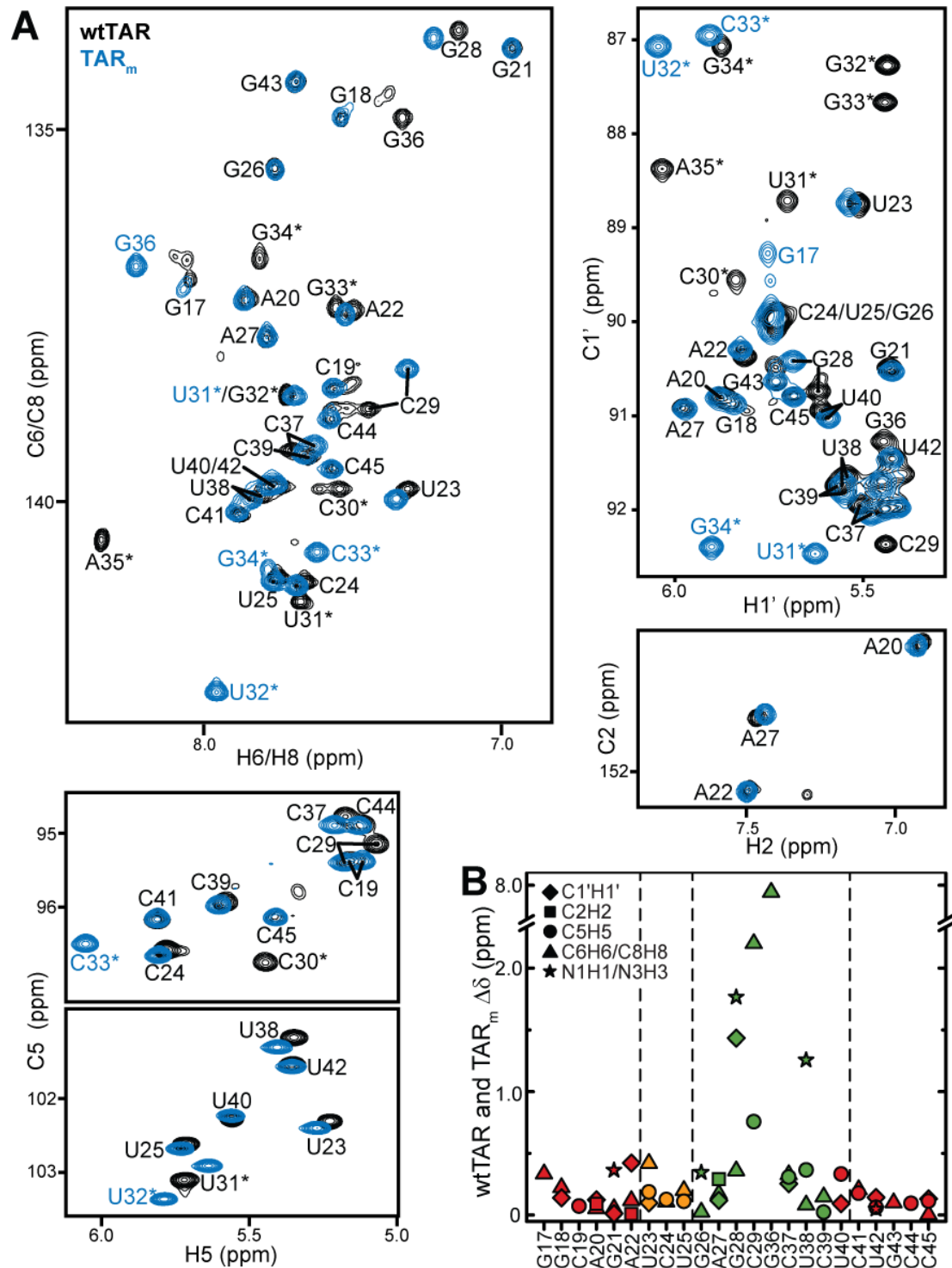
Figure 2.3: Comparison on wtTAR and TAR_m chemical shifts. (A) 2D ^1H-^{13}C HSQC spectra of wtTAR (black) overlaid on corresponding spectra of TAR_m (blue). Asterisks and colored labels denote resonances that are different between wtTAR and TAR_m due to the loop mutation. (B) Chemical shift differences between wtTAR and TAR_m plotted by residue. See inset for key.
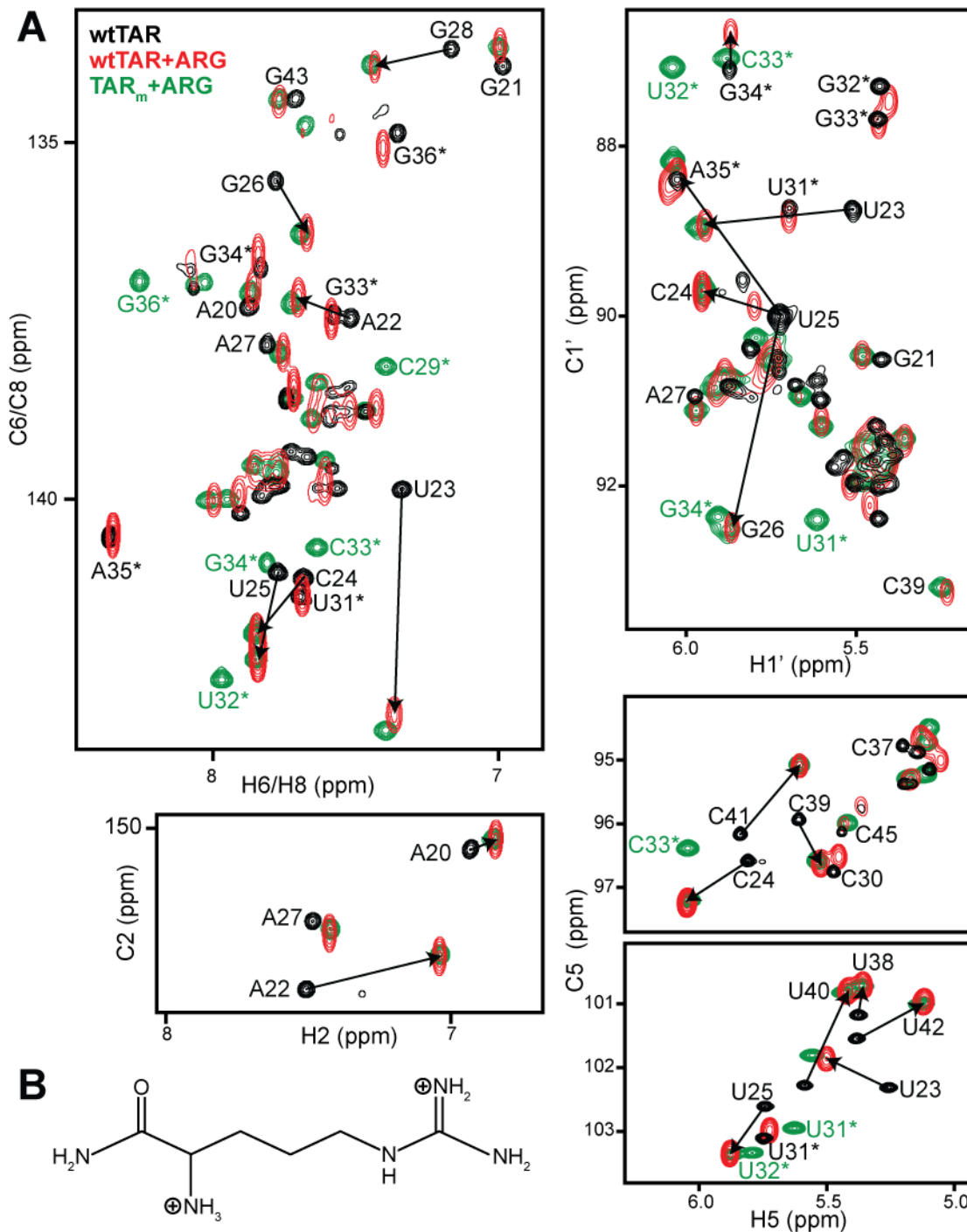
Figure 2.4: Comparison of the chemical shift perturbations in wtTAR and TAR_m induced by argininamide (ARG) binding. (A) 2D ^1H-^{13}C spectra of wtTAR (black) overlaid on corresponding spectra of wtTAR+ARG (red) and TAR_m+ARG (green). Asterisks and colored labels denote resonances that are different between wtTAR and TAR_m due to the loop mutation. Arrows indicate wtTAR chemical shift perturbations upon ARG binding. (B) Chemical structure of argininamide.
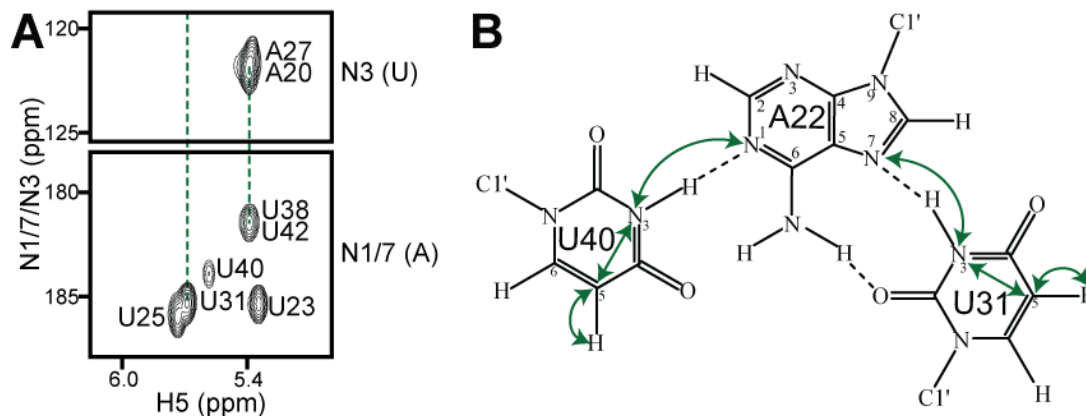
Figure 2.5: *Trans*-hydrogen bond detection in wtTAR. (A) J_{NN} H5(C5)NN[68] *trans*-hydrogen bond spectrum showing no evidence for a U31•A22-U40 base triple. (B) Proposed U31•A22-U40 base triple and magnetization transfer pathway (indicated by green arrows) [38].

2.3.2 Comparison of wtTAR and TAR$_m$ Structural Dynamics using RDCs

The small differences between spectra of TAR$_m$ and wtTAR raise the possibility that the apical loop influences the bulge conformation and therefore the relative orientation and dynamics of the TAR helices. To this end, we characterized the wtTAR inter-helical conformation using one bond C-H ($^1D_{CH}$) and N-H ($^1D_{NH}$) RDCs [51,69,70] and compared it to the TAR$_m$ conformation reported previously [30]. The relative orientation and dynamics of helices was determined using an order tensor analysis [55,57] as reported previously for TAR$_m$ [53]. Key statistics for this analysis are summarized in Table 2.2. Here, order tensors describing partial alignment relative to the magnetic field were determined for each helix by fitting the RDCs measured in non-terminal Watson-Crick basepairs to an idealized A-form helix geometry [53,57]. The program AFORM-RDC was used to estimate errors in the order tensor elements arising from "structural noise" and RDC measurement uncertainty [53]. An excellent fit was observed for the two helices (Figure 2.6A) indicating that the two helices adopt a canonical A-form geometry, as previously reported for TAR$_m$ [53].

Superposition of the order tensor frames obtained for each helix allowed determination of their average relative orientation [53,57]. As shown in Figure 2.6B, the average inter-helical bend (θ) and twist (ξ) angles obtained for wtTAR are, within experimental error, identical to those determined for TAR$_m$ [30]. The order tensor analysis also yields the amplitude of inter-helical motions as characterized by an internal
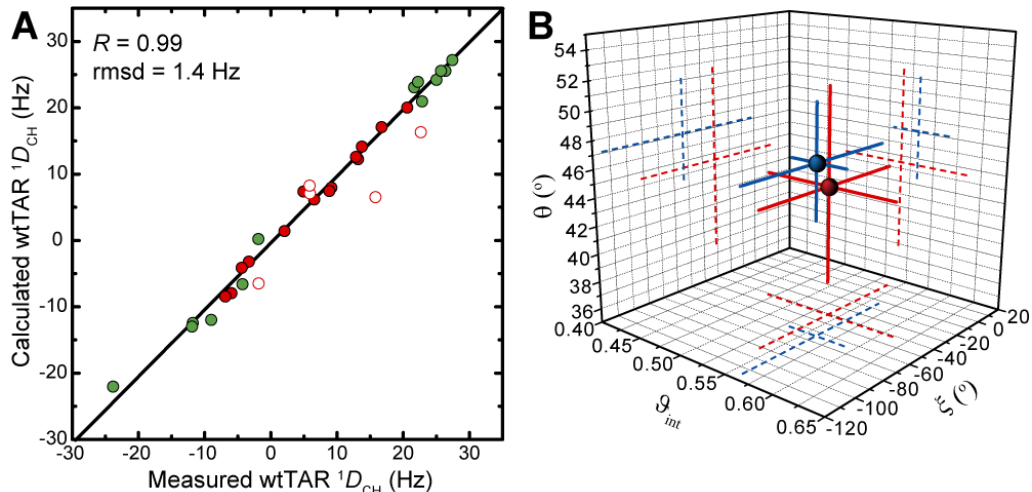
Figure 2.6: RDC-based comparison of the wtTAR and TAR$_m$ conformational dynamics. (*a*) Correlation plot between measured and back-calculated wtTAR RDCs when each stem order tensor is independently fit to an idealized A-form geometry. Domain II is shown in green, and domain I is shown in red, with A22 and U40 represented by open red circles. (*b*) The values for the inter-helical bend (θ), twist angle (ξ), and the inter-helical mobility (ϑ_{int}) determined for wtTAR (red) and TAR$_m$ (blue). Experimental errors are indicated by dashed lines.

generalized degree of order, ϑ_{int}, which varies between 0 and 1 for maximum and minimum inter-helical motional amplitudes, respectively [71]. As shown in Figure 2.6B and Table 2.2, $\vartheta_{int} = 0.54 \pm 0.07$ for wtTAR is in excellent agreement with $\vartheta_{int} = 0.56 \pm 0.03$ for TAR$_m$ [30,53]. Thus, wtTAR and TAR$_m$ undergo similar amplitude inter-helical motions at sub-millisecond timescales. This, together with the similar RDCs observed for the bulge and junctional A22-U40 basepair (data not shown) suggest that TAR$_m$ and wtTAR adopt very similar conformations in solution and therefore that the apical loop does not significantly impact the conformational dynamics of the bulge.

2.3.3 Structural Dynamics by ^{13}C Spin Relaxation and Domain Elongation

   To further characterize the conformational dynamics of wtTAR at ps-ns timescales, we used a domain-elongation strategy to decouple internal motions from overall reorientation [29] (Figure 2.7). The elongation slows down overall rotational diffusion allowing its separation from internal motions occurring at timescales comparable to overall molecular tumbling of the non-elongated RNA [72]. To avoid increasing NMR spectral overlap, two RNA constructs were prepared that are elongated
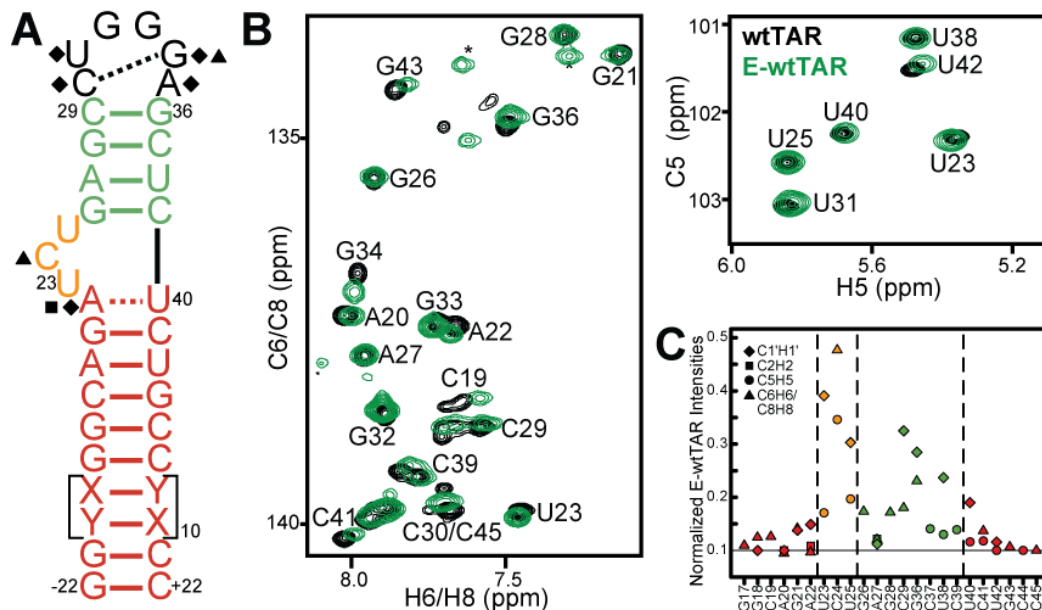
Figure 2.7: Chemical shift comparison of E-wTAR and wtTAR. (A) Elongated construct of wtTAR in which XY are unlabelled CG (E-GC-wtTAR) and UA (and E-AU-wtTAR) residues. Residues that undergo significant chemical shift perturbations upon elongation are highlighted on the secondary structure using black filled symbols. See inset in Fig 2.3B for symbol key. (B) 2D HSQC spectra of wtTAR (black) overlaid on corresponding spectra of E-AU-wtTAR and E-GC-wtTAR (green). (C) Normalized resonance intensities (peak heights) of helix measured from 2D HSQC spectra of E-wTAR. See inset for key.

using stretches of either unlabeled A–U (E-AU-wtTAR) or G–C (E-GC-wtTAR) basepairs in a background of uniformly $^{13}C/^{15}N$-labeled G–C or A–U nucleotides, respectively, as described previously [29] (Figure 2.7A).

The E-wtTAR construct more closely resembles TAR in the HIV-1 leader context, where its lower helix is nineteen basepairs long. Interestingly, unlike for TAR_m, elongation of wtTAR resulted in significant chemical shift changes in residues that are distant from the elongation site, including residues in and around the bulge (A22 and C24) and even the very distant apical loop (C30, U31, G34, and A35) (Figure 2.7B). In contrast, chemical shift perturbations are not observed upon elongation for UUCG loop resonances in TAR_m [73]. Compared to the short constructs, spectra of E-wtTAR are in better agreement with those of E-TAR_m, which are in turn very similar to those of TAR_m [29]. Thus, domain elongation diminishes the spectral discrepancies observed in non-elongated wtTAR One possible explanation is that the wt apical loop transiently interacts with the flexible terminal basepairs in domain I in short wtTAR and that this interaction is disrupted or altered upon elongation.

We observe such a long-range interaction in a 65 ns MD simulation of wtTAR [39] in which apical loop residue A35 for which we see largest differences in chemical shifts between short and elongated wtTAR coming within <3Å of G17 at the terminal end of the lower helix I (Figure 2.8B). The ensemble of NOE NMR structures of unbound wtTAR also include conformations in which the same apical loop residue A35 is in direct contact with the lower region of domain I [32] (Figure 2.8A). Such a long-range interaction could indirectly influence the bulge conformation and thereby give rise to the motional correlations between the bulge and apical loop dynamics observed in the iRED study [39]. However, we were unable to detect NOE cross peaks between the apical loop and terminal resides of domain I in wtTAR. Overall, the excellent agreement between the RDC-derived conformations of wtTAR and TAR_m (Figure 2.6B) indicate that any such long-range interaction does not significantly alter the global dynamics of wtTAR relative to TAR_m. Note that even if present transiently, such a long-range interaction should be considered an artifact of the wtTAR construct design since it is apparently diminished in elongated wtTAR which more accurately mimics TAR in its native viral context.

We recently reported ^{15}N relaxation measurements on E-wtTAR [74]. The goal of the latter study was to quantitatively compare the dynamics measured by NMR with those computed by MD. The imino nitrogen spin relaxation order parameters (S) describing the amplitude of fast (S_f) and slow (S_s) ps-ns motions obtained for E-wtTAR were very similar to those determined for E-TAR_m, indicating that the two RNAs have very similar dynamical properties at ps-ns timescales. To more fully characterize the dynamics of E-wtTAR, we measured the C-H resonance intensities, which, ignoring chemical exchange,
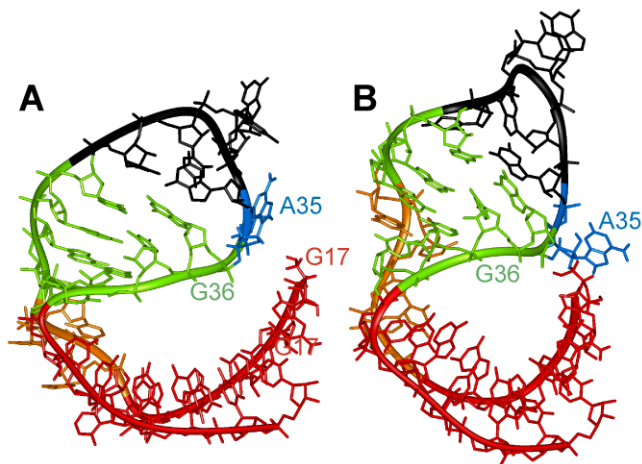
Figure 2.8: Evidence for long-range interaction beween loop and helix I. (A) Structure 18 from free TAR NMR structure ensemble (PDB 1ANR) and (B) snapshot from an MD trajectory of wtTAR. Helix I, the bulge, helix II, and the loop are colored red, orange, green, and black, respectively. A35 is in blue.
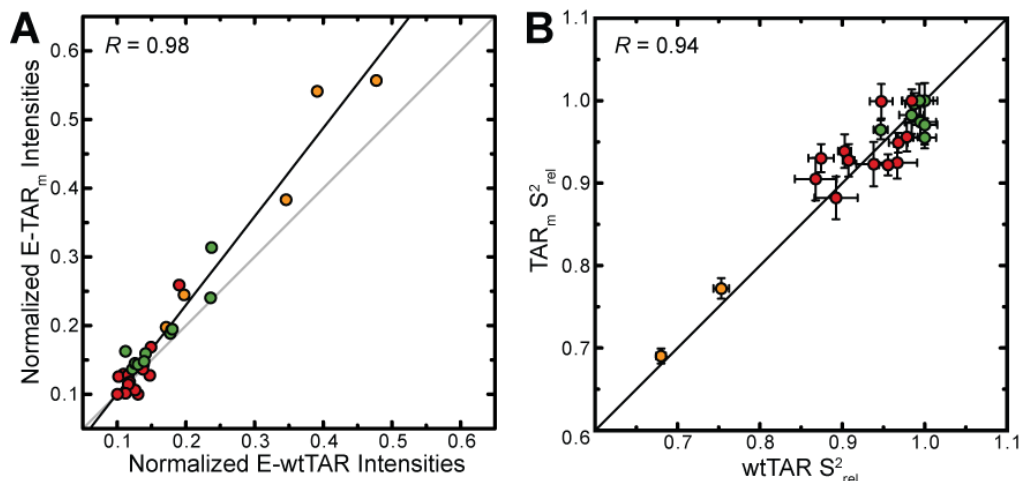
Figure 2.9: Spin relaxation-based comparison of E-wtTAR and E-TAR$_m$ conformational dynamics. (A) Correlation plot between the E-wtTAR and E-TAR$_m$ normalized intensities. The correlation coefficient (R) is shown. (B) Correlation plot between nucleobase carbon S^2_{rel} values measured in non-elongated wtTAR and TAR$_m$. Values for domain I, bulge, and domain II are shown in red, orange, and green circles, respectively.

provide a qualitative measure of internal motions occurring at nanosecond and faster timescales [29,60]. The normalized E-wtTAR intensities were very similar to those of E-TAR$_m$ (Figure 2.9A), again indicating that the two RNAs have very similar dynamical properties. However, the slope of the line deviates from unity and the intensities measured in E-wtTAR and particularly the bulge and domain II are consistently lower than those measured in E-TAR$_m$ (Figure 2.9A). Though the similar RDCs observed at the bulge and neighboring residues in wt-TAR and TAR$_m$ rule out substantial differences in motional amplitudes at sub-millisecond timescales, other factors could influence the observed resonances intensities. The bulge and domain II residues in wtTAR may experience a greater deal of chemical exchange broadening. However, based on carbon $R_{1\rho}$ relaxation dispersion experiments, the only significant exchange contribution is observed in the apical loop of wtTAR, with little to no exchange observed in the bulge region in both TAR$_m$ and wtTAR (discussed in later section).

We also measured longitudinal (R_1) and transverse (R_2) spin relaxation rates for the nucleobase carbons C2, C6 and C8 in the short wtTAR and TAR$_m$. Qualitative relative order parameters (S^2_{rel}) were determined based on computation of $2R_2$-R_1 (see Materials and Methods) which is largely independent of motional timescales [59,60]. Good agreement is observed between the S^2_{rel} values obtained for TAR$_m$ and wtTAR,

70

which show little to no systematic differences (Figure 2.9B). The slope deviation in the resonance intensities observed in the elongated constructs (Figure 2.9A) may be due to slower motional timescales for wtTAR possibly arising due to its larger and more flexible apical loop. Indeed, analysis of ^{15}N relaxation data indicates that the domain motions are slightly slower in E-wtTAR (τ_s = 20.7 ns) [74] compared to E-TAR$_m$ (τ_s = 18.9 ns) [29].

2.3.4 Complex Dynamics in the Apical Loop

Our NMR data revealed complex dynamics over a broad range of timescales in the apical loop. The apical loop RDCs yielded very poor fits to available HIV-1 TAR structures (1ANR, 1ARJ, 1AJU, 1QD3, 1UTS, 1LVJ, 1AKX, 1UUI, and 1UUD [22-25,32,75,76] likely due to the presence of internal motional averaging contributions (data not shown).

Very high resonance intensities in E-wtTAR (Figure 2.10A), low $2R_2$-R_1 values (Figure 2.10A), and near zero RDCs (Figure 2.10B) indicative of high internal mobility at the ps-ns timescales were observed for both base and sugar vectors in G32. In most NMR structures of HIV-1 TAR [22-25,32,75,76], G32 adopts a looped out conformation. CycT1 interacts with the U31 side of the apical loop [6], so the elevated mobility at G32 may be important for adaptive binding. The looped out G32 may also be involved in binding the NC protein which is known to interact with looped out guanine bases [77]. Interestingly, a rather abrupt increase in stability is observed in the neighboring G33 residue (Figure 2.10A,B) particularly for its nucleobase which in MD simulations also shows limited mobility and occasional stacking interactions.

