Targeting Dynamic Structures of RNA Using Experiment and Computation

by

Janghyun Lee

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2014

Doctoral Committee:

Professor Hashim M. Al-Hashimi, Chair
Professor Charles L. Brooks III
Professor Anna K. Mapp
Assistant Professor Sarah L. Veatch

Acknowledgements

I would like to first and foremost thank Prof. Hashim Al-Hashimi. It is lucky for me to learn scientific thinking and problem solving skills along with his great enthusiasm about science.

I would like to thank everyone from Al-Hashimi group, especially Andrew Stelzer, Jeremy Kratz, Elizabeth Dethoff, and Evgenia Nikolova. Andrew taught me the computational docking and the analysis of chemical structures of small molecules. Jeremy helped me to inherit the experimental techniques that Andrew developed after he graduated from the lab. Jeremy and Elizabeth taught me how to make my first RNA sample. Elizabeth also provided me valuable NMR experiments and her data that founded a part of my thesis. Evgenia taught me NMR R1ρ relaxation dispersion experiment, which is a key experiment in a chapter of my thesis.

I would like to thank the collaborators, especially Martha Larsen and Steve vander Roest at Center for Chemical Genomics (CCG) in University of Michigan. Without them, I would not have been able to perform the high throughput screening against TAR. I also would like to thank Dr. Marta Gonzalez-Hernandez from the Markovitz lab (University of Michigan) for teaching me cell-based assays and I would like to thank Dr. Steven Soll from the Bieniasz lab (The Rockefeller University) for performing the cell-based assays.

Table of Contents

List of Tables

List of Figures

Abstract

RNA is growing in its importance as a drug target and new approaches are needed in order to efficiently screen small molecules that bind and modulate the function of RNA. Targeting RNA with small molecules opens a door to the therapeutics of diseases that have not been approachable by targeting proteins and can also increase the effectiveness of existing treatments. In the present thesis, we used experimental high throughput screening utilizing a fluorescence-based displacement assay and ensemble-targeted virtual screening to screen a library of ~100,000 small molecules in search of compounds that inhibit the interaction between the transactivation response element of HIV-1 (TAR) and the Tat protein, a viral RNA-protein complex that is responsible for the transcription elongation. We showed that dimethyl sulfoxide (DMSO), a commonly used solvent to prepare small molecules for biochemical assays, has a small but significant effect on RNA structure and RNA-ligand binding that involves local melting of base-pairs near non-canonical motifs. The experimental and computational screens yielded complimentary small molecule hits, which includes 11 compounds with <+2 formal charge at neutral pH, average molecular weight of 480 g/mol, and IC50s ranging between 30 μM and 178 μM. One of the compounds showed significant transcriptional inhibition in cell-based assays utilizing TZM-bl reporter cell lines. Finally, we characterized a transient state structure of HIV-1 TAR using NMR relaxation dispersion experiment and mutagenesis that simultaneously modifies both a bulge and apical loop

motif that are separated by four base-pairs. This transient structure provides the basis for long-range communication between remotely positioned motifs and represents a new TAR target for developing ant-HIV therapeutics. The combination of experimental and virtual screening provides a general strategy for rapidly and effectively screening broad regions of chemical space in search of compounds that bind RNA and modulate their activity.

Chapter 1

Introduction

1.1 RNA Dynamics in Biology

Ever since the discovery of transfer RNA (tRNA) (1-3), the first non-coding RNA that departed from being a mere template for protein synthesis, myriad of non-coding RNAs has been discovered and understood to be involved in the regulation of RNA/DNA modifications (4), gene silencing (5, 6), transcription (7, 8) and translation (7). For example, HOTAIR (9) (HOX antisense intergenic RNA) is a long non-coding RNA that regulates the gene expression. It is particularly important for epigenetic differentiation of skin in different parts of the body, and the overexpression of HOTAIR is an indicator of metastasis in breast cancer (10). MicroRNAs and small interfering RNA (siRNA) are known to selectively silence the expression of genes by cleaving the complementary strand in the gene or by forming mismatches that blocks the transcription of the genomic RNA (6). Varying sizes, from as small as ~20 nucleotides to as large as few kilobases, the non-coding RNAs (Table 1.1) participate in many cellular processes and are linked to diseases (11, 12).

Type	**Associated Function**
Messenger RNA (mRNA)	Codes for protein
Ribosomal RNA (rRNA)	Translation
Y RNA	RNA processing; DNA replication
Telomerase RNA	Telomerase synthesis

Antisense RNA (aRNA)	Gene regulation
CRISPR RNA (crRNA)	Immune system; protects from parasites
Long noncoding RNA (lncRNA)	Varius functions
MicroRNA (miRNA)	Gene regulation
Piwi-interacting RNA (piRNA)	Gene silencing
Small interfering RNA (siRNA)	Gene regulation
7SK RNA (7SK)	Transcription control
Signal recognition particle RNA (SRP RNA)	Protein trafficking
Transfer RNA (tRNA)	Translation
Small nuclear RNA (snRNA)	Splicing
Small nucleolar RNA (snoRNA)	RNA nucleotide modification
SmY RNA (SmY)	Splicing
Guide RNA (gRNA)	RNA nucleotide modification
Ribonuclease P (Rnase P)	tRNA maturation
Ribonuclease MRP (Rnase MRP)	rRNA maturation, DNA replication
Circular RNA (CircRNA)	Gene regulation

Table 1.1 List of functional non-coding RNAs (11, 12)

Many of the newly discovered non-coding RNAs have functions that require large changes from their native conformation (13-16). These RNA conformational transitions take place in response to various cellular cues such as changes in pH and temperature, and binding to ligands (e.g. metabolites), proteins, and RNAs (13, 17-23). RNA conformational changes can involve re-arrangements in secondary structure that expose or sequester key functional elements(24-26); formation/ disruption of tertiary contacts that stabilize functional 3D structure(27, 28); and rapid librations of unpaired residues that drive conformational changes during adaptive recognition(21) (29, 30).

A well-described class of dynamic regulatory RNAs are riboswitches which represent ~50% of RNA structures solved to-date (24-26). The riboswitches are mainly found in prokaryotes, regulating gene expressions, thus are promising targets for development of antibiotics (24, 25, 31). Riboswitches are composed of an aptamer domain where a metabolite binds and consequently generating an expression platform. Binding of a metabolite to the aptamer domain ultimately induces the changes in the secondary structure of expression platform that facilitate or inhibit the transcription or the translation, thus modulating gene expression through modulation of RNA dynamics (Figure 1.1).

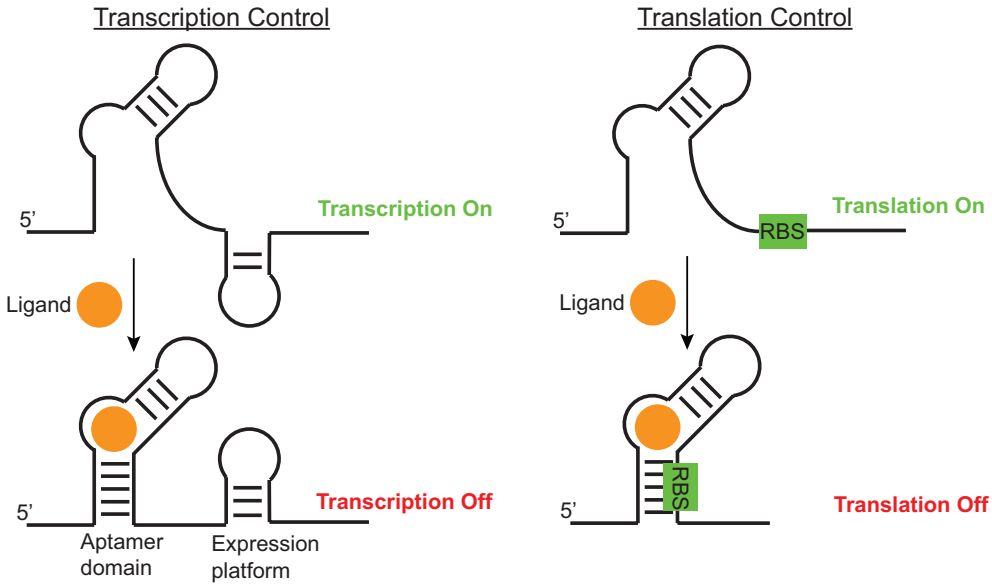

Figure 1.1 Conformational changes in regulatory RNAs. Illustration of transcription/translation control by a riboswitch

The assembly of RNA-protein complexes (RNP) is also a dynamic process that often involves sequential change in RNA structure induced by successive protein binding events (28, 32, 33). Such protein-induced conformational changes are thought to

direct the order of the assembly of ribonucleoprotein complexes. For example, the transactivation response element (TAR) (34) from human immunodeficiency virus type-1 (HIV-1) interacts with a complex comprising of transactivator protein, Tat, and positive transcription elongation factor b (P-TEFb) to promote the transcription elongation of viral RNA. The binding of Tat rearranges the helices of TAR from bent to a more linear conformation (21). Consequently this permits for binding to the positive transcription elongation factor b (PTEF-b), which is composed of cyclin T1 (cycT1) and cyclin-dependent kinase 9 (CDK9), and hyperphosphorylate the RNA polymerase II to promote the transcription elongation (35, 36). The conformational change of RNA is also important in ribosome assembly (37). For example, the binding of ribosomal protein S15 to 16S RNA ribosomal RNA re-orients the helices of 16S RNA that favors the subsequent binding of ribosomal protein S6 and S18 (38).

1.2 Non-coding RNA and Disease

A growing number of regulatory RNA elements are being linked to disease states, the list including viral and bacterial infections, cancer, genetic disorders, and neurodegenerative diseases (39-41) (Table 1.2). RNA viruses have non-coding RNA elements in the 5′- or 3′-ends of their genomes, that are indispensable to the viral life cycle, which offer novel targets for development of anti-viral compounds (42-44). Likewise, many riboswitches are targets for developing antibiotics (31, 45). MicroRNAs are deficient or overexpressed in cancer cells, providing new biomarkers to diagnose and treat the tumors (46, 47). All these non-coding RNA are potential therapeutic targets (Table 1.2).

Non-coding RNA	Affected function	Disease
RPS19, RPS24	ribosome biogenesis	Diamond-Blackfan anemia
DMPK (gain of function)	protein kinase	Myotonic dystrophy, type 1
ZNF9 (gain of function)	RNA binding	Myotonic dystrophy, type 2
JPH3 (gain of function)	ion channel function	Huntington's disease-like 2
FMR1	translation	*Fragile X syndrome*
TARDBP	splicing, transcription	Amyotrophic lateral sclerosis
ELF2AK3	translation (protease)	*Wolcott-Rallison syndrome*
miR-17-92 cluster	RNA interference	Cancer
RBM5	splicing	Cancer
16S ribosomal A-site	translation	Bacterial infection
Riboswitch	transcription, translation, splicing, mRNA degradation	Bacterial infection
Transactivation response element	viral transcription	HIV-1
Rev-response element	viral genome export	HIV-1
Dimerization initiation site	viral genome packaging	HIV-1

Table 1.2 Examples of diseases caused by non-coding RNA (31, 41, 42, 48-50)

1.3 Non-coding RNAs as a Drug Target

Antibiotics, such as aminoglycosides (51-53), tetracyclines(54), macrolides(55), and oxazolidinones(56), that target ribosomal RNAs are the most successful case of RNA-targeted drug discovery. These FDA-approved aminoglycosides, tetracyclines, macrolides, and oxazolidinones bind 16S ribosomal A-site, 30S ribosomal subunit, P-site

of 50S subunit, and 23S subunit, respectively. The molecular mechanism of action of these antibiotics have been unveiled recently by X-ray crystal structures of ribosomes (57-59) bound to antibiotics (60-62). For example, the bacterial 16S ribosomal A-site decodes the messenger RNA (mRNA) by flipping out two adenine bases in its internal loop (63, 64). These flipped out bases stabilize the nearby codon-anticodon helix, formed between the aminoacyl tRNA and the mRNA, facilitating the correct incorporation of amino acids by a ribosome (Figure 1.2A). X-ray crystallography and NMR studies (51, 65-67) revealed that the binding of an antibiotic, paromomycin, to the 16S ribosomal A-site induces the flipping of adenine bases (A1492 and A1493) in the absence of the correct codon-anticodon helix (Figure 1.2B), resulting in a misincorporation of amino acids which eventually kills a bacteria. Since then, high throughput assays targeting the 16S ribosomal A-site (68, 69) have been developed to discover novel antibiotics. Moreover, the success of targeting ribosomal RNA motivated targeting other regulatory RNA elements including various repeats, riboswitches, viral RNA and microRNA for bacterial infection, genetic disorder, viral infection, and cancer.

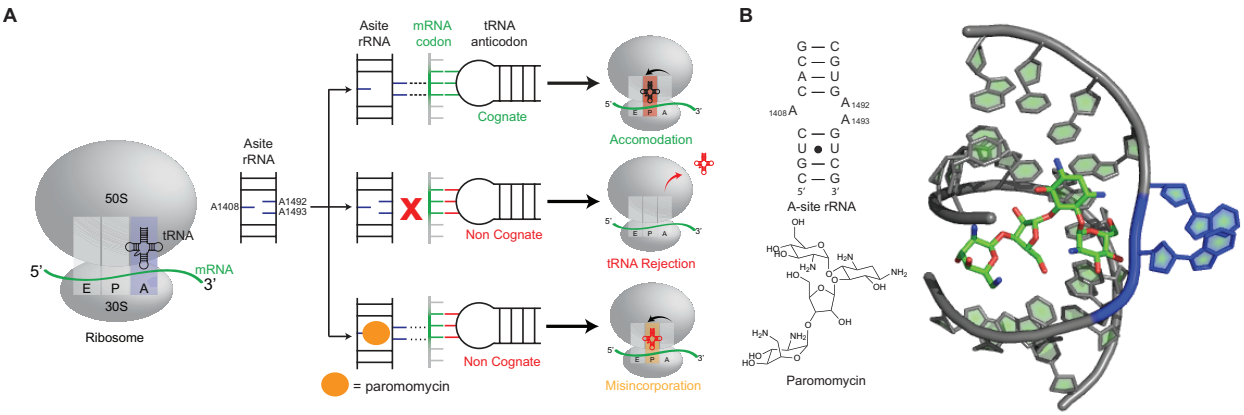

Figure 1.2. Mechanisms of antibiotics targeting 16S ribosomal A-site RNA. A. Mechanism of paromomycin **B.** X-ray crystal structure of paromomycin-bound A-site RNA. (PDB: 1J7T)

Riboswitches are emerging as a promising anti-bacterial target that potentially will broaden the chemical spectrum of antibiotics to fight off the bacterial resistance (26, 31, 70). They are mainly found in bacterial mRNA and have a well-defined binding pocket for a small metabolite in the aptamer domain, providing a nice pre-requisite for selective small molecule inhibitors. Currently, several historically known antibacterial compounds are suggested to be targeting riboswitches. For example, pyrithiamine (71, 72), an analogue of thiamine, inhibits the growth of bacteria and fungi. It was recently shown that pyrithiamine is readily phosphorylated in cells to become pyrithiamine pyrophosphate (73), which binds to the TPP riboswitch *in vitro (74, 75)*. Other similar cases exist such as for a lysine-analogue, L-aminoethylcystein (76) and DL-4-oxalysine (77) binding to lysine riboswitch (45) and for riboflavin-analogue, roseoflavin (78), binding to flavin mononucleotide (FMN) riboswitch (79). Certain riboswitches, for example glucosamine-6-phosphate riboswitch (80), are ribozymes in the presence of ligand, which offers convenience to design HTS assays to explore the chemical space of

small molecule inhibitors. Two assays have been developed utilizing Fluorescence Polarization (FP) (81) and fluorescence/Förster Resonance Energy Transfer (FRET) (82).

Mutations in certain alleles can result in inducing functionality for a RNA that is overexpressed and pathogenic. For example, myotonic dystrophy type 1 is caused by the overexpression of CUG repeats (83, 84) from 50 to 2500 repeats in 3'-untranslated region (UTR) of dystrophia myotonia protein kinase (*DMPK*) mRNA. The overexpression of CUG repeats sequesters CUG binding protein 1 (CUGBP1) and muscleblind-like 1 (MBNL1) proteins that results in the abnormal MBNL1-sensitive splicing events (85, 86) due to accumulation of MBNL1 proteins in the nuclear foci and down regulation of *DMPK* mRNA translation. Recently, Disney, M. D. and co-workers designed small molecules using the multivalency theorem (87-90) (Figure 1.3), linking 2 - 5 weakly binding chemical moieties to increase the overall binding affinity to the target. The small molecule inhibitor bound to CUG repeats with K_d ranging from 13 nM to 100 nM and was shown to alleviate the splicing defects and increase translation of *DMPK* mRNA (89). This strategy of optimizing small molecules using the multivalency theorem could be applicable to other pathogenic RNA repeats.

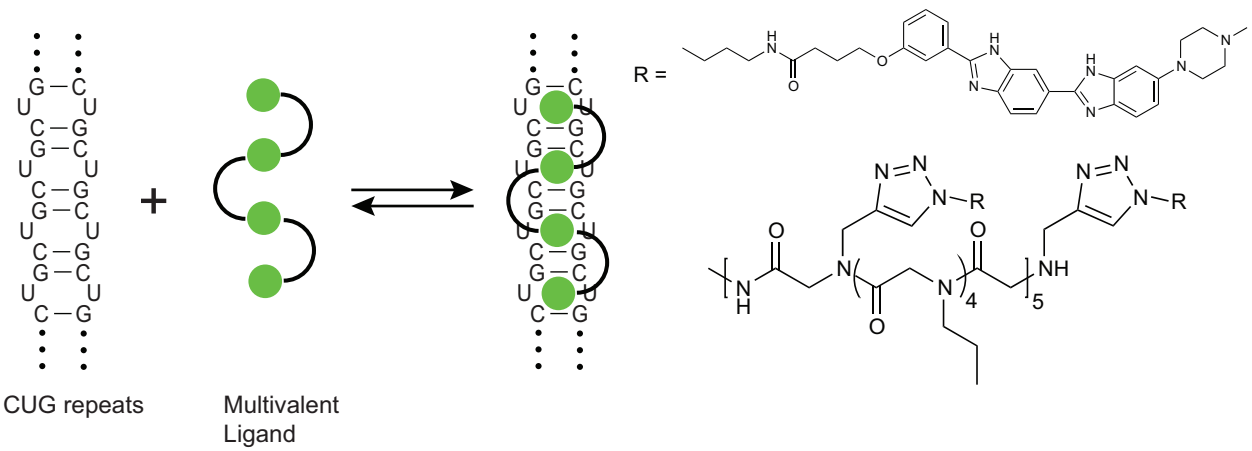

Figure 1.3 Targeting RNA repeats by a small molecule. A strategy to target CUG repeats by a multivalent ligand. An example multivalent ligand that binds CUG repeats is shown.

HIV-1 TAR is an example of a viral non-conding RNA drug target (34, 42). The current Highly Active AntiRetroviral Therapy (HAART) treatment program is mainly composed of reverse transcriptase inhibitors, protease inhibitors, and integrase inhibitors. The challenges of HAART treatment is that the HIV-1 strains rapidly gain resistance to the drugs, and the treatment plan currently is to change different drugs at regular intervals that pose a high risk of potential side-effects for the various drugs employed. In addition, the treatment has to continue on a long-term basis, as the latent viral population remains unaffected by the treatment and these dormant viruses are activated once the treatment stops (91, 92). HIV-1 TAR is an attractive therapeutic target as it participates in several steps in HIV-1 life cycle including transcription (36, 93-95), translation (96, 97), dimerization (98, 99), packaging (98, 100), and even viral latency (101, 102).

HIV-1 TAR is located 5′-UTR of viral RNA and it is the first element transcribed by the RNA polymerase II in HIV-1 RNA (103). Upon binding of transactivator protein

(Tat), the P-TEFb complex is recruited (104) and stimulates the elongation of viral RNA by hyperphosphorylating the RNA polymerase II (93, 105) (Figure 1.4A). Targeting TAR with small molecules has been focused on interfering TAR-Tat interaction. Interfering Tat and P-TEFb complex instead is more challenging due to the large protein-protein interaction area(106). So far approximately ~120 small molecules with 14 unique chemotypes including quinolone derivatives(107-110), polyamidoamines (111, 112), arginine-aminoglycoside conjugates (113, 114), and peptoids (115, 116), were discovered to inhibit TAR-Tat interactions with IC_{50} between 22 nM and 1.2 mM. Only two compounds were potent enough to proceed to the animal studies (117) (Figure 1.4B), but unfortunately none of them have reached to the clinical trials. Nevertheless, these studies suggest that there is high potential to target HIV-1 TAR with small molecules.