Greater structural stability at the ps-ns timescale is observed for C30 and G34, which exhibit low intensities and sizeable RDCs and S_{rel}^2 values (Figure 2.10A,B). However, the nucleobase of G34 also shows significant dispersion, consistent with a slower dynamical process occurring with a time constant of 30.9 ± 7.1 µs (Figure 2.10C). Biochemical studies [78] and MD simulations [53,79,80] suggest that C30 and G34 form a Watson-Crick basepair that is important for transcriptional activation [7]. Although we were unable to observe the G34 imino proton, fitting of the C30 and G34 RDCs to available NMR structures of TAR yield better agreement with structures in which the two bases are hydrogen bonded (data not shown). In MD simulations of TAR, the C30-G34
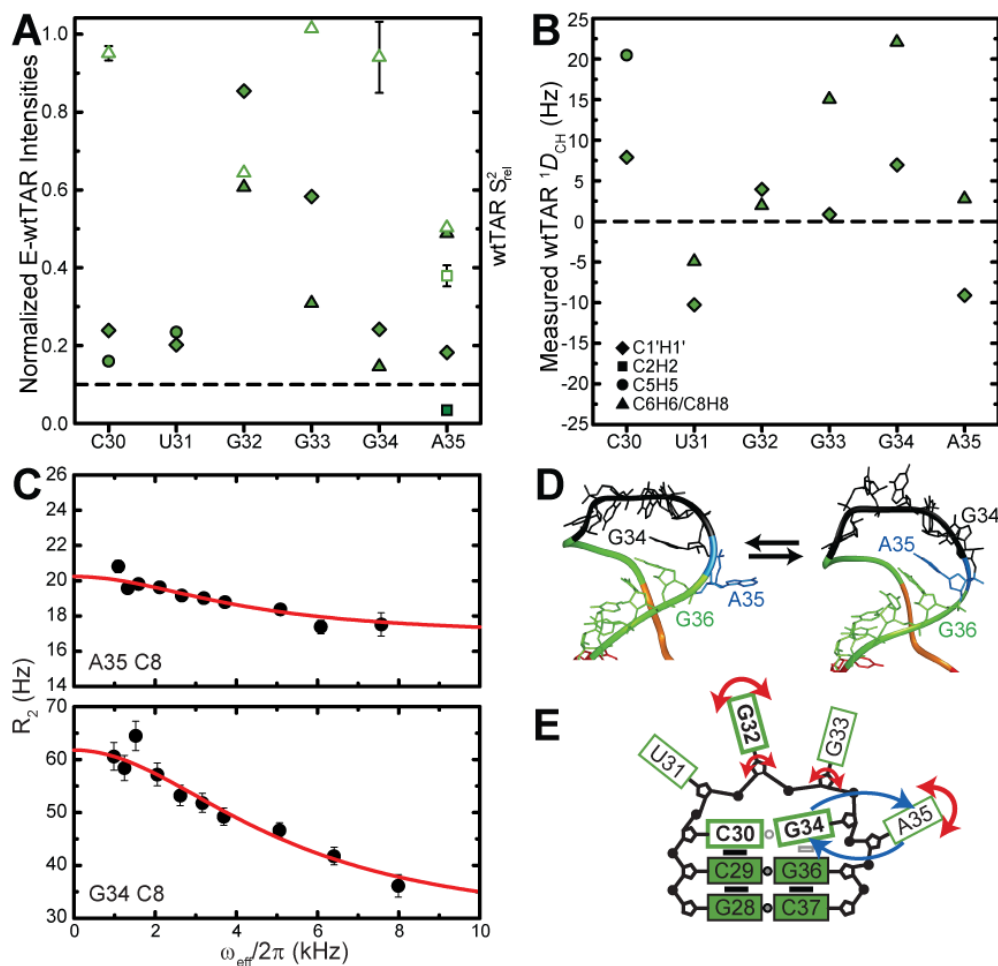
Figure 2.10: Dynamics of the wtTAR loop. Shown are the (A) normalized intensities for E- wtTAR (filled symbols) and S_{rel}^2 values for non-elongated wtTAR (open symbols) and (B) one bond C-H RDCs measured in the wtTAR loop (see inset for key). (C) Carbon $R_{1\rho}$ power dependence for residues G34 (C8) and A35 (C8). (D) Two snapshots from the 65 ns MD simulation of wtTAR illustrating the looping in and out of A35 (in blue). (E) Schematic diagram of the observed wtTAR apical loop dynamics. Nucleobases, ribose moieties, and phosphate groups are represented by green rectangles, black pentagons, and filled black circles, respectively. Filled black rectangles denote base-base stacking. Gray circles indicate hydrogen bonding between bases. The open gray rectangle indicates transient base-base stacking. The open gray circle indicates a transient hydrogen bond across the loop. Fast motions are indicated by red arrows. The looping in and out of A35 and G34 is indicated by blue arrows. Functionally important loop nucleotides are indicated by bolded labels and rectangles.

Watson-Crick basepair is observed to form transiently [39,53] and the observed G34 dispersion likely reflects the dynamics of basepair formation (Figure 2.10C).

Residue A35, which has been shown to be a critical spacer for transactivation [7] exhibits the most complex dynamics (Figure 2.10A,B). The near zero RDCs observed for

both the C2 and C8 vectors suggest they undergo very large motional amplitudes (Figure 2.10B). Though the low S^2_{rel} values observed for these carbons are consistent with fast ps-ns motions, these carbon nuclei also exhibit significant relaxation dispersion or reduced intensities in E-wtTAR, consistent with a slower dynamical process occurring with a time constant of 37.7 ± 10.2 µs (Figure 2.10A,C). The similarity of the timescales for A35 (37.7 µs) and G34 (30.9 µs) suggests that they may be undergoing a concerted dynamical process. In MD simulations, we observe a concerted looping in and out motions in which G34 and A35 displace one another from a looped in conformation (Figure 2.10D). More limited motions are observed for U31, which in NMR structures and MD simulations of TAR is involved in stacking interactions with G32 for the majority of the time (Figure 2.10B).

2.4 Conclusion

Despite its small size, TAR exhibits a remarkable degree of structural and dynamical complexity that is important for recognition of diverse targets. So far, the majority of biophysical studies have focused on characterizing dynamics at the UCU bulge and its importance in adaptive binding. However, few studies have quantitatively characterized the dynamics of the CUGGGA apical loop binding site and its potential role in adaptive recognition. The high flexibility at these two distinct binding sites may also give rise to motional correlations between them as suggested previously [38,39]. Such long-range motional correlations are increasingly implicated in the functions of RNA and are notoriously difficult to characterize experimentally.

Our study outlines a general NMR approach that can be used for studying long-range motional correlations. Here, strategic mutations are introduced at a given site and the consequent effects on structural dynamics at distant sites quantified site-specifically using NMR measurements. Our results suggest that the TAR apical loop and bulge largely act as independent dynamical centers but do not entirely rule out the presence of transient long-range interactions between the apical loop and lower stem seen in MD simulations [39] and the NOE-based NMR structure of TAR [32], which may potentially indirectly affect the conformation of the intervening bulge linker. Such long-range

interactions could provide a general mechanism for creating motional couplings between the apical loops and internal loop junctions.

The TAR apical loop is structurally heterogeneous and undergoes highly complex dynamics with different residues exhibiting very distinct dynamical behaviors over ps-ms timescales (Figure 2.10E). These motions render the apical loop highly malleable and ready for conformational adaptation that may allow recognition of diverse targets. Insights into the structural adaptability of the apical loop are beginning to emerge from structures of TAR complexes. For example, an NMR structure of HIV-2 TAR bound to a cyclic neooligoaminodeoxysaccharide shows that the loop undergoes a large conformational change particularly at residues G32 and G33 and which generally becomes ordered upon binding [36]. Structures of TAR kissing complexes also show the apical loop residues involved in basepairing and assuming a geometry similar to that of an A-form helix [37,81]. Future studies will likely show that in analogy to the highly flexible TAR bulge, the apical loop can adopt highly different conformations and thereby provide a molecular basis for binding a variety of protein targets. The flexibility of the TAR apical loop uncovered in this study should also provide useful guidelines for rationally designing TAR targeting anti-HIV therapeutics.

2.5 References

[1] M.A. Muesing, D.H. Smith, and D.J. Capon, Regulation of mRNA accumilation by a human immunodeficiency virus trans-activator protein. Cell 48 (1987) 691-701.
[2] S. Bannwarth, and A. Gatignol, HIV-1 TAR RNA: the target of molecular interactions between the virus and its host. Curr HIV Res 3 (2005) 61-71.
[3] J. Karn, Tackling Tat. Journal of Molecular Biology 293 (1999) 235-254.
[4] K.A. Jones, Taking a new TAK on tat transactivation. Genes Dev 11 (1997) 2593-9.
[5] B. Majello, G. Napolitano, A. Giordano, and L. Lania, Transcriptional regulation by targeted recruitment of cyclin-dependent CDK9 kinase in vivo. Oncogene 18 (1999) 4598-605.
[6] S. Richter, Y.H. Ping, and T.M. Rana, TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. Proc Natl Acad Sci U S A 99 (2002) 7928-33.
[7] S. Richter, H. Cao, and T.M. Rana, Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. Biochemistry 41 (2002) 6391-7.

[8] I. Kim, C.W. Liu, and J.D. Puglisi, Specific recognition of HIV TAR RNA by the dsRNA binding domains (dsRBD1-dsRBD2) of PKR. J Mol Biol 358 (2006) 430-42.

[9] Y.N. Chang, D.J. Kenan, J.D. Keene, A. Gatignol, and K.T. Jeang, Direct interactions between autoantigen La and human immunodeficiency virus leader RNA. J Virol 68 (1994) 7008-20.

[10] M. Erard, D.G. Barker, F. Amalric, K.T. Jeang, and A. Gatignol, An Arg/Lys-rich core peptide mimics TRBP binding to the HIV-1 TAR RNA upper-stem/loop. J Mol Biol 279 (1998) 1085-99.

[11] I. Kanevsky, F. Chaminade, D. Ficheux, A. Moumen, R. Gorelick, M. Negroni, J.L. Darlix, and P. Fosse, Specific interactions between HIV-1 nucleocapsid protein and the TAR element. J Mol Biol 348 (2005) 1059-77.

[12] Y. Ben-Asouli, Y. Banai, H. Hauser, and R. Kaempfer, Recognition of 5'-terminal TAR structure in human immunodeficiency virus-1 mRNA by eukaryotic translation initiation factor 2. Nucleic Acids Res 28 (2000) 1011-8.

[13] E.S. Andersen, S.A. Contera, B. Knudsen, C.K. Damgaard, F. Besenbacher, and J. Kjems, Role of the trans-activation response element in dimerization of HIV-1 RNA. J Biol Chem 279 (2004) 22243-9.

[14] R. Song, J. Kafaie, and M. Laughrea, Role of the 5' TAR Stem-Loop and the U5-AUG Duplex in Dimerization of HIV-1 Genomic RNA. Biochemistry 47 (2008) 3283-93.

[15] J.D. Puglisi, R. Tan, B.J. Calnan, A.D. Frankel, and J.R. Williamson, Conformation of the TAR RNA-arginine complex by NMR spectroscopy. Science 257 (1992) 76-80.

[16] J.R. Williamson, Induced fit in RNA-protein recognition. Nature Structural Biology 7 (2000) 834-837.

[17] J.M. Perez-Canadillas, and G. Varani, Recent advances in RNA-protein recognition. Current Opinion in Structural Biology 11 (2001) 53-58.

[18] K.S. Long, and D.M. Crothers, Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. Biochemistry 38 (1999) 10059-10069.

[19] M. Zacharias, and P.J. Hagerman, Bulge-induced bends in RNA: quantification by transient electric birefringence. J Mol Biol 247 (1995) 486-500.

[20] J.A. Ippolito, and T.A. Steitz, A 1.3-angstrom resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion- dependent bulge conformation. Proceedings of the National Academy of Sciences of the United States of America 95 (1998) 9819-9824.

[21] T.E. Edwards, and S.T. Sigurdsson, EPR spectroscopic analysis of TAR RNA-metal ion interactions. Biochem Biophys Res Commun 303 (2003) 721-5.

[22] C. Faber, H. Sticht, K. Schweimer, and P. Rosch, Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. J Biol Chem 275 (2000) 20660-6.

[23] Z. Du, K.E. Lind, and T.L. James, Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. Chem Biol 9 (2002) 707-12.

[24] B. Davis, M. Afshar, G. Varani, A.I. Murchie, J. Karn, G. Lentzen, M. Drysdale, J. Bower, A.J. Potter, I.D. Starkey, T. Swarbrick, and F. Aboul-ela, Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". J Mol Biol 336 (2004) 343-56.

[25] A.I. Murchie, B. Davis, C. Isel, M. Afshar, M.J. Drysdale, J. Bower, A.J. Potter, I.D. Starkey, T.M. Swarbrick, S. Mirza, C.D. Prescott, P. Vaglio, F. Aboul-ela, and J. Karn, Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. J Mol Biol 336 (2004) 625-38.

[26] G.L. Olsen, D.C. Echodu, Z. Shajani, M.F. Bardaro, Jr., G. Varani, and G.P. Drobny, Solid-State Deuterium NMR Studies Reveal mus-ns Motions in the HIV-1 Transactivation Response RNA Recognition Site. J Am Chem Soc (2008).

[27] T.E. Edwards, and S. Th Sigurdsson, Electron paramagnetic resonance dynamic signatures of TAR RNA-small molecule complexes provide insight into RNA structure and recognition. Biochemistry 41 (2002) 14843-7.

[28] Q. Zhang, A.C. Stelzer, C.K. Fisher, and H.M. Al-Hashimi, Visualizing spatially correlated dynamics that directs RNA conformational transitions. Nature 450 (2007) 1263-7.

[29] Q. Zhang, X. Sun, E.D. Watt, and H.M. Al-Hashimi, Resolving the motional modes that code for RNA adaptation. Science 311 (2006) 653-6.

[30] H.M. Al-Hashimi, Y. Gosser, A. Gorin, W. Hu, A. Majumdar, and D.J. Patel, Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. J Mol Biol 315 (2002) 95-102.

[31] K.T. Dayie, A.S. Brodsky, and J.R. Williamson, Base Flexibility in HIV-2 TAR RNA Mapped by Solution (15)N, (13)C NMR Relaxation. J Mol Biol 317 (2002) 263-78.

[32] F. Aboul-ela, J. Karn, and G. Varani, Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. Nucleic Acids Research 24 (1996) 3974-3981.

[33] H.M. Al-Hashimi, A. Gorin, A. Majumdar, Y. Gosser, and D.J. Patel, Towards structural genomics of RNA: rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings. J Mol Biol 318 (2002) 637-49.

[34] H.Y. Mei, M. Cui, A. Heldsinger, S.M. Lemrow, J.A. Loo, K.A. Sannes-Lowery, L. Sharmeen, and A.W. Czarnik, Inhibitors of protein-RNA complexation that target the RNA: specific recognition of human immunodeficiency virus type 1 TAR RNA by small organic molecules. Biochemistry 37 (1998) 14204-12.

[35] A. Litovchick, A.G. Evdokimov, and A. Lapidot, Aminoglycoside-arginine conjugates that bind TAR RNA: synthesis, characterization, and antiviral activity. Biochemistry 39 (2000) 2838-52.

[36] D. Raghunathan, V.M. Sanchez-Pedregal, J. Junker, C. Schwiegk, M. Kalesse, A. Kirschning, and T. Carlomagno, TAR-RNA recognition by a novel cyclic aminoglycoside analogue. Nucleic Acids Res 34 (2006) 3599-608.

[37] I. Lebars, T. Richard, C. Di Primo, and J.J. Toulme, NMR structure of a kissing complex formed between the TAR RNA element of HIV-1 and a LNA-modified aptamer. Nucleic Acids Res 35 (2007) 6103-14.

[38] H. Huthoff, F. Girard, S.S. Wijmenga, and B. Berkhout, Evidence for a base triple in the free HIV-1 TAR RNA. Rna 10 (2004) 412-23.

[39] C. Musselman, H.M. Al-Hashimi, and I. Andricioaei, iRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. Biophys J 93 (2007) 411-22.

[40] R.A. Tinsley, and N.G. Walter, Long-range impact of peripheral joining elements on structure and function of the hepatitis delta virus ribozyme. Biol Chem 388 (2007) 705-15.

[41] D. Rueda, G. Bokinsky, M.M. Rhodes, M.J. Rust, X. Zhuang, and N.G. Walter, Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. Proc Natl Acad Sci U S A 101 (2004) 10066-71.

[42] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, Nmrpipe - a Multidimensional Spectral Processing System Based On Unix Pipes. Journal of Biomolecular Nmr 6 (1995) 277-293.

[43] T.D. Goddard, and D.G. Kneller, SPARKY 3, University of California, San Franscisco.

[44] B. Furtig, C. Richter, J. Wohnert, and H. Schwalbe, NMR spectroscopy of RNA. Chembiochem 4 (2003) 936-62.

[45] V. Sklenar, R.D. Peterson, M.R. Rejante, and J. Feigon, Two- and three-dimensional HCN experiments for correlating base and sugar resonances in 15N,13C-labeled RNA oligonucleotides. J Biomol NMR 3 (1993) 721-7.

[46] J. Wohnert, M. Gorlach, and H. Schwalbe, Triple resonance experiments for the simultaneous correlation of H6/H5 and exchangeable protons of pyrimidine nucleotides in 13C,15N-labeled RNA applicable to larger RNA molecules. J Biomol NMR 26 (2003) 79-83.

[47] A. Meissner, and O.W. Sorensen, The role of coherence transfer efficiency in design of TROSY- type multidimensional NMR experiments. Journal of Magnetic Resonance 139 (1999) 439-442.

[48] S.W. Pitt, Q. Zhang, D.J. Patel, and H.M. Al-Hashimi, Evidence that electrostatic interactions dictate the ligand-induced arrest of RNA global flexibility. Angew Chem Int Ed Engl 44 (2005) 3412-5.

[49] M.R. Hansen, P. Hanson, and A. Pardi, Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. Methods in Enzymology 317 (2000) 220-240.

[50] G.M. Clore, M.R. Starich, and A.M. Gronenborn, Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod- shaped viruses. Journal of the American Chemical Society 120 (1998) 10571-10572.

[51] M. Getz, X. Sun, A. Casiano-Negroni, Q. Zhang, and H.M. Al-Hashimi, Review NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. Biopolymers 86 (2007) 384-402.

[52] A. Casiano-Negroni, X. Sun, and H.M. Al-Hashimi, Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. Biochemistry 46 (2007) 6525-35.

[53] C. Musselman, S.W. Pitt, K. Gulati, L.L. Foster, I. Andricioaei, and H.M. Al-Hashimi, Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. J Biomol NMR 36 (2006) 235-49.

[54] S. Neidle, Oxford Handbook of Nucleic Acid Structure, Oxford University Press, New York, 1999.

[55] J.A. Losonczi, M. Andrec, M.W.F. Fischer, and J.H. Prestegard, Order matrix analysis of residual dipolar couplings using singular value decomposition. Journal of Magnetic Resonance 138 (1999) 334-342.

[56] A.L. Hansen, and H.M. Al-Hashimi, Insight into the CSA tensors of nucleobase carbons in RNA polynucleotides from solution measurements of residual CSA: towards new long-range orientational constraints. J Magn Reson 179 (2006) 299-307.

[57] M.H. Bailor, C. Musselman, A.L. Hansen, K. Gulati, D.J. Patel, and H.M. Al-Hashimi, Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. Nat Protoc 2 (2007) 1536-46.

[58] A.G. Palmer, 3rd, and F. Massi, Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. Chem Rev 106 (2006) 1700-19.

[59] D. Fushman, N. Tjandra, and D. Cowburn, An approach to direct determination of protein dynamics from N-15 NMR relaxation at multiple fields, independent of variable N-15 chemical shift anisotropy and chemical exchange contributions. Journal of the American Chemical Society 121 (1999) 8577-8582.

[60] A.L. Hansen, and H.M. Al-Hashimi, Dynamics of large elongated RNA by NMR carbon relaxation. J Am Chem Soc 129 (2007) 16072-82.

[61] J.A. Jones, Optimal sampling strategies for the measurement of relaxation times in proteins. Journal Of Magnetic Resonance 126 (1997) 283-286.

[62] A.D. MacKerell, N. Banavali, and N. Foloppe, Development and current status of the CHARMM force field for nucleic acids. Biopolymers 56 (2000) 257-265.

[63] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water. Journal of Chemical Physics 79 (1983) 926-935.

[64] C.L. Brooks, and M. Karplus, Deformable Stochastic Boundaries in Molecular-Dynamics. Journal of Chemical Physics 79 (1983) 6312-6325.

[65] S. Nose, A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods. Journal of Chemical Physics 81 (1984) 511-519.

[66] W.G. Hoover, Canonical Dynamics - Equilibrium Phase-Space Distributions. Physical Review A 31 (1985) 1695-1697.

[67] K.M. Weeks, C. Ampe, S.C. Schultz, T.A. Steitz, and D.M. Crothers, Fragments of the HIV-1 Tat protein specifically bind TAR RNA. Science 249 (1990) 1281-5.

[68] S.W. Pitt, A. Majumdar, A. Serganov, D.J. Patel, and H.M. Al-Hashimi, Argininamide binding arrests global motions in HIV-1 TAR RNA: comparison with Mg2+-induced conformational stabilization. J Mol Biol 338 (2004) 7-16.

[69] N. Tjandra, and A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278 (1997) 1111-1114.

[70] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, and J.H. Prestegard, Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information For Structure Determination in Solution. Proceedings of the National Academy of Sciences of the United States of America 92 (1995) 9279-9283.

[71] J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, and J.H. Prestegard, Structural and dynamic analysis of residual dipolar coupling data for proteins. Journal of the American Chemical Society 123 (2001) 1416-1424.

[72] G. Lipari, and A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. Journal of the American Chemical Society 104 (1982) 4546-4559.

[73] M.M. Getz, A.J. Andrews, C.A. Fierke, and H.M. Al-Hashimi, Structural plasticity and Mg2+ binding properties of RNase P P4 from combined analysis of NMR residual dipolar couplings and motionally decoupled spin relaxation. Rna (2006).

[74] C. Musselman, Q. Zhang, H.M. Al-Hashimi, and I. Andricioaei, A domain-anchored referencing strategy for the direct comparison of NMR and MD motional parameters in RNA. (Submitted).

[75] F. Aboul-ela, J. Karn, and G. Varani, The Structure of the Human-Immunodeficiency-Virus Type-1 Tar RNA Reveals Principles of RNA Recognition By Tat Protein. Journal of Molecular Biology 253 (1995) 313-332.

[76] A.S. Brodsky, and J.R. Williamson, Solution structure of the HIV-2 TAR-argininamide complex. Journal of Molecular Biology 267 (1997) 624-639.

[77] R.N. De Guzman, Z.R. Wu, C.C. Stalling, L. Pappalardo, P.N. Borer, and M.F. Summers, Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. Science 279 (1998) 384-8.

[78] T. Kulinski, M. Olejniczak, H. Huthoff, L. Bielecki, K. Pachulska-Wieczorek, A.T. Das, B. Berkhout, and R.W. Adamiak, The apical loop of the HIV-1 TAR RNA hairpin is stabilized by a cross-loop base pair. J Biol Chem 278 (2003) 38892-901.

[79] A.D. Critchley, I. Haneef, D.J. Cousens, and P.G. Stockley, Modeling and solution structure probing of the HIV-1 TAR stem-loop. J Mol Graph 11 (1993) 92-7, 124.

[80] R. Nifosi, C.M. Reyes, and P.A. Kollman, Molecular Dynamics Studies of the HIV-1 TAR and its Complex with Argininamide. Nucleic Acids Research (2000) 4944-4955.

[81] K.Y. Chang, and I. Tinoco, Jr., The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement. J Mol Biol 269 (1997) 52-66.

Chapter 3

Variable helix elongation as a tool to modulate RNA alignment and motional couplings

3.1 Introduction

There is great interest in utilizing NMR residual dipolar couplings (RDCs) [1,2] to characterize the structure and dynamics of biomolecules [3-5]. RDCs can be measured in molecules that are partially aligned, either spontaneously when they have a significant magnetic susceptibility anisotropy [1,6] or, more commonly, by dissolution in an appropriate ordering medium [2,7-10]. Several studies have established the utility of measuring RDCs under multiple linearly independent alignment conditions [11] to increase the spatial resolution with which structure and dynamics can be simultaneously characterized [12-18].

Two challenges can arise when using RDCs to characterize the conformational dynamics of nucleic acids, particularly globally flexible RNAs. First, it cannot be generally assumed, as is done in most formalisms, that internal motions do not lead to coupled changes in overall alignment [12,13,19]. Rather, rigid-body collective movements of helical domains about flexible junctions can lead to significant changes in the RNA overall molecular shape and thus global alignment [20,21]. Second, though the alignment of proteins can be modulated by changing the alignment medium and altering the balance between electrostatic and steric forces governing alignment [11,22], this has proven difficult if not impossible for nucleic acids because their uniform charge distribution closely follows that of their molecular shape [23-28]. Although spontaneous magnetic field alignment can yield one additional independent alignment [26,28], the degree of order generally remains unfavorably small under current magnetic field strengths.

Recently, we showed that correlations between internal motions and overall alignment could be reduced and overall alignment altered by chemically perturbing the overall RNA molecular shape [20]. In particular, the alignment of the transactivation response element (TAR) RNA from HIV-1 dissolved in the Pf1 phage ordering medium [8,9] was modulated by independently elongating each of its two helices by twenty-two Watson-Crick base-pairs [20,29]. The elongation renders the overall molecular shape, and consequently overall alignment, far less sensitive to collective motions of helices. It also serves to predefine overall alignment by fixing the principal axis of order (S_{zz}) to be nearly parallel to the elongated helix axis. A similar mutagenesis strategy that serves to alter the surface electrostatic properties has successfully been used by Bax and co-workers to modulate the alignment of proteins [30].

In RNA, extensive helix elongation (>20 base-pairs) is generally required to adequately decouple internal and overall motions. However, this results in an unfavorable increase in the size of the RNA under study and also necessitates preparation of two specifically labeled samples to minimize spectral overcrowding from elongation residues [29]. Many RNAs, including TAR bound to ligands [31] or at high monovalent and divalent ion concentrations [32], are globally rigid and may not require extensive helix elongation to decouple motions. In these cases, moderate degrees of elongation may offer an approach for modulating alignment. Even for globally flexible RNAs, moderate elongation may help expose collective helix motions and provide physical insights into the motional couplings themselves [33]. Here, we examine the utility of moderate degrees of helix elongation in modulating both the alignment and degree of motional couplings in a TAR mutant, EI(3)-TAR$_m$, in which the apical wild-type loop has been replaced with a UUCG loop and in which the terminal helix is elongated by three base-pairs (Figure 3.1A).

3.2 Materials and Methods

3.2.1 Preparation and Purification of Uniformly ^{13}C/^{15}N Labeled RNA

We prepared uniformly ^{13}C/^{15}N labeled EI(3)-TAR$_m$ by *in vitro* transcription as previously described previously [34] and in Chapter 2. EI(3)-TAR$_m$ was prepared by *in vitro* transcription using T7 RNA polymerase (Takara Mirus Bio, Inc.), uniformly ^{13}C/^{15}N

labeled nucleotide triphosphates (ISOTEC, Inc.), and a synthetic DNA template (Integrated DNA Technologies, Inc.) containing the T7 promoter and sequence of interest. The RNA was purified using 20% (w/v) denaturing polyacrylamide gel electrophoresis, eluted from the gel in 20 mM Tris pH 8 buffer followed by ethanol precipitation. The RNA pellet was dissolved in water and exchanged into NMR buffer (15 mM sodium phosphate, 0.1 mM EDTA, and 25 mM NaCl at pH ~6.4) multiple times using a Centricon Ultracel YM-3 concentrator (Millipore Corp.). The final RNA concentration was ~1.0 mM. The aligned NMR sample was prepared by adding Pf1 phage (Asla Biotech Ltd.) in NMR buffer to yield a final Pf1 phage concentration of ~22 mg/mL and a final RNA concentration of ~0.6mM. The addition of phage did not affect the average structure of RNA as judged from careful comparison of the chemical shifts in the absence and presence of phage (data not shown).

3.2.2 Resonance Assignments

All EI(3)-TAR$_m$ experiments were conducted in NMR buffer at 298 K on an Avance Bruker 600 MHz NMR spectrometer equipped with a triple-resonance 5 mm cryogenic probe. NMR spectra were analyzed using NMR Draw [35] and Sparky 3 [36]. ^{1}H, ^{13}C, and ^{15}N resonances were assigned by spectra overlay [29,37] and using standard homonuclear and heteronuclear 2D experiments, including an exchangeable NOESY and a 2D HCN experiment that correlates intranucleotide H8/H6 to N1/N9 to H1′. The exchangeable NOE walk was uninterrupted from G15 to G21 and G26 to G28, indicating a continuous helical conformation in these regions. The NOE between G28 and G36 was not observed due to spectral overlap.

3.2.3 Measurement and Order Tensor Analysis of RDCs

2D ^{13}C–^{1}H (or ^{15}N-^{1}H) S^{3}E HSQC experiments[31,38] were used to measure one-bond $^1D_{C6H6}$, $^1D_{C8H8}$, $^1D_{C5H5}$, $^1D_{C2H2}$, $^1D_{C1'H1'}$, and $^1D_{N1/3H1/3}$ RDCs in EI(3)-TAR$_m$ using ~22 mg/ml of Pf1 phage[8,9] as an ordering medium. RDC measurement error was estimated from two sets of experiments that yielded either splittings along the ^{1}H dimension or ^{15}N/^{13}C dimension, as previously described [5,32]. The measured EI(3)-TAR$_m$ RDCs are listed in Table 3.1.

Residue (bond vector)	RDC (Hz)	Residue (bond vector)	RDC (Hz)
G15 (C8H8)	-3.8	*U32 (C6H6)*	*19.5*
G15 (N1H1)	0.7	*U32 (C1'H1')*	*16.5*
A16 (C8H8)	-9	*C33 (C5H5)*	*21.2*
C17 (C6H6)	16	*C33 (C6H6)*	*16.1*
U18 (C5H5)	0.2	*C33 (C1'H1')*	*7.8*
U18 (N3H3)	-15	G34 (C1'H1')	23.7
C19 (C6H6)	31.9	G34 (N1H1)	-13.2
A20 (C2H2)	34.3	G36 (C8H8)	26.4
A20 (C8H8)	26.4	G36 (N1H1)	-13.2
G21 (C8H8)	21.2	C37 (C5H5)	31.1
G21 (C1'H1')	-4.7	C37 (C6H6)	31.7
A22 (C2H2)	*15.8*	C37 (C1'H1')	-4
A22 (C8H8)	*19.4*	U38 (C5H5)	26.2
A22 (C1'H1')	*11.9*	U38 (C6H6)	32.4
U23 (C5H5)	*1.8*	U38 (N3H3)	-15.2
U23 (C6H6)	*22.4*	C39 (C5H5)	30.2
U23 (C1'H1')	*4.8*	*U40 (C5H5)*	*28.4*
C24 (C5H5)	*5.8*	*U40 (C1'H1')*	*-11.9*
C24 (C6H6)	*5.5*	C41 (C5H5)	31.3
U25 (C5H5)	*-1.5*	U42 (C5H5)	23.7
U25 (C6H6)	*11.2*	U42 (C1'H1')	2.8
G26 (C8H8)	25.4	U42 (N3H3)	-8.5
G26 (N1H1)	-14	G43 (C8H8)	26.8
A27 (C2H2)	23	G43 (N1H1)	-14.9
A27 (C8H8)	32.2	A44 (C8H8)	34.3
G28 (C8H8)	33.1	A44 (C2H2)	25.1
C29 (C6H6)	32.9	G45 (C8H8)	30.7
U31 (C5H5)	27.7	G45 (C1'H1')	-1.6
U31 (C6H6)	28.2	U46 (C5H5)	33.8
U31 (C1'H1')	-1	U46 (N3H3)	-4
U32 (C5H5)	*7.6*	C47 (C5H5)	26

Table 3.1: RDCs measured in EI(3)-TAR$_m$. Values shown in italic were excluded from the order tensor analysis. RDCs from the locally flexible residues A22-U40 and apical loop residues U32 and C33 [39] were omitted from analysis.

The RDCs measured in Watson-Crick base pairs in domains I and II were subjected to an order tensor analysis using idealized A-form helices as input coordinates [40] as described in detail in Chapter 2. Briefly, the measured RDCs and idealized A-

form helices were used to determine best-fit order tensors for each domain using singular value decomposition [41] implemented by the in-house written program RAMAH [42]. An in-house program (AFORM-RDC) was used to determined the error in the order tensor errors due to "A-form structural noise" and RDC uncertainty [40]. The overall RNA structure was assembled by rotating each domain into the principal axis system (PAS) of its best-fitted order tensor and then assembling the two helices using the program Insight II (Molecular Simulations, Inc). RDCs from flexible residues and the terminal end were excluded from analysis [40,43]. Inter-helical angles for all RNA structures were calculated using the in-house program Euler-RNA as previously described [44].

3.3 Results and Discussion

As shown in Figure 3.1B, spectra of EI(3)-TAR$_m$ are in excellent agreement with those of non-elongated TAR$_m$, indicating that the elongation does not affect its structural and dynamical integrity, as reported previously for a twenty-two base-pair elongated TAR sample (EI(22)-TAR$_m$, (Figure 3.1A) [20,29]. A total of 9 N-H and 53 C-H RDCs (Table 3.1) were measured in sugar (C1'H1') and base (C2H2, C5H5, C6H6, C8H8, N1H1, and N3H3) moieties of EI(3)-TAR$_m$ using phage as an ordering medium [8,9,45].
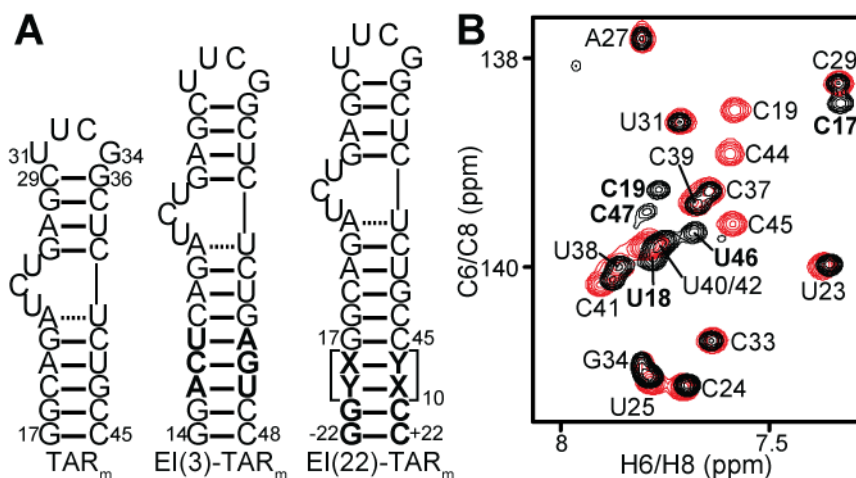
Figure 3.1: TAR$_m$ constructs. (a) Secondary structures of TAR$_m$, EI(3)-TAR$_m$, and EI(22)-TAR$_m$. Differences between constructs are shown in bold. (b) 2D ^{1}H-^{13}C HSQC spectrum of EI(3)-TAR$_m$ (black) overlaid on the corresponding spectrum of TAR$_m$ (red). TAR$_m$ resonances are labeled in black, while peaks that belong only to EI(3)-TAR$_m$ are shown in bold.
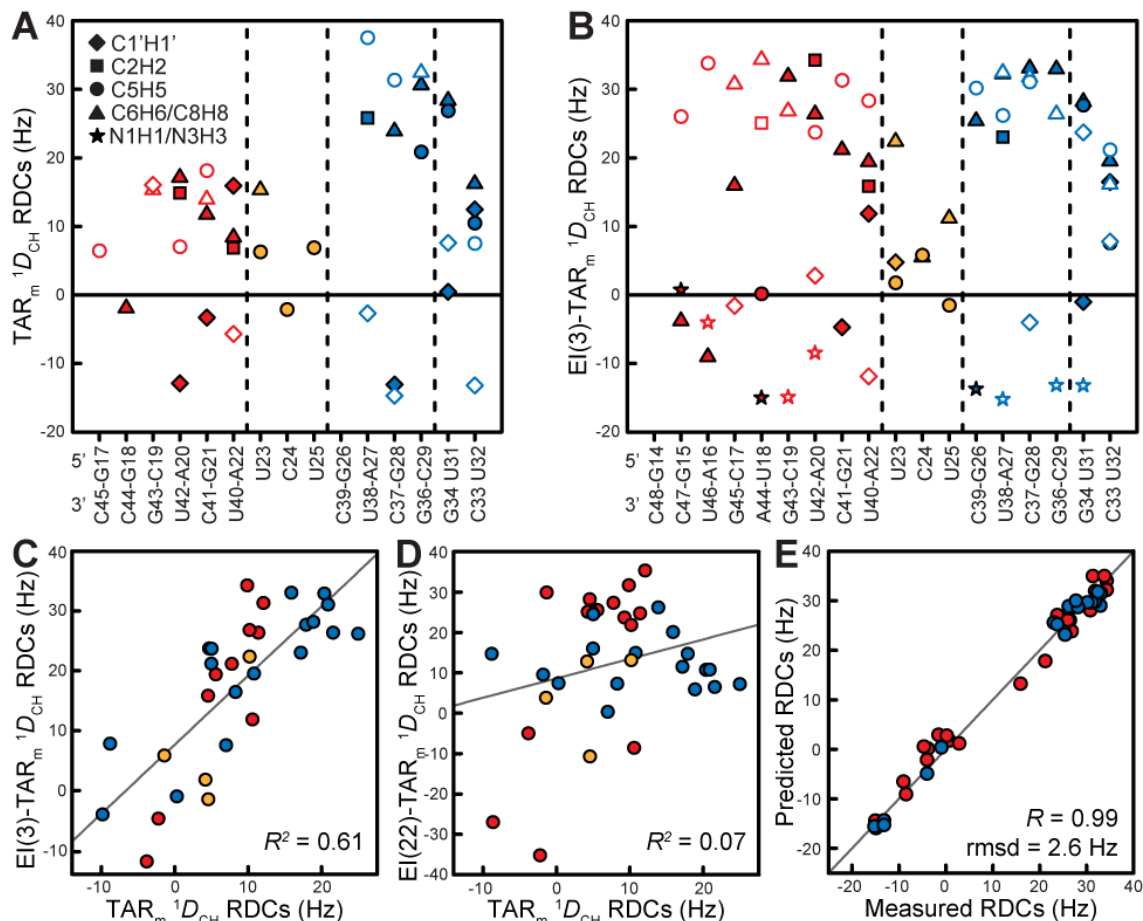
Figure 3.2: EI(3)-TAR$_m$ RDCs. One bond RDCs ($^1D_{CH}$ and $^1D_{NH}$) measured in (A) TAR$_m$ and (B) EI(3)-TAR$_m$ as a function of residue/secondary structure. Filled and open circles represent RDCs measured on the 5′ and 3′ strand, respectively. RDCs have been normalized to those of EI(3)-TAR$_m$ by the ratio of the ϑ_{II} values of EI(3)-TAR$_m$ and TAR$_m$. See inset in Figure 2A for key. (C-D) Correlation plots between (C) TAR$_m$ and EI(3)-TAR$_m$ and (D) TAR$_m$ and EI(22)-TAR$_m$ RDCs. (E) Comparison of EI(3)-TAR$_m$ RDCs measured in helix I (red) and helix II (blue) with values back-predicted using the best-fit order tensor and an idealized A-form helix and X-ray structure of the loop.

A plot of RDCs as a function of secondary structure is shown in Figure 3.2A and 2B for TAR$_m$ and EI(3)-TAR$_m$, respectively. The RDCs measured in helix I of TAR$_m$ are consistently attenuated relative to counterparts in helix II. This has previously been attributed to inter-helical motions and a smaller degree of order for helix I compared to helix II [46]. Differences between helix RDCs are far less pronounced in EI(3)-TAR$_m$, indicating that elongation of helix I by three base-pairs increases its degree of order relative to helix II. The poor agreement (R^2=0.61) observed between the EI(3)-TAR$_m$ and TAR$_m$ RDCs (Figure 3.2C) qualitatively suggests that the three-base pair elongation

modulates overall alignment and/or motional couplings. An even poorer correlation ($R^2=0.07$) is observed upon elongating helix I by twenty-two base-pairs (Figure 3.2D), consistent with elongation-dependent modulation of alignment and/or motional couplings.

To characterize the alignment of EI(3)-TAR$_m$ in Pf1 phage, we used the RDCs measured in non-terminal Watson-Crick base-pairs and structurally stable loop residues to determine order tensors for each helix. The order tensors were determined by fitting measured RDCs to an idealized A-form helical geometry and an X-ray structure of the loop [47] as described previously [40,44,48]. An excellent fit was observed between the measured and back-predicted RDCs, indicating that both EI(3)-TAR$_m$ helices are accurately modeled using an idealized A-form geometry as described previously for non-elongated [46] and elongated [20] TAR (Figure 3.2E).

In Figure 3.3, we use a Sauson-Flamsteed map to depict the best-fit principal orientational solutions (S_{xx}, S_{yy}, and S_{zz}) obtained for each helix in EI(3)-TAR$_m$. The orientational solutions are shown relative to a molecular frame in which the helix axis is oriented along the molecular z direction. For comparison, the corresponding order tensors reported previously for TAR$_m$ [46] and EI(22)-TAR$_m$ [20] are also shown. The estimated uncertainty in the S_{zz} orientations determined for helix I/II using the program Aform-RDC [40] is $7^o/6^o$, $2^o/2^o$, and $2^o/4^o$ for TAR$_m$, EI(3)-TAR$_m$, and EI(22)-TAR$_m$, respectively. In TAR$_m$, S_{zz} deviates by only ~7° from the helix II axis but deviates by ~40° from the helix I axis. As expected, elongation of helix I by three base-pairs tips S_{zz} towards its helical axis such that S_{zz} deviates by only ~17° and is nearly fully aligned with the axis following a twenty-two base-pair elongation (~7°). The three S_{zz} directions in TAR$_m$, EI(3)-TAR$_m$, and EI(22)-TAR$_m$ do not fall along a common plane, most likely because the A-form helices deviate from perfect cylindrical axial symmetry (Figure 3.3). The three base-pair elongation has an opposite, albeit smaller effect on helix II. The angle between S_{zz} and the helix II axis increases from ~7° to ~16° to ~32° upon elongation by three and twenty-two base-pairs, respectively (Figure 3.3). Changes are also observed in the S_{xx} and S_{yy} orientations (Figure 3.3) and the asymmetry parameter (η) (Figure 3.4A and 4B), indicating that elongation modulates all five elements of the order tensor.
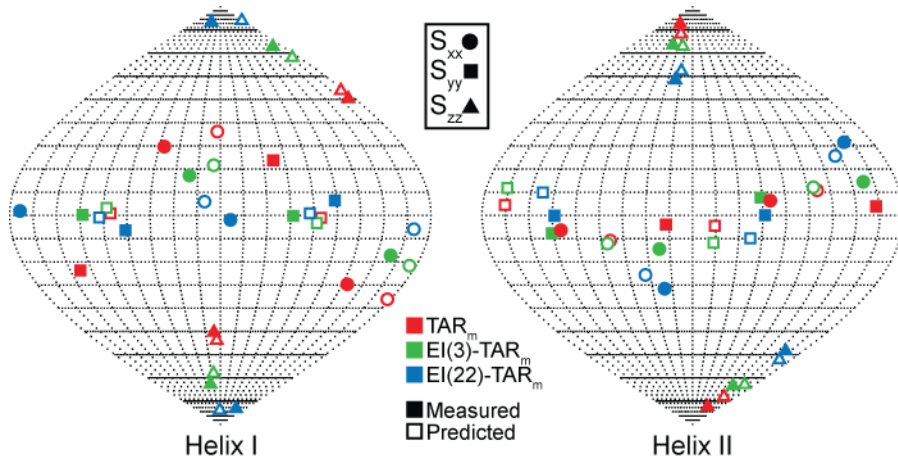
Figure 3.3: Measured and predicted alignment of variable elongated TAR. Sauson-Flamsteed maps showing the best-fit measured (filled symbols) and PALES-predicted (open symbols) order tensor frames (S_{xx}, S_{yy}, and S_{zz}) for helices I and II. Solutions are depicted relative to a molecular frame in which the helix axis is oriented along the z direction.

Elongation of helix I leads to changes in relative degrees of helix order that mirror those observed for the principal S_{zz} direction. In TAR_m, helix II dominates overall alignment, yielding an internal generalized degree of order $\vartheta_{int} = \vartheta_I/\vartheta_{II} = 0.56\pm0.1$ that is consistent with large amplitude, inter-helical motions [46]. Elongating helix I by three base-pairs increases its level of order compared to helix II such that both helices now have comparable degrees of order ($\vartheta_{int} = \vartheta_I/\vartheta_{II} = 0.92\pm0.04$). This, together with the similar angles between the S_{zz} direction and the two helical axes (Figure 3.3), strongly suggests that the three base-pair elongation equalizes contributions of the two helices to overall alignment and pushes TAR into the motional coupling limit. Extending the elongation of helix I to twenty-two base-pairs drives the system towards the decoupling limit in which helix I dominates alignment ($\vartheta_{int} = \vartheta_{II}/\vartheta_I = 0.45\pm0.05$). Because RNA helices are often not coaxially stacked, moderate degrees of elongation are expected to modulate alignment. This was verified using PALES simulations on a variety of RNAs following elongation of the terminal helix by three base-pairs (data not shown). The combination of helix-elongation with measurements of magnetic field-induced RDCs [26,28] may make it possible to access the maximum of five linearly independent sets of RDCs in nucleic acids. Access to such measurements will make it possible to more fully characterize the structural dynamics of RNA.

To rationalize the observed trends in the alignment of TAR as a function of elongation, we used the three-state dynamical ensemble determined previously using RDCs [20] in conjunction with electrostatic-induced alignment simulations using the program PALES [23,49] to simulate TAR alignment in phage at varying lengths of helix I elongation. The three-state ensemble is a discrete approximation of what is likely a continuous distribution involving many conformations. While we do not expect to observe quantitative agreement between measured alignment parameters and values predicted using this three-state ensemble approximation, we would like to obtain insights into the general trends observed for orientation and alignment upon helix elongation. An elongated helix I was superimposed onto each of the three TAR conformers of the ensemble, and PALES was used to predict the RDCs for each conformer. Bulge residues were not included in the three-state dynamical ensemble, so they were not included in PALES calculations. The three sets of RDCs simulated for each conformer were then averaged and used to compute order tensors for each helix. As shown in Figure 3.3, we observe very good agreement between measured and predicted S_{zz} directions. The deviations range between 2-11° and 6-8°, for helices I and II, respectively. The simulations reproduce the observed non-planar approach of the S_{zz} direction towards the helix I axis with increasing elongation. Interestingly, the S_{xx} and S_{yy} directions (Figure 3.3) and the asymmetry (Figure 3.4A and 4B) are also reasonably well reproduced considering their much larger experimental uncertainty and given that exclusion of the bulge has a strong effect on these predicted values (data not shown). The simulations reproduce the increase in the helix I order compared to helix II upon helix I elongation (Figure 3.4C). Also, the simulated ϑ_{int} values reflect the observed trend. The larger discrepancy in TAR_m could arise from neglect of the bulge and fraying motions at terminal base-pairs which are likely to have a bigger effect on the small TAR_m compared to the other elongated constructs. Indeed, better agreement is observed for TAR_m when excluding the two terminal base-pairs in helix I (Figure 3.4C).

Independent RDC data sets are often used to overcome the 4^{n-1} fold orientational degeneracy arising from superposition of the order tensors of *n* domains [11,22]. One of the consequences of having differential motional averaging effects due to inter-domain motions is that two sets of RDCs may not be reconcilable with a single inter-domain
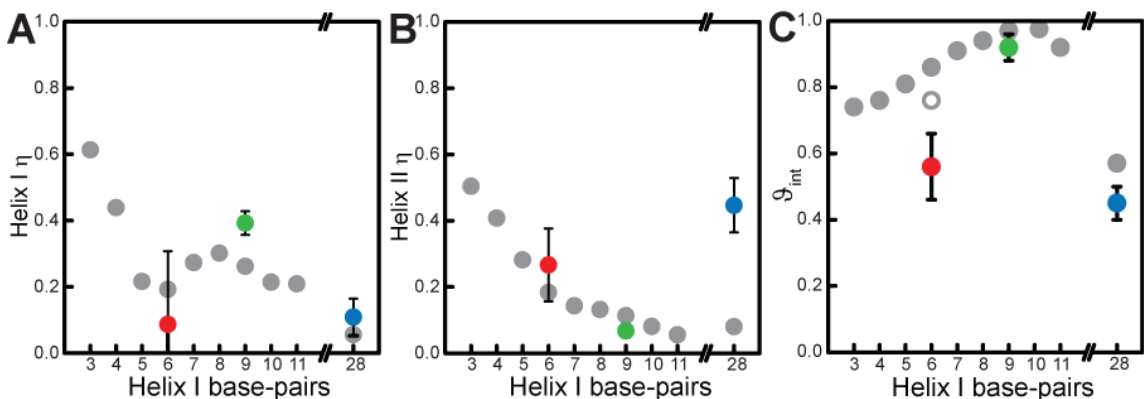
Figure 3.4: Order tensor parameters as a function of helix length. The measured and PALES-predicted η parameters for (A) helix I and (B) helix II. (C) ϑ_{int} values ($\vartheta_{int} = \vartheta_{II}/\vartheta_{I}$) as a function of helix I length. The value for EI(22)-TAR$_m$ corresponds to $\vartheta_{int} = \vartheta_{I}/\vartheta_{II}$. The open circle corresponds to the predicted ϑ_{int} value of TAR$_m$ when the two terminal base-pairs are excluded. Measured values and PALES-predicted values are in color and gray, respectively. See Figure 3.3 for key.

orientation [50]. This is the case for the RDC datasets measured in TAR$_m$, EI(3)-TAR$_m$, and EI(22)-TAR$_m$. As shown in Figure 3.5A, none of the four inter-helix structures assembled using the RDCs measured in TAR$_m$ satisfy the orientational solutions obtained using the EI(3)-TAR$_m$ RDCs. Similarly, a common solution does not exist when using RDCs from EI(22)-TAR$_m$ (Figure 3.5B). Thus, the inability to reconcile RDCs measured in differentially elongated RNAs in this manner may be an indication that inter-helix motions are present. However, other sources of experimental uncertainty should also be ruled out.

3.4 Conclusion

In conclusion, we show that modest degrees of helix-elongation can be used to modulate both overall alignment of RNA and the degree of motional couplings in a semi-predictable manner. Our results underscore the importance of exercising caution in interpreting similar levels of order for two domains ($\vartheta_{int} \sim 1$) in terms of inter-domain rigidity. In general, motional couplings obscure inter-domain motion and will often result in underestimated dynamics. Our data suggest that even moderate degrees of helix elongation may be used to push an RNA system outside the motionally coupled regime, though the degree of elongation needed will obviously vary from RNA to RNA.
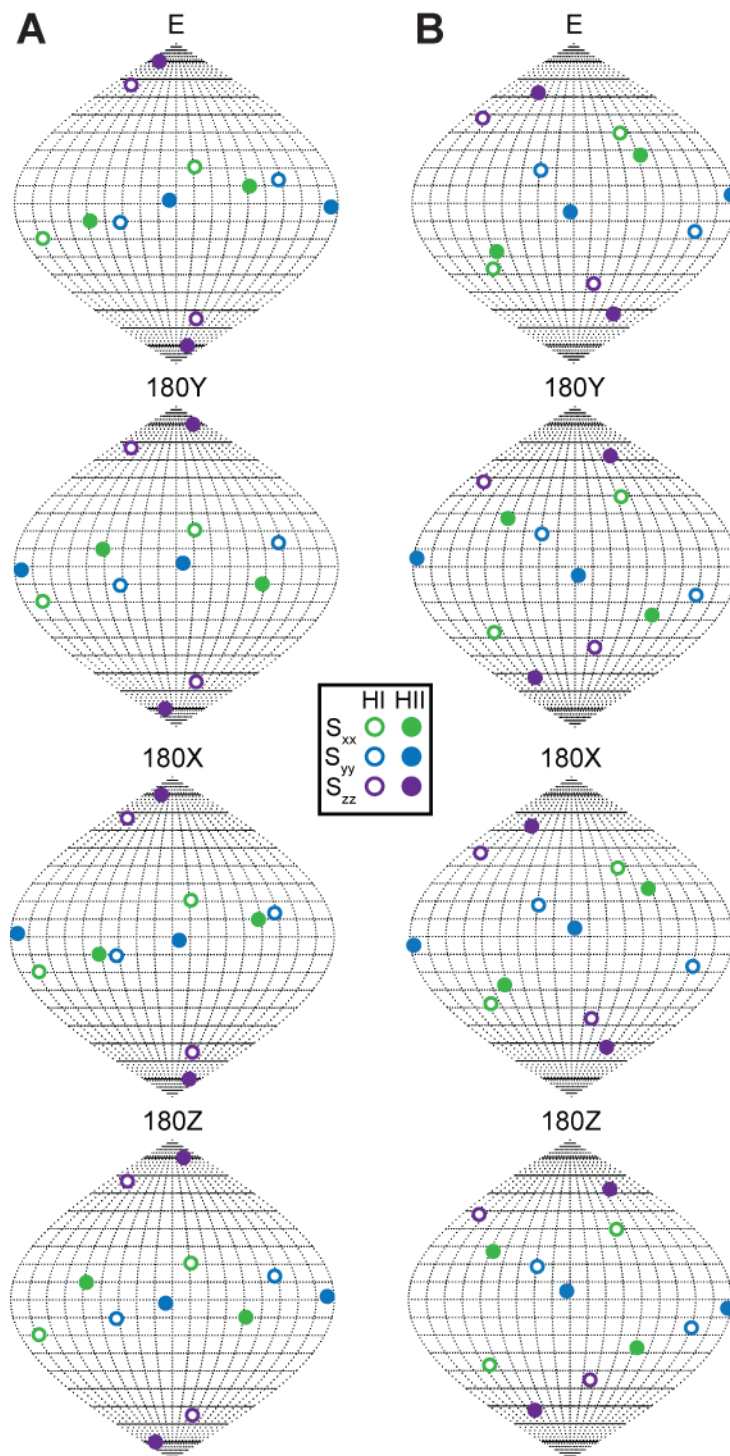
Figure 3.5: Difficulty in resolving degeneracies in determining the average orientation of TAR helices using multiple RDC data sets. Shown are the experimental order tensor frames (S_{xx}, S_{yy}, and S_{zz}, see inset for key) determined for helix I and II (open and filled circles, respectively) using (A) EI(3)-TAR$_m$ and (B) EI(22)-TAR$_m$ RDCs. Solutions are shown relative to the TAR$_m$ PAS. The different orientational solutions were generated by rotating helix II in the TAR$_m$ PAS 180° degrees about the x, y, and z axis.

This work has been published in the Journal of Magnetic Resonance. The idea was conceived by Al-Hashimi, H. M., Qi Zhang, and Dethoff, E.A.. The EI(3)-TAR$_m$ oligonucleotides were synthesized by Dethoff, E.A. Hansen, A.L. recorded and analyzed the phage spectra. All other NMR data was recorded and analyzed by Dethoff, E.A.

3.5 References

[1] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, and J.H. Prestegard, Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information For Structure Determination in Solution. Proceedings of the National Academy of Sciences of the United States of America 92 (1995) 9279-9283.

[2] N. Tjandra, and A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278 (1997) 1111-1114.

[3] J.R. Tolman, and K. Ruan, NMR residual dipolar couplings as probes of biomolecular dynamics. Chem Rev 106 (2006) 1720-36.

[4] G. Bouvignies, P.R. Markwick, and M. Blackledge, Simultaneous definition of high resolution protein structure and backbone conformational dynamics using NMR residual dipolar couplings. Chemphyschem 8 (2007) 1901-9.

[5] M. Getz, X. Sun, A. Casiano-Negroni, Q. Zhang, and H.M. Al-Hashimi, Review NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. Biopolymers 86 (2007) 384-402.

[6] A.A. Bothner-By, Magnetic field induced alignment of molecules. in: D.M. Grant, and R.K. Harris, (Eds.), Encyclopedia of Nuclear Magnetic Resonance, Wiley, Chichester, 1995, pp. 2932-2938.

[7] A. Bax, and A. Grishaev, Weak alignment NMR: a hawk-eyed view of biomolecular structure. Curr Opin Struct Biol 15 (2005) 563-70.

[8] G.M. Clore, M.R. Starich, and A.M. Gronenborn, Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod- shaped viruses. Journal of the American Chemical Society 120 (1998) 10571-10572.