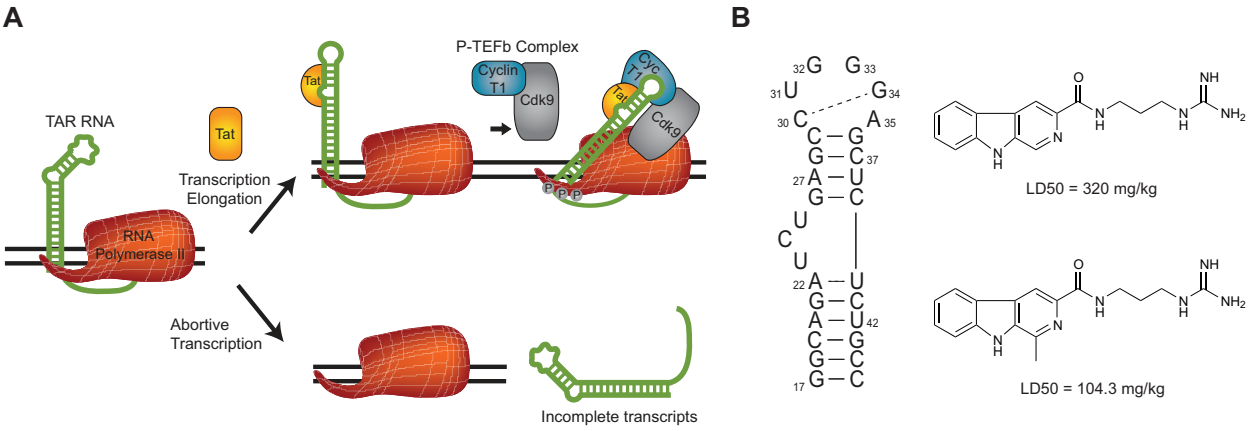

Figure 1.4 Tat-mediated transactivation in HIV-1 RNA. A. Mechanism of Tat-mediated transactivation involving TAR. **B.** Secondary structures of TAR and the structures of two small molecules that showed activity in mice (117).

MicroRNA (miRNA) also gained attention as novel therapeutic targets for treating cancers. With similar mechanism to siRNA, miRNA regulate the transcription and translation of genes, and misregulation of miRNA results in cancer. Specifically, certain

microRNAs are under-expressed or over-expressed in many cancer cells (47). These changes in the expression levels of miRNA are closely related to the oncogenic or tumor-suppressing activities and provide a ideal biomarker for cancer diagnosis (40). For example, the low expression level of miR-122 in liver cancer is (118) results in the down-regulation of cell's apoptotic pathway (119). Recently, small molecule modulators of miRNA expression in cancer cells were discovered targeting miR-122 for liver cancer (120). Interestingly, the small molecules reverted the low expression level of miR-122 in liver cancer and re-activated the apoptosis of the cells. Similar reverse effect was also observed for other small molecules targeting miR-21 (121), which is involved with breast, ovarian, and lung cancers. More recently, another study (122) showed that combining the microarray-based screening and ligand-based computational screening, a small molecule was targeted specifically for inhibiting the cleavage of miRNA by Drosha or Dicer protein complexes, thus suppressing the expression level of miRNAs in cells. These studies indicate the promise/potential of miRNA as a drug target for various types of cancer.

1.4 Principles for RNA-small molecule recognition

Early studies on the nature of RNA-small molecule recognition dating back to 1990s showed that positively charged small molecules, such as aminoglycosides and amino acid groups such as lysine and arginine in proteins, can bind a variety of regulatory RNAs including 16S ribosomal RNA, viral RNA, and ribozymes and modulate the cellular processes. Soon after, NMR and X-ray crystal structures of RNA-small molecule complexes provided detailed structural insights into RNA-small molecule recognition (21, 65, 123) (124). These studies, in addition to providing structural insights of the interactions between RNA and small molecules, also revealed

that RNA-small molecule recognition is often accompanied by significant/substantial changes in RNA conformation referred to as "adaptative recognition" (21, 26, 29, 30, 51) (Figure 1.5)

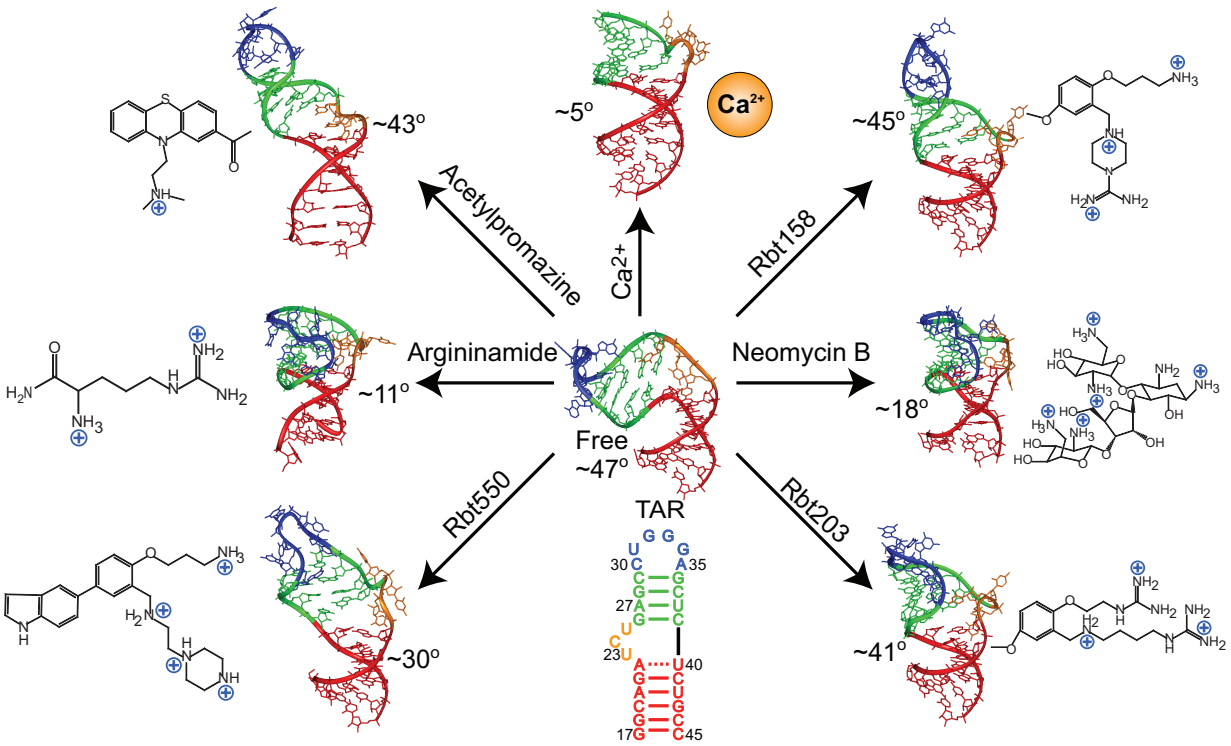

Figure 1.5 Examples of structural adaption by TAR upon binding of ligands (125)

For example, the structure of HIV-1 TAR bound with argininamide (21), an analogue of arginine, showed that TAR undergoes a large structural change upon binding, resulting in a specific conformation. First, the global orientation of two helices connected by a trinucleotide bulge rearranges from a bent state to a collinearly stacked state resulting due to the change of the inter-helical angle from 47° to ~11° (21, 126, 127). This global conformational change is brought about by U23, the first nucleotide in the bulge, transitioning from being stacked with A22 to positioning into the major groove

of A27-U38 to interact with the Hoogsteen interface of A27 and forms an A22-A27-U38 base triplet (21, 128, 129). This unique conformation of TAR allows argininamide to improve the binding specificity by stacking with U23 and forming hydrogen bonds with the phosphate backbone of U23 and G26 base. Additional structural studies such as ribosomal A-site RNA bound with aminoglycosides (51, 65, 130) and ribozymes (131, 132), and supplementary biochemical studies with modified aminoglycosides (133) further supported the importance of electrostatics, hydrogen bonding and stacking interactions in RNA-small molecule recognition.

The aptamer domains of metabolite-sensing riboswitches extensively utilize hydrogen bond networks to bind ligands with exquisite selectivity and affinity. The aptamer domains of riboswitches are often disordered in the absence of ligand but assume intricate 3D structures on adaptive ligand recognition, forming deep pockets that engulf ligands such that that almost every single atom of the ligand interacts with the nucleotides in the riboswitch via hydrogen bonding or Van der Waals interactions (134). For example, guanine binds to a guanine riboswitch, forming 8 out of 9 possible hydrogen bonds directly with the nucleobases present in the riboswitch (135, 136). The mutations of the nucleotides that form a binding pocket can entirely change the binding profile of the riboswitch. For example, a mutant C74U guanine riboswitch shifts the a specificity profile from guanine to adenine and pyrimidine-derivatives (137, 138). It is also interesting to note that merely increasing the number of hydrogen bond donors and acceptors in a small molecule does not improve their binding affinity to the riboswitch (138). Thus, the specific positions of hydrogen bond donors and acceptors are important for binding specifically and tightly. Other than the direct hydrogen bonds, the indirect hydrogen bonds between RNA and a small molecule via water

molecules were also observed in the small molecules bound to the solvent-exposed binding pocket, especially in aminoglycoside-bound ribosomal A-site RNA (65).

The stacking interactions also play an important role in small molecule binding to RNA (65, 139). A well-known stacking interaction is π-π stacking interaction that is common among aromatic compounds such as nucleobases. It is in fact the major driving force of the nucleic acid secondary structure formation (140, 141). The extensive stacking interactions observed in riboswitches where the binding site almost entirely engulfs the purine or pyrimidine ligands (135). NMR structure of mitoxantrone, an anthracenedione drug, bound to *tau* splicing regulatory element showed an extensive stacking of mitoxantrone between base pairs of G17-C3 and C16-G1 (139). Other stacking interactions such as nonionic stacking and cation-π interactions are also observed in glycosidic groups and guanidium groups, respectively, of small molecules such as aminoglycosides and argininamide bound to RNA (21, 65). The guanidium functional group of arginine is one of the most favorable RNA binding chemotype with 5 hydrogen bond donors and acceptors and an ability to stack with nucleobases.

The electrostatic interactions also play vital role in small molecule binding to RNA. Due to intrinsic nature of negatively charged phosphate groups in nucleic acids, it is intuitive to see positively charged small molecules would favor binding to RNA. Indeed, studies have shown that aminoglycosides compete for divalent cation binding to tRNA (142) and self-splicing group I intro ribozyme (143, 144). In another study (133), deletion of hydroxyl groups adjacent to the primary amine groups, to increase their pKa, in aminoglycosides showed an increase in their binding affinity to hammerhead ribozyme. Additional studies also showed that the binding affinities of aminoglycosides enhances with increase in number of positively charged amine groups (131). Moreover,

the binding affinities of aminoglycosides diminish with increase in the concentration of monovalent or divalent cations in a solution. A recent study (145) showed that the aminoglycosides with the same number of primary amines have different salt dependencies on their binding affinity. This suggests that the electrostatic interactions are not non-specific that merely depend on the number of positive charges and therefore can be exploited to achieve the specific binding of small molecules by positioning them appropriately.

Unlike B-form DNA which forms a binding site for a small molecule at the major and minor grooves, the RNA A-form helix accommodates small molecule binding less favorably(146), because the major groove is narrow with width of 4 Å and the minor groove is wide but exceedingly shallow for small molecule binding. Nevertheless, the RNA major groove is a key site for small molecule recognition (146, 147). The major groove of RNA possesses more hydrogen bond donors and acceptors than the minor groove, and is relatively more electronegative than the minor groove (148). In fact, many RNAs including tRNA, group I intron, and HIV-1 TAR recognize other molecules via interactions at the major groove. The presence of un-paired nucleotides that form bulges, internal loops, and apical loops assist in perturbing the A-form helix by widening the major groove for recognition by small molecules, proteins, and other RNAs. This perturbation also results in an electronegative bias for certain major grooves over others creating "hot spots". These hot spots are often observed to provide the binding sites for metal ions and small molecules and thus could be exploited to guide the rational design of small molecules binding to RNA (149).

1.5 Experimental Methods to target RNA

Assays for measuring binding of small molecules to RNA

Although there are many experimental approaches that are employed to measure the binding affinity and inhibitory activity of small molecules targeting proteins, developing assays for small molecule binding or inhibiting RNA-protein complexes is impeded due to several reasons. Firstly, many non-coding RNAs are non-enzymatic therefore cannot be assayed based on the perturbation of activity. Secondly, RNA is flexible and this often introduces challenges in labeling reporters that probe the small molecule binding. Nevertheless, several assays have been developed to probe the RNA-ligand interactions.

Fluorescence spectroscopy is the widely employed spectroscopic technique to probe RNA-ligand interaction and to measure the dissociation constant (K_d) of small molecule binding to biomolecules. Stable fluorophores have been developed for studying small molecule binding to RNA. Amongst various fluorophores, 2-aminopurine (2-AP) is the commonly used fluorophore for nucleic acids assays. 2-AP is a base analogue of adenine (6-aminopurine) and it has been demonstrated to cause minimal perturbation to the RNA structure and stability upon incorporation. The changes of fluorescence intensity of 2-AP is predominantly contributed by stacking interactions (150-152), making it an ideal fluorophore to study the changes in RNA secondary structure by fluorescence spectroscopy. For example, the argininamide binding of HIV-1 TAR has been studied with 2-AP labeled at C24 or U25 in the trinucleotide bulge region (153). The NMR studies(21, 154) have shown that the nucleobases of the bulge residues C24 and U25 flip out upon binding of argininamide to TAR, and as predicted the fluorescence intensity of 2-AP labeled at C24 or U25 increased upon binding (Figure 1.6A). Similarly, the 2AP-based assay was developed for 16S ribosomal A-site RNA (155). Again, the structural analysis (63, 65) of A-site in the absence and the presence of aminoglycosides showed that two adenine residues,

A92 and A93, flipped out as the aminoglycosides bound. As expected, substituting A92 with 2AP showed the increase in the fluorescence intensity upon binding of paromomycin (Figure 1.6B).

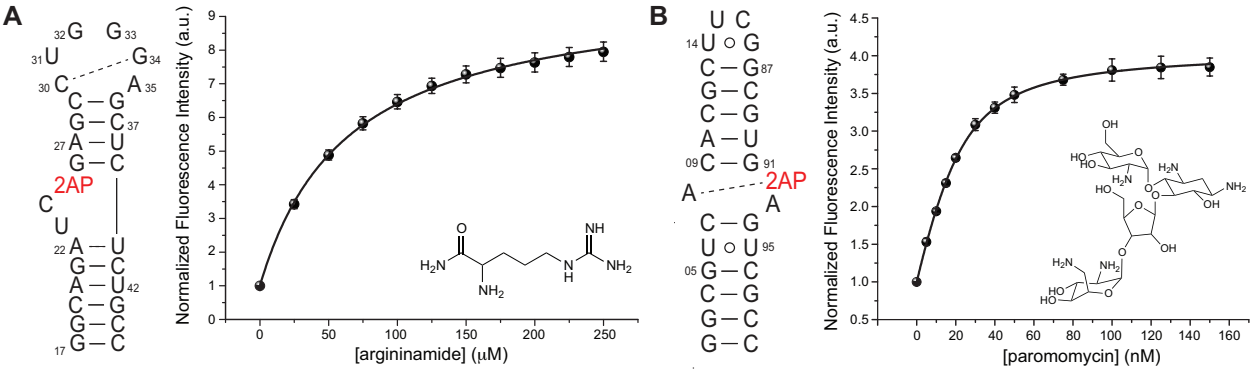

Figure 1.6 2AP binding assay for RNA. A. Measurement of argininamide binding to TAR by 2AP assay. **B.** Measurement of paromomycin binding to A-site by 2AP assay.

Despite its great adaptability to RNA structures, the application of 2-AP based fluorescence assay precludes implementation in a high throughput mode, since the excitation and emission wavelengths of 2-AP lie in the UV range resulting in interference from any small molecule prospects that contain aromatic ring moieties and thereby restricting its application to a limited chemical space of small molecules. Other fluorophores that mimic nucleobases have been developed(156), but were not successful in resolving the problem. Alternatively, the fluorescein is occassionally used as its excitation and emission wavelengths are in the visible range (485nm to 520nm). However, fluorescein based implementation are limited by the fact that they can be labeled only at the terminal ends of RNA, owing to their bulky structure, since terminal ends are generally insensitive to small molecule binding. Thus, the development of a novel fluorophore that fluoresces in the visible light range to study RNA-small molecule interaction is attractive and would prove beneficial.

As an alternative to fluorescence-based assays, other techniques such as mass spectroscopy (MS), nuclear magnetic resonance spectroscopy (NMR), isothermal calorimetry (ITC), surface plasmon resonance (SPR) and gel electrophoresis have been employed to probe small molecule binding to RNA (157-160). Although they can be used to confirm the binding and measuring the affinity, however they cannot be employed for HTS.

High Throughput Screening Targeting RNA

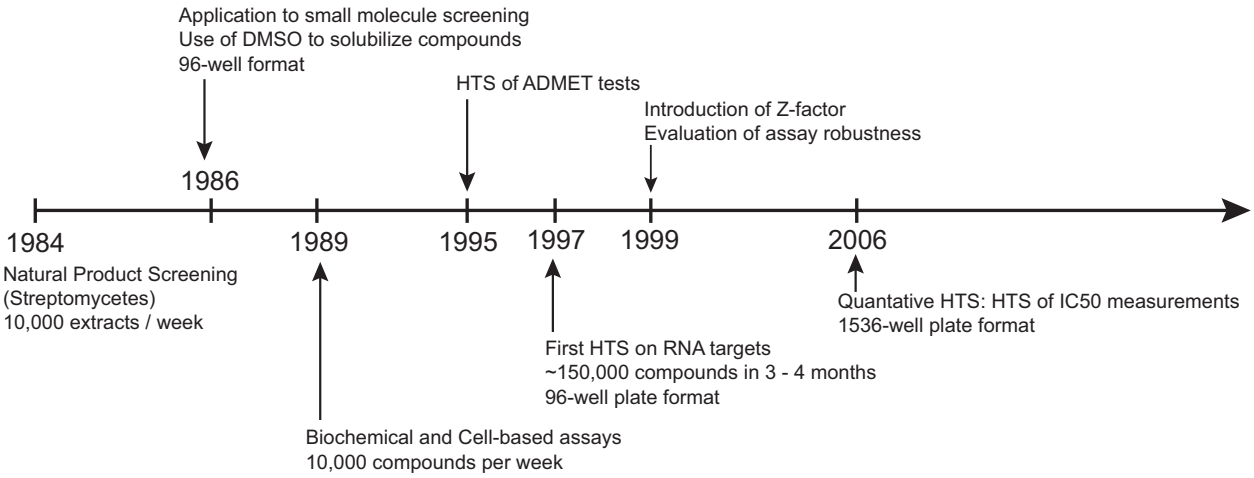

Figure 1.7 Brief timeline of technical evolution in HTS (161-166).

HTS is the optimal way to explore the chemotypes of synthetic compounds that modulate the biological target of interest and Figure 1.7 shows a brief timeline of HTS development. The origin of HTS is controversial, but Pfizer was certainly one of the early adopters of HTS for natural product screening of 100,000 cloned DNA from Streptomycetes in 1984(161). Using 24-well microplates, the researchers were able to increase the throughput of the HTS to 10,000 extracts per week. The automated systems

for the natural product screening have been implemented to the small molecule screening by Pfizer in 1986, using dimethyl sulfoxide (DMSO) to solubilize the compound library (161). Since then, the HTS have been quite successful on many protein targets yielding ~90 drug candidates from HTS that are in clinical trials and 15 drugs that are approved for market(167).

Despite the rapid development of HTS formats over the last three decades along with many successes in drug discovery, proteins have remained the primary targets of HTS and it was only in 1996 the first RNA-targeted HTS was published by Czarnik, A. W. and co-workers (164) for two different ribozymes, self-splicing group I introns and self-assembled ribozyme. Using the radio-labeled GTP (α-^{32}P-labeled GTP) and high throughput filtration method, the small molecules that inhibited the GTP-mediated cleavage by the group I introns were identified by measuring the radioactivity of the cleaved RNA products in 96-well microplates. Similarly, the small molecule inhibitors of self-assembled ribozyme were identified using the radio-labeled ribozyme. Among the 80 compounds tested, 2 compounds showed >70% inhibition of self-splicing group I introns. However, the size of the chemical library was not significantly big (80 compounds; a single 96-well microplate) to be called a HTS.