[9] M.R. Hansen, P. Hanson, and A. Pardi, Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. Methods in Enzymology 317 (2000) 220-240.

[10] M. Ruckert, and G. Otting, Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. Journal of the American Chemical Society 122 (2000) 7793-7797.

[11] B.E. Ramirez, and A. Bax, Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. Journal of the American Chemical Society 120 (1998) 9106-9107.

[12] K. Ruan, K.B. Briggman, and J.R. Tolman, De novo determination of internuclear vector orientations from residual dipolar couplings measured in three independent alignment media. J Biomol NMR 41 (2008) 61-76.

[13] W. Peti, J. Meiler, R. Bruschweiler, and C. Griesinger, Model-free analysis of protein backbone motion from residual dipolar couplings. J Am Chem Soc 124 (2002) 5822-33.

[14] G. Bouvignies, P. Markwick, R. Bruschweiler, and M. Blackledge, Simultaneous determination of protein backbone structure and dynamics from residual dipolar couplings. J Am Chem Soc 128 (2006) 15100-1.

[15] G.M. Clore, and C.D. Schwieters, Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. Biochemistry 43 (2004) 10678-91.

[16] O.F. Lange, N.A. Lakomek, C. Fares, G.F. Schroder, K.F. Walter, S. Becker, J. Meiler, H. Grubmuller, C. Griesinger, and B.L. de Groot, Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 320 (2008) 1471-5.

[17] J.R. Tolman, A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. J Am Chem Soc 124 (2002) 12020-30.

[18] J. Meiler, J.J. Prompers, W. Peti, C. Griesinger, and R. Bruschweiler, Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. Journal of the American Chemical Society 123 (2001) 6098-6107.

[19] K.B. Briggman, and J.R. Tolman, De novo determination of bond orientations and order parameters from residual dipolar couplings with high accuracy. J Am Chem Soc 125 (2003) 10164-5.

[20] Q. Zhang, A.C. Stelzer, C.K. Fisher, and H.M. Al-Hashimi, Visualizing spatially correlated dynamics that directs RNA conformational transitions. Nature 450 (2007) 1263-7.

[21] Q. Zhang, R. Throolin, S.W. Pitt, A. Serganov, and H.M. Al-Hashimi, Probing motions between equivalent RNA domains using magnetic field induced residual dipolar couplings: accounting for correlations between motions and alignment. J Am Chem Soc 125 (2003) 10530-1.

[22] H.M. Al-Hashimi, H. Valafar, M. Terrell, E.R. Zartler, M.K. Eidsness, and J.H. Prestegard, Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. Journal of Magnetic Resonance 143 (2000) 402-406.

[23] M. Zweckstetter, and A. Bax, Predicition of sterically induced alignment in a dilute liquid crystalline phase; aid to protein structure determination by NMR. J . Am. Chem. Soc. 122 (2000) 3791-3792.

[24] M. Zweckstetter, G. Hummer, and A. Bax, Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. Biophys J 86 (2004) 3444-60.

[25] B. Wu, M. Petersen, F. Girard, M. Tessari, and S.S. Wijmenga, Prediction of molecular alignment of nucleic acids in aligned media. J. Of Biomol. Nmr 35 (2006) 103-115.

[26] H.M. Al-Hashimi, A. Majumdar, A. Gorin, A. Kettani, E. Skripkin, and D.J. Patel, Field- and phage-induced dipolar couplings in a homodimeric DNA quadruplex, relative orientation of G center dot(C-A) triad and G-tetrad motifs and direct determination of C2 symmetry axis orientation. Journal of the American Chemical Society 123 (2001) 633-640.

[27] A. Vermeulen, S.A. McCallum, and A. Pardi, Comparison of the global structure and dynamics of native and unmodified tRNAval. Biochemistry 44 (2005) 6024-33.

[28] M.P. Latham, P. Hanson, D.J. Brown, and A. Pardi, Comparison of alignment tensors generated for native tRNA(Val) using magnetic fields and liquid crystalline media. J Biomol NMR 40 (2008) 83-94.

[29] Q. Zhang, X. Sun, E.D. Watt, and H.M. Al-Hashimi, Resolving the motional modes that code for RNA adaptation. Science 311 (2006) 653-6.

[30] L. Yao, and A. Bax, Modulating protein alignment in a liquid-crystalline medium through conservative mutagenesis. J Am Chem Soc 129 (2007) 11326-7.

[31] S.W. Pitt, Q. Zhang, D.J. Patel, and H.M. Al-Hashimi, Evidence that electrostatic interactions dictate the ligand-induced arrest of RNA global flexibility. Angew Chem Int Ed Engl 44 (2005) 3412-5.

[32] A. Casiano-Negroni, X. Sun, and H.M. Al-Hashimi, Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. Biochemistry 46 (2007) 6525-35.

[33] Q. Zhang, and H.M. Al-Hashimi, Extending the NMR spatial resolution limit for RNA by motional couplings. Nat Methods 5 (2008) 243-5.

[34] E.A. Dethoff, A.L. Hansen, C. Musselman, E.D. Watt, I. Andricioaei, and H.M. Al-Hashimi, Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. Biophys J 95 (2008) 3906-15.

[35] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, Nmrpipe - a Multidimensional Spectral Processing System Based On Unix Pipes. Journal of Biomolecular Nmr 6 (1995) 277-293.

[36] T.D. Goddard, and D.G. Kneller, SPARKY 3, University of California, San Franscisco.

[37] X. Sun, Q. Zhang, and H.M. Al-Hashimi, Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg2+ binding: role in the kissing-duplex structural transition. Nucleic Acids Res 35 (2007) 1698-713.

[38] A. Meissner, and O.W. Sorensen, The role of coherence transfer efficiency in design of TROSY- type multidimensional NMR experiments. J of Mag. Res. 139 (1999) 439-442.

[39] A.L. Hansen, and H.M. Al-Hashimi, Dynamics of large elongated RNA by NMR carbon relaxation. Journal of the American Chemical Society 129 (2007) 16072-16082.

[40] C. Musselman, S.W. Pitt, K. Gulati, L.L. Foster, I. Andricioaei, and H.M. Al-Hashimi, Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. J Biomol NMR 36 (2006) 235-49.

[41] J.A. Losonczi, M. Andrec, M.W.F. Fischer, and J.H. Prestegard, Order matrix analysis of residual dipolar couplings using singular value decomposition. Journal of Magnetic Resonance 138 (1999) 334-342.

[42] A.L. Hansen, and H.M. Al-Hashimi, Insight into the CSA tensors of nucleobase carbons in RNA polynucleotides from solution measurements of residual CSA: towards new long-range orientational constraints. J Magn Reson 179 (2006) 299-307.

[43] A.L. Hansen, and H.M. Al-Hashimi, Dynamics of large elongated RNA by NMR carbon relaxation. J Am Chem Soc 129 (2007) 16072-82.

[44] M.H. Bailor, C. Musselman, A.L. Hansen, K. Gulati, D.J. Patel, and H.M. Al-Hashimi, Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. Nat Protoc 2 (2007) 1536-46.

[45] M.R. Hansen, L. Mueller, and A. Pardi, Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nature Structural Biology 5 (1998) 1065-1074.

[46] H.M. Al-Hashimi, Y. Gosser, A. Gorin, W. Hu, A. Majumdar, and D.J. Patel, Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. J Mol Biol 315 (2002) 95-102.

[47] E. Ennifar, A. Nikulin, S. Tishchenko, A. Serganov, N. Nevskaya, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov, and P. Dumas, The crystal structure of UUCG tetraloop. Journal of Molecular Biology 304 (2000) 35-42.

[48] H.M. Al-Hashimi, A. Gorin, A. Majumdar, Y. Gosser, and D.J. Patel, Towards structural Genomics of RNA: Rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings. Journal of Molecular Biology 318 (2002) 637-649.

[49] M. Zweckstetter, NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 3 (2008) 679-90.

[50] J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, and J.H. Prestegard, Structural and dynamic analysis of residual dipolar coupling data for proteins. Journal of the American Chemical Society 123 (2001) 1416-1424.

Chapter 4

Quantitative insights into TAR apical loop dynamics by domain elongation, RDC, and spin relaxation NMR

4.1 Introduction

As described in Chapter 1, NMR is a particularly attractive technique for the study of nucleic acid dynamics, for it provides both structural and dynamic information on the atomic scale and can probe dynamics on the picoseconds to seconds timescale. NMR can probe a large range of motions, from the reorientation of entire helices [1-3] to the flipping of a single base or base pair [4,5]. However, the study of RNA dynamics by NMR is complicated by many factors that are relative non-issues in the study of protein dynamics, including carbon-carbon coupling [6], limited chemical shift dispersion, the inability to obtain multiple independent alignments in residual dipolar couplings (RDC) studies despite using a variety of alignment media [7], and violation of the motional coupling approximation [8].

As described in detail in section 1.3.5, the Al-Hashimi group has devised an elongation strategy to overcome the issue of motional coupling. This elongation technique has been used to study the dynamics of multiple RNAs by residual dipolar couplings (RDCs), including the transactivation response element (TAR) of HIV-2, HIV-1 TAR bound to argininamide [9], and a TAR mutant designed to capture a bound state [10]. Spin relaxation parameters under this decoupling scheme have been measured on TAR [6], P4 of RNAse P [11], and DIS of HIV-1 [12]. Varani and coworkers have measured resonance intensities under this elongation strategy to determine the local and interhelical dynamics of an RNA bound to the U1A protein and of TAR on binding various ligands [13,14].

However, the elongation strategy has its own obstacles. First, the technique requires the synthesis of two elongated samples to minimize spectral overcrowding from

the elongation residues [8]. In the two samples, stretches of unlabeled A-U (or G-C) base pairs are used to elongate the lower stem, with ^{15}N/^{13}C-labeled G-C (or A-U) nucleotides. This can be expensive and time-consuming, especially because the yield of elongated RNAs tend to be much lower than their shorter counterparts. The elongation of the sample also causes significant peak broadening and overlap that is not present in the short RNA spectra. Finally, combining RDCs from two separate RNA samples requires careful calibration to account for possible differences in the degree of alignment [8-10]. It is therefore of great interest to explore whether RNA motional decoupling and alignment can be modulated by a smaller degree of elongation which does not necessitate the preparation of two NMR samples.

Figure 4.1: Secondary structures of wtTAR, EII-wtTAR-GC, and EII-wtTAR. EII-wtTAR-GC is GC-labeled only, as indicated by the residues in gray. The other samples are fully isotopically labeled.

In this chapter, we study an RNA construct that allows for the quantitative characterization of the wtTAR apical loop dynamics under the decoupling approximation (Figure 4.1). The apical loop is studied independently from the bulge, thus eliminating the large interhelical motions observed in wtTAR that lead to motional coupling. A shorter elongation is sufficient to align the RNA preferentially along its helical axis, suggesting that conservative elongation is adequate to achieve decoupling when studying small fragments of an RNA, such as a stem-loop structure. This strategy requires the synthesis of only one elongated NMR sample. Our results also provide characterization of the complex dynamics of the wtTAR apical loop on the ps-ms timescale.

4.2 Materials and Methods

4.2.1 Preparation and Purification of NMR Samples

EII-wtTAR-GC and EII-wtTAR samples (Figure 4.1) were prepared by *in vitro* transcription using T7 RNA polymerase (Takara Mirus Bio, Inc.), uniformly $^{13}C/^{15}N$ labeled nucleotide triphosphates (ISOTEC, Inc.), unlabeled nucleotide triphosphates (Sigma), and synthetic DNA templates (Integrated DNA Technologies, Inc.) containing the T7 promoter and sequence of interest. EII-wtTAR-GC contains only $^{15}N/^{13}C$-labeled G-C nucleotides, while EII-wtTAR is fully isotopically labeled. Both RNAs were purified using 15% (w/v) denaturing polyacrylamide gel electrophoresis with 8M urea and TBE. The RNA was eluted from the gel in 20 mM Tris pH 8 buffer followed by ethanol precipitation. The RNA pellet was dissolved in water and exchanged into NMR buffer (15 mM sodium phosphate, 0.1 mM EDTA, and 25 mM NaCl at pH ~6.4) multiple times using an Ultra-4 amicon (Millipore Corp.). All samples contained 10% D_2O. The samples were annealed at 95°C for 10 min before being exchanged into NMR buffer. The final RNA concentration of the EII-wtTAR-GC and EII-wtTAR samples were ~0.6 mM and ~0.7mM, respectively. The aligned NMR samples of EII-wtTAR-GC and EII-wtTAR were prepared by adding phage (Asla Biotech Ltd.) in NMR buffer to yield a final Pf1-phage concentration of ~16 mg/mL and ~17 mg/mL, respectively, and a final RNA concentration of 0.37mM and 0.47mM, respectively. The addition of phage did not affect the average structure of the RNA as judged from careful comparison of the chemical shifts in the absence and presence of phage (data not shown).

4.2.2 Resonance Assignments

All NMR experiments were performed at 298 K on an Avance Bruker 600 MHz spectrometer equipped with a triple-resonance cryogenic (5mm) probe. NMR spectra were analyzed using NMR-Draw [15] and Sparky 3 [16]. The ^1H, ^{13}C, and ^{15}N resonances in EII-wtTAR were assigned using standard homonuclear and heteronuclear 2D experiments [17] on the uniformly $^{13}C/^{15}N$ labeled EII-wtTAR and partially labeled EII-wtTAR-GC sample and by comparison to wtTAR resonance assignments. The exchangeable NOE walk of EII-wtTAR was uninterrupted from G36 to U42 and G14 to U16, indicating a continuous helical conformation in these regions. The NOE walk was

interrupted at U22 due to spectral overlap and became convoluted at the terminal ends due to peak multiplicity and end-fraying effects. The ^{15}N-edited NOESY data was supplemented with 3D HCN experiments [18] for intranucleotide H8/H6-N1/N9-H1' correlations and a 2D HCCH-COSY experiment [19] was used to correlate H2 to H8 in adenosines [20].

4.2.3 Measurement and Order Tensor Analysis of RDCs

2D ^{13}C–^{1}H S^{3}E HSQC experiments [21,22] were used to measure one bond $^{1}D_{C6H6}$, $^{1}D_{C8H8}$, $^{1}D_{C5H5}$, $^{1}D_{C2H2}$, $^{1}D_{C1'H1'}$, and $^{1}D_{N1/3H1/3}$ RDCs in EII-wtTAR-GC and EII-wtTAR using ~16 mg/ml and 17 mg/ml of Pf1-phage [23,24], respectively, as an ordering medium. RDC measurement error was estimated from duplicate experiments that yielded splittings along the ^{1}H dimension or ^{15}N/^{13}C dimension, as previously described [25,26]. The measured RDCs for EII-wtTAR-GC and EII-wtTAR are listed in Table 4.1.

The RDCs measured in Watson-Crick basepairs in EII-wtTAR were used to calculate the order tensor, using an idealized A-form helix as an input coordinate [27]. The helices were constructed using the Biopolymers module in Insight II (Molecular Simulations, Inc.), followed by a correction of the propeller twist angles from +15° to the idealized A-form value of -15° [27,28]. RDCs from the flexible residues G13-C52 were excluded from analysis [27]. The measured RDCs and idealized A-form helix were used to determine a best-fit order tensor using singular value decomposition [29] implemented by the in-house written program RAMAH [30]. Order tensor errors due to "A-form structural noise" and RDC uncertainty were estimated using the program AFORM-RDC [27]. The overall RNA structure was assembled by rotating the helix into the principal axis system (PAS) of the best-fitted order tensor. Statistics for the analysis are summarized in Table 4.2.

Residue (bond vector)	EII-wtTAR-GC RDC (Hz)	EII-wtTAR RDC (Hz)
G13 (C8H8)	34.197	*38.218*
G14 (C8H8)	31.519	36.501
G14 (C1'H1')	-33.365	*OV*

A15 (C2H2)	*NA*	40.104
A27 (C2H2)	*NA*	43.346
G28 (C8H8)	35.479	40.488
C29 (C6H6)	28.157	*OV*
C29 (C1'H1')	-19.935	-23.209
C30 (C5H5)	*31.018*	*33.956*
C30 (C1'H1')	*9.44*	*11.305*
U31 (C6H6)	*NA*	*-8.338*
U31 (C5H5)	*NA*	*7.387*
U31 (C1'H1')	*NA*	*-15.499*
G32 (C8H8)	*2.664*	*5.2*
G32 (C1'H1')	*4.342*	*5.047*
G33 (C8H8)	*24.442*	*28.272*
G33 (C1'H1')	*1.554*	*2.153*
G34 (C8H8)	*32.713*	*35.371*
G34 (C1'H1')	*12.562*	*14.327*
A35 (C8H8)	*NA*	*4.627*
A35 (C1'H1')	*NA*	*-19.445*
C37 (C5H5)	19.074	23.408
U38 (C5H5)	*NA*	27.897
C39 (C5H5)	31.508	32.9
A41 (C2H2)	*NA*	42.337
C51 C6H6	28.276	*OV*
C52 C1'H1'	*0.756*	*0.89*

Table 4.1: RDCs of EII-wtTAR-GC and EII-wtTAR. Values shown in italic were excluded from the EII-wtTAR order tensor analysis. "NA" designates an RDC that could not be measured because the residue was not isotopically labeled. Peaks that are overlapped in EII-wtTAR due to the addition of AU labeling are indicated by "OV".

N	CN	RMSD (Hz)	R	η	ϑ x 10^{-3}	θ
9	5.2	2.6		0.99	0.20 ± 0.07	2.2 ± 0.1	8 ± 3

Table 4.2: Order tensor analysis of RDCs measured in the EII-wtTAR helical stem. Here, θ is the angle of the helical axis relative to the magnetic field.

4.2.4 Measurement and Data Analysis of ^{13}C Relaxation

Longitudinal (R_1) and rotating frame ($R_{1\rho}$) relaxation rates were measured for C2, C5, C6, C8, and C1' carbons using non-TROSY detected experiments [6]. In the $R_{1\rho}$ experiments, high-power off-resonance spinlocks were used to both reduce potential chemical exchange and suppress Hartman-Hahn type transfers to scalar-coupled carbons.

Spinlock powers were calibrated as described previously [31] by A.L. Hansen [6]. Setup and acquisition parameters similar to those reported previously [6] were used and are reported in Table 4.3.

Maximum delays were chosen by recording quick 1D spectra of the 2D relaxation experiment to estimate the decay time at which the signal is reduced by one third relative to T_{min} [6]. A table of delays is listed in Table 4.4. Relaxation rates were determined by fitting peak intensities to a mono-exponential decay using in-house software. Measured $R_{1\rho}$ values had R_1 contributions due to off-resonance effects and so had to be corrected according to $R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta$, where $\theta = \arctan(\omega_{SL}/\Omega)$ is the effective tilt angle of the spinlock field from the static field, ω_{SL} is the amplitude of the spinlock field (in Hz), and Ω is the resonance offset from the spinlock carrier frequency (in Hz).

Pulse	Parameter	C2 C6 C8	C1'	C5
On-resonance inversion	Shape	Q3	Q3	Q3
	Pulse width (ms)	750	2000	1000
On-resonance refocusing	Shape	Q3	Q3	Q3
	Pulse width (ms)	600	1250	700
Off-resonance inversion	Shape	Q3	Q3	Q3
	Pulse width (ms)	700	1500	550
	Offset (ppm)	99	73	153
	Carrier (ppm)	142.5	91	99
	Sweep width (ppm)	24	20	20
	Spinlock offset (Hz)	+1700	+2000	-1750
	Spinlock power (Hz)	3484 ± 27	2245 ± 19	3484 ± 27

Table 4.3: Parameters used in R_1 and $R_{1\rho}$ experiments.

	C2/C6/C8 delays (ms)	C1' delays (ms)	C5 delays (ms)
R_1	20, 130, 290, 550 (x3)	20, 160, 360, 640 (x3)	20, 170, 400, 700 (x3)
$R_{1\rho}$	4,9,20,40 (x3)	4, 14, 33, 60 (x3)	4,10,20,44 (x3)

Table 4.4: Relaxation delays used in R_1 and $R_{1\rho}$ experiments. Triplicate measurements used to estimate error are indicated by (x3).

The R_1 and R_2 relaxation rates listed in Table 4.5 were used to calculate $2R_2$-R_1, which is approximately proportional to $S^2 \cdot J(0)$, where $J(0)$ is the spectral density function at zero field and S^2 is the order parameter describing motions occurring faster than the overall tumbling of the RNA [6,32,33]. Relative S^2 (S^2_{rel}) values were calculated by

normalizing all $2R_2$-R_1 values to that of the largest value measured in the well-structured A-form helix. Each group of bond vector types (i.e. C8H8, C6H6, C1'H1', and C5H5) was normalized separately.

Residue	Carbon	R_1 (Hz)			R_2 (Hz)		
G13	C8	1.488	±	0.006	37.3	±	0.5
	C1'	1.229	±	0.023	49.3	±	1.5
G14	C8	1.479	±	0.008	40.3	±	0.7
A15	C2	1.480	±	0.008	52.4	±	0.8
	C1'	1.038	±	0.011	33.6	±	0.7
A23	C2	1.323	±	0.006	52.5	±	0.6
C25	C6	1.826	±	0.013	59.8	±	1.4
	C5	1.715	±	0.007	54.3	±	0.6
A27	C2	1.399	±	0.007	52.5	±	0.7
	C1'	1.110	±	0.014	33.7	±	0.7
G28	C8	1.384	±	0.007	41.6	±	0.7
C29	C5	1.825	±	0.007	52.7	±	0.5
	C1'	1.153	±	0.011	30.9	±	0.6
C30	C5	1.892	±	0.008	53.5	±	0.6
	C1'	1.413	±	0.019	46.5	±	1.2
U31	C6	2.367	±	0.006	37.4	±	0.5
	C5	2.020	±	0.003	33.6	±	0.3
	C1'	1.726	±	0.017	39.2	±	0.9
G32	C8	1.905	±	0.003	25.1	±	0.3
	C1'	1.670	±	0.007	23.4	±	0.5
G33	C8	1.688	±	0.005	38.4	±	0.5
	C1'	1.503	±	0.007	25.0	±	0.5
G34	C8	1.578	±	0.008	54.6	±	1.0
	C1'	1.501	±	0.014	45.8	±	1.2
A35	C8	1.890	±	0.002	20.5	±	0.2
	C1'	1.891	±	0.023	39.1	±	1.0
G36	C8	1.433	±	0.005	42.2	±	0.7
C37	C6	1.907	±	0.014	56.2	±	1.2
	C5	1.828	±	0.006	52.4	±	0.5
	C1'	1.082	±	0.106	33.8	±	0.6
U38	C5	1.756	±	0.006	53.3	±	0.5
C39	C6	1.794	±	0.013	57.9	±	1.4
	C5	1.797	±	0.006	51.9	±	0.5
A41	C2	1.372	±	0.007	54.3	±	0.7

A49	C2	1.413	±	0.007	52.9	±	0.6
C51	C5	1.838	±	0.009	50.6	±	0.6
	C1'	1.163	±	0.013	31.0	±	0.6
C52	C5	1.851	±	0.009	49.7	±	0.6
	C1'	1.476	±	0.009	24.0	±	0.5

Table 4.5: R_1 and R_2 measured for EII-wtTAR using non-TROSY experiments.

4.3 Results and Discussion

4.3.1 Design and Resonance Assignment of EII-wtTAR.

Our goal is to create an RNA construct that allows for the characterization of the wtTAR apical loop dynamics under the decoupling approximation. To this end, simulations using the program PALES [34] were completed using multiple apical loop structures [35,36] appended on helices of varying lengths. Results showed that elongating domain II by as little as thirteen base pairs was sufficient to align the RNA preferentially along the helical axis of the elongated stem (data not shown). A series of repeated AU base pairs was chosen to elongate the helix because domain II of wtTAR (Figure 4.1) contains only one AU base pair, thus minimizing the number of peaks that could potentially be overlapped by the AU repeat. Two elongated samples were prepared: one in which only Gs and Cs were isotopically labeled (EII-wtTAR-GC) and another which was fully labeled (EII-wtTAR) (Figure 4.1). The GC-labeled sample was prepared solely for comparison purposes to determine how many peaks are lost to spectral overlap due to the labeled AU elongation. Both samples were prepared and assigned as described in Materials and Methods (section 4.2.1). Assignments for both spectra are shown in Figure 4.2.

Isotopically labeling the adenosines (As) and uridines (Us) results in a loss of some peaks due to spectral overlap. For example, in the C6H6/C8H8 spectrum, the Us from the elongation (E-Us) overlap C29 C6H6, and C51 C6H6 is overlapped by the now visible U38 C6H6. In the C1'H1' spectrum, the As from the elongation (E-As) overlap G14 C1'H1' while the E-Us overlap G36 C1'H1'. However, the fully labeled sample provides new peaks as well, such as the C2H2s of As and C5H5s of Us. Most importantly, the fully labeled sample provides peaks for U31 and A35 of the loop. Finally, although the additional C25-G40 base pair was added to serve as an additional
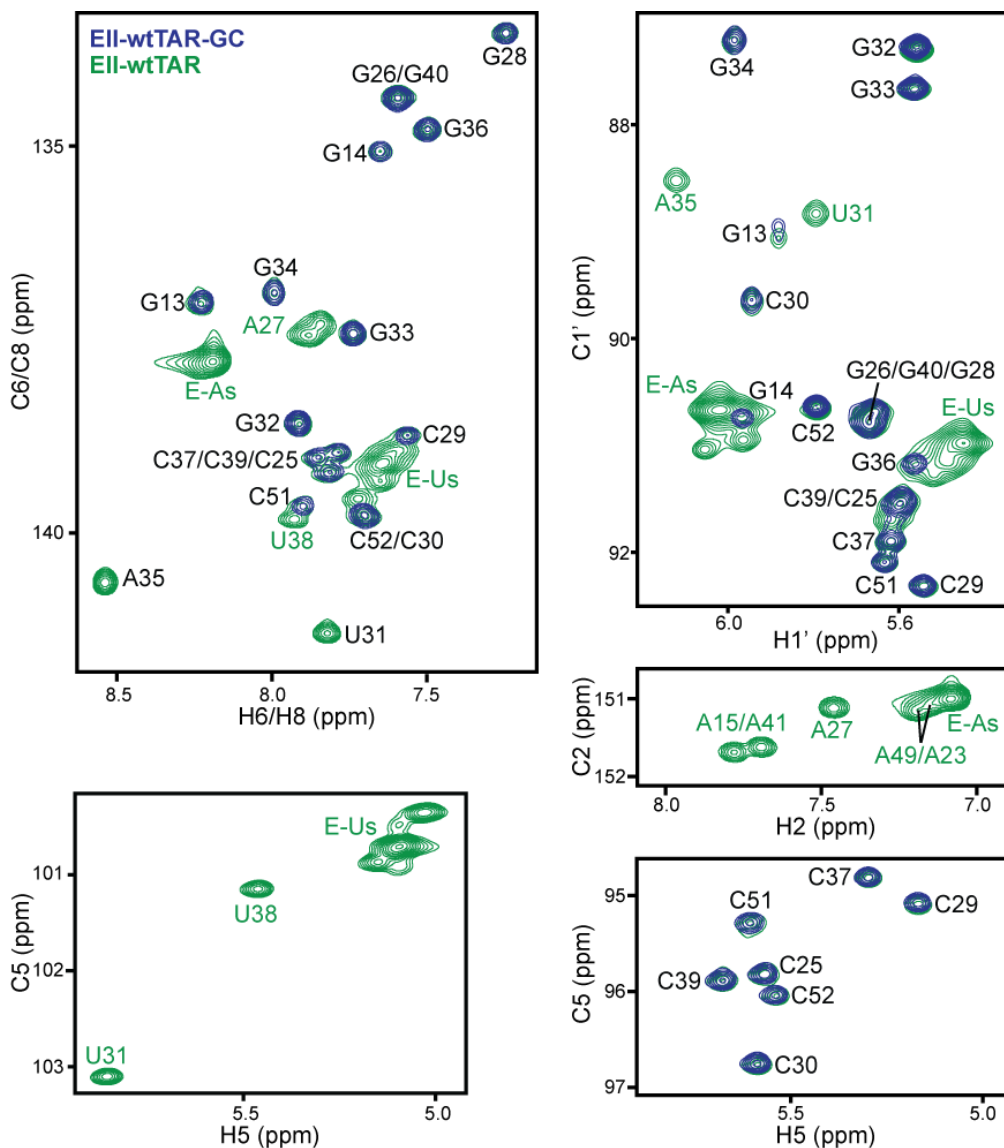
Figure 4.2: Resonance assignments of EII-wtTAR-GC and EII-wtTAR. 2D ^{1}H-^{13}C spectra of EII-wtTAR-GC (dark blue) overlaid on corresponding spectra of EII-wtTAR (green). Green labels denote resonances that only appear in EII-wtTAR due to the addition of isotopically labeled As and Us.

probe in the GC-labeled sample, the peaks of this base pair overlap almost perfectly with the base pair above it, G26-C39. Despite the fact that labeling the As and Us creates more overlap, enough helical peaks remain to complete RDC and spin relaxation analysis. It should be noted that the AU labeling does not cause any overlap of the loop residues, which are of our primary interest.

In Chapter 2 it was shown that mutation of the hexanucleotide loop of wtTAR to a spectroscopically stable UUCG loop does not change the average interhelical angle or
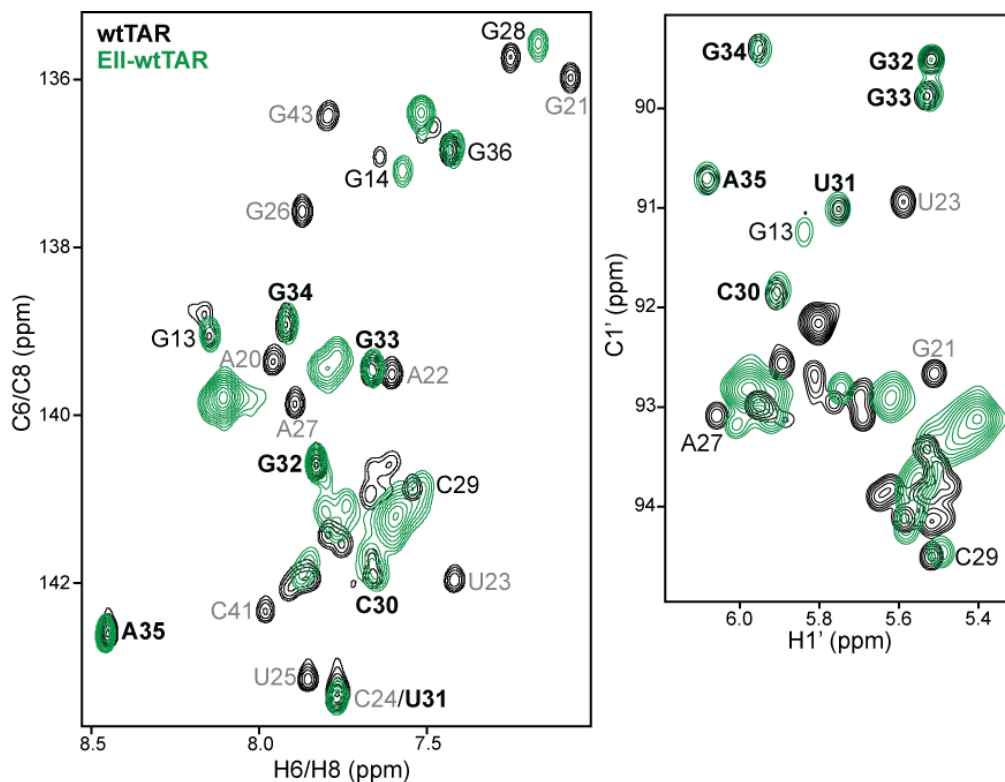
Figure 4.3: Comparison of wtTAR and EII-wtTAR chemical shifts. 2D ^{1}H-^{13}C spectra of EII-wtTAR (green) overlaid on corresponding spectra of wtTAR (black). Loop resonances are in bold. Resonances that only exist in wtTAR are in gray.

interhelical dynamics, suggesting that the bulge and loop are independent dynamical entities. This was further verified by comparing the chemical shifts of wtTAR and EII-wtTAR, in which the latter contains no trinucleotide bulge (Figure 4.3). Examination of the loop peaks of both samples shows that they are in excellent agreement. The only major differences in chemical shifts between the two samples occur near or around the bulge due to its deletion in EII-wtTAR. Deletion of the bulge does not appear to affect the average structure or dynamics of the apical loop. This justifies studying the dynamics of the apical loop in isolation from the bulge.

4.3.2 Order Tensor Analysis of EII-wtTAR RDCs

The orientation and degree of alignment of the EII-wtTAR helix was determined through order tensor analysis [29,37] as reported previously for wtTAR [33] in Chapter 2. However, in this case there is only a single helix to align as the bulge has been deleted. Briefly, an order tensor describing the partial alignment relative to the magnetic field was

determined for the EII-wtTAR elongated helix by fitting measured RDCs of non-terminal Watson-Crick base pairs to an idealized A-form helix geometry [27,37]. An in-house program, AFORM-RDC, was used to estimate errors in the order tensor elements that originate from "structural noise" and uncertainty in the RDC measurements. The excellent fit between measured RDCs and those back-predicted using the best-fit order tensor and idealized A-form helix indicates that the helix adopts a canonical A-form geometry (Figure 4.4A), as previously reported in Chapter 2 for wtTAR [33].

Crucial statistics of the EII-wtTAR order tensor analysis are summarized in Table 4.2. Although only nine RDCs were used to determine the order tensor, the results do not change when single RDCs are deleted one at a time (data not shown). Also, three RDCs that were measured in EII-wtTAR-GC but were overlapped in EII-wtTAR were added to the order tensor analysis for a total of 12 RDCs, and the results remain unchanged (data not shown), further indicating that the order tensor analysis using only nine EII-wtTAR RDCs is valid.

The analysis indicates that the angle of the helical axis of EII-wtTAR relative to the magnetic field (θ) is ~8° from coincidence with the magnetic field. This compares to ~11° as determined for wtTAR, for which domain II dominates alignment. The orientations of domain II of both RNAs are visualized in a Sauson-Flamsteed map in Figure 4.4B. An analogous effect was seen in a previous study by the Al-Hashimi group in which domain II of a TAR mutant with a UUCG tetraloop was elongated by twenty-
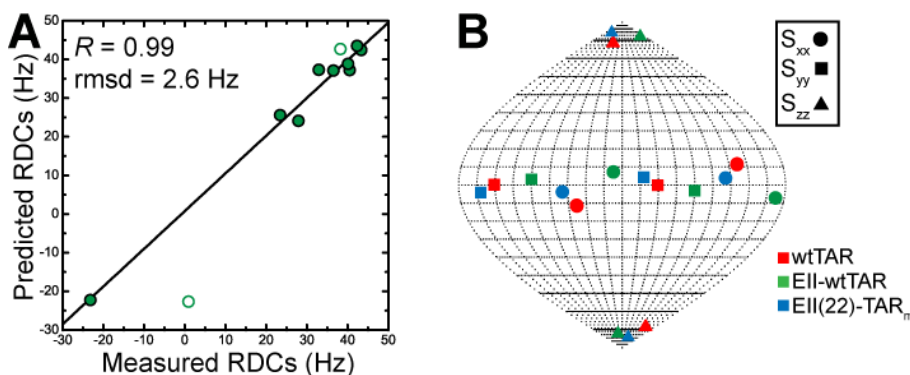
Figure 4.4: Order tensor analysis of EII-wtTAR. (A) Correlation plot between measured and back-calculated EII-wtTAR RDCs when the order tensor of the elongated helix is fit to an idealized A-form geometry. RDCs of terminal end residues that were not used in the analysis are shown in open circles. (B) Sauson-Flamsteed maps showing the best-fit measured order tensor frames (S_{xx}, S_{yy}, S_{zz}) for domain II. Solutions are depicted relative to a molecular frame in which the helix axis is oriented along the z direction.

two base pairs, referred to here as EII(22)-TAR$_m$ [9]. The helical axis of the elongated domain II is oriented ~6° from coincidence with the magnetic field (Figure 4.4B), which is also very similar to the orientation of the non-elongated TAR$_m$ (data not shown). The Sauson-Flamsteed map shows that the elongation in EII-wtTAR orients domain II along the magnetic field to a similar degree as the twenty-two base pair elongation of EII(22)-TAR$_m$. Thus, to a good approximation, the thirteen base pair elongation is sufficient to orient domain II parallel to the magnetic field.

Although θ for wtTAR and EII-wtTAR are similar, the alignment of domain II is modulated through a difference in the twist of the helix of ~40° (data not shown). Thus, this elongation could potentially be used to achieve a second, independent alignment tensor for RDC analysis.

4.3.3 Interpretation of RDCs Values and ^{13}C Relaxation

Measured R$_1$ and R$_2$ relaxation rate constants for C2, C5, C6, C8, and C1' carbons (Table 4.5) were used to calculate pseudo-order parameters (S^2_{rel}) for the C-H bond vectors by normalizing $2R_2$-R_1 values (see section 4.2.4). S^2_{rel} reports on the amplitude of internal motions on the ps-ns timescale and ranges from 0 to 1, where 0 indicates maximum motion and 1 represents a rigid bond vector. However, interpretation of S^2_{rel} values is complicated by the anisotropic tumbling induced by the elongation. For anisotropic rotational diffusion, spin relaxation data is strongly dependent on the orientation of the principal axis of overall rotational diffusion (D$_{zz}$) relative to the relevant dipolar and CSA interaction tensors [6]. For example, a small S^2_{rel} in a non-helical region cannot definitely be attributed to local motions, as attenuation could also arise from a unique static structure.

However, analysis can be partially simplified because we have determined the orientation of the helical axis via RDC order tensor analysis. In general, it can be assumed that the principal axis of alignment (S$_{zz}$) in phage ordering media is identical to D$_{zz}$, because the alignment in both conditions is determined by the global molecular shape of the RNA [6]. Thus, it can be assumed that domain II is aligned approximately parallel to the magnetic field so that base moieties in the helix are oriented perpendicular to the magnetic field. The S^2_{rel} values from the helical region then correspond to

minimum values for a "static" bond vector. This means that if an S^2_{rel} of a base C-H bond observed in the loop is smaller than its bond vector counterpart in the helical region, the attenuation can only be attributed to the presence of dynamics on the ps-ns timescale. Unfortunately, the same simplification cannot be made for the sugar carbons, because the C1'H1' bonds are not oriented perpendicular to the magnetic field under this alignment. The observation of a small S^2_{rel} value in the loop cannot be solely attributed to fast dynamics. However, comparison of both S^2_{rel} and RDCs values will likely reveal any fast motions present in the loop.

Although RDC values are used to determine the static average of a bond vector relative to the magnetic field, they also report on dynamics on the sub-ms timescale. Similar to how an RDC value averages to zero under totally isotropic conditions, the RDC of a bond vector is also attenuated by local dynamics. However, due to the dependence of RDCs on orientation, a small RDC in a non-helical region cannot be attributed solely to local dynamics, because a similar value could arise from a unique static structure. Similar to the relaxation data analysis, interpretation of raw RDC values is also simplified if the orientation of the helix is known. For example, the bases of the helix, to a good approximation, represent minimum values for a "static" bond vector. Thus, any base RDCs in the loop with an RDC smaller than its helical counterpart can be attributed to local dynamics. The same interpretation cannot be extended to the sugar carbons, but incorporation of both RDCs and relaxation data will likely reveal any dynamics in these carbons.

4.3.4 Dynamics of EII-wtTAR on the sub-ms Timescale

As expected, all helical RDCs of base moieties fall between 20-45Hz (Figure 4.5A), and S^2_{rel} values for bonds vectors in the helix are approximately equal to 1 (Figure 4.5B). The only deviations in the helical region fall at the terminal ends, likely due to end fraying effects. In the apical loop, there are multiple resonances with low S^2_{rel} values and near-zero RDCs, which is generally indicative of high internal mobility on the ps-ns timescales. For example, C6H6 and C5H5 of U31, C8H8 of G32, and C8H8 of A35 all have base RDCs between -10 to10 Hz (Figure 4.5A). Likewise, the S^2_{rel} values of these residues are also low: 0.62, 0.61, 0.59, and 0.48, respectively (Figure 4.5B).

Conversely, there are a few resonances in the loop that do not appear to be significantly dynamic on the ps-ns timescale. RDCs of C30 C5H5, G34 C8H8, and, to a lesser extent, G33 C8H8 are all larger than 25Hz, which is similar to the magnitude of the helical RDCs (Figure 4.5A). Correspondingly, the S^2_{rel} values of C30 C5 and G33 C8 are close to 1 (0.97 and 0.92, respectively) (Figure 4.5B), and G34 C8 has an S^2_{rel} greater than 1. An S^2_{rel} greater than 1 could be attributed to either a unique orientation relative to the magnetic field or chemical exchange that is not fully suppressed during $R_{1\rho}$ relaxation experiments. In Chapter 2, results from relaxation dispersion experiments showed that this residue experiences exchange on the microsecond timescale [33].
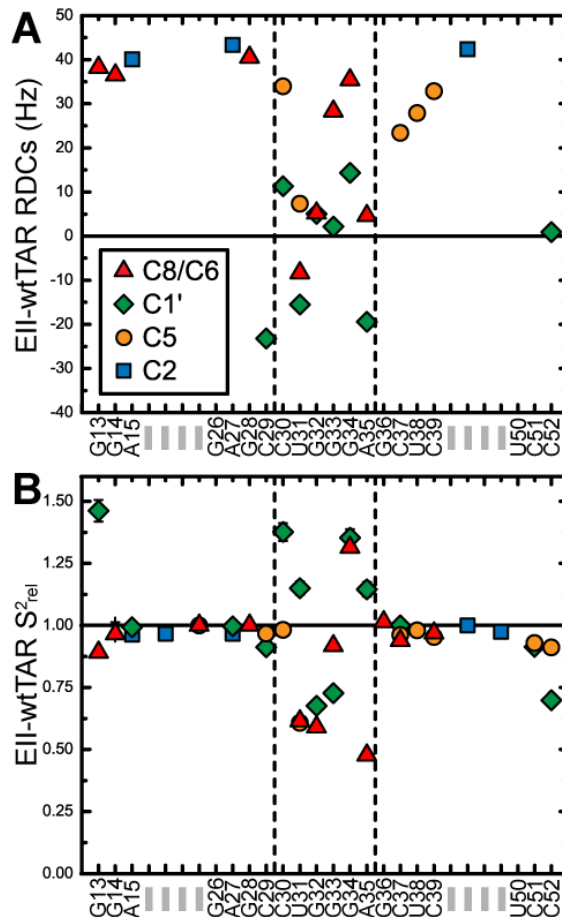
Figure 4.5: Sub-ms dynamics of EII-wtTAR. (A) EII-wtTAR RDCs and (B) normalized $2R_2$-R_1 (S^2_{rel}) values as a function of secondary structure. See inset in for key.

Complications arising from the anisotropy of EII-wtTAR make interpretation of the sugar dynamics difficult. However, it is clear that the C1'H1' of G32 and G33 are likely very dynamic on the ps-ns timescale, because both have small RDCs of less than 10 Hz and have attenuated S^2_{rel} values of 0.68 and 0.73, respectively (Figure 4.5). C30, U31, G34, and A35 sugars all have S^2_{rel} values greater than one, which could be attributed to either a unique orientation or chemical exchange. HSQC spectra of wtTAR and EII-wtTAR recorded at a lower temperature reveal that these peaks are broadened, which is indicative of chemical exchange (data not shown), suggesting the that the high S^2_{rel} values are due at least in part to dynamics on the microsecond to millisecond timescale. The biological importance of these dynamics has already been discussed at length in Chapter 2 and so will only be discussed briefly in this section.

NMR data suggest that the base of U31 and G32 are dynamic on the ps-ns timescale. DMS and kethoxal reactivity data show that U31 and G32 functional groups are exposed to solvent, respectively [38]. Thus, it is likely that these bases are flipped out in solution, as seen in a variety of HIV-1 TAR NMR structures [35,36][39-42]. Interestingly, cross-linking experiments have shown that CycT1 interacts with the U31 side of the apical loop [43], so the elevated degree of mobility at these bases may be important for adaptive binding to the protein.

The other base in the loop with significant fast dynamics on the ps-ns timescale, A35, has been shown to be a critical spacer residue for transactivation [44]. The unique chemical shift of A35 C8H8 and C2H2 and its very small S_{rel}^2 value suggest that A35 is flipped out into solution undergoing large amplitude motions. This agrees with DMS reactivity data that confirms the functional groups of A35 are exposed to solvent [38]. Although its identity is not critical for transcription activation [44], it is strongly conserved across HIV-1 sequences [38]. Thus, A35 may play a role in the life cycle of HIV-1 outside of transactivation of the viral genome.

Our data suggests that C30, G34, and, to a lesser extent, G33 are not dynamic on the ps-ns timescale. This is in agreement with kethoxal and DMS reactivity data that show all three are fairly protected, with G33 being more reactive than G34 [38]. Furthermore, this supports the existence of a C30-G34 cross-strand base pair proposed by both biochemical studies [38] and observed in MD simulations [27,45,46]. This base pair is likely important for transcriptional activation [44].

4.3.5 Comparison of wtTAR and EII-wtTAR Dynamics.

In Figure 4.6 we compare the RDCs and S_{rel}^2 values of domain II and the loop of wtTAR and EII-wtTAR. The RDCs between the two are relatively similar, with the loop RDCs in better agreement as compared to the stem RDCs (Figure 4.6A). This slight difference in the stem RDCs is likely due to the fact that the stems are oriented in slightly different orientations due to a difference in twist about the helix relative to one another. Meanwhile, the magnitude of the loop RDCs are determined primarily by the presence or absence of motions. In other words, dynamic bonds vectors are averaged out to near-zero in both samples, while "rigid" bond vectors have larger RDCs in both samples.
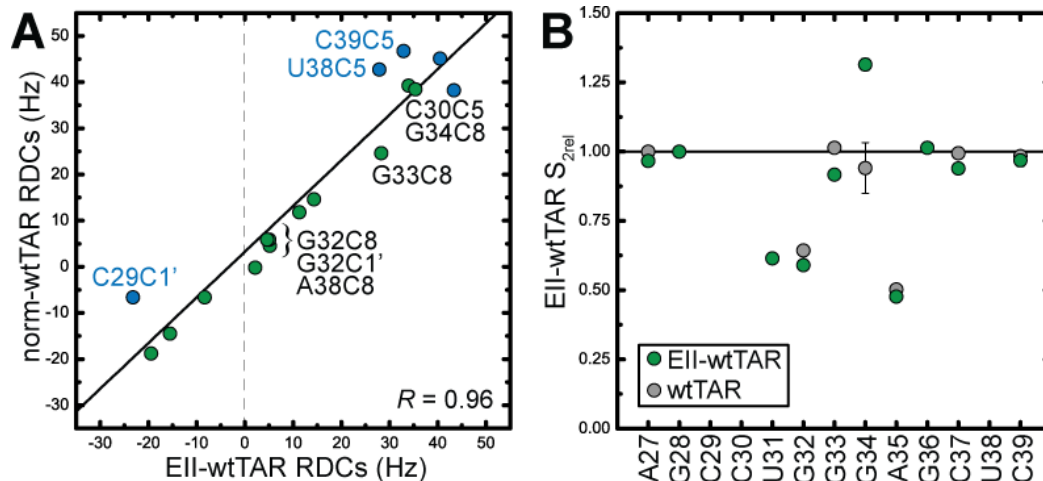
Figure 4.6: Comparison of wtTAR and EII-wtTAR data. (A). Correlation graph of EII-wtTAR and normalized wtTAR RDCs, with helix and loop RDCs in blue and green, respectively. wtTAR RDCs were normalized by multiplying values by $\vartheta_{\text{EII-wtTAR}}/\vartheta_{\text{wtTAR}}$ to compensate for different degrees of alignment. (B) S_{rel}^2 values of aromatic carbons (C2/C6/C8) as a function of residue/secondary structure for wtTAR (black) and EII-wtTAR (green).

The S_{rel}^2 values of the aromatic carbons reported in Figure 4.6B show similar trends in the fast dynamics in the wtTAR and EII-wtTAR apical loops. For example, data for both suggest that the G32 and A35 bases are dynamic, with S_{rel}^2 values of less than 0.65 and 0.51, respectively. However, the S_{rel}^2 values of EII-wtTAR of dynamic bond vectors tend to be slightly more attenuated than their wtTAR counterparts. This likely reflects a greater sensitivity to internal motions due to the EII-wtTAR elongation. The most extreme example is that of G33 C8, which does not appear to experience dynamics on the ps-ns timescale in the wtTAR sample (wtTAR and EII-wtTAR S_{rel}^2 values of ~1 and 0.92, respectively). Finally, wtTAR and EII-wtTAR report different values for G34 C8: 0.95 and 1.3, respectively. As discussed above in section 4.3.4, the S_{rel}^2 value of G34 C8 is likely due to chemical exchange. Also, the wtTAR relaxation experiments used to calculate S_{rel}^2 were measured at a higher spinlock (5 kHz) than those of EII-wtTAR (~3.5 kHz), which suppressed chemical exchange more effectively. Finally, it may also be possible that the apical loop dynamics in wtTAR are coupled to interhelical motions. The resulting coupling of two internal motional modes may lead to different trends in the NMR data.

4.4 Conclusions

In Chapter 2, we showed that the wtTAR apical loop and bulge act as independent dynamical centers with no long-range communication. This was further corroborated by the comparison of wtTAR and EII-wtTAR, the latter of which lacks the trinucleotide bulge. The chemical shifts of the loop resonances for the two RNAs are in excellent agreement (Figure 4.3), which suggests that there is no major long-range interaction between the loop and bulge or lower stem as suggested from sequence covariance studies [47]. However, our results do not rule out the possibility of a transient interaction between the loop and another part of the wtTAR RNA.

The results in Chapter 2 also showed that the apical loop undergoes complex dynamics at multiple timescales that are likely important for adaptive recognition of Tat and P-TEFb. However, these results were relatively qualitative, because all RDC and relaxation measurements were made on a system in which global and local motions were not motionally decoupled. Studying an elongated wtTAR sample allows us to make more quantitative measurements of the wtTAR apical loop dynamics as outlined in this chapter.

We show that the apical loop is structurally heterogeneous, with multiple residues undergoing distinct dynamics over a range of timescales. For example, the base of U31, G32, and A35 all undergo fast dynamics on the ps-ns timescale. Examination of reactivity data [38] and NMR structures of the wtTAR loop [35,36,39-42] suggest that all three of these bases are looped out into solution, and this elevated mobility in U31 and G32 may play a role in the binding of CycT1. Our data also supports the existence of a C30-G34 cross-strand base pair proposed to be important for transcriptional activation [44].

Future studies of dynamic regions of RNA, such as bulges, junctions, and apical loops, will likely use numerous elongation schemes in an attempt to achieve five independent alignments under the decoupling approximation. This chapter outlines a general scheme for the design of one particular type of elongation and our results demonstrate the advantage of using an elongation to study the dynamics of an apical loop when isolated from other dynamical centers.

Most importantly, the removal of the bulge abolishes the motional coupling between interhelical and local dynamics that normally complicate NMR data interpretation. Secondly, the elongation is sufficient to anchor the principal axis of alignment in phage ordering media along the helical axis, which simplifies NMR data analysis and the computation of NMR parameters from MD trajectories [48,49]. Thirdly, the use of a elongation instead of the full twenty-two base pairs used previously [8,9] obviates the need for preparation of two elongated samples and the combination of data from two separate RNA constructs. Finally, this simple elongation may serve as a second independent alignment for characterizing RNA structural dynamics.

The idea was conceived by Al-Hashimi, H. M. and Dethoff, E.A.. The EII-wtTAR-GC and EII-wtTAR RNA oligonucleotides were synthesized by Dethoff, E.A. and Gulati, N., respectively. All EII-wtTAR-GC and EII-wtTAR NMR data were recorded and analyzed by Dethoff, E.A.

4.5 References

[1] D.W. Staple, and S.E. Butcher, Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. J Mol Biol 349 (2005) 1011-23.
[2] R.J. Richards, H. Wu, L. Trantirek, C.M. O'Connor, K. Collins, and J. Feigon, Structural study of elements of Tetrahymena telomerase RNA stem-loop IV domain important for function. Rna 12 (2006) 1475-85.
[3] H.M. Al-Hashimi, Y. Gosser, A. Gorin, W. Hu, A. Majumdar, and D.J. Patel, Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. J Mol Biol 315 (2002) 95-102.
[4] H. Blad, N.J. Reiter, F. Abildgaard, J.L. Markley, and S.E. Butcher, Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. J Mol Biol 353 (2005) 540-555.
[5] H.M. Al-Hashimi, E.N. Nikolova, E. Kim, A.A. Wise, P.J. O'Brien, and I. Andricioaei, Transient Hoogsteen base pairs in canonical duplex DNA. Nature 470 (2011) 498-U84.
[6] A.L. Hansen, and H.M. Al-Hashimi, Dynamics of large elongated RNA by NMR carbon relaxation. J Am Chem Soc 129 (2007) 16072-82.
[7] M.P. Latham, P. Hanson, D.J. Brown, and A. Pardi, Comparison of alignment tensors generated for native tRNA(Val) using magnetic fields and liquid crystalline media. J Biomol NMR 40 (2008) 83-94.
[8] Q. Zhang, X. Sun, E.D. Watt, and H.M. Al-Hashimi, Resolving the motional modes that code for RNA adaptation. Science 311 (2006) 653-6.
[9] Q. Zhang, A.C. Stelzer, C.K. Fisher, and H.M. Al-Hashimi, Visualizing spatially correlated dynamics that directs RNA conformational transitions. Nature 450 (2007) 1263-7.
[10] A.C. Stelzer, J.D. Kratz, Q. Zhang, and H.M. Al-Hashimi, RNA Dynamics by Design: Biasing Ensembles Towards the Ligand-Bound State. Angewandte Chemie-International Edition 49 (2010) 5731-5733.
[11] M.M. Getz, A.J. Andrews, C.A. Fierke, and H.M. Al-Hashimi, Structural plasticity and Mg2+ binding properties of RNase P P4 from combined analysis of NMR residual dipolar couplings and motionally decoupled spin relaxation. RNA (2006).

[12] X. Sun, Q. Zhang, and H.M. Al-Hashimi, Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg2+ binding: role in the kissing-duplex structural transition. Nucleic Acids Res 35 (2007) 1698-713.

[13] M.F. Bardaro, Z. Shajani, K. Patora-Komisarska, J.A. Robinson, and G. Varani, How binding of small molecule and peptide ligands to HIV-1 TAR alters the RNA motional landscape. Nucleic Acids Research 37 (2009) 1529-1540.

[14] Z. Shajani, G. Drobny, and G. Varani, Binding of U1A Protein Changes RNA Dynamics As Observed by (13)C NMR Relaxation Studies(,). Biochemistry 46 (2007) 5875-83.

[15] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, Nmrpipe - a Multidimensional Spectral Processing System Based On Unix Pipes. Journal of Biomolecular Nmr 6 (1995) 277-293.

[16] T.D. Goddard, and D.G. Kneller, SPARKY 3, University of California, San Franscisco.

[17] B. Furtig, C. Richter, J. Wohnert, and H. Schwalbe, NMR spectroscopy of RNA. Chembiochem 4 (2003) 936-62.

[18] V. Sklenar, R.D. Peterson, M.R. Rejante, and J. Feigon, Two- and three-dimensional HCN experiments for correlating base and sugar resonances in 15N,13C-labeled RNA oligonucleotides. J Biomol NMR 3 (1993) 721-7.