A year later, Czarnik, A. W. and co-workers simultaneously published two HTS campaigns (165, 166) screening 150,000 small molecules against the aforementioned ribozymes (self-splicing group I intron and self-assembled ribozyme) and HIV-1 TAR-Tat interaction being the first non-enzymatic RNA target for HTS (165). Using the Scintillation Proximity Assay (SPA) and filtration assay, 150,000 small molecules were screened against HIV-1 TAR. From the initial screen, 3,200 and 2,000 compounds were found to be active in SPA and 2,000 in filtration assay. Subsequent confirmation assays

for reproducibility using freshly prepared compounds confirmed one third of the initial hits as active (~500 compounds). These 500 compounds were then subjected to the cellular assays resulting in ~20 non-cytotoxic compounds with IC_{50} < 50 μM. They were finally identified as inhibitors of TAR-Tat interaction from the HTS, yielding the final hit rate of 0.01%. It is to be noted that the molecular structures of these inhibitors or any other hits from this study were not disclosed. Using the aforementioned assays (164) for ribozymes, ~150,000 compounds were screened against the self-splicing group I introns in 96-well microplate format(166). Including the known inhibitors such as aminoglycosides and guanosine, ~1,000 compounds were identified as hits. Three new chemotypes with IC_{50}s of 5 – 20 μM were identified from the screen. Since then, few more HTS campaigns targeted 16S ribosomal A-site RNA (68), HCV SL-3e RNA (168), and HIV-1 TAR (using different HTS assays) (169), but in general only a few HTS campaigns for RNA are published in the literature.

While the technical advances boosted the screening throughput, statistically evaluation of the results from the HTS (reviewed in (170)) became essential. Z-factor (163) is a commonly used parameter to evaluate the quality of a HTS assay and is a ratio between the standard deviations from individual measurements of controls and the dynamic range of the assay as given in Eq. 1.

$$Z = \frac{3(\sigma_{\text{negative}} + \sigma_{\text{positive}})}{|S_{\text{negative}} - S_{\text{positive}}|} \qquad \text{(Eq. 1)}$$

where σ_{negative} and σ_{positive} are standard deviations of fluorescence intensity measured from the negative and positive control wells, respectively, and S_{negative} and S_{positive} are the averaged value of the signal intensities from negative and positive control wells, respectively. For each plate, one or two columns (16-32 wells in 384-well microplates)

are used for positive controls and negative controls, and Z-factor is calculated for every plate during the HTS campaign. The Z-factor ranges between zero and one, with Z = 0 representing the least Z-factor for an assay to be considered in HTS setting while Z = 1 an ideal case with zero standard deviations of the measurements in the assay. An assay is deemed impractical when Z < 0 as it reflects significant overlap between the standard deviations of the positive controls and the negative controls. Typically for HTS a Z > 0.5 is required to characterize compounds as active or inactive. Figure 1.8 shows the simulated data points for various cases of Z ranging between 0 and 1.

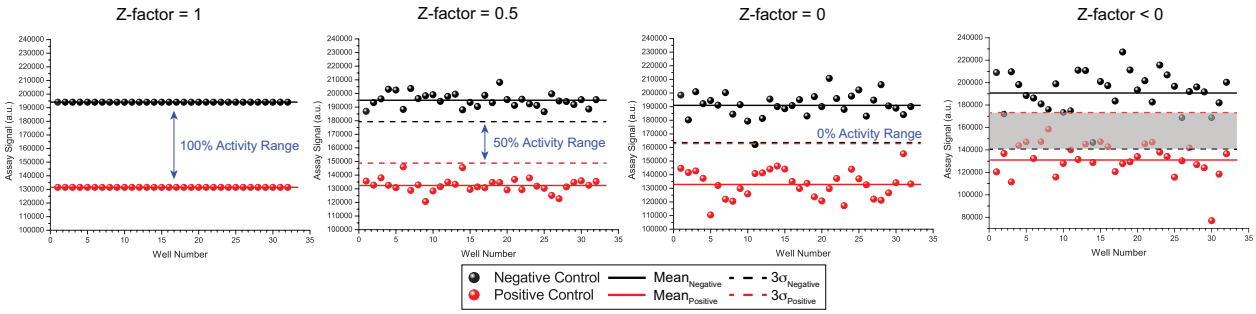

Figure 1.8 Simulated data points for positive and negative controls in HTS. The negative control wells and the positive control wells are shown in black and red, respectively. The solid lines represent the mean of the control wells, and the dotted lines represents the 3σ of the control wells assuming the normal distribution.

1.6 Computational Docking to Target RNA

HTS is expensive, time-consuming and is often challenging when there are no suitable assays for examining binding. Moreover, the cost to set up the instruments for HTS is extremely expensive which is the foremost reason why HTS is rarely performed in academia until recently. As the structure determination of proteins and nucleic acids by X-ray crystallography and NMR is becoming routine, computational docking becomes an economical and viable alternative to HTS. Computational docking

eliminates a need to solubilize the small molecule library into DMSO that sometimes affects the structure and dynamics of biomolecules (171-173). The only constraint for docking however is the availability of high resolution 3D structures. The structures of small molecules are typically available in many community-based databases and from commercial vendors. The computational docking can offer a basis for guiding synthesis of novel RNA-targeting compounds as new small molecules can be screened even before they are synthetically available.

Computational docking is essentially an optimization problem. The docking program first attempts to sample all possible ligand poses in the binding pocket of the receptor, and evaluates the poses of the ligand using a scoring function that describes the various energy terms involved in binding, such as hydrogen bonding, van der Waals interactions, electrostatic interactions, and solvation energy of a ligand, for an optimal ligand pose. Computational docking can be used to screen millions of small molecules to find a novel chemotype that binds to a receptor, namely virtual screening (VS). Computational docking on proteins has been quite successful resulting in discovering novel lead compounds targeting tyrosine phosphatase (174), DNA gyrase (175), and HIV integrase (176, 177) that ultimately led to the development of Raltegravir, the first FDA-approved antiretroviral drug targeting HIV-1 integrase. However, the computational docking against RNA has not been quite as successful as proteins due to various challenges as discussed.

One challenge of docking is that it does not account for the flexibility of a receptor, which is critical in the case of RNAs as they are known to be flexible as discussed in **1.1**. If a receptor is flexible, the binding of small molecules can induce different conformations of the receptor (Figure 1.5), and a single static structure might not be able

to predict the ligand bound poses of small molecules, leading to false positives or negatives. This is particularly an issue when targeting RNA due to dramatic conformational changes that RNA undergoes upon binding of small molecules (21, 65, 125). To address the receptor flexibility problem, several schemes, for example docking against multiple receptor structures, have been developed for proteins (for reviews (178, 179)) as well as for RNA (180-182).

Despite the structural flexibility, virtual screens targeting RNA have been reported showing some success (183-185). Using DOCK (186, 187) and ICM (188), James and co-workers screened 153,000 small molecules against HIV-1 TAR (183). Among the 8 best-scored compounds, 2 compounds yielded IC_{50}s of 1 μM in *in-vitro* scintillation proximity assay. In a subsequent study (184), a slightly larger compound library (181,000 compounds) was screened using the same approach. Testing 43 top-ranked small molecules, 11 compounds were confirmed as hits (hit rate ~ 25%), one of which also exhibited inhibitory activity on Tat-mediated transactivation *in vivo*. Other RNA targets such as human telomerase RNA have also been virtually screened (189).

Recently, Al-Hashimi and co-workers determined the ensemble of 20 HIV-1 TAR conformers from NMR and molecular dynamics (MD) simulations (190, 191) capturing the inherent flexibility of the molecule. They showed that docking against the ensemble of structures from NMR and MD resulted in an improved prediction of small molecule binding affinities in comparison to the predictions from docking against single static structure or NOE derived NMR structures (181). In this study, ~51,000 small molecules were virtually screened against the 20 TAR conformers and 58 compounds from this set were experimentally tested. Among these, 7 compounds were confirmed as hits (hit rate ~12%) including spermine, a previously known binder and 6 new inhibitors. Their

binding and specificity were further validated by NMR chemical shift perturbations, fluorescence binding assays, and cell assays. Among the hits, one compound, netilmicin, was selective to HIV-1 TAR among other RNA therapeutic targets and actively inhibited HIV-1 viral replication in cells.

These successes in virtual screening against RNA showed the great potential of a computational docking as a method to explore the chemical space that favors binding to RNA. However, a question remains as to how does the VS compared with respect to the experimental HTS. Several studies (174, 192) that screened library compounds against protein targets have argued that the HTS and VS are complementary methods, as VS avoids the experimental complications that HTS have and HTS eliminates the issues related with the accuracy of the binding affinity prediction that VS has. One of the objectives of this thesis is to investigate the specific issues arising from screening RNA targets and evaluate the performance and complementarity of HTS and VS methods targeting RNA.

1.7 Characterizing RNA Dynamic Structures as Potential Therapeutic Targets

RNA dynamics play a crucial role in regulating cellular processes and under the effect of cellular cues RNA undergoes various conformational changes that affect various cellular processes. These effective conformations form transiently and are encoded in the dynamics of RNA (193-198). These transient states often occur at μs – ms time scales involving base opening(199) and flipping(197), sugar repuckering(198), and ligand binding(200), and they are vital for biological functions. Thus, characterizing these transient state conformations of RNA will open a new opportunity to use small molecules to target the transient states of RNA.

NMR spectroscopy is a powerful tool not only to determine 3D structures at atomic resolution under solution conditions, but also to characterize the dynamics over timescales ranging between picoseconds (ps) and seconds (s) (194, 201-204). Various NMR parameters, such as spin relaxation, residual dipolar coupling, and chemical shifts, are used to characterize the motions occurring at ps – s time scale. In this study, we use rotating frame R1ρ relaxation dispersion experiment (205-208) to characterize the solution state dynamics of HIV-1 TAR. The R1ρ relaxation dispersion experiment uses a spinlock field to spin lock the magnetization of interest along an effective field along or tilted relative to the x-y plane. One then measures the R1ρ relaxation rates as a function of varying the strength of the spinlock pulse as well as the offset frequency. The limitation of the spinlock pulse strengths determines the detection limit of time scales in R1ρ relaxation dispersion.

The conventional 2D R1ρ relaxation dispersion experiment was limited to spinlock fields > 1000 Hz, which limited the R1ρ relaxation dispersion experiment capable of detecting motions slower than μs time scale. A recent development in pulse sequence design (206) involving selective excitation of single resonances overcomes the limitations posed by the spinlock power and makes it possible to use much lower spinlock powers of ~25 Hz in proteins and up to 100 Hz in nucleic acids (206, 208) alleviating the limitation and extending the detection limit up to milliseconds.

The R1ρ relaxation rate is described mathematically as below,

$$R1\rho = R_1 cos^2\theta + R_2 sin^2\theta + R_{ex} sin^2\theta \qquad \text{(Eq. 2)}$$

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

where R_1 and R_2 are intrinsic longitudinal and transverse relaxation rates, respectively, and R_{ex} is the relaxation rate due to chemical exchange. θ is the effective tilt angle, where ω_1 is the spinlock power and Ω is the offset of the spin from the spinlock frequency. The chemical exchange increases the overall R1ρ relaxation rate, and one can observe the chemical exchange term by changing the spinlock power and the offset.

The relaxation rate due to chemical exchange (R_{ex}) can be mathematically described by Bloch-McConnell equation (209). Because Bloch-McConnell equation is 6x6 matrix form, which challenges to fit the experimental data, much simpler approximations to the solution of Bloch-McConnell equation are used (reviewed in (205)). One of the important assumptions in the simplified equations is that the intrinsic R_1 and R_2 are separable from R_{ex}, which is valid under most experimental conditions where $\tau_c k_{ex} \ll 1$, $|R_{1a} - R_{1b}| \ll k_{ex}$, $|R_{2a} - R_{2b}| \ll k_{ex}$, and $R_2 - R_1 \ll k_{ex}$. The term τ_c is the correlation time of overall rotational diffusion of a molecule, k_{ex} is the exchange rate constant, and R_{1a}, R_{2a}, R_{1b}, and R_{2b} are the individual relaxation rates (R_1 and R_2) of two species undergoing chemical exchange in the simplest 2-state model of chemical exchange,

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} B$$

where k_1 is forward rate constant and k_{-1} is backward rate constant of the transition between state A and state B. The sum of k_1 and k_{-1} yields the exchange rate constant (k_{ex}).

The most accurate approximation to the solution of Bloch-McConnell equation uses Laguerre's method for polynomial root finding (209). In the simple 2-state model, the Laguerre's approximation of the relaxation rate due to chemical change is,

$$R_{ex} = \frac{p_a p_b \Delta\omega^2 k_{ex}}{\omega_a^2\omega_b^2 \Big/ \omega_e^2 + k_{ex}^2 - sin^2\theta p_a p_b \Delta\omega^2 (1 + \frac{2k_{ex}^2(p_a\omega_a^2 + p_b\omega_b^2)}{\omega_a^2\omega_b^2 + \omega_e^2 k_{ex}^2})}$$ (eq. 3)

where p_a and p_b are populations of state A and state B, $\Delta\omega^2$ is the chemical shift difference between state A and B, ω_a and ω_b are the effective magnetization of state A and state B, and ω_e is the effective magnetization of the offset between the state A and B. By fitting this equation, the chemical shift difference ($\Delta\omega$), the population of the transient state (p_b) and the exchange rate constant (k_{ex}) can be extracted. There are other exchange models with more than 2-states(205), and F-test and/or Akaike test can be used to compare and select the appropriate model that best fits to the experimental data.

1.8 Objectives of this Dissertation

The objective of this dissertation is to explore experimental and computational methods to discover small molecule modulators targeting HIV-1 TAR. First, the effect of DMSO, a universal solvent used in HTS, is evaluated against RNA targets (Chapter 2). The screening results from HTS and ensemble-based VS targeting HIV-1 TAR and their performances are evaluated (Chapter 3). The transient states of HIV-1 TAR are characterized using NMR R1ρ relaxation dispersion experiment (Chapter 4). These transient states are biologically significant in regulatory functions of RNA and can be potentially targeted by small molecules.

1.9 References

1. Hoagland MB, Stephenson ML, Scott JF, Hecht LI, & Zamecnik PC (1958) A Soluble Ribonucleic Acid Intermediate in Protein Synthesis. *J. Biol. Chem.* 231(1):241-257.
2. Holley RW, Everett GA, Madison JT, & Zamir A (1965) Nucleotide Sequences in the Yeast Alanine Transfer Ribonucleic Acid. *J. Biol. Chem.* 240(5):2122-2128.

3. Holley RW, *et al.* (1965) Structure of a Ribonucleic Acid. *Science* 147(3664):1462-1465.
4. Decatur WA & Fournier MJ (2003) RNA-guided Nucleotide Modification of Ribosomal and Other RNAs. *J. Biol. Chem.* 278(2):695-698.
5. Wiedenheft B, Sternberg SH, & Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482:331-338.
6. Dong H, *et al.* (2013) MicroRNA: Function, Detection, and Bioanalysis. *Chem. Rev.* 113(8):6207-6233.
7. Serganov A & Patel DJ (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* 8:776-790.
8. Zhou Q, Li T, & Price DH (2012) RNA Polymerase II Elongation Control. *Annu. Rev. Biochem.* 81(1):119-143.
9. Rinn JL, *et al.* (2007) Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* 129(7):1311-1323.
10. Gupta RA, *et al.* (2010) Long noncoding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464(7291):1071-1076.
11. Hannon GJ, Rivas FV, Murchison EP, & Steitz JA (2006) The Expanding Universe of Noncoding RNAs. *Cold Spring Harb. Symp. Quant. Biol.* 71:551-564.
12. Bompfunewere AF, *et al.* (2005) Evolutionary patterns of non-coding RNAs. *Theory Biosci.* 123(4):301-369.
13. Dethoff EA, Chugh J, Mustoe AM, & Al-Hashimi HM (2012) Functional complexity and regulation through RNA dynamics. *Nature* 482:322-330.
14. Boehr DD, Nussinov R, & Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5(11):789-796.
15. Al-Hashimi HM & Walter NG (2008) RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* 18(3):321-329.
16. Cruz JA & Westhof E (2009) The Dynamic Landscapes of RNA Architecture. *Cell* 136(4):604-609.
17. Haller A, Souli√®re MF, & Micura R (2011) The Dynamic Nature of RNA as Key to Understanding Riboswitch Mechanisms. *Acc. Chem. Res.* 44(12):1339-1348.
18. Winkler W, Nahvi A, & Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419:952-956.
19. Cromie MJ, Shi Y, Latifi T, & Groisman EA (2006) An RNA sensor for intracellular Mg2+. *Cell* 125:71-84.
20. Nechooshtan G, Elgrably-Weiss M, Sheaffer A, Westhof E, & Altuvia S (2009) A pH-responsive riboregulator. *Genes Dev.* 23:2650-2662.

21. Puglisi JD, Tan R, Calnan BJ, Frankel AD, & Williamson JR (1992) Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy. *Science* 257(5066):76-80.
22. Nocker A (2001) A mRNA-based thermosensor controls expression of rhizobial heat shock genes. *Nucleic Acids Res.* 29:4800-4807.
23. Pyle AM & Green JB (1995) RNA folding. *Curr. Opin. Struct. Biol.* 5:303-310.
24. Kim JN & Breaker RR (2008) Purine sensing by riboswitches. *Biol. Cell* 100(1):1-11.
25. Montange RK & Batey RT (2008) Riboswitches: Emerging Themes in RNA Structure and Function. *Annu. Rev. Biophys.* 37(1):117-133.
26. Roth A & Breaker RR (2009) The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annu. Rev. Biochem.* 78(1):305-334.
27. Giege R (2008) Toward a more complete view of tRNA biology. *Nat Struct Mol Biol* 15(10):1007-1014.
28. Mulder AM (2010) Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science* 330:673-677.
29. Patel DJ (1999) Adaptive recognition in RNA complexes with peptides and protein modules. *Curr. Opin. Struct. Biol.* 9(1):74-87.
30. Hermann T & Patel DJ (2000) Adaptive Recognition by Nucleic Acid Aptamers. *Science* 287(5454):820-825.
31. Blount KF & Breaker RR (2006) Riboswitches as antibacterial drug targets. *Nat. Biotechnol.* 24:1558-1564.
32. Egea PF, Stroud RM, & Walter P (2005) Targeting proteins to membranes: structure of the signal recognition particle. *Curr. Opin. Struct. Biol.* 15(2):213-220.
33. Stone MD, *et al.* (2007) Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* 446(7134):458-461.
34. Bannwarth S & Gatignol A (2005) HIV-1 TAR RNA: the target of molecular interactions between the virus and its host. *Curr. HIV Res.* 3(1):61-71.
35. Richter S, Cao H, & Rana TM (2002) Specific HIV-1 TAR RNA Loop Sequence and Functional Groups Are Required for Human Cyclin T1-Tat-TAR Ternary Complex Formation. *Biochemistry* 41(20):6391-6397.
36. Laspia MF, Rice AP, & Mathews MB (1989) HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation. *Cell* 59(2):283-292.
37. Held WA, Ballou B, Mizushima S, & Nomura M (1974) Assembly Mapping of 30 S Ribosomal Proteins from Escherichia coli : FURTHER STUDIES. *J. Biol. Chem.* 249(10):3103-3111.
38. Agalarov SC, *et al.* (2000) Structure of the S15,S6,S18-rRNA Complex: Assembly of the 30S Ribosome Central Domain. *Science* 288(5463):107-112.

39. Ciccodicola A, Ambrosio M, Scarpato M, & Costa V (2012) Non-coding RNA in Neurodegeneration. (Translated from English) *Curr. Transl. Geriatr. Exp. Gerontol. Rep.* 1(4):219-228 (in English).
40. Sassen S, Miska E, & Caldas C (2008) MicroRNA-implications for cancer. (Translated from English) *Virchows Arch* 452(1):1-10 (in English).
41. Cooper TA, Wan L, & Dreyfuss G (2009) RNA and Disease. *Cell* 136(4):777-793.
42. Stevens M, De Clercq E, & Balzarini J (2006) The regulation of HIV-1 transcription: Molecular targets for chemotherapeutic intervention. *Med. Res. Rev.* 26(5):595-625.
43. Guan L & Disney MD (2012) Recent Advances in Developing Small Molecules Targeting RNA. *ACS Chem. Biol.* 7(1):73-86.
44. Thomas JR & Hergenrother PJ (2008) Targeting RNA with Small Molecules. *Chem. Rev.* 108(4):1171-1224.
45. Blount KF, Wang JX, Lim J, Sudarsan N, & Breaker RR (2007) Antibacterial compounds that target lysine riboswitches. *Nat. Chem. Biol.* 3(1):44-49.
46. Vannini I, Fanini F, & Fabbri M (2013) MicroRNAs as lung cancer biomarkers and key players in lung carcinogenesis. *Clin. Biochem.* 46(10-11):918-925.
47. Jiang Q*, et al.* (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 37(suppl 1):D98-D104.
48. Hermann T (2005) Drugs targeting the ribosome. *Curr. Opin. Struct. Biol.* 15(3):355-366.
49. Ennifar E*, et al.* (2006) Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell. *Nucleic Acids Res.* 34(8):2328-2339.
50. Hamasaki K, Woo M-C, & Ueno A (2000) An aminoglycoside antibiotic, neamine, and its aromatic ring-substituted derivatives as potential inhibitors for HIV-1 RRE-Rev. *Tetrahedron Lett.* 41(43):8327-8332.
51. Fourmy D, Recht MI, Blanchard SC, & Puglisi JD (1996) Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* 274(5291):1367-1371.
52. Tor Y (2006) The ribosomal A-site as an inspiration for the design of RNA binders. *Biochimie* 88(8):1045-1051.
53. Vicens Q & Westhof E (2003) RNA as a Drug Target: The Case of Aminoglycosides. *ChemBioChem* 4(10):1018-1023.
54. Brodersen DE*, et al.* (2000) The Structural Basis for the Action of the Antibiotics Tetracycline, Pactamycin, and Hygromycin B on the 30S Ribosomal Subunit. *Cell* 103(7):1143-1154.

55. Brisson-Noel A, Trieu-Cuot P, & Courvalin P (1988) Mechanism of action of spiramycin and other macrolides. *J. Antimicrob. Chemother.* Suppl B:13-23.
56. Bozdogan B & Appelbaum PC (2004) Oxazolidinones: activity, mode of action, and mechanism of resistance. *Int J. Antimicrob. Agents* 23(2):113-119.
57. Ban N, Nissen P, Hansen J, Moore PB, & Steitz TA (2000) The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 angstrom Resolution. *Science* 289(5481):905-920.
58. Schluenzen F, *et al.* (2000) Structure of Functionally Activated Small Ribosomal Subunit at 3.3 angstrom Resolution. *Cell* 102(5):615-623.
59. Wimberly BT, *et al.* (2000) Structure of the 30S ribosomal subunit. *Nature* 407(6802):327-339.
60. Murray JB, *et al.* (2006) Interactions of Designer Antibiotics and the Bacterial Ribosomal Aminoacyl-tRNA Site. *Chem. Biol.* 13(2):129-138.
61. Borovinskaya MA, *et al.* (2007) Structural basis for aminoglycoside inhibition of bacterial ribosome recycling. *Nat. Struct. Mol. Biol.* 14(8):727-732.
62. Borovinskaya MA, Shoji S, Holton JM, Fredrick K, & Cate JHD (2007) A Steric Block in Translation Caused by the Antibiotic Spectinomycin. *ACS Chem. Biol.* 2(8):545-552.
63. Yoshizawa S, Fourmy D, & Puglisi JD (1999) Recognition of the Codon-Anticodon Helix by Ribosomal RNA. *Science* 285(5434):1722-1725.
64. Schmeing TM & Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461:1234-1242.
65. Vicens Q & Westhof E (2001) Crystal Structure of Paromomycin Docked into the Eubacterial Ribosomal Decoding A Site. *Structure* 9(8):647-658.
66. Lynch SR, Gonzalez Jr RL, & Puglisi JD (2003) Comparison of X-Ray Crystal Structure of the 30S Subunit-Antibiotic Complex with NMR Structure of Decoding Site Oligonucleotide-Paromomycin Complex. *Structure* 11(1):43-53.
67. Fourmy D, Yoshizawa S, & Puglisi JD (1998) Paromomycin binding induces a local conformational change in the A-site of 16 s rRNA. *J. Mol. Biol.* 277(2):333-345.
68. Hofstadler SA, *et al.* (1999) Multiplexed Screening of Neutral Mass-Tagged RNA Targets against Ligand Libraries with Electrospray Ionization FTICR MS: A Paradigm for High-Throughput Affinity Screening. *Anal. Chem.* 71(16):3436-3440.
69. Disney MD & Seeberger PH (2004) Aminoglycoside Microarrays To Explore Interactions of Antibiotics with RNAs and Proteins. *Chemistry* 10(13):3308-3314.
70. Deigan KE & Ferre-D'Amare AR (2011) Riboswitches: Discovery of Drugs That Target Bacterial Gene-Regulatory RNAs. *Acc. Chem. Res.* 44(12):1329-1338.

71. Robbins WJ (1941) The pyridine analog of thiamin and the growth of fungi. *Proc. Natl. Acad. Sci. U. S. A.* 27(9):419-422.
72. Woolley DW & White AGC (1943) Selective Reversible Inhibition of Microbial Growth with Pyrithiamine. *J. Exp. Med.* 78(6):489-497.
73. Iwashima A, Wakabayashi Y, & Nose Y (1976) Formation of Pyrithiamine Pyrophosphate in Brain Tissue. *J. Biochem.* 79(4):845-847.
74. Sudarsan N, Cohen-Chalamish S, Nakamura S, Emilsson GM, & Breaker RR (2005) Thiamine Pyrophosphate Riboswitches Are Targets for the Antimicrobial Compound Pyrithiamine. *Chem. Biol.* 12(12):1325-1335.
75. Thore Sp, Leibundgut M, & Ban N (2006) Structure of the Eukaryotic Thiamine Pyrophosphate Riboswitch with Its Regulatory Ligand. *Science* 312(5777):1208-1211.
76. Shiota T, Folk JE, & Tietze F (1958) Inhibition of lysine utilization in bacteria by S-(beta-aminoethyl)cysteine and its reversal by lysine peptides. *Arch. Biochem. Biophys.* 77(2):372-377.
77. McCord TJ, Ravel JM, Skinner CG, & Shive W (1957) DL-4-Oxalysine, an Inhibitory Analog of Lysine. *J. Am. Chem. Soc.* 79(21):5693-5696.
78. Matsui K, *et al.* (1982) Riboflavin Production by Roseoflavin-resistant Strains of Some Bacteria. *Agric. Biol. Chem.* 46(8):2003-2008.
79. Lee ER, Blount KF, & Breaker RR (2009) Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol.* 6(2):187-194.
80. Winkler WC, Nahvi A, Roth A, Collins JA, & Breaker RR (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428(6980):281-286.
81. Mayer G & Famulok M (2006) High-Throughput-Compatible Assay for glmS Riboswitch Metabolite Dependence. *ChemBioChem* 7(4):602-604.
82. Blount K, Puskarz I, Penchovsky R, & Breaker R (2006) Development and Application of a High-Throughput Assay for glmS Riboswitch Activators. *RNA Biol.* 3(2):77-81.
83. Timchenko LT, *et al.* (1996) Identification of a (CUG)n Triplet Repeat RNA-Binding Protein and Its Expression in Myotonic Dystrophy. *Nucleic Acids Res* 24(22):4407-4414.
84. Wang J, *et al.* (1995) Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation. *Hum. Mol. Genet.* 4(4):599-606.
85. O'Rourke JR & Swanson MS (2009) Mechanisms of RNA-mediated Disease. *J. Biol. Chem.* 284(12):7419-7423.

86. Pascual M, Vicente M, Monferrer L, & Artero R (2006) The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation* 74(2-3):65-80.
87. Disney MD, Lee MM, Pushechnikov A, & Childs-Disney JL (2010) The Role of Flexibility in the Rational Design of Modularly Assembled Ligands Targeting the RNAs that Cause the Myotonic Dystrophies. *ChemBioChem* 11(3):375-382.
88. Parkesh R, *et al.* (2012) Design of a Bioactive Small Molecule That Targets the Myotonic Dystrophy Type 1 RNA via an RNA Motif-Ligand Database and Chemical Similarity Searching. *J. Am. Chem. Soc.* 134(10):4731-4742.
89. Childs-Disney JL, Hoskins J, Rzuczek SG, Thornton CA, & Disney MD (2012) Rationally Designed Small Molecules Targeting the RNA That Causes Myotonic Dystrophy Type 1 Are Potently Bioactive. *ACS Chem. Biol.* 7(5):856-862.
90. Kitov PI & Bundle DR (2003) On the Nature of the Multivalency Effect: A Thermodynamic Model. *J. Am. Chem. Soc.* 125(52):16271-16284.
91. Embretson J, *et al.* (1993) Massive covert infection of helper T lymphocytes and macrophages by HIV during the incubation period of AIDS. *Nature* 362(6418):359-362.
92. Wei X, *et al.* (1995) Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 373:117-122.
93. Bieniasz PD, Grdina TA, Bogerd HP, & Cullen BR (1999) Recruitment of cyclin T1/P-TEFb to an HIV type 1 long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. *Proc. Natl. Acad. Sci. U.S.A.* 96(14):7791-7796.
94. Roy S, Delling U, Chen CH, Rosen CA, & Sonenberg N (1990) A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev.* 4:1365-1373.
95. Selby MJ, Bain ES, Luciw PA, & Peterlin BM (1989) Structure, sequence, and position of the stem-loop in tar determine transcriptional elongation by tat through the HIV-1 long terminal repeat. *Genes Dev.* 3:547-558.
96. Benkirane M, *et al.* (1997) Oncogenic potential of TAR RNA binding protein TRBP and its regulatory interaction with RNA-dependent protein kinase PKR. *EMBO J.* 16(3):611-624.
97. Carpick BW, *et al.* (1997) Characterization of the Solution Complex between the Interferon-induced, Double-stranded RNA-activated Protein Kinase and HIV-I Trans-activating Region RNA. *J. Biol. Chem.* 272(14):9510-9516.

98. Jalalirad M, Saadatmand J, & Laughrea M (2012) Dominant Role of the 5' TAR Bulge in Dimerization of HIV-1 Genomic RNA, but No Evidence of TAR-TAR Kissing during in Vivo Virus Assembly. *Biochemistry* 51(18):3744-3758.
99. Das A, Vrolijk M, Harwig A, & Berkhout B (2012) Opening of the TAR hairpin in the HIV-1 genome causes aberrant RNA dimerization and packaging. *Retrovirology* 9(1):59.
100. Helga-Maria C, Hammarskjöld M-L, & Rekosh D (1999) An Intact TAR Element and Cytoplasmic Localization Are Necessary for Efficient Packaging of Human Immunodeficiency Virus Type 1 Genomic RNA. *J. Virol.* 73(5):4127-4135.
101. Klase Z, *et al.* (2007) HIV-1 TAR element is processed by Dicer to yield a viral micro-RNA involved in chromatin remodeling of the viral LTR. *BMC Mol. Biol.* 8(1):63.
102. Wagschal A, *et al.* (2012) Microprocessor, Setx, Xrn2, and Rrp6 Co-operate to Induce Premature Termination of Transcription by RNAPII. *Cell* 150(6):1147-1157.
103. Watts JM, *et al.* (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460(7256):711-716.
104. Zhu Y, *et al.* (1997) Transcription elongation factor P-TEFb is required for HIV-1 Tat transactivation in vitro. *Genes Dev.* 11(20):2622-2632.
105. Majello B, Napolitano G, Giordano A, & Lania L (1999) Transcriptional regulation by targeted recruitment of cyclin-dependent CDK9 kinase in vivo. *Oncogene* 18(32):4598-4605.
106. Tahirov TH, *et al.* (2010) Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature* 465(7299):747-751.
107. Okamoto H, *et al.* (2000) Inhibition of the RNA-Dependent Transactivation and Replication of Human Immunodeficiency Virus Type 1 by a Fluoroquinoline Derivative K-37. *Virology* 272(2):402-408.
108. Richter S, Parolin C, Palumbo M, & Palu G (2004) Antiviral properties of quinolone-based drugs. *Curr. Drug Targets Infect. Disord.* 4(2):111-116.
109. Parolin C, *et al.* (2003) New Anti-Human Immunodeficiency Virus Type 1 6-Aminoquinolones: Mechanism of Action. *Antimicrob. Agents Chemother.* 47(3):889-896.
110. Cecchetti V, *et al.* (2000) 6-Aminoquinolones as New Potential Anti-HIV Agents. *J. Med. Chem.* 43(20):3799-3802.
111. Zhao H, Li J, Xi F, & Jiang L (2004) Polyamidoamine dendrimers inhibit binding of Tat peptide to TAR RNA. *FEBS Lett.* 563(1-3):241-245.

112. Zhao H, Dai D, Li J, Chen Y, & Jiang L (2003) Quantitative study of HIV-1 Tat peptide and TAR RNA interaction inhibited by poly(allylamine hydrochloride). *Biochem. Biophys. Res. Commun.* 312(2):351-354.
113. Litovchick A, Evdokimov AG, & Lapidot A (2000) Aminoglycoside-Arginine Conjugates That Bind TAR RNA: Synthesis, Characterization, and Antiviral Activity. *Biochemistry* 39(11):2838-2852.
114. Litovchick A, Lapidot A, Eisenstein M, Kalinkovich A, & Borkow G (2001) Neomycin B-Arginine Conjugate, a Novel HIV-1 Tat Antagonist: Synthesis and Anti-HIV Activities. *Biochemistry* 40(51):15612-15623.
115. Hamy F, *et al.* (1997) An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. *Proc. Natl Acad. Sci. USA* 94(8):3548-3553.
116. Kesavan V, Tamilarasu N, Cao H, & Rana TM (2002) A New Class of RNA-Binding Oligomers:, Peptoid Amide and Ester Analogues. *Bioconjug. Chem.* 13(6):1171-1175.
117. Yu X, Lin W, Pang R, & Yang M (2005) Design, synthesis and bioactivities of TAR RNA targeting beta-carboline derivatives based on Tat-TAR interaction. *Eur. J. Med. Chem.* 40(9):831-839.
118. Chang J, *et al.* (2004) miR-122, a Mammalian Liver-Specific microRNA, is Processed from hcr mRNA and MayDownregulate the High Affinity Cationic Amino Acid Transporter CAT-1. *RNA Biol.* 1(2):106-113.
119. Gramantieri L, *et al.* (2007) Cyclin G1 Is a Target of miR-122a, a MicroRNA Frequently Down-regulated in Human Hepatocellular Carcinoma. *Cancer Res.* 67(13):6092-6099.
120. Young DD, Connelly CM, Grohmann C, & Deiters A (2010) Small Molecule Modifiers of MicroRNA miR-122 Function for the Treatment of Hepatitis C Virus Infection and Hepatocellular Carcinoma. *J. Am. Chem. Soc.* 132(23):7976-7981.
121. Gumireddy K, *et al.* (2008) Small-Molecule Inhibitors of MicroRNA miR-21 Function. *Angew. Chem. Int. Ed.* 47(39):7482-7484.
122. Velagapudi SP, Gallo SM, & Disney MD (2014) Sequence-based design of bioactive small molecules that target precursor microRNAs. *Nat. Chem. Biol.* 10:291-297.
123. François B, *et al.* (2005) Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucleic Acids Res.* 33(17):5677-5690.

124. Hermann T & Westhof E (1998) Aminoglycoside binding to the hammerhead ribozyme: a general model for the interaction of cationic antibiotics with RNA. *J. Mol. Biol.* 276(5):903-912.
125. Casiano-Negroni A, Sun X, & Al-Hashimi HM (2007) 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(22):6525-6535.
126. Zhang Q, Stelzer AC, Fisher CK, & Al-Hashimi HM (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* 450(7173):1263-1267.
127. Al-Hashimi HM, *et al.* (2002) 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(2):95-102.
128. Puglisi JD, Chen L, Frankel AD, & Williamson JR (1993) Role of RNA structure in arginine recognition of TAR RNA. *Proc. Natl. Acad. Sci. U. S. A.* 90:3680-3684.
129. Pitt SW, Majumdar A, Serganov A, Patel DJ, & Al-Hashimi HM (2004) Argininamide Binding Arrests Global Motions in HIV-1 TAR RNA: Comparison with Mg2+-induced Conformational Stabilization. *J. Mol. Biol.* 338(1):7-16.
130. Fourmy D, Recht MI, & Puglisi JD (1998) Binding of neomycin-class aminoglycoside antibiotics to the A-site of 16 s rRNA. *J. Mol. Biol.* 277(2):347-362.
131. Tor Y, Hermann T, & Westhof E (1998) Deciphering RNA recognition: aminoglycoside binding to the hammerhead ribozyme. *Chem. Biol.* 5(11):R277-R283.
132. Hermann T & Westhof E (1998) Aminoglycoside binding to the hammerhead ribozyme: a general model for the interaction of cationic antibiotics with RNA. *Journal of Molecular Biology* 276(5):903-912.
133. Wang H & Tor Y (1997) Electrostatic Interactions in RNA Aminoglycosides Binding. *J. Am. Chem. Soc.* 119:8734-8735.
134. Mandal M & Breaker RR (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* 11(1):29-35.
135. Serganov A, *et al.* (2004) Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. *Chem. Biol.* 11(12):1729-1741.
136. Batey RT, Gilbert SD, & Montange RK (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* 432(7015):411-415.

137. Gilbert SD, Stoddard CD, Wise SJ, & Batey RT (2006) Thermodynamic and Kinetic Characterization of Ligand Binding to the Purine Riboswitch Aptamer Domain. *J. Mol. Biol.* 359(3):754-768.
138. Gilbert SD, Mediatore SJ, & Batey RT (2006) Modified Pyrimidines Specifically Bind the Purine Riboswitch. *J. Am. Chem. Soc.* 128(44):14214-14215.
139. Zheng S, Chen Y, Donahue CP, Wolfe MS, & Varani G (2009) Structural Basis for Stabilization of the Tau Pre-mRNA Splicing Regulatory Element by Novantrone (Mitoxantrone). *Chem. Biol.* 16(5):557-566.
140. Zimm BH & Kallenbach NR (1962) Selected Aspects of the Physical Chemistry of Polynucleotides and Nucleic Acids. *Annu. Rev. Phys. Chem.* 13(1):171-194.
141. Helmkamp GK & Ts'o POP (1961) The Secondary Structures of Nucleic Acids in Organic Solvents. *J. Am. Chem. Soc.* 83:138-142.
142. Walter F, Putz J, Giege R, & Westhof E (2002) Binding of tobramycin leads to conformational changes in yeast tRNAAsp and inhibition of aminoacylation. *EMBO J.* 21(4):760-768.
143. Hoch I, Berens C, Westhof E, & Schroeder R (1998) Antibiotic inhibition of RNA catalysis: neomycin B binds to the catalytic core of the td group I intron displacing essential metal ions. *Journal of Molecular Biology* 282(3):557-569.
144. Zhang Y, Li Z, Pilch DS, & Leibowitz MJ (2002) Pentamidine inhibits catalytic activity of group I intron Ca.LSU by altering RNA folding. *Nucleic Acids Res* 30(13):2961-2971.
145. Smith AL, Kassman J, Srour KJ, & Soto AM (2011) Effect of Salt Concentration on the Conformation of TAR RNA and Its Association with Aminoglycoside Antibiotics. *Biochemistry* 50(44):9434-9445.
146. Weeks KM & Crothers DM (1993) Major Groove Accessibility of RNA. *Science* 261:1574-1577.
147. Hermann T & Patel DJ (1999) Stitching together RNA tertiary architectures. *J. Mol. Biol.* 294(4):829-849.
148. Chin K, Sharp KA, Honig B, & Pyle AM (1999) Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nat. Struct. Biol.* 6(11):1055-1061.
149. Davis B, *et al.* (2004) Rational Design of Inhibitors of HIV-1 TAR RNA through the Stabilisation of Electrostatic "Hot Spots". *J. Mol. Biol.* 336(2):343-356.
150. Rachofsky EL, Osman R, & Ross JBA (2001) Probing Structure and Dynamics of DNA with 2-Aminopurine: Effects of Local Environment on Fluorescence. *Biochemistry* 40(4):946-956.