[19] J. Wohnert, M. Gorlach, and H. Schwalbe, Triple resonance experiments for the simultaneous correlation of H6/H5 and exchangeable protons of pyrimidine nucleotides in 13C,15N-labeled RNA applicable to larger RNA molecules. J Biomol NMR 26 (2003) 79-83.

[20] B. Simon, K. Zanier, and M. Sattler, A TROSY relayed HCCH-COSY experiment for correlating adenine H2/H8 resonances in uniformly C-13-labeled RNA molecules. Journal of Biomolecular Nmr 20 (2001) 173-176.

[21] A. Meissner, and O.W. Sorensen, The role of coherence transfer efficiency in design of TROSY- type multidimensional NMR experiments. Journal of Magnetic Resonance 139 (1999) 439-442.

[22] S.W. Pitt, Q. Zhang, D.J. Patel, and H.M. Al-Hashimi, Evidence that electrostatic interactions dictate the ligand-induced arrest of RNA global flexibility. Angew Chem Int Ed Engl 44 (2005) 3412-5.

[23] M.R. Hansen, P. Hanson, and A. Pardi, Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. Methods in Enzymology 317 (2000) 220-240.

[24] G.M. Clore, M.R. Starich, and A.M. Gronenborn, Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod- shaped viruses. Journal of the American Chemical Society 120 (1998) 10571-10572.

[25] M. Getz, X. Sun, A. Casiano-Negroni, Q. Zhang, and H.M. Al-Hashimi, Review NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. Biopolymers 86 (2007) 384-402.

[26] A. Casiano-Negroni, X. Sun, and H.M. Al-Hashimi, Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. Biochemistry 46 (2007) 6525-35.

[27] C. Musselman, S.W. Pitt, K. Gulati, L.L. Foster, I. Andricioaei, and H.M. Al-Hashimi, Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. J Biomol NMR 36 (2006) 235-49.

[28] S. Neidle, Oxford Handbook of Nucleic Acid Structure, Oxford University Press, New York, 1999.

[29] J.A. Losonczi, M. Andrec, M.W.F. Fischer, and J.H. Prestegard, Order matrix analysis of residual dipolar couplings using singular value decomposition. Journal of Magnetic Resonance 138 (1999) 334-342.

[30] A.L. Hansen, and H.M. Al-Hashimi, Insight into the CSA tensors of nucleobase carbons in RNA polynucleotides from solution measurements of residual CSA: towards new long-range orientational constraints. J Magn Reson 179 (2006) 299-307.

[31] A.G. Palmer, 3rd, C.D. Kroenke, and J.P. Loria, Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. Methods Enzymol 339 (2001) 204-38.

[32] D. Fushman, N. Tjandra, and D. Cowburn, An approach to direct determination of protein dynamics from N-15 NMR relaxation at multiple fields, independent of variable N-15 chemical shift anisotropy and chemical exchange contributions. Journal of the American Chemical Society 121 (1999) 8577-8582.

[33] E.A. Dethoff, A.L. Hansen, C. Musselman, E.D. Watt, I. Andricioaei, and H.M. Al-Hashimi, Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. Biophys J 95 (2008) 3906-15.

[34] M. Zweckstetter, and A. Bax, Prediction of sterically induced alignment in a dilute liquid crystalline phase; aid to protein structure determination by NMR. J . Am. Chem. Soc. 122 (2000) 3791-3792.

[35] F. Aboul-ela, J. Karn, and G. Varani, The Structure of the Human-Immunodeficiency-Virus Type-1 Tar RNA Reveals Principles of RNA Recognition By Tat Protein. Journal of Molecular Biology 253 (1995) 313-332.

[36] F. Aboul-ela, J. Karn, and G. Varani, Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. Nucleic Acids Research 24 (1996) 3974-3981.

[37] M.H. Bailor, C. Musselman, A.L. Hansen, K. Gulati, D.J. Patel, and H.M. Al-Hashimi, Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. Nat Protoc 2 (2007) 1536-46.

[38] T. Kulinski, M. Olejniczak, H. Huthoff, L. Bielecki, K. Pachulska-Wieczorek, A.T. Das, B. Berkhout, and R.W. Adamiak, The apical loop of the HIV-1 TAR RNA hairpin is stabilized by a cross-loop base pair. J Biol Chem 278 (2003) 38892-901.

[39] C. Faber, H. Sticht, K. Schweimer, and P. Rosch, Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. J Biol Chem 275 (2000) 20660-6.

[40] Z. Du, K.E. Lind, and T.L. James, Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. Chem Biol 9 (2002) 707-12.

[41] B. Davis, M. Afshar, G. Varani, A.I. Murchie, J. Karn, G. Lentzen, M. Drysdale, J. Bower, A.J. Potter, I.D. Starkey, T. Swarbrick, and F. Aboul-ela, Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". J Mol Biol 336 (2004) 343-56.

[42] A.I. Murchie, B. Davis, C. Isel, M. Afshar, M.J. Drysdale, J. Bower, A.J. Potter, I.D. Starkey, T.M. Swarbrick, S. Mirza, C.D. Prescott, P. Vaglio, F. Aboul-ela, and J. Karn, Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. J Mol Biol 336 (2004) 625-38.

[43] S. Richter, Y.H. Ping, and T.M. Rana, TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. Proc Natl Acad Sci U S A 99 (2002) 7928-33.

[44] S. Richter, H. Cao, and T.M. Rana, Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. Biochemistry 41 (2002) 6391-7.

[45] A.D. Critchley, I. Haneef, D.J. Cousens, and P.G. Stockley, Modeling and solution structure probing of the HIV-1 TAR stem-loop. J Mol Graph 11 (1993) 92-7, 124.

[46] R. Nifosi, C.M. Reyes, and P.A. Kollman, Molecular Dynamics Studies of the HIV-1 TAR and its Complex with Argininamide. Nucleic Acids Research (2000) 4944-4955.

[47] H. Huthoff, F. Girard, S.S. Wijmenga, and B. Berkhout, Evidence for a base triple in the free HIV-1 TAR RNA. Rna 10 (2004) 412-23.

[48] C. Musselman, H.M. Al-Hashimi, and I. Andricioaei, iRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. Biophys J 93 (2007) 411-22.

[49] A.T. Frank, A.C. Stelzer, H.M. Al-Hashimi, and I. Andricioaei, Constructing RNA Dynamical Ensembles by Combining MD and Motionally Decoupled NMR RDCs: New insights into RNA Dynamics and Adaptive Ligand Recognition. Nucleic Acids Res 37 (2009) 3670-9.

Chapter 5

Characterizing the excited state of the wtTAR apical loop by NMR and mutagenesis

5.1 Introduction

Spin relaxation experiments are useful when studying the dynamics of a biomolecule on the ps-ns timescale. However, processes such as base pair opening [1], repuckering of the sugar moiety [2], and base flipping [3] commonly occur on slower timescales (μs-ms) and will likely go undetected. As discussed in section 1.3.4, motional modes on this timescale can manifest as chemical exchange and are commonly studied using CPMG and $R_{1\rho}$ dispersion experiments. A recent ^{13}C $R_{1\rho}$ dispersion experiment developed by Al-Hashimi and coworkers extends the measureable timescale of traditional $R_{1\rho}$ experiments to milliseconds by using selective Hartman-Hahn polarization transfers to excite specific spins of interest [4]. This experiment can be used to determine not only the rate of exchange but the population and chemical shift of the lowly-populated, "excited" state.

In the last chapter, ^{13}C relaxation parameters of the wtTAR apical loop suggested that multiple carbons may experience chemical exchange between two or more chemically distinct states. Specifically, G34 C8 and the C1' sugars of C30, U31, G34, and A35 likely experience exchange on the μs-ms timescale (Figure 5.1). In this chapter, we use a combination of $R_{1\rho}$ relaxation dispersion and mutagenesis to characterize the ground and excited state of the wtTAR apical loop. We find that all five of the residues listed above share a common two-state process occurring on the microsecond timescale. Ground state (GS) and excited state (ES) mimics
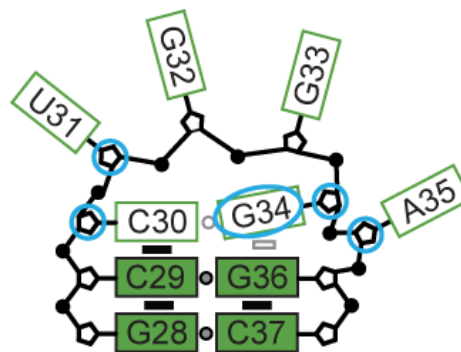

Figure 5.1: Apical loop of wtTAR. Possible sites of chemical exchange as identified in Chapter 4 are circled in blue.

suggest that A35 is looped out in solution in the ground state and that the excited state of the apical loop likely involves the flipping in of A35 to stack on G36 and possibly base pair with C30. Additionally, A35 C8, G33 C8, and G33 C1' experience exchange on the millisecond timescale, but the nature of this process is not yet clear.

5.2 Materials and Methods

5.2.1 Preparation and Purification of RNA Samples

wtTAR was prepared as described in Chapter 2. The RNA pellet was exchanged into NMR buffer (15mM sodium phosphate, 0.1mM EDTA, and 25mM NaCl at pH ~6.4) multiple times using an Ultra-4 amicon (Millipore Corp.). The final RNA concentration was ~0.7 mM. Unlabeled wtTAR-U30 and wtTAR-DMA35 were ordered from Dharmacon (Thermo Fisher Scientific), and wtTAR-G35 was ordered from Integrated DNA Technologies, Inc. All three samples were dissolved in water and annealed at 95°C for 10 min before being exchanged into NMR buffer. The final RNA concentrations were ~1.2, 1.5, and 1.7mM.

5.2.2 Resonance Assignments

All NMR experiments were performed at 298K on an Avance Bruker 600 MHz spectrometer equipped with a triple-resonance cryogenic (5mm) probe. NMR spectra were analyzed using NMR-Draw [5] and Sparky 3 [6]. In addition to the wtTAR assignments in Chapter 2, an HCCH-COSY was used to correlate H2 to H8 in adenosines [7] and to obtain the chemical shifts of C4, C5 and C6 of adenosines. A COSY was used to assign the C2' carbons [8]. The three excited-state mimics were assigned as described previously [9].

5.2.3 Measurement of Relaxation Dispersion

The relaxation parameter $R_{1\rho}$ was measured using a recently developed 1D selective ^{13}C $R_{1\rho}$ relaxation dispersion experiment [4]. In this experiment, selective Hartman-Hahn (HH) polarization transfers are used to excite specific carbons of interest to avoid carbon-carbon scalar coupling. Before $R_{1\rho}$ measurements are recorded, care must be taken to avoid HH matching conditions between scalar-coupled carbons using

equation 1.20 in Chapter 1 [10]. The following coupling constants were used in the calculations: $J_{C8-C5} = 1Hz$, $J_{C8-C4/6} = 10Hz$, $J_{C1'-C2'} = 40Hz$, $J_{C1'-C8} = 2Hz$ [11]. To calculate the effective spinlock of a particular spin (see eq 1.17), the chemical shifts of all relevant carbons must first be determined, a nontrivial task for carbons in the adenine and guanine bases. The chemical shift of these carbons were determined using a TROSY-relayed HCCH-COSY experiment [7]. The chemical shifts of C2s and C5s were used to calculate HH transfers for C1's and C6s, respectively. Spin lock and offset conditions that resulted in an A_{HAHA} transfer efficiency of greater than 1% were avoided. Generally, spin lock powers greater than 2500Hz could not be used for C1' measurements due to significant HH transfer complications.

Next, a series of "on-resonance" experiments, with the offset carrier set to the frequency of the peak of interest, were run at the following powers: 200(x2), 300, 400, 500, 600, 800(x2), 1000, 1500, 2000, 2500, 3000, and 3500(x2) Hz, where x2 indicates repeated measurements used for error assessment. These powers were calibrated as described previously [12]. A number of delay points were measured to determine the $R_{1\rho}$ relaxation rate at each power. Similar to the R_1 and $R_{1\rho}$ relaxation experiments described in Chapter 4, maximum delays were chosen by recording quick 1D spectra to estimate the decay time at which the signal is reduced by one third relative to T_{min} [12]. For a particular carbon, the same delay points were used for on- and off-resonance measurements. Each set included four to seven points, with at least two repeats for estimation of error. A table of delays is listed in Appendix 2 (Table A2.1). The number of scans was chosen so that the signal-to-noise of the longest delay was at least 40. Relaxation rates were determined by fitting peak intensities to a mono-exponential decay using the program Mathematica [13]. See Figure A2.1 for representative monoexponential decays. Off-resonance powers and offsets were chosen according to the on-resonance dispersion profile. See Table A2.2 for representative off-resonance offsets and powers.

5.2.4 Data Analysis of Relaxation Dispersion Data

Initially, on-resonance dispersion data was measured and fit to a simplified version of the fast exchange equation (described in section 1.3.4).

$$R_{1\rho} = R_2 + \frac{\phi_{ex}k_{ex}}{k_{ex}^2 + \omega_1^2} \ , \ \phi_{ex} = p_a p_b \Delta\omega^2 \qquad\qquad (5.1)$$

Here, ω_1 is the spin lock power, k_{ex} is the sum of the forward and reverse rate constants, $\Delta\omega$ is the chemical shift difference between the two states, and p_a and p_b are the populations of state a and b, respectively. In these experiments there is no contribution from R_1, so the measured rates are simply a combination of R_2 and R_{ex}. Data was fit to both a straight line (representing no dispersion) and to equation 5.1, where R_2, ϕ, and kex were allowed to float. The best model (fast chemical exchange or no exchange) was determined from F-statistics at the 95% confidence level. On-resonance profiles are shown in Figure 5.2. To obtain a more detailed characterization of the exchange process, off-resonance measurements were run for any carbon that displayed relaxation dispersion in the on-resonance profile (Figure 5.2).

For each particular carbon of interest, all on-resonance and off-resonance data were fit to the Fast (eq. 1.17) and Laguerre (eq. 1.19) equations described in section 1.3.4, and R_1, R_2, p_b, k_{ex}, and $\Delta\omega$ (or ϕ) were allowed to vary [14]. Carbons that had similar exchange rates were also fit globally using both equations. In a global fit, p_b and k_{ex} were shared among all the residues but allowed to float, while all other parameters were allowed to vary individually. The best-fit model was determined using F-statistics at the 95% confidence level. The fitted parameters and F-test statistics are listed in Table 5.1 and 5.2. Off-resonance $R_{1\rho}$ values had R_1 contributions due to off-resonance effects and had to be corrected according to equation 1.14 to calculate $R_2 + R_{ex}$.

	N	F-stat	p-value	R_2 (Hz)	τ_{ex} (ms)	p_b (%)	$\Delta\omega$ (ppm)
G33 C1'	126	245	0.0	23.2 ± 0.1	--	--	--
				22.16 ± 0.08	2.1 ± 0.6	0.32 ± 0.07	2.58 ± 0.06
G33 C8	132	245	0.0	34.8 ± 0.2	--	--	--
				33.83 ± 0.09	1.9 ± 0.4	0.54 ± 0.09	2.20 ± 0.07
A35 C8	127	532	0.0	19.6 ± 0.1	--	--	--
				18.58 ± 0.05	2.2 ± 0.5	0.40 ± 0.08	-2.48 ± 0.05

Table 5.1: Chemical exchange parameters and statistical analysis of millisecond exchange. Individual carbons were fit to a "no exchange" and the Laguerre equation. N is the number of data points.

	N	F-stat	p-value	R_2 (Hz)	τ_{ex} (us)	p_b (%)	$\Delta\omega$ (ppm)	ϕ (x10^3)
Individual								
C30 C1'	156	6.3	1.3E-02	32 ± 2	38 ± 3	--	--	497 ± 71
				32 ± 2	37 ± 2	22 ± 8	1.8 ± 0.2	--
U31 C1'	153	17.2	5.8E-05	29 ± 1	38 ± 3	--	--	444 ± 64
				29 ± 1	40 ± 3	9 ± 3	2.4 ± 0.4	--
G34 C1'	142	13.6	3.2E-04	35 ± 1	45 ± 2	--	--	409 ± 43
				35 ± 1	45 ± 2	14 ± 5	1.9 ± 0.3	--
A35 C1'	98	29.8	4.0E-07	18 ± 3	33 ± 2	--	--	934 ± 140
				20 ± 2	37 ± 2	11 ± 2	3.0 ± 0.3	--
G34 C8	158	26.2	9.2E-07	38 ± 2	38 ± 3	--	--	666 ± 98
				38 ± 2	38 ± 2	11 ± 3	2.7 ± 0.3	--
U31 C6	*116*	*0.04*	*0.85*	*39 ± 1*	*43 ± 9*	*--*	*--*	*145 ± 57*
				39 ± 1	*43 ± 9*	*30 ± 70*	*-0.9 ± 0.6*	*--*
Global (shared p_b, k_{ex})								
C30 C1'				32.3 ± 0.7			2.1 ± 0.1	
U31 C1'				28.7 ± 0.7			2.1 ± 0.1	
G34 C1'	707	1.47	0.166	31.4 ± 0.8	39 ± 1	13 ± 2	2.3 ± 0.1	
A35 C1'				22 ± 1			2.6 ± 0.2	
G34 C8				38.5 ± 0.9			2.5 ± 0.1	
U31 C6				*38.5 ± 0.3*			*~1.3*	

Table 5.2: Chemical exchange parameters and statistical analysis of microsecond exchange. Individual carbons were fit to the Fast and Laguerre equations. U31 C6 data were measured from the EII-wtTAR construct (see Chapter 4), as indicated by italics. All data (except U31 C6) were fit globally to the Laguerre equation. Global fit parameters (p_b and k_{ex}) were used to calculate $\Delta\omega$ of U31 C6.
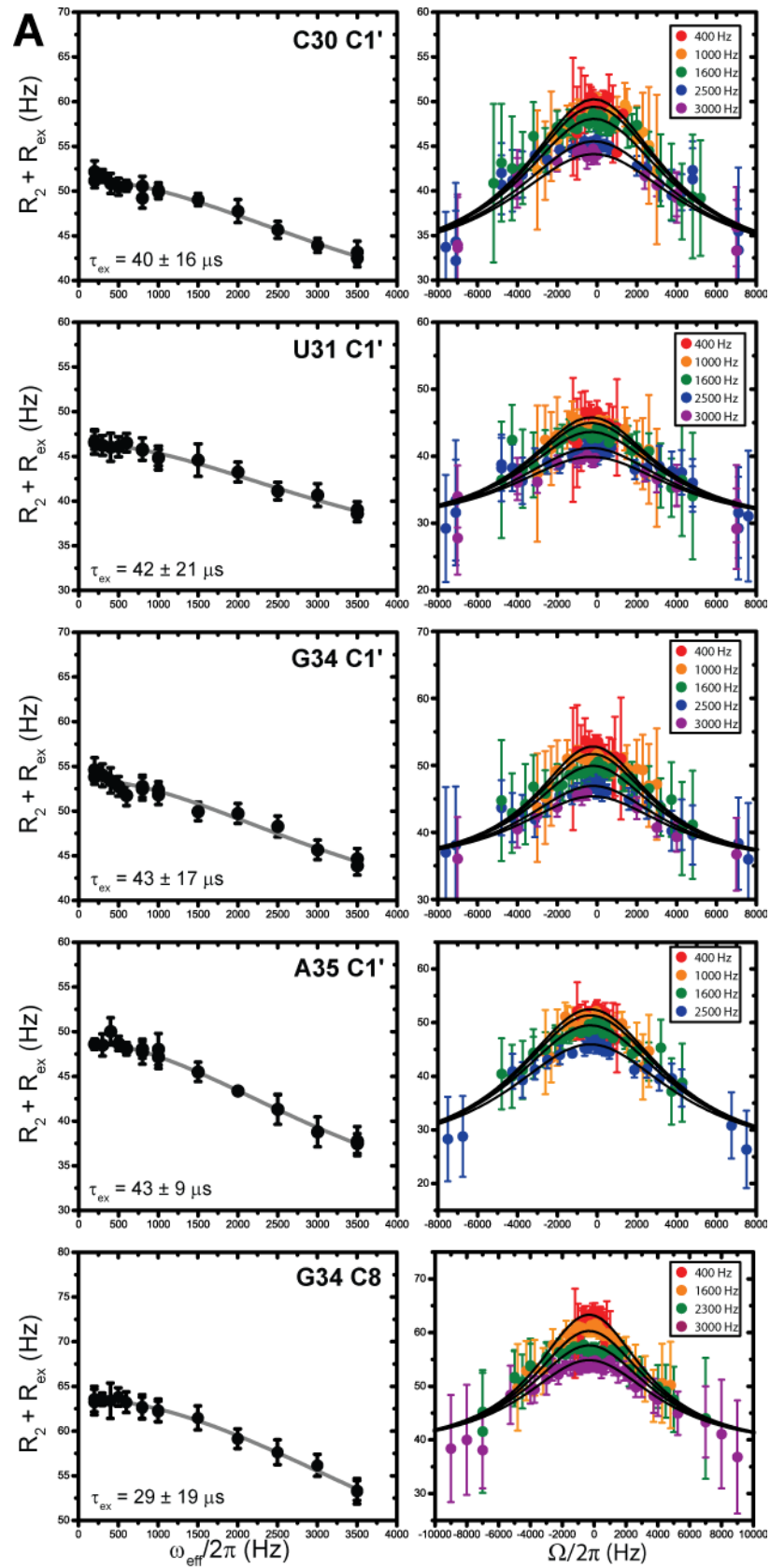
5.3 Results and Discussion

5.3.1 Dynamics of the wtTAR Apical Loop on the μs-ms Timescale

A previous ^{13}C relaxation study of the dynamics of the wtTAR apical loop (described in Chapter 4) suggests that G34 C8 and the C1' carbons of C30, U31, G34, and A35 experience chemical exchange between two or more chemically distinct states on the μs-ms timescale. This is further supported by exchange broadening seen in many of the loop residues in spectra run at low temperatures (data not shown). Even at 25°C, the C2H2 peak of A35 is exchange broadened to the point that it is barely visible above the noise. To characterize these potential slow motions, we measured $R_{1\rho}$ relaxation dispersion on all resolved C6/C8 and C1' peaks of the apical loop.

On-resonance dispersion profiles reveal that indeed all five of these carbons (G34 C8 and C1' of C30, U31, G34, A35) experience chemical exchange on the microsecond timescale (Figure 5.2A). In addition, on-resonance profiles of A35 C8, G33 C8, and G33 C1' indicate these residues likely experience chemical exchange on the millisecond timescale (Figure 5.2B). G32 C8 is included as a representative of a profile with no visible dispersion (Figure 5.2B). Unfortunately, dispersion could not be measured on C30 C6 or U31 C6 in the wtTAR sample due to spectral overlap. However, U31 C6 is not overlapped in EII-wtTAR (described in Chapter 4), so this sample was used to measure U31 C6 data in this chapter. A35 C2 was too exchange broadened to measure accurately.

To obtain more accurate rates of exchange and determine populations and the chemical shift difference between the different states, off-resonance decays were measured and fit to the two-state Fast and Laguerre equation as described in section 1.3.4. The best fit parameters for individual and global fits are included in Tables 5.1 and 5.2. All data fit well to a two-state model and do not appear to require a higher order state model. The residues undergoing slow exchange at the millisecond timescale were fit individually using the Laguerre equation (eq 1.19). The population of the excited states of G33 C1' and C8 and A35 C8 are all less than 1% (0.32%, 0.54%, 0.40%, respectively). The carbon chemical shifts of the excited states of G33 C1' and G33 C8 are located downfield by 2.58ppm and 2.20ppm, respectively (Figure 5.3). The excited state of A35 C8 is shifted upfield by 2.48ppm.

All residues undergoing slow exchange at the microsecond timescale (except U31 C6) fit best to the two-state Laguerre equation according to F-test statistics at the 95% confidence level (Table 5.2) and do not appear to require a higher order state model. Interestingly, F-test statistics at the 95% confidence level reveal that all of the microsecond motions (except U31 C6) can be fit simultaneously to a single exchange equation, suggesting that these fast motions likely report on the same two-state process. The fit yields a timescale of $39 \pm 1\mu s$ and the population of the excited state is ~13%. The excited state C1' carbon chemical shifts of C30, U31, G34, and A35 are all located downfield by approximately 2.1ppm, 2.1ppm, 2.3ppm, 2.6ppm, respectively (Figure 5.3). Surprisingly, the excited state of G34 C8, which is already shifted quite far downfield from the helical C8s, shifts even further downfield by 2.5ppm.
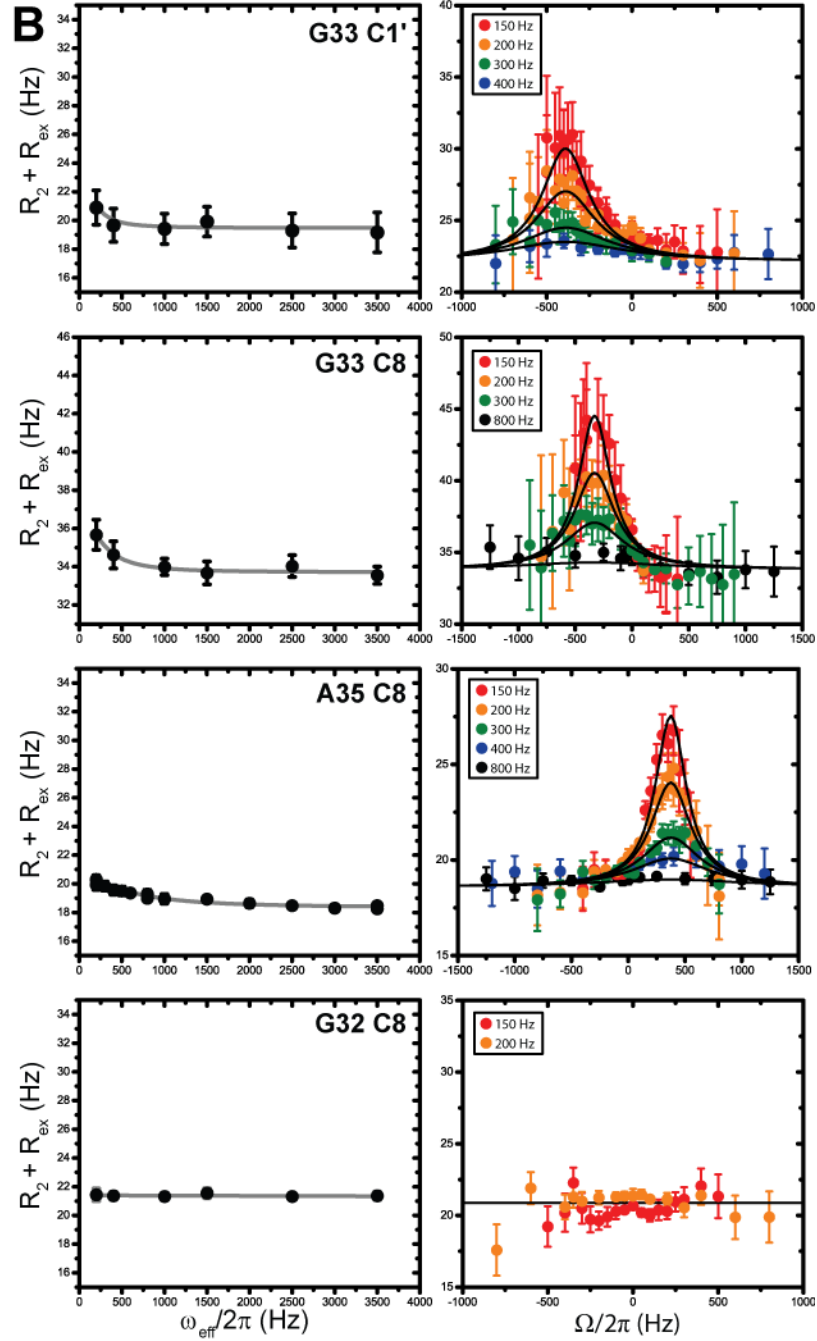
A

C30 C1'
$\tau_{ex} = 40 \pm 16\ \mu s$

U31 C1'
$\tau_{ex} = 42 \pm 21\ \mu s$

G34 C1'
$\tau_{ex} = 43 \pm 17\ \mu s$

A35 C1'
$\tau_{ex} = 43 \pm 9\ \mu s$

G34 C8
$\tau_{ex} = 29 \pm 19\ \mu s$

$\omega_{eff}/2\pi$ (Hz)

$\Omega/2\pi$ (Hz)

$R_2 + R_{ex}$ (Hz)

120

Figure 5.2: Characterization of slow motions in the wtTAR apical loop. On-resonance profiles (*left*) and off-resonance profiles (*right*) for (A) microsecond motions and (B) millisecond motions, except G32 C8, which has no visible dispersion. See insets for keys to spin lock powers ($\omega_{eff}/2\pi$).
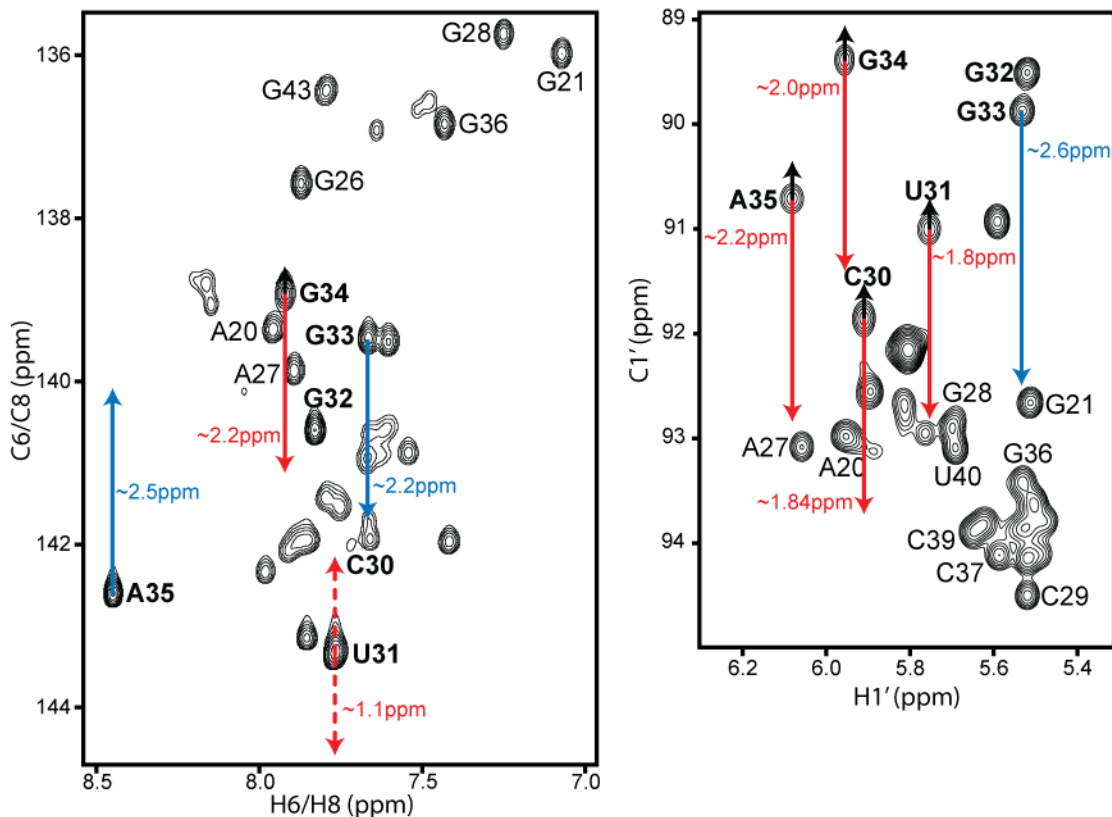
Figure 5.3: Chemical shifts of the wtTAR apical loop excited state. Red and black arrows indicate the location of the excited and ground state chemical shifts of microsecond motions, respectively. Blue arrows indicate the excited state of millisecond motions. The directionality of the U31 C6 chemical shift is not known. Loop resonances are bolded.