151. Jean JM & Hall KB (2001) 2-Aminopurine fluorescence quenching and lifetimes: Role of base stacking. *Proc. Natl. Acad. Sci. U. S. A.* 98(1):37-41.
152. Stivers JT (1998) 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. *Nucleic Acids Res.* 26(16):3837-3844.
153. Bradrick TD & Marino JP (2004) Ligand-induced changes in 2-aminopurine fluorescence as a probe for small molecule binding to HIV-1 TAR RNA. *RNA* 10(9):1459-1468.
154. Aboul-ela F, Karn J, & Varani G (1995) The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *J. Mol. Biol.* 253(2):313-332.
155. Kaul M, Barbieri CM, & Pilch DS (2004) Fluorescence-Based Approach for Detecting and Characterizing Antibiotic-Induced Conformational Changes in Ribosomal RNA: Comparing Aminoglycoside Binding to Prokaryotic and Eukaryotic Ribosomal RNA Sequences. *J. Am. Chem. Soc.* 126(11):3447-3453.
156. Wilhelmsson LM (2010) Fluorescent nucleic acid base analogues. *Q. Rev. Biophys.* 43(02):159-183.
157. Griffey RH, Hofstadler SA, Sannes-Lowery KA, Ecker DJ, & Crooke ST (1999) Determinants of aminoglycoside-binding specificity for rRNA by using mass spectrometry. *Proc. Natl Acad. Sci. USA* 96(18):10129-10133.
158. Suryawanshi H, Sabharwal H, & Maiti S (2010) Thermodynamics of Peptide-RNA Recognition: The Binding of a Tat Peptide to TAR RNA. *J. Phys. Chem. B* 114(34):11155-11163.
159. Latham MP, Brown DJ, McCallum SA, & Pardi A (2005) NMR Methods for Studying the Structure and Dynamics of RNA. *ChemBioChem* 6(9):1492-1505.
160. Hendrix M, Priestley ES, Joyce GF, & Wong C-H (1997) Direct Observation of Aminoglycoside-RNA Interactions by Surface Plasmon Resonance. *J. Am. Chem. Soc.* 119(16):3641-3648.
161. Pereira DA & Williams JA (2007) Origin and evolution of high throughput screening. *Br. J. Pharmacol.* 152(1):53-61.
162. Inglese J, *et al.* (2006) Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc. Natl Acad. Sci. USA* 103(31):11473-11478.
163. Zhang J-H, Chung TDY, & Oldenburg KR (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen* 4(2):67-73.

164. Mei H-Y, *et al.* (1996) Inhibition of Self-Splicing Group I Intron RNA: High-Throughput Screening Assays. *Nucleic Acids Res* 24(24):5051-5053.
165. Mei H-Y, *et al.* (1997) Discovery of selective, small-molecule inhibitors of RNA complexes, 1. The tat protein/TAR RNA complexes required for HIV-1 transcription. *Bioorg. Med. Chem.* 5(6):1173-1184.
166. Mei H-Y, Cui M, Lemrow SM, & Czarnik AW (1997) Discovery of selective, small-molecule inhibitors of RNA complexes, II. Self-splicing group I intron ribozyme. *Bioorg. Med. Chem.* 5(6):1185-1195.
167. Fox S, *et al.* (2006) High-Throughput Screening: Update on Practices and Success. *J. Biomol. Screen* 11(7):864-869.
168. Gooding KB, *et al.* (2004) High throughput screening of library compounds against an oligonucleotide substructure of an RNA target. *J. Am. Soc. Mass. Spectrom.* 15(6):884-892.
169. Hwang S, *et al.* (2003) Discovery of a Small Molecule Tat-trans-Activation-responsive RNA Antagonist That Potently Inhibits Human Immunodeficiency Virus-1 Replication. *J. Biol. Chem.* 278(40):39092-39103.
170. Malo N, Hanley JA, Cerquozzi S, Pelletier J, & Nadon R (2006) Statistical practice in high-throughput screening data analysis. *Nat. Biotechnol.* 24(2):167-175.
171. Tjernberg A, Markova N, Griffiths WJ, & Hallén D (2006) DMSO-Related Effects in Protein Characterization. *J. Biomol. Screening* 11(2):131-137.
172. Cubrilovic D & Zenobi R (2013) Influence of Dimehylsulfoxide on Protein-Ligand Binding Affinities. *Anal. Chem.* 85(5):2724-2730.
173. Lee J, Vogt CE, McBrairty M, & Al-Hashimi HM (2013) Influence of Dimethylsulfoxide on RNA Structure and Ligand Binding. *Anal. Chem.* 85(20):9692-9698.
174. Doman TN, *et al.* (2002) Molecular Docking and High-Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase-1B. *J. Med. Chem.* 45(11):2213-2221.
175. Boehm H-J, *et al.* (2000) Novel Inhibitors of DNA Gyrase: 3D Structure Based Biased Needle Screening, Hit Validation by Biophysical Methods, and 3D Guided Optimization. A Promising Alternative to Random Screening. *J. Med. Chem.* 43(14):2664-2674.
176. Schames JR, *et al.* (2004) Discovery of a Novel Binding Trench in HIV Integrase. *J. Med. Chem.* 47(8):1879-1881.
177. Sotriffer CA, Ni H, & McCammon JA (2000) Active Site Binding Modes of HIV-1 Integrase Inhibitors. *J. Med. Chem.* 43(22):4109-4117.

178. Totrov M & Abagyan R (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* 18(2):178-184.
179. Teodoro ML & Kavraki LE (2003) Conformational Flexibility Models for the Receptor in Structure Based Drug Design. *Curr. Pharm. Des.* 9(20):1635-1648.
180. Guilbert C & James TL (2008) Docking to RNA via Root-Mean-Square-Deviation-Driven Energy Minimization with Flexible Ligands and Flexible Targets. *J. Chem. Inf. Model.* 48(6):1257-1268.
181. Stelzer AC, *et al.* (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* 7(8):553-559.
182. Kruger DM, Bergs J, Kazemi S, & Gohlke H (2011) Target Flexibility in RNA-Ligand Docking Modeled by Elastic Potential Grids. *ACS Med. Chem. Lett.* 2(7):489-493.
183. Filikov A, *et al.* (2000) Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. (Translated from English) *J. Comput. Aided. Mol. Des.* 14(6):593-610 (in English).
184. Lind KE, Du Z, Fujinaga K, Peterlin BM, & James TL (2002) Structure-Based Computational Database Screening, In Vitro Assay, and NMR Assessment of Compounds that Target TAR RNA. *Chem. Biol.* 9(2):185-193.
185. Du Z, Lind KE, & James TL (2002) Structure of TAR RNA Complexed with a Tat-TAR Interaction Nanomolar Inhibitor that Was Identified by Computational Screening. *Chem. Biol.* 9(6):707-712.
186. Meng EC, Shoichet BK, & Kuntz ID (1992) Automated docking with grid-based energy evaluation. *J. Comput. Chem.* 13(4):505-524.
187. Shoichet BK, Kuntz ID, & Bodian DL (1992) Molecular docking using shape descriptors. *J. Comput. Chem.* 13(3):380-397.
188. Abagyan R, Totrov M, & Kuznetsov D (1994) ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* 15(5):488-506.
189. Gomez Pinto I, Guilbert C, Ulyanov NB, Stearns J, & James TL (2008) Discovery of Ligands for a Novel Target, the Human Telomerase RNA, Based on Flexible-Target Virtual Screening and NMR. *J. Med. Chem.* 51(22):7205-7215.
190. Frank AT, Stelzer AC, Al-Hashimi HM, & Andricioaei I (2009) 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(11):3670-3679.

191. Stelzer AC, Frank AT, Bailor MH, Andricioaei I, & Al-Hashimi HM (2009) Constructing atomic-resolution RNA structural ensembles using MD and motionally decoupled NMR RDCs. *Methods* 49(2):167-173.
192. Babaoglu K, *et al.* (2008) Comprehensive Mechanistic Analysis of Hits from High-Throughput and Docking Screens against beta-Lactamase. *J. Med. Chem.* 51(8):2502-2511.
193. Dethoff EA, Petzold K, Chugh J, Casiano-Negroni A, & Al-Hashimi HM (2012) Visualizing transient low-populated structures of RNA. *Nature* 491(7426):724-728.
194. Sekhar A & Kay LE (2013) NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci. U.S.A.* 110(32):12867-12874.
195. Palmer Iii AG (2014) Chemical exchange in biomacromolecules: Past, present, and future. *J. Magn. Reson.* 241(0):3-17.
196. Hoogstraten CG, Wank JR, & Pardi A (2000) Active Site Dynamics in the Lead-Dependent Ribozyme. *Biochemistry* 39(32):9951-9958.
197. Blad H, Reiter NJ, Abildgaard F, Markley JL, & Butcher SE (2005) Dynamics and Metal Ion Binding in the U6 RNA Intramolecular Stem-Loop as Analyzed by NMR. *J. Mol. Biol.* 353(3):540-555.
198. Johnson JE & Hoogstraten CG (2008) Extensive Backbone Dynamics in the GCAA RNA Tetraloop Analyzed Using 13C NMR Spin Relaxation and Specific Isotope Labeling. *J. Am. Chem. Soc.* 130(49):16757-16769.
199. Lee J-H & Pardi A (2007) Thermodynamics and kinetics for base-pair opening in the P1 duplex of the Tetrahymena group I ribozyme. *Nucleic Acids Res.* 35(9):2965-2974.
200. Latham MP, Zimmermann GR, & Pardi A (2009) NMR Chemical Exchange as a Probe for Ligand-Binding Kinetics in a Theophylline-Binding RNA Aptamer. *J. Am. Chem. Soc.* 131(14):5052-5053.
201. Bothe JR, *et al.* (2011) Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat Meth* 8(11):919-931.
202. Mittermaier AK & Kay LE (2009) Observing biological dynamics at atomic resolution using NMR. *Trends Biochem. Sci.* 34(12):601-611.
203. Palmer III AG, Kroenke CD, & Loria JP (2001) [10] Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods in Enzymology*, eds Thomas L. James VDt & Uli S (Academic Press), Vol Volume 339, pp 204-238.
204. Palmer III AG (2004) NMR Characterization of the Dynamics of Biomacromolecules. *Chem. Rev.* 104(8):3623-3640.

205. Palmer AG & Massi F (2006) Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy. *Chem. Rev.* 106(5):1700-1719.
206. Hansen AL, Nikolova EN, Casiano-Negroni A, & Al-Hashimi HM (2009) 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(11):3818-3819.
207. Massi F, Johnson E, Wang C, Rance M, & Palmer AG (2004) NMR R1 rho Rotating-Frame Relaxation with Weak Radio Frequency Fields. *J. Am. Chem. Soc.* 126(7):2247-2256.
208. Korzhnev DM, Orekhov VY, & Kay LE (2004) 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(2):713-721.
209. Miloushev VZ & Palmer III AG (2005) R(1rho) relaxation for two-site chemical exchange: General approximations and some exact solutions. *J. Magn. Reson.* 177(2):221-227.

Chapter 2

Influence of Dimethylsulfoxide on RNA Structure and Ligand Binding

2.1 Introduction

Many regulatory RNA elements are emerging as new potential drug targets for treating a wide range of diseases(1-5). For example, riboswitches are RNA-based gene regulatory elements in bacteria that form complex 3D structures, and are being targeted in the development of antibiotics(6-8). Many RNA hairpin structures, including bacterial(9, 10) and retroviral elements(11-15), human micro-RNAs(16), various repeats(17, 18) as well as more complex pseudoknots(19) are being targeted in the development of therapeutics against infectious diseases, diabetes, various genetic disorders, and cancer. There is growing interest in using experimental and computational high throughput screens (HTS) to identify small molecules that can target RNA(2-5) and that overcome delivery limitations inherent to large molecular weight RNA-based therapeutics such as antisense(20) and small interfering RNAs(21).

Dimethyl sulfoxide (DMSO) is widely used as a universal solvent in HTS due to its miscibility with water, non-reactivity, and ability to dissolve hydrophobic compounds(22). Although it is well established that DMSO concentrations (<10%) typically used in HTS can have a significant effect on the stability of certain proteins(22) and reduce ligand binding affinities by as much as 10-fold(23), few studies have examined the impact of DMSO (<10%) on RNA structure and ligand-binding. Such studies are needed given that many RNA targets are highly flexible(24, 25) and

therefore potentially highly susceptible to perturbations by external chemical agents.(26) Indeed, high DMSO concentrations (>75%) have previously been shown to disrupt the structure and stability of RNA(27, 28) and DNA.(29-31) In particular, NMR studies of yeast $tRNA^{phe}$ showed that high concentrations of DMSO (up to 83%) resulted in changes in the ^{1}H spectra that were consistent with disruption of base-stacking and increased flexibility(28). Moreover, in a recent virtual screen targeting the transactivation response element (TAR) RNA from the human immunodeficiency virus type I (HIV-1), we found that the vast majority of false positive hits were water insoluble compounds that required DMSO for dissolution.(13) Therefore, it is important to verify that DMSO does not affect the structure of TAR and other RNAs or interfere with their ability to bind small molecules in ways that are not accounted for in the virtual screen.

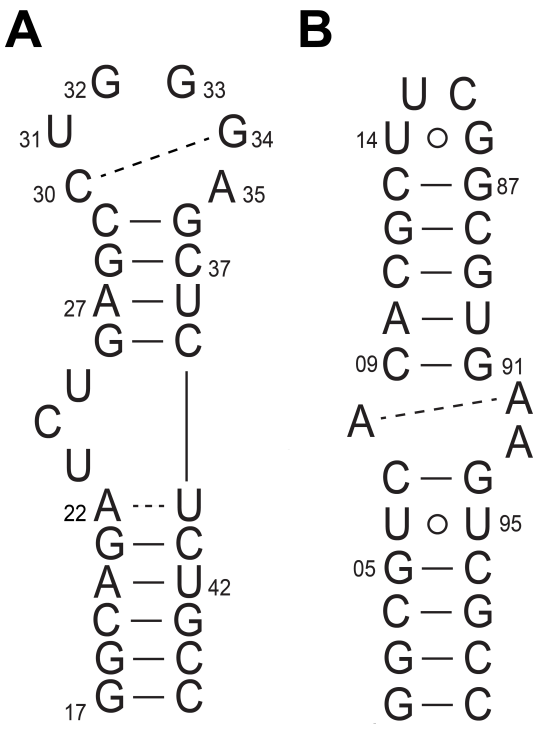

Figure 2.1 Secondary structure of RNA used in this study. Solid and dashed lines denote Watson-Crick and flexible base pairs, respectively. **A.** HIV-1 TAR **B.** Bacterial ribosomal A-site. The open circle denotes a U-U wobble base pair.

Here we use a combination of nuclear magnetic resonance (NMR) and fluorescence spectroscopy to explore how DMSO affects the structure and ligand-binding properties of two flexible hairpin RNA structures; HIV-1 TAR (32, 33) and the ribosomal A-site(34, 35) (Figure 2.1). These two RNAs have served as model systems for

understanding RNA-small molecule targeting and provide an excellent opportunity, to examine in depth, the impact of DMSO on RNA structure and ligand binding.

2.2 Materials and Methods

Preparation of RNA samples and small molecules

2-aminopurine (2AP)-labeled TAR and A-site RNA were purchased from Dharmacon (Lafayette, CO). Uniformly $^{13}C/^{15}N$ labeled TAR RNA and A-site RNA (Figure 2.1) were prepared by *in vitro* transcription using double stranded DNA encoding the RNA sequence of interest and containing the T7 promoter at 5'-end (Integrated DNA Technologies). T7 RNA polymerase (Takara Mirus Bio, Inc.) was used to transcribe the DNA sequence in the presence of $^{13}C/^{15}N$ labeled nucleotide triphosphates (ISOTEC, Inc. and Cambridge Isotope Laboratories, Inc). The RNA was purified using 20% (w/v) denaturing polyacrylamide gel electrophoresis (PAGE) in 8M urea and 1X TBE. The RNA was electroeluted in 20mM Tris (pH 8) buffer and then precipitated in ethanol. The purified RNA pellet was dissolved and exchanged into NMR buffer (15mM sodium phosphate, 25mM sodium chloride, 0.1mM EDTA, 10% (v/v) D_2O, and pH ~ 6.4) using a centricon ultracel YM-3 concentrator (Millipore Corp.). The compounds L-argininamide (A3913), mitoxantrone (M6545), kanamycin B (B5264), and paromomycin (P9297) were purchased from Sigma-Aldrich (St. Louis, MO). 2-aminopurine base (276560500) was purchased from Acros Organics (Geel, Belgium). DMSO was purchased from Amresco (Solon, OH).

NMR spectroscopy

All NMR experiments were performed at 298K on 600 MHz Avance Bruker or Agilent spectrometers equipped with 5mm triple-resonance cryogenic probe. 5% and 10% DMSO were added volume by volume (v/v). All NMR spectra were processed using NMRPipe(36) and SPARKY 3(37). The overall chemical shift perturbations ($\Delta\delta_{overall}$) upon incremental addition of DMSO in 2D-HSQC spectra of TAR and A-site were calculated using the following equation:,

$$\Delta\delta_{\text{overall}} = \sqrt{(\Delta\delta_{\text{H}})^2 + (\frac{\gamma_{\text{C}}}{\gamma_{\text{H}}}(\Delta\delta_{\text{C}}))^2} \quad (1)$$

where $\Delta\delta_H$ and $\Delta\delta_C$ are the changes in ^1H and ^{13}C chemical shift (in ppm), respectively, and γ_H and γ_C are gyromagnetic ratios for hydrogen and carbon, respectively. For N-H HSQC spectra, γ_C and $\Delta\delta_C$ are replaced with γ_N and $\Delta\delta_N$, respectively, which are gyromagnetic ratio for nitrogen and the changes in ^{15}N chemical shift (in ppm), respectively.

Fluorescence Spectroscopy

The fluorescence-based binding assays employed TAR RNA with 2AP labeled at the bulge residue U25 (2AP-U25-TAR) and ribosomal A-site RNA with 2AP labeled at the internal loop residue A92 (2AP-A92-Asite). 10 μM of 2AP-labeled RNA was annealed by heating to 95°C for 5 minutes and then cooled on ice for an hour prior to use. The folded RNA was diluted to 20 nM concentration in assay buffer (10mM sodium phosphate, 50mM NaCl, 0.1mM EDTA, and pH ~6.8). The temperature of the cuvette holder was maintained at 25°C with a water-cooling system. DMSO was added volume by volume (v/v). Time-resolved fluorescence intensity measurements were collected using a Fluoromax-2 fluorimeter at 320 nm excitation and 370 nm emission wavelengths following incremental addition of small molecules in 1:1000 dilutions. Each titration

point with small molecule was averaged over 15 seconds of fluorescence intensity measurement with 0.1 second time interval. The slit width was 10 nm. All measurements were triplicated. The fluorescence emission intensities were normalized with respect to the fluorescence emission intensity in the absence of small molecules. With the exception of kanamycin B binding to TAR, all dissociation constants were computed by fitting to the following equation using the Origin software (Origin Lab Corporation),

$$F = A \times [RNA]_T + B \times \left\{ \frac{([RNA]_T + [L]_T + K_d) - \sqrt{([RNA]_T + [L]_T + K_d)^2 - (4 \times [RNA]_T \times [L]_T)}}{2} \right\}$$

(eq. 2)

where $[RNA]_T$ is the total RNA concentration, $[L]_T$ is the total small molecule concentration, and A and B are fluorescence contribution factors that account for relative fluorescence intensities in the free and bound state, respectively(38). The titration data for kanamycin B binding to TAR exhibited two inflection points characteristic of two-site binding. This data did not fit well to Equation 2 (reduced χ^2 ~ 44) and was instead fitted to Equation 3 (reduced χ^2 ~ 0.2) which assumes two independent site binding:

$$F = A \times [RNA]_T + B \times \left\{ \frac{([RNA]_T + [L]_T + K_{d,1}) - \sqrt{([RNA]_T + [L]_T + K_{d,1})^2 - (4 \times [RNA]_T \times [L]_T)}}{2} \right\} +$$
$$C \times \left\{ \frac{([RNA]_T + [L]_T + K_{d,2}) - \sqrt{([RNA]_T + [L]_T + K_{d,2})^2 - (4 \times [RNA]_T \times [L]_T)}}{2} \right\}$$

(eq. 3)

The lower affinity binding site is not observed under higher ionic strength conditions (150mM NaCl or 3mM $MgCl_2$) and therefore likely reflects non-specific binding to the RNA.