U31 C6 was found to fit the Fast equation best. Unfortunately, under the fast exchange limit, $R_{1\rho}$ is not dependent on the individual parameters p_a, p_b, and $\Delta\omega$ but on their combination (ϕ_{ex}); therefore, the chemical shift difference and populations of each state of U31 C6 cannot be obtained without additional information. Thus, the shared parameters from the global fit (p_b and k_{ex}) were used to calculate $\Delta\omega$ of U31 C6, giving a chemical shift difference of approximately 1.3ppm. However, because U31 C6 is in the fast exchange regime, the directionality of the shift cannot be determined without additional information (Figure 5.3).

It should be noted that although $\Delta\omega$ is the chemical shift difference between the excited and ground state, it can sometimes be treated as the difference between the observed chemical shift and the excited state. This simplification can be made for the millisecond motions that have very lowly populated excited states. However, the

populations of the microsecond motions are significant, so the observed chemical shift is population-weighted. Thus, the population of the excited and ground state must be taken into account to calculate their absolute chemical shifts.

The rates of exchange and chemical shift differences of the ground and excited states can be used to construct a probable excited state conformation of the apical loop. In this chapter we will focus on the microsecond motions of the loop, because this timescale has more probes that likely report on the same concerted process. The same cannot be said about the millisecond motions, which did not fit globally to any equation according to F-test statistics at the 95% confidence level. However, before characterizing the excited state of the apical loop, we must first define the ground state.

5.3.2 Ground State Conformation of the wtTAR Apical Loop.

The ground state C1' chemical shifts of all loop residues are located upfield of helical resonances, especially G32, G33, G34, and A35. Studies of the relationship between chemical shift and sugar pucker in RNA nucleotides have shown that an upfield C1' chemical shift signifies a C2'-endo sugar pucker, while C3'-endo sugar puckers have a downfield chemical shift [15,16]. Additionally, DQF-COSY experiments of HIV-1 TAR (data not shown) and HIV-TAR in similar buffer conditions [17] indicate that all the residues in the apical loop either adopt a C2'-endo or mixed C2′-endo/C3′-endo sugar puckers. Thus, in the ground state, it is likely that all of the loop residues have at least some percentage of C2'-endo sugar pucker, with G32 and G33 almost exclusively C2'-endo.

Previous NMR-solved structures of TAR and the RDC and relaxation study in Chapter 4 of the wtTAR apical loop suggest that U31, G32, and A35 bases are flexible and therefore possibly looped out in solution [18,19]. This agrees with previous chemical probing assays that show all three bases are exposed to solvent [20]. C30 and G34 do not experience significant dynamics on the ps-ns timescale, indicating that they are possibly located inside the loop and stacked on the C29-G36 base pair below the loop or involved in some other interaction, such as the proposed C30-G34 cross-loop base pair [20]. G33 is slightly more dynamic than the C30 and G34 bases, so it is unclear whether G33 is

looped out into solution or involved in a stabilizing interaction, such as base stacking on G34.

5.3.3 Excited State Conformation of the wtTAR Apical Loop.

The C1' carbon chemical shift of the excited state of C30, U31, G34, and A35 are all located downfield of the ground state, suggesting that the excited state may involve a sugar repuckering to the helical A-form C3'-endo conformation [15] (Figure 5.3). G34 C1' does not shift as far downfield as the other apical loop residues, so it is unclear if the sugar of G34 is C2'- or C3'-endo in the excited state. Even more interesting is the chemical shift of G34 C8, which shifts very far downfield, ~4ppm compared to helical guanine C8s. Guanine C8 chemical shifts in this range are commonly associated with a base in the *syn* conformation, as seen for example in the *syn* G of the UUCG tetraloop [16]. An *ab inito* calculation of carbon chemical shifts has qualitatively shown a similar trend [21] and *syn* Gs of G-tetrad DNA also have far downfield chemical shifts [22]. A recent study of a transient Hoogsteen base pair in DNA uses DFT calculations to show that a *syn* G is located ~3ppm downfield from a GC Watson Crick base pair [23]. However, a separate DFT calculation of RNA chemical shifts has reported contradicting results, with the C8 chemical shift moving upfield when *syn* [15].

To verify the trend using actual measured data, we examined the chemical shifts reported in the Biological Magnetic Resonance Bank (BMRB) [24] of *syn* Gs, which were identified using RNA FRABASE 2.0 [25]. Of the ~200 *syn* Gs solved by NMR and reported in the RCSB Protein Data Bank (PDB), only fourteen have carbon chemical shifts in the BMRB and three are UUCG tetraloops. Generally, we found that although *syn* Gs usually have a downfield chemical shift compared to helical Gs, they are not always as far downfield as the excited state of wtTAR G34. Instead, Gs with a chemical shift this far downfield tend to be stabilized by a cross-loop base pair, such as the GU reverse wobble observed in both a mutant of the P2b hairpin from telomerase RNA and in the UUCG tetraloop [26,27].

The rate of exchange provides additional insight into the exchange process. Johnson and Hoogstraten studied the dynamics of the sugars of the GCAA tetraloop [2]. They observed motions occurring on the 25-50μs timescale, which they attributed to

sugar repuckering. This timescale is very similar to the exchange rate of the C1' carbons of C30, U31, G34, and A35. NMR studies suggest that timescales of *anti*/*syn* transitions can range from nanoseconds to milliseconds, depending on whether the nucleotide is looped out in solution like the second U in the UUCG tetraloop (nanoseconds) [27] or is part of a G-G base pair (~1ms or faster) [28]. It stands to reason that if G34 is undergoing an *anti*/*syn* transition the timescale would fall somewhere between these two rates, because although G34 is involved in tertiary

Figure 5.4: Possible excited state of the wtTAR apical loop. *Syn* conformation of G34 is indicated by a red sugar moiety.

interactions, they are likely transient. Thus, the excited state chemical shifts and rates of exchange all support a sugar repuckering of C30, U31, A35 and stabilization of G34 in the *syn* conformation.
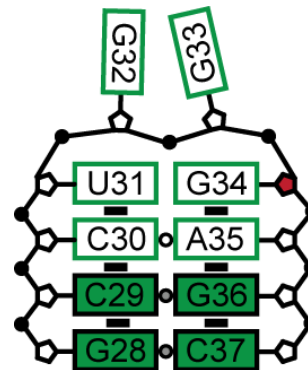
We must create an overall description of the apical loop excited state that appeases all the individual conformational changes described above. If we insist that residues C30, U31, A35, and possibly G34 adopt a helical conformation with a C3'-endo sugar, then one naturally generates a secondary structure involving a C30-A35 base pair (Figure 5.4). This leaves G34 in the YNMG-type tetraloop configuration, of which the G is often *syn* [27]. Formation of a C30-A35 base pair would require the breaking of the transient cross-loop base pair between C30 and G34 and would be accompanied by the flipping in of A35. This would propagate the A-form structure further up the helix, which would explain a sugar repuckering from C2'-endo to C3'-endo.

5.3.4 Design of Ground and Excited State Mimics.

It is difficult to further characterize the structure of the excited state because the excited state is "invisible" to traditional NMR structure-determination methods, such as NOE distance constraints. Thus, to further test the existence of a C30-A35 base pair, we have created a series of TAR mutants that attempt to trap the ground or excited state of the apical loop. This allowed us to compare the carbon chemical shift signatures of the wtTAR ground and excited state to those of the ground and excited state mimics. A

similar strategy was recently used by Al-Hashimi and corworkers to probe the existence of a transient Hoogsteen base pair in canonical DNA [23].

Figure 5.5: Ground and excited state mimics of the wtTAR apical loop. DMA substitution at position A35 is indicated by an asterisk.

To trap the excited state or at least push the equilibrium toward the excited state, two mimics were created that make the formation of a X30-Y35 base pair more favorable: TARU30 and TARG35 (Figure 5.5). If either of the mimics captures the excited state, the chemical shifts would be similar to those of the excited state. Note that we do not expect perfect agreement with respect to the magnitudes of the chemical shifts, because the mutations of the loop will likely affect the chemical shifts of surrounding residues.

As an inverse experiment, TARDMA35 was created to mimic the ground state (Figure 5.5). In this sample, A35 was substituted with N6, N6-dimethyl-adeosine (DMA). With two methyl group substitutions, DMA is bulkier than adenosine and less likely to flip into the loop to base pair with C30. If this construct successfully captures the ground state, we would expect the chemical shifts to be very similar to those of the ground state and there would be little or no dispersion on the microsecond timescale.

Due to complications arising from spectral overlap and the cost of full-length RNA, two of the constructs (TARDMA35 and TARU30) do not contain the bulge and are only 16nt in length. The removal of the bulge is justified by the results of Chapter 4, which show that there is likely no long-range interaction between the bulge and loop. The three RNA constructs were prepared and assigned as described in Materials and Methods (section 5.2). Assignments of TARU30, TARG35, and TARDMA35 are included in Appendix 2 (Figure A2.2 and A2.3 and A2.4, respectively)

5.3.5 Comparison of Excited State Chemical Shifts to ES and GS Mimics.

Comparison of the excited state (ES) chemical shifts to those of the ES mimics is very revealing (Figure 5.6). The directions of the ES chemical shifts agree with both
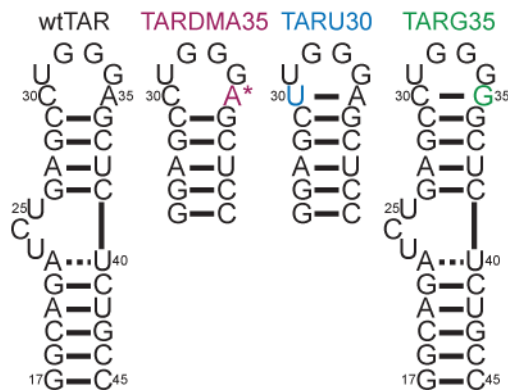
TARU30 and TARG35. Strikingly, the magnitudes of the shifts are in excellent agreement with those of TARU30. For example, U31 C1', G34 C1', and G34 C8 ES chemical shifts are all within 0.5ppm of TARU30 chemical shifts. The ES mimics also suggest that the direction of the U31 C6 ES, which could not be determined due to fast exchange, is upfield (Figure 5.6) and it is in excellent agreement with TARU30. Additionally, resonances that do not show large carbon chemical shift differences (>1ppm) between wtTAR and the ES mimics have no visible dispersion on the microsecond timescale. Examples include C1' and C8 of both G32 and G33 and G36 C8 (see Figure A2.3).

The only discrepancies occur in C30 C1' and A35 C1', which appear to agree better with TARG35. However, disparities in the chemical shifts of these two resonances may arise for two reasons. First, a direct comparison of the ES of C30 C1' and TARU30 and comparison of the ES of A35 C1' and TARG35 cannot be made because these are the sites of mutation (see open circles in Figure 5.6). In other words, it is inaccurate to compare the ES chemical shift of C30 to that of U30 in TARU30. Secondly, each of these carbons is involved in a mutated base pair (see asterisks in Figure 5.6). For example, in the ES of wtTAR, C30 is likely base paired to A35, but C30 is base paired to G35 in TARG35, which could cause a deviation in its chemical shift compared to the wtTAR ES. Overall, comparison of the wtTAR ES chemical shifts to those of TARU30 provides strong support for a cross-strand base pair between C30 and A35 in the excited state of the apical loop.

Comparison of the ground state (GS) chemical shifts to those of the GS mimic (TARDMA35) show that both the direction and magnitudes of the shifts are in excellent agreement, with all chemical shifts within 0.5 ppm of each other. The lack of chemical exchange in TARDMA35 is also revealing. Preliminary on-resonance dispersion data suggests that G34 C8 of TARDMA35 has little to no dispersion (Figure A.2.5A) on the microsecond timescale. Also, while A35 C2 is extremely exchange broadened in wtTAR, this exchange disappears on mutation of A35 to DMA (Figure A.2.5B). This, along with the chemical shift data, suggests that TARDMA35 does not sample the excited state to the same extent as wtTAR due to the A35 mutation.
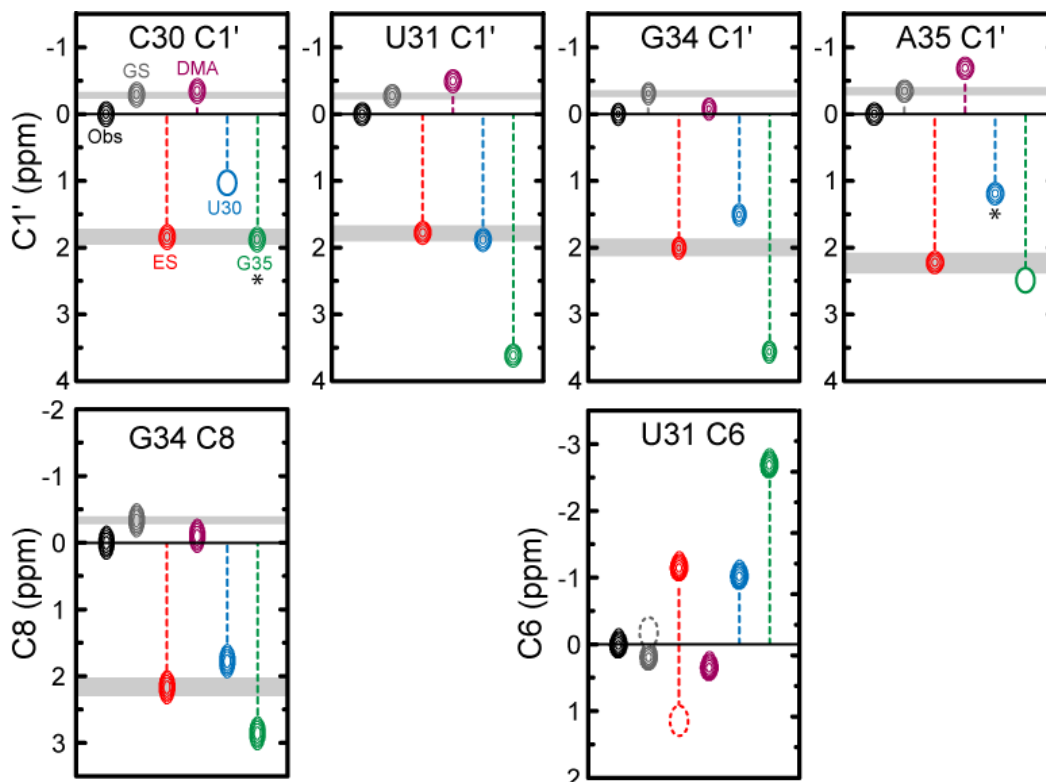
Figure 5.6: Chemical shift comparison of wtTAR ES and ES and GS mimics. Shown are the wtTAR population-weighted observed (black), ground state (gray), and excited state (red) chemical shifts. TARU30 (blue), TARG35 (green), and TARDMA35 (purple), chemical shifts are also shown. Open circles (such as U30 in the C30 C1' box) indicate a site of mutation. Direct comparisons should not be made with these peaks and were only included to show the directionality of the chemical shift. Asterisks denote carbons that participate in a mutated base pair. Dotted open circles indicate the degeneracy of the U31 ES and GS chemical shifts due to fast exchange. See C30 C1' box for labels.

5.4 Conclusions

In this chapter, we used a combination of $R_{1\rho}$ relaxation dispersion and ES and GS mimics to characterize the ground and excited state of the wtTAR apical loop. We have determined that G34 G8 and C1' of C30, U31, G34, and A35 all share a common two-state process occurring on the microsecond timescale. The chemical shifts of the ES and GS mimics and preliminary TARDMA35 dispersion data suggest that in the ground state A35 is looped out in solution as depicted in Figure 5.1. The excited state likely involves the flipping in of A35 to form a C30-A35 base pair accompanied by the repuckering of C30, U31, and A35 to a C3'-endo conformation and stabilization of G34 in the *syn* conformation (Figure 5.4). It is unclear whether G34 is C2'- or C3'-endo,

128

because both a transition to a *syn* conformation and sugar repuckering to a C3'-endo conformation result in a downfield chemical shift of the C1' carbon [15,23].

In one of the first NMR studies of TAR it was proposed that the apical loop could potentially be stabilized by the formation of a C30-A35[+] and U31-G34 base pair, but there was no compelling evidence for this structure at the time [29]. Chemical shifts from preliminary pH titrations indicate that a C30-A35[+] AC wobble base pair, in which the A35 N1 is protonated, is a possible candidate for the C30-A35 base pair of the wtTAR excited state (data not shown). Alternatively, other possible base pairs include the AC reverse hoogsteen or AC reverse wobble, both of which involve an interaction with the proton of N4, which is the site of mutation in the TARDMA35 mimic [30]. Additional pH titrations, more complex mutation studies, and DFT calculations will be necessary to determine the type of C30-A35 base pair of the excited state.

The relationship of the excited state of the wtTAR apical loop and biological function is not yet well understood. However, it does appear that loop mutations that promote the formation of the excited state adversely affect the ability of TAR to bind Tat and CycT1 [31]. For example, mutation of C30 to G30 results in a 13% decrease in binding capacity, likely due to disruption of the C30-G34 base pair, while mutation of C30 to U30 decreases binding by 31% compared to wtTAR. Similarly, while mutation of A35 to C35 has no significant affect, mutation to G35 decreases binding capacity to 24%. These results suggest that A35 must be looped out in solution or transitioning between a looped in and looped out state for biological function.

In fact, it was previously proposed that Tat may preorganize the apical loop for binding of CycT1 by stabilizing A35 in a looped out conformation [32]. In this study, Carlomagno and coworkers found that upon binding a neooligoaminodeoxysaccharide, the apical loop of HIV-2 TAR undergoes a conformational change, including the stabilization of a looped out A35 and efficient stacking of G34 on G36. The authors suggest Tat may induce a similar change, and that this reorganization in the loop would position the O6 and N7 atoms of G34 in the major groove where they would be accessible for CycT1 binding. In this way, the dynamic nature of A35 would serve as a regulation mechanism for the formation of the Tat:TAR:CycT1 complex.

The idea was conceived by Al-Hashimi, H. M. and Dethoff, E.A. The wtTAR RNA oligonucleotide was synthesized by Gulai N. All NMR data was recorded and analyzed by Dethoff, E.A.

5.5 References

[1] J.H. Lee, and A. Pardi, Thermodynamics and kinetics for base-pair opening in the P1 duplex of the Tetrahymena group I ribozyme. Nucleic Acids Research 35 (2007) 2965-2974.

[2] J.E. Johnson, and C.G. Hoogstraten, Extensive Backbone Dynamics in the GCAA RNA Tetraloop Analyzed Using C-13 NMR Spin Relaxation and Specific Isotope Labeling. Journal of the American Chemical Society 130 (2008) 16757-16769.

[3] H. Blad, N.J. Reiter, F. Abildgaard, J.L. Markley, and S.E. Butcher, Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. Journal of Molecular Biology 353 (2005) 540-555.

[4] A.L. Hansen, E.N. Nikolova, A. Casiano-Negroni, and H.M. Al-Hashimi, Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. J Am Chem Soc 131 (2009) 3818-9.

[5] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, Nmrpipe - a Multidimensional Spectral Processing System Based On Unix Pipes. Journal of Biomolecular Nmr 6 (1995) 277-293.

[6] T.D. Goddard, and D.G. Kneller, SPARKY 3, University of California, San Franscisco.

[7] B. Simon, K. Zanier, and M. Sattler, A TROSY relayed HCCH-COSY experiment for correlating adenine H2/H8 resonances in uniformly C-13-labeled RNA molecules. Journal of Biomolecular Nmr 20 (2001) 173-176.

[8] A.E. Derome, and M.P. Williamson, Rapid-pulsing artifacts in double-quantum-filtered COSY. Journal of Magnetic Resonance 88 (1990) 177-185.

[9] E.A. Dethoff, A.L. Hansen, C. Musselman, E.D. Watt, I. Andricioaei, and H.M. Al-Hashimi, Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. Biophys J 95 (2008) 3906-15.

[10] A. Bax, and D.G. Davis, Practical Aspects of Two-Dimensional Transverse Noe Spectroscopy. J. Mag. Reson 63 (1985) 207-213.

[11] R. Fiala, M.L. Munzarova, and V. Sklenar, Experiments for correlating quaternary carbons in RNA bases. Journal of Biomolecular Nmr 29 (2004) 477-490.

[12] A.L. Hansen, and H.M. Al-Hashimi, Dynamics of large elongated RNA by NMR carbon relaxation. J Am Chem Soc 129 (2007) 16072-82.

[13] L. Spyracopoulos, A suite of Mathematica notebooks for the analysis of protein main chain 15N NMR relaxation data. J Biomol NMR 36 (2006) 215-24.

[14] V.Z. Miloushev, and A.G. Palmer, R-1p relaxation for two-site chemical exchange: General approximations and some exact solutions. Journal of Magnetic Resonance 177 (2005) 221-227.

[15] X.P. Xu, and S.C.F. Au-Yeung, Investigation of chemical shift and structure relationships in nucleic acids using NMR and density functional theory methods. Journal of Physical Chemistry B 104 (2000) 5641-5650.

[16] G. Varani, C.J. Cheong, and I. Tinoco, Structure of an Unusually Stable Rna Hairpin. Biochemistry 30 (1991) 3280-3289.

[17] T. Carlomagno, I. Amata, J.R. Williamson, and M. Hennig, NMR assignments of HIV-2 TAR RNA. Biomolecular Nmr Assignments 2 (2008) 167-169.

[18] A.D. Critchley, I. Haneef, D.J. Cousens, and P.G. Stockley, Modeling and solution structure probing of the HIV-1 TAR stem-loop. J Mol Graph 11 (1993) 92-7, 124.

[19] M.J. Michnicka, J.W. Harper, and G.C. King, Selective Isotopic Enrichment of Synthetic Rna - Application to the Hiv-1 Tar Element. Biochemistry 32 (1993) 395-400.

[20] T. Kulinski, M. Olejniczak, H. Huthoff, L. Bielecki, K. Pachulska-Wieczorek, A.T. Das, B. Berkhout, and R.W. Adamiak, The apical loop of the HIV-1 TAR RNA hairpin is stabilized by a cross-loop base pair. J Biol Chem 278 (2003) 38892-901.

[21] R. Ghose, J.P. Marino, K.B. Wiberg, and J.H. Prestegard, Dependence of C-13 Chemical-Shifts On Glycosidic Torsional Angles in Ribonucleic-Acids. Journal of the American Chemical Society 116 (1994) 8827-8828.

[22] K.L. Greene, Y. Wang, and D. Live, Influence of the Glycosidic Torsion Angle on C-13 and N-15 Shifts in Guanosine Nucleotides - Investigations of G-Tetrad Models with Alternating Syn and Anti Bases. Journal of Biomolecular Nmr 5 (1995) 333-338.

[23] H.M. Al-Hashimi, E.N. Nikolova, E. Kim, A.A. Wise, P.J. O'Brien, and I. Andricioaei, Transient Hoogsteen base pairs in canonical duplex DNA. Nature 470 (2011) 498-U84.

[24] B.R. Seavey, E.A. Farr, W.M. Westler, and J.L. Markley, A relational database for sequence-specific protein NMR data. J Biomol NMR 1 (1991) 217-36.

[25] M. Popenda, M. Szachniuk, M. Blazewicz, S. Wasik, E.K. Burke, J. Blazewicz, and R.W. Adamiak, RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. BMC Bioinformatics 11 231.

[26] C.A. Theimer, L.D. Finger, and J. Feigon, YNMG tetraloop formation by a dyskeratosis congenita mutation in human telomerase RNA. Rna-a Publication of the Rna Society 9 (2003) 1446-1455.

[27] S. Nozinovic, B. Furtig, H.R.A. Jonker, C. Richter, and H. Schwalbe, High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. Nucleic Acids Research 38 (2010) 683-694.

[28] M.E. Burkard, and D.H. Turner, NMR structures of r(GCA(G)under-barGC(G)under-barUGC)(2) and determinants of stability for single guanosine-guanosine base pairs. Biochemistry 39 (2000) 11748-11762.

[29] J.A. Jaeger, J. SantaLucia, Jr., and I. Tinoco, Jr., Determination of RNA structure and thermodynamics. Annu Rev Biochem 62 (1993) 255-87.

[30] U. Nagaswamy, M. Larios-Sanz, J. Hury, S. Collins, Z.D. Zhang, Q. Zhao, and G.E. Fox, NCIR: a database of non-canonical interactions in known RNA structures. Nucleic Acids Research 30 (2002) 395-397.

[31] S. Richter, H. Cao, and T.M. Rana, Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. Biochemistry 41 (2002) 6391-7.

[32] D. Raghunathan, V.M. Sanchez-Pedregal, J. Junker, C. Schwiegk, M. Kalesse, A. Kirschning, and T. Carlomagno, TAR-RNA recognition by a novel cyclic aminoglycoside analogue. Nucleic Acids Res 34 (2006) 3599-608.

Chapter 6

Conclusions and Future Directions

6.1 Modulating Alignment and Motional Couplings by RNA Mutagenesis

RDCs are very powerful for probing the structural and dynamic properties of a biomolecule; however, in RNA, these measurements are often complicated by motional couplings. Our work in Chapter 3 provides insight into motional couplings and suggests that a simple structural modification, helix elongation, can give access to independent sets of RDCs, a necessity for full characterization of structural dynamics by RDCs which has not yet been fully realized for RNA. Furthermore, in Chapter 4, we demonstrate that when studying a small RNA fragment, such as an apical loop, a conservative elongation is adequate to achieve decoupling and allow for quantitative measurements of dynamics.

Additional mutations of the RNA sequence can potentially be utilized to achieve five independent alignments (Figure 6.1). For example, in addition to increasing the length of the lower helix, the length (X_m) and asymmetry (Y_n) of the bulge could be modified to change the overall alignment of the RNA of interest located above the bulge. A recent study by Al-Hashimi and coworkers demonstrated that the secondary structure of bulges encodes for topological constraints of the connected helices and that these constraints can be used to accurately predict the three-dimensional orientation of the two helices [1]. This knowledge could be used to effectively design bulges with the correct length and asymmetry to achieve unique alignments. In addition, even simple modifications, such as the addition of a single base pair above the bulge may change the phase of the helix such that the RNA of interest is rotated by 33° relative to the magnetic field.
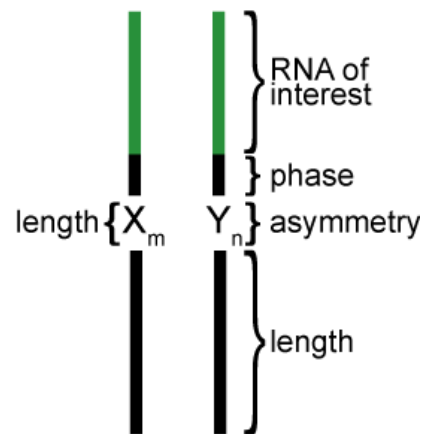

Figure 6.1: Modulation of RNA alignment and motional couplings by sequence mutagenesis.

6.2 Characterizing Dynamics of Apical Loops

In Chapter 2 and 4 we used RDCs, ^{13}C relaxation, and mutagenesis to characterize global and local motions occurring on the sub-millisecond timescale and found that the apical loop is structurally heterogeneous over a range of timescales. These dynamics are summarized in Figure 6.2. For example, the base of U31, G32, and A35 all undergo fast dynamics on the ps-ns timescale and are therefore likely looped out in solution, which agrees with chemical reactivity data [2] and previously solved NMR structures of the wtTAR loop [3-8]. We do not see direct evidence for a C30-G34 base pair, but our dynamic studies of the loop support the existence of this cross-loop base pair, which is proposed to be important for transcriptional activation [9]. We also demonstrated that although the wtTAR apical loop and bulge are complex dynamic centers undergoing complicated dynamics at multiple timescales, there is no detectable long-range interaction between the two. This justifies studying the dynamics of the loop separately from the bulge and vice versa.

In Chapter 5 we studied the slow motions of the apical loop (Figure 6.2) using ^{13}C $R_{1\rho}$ relaxation dispersion and mutagenesis to generate a picture of the ground and excited state conformations assuming a two-state model. The excited state of the loop likely involves a C30-A35 cross-loop base pair and stabilization of G34 in the *syn* conformation. This study represents the first-ever dynamic characterization of an "invisible" state of an RNA apical loop by NMR. Overall, our results establish a broadly useful approach for characterizing the dynamics of RNA loops that can be used to shed light on the relationship between dynamics and biological function.
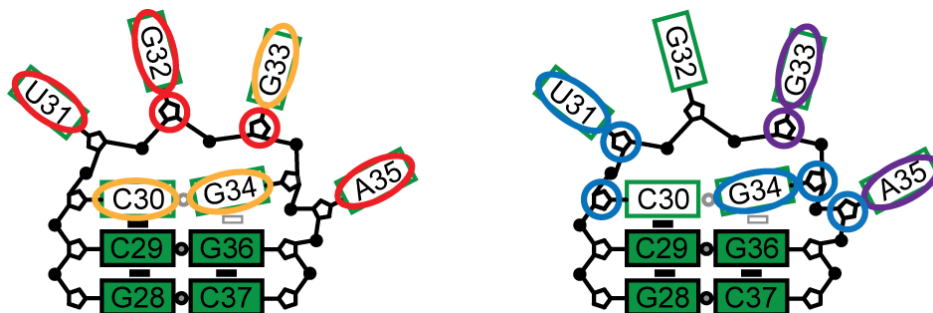

Figure 6.2: Summary of wtTAR apical loop dynamics. On the left, red circles indicate resonances that undergo significant dynamics on the ps-ns timescale, while yellow circles indicate resonances that do not experience significant dynamics on this timescale. On the right, blue and dark purple circles highlight resonances that experience chemical exchange on the microsecond and millisecond timescale, respectively.

133

The proposed excited state structure of the TAR apical loop (Figure 5.4) is most likely not the biologically active structure in the transactivation step of the HIV-1 genome, because the two mutations (C30 to U30 and A35 to G35) that were used to trap the excited state in Chapter 5 decrease the ability of TAR to bind Tat and CycT1 by 31% and 24%, respectively, in comparison to the wild-type loop [9]. This would result in a decrease in the ability of the TAR mutants to activate transcription. However, the excited state structure is still relevant in that it can be targeted by anti-HIV drug design strategies. For example, a small molecule or peptide could be designed that captures the biologically inactive excited state, thus inhibiting the interaction with Tat and CycT1 and formation of the transactivation RNP complex. Also, very little is known about the function of the apical loop in the other biological roles of TAR in the HIV-1 life cycle. It is possible that the proposed excited state structure is a biologically active structure in the interactions with proteins such as TRBP [10], PKR [11], or NC [12] or in the dimerization of the HIV-1 genome [13].

While the characterized dynamics and the two-state model of the excited and ground state of the apical loop are revealing, the wtTAR apical loop is ultimately undergoing very complex dynamics over a wide timescale and likely requires more than a simple two-state description. Unfortunately, while NMR spectroscopy can be used to detect motions over a vast timescale, an all-atom dynamic model cannot be constructed based solely on this data. In contrast, molecular dynamics (MD) simulations provide an all-atom description of motions, but the motional timescales are usually limited to hundreds of nanoseconds. The Sample-and-Select (SAS) method is a hybrid NMR-MD approach that constructs unbound RNA structural ensembles at atomic resolution [14]. Thus, a natural next step in the characterization of the wtTAR apical loop is the generation of an all-atom model of motion. The ensemble can be generated by incorporation of RDCs, relaxation parameters, and even chemical shifts. Not only will this approach provide insight into the motions of the loop, but it may aid in the identification of drugs that target the apical loop [15], a region that is commonly overlooked in anti-HIV drug discovery studies.

134

6.3 References

[1] M.H. Bailor, X. Sun, and H.M. Al-Hashimi, Topology links RNA secondary structure with global conformation, dynamics, and adaptation. Science 327 (2010) 202-6.

[2] T. Kulinski, M. Olejniczak, H. Huthoff, L. Bielecki, K. Pachulska-Wieczorek, A.T. Das, B. Berkhout, and R.W. Adamiak, The apical loop of the HIV-1 TAR RNA hairpin is stabilized by a cross-loop base pair. J Biol Chem 278 (2003) 38892-901.

[3] F. Aboul-ela, J. Karn, and G. Varani, The Structure of the Human-Immunodeficiency-Virus Type-1 Tar RNA Reveals Principles of RNA Recognition By Tat Protein. J of Mol Biol 253 (1995) 313-332.

[4] F. Aboul-ela, J. Karn, and G. Varani, Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. Nuc. Acids Res. 24 (1996) 3974-3981.

[5] C. Faber, H. Sticht, K. Schweimer, and P. Rosch, Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. J Biol Chem 275 (2000) 20660-6.

[6] Z. Du, K.E. Lind, and T.L. James, Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. Chem Biol 9 (2002) 707-12.

[7] B. Davis, M. Afshar, G. Varani, A.I. Murchie, J. Karn, G. Lentzen, M. Drysdale, J. Bower, A.J. Potter, I.D. Starkey, T. Swarbrick, and F. Aboul-ela, Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". J Mol Biol 336 (2004) 343-56.

[8] A.I. Murchie, et. al., Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. J Mol Biol 336 (2004) 625-38.

[9] S. Richter, H. Cao, and T.M. Rana, Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. Biochemistry 41 (2002) 6391-7.

[10] M. Erard, D.G. Barker, F. Amalric, K.T. Jeang, and A. Gatignol, An Arg/Lys-rich core peptide mimics TRBP binding to the HIV-1 TAR RNA upper-stem/loop. J Mol Biol 279 (1998) 1085-99.

[11] G. Clerzius, J.F. Gelinas, and A. Gatignol, Multiple levels of PKR inhibition during HIV-1 replication. Reviews in Medical Virology 21 (2011) 42-53.

[12] I. Kanevsky, F. Chaminade, D. Ficheux, A. Moumen, R. Gorelick, M. Negroni, J.L. Darlix, and P. Fosse, Specific interactions between HIV-1 nucleocapsid protein and the TAR element. J Mol Biol 348 (2005) 1059-77.

[13] E.S. Andersen, S.A. Contera, B. Knudsen, C.K. Damgaard, F. Besenbacher, and J. Kjems, Role of the trans-activation response element in dimerization of HIV-1 RNA. J Biol Chem 279 (2004) 22243-9.

[14] A.C. Stelzer, A.T. Frank, M.H. Bailor, I. Andricioaei, and H.M. Al-Hashimi, Constructing atomic-resolution RNA structural ensembles using MD and motionally decoupled NMR RDCs. Methods 49 (2009) 167-73.

[15] A.C. Stelzer, A.T. Frank, J.D. Kratz, M. Swanson, M.J. Gonzalez-Hernandez, J. Lee, I. Andricioaei, D.M. Markovitz, and H.M. Al-Hashimi, Discovery of Selective Bioactive Small Molecules by Targeting an RNA Dynamic Ensemble. Nat. Chem. Biol. In Press (2011).

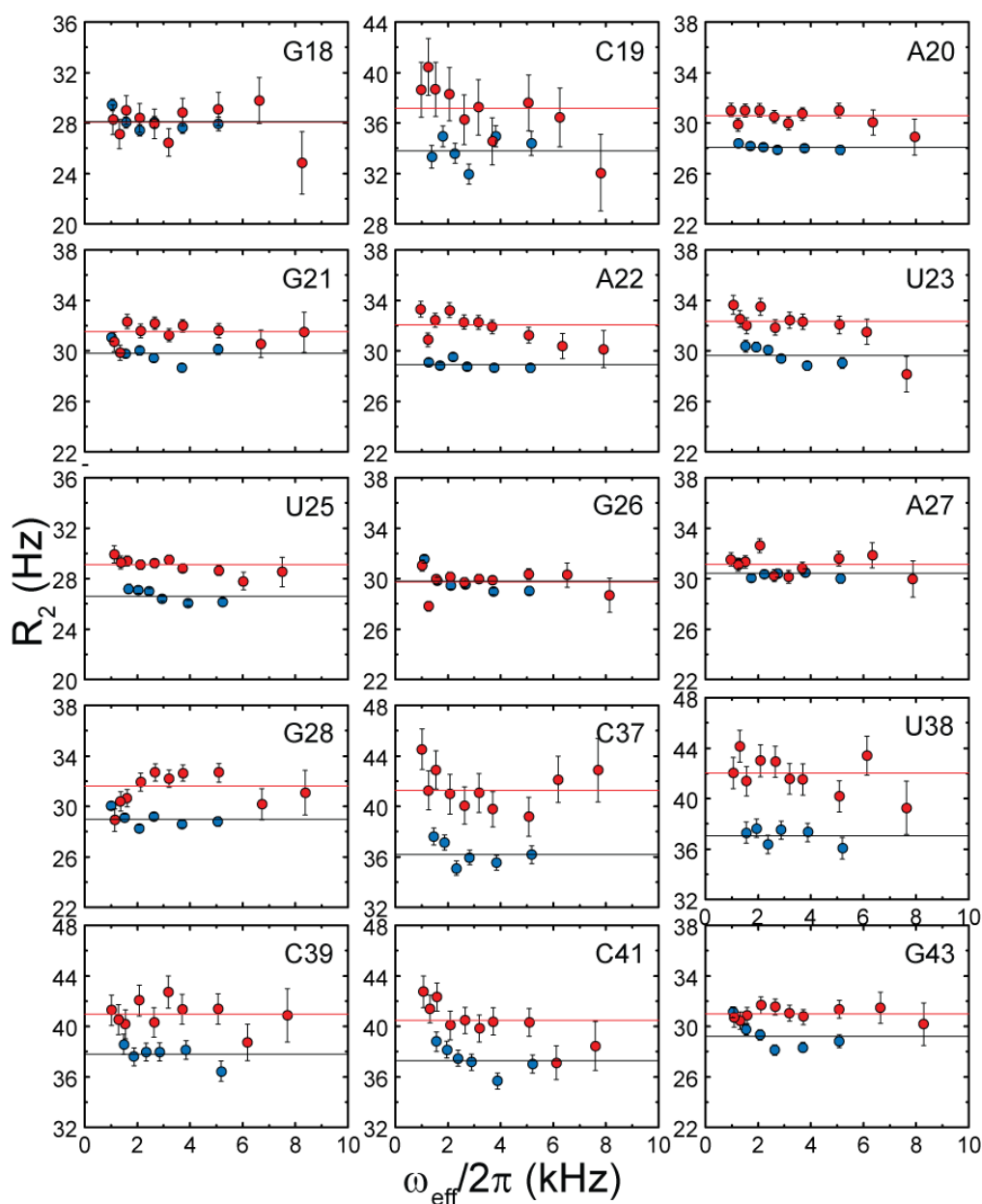**Summary of ^{13}C relaxation measurements of wtTAR and TAR$_m$**

Figure A1.1: Relaxation dispersion curves showing $R2$ as a function of field strength (ω_{eff}) and the absence of exchange for the majority of resonances in wtTAR (red) and TAR, (blue) as determined using F-statistics at the 99% confidence level using pulse sequence described in Chapter 2.

Appendix 2

^{13}C R1ρ relaxation dispersion measurements of wtTAR

Carbon	Delays (ms)
C30/U31/G34/A35 C1'	0,6(x2),14,24(x2)
G32/G33 C1'	0,14(x2),34,55(x2)
G32 C8	0,10(x2),21,34,45,55 (x2)
G33 C8	0,7(x2),13,21,28,35 (x2)
G34 C8	0,5(x2),11,20(x2)
A35 C8	0,10(x2),21,34,45,55 (x2)

Table A2.1: Delays used in $R_{1\rho}$ relaxation experiments of wtTAR.

Spin lock power (Hz)	Offset (Hz)	Spin lock power (Hz)	Offset (Hz)	Spin lock power (Hz)	Offset (Hz)
400	-1200	1600	-4800	2500	-7600
400	-1000	1600	-4270	2500	-7100
400	-800	1600	-3600	2500	-4800
400	-600	1600	-3200	2500	-4250
400	-400	1600	-2600	2500	-3750
400	-300	1600	-2000	2500	-3125
400	-200	1600	-1750	2500	-2500
400	-100	1600	-1400	2500	-1875
400	-50	1600	-1200	2500	-1250
400	0	1600	-1000	2500	-1000
400	50	1600	-800	2500	-625
400	100	1600	-600	2500	-500
400	200	1600	-400	2500	-375
400	300	1600	-300	2500	-250
400	400	1600	-200	2500	-125
400	600	1600	-100	2500	-50
400	800	1600	-50	2500	0
400	900	1600	0	2500	50
400	1200	1600	50	2500	125
		1600	100	2500	250
		1600	200	2500	375
		1600	300	2500	500

1000	-3000	1600	400	2500	1600
1000	-2600	1600	600	2500	1875
1000	-2300	1600	700	2500	2500
1000	-2000	1600	1300	2500	3125
1000	-1500	1600	1400	2500	3750
1000	-1250	1600	1750	2500	4250
1000	-1000	1600	2000	2500	4800
1000	-750	1600	2600	2500	7100
1000	-500	1600	3200	2500	7600
1000	-250	1600	3740		
1000	-100	1600	4270	3000	-9000
1000	-50	1600	4800	3000	-8000
1000	0			3000	-7000
1000	50			3000	-4000
1000	100			3000	-3000
1000	250			3000	-2000
1000	500			3000	-1000
1000	750			3000	-750
1000	800			3000	-500
1000	1400			3000	2000
1000	1500			3000	3000
1000	2000			3000	4000
1000	2300			3000	7000
1000	2600			3000	8000
1000	3000			3000	9000

Table A2.2: Example offsets and powers for off-resonance experiments of resonance with microsecond exchange (G34 C1').
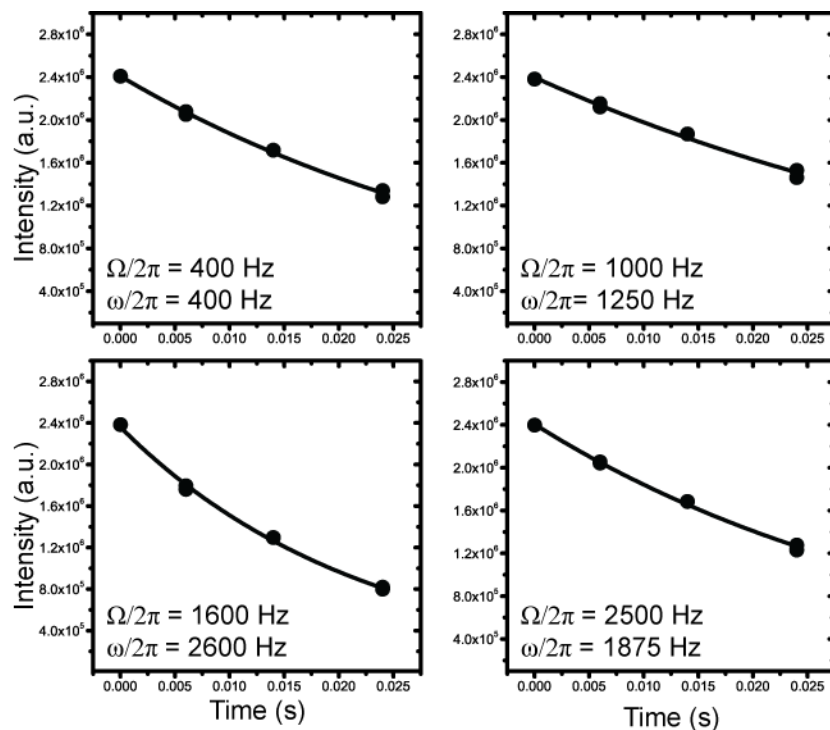
Figure A2.1: Representative mono-exponential decays for A35 C1' at varying offsets (Ω/2π) and spin lock powers (ω/2π).
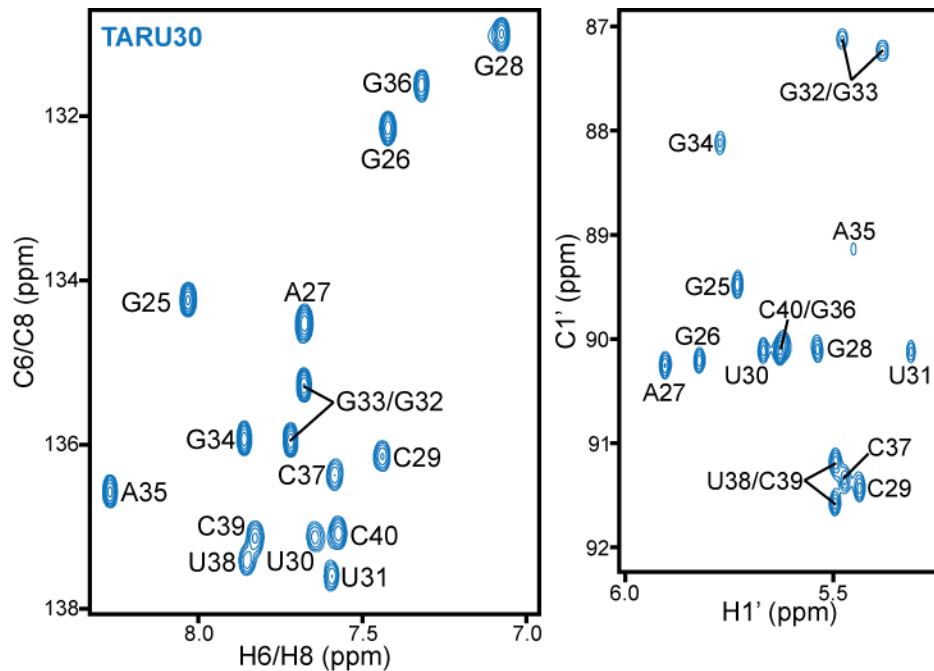

Figure A2.2: Resonance assignments of TARU30.

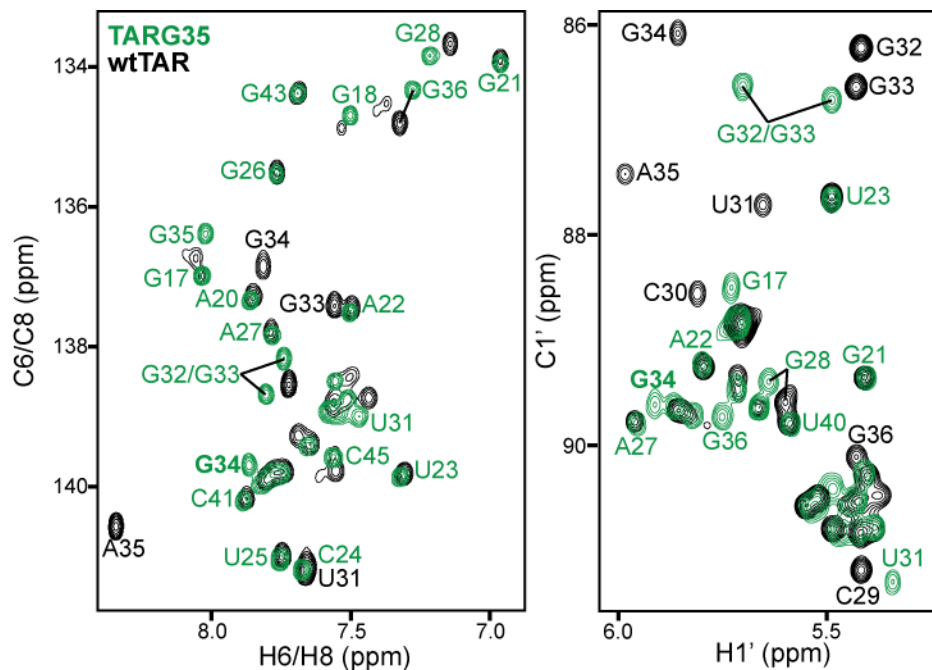**Figure A2.3:** Spectra overlay of TARG35 (green) on wtTAR (black). Resonance assignments of TARG35 and wtTAR are in green and black, respectively.
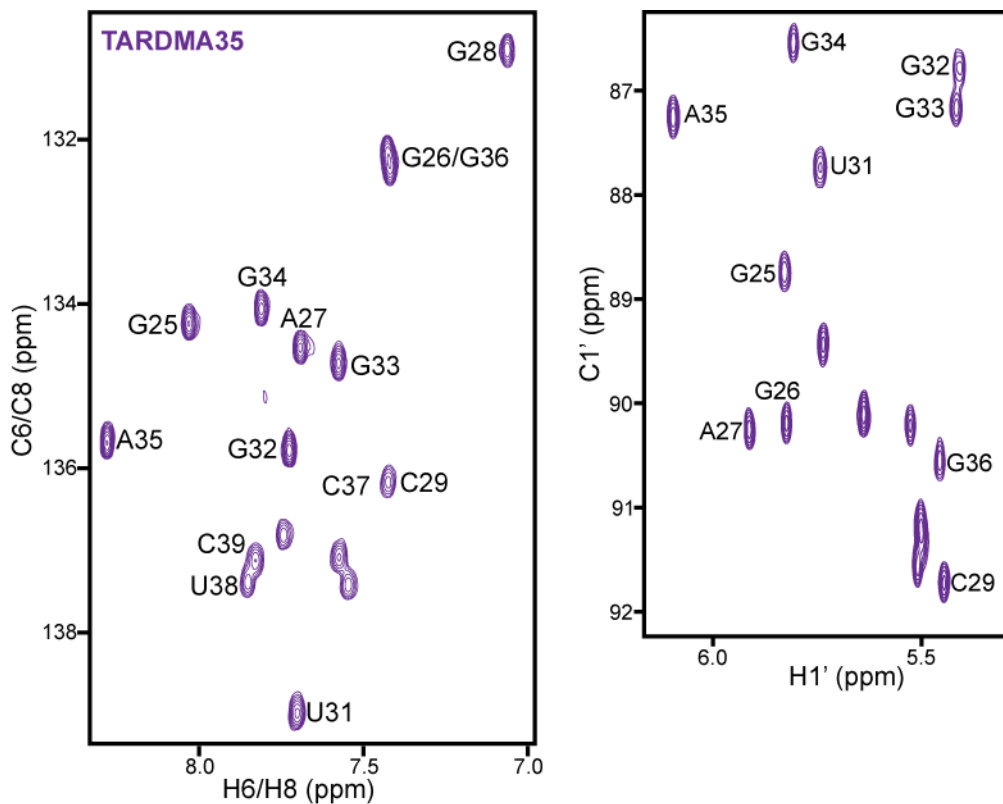

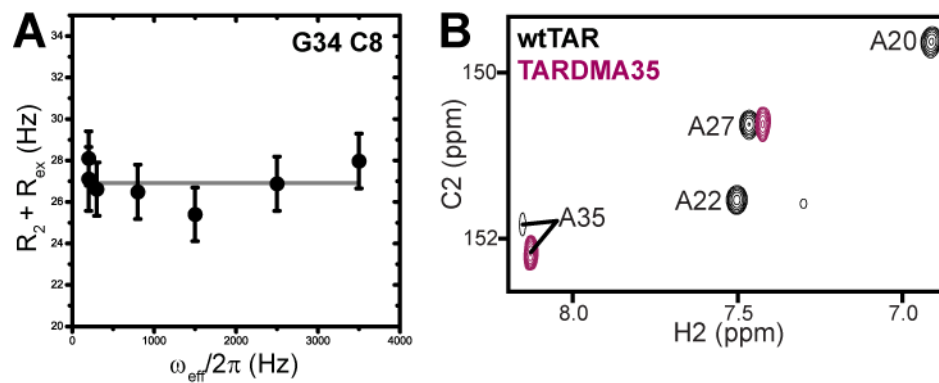**Figure A2.4:** Resonance assignments of TARDMA35.

Figure A2.5: Comparison of wtTAR and TARDMA35chemical exchange. (A) On-resonance dispersion profile of G34 C8. (B) Disappearance of exchange broadening of A35 C2 on mutation to DMA.