The fluorescence emission spectra of 2AP-U25-TAR, 2AP-A92-Asite, and 2-aminopurine base were measured using Fluoromax-2 fluorimeter in the same assay buffer used in the binding assays. The fluorescence emission was measured from 350 nm to 400 nm with excitation at 320 nm. All spectra were averaged over 3 scans.

2.3 Results and Discussion

Impact of DMSO on the structure and dynamics of RNA by NMR

We investigated how DMSO affects the structure of HIV-1 TAR (Figure 2.1A) and bacterial ribosomal A-site RNA (Figure 2.1B) using NMR chemical shift titrations, where we acquired 2D N-H and C-H HSQC spectra of uniformly $^{13}C/^{15}N$ labeled RNA following the incremental addition of DMSO up to 10% concentration, which corresponds to the higher end of DMSO concentrations typically used in ligand-binding assays.(39)

Increasing concentrations of DMSO resulted in specific perturbations in the TAR chemical shifts at residues located in and around the flexible bulge and apical loop (Figure 2.2A). This data suggests that DMSO affected the TAR conformation and that the transition between free and "DMSO-bound" TAR occurs in fast exchange relative to the NMR timescale. The most significant ($\Delta\delta_{overall}$ > 0.1 ppm) perturbations were observed for residues A22 and U23 in and around bulge and the apical loop residue A35 (Figure 2.2B), which are all flexible residues that adopt partially stacked conformations.(40-42)

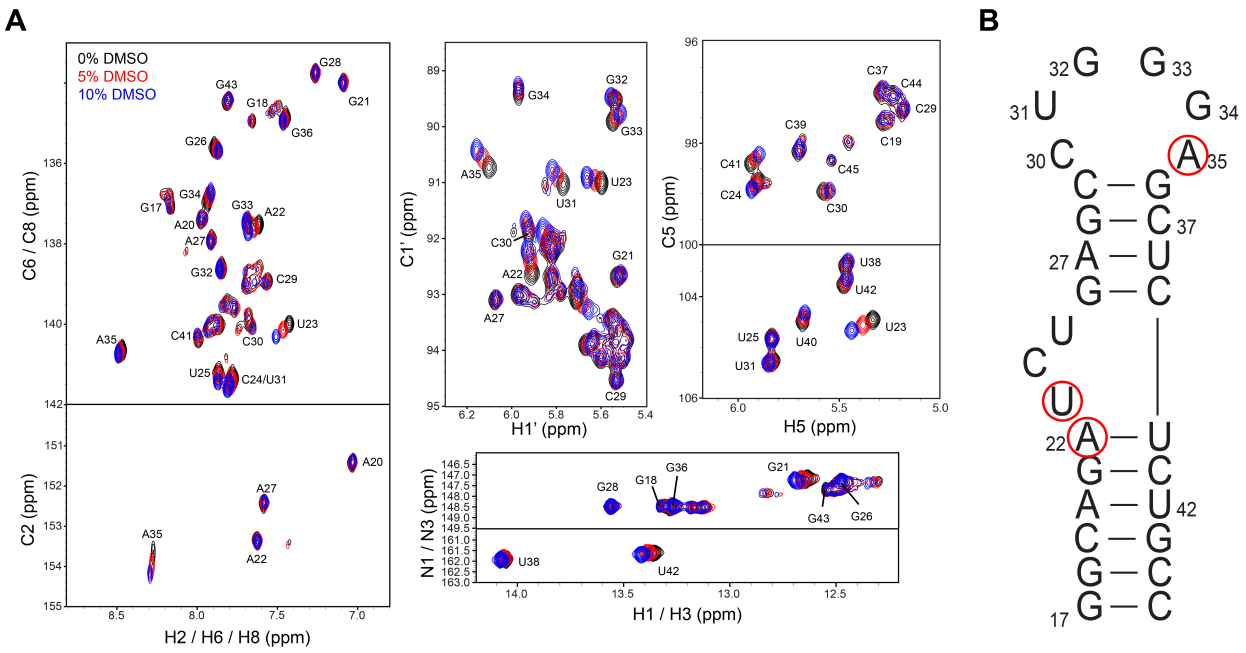

Figure 2.2 Examining impact of DMSO on HIV-1 TAR conformation by NMR. A. 2D C-H and N-H HSQC spectra of TAR in increasing DMSO concentration. **B.** Secondary structure of highlighting residues (in red circles) that undergo the largest DMSO induced chemical shift perturbations ($\Delta\delta_{overall}$ >0.1 ppm).

The downfield perturbations in the nucleobase (C6/C8) carbon resonances induced by DMSO at the bulge (A22(C8H8), U23(C6H6), and U25(C6H6)) suggest loss of stacking interactions(42-44). Interestingly, all of the DMSO-induced chemical shift perturbations in and around the bulge, including for A22, U23, C24, U25 and U40 were similar to those induced by increasing Mg^{2+} and Na^{+} concentrations (Figure 2.3), which were previously shown to stabilize a co-axial TAR conformation in which all three bulge residues are flipped out and flexible.(45, 46) Indeed, DMSO induced a gradual increase in the normalized resonance intensities(47) in the bulge resonances (U23, C24,

U25) (Figure 2.4), consistent with an increase in picosecond-to-nanosecond timescale motions due to loss of stacking interactions.

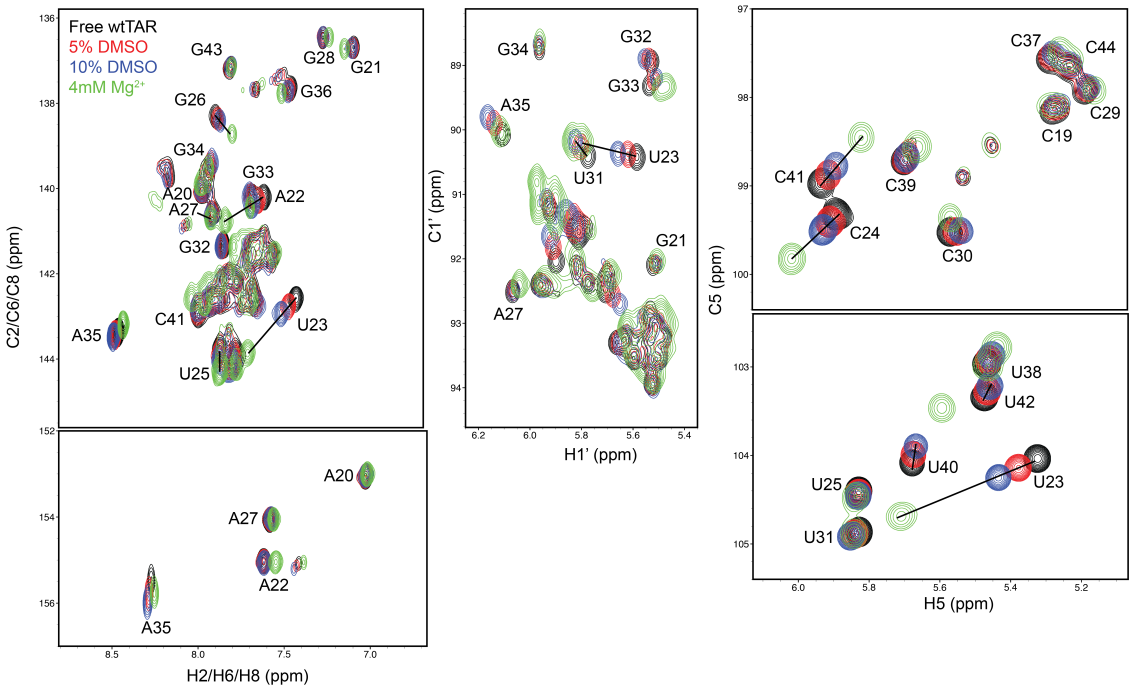

Figure 2.3 Comparison of 2D C-H HSQC spectra of HIV-1 TAR in 0-10% DMSO and 4mM $MgCl_2$.

In contrast, the DMSO-induced perturbations at apical loop residue A35(C1′H1′) were not observed with increasing Mg^{2+} (Figure 2.3) and there were many resonances which showed significant perturbations with Mg^{2+} that showed little to no perturbations with DMSO (Figure 2.3). The carbon perturbations for G34(C8H8), A35(C1′H1′), and A35(C8H8) suggest stabilization of A35 in a flipped out conformation and G34 in an intra-helical stacked conformation. Prior studies have shown that A35 and G34 exist in a dynamic equilibrium in which they inter-change stacking interactions, with one residue flipping in while the other flips out, with the dominant form being a conformation in which A35 is flipped out and G34 flipped in(42). The NMR data suggests that DMSO

favors this dominant conformation. This is also consistent with the gradual increase in measured normalized resonance intensities (Figure 2.4) seen for A35(C1′H1′) and A35(C8H8) upon addition of DMSO.

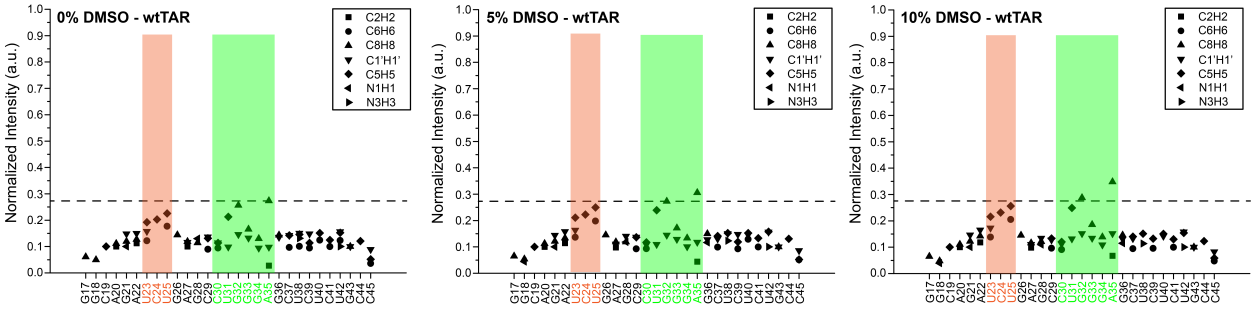

Figure 2.4 Normalized resonance intensities measured from 2D-HSQC spectra of TAR as a function of DMSO concentration. Orange residues and green residues belong to the bulge and apical loop respectively. The dotted line shows the normalized intensity of A35(C8H8).

In contrast, little to no chemical shift perturbations or changes in resonance intensities were observed for more stable helical residues. This is also consistent with the imino N1-H1 and N3-H3 HSQC spectra (Figure 2.2A), which showed little to no changes on addition of DMSO. Thus, the more stable helical residues seem to be more shielded from the effects of DMSO.

Addition of DMSO to A-site also resulted in significant chemical shift perturbations, which were localized in flexible internal loop (A08, A92 and A93) and nearby residues (G91 and G94), indicating that DMSO also induced changes in the structure of A-site (Figure 2.5A). Once again, the gradual changes in the NMR resonance positions with increasing DMSO suggests that any transition between free and "DMSO-bound" A-site occurred in fast exchange relative to the NMR timescale. The largest perturbations ($\Delta\delta_{overall}$ > 0.1 ppm) were observed for the flexible and intra-

helically stacked A92 (Figure 2.5B). The downfield shifted ^{13}C chemical shift particularly for A92(C8H8) is consistent with the loss of stacking and flipping out of the adenosine base. Indeed, similar perturbations were observed for A92 upon the addition of the aminoglycoside paromomycin, which has been known to promote the flipping out of A92.(34, 48, 49) This is also consistent with the gradual increase in measured normalized resonance intensities(47) (Figure 2.6) for A92 and A93 upon addition of DMSO, consistent with an increase in picosecond-to-nanosecond timescale motions due to loss of stacking interactions.

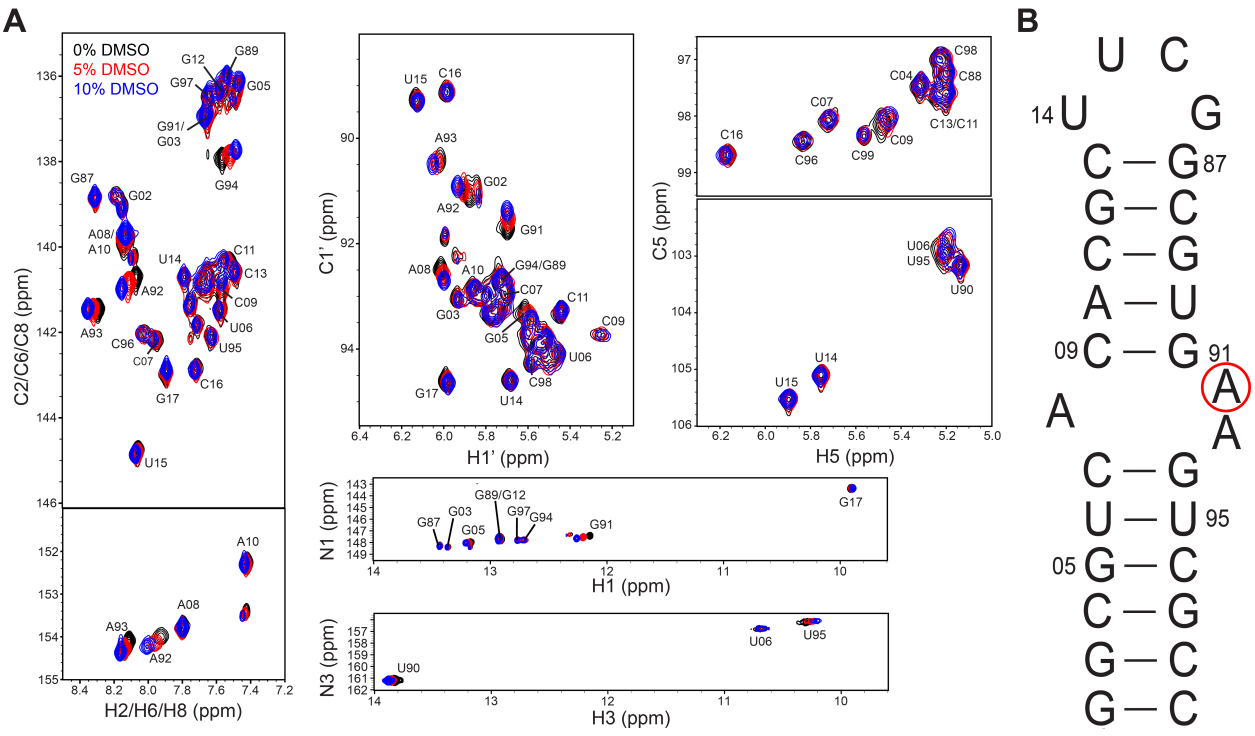

Figure 2.5 Examining impact of DMSO on bacterial A-site conformation by NMR. A. 2D C-H and N-H HSQC spectra of A-site in increasing DMSO concentration. **B.** Secondary structure of A-site highlighting residues (in red circles) that undergo the largest DMSO induced chemical shift perturbations ($\Delta\delta_{overall}$ >0.1 ppm).

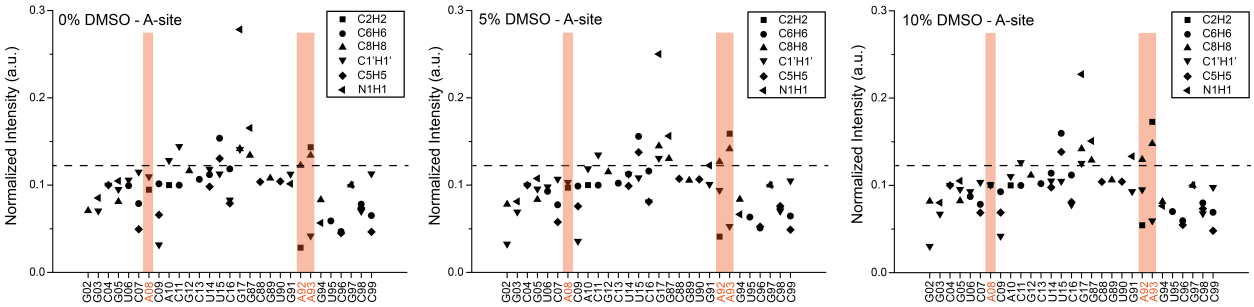

Figure 2.6. Normalized resonance intensities measured from 2D-HSQC spectra of A-site as a function of DMSO concentration. Orange residues denote the internal loop residues. The dotted line shows the normalized intensity of A92(C8H8).

DMSO did not induce any significant chemical shift perturbations in the thermodynamically stable UUCG apical loop(50). However, in contrast to TAR, DMSO did have an effect on the 2D N-H HSQC spectrum of A-site. In particular, DMSO induced a significant downfield shift in the imino proton of G91-H1, suggesting that DMSO affects the C09-G91 base-pair, which is near the non-canonical residues. Once again, canonical base-pairs embedded within helices did not exhibit significant chemical shift perturbations with DMSO.

Impact of DMSO on stacking interactions using 2-AP fluorescence

We used 2-aminopurine (2AP) fluorescence to further examine the impact of DMSO on stacking interactions. 2AP is widely used as a fluorescent reporter of stacking interactions in nucleic acids since the fluorescence emission intensity is highly sensitive to the details of the stacking interactions, and generally increases upon transitioning from a stacked to an unstacked conformation.(51-53) For these studies, we used a TAR construct (2AP-U25-TAR)(38) in which the flipped out bulge residue U25 is labeled with

2-AP and an A-site construct (2AP-A92-Asite)(54) in which A92 is labeled with 2-AP (Figure 2.7). Both of these constructs have previously been used in ligand binding studies(38, 54).

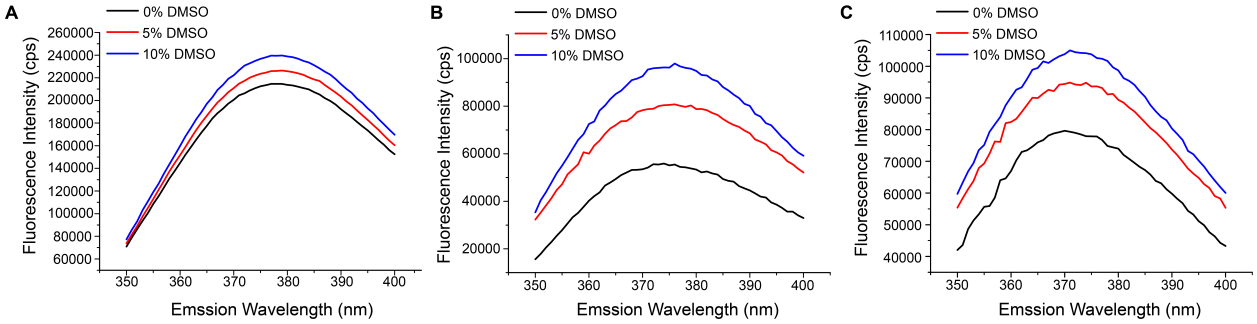

Figure 2.7 Examining impact of DMSO on RNA stacking interactions using 2AP fluorescence. Shown is the emission spectrum of 2AP at max wavelength with increasing DMSO concentration **A**. 2-AP **B**. 2AP-U25-TAR. **C**. 2AP-A92-Asite.

As a control, we first examined how DMSO affects the fluorescence intensity of the 2-AP base. Upon addition of 5% and 10% DMSO, we observed a small increase in the fluorescence intensity of 2-aminopurine base at maximum emission wavelength (λ_{max} = 378nm) of 5% and 11%, respectively (Figure 2.7A). This indicates that DMSO did not significantly alter the fluorescence properties of 2-AP. Moreover, addition of 10% DMSO did not shift the maximum emission wavelengths of 2-aminopurine base and 2AP-labeled RNA constructs.

Addition of 5%-10% DMSO increased the fluorescence intensity of 2AP-U25-TAR at maximum emission wavelength (λ_{max} = 374nm) by 44%-75% (Figure 2.7B). For A-site, 5%-10% DMSO increased the fluorescence intensity at maximum emission wavelength (λ_{max} = 370nm) by 19% and 32%, respectively (Figure 2.7C). The observed increase in the fluorescence intensities is consistent with the DMSO-induced loss of stacking

interactions within the TAR bulge and A-site internal loop, as suggested independently by the NMR data. The smaller DMSO-induced increase in fluorescence intensity observed for A-site as compared to TAR is likely because the purine A92 retains stacking interactions with the flipped out A93 as observed in structures of A-site bound to aminoglycosides(48, 54).

Impact of DMSO on ligand binding affinity

We used 2-AP fluorescence to examine the impact of DMSO on the binding affinity of small molecules that bind to TAR and A-site. We measured the dissociation constants (K_d) for argininamide (ARG) (Figure 2.8A), kanamycin B (Figure 2.8B), and mitoxantrone (Figure 2.8C) binding to TAR and for paromomycin (Figure 2.9A) and mitoxantrone (Figure 2.9B) binding to A-site. ARG is a ligand mimic of TAR's cognate protein target, the transactivator protein, which binds TAR with micromolar affinity and which has been shown to recapitulate many essential features of TAR-Tat recognition(55-57). Kanamycin B and paromomycin are example aminoglycosides that bind RNA in the nanomolar range and in a manner strongly dependent on electrostatic interactions.(58-61) Mitoxantrone is a newly identified intercalator that binds non-specifically to RNA with affinities on the nanomolar range(13).

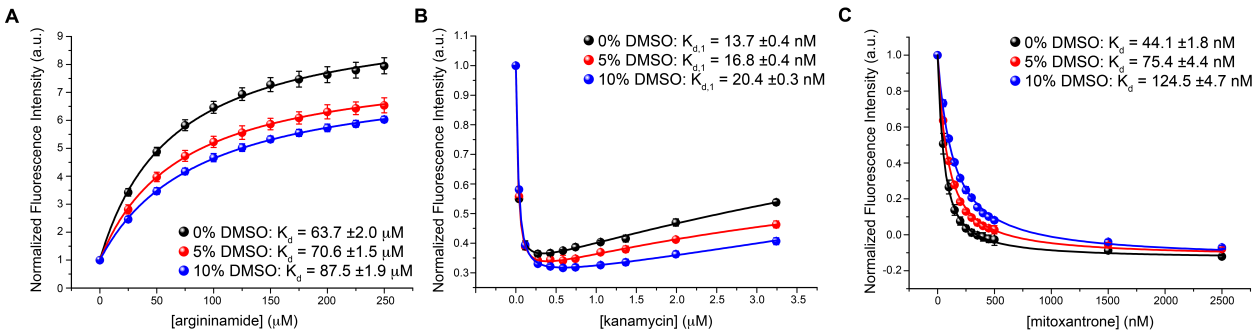

Figure 2.8. Impact of DMSO on TAR-small molecule binding affinities. Shown are 2-AP fluorescence intensity titration curves for small molecule binding to 2AP-U25-TAR with varying DMSO concentration. **A.** Argininamide. **B**. Kanamycin. **C**. Mitoxantrone. Error bars are obtained from repeating the measurements in triplicate.

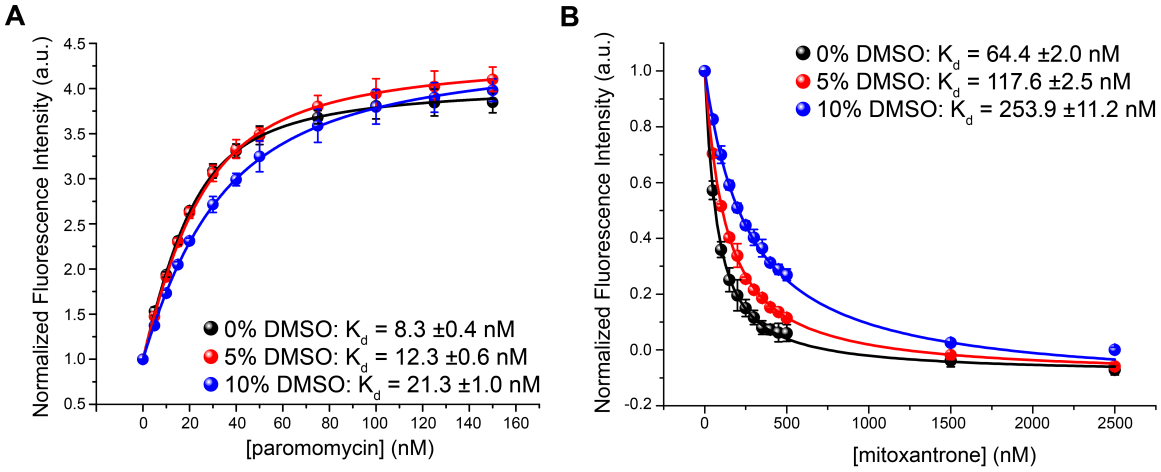

Figure 2.9. Impact of DMSO on A-site-small molecule binding affinities. Shown are 2-AP fluorescence intensity titration curves for small molecule binding to 2AP-A92-A-site with varying DMSO concentration. **A.** Paromomycin. **B**. Mitoxantrone. Error bars are obtained from repeating the measurements in triplicate.

All of the molecules tested resulted in significant changes in fluorescence intensity upon binding to their RNA targets allowing the accurate determination of K_ds. In all cases, the addition of DMSO slightly weakened the RNA-ligand binding affinity and resulted in an increase in the measured K_d (Table 2.1). This increase was small for 5%

DMSO but became significant at 10% DMSO. For example, while the addition of 5% DMSO increased the measured K_d for ARG, kanamycin B, and mitoxantrone by only 1.1, 1.2 and 1.7 fold, respectively, 10% DMSO resulted in much larger increases of 1.4, 1.5, and 2.8 fold, respectively. It should be noted that kanamycin B exhibited a second weaker and non-specific binding to TAR which was not observable in the presence of 10% DMSO. A more significant 1.7-2.8 fold increase in K_d was observed for the intercalator, mitoxantrone (Table 2.1). Similar results and trends were observed with A-site, where the addition of 5% DMSO increased the measured K_d for paromomycin and mitoxantrone by 1.5 and 1.8 fold respectively, whereas 10% DMSO resulted in much larger increases of 2.6 and 3.9 fold respectively (Table 2.1).

2AP-U25-TAR			
Small Molecule	**0% DMSO**	**5% DMSO**	**10% DMSO**
Argininamide	63.7 ±2.0 mM	70.6 ±1.5 mM	87.5 ±1.9 mM
Kanamycin B	13.7 ±0.4 nM 10.2 ±4.1 mM	16.8 ±0.4 nM 5.5 ±1.1 mM	20.4 ±0.3 nM N/A
Mitoxantrone	44.1 ±1.8 nM	75.4 ±4.4 nM	124.5 ±4.7 nM
2AP-A92-Asite			
Small Molecule	**0% DMSO**	**5% DMSO**	**10% DMSO**
Paromomycin	8.3 ±0.4 nM	12.3 ±0.6 nM	21.3 ±1.0 nM
Mitoxantrone	64.4 ±2.0 nM	117.6 ±2.5 nM	253.9 ±11.2 nM

Table 2.1 Dissociation constants (K_ds) for RNA-small molecule binding determined using 2-AP fluorescence at different DMSO concentrations.

In general, the weakened binding affinity observed with DMSO was not as dramatic as that reported for some proteins(23). However, our study focused on hairpin structures; additional studies are required with more complex RNA structures, such as riboswitches, that have deeper and often more hydrophobic binding pockets that are more similar to typical protein binding sites.

Impact of DMSO on RNA ligand binding mode

To further examine whether DMSO affects the RNA-ligand binding kinetics and the RNA-ligand bound structure, we performed NMR chemical shift titrations in which spectra of TAR or A-site were recorded upon addition of a small molecule in the presence of 5% DMSO and compared the NMR titration profiles with counterparts observed in the absence of DMSO (Figure 2.10). In both cases, similar chemical shift perturbations were observed in the absence and presence of DMSO, indicating that the exchange kinetics between free and ligand-bound RNA structure, and the ligand-binding mode were not significantly affected by DMSO. Comparison of the spectra of TAR-ARG and A-site-paromomycin complexes with and without 5% DMSO (Figure 2.11) revealed even smaller differences than observed for the unbound RNA, indicating that DMSO has a smaller effect on the structures of RNA-ligand complexes.

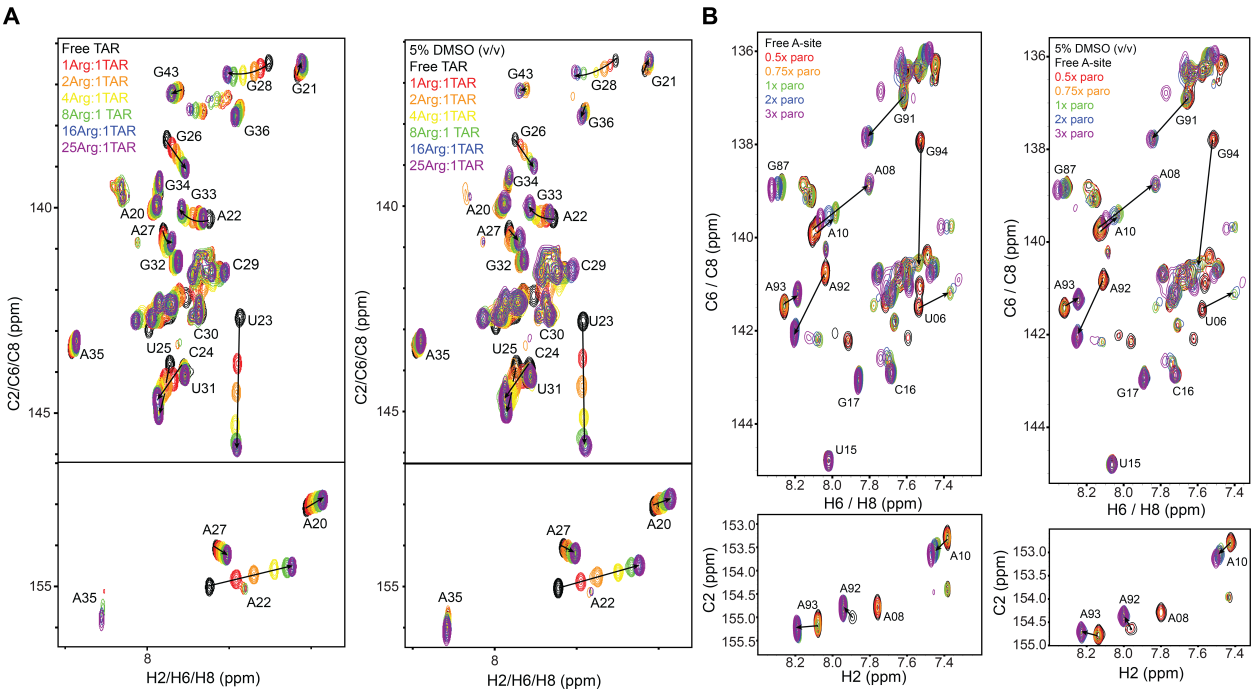

Figure 2.10. Examining small molecule binding to RNA in DMSO by NMR. Shown are 2D C-H HSQC spectra of RNA with small molecules titrated in the absence and presence of 5% DMSO. **A.** Argininamide binding to TAR **B.** Paromomycin binding to A-site

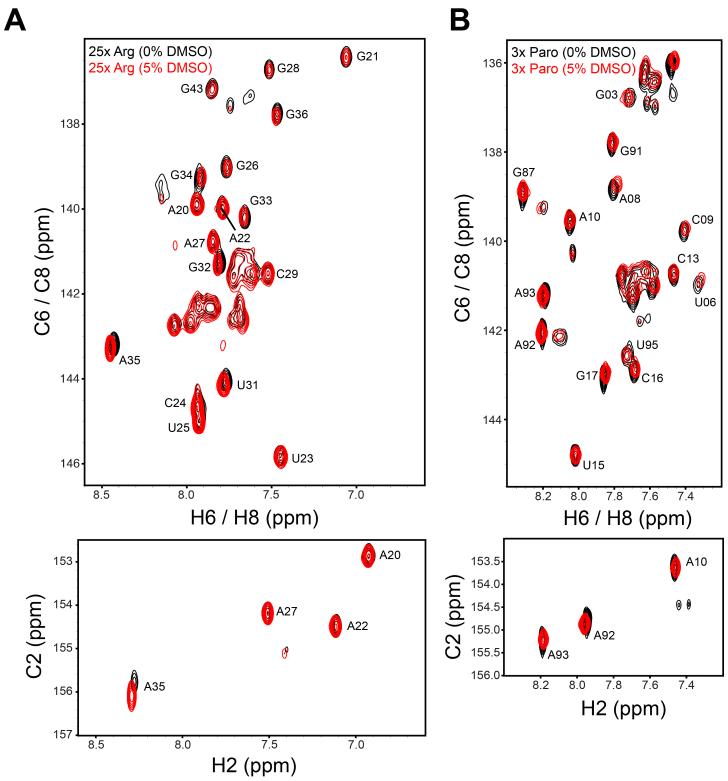

Figure 2.11. Examining impact of DMSO on RNA-ligand complexes by NMR. Shown are 2D C-H HSQC spectra of RNA-small molecules complexes in absence and presence of DMSO. **A.** TAR-argininamide at 1:25 molar ratio and **B.** A-site-paromomycin at 1:3 molar ratio.

Mechanism of DMSO perturbations and implications for experimental and virtual screens

Our results show that DMSO promotes the flipping out of the flexible non-canonical residues. This effect is likely due to the increased hydrophobicity of the solvent, which favors solvation of the hydrophobic nucleobase as compared to the polar water solvent, and possibly due to DMSO forming hydrogen bonds with nucleobases and other hydration effects. Prior studies showed that organic solvents, including N,N-dimethylformamide, DMSO, tetramethylurea, methanol, ethanol, and ethylene glycol

denature duplex DNA in a manner strongly dependent on their degree of hydrophobicity(30). Another study has shown that replacement of the two methyl groups in DMSO with ethyl groups results in a more hydrophobic solvent (diethylsulfoxide or DESO) that is more effective at denaturing duplex DNA(31). While the DMSO concentrations (5%-10%) used in binding assays are not sufficient to disrupt duplex structures, it can promote destacking and 'local melting' of flexible non-canonical residue, that are often the sites of small molecule binding. These effects are expected to become even more pronounced for more flexible RNAs containing single-strands. While the effects of DMSO on RNA observed here are smaller than those reported for certain proteins(23), they warrant careful attention before proceeding with ligand binding assays and high throughput screening campaigns.

Our results suggest that DMSO has a smaller effect on the ligand binding affinities of polycationic aminoglycosides that bind TAR and A-site primarily through electrostatic interactions, as compared to the intercalating mitoxantrone that binds primarily via hydrophobic stacking interactions. One possibility would be that DMSO more strongly perturbs the RNA binding site for mitoxantrone as compared to the aminoglycosides. However, comparison of the chemical shift perturbations induced by DMSO and the small molecules does not provide any evidence that the DMSO more greatly affects the mitoxantrone binding site. Rather, this effect is likely due to the increase in the hydrophobicity of the solvent with DMSO, which favorably solvates the hydrophobic mitoxantrone. The slightly weakened K_d of paromomycin in 10% DMSO could also be explained by the weakened stacking interaction between ring I of paromomycin and G91 of A-site(48) due to increase in the hydrophobicity of the solvent and possibly due to competing h-bonding interactions among DMSO, water, and the A-site nucleobases.

Our study was motivated in large part by a recent virtual screen targeting HIV-1 TAR(13). Among 58 top-scoring compounds tested, only 14 compounds were water-soluble and did not require DMSO for solubilization, and among these compounds, 7 were active in *in vitro* binding assays. In stark contrast, among 44 water insoluble compounds that required DMSO for solubilization, none were active in *in vitro* binding assays. The overall hit rate for the virtual screen was 12%, but improved to 50% when excluding compounds that required DMSO for their solubilization. Our results suggest that DMSO is unlikely to have affected TAR-ligand binding to such an extent as to completely abrogate binding. Rather, it is more likely that the docking scoring function used in the virtual screen generates false positives that are biased toward hydrophobic compounds. Indeed, preliminary analysis of a virtual screen using a distinct docking program and scoring function reveals a smaller level enrichment with hydrophobic compounds among the top virtual screening hits (data not shown).

2.4 Conclusion

We have investigated how DMSO affects the structure and ligand binding properties of two well-studied RNA targets, HIV-1 TAR and ribosomal A-site. In both cases, typical concentrations (5%-10%) of DMSO used in ligand binding assays and in high throughput screens destabilized non-canonical residues within bulges and loops and resulted in a 0.3–4 fold reduction in the measured binding affinities for different small molecules, with the greatest reduction observed for an intercalating hydrophobic compound that binds RNA non-specifically. Our results suggest that by competing for hydrophobic interactions, DMSO can have some effects on RNA structure and ligand binding. These effects should be considered when developing ligand binding assays and high throughput screens.

2.5 References

1. Cooper TA, Wan L, & Dreyfuss G (2009) RNA and Disease. *Cell* 136(4):777-793.
2. Thomas JR & Hergenrother PJ (2008) Targeting RNA with Small Molecules. *Chem. Rev.* 108(4):1171-1224.
3. Hermann T (2000) Strategies for the Design of Drugs Targeting RNA and RNA–Protein Complexes. *Angew. Chem. Int. Ed.* 39(11):1890-1904.
4. Guan L & Disney MD (2012) Recent Advances in Developing Small Molecules Targeting RNA. *ACS Chem. Biol.* 7(1):73-86.
5. Ferner J-P, *et al.* (2012) RNA as a Drug Target. *NMR of Biomolecules*, (Wiley-VCH Verlag GmbH & Co. KGaA), pp 298-313.
6. Serganov A & Patel DJ (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* 8:776-790.
7. Blount KF & Breaker RR (2006) Riboswitches as antibacterial drug targets. *Nat. Biotechnol.* 24:1558-1564.
8. Schwalbe H, Buck J, Fürtig B, Noeske J, & Wöhnert J (2007) Structures of RNA Switches: Insight into Molecular Recognition and Tertiary Structure. *Angew. Chem. Int. Ed.* 46(8):1212-1219.
9. Li L, *et al.* (2013) BSRD: a repository for bacterial small regulatory RNA. *Nucleic Acids Res.* 41(D1):D233-D238.
10. Hermann T (2005) Drugs targeting the ribosome. *Curr. Opin. Struct. Biol.* 15(3):355-366.
11. Zhou S, Rynearson KD, Ding K, Brunn ND, & Hermann T (2013) Screening for inhibitors of the hepatitis C virus internal ribosome entry site RNA. *Bioorg. Med. Chem.* (0):DOI: 10.1016/j.bmc.2013.1003.1054.
12. Marcheschi RJ, Mouzakis KD, & Butcher SE (2009) Selection and Characterization of Small Molecules That Bind the HIV-1 Frameshift Site RNA. *ACS Chem. Biol.* 4(10):844-854.
13. Stelzer AC, *et al.* (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* 7(8):553-559.
14. Zapp ML, Stern S, & Green MR (1993) Small molecules that selectively block RNA binding of HIV-1 rev protein inhibit rev function and viral production. *Cell* 74(6):969-978.
15. Ennifar E, *et al.* (2006) Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell. *Nucleic Acids Res.* 34(8):2328-2339.
16. Jiang Q, *et al.* (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 37(suppl 1):D98-D104.

17. Childs-Disney JL, Hoskins J, Rzuczek SG, Thornton CA, & Disney MD (2012) Rationally Designed Small Molecules Targeting the RNA That Causes Myotonic Dystrophy Type 1 Are Potently Bioactive. *ACS Chem. Biol.* 7(5):856-862.
18. Ranum LPW & Day JW (2004) Pathogenic RNA repeats: an expanding role in genetic disease. *Trends Genet.* 20(10):506-512.
19. Staple DW & Butcher SE (2005) Pseudoknots: RNA Structures with Diverse Functions. *PLoS Biol* 3(6):e213.
20. Akhtar S, *et al.* (2000) The delivery of antisense therapeutics. *Adv. Drug Deliv. Rev.* 44(1):3-21.
21. Aagaard L & Rossi JJ (2007) RNAi therapeutics: Principles, prospects and challenges. *Adv. Drug Deliv. Rev.* 59(2,Äì3):75-86.
22. Tjernberg A, Markova N, Griffiths WJ, & Hallén D (2006) DMSO-Related Effects in Protein Characterization. *J. Biomol. Screening* 11(2):131-137.
23. Cubrilovic D & Zenobi R (2013) Influence of Dimehylsulfoxide on Protein,ÄìLigand Binding Affinities. *Anal. Chem.* 85(5):2724-2730.
24. Hermann T (2002) Rational ligand design for RNA: the role of static structure and conformational flexibility in target recognition. *Biochimie* 84(9):869-875.
25. Dethoff EA, Chugh J, Mustoe AM, & Al-Hashimi HM (2012) Functional Complexity and Regulation through RNA Dynamics. *Nature* 482:322-330.
26. Casiano-Negroni A, Sun X, & Al-Hashimi HM (2007) Probing Na+-Induced Changes in the HIV-1 TAR Conformational Dynamics Using NMR Residual Dipolar Couplings:,Äâ New Insights into the Role of Counterions and Electrostatic Interactions in Adaptive Recognition,Ä†. *Biochemistry* 46(22):6525-6535.
27. Strauss Jr. JH, Kelly RB, & Sinsheimer RL (1968) Denaturation of RNA with Dimethyl Sulfoxide. *Biopolymers* 6:793-807.
28. Crawford JE, Chan SI, & Schweizer MP (1971) NMR studies of organic solvent denatured yeast phenylalanyl transfer RNA at 220 MHZ. *Biochem. Biophys. Res. Commun.* 44(1):1-7.
29. Bonner G & Klibanov AM (2000) Structural stability of DNA in nonaqueous solvents. *Biotechnol. Bioeng.* 68(3):339-344.
30. Herskovits TT (1962) Nonaqueous solutions of DNA: Factors determining the stability of the helical configuration in solution. *Arch. Biochem. Biophys.* 97(3):474-484.
31. Markarian SA, Asatryan AM, Grigoryan KR, & Sargsyan HR (2006) Effect of diethylsulfoxide on the thermal denaturation of DNA. *Biopolymers* 82(1):1-5.

32. Muesing MA, Smith DH, & Capon DJ (1987) Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. *Cell* 48(4):691-701.
33. Feng S & Holland EC (1988) HIV-1 tat trans-activation requires the loop sequence within tar. *Nature* 334:165-167.
34. Yoshizawa S, Fourmy D, & Puglisi JD (1999) Recognition of the Codon-Anticodon Helix by Ribosomal RNA. *Science* 285(5434):1722-1725.
35. Schmeing TM & Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461:1234-1242.
36. Delaglio F, *et al.* (1995) Nmrpipe - a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR* 6(3):277-293.
37. Goddard TD & Kneller DG (SPARKY 3 (University of California, San Francisco).
38. Bradrick TD & Marino JP (2004) Ligand-induced changes in 2-aminopurine fluorescence as a probe for small molecule binding to HIV-1 TAR RNA. *RNA* 10(9):1459-1468.
39. Renner S, *et al.* (2005) New Inhibitors of the Tat–TAR RNA Interaction Found with a "Fuzzy" Pharmacophore Model. *ChemBioChem* 6(6):1119-1125.
40. Dethoff EA, *et al.* (2008) 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(8):3906-3915.
41. Hansen AL & Al-Hashimi HM (2007) Dynamics of Large Elongated RNA by NMR Carbon Relaxation. *J. Am. Chem. Soc.* 129(51):16072-16082.
42. Dethoff EA, Petzold K, Chugh J, Casiano-Negroni A, & Al-Hashimi HM (2012) Visualizing transient low-populated structures of RNA. *Nature* 491:724-728.
43. Nikolova EN, *et al.* (2011) Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* 470:498-502.
44. Xu X-P & Au-Yeung SCF (2000) Investigation of Chemical Shift and Structure Relationships in Nucleic Acids Using NMR and Density Functional Theory Methods. *J. Phys. Chem. B* 104(23):5641-5650.
45. Pitt SW, Majumdar A, Serganov A, Patel DJ, & Al-Hashimi HM (2004) Argininamide Binding Arrests Global Motions in HIV-1 TAR RNA: Comparison with Mg2+-induced Conformational Stabilization. *J. Mol. Biol.* 338(1):7-16.
46. Ippolito JA & Steitz TA (1998) A 1.3-Å resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation. *Proc. Natl. Acad. Sci. U. S. A.* 95:9819-9824.
47. Zhang Q, Sun X, Watt ED, & Al-Hashimi HM (2006) Resolving the Motional Modes That Code for RNA Adaptation. *Science* 311(5761):653-656.

48. Vicens Q & Westhof E (2001) Crystal Structure of Paromomycin Docked into the Eubacterial Ribosomal Decoding A Site. *Structure* 9(8):647-658.
49. Fourmy D, Yoshizawa S, & Puglisi JD (1998) Paromomycin binding induces a local conformational change in the A-site of 16 s rRNA. *J. Mol. Biol.* 277(2):333-345.
50. Varani G (1995) Exceptionally Stable Nucleic Acid Hairpins. *Annu. Rev. Biophys. Biomol. Struct.* 24(1):379-404.
51. Rachofsky EL, Osman R, & Ross JBA (2001) Probing Structure and Dynamics of DNA with 2-Aminopurine:‚Äâ Effects of Local Environment on Fluorescence‚Ä†. *Biochemistry* 40(4):946-956.
52. Jean JM & Hall KB (2001) 2-Aminopurine fluorescence quenching and lifetimes: Role of base stacking. *Proc. Natl. Acad. Sci. U. S. A.* 98(1):37-41.
53. Stivers JT (1998) 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. *Nucleic Acids Res.* 26(16):3837-3844.
54. Kaul M, Barbieri CM, & Pilch DS (2004) Fluorescence-Based Approach for Detecting and Characterizing Antibiotic-Induced Conformational Changes in Ribosomal RNA:‚Äâ Comparing Aminoglycoside Binding to Prokaryotic and Eukaryotic Ribosomal RNA Sequences. *J. Am. Chem. Soc.* 126(11):3447-3453.
55. Aboul-ela F, Karn J, & Varani G (1995) The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *J. Mol. Biol.* 253(2):313-332.
56. Puglisi J, Chen L, Frankel A, & Williamson J (1993) Role of RNA structure in arginine recognition of TAR RNA. *Proc. Natl. Acad. Sci. U. S. A.* 90:3680-3684.
57. Puglisi J, Tan R, Calnan B, Frankel A, & Williamson (1992) Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* 257(5066):76-80.
58. Smith AL, Kassman J, Srour KJ, & Soto AM (2011) Effect of Salt Concentration on the Conformation of TAR RNA and Its Association with Aminoglycoside Antibiotics. *Biochemistry* 50(44):9434-9445.
59. Stelzer AC, Kratz JD, Zhang Q, & Al-Hashimi HM (2010) RNA Dynamics by Design: Biasing Ensembles Towards the Ligand-Bound State. *Angew. Chem. Int. Ed.* 49(33):5731-5733.
60. François B, *et al.* (2005) Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucleic Acids Res.* 33(17):5677-5690.

61. Hermann T (2007) Aminoglycoside antibiotics: old drugs and new therapeutic approaches. (Translated from English) *Cell. Mol. Life Sci.* 64(14):1841-1852 (in English).

Chapter 3

Experimental High Throughput and Virtual Screening Targeting HIV-1 TAR RNA

3.1 Introduction

In the present study, we screened a ~100,000 small molecule library using the fluorescence-based high throughput screening and the ensemble-based virtual screening in order to identify compounds that disrupts HIV-1 TAR-Tat interaction. We have identified total 11 inhibitors of TAR-Tat interaction from HTS and VS, one of them showing a promising anti-HIV activity despite high cytotoxicity. We also showed that combining HTS and VS could enable the efficient search for small molecule modulators.

3.2 Materials and Methods

Preparation of HIV-1 TAR RNA and Tat peptide

The HIV-1 TAR RNA was prepared by *in vitro* transcription using double stranded DNA encoding the RNA sequence of interest and containing the T7 promoter at 5'-end (Integrated DNA Technologies). T7 RNA polymerase (Takara Mirus Bio, Inc.) was used to transcribe the DNA sequence in the presence of $^{13}C/^{15}N$ labeled nucleotide triphosphates (ISOTEC, Inc. and Cambridge Isotope Laboratories, Inc) or unlabeled nucleotide triphosphates. The RNA was purified using 20% (w/v) denaturing polyacrylamide gel electrophoresis (PAGE) in 8M urea and 1X TBE. The RNA was electroeluted in 20mM Tris (pH 8) buffer and then precipitated in ethanol. The purified RNA pellet was dissolved and exchanged into the phosphate buffer (15mM sodium

phosphate, 25mM sodium chloride, 0.1mM EDTA, and pH ~ 6.4) using a centricon ultracel YM-3 concentrator (Millipore Corp.). For the fluorescence assay, the RNA was diluted to 150nM in Tris-HCl buffer (50 mM Tris-HCl, 50 mM potassium chloride, 0.01% (v/v) Triton X-100 at pH ~7.4).

The arginine-rich motif (RKKRRQRRR) of Tat is doubly labeled with fluorescein at N-terminus and TAMRA at C-terminus. Three alanines were placed between the fluorophores and the arginine-rich motif to avoid potential interference on the binding affinity against TAR. The sequence used in this study is, N-(5-FAM)-AAARKKRRQRRRAAAK(TAMRA)-C, where N and C denote the N-terminus and C-terminus, respectively. This Tat peptide was purchased from Lifetein (Hillsborough, NJ). The purity of the peptide was > 95%.

TAR-Tat Displacement Assay

The assay was first developed by Matsumoto et al. for HIV-1 TAR (1). The assay utilizes the intramolecular quenching due to formation of dimers between fluorescein and TAMRA and the fluorescence resonance energy transfer (FRET) effect between fluorescein and TAMRA (Figure 3.1). When the Tat peptide is in free solution, the previous CD and NMR studies(2) (3) showed that the peptide does not form typical protein secondary structures (α-helix or β-sheets), but rather behave as an intrinsically disordered protein. The dynamic nature of the Tat peptide causes two fluorophores (fluorescein and TAMRA) to form a dimer each other, quenching the fluorescence of both fluorophores. The mechanism of quenching is called static quenching that occurs when the two fluorophores approach closer than 20 angstroms. When the Tat peptide is bound to the bulge of TAR, the circular dichroism (CD) spectroscopy confirmed that the peptide transitioned to the extended conformation(4). This extended conformation

creates a distance between the two fluorophores, fluorescein and TAMRA, and the FRET can be observed when the fluorescein is excited at 485 nm and TAMRA emission at 590 nm is detected. Experimentally, the binding of the Tat peptide to TAR increases the fluorescence emission from TAMRA and the displacement of Tat peptide by small molecule inhibitors decreases the fluorescence intensity.

The dose-dependent curve was fitted using the following equation (eq. 1) to obtain the half maximal inhibitory concentration (IC50).

$$Y = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{(1+10^{(\log \text{IC50} - x)*n})} \qquad \text{(eq. 1)}$$

, where "Bottom" is the lowest signal (due to inhibition), "Top" is the highest signal, and n is the hill slope. All variables (Top, Bottom, IC50, and n) are allowed to float during the fitting. Prism (GraphPad, Inc) was used to fit the dose-response curve in this study.

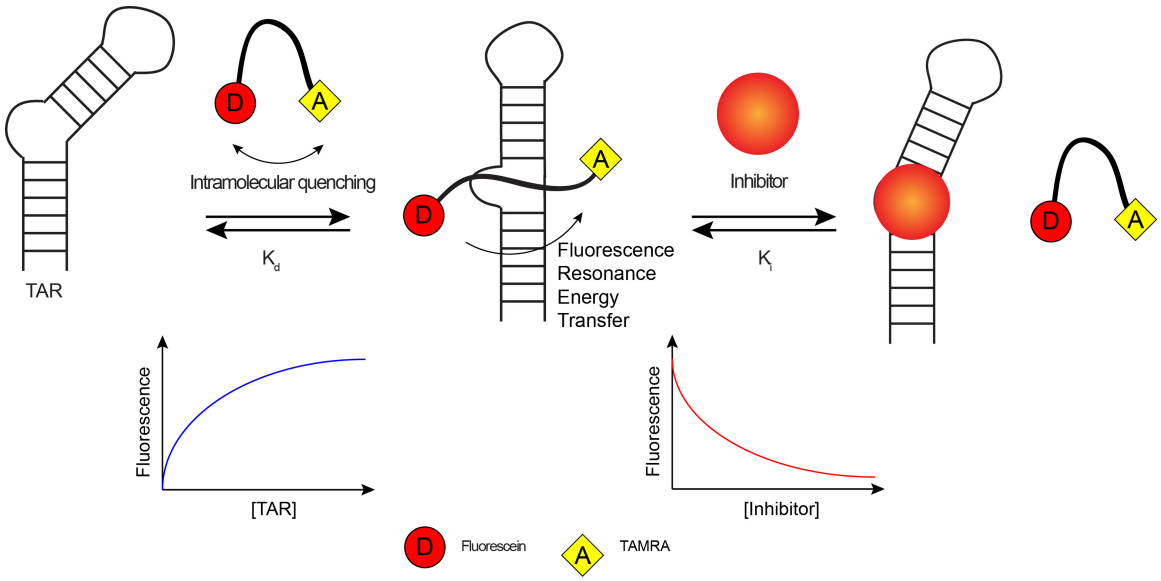

Figure 3.1 TAR-Tat displacement assay using fluorescence resonance energy transfer.

Small molecule library composition and high throughput screening

The small molecule library was obtained from the Center for Chemical Genomics (CCG) at University of Michigan – Ann Arbor. A total of 103,146 compounds were screened. Compounds were stored as stock solutions of 2 – 5 mM compound in DMSO for ~3 years. The 100,000 compounds were synthetic organic molecules with drug-like properties purchased from ChemDiv by CCG. Another 3,146 compounds consistent of 2,000 bioactive compounds from MicroSource, 446 compounds from National Institute of Health (NIH) clinical collection, and 700 active compounds against targets that CCG screened previously.

The high throughput screening was performed in the 384-well format (Figure 3.2). The small molecules were pin-tooled (200nL) into the wells by Biomek FX 384-well nanoliter HDR (Beckman) and Mosquito X1 (TTP Labtech). 10 μL each of TAR and Tat were dispensed to the entire plate by using Multidrop reagent dispenser (Thermo Scientific). The final volume in each well was 30 μL with 50 nM of TAR, 20 nM of Tat, and 13 μM of small molecules. The microplates were incubated at room temperature for 10 – 15 minutes. The microplates were then screened using Pherastar plate reader with an optical filter for the assay (excitation: 485 nm, emission: 590 nm). For each microplate, the Z-factor(5) was calculated. The average Z-factor throughout the HTS campaign was 0.71.

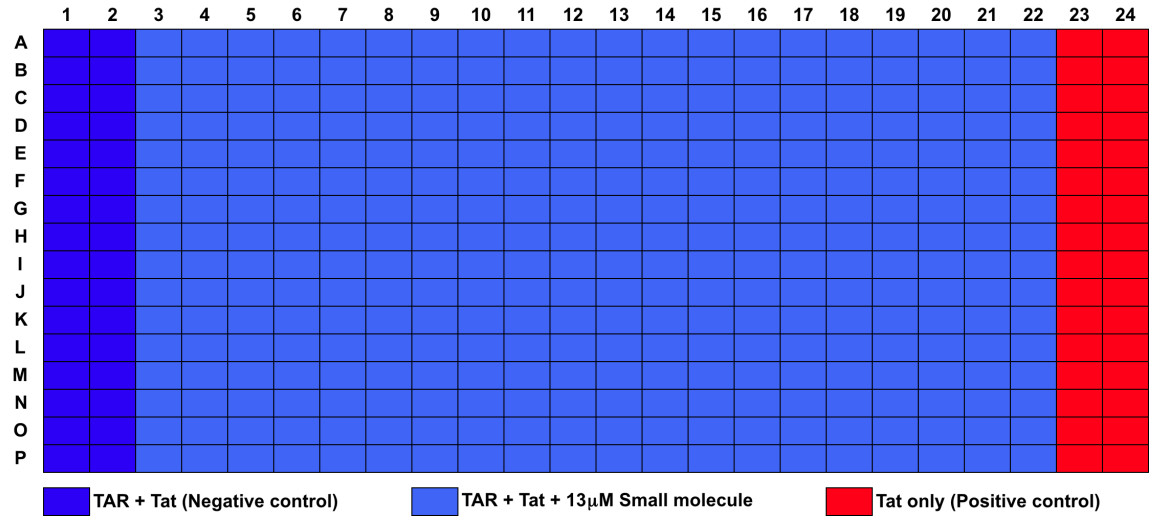

Figure 3.2. 384-well microplate layout for HTS

NMR Spectroscopy

All NMR experiments were performed at 298K on 600 MHz and 800 MHz Agilent spectrometers equipped with 5mm triple-resonance cryogenic probe. $^{13}C/^{15}N$-labeled TAR was exchanged to the NMR buffer (15 mM sodium phosphate, 25 mM NaCl, 0.1 mM EDTA at pH ~6.4). Up to 16 μL of concentrated small molecules was added to TAR to avoid the dilution of TAR and the buffer components by addition of small molecules. The aromatic SOFAST-HMQC spectra (6) were recorded and processed using nmrPipe (7) and SPARKY (8).

Cell-based Assays

The cell assays were performed in collaboration with Bieniasz lab at Rockefeller University. A recombinant HIV-1 of NL4-3 and HXB2 recombinant strains that had green fluorescence protein (GFP) in place of *Nef* gene was used in all cell assays. The virus was produced by transfecting 293T cells using PEI (6 μg DNA on 10cm plate). For

all cell assays, any data points or spreading infections that arise from toxicity of the small molecules were excluded.

For luciferase assay using TZM-bl cell line, 10,000 cells / well were seeded in a 96-well plate 24 hours before the infection. Right before infection, the small molecules were diluted in water, and then a serial dilution was done with media. The media on the TZM-bl cells was replaced with the molecule-containing media or a solvent control. HIV-1 virus was infected such that 20% of the cells were infected. A luciferase assay was performed 48 hours after infection.

For single-cycle infection with CEMx 174, and SUPT1 T cell lines, a similar protocol to TZM-bl was used except that 20,000 cells / well were seeded. After 48 hours from infection, the cells were fixed in 2% paraformaldehyde and GFP+ cells were detected by flow cytometry using a GUAVA easycyte.

For CEMx 174 spreading infections, 200,000 cells / well were seeded in 12-well plate with 1mL media with indicated concentration of compounds. The virus was added such that 0.2% of the cells were initially infected. 24 hours after infection, 100 μL of cells were taken from the cultures and were incubated with 100 μg/mL dextran sulfate to stop further replication. The cells were fixed and the GFP+ cells were detected by flow-cytometry. Every three days, cells were split 1:3 (disposing of 2/3 of the cells) with fresh media and compounds added.

Virtual Screening using ICM

The docking program Internal Coordinate Mechanics (ICM, Molsoft)(9) was used for virtual screening. ICM uses internal coordinates (bond lengths, bond angles, torsion angles, and phase angles) instead of conventional XYZ coordinates thus simplifying the

number of variables to depict the positions of atoms in space, and reducing the computation time to calculate the small molecule conformation bound to the receptor and the interaction energies. ICM utilizes the Monte Carlo simulated annealing scheme (10) to search for the optimal conformations of a small molecule and uses the following empirical scoring function (eq. 2) to estimate the binding affinity,

$$\text{Score} = A_1 \times E_{VDW} + A_2 \times N_{Rot} + A_3 \times E_{el} + A_4 \times E_{HB} + A_5 \times E_{phob} + A_6 \times E_{desolv} \qquad \text{(eq. 2)}$$

where E_{VDW} is the van der Waals interaction energy, N_{Rot} is number of rotatable bonds in a ligand, E_{el} is electrostatic interaction energy, E_{HB} is hydrogen bonding energy term, E_{phob} is the hydrophobic energy proportional to buried surface area of a ligand upon binding, and E_{desolv} is the desolvation energy that accounts for a loss of hydrogen bonding between a ligand and a solvent. The weighting terms, $A_1 - A_6$, are 1, 0.6, 0.66, 2.53, 0.67, and 0.75, respectively.

The choice of TAR structure and the ensemble-based docking approach have been described previously(11)(12). The same set of TAR dynamic ensemble comprised of 20 conformations was used in the virtual screening. For each compound, the minimum score among the 20 conformations was assigned and used for analysis.

The same exact small molecule library used in the HTS was downloaded from CCG. The protonation and energy-minimization of small molecules were completed by calculator plugins from JChem packages (ChemAxon). The protonation states of small molecules were predicted at pH 5.4, 7.4, and 9.4. The partial charges and the explicit protons on the small molecules were assigned by ICM.

Clustering Analysis

Sphere exclusion clustering is performed with JKlustor package (ChemAxon, Inc). The clustering begins with the initial structure from the small molecule library. The next structure is randomly chosen and the chemical similarity between two molecules is calculated. If the similarity is within the threshold, these two are clustered together. If not, the second cluster is generated. The similarity threshold used in this study is 80%. Compounds that are 80% or more similar are clustered together.

The chemical similarity is calculated by Tanimoto similarity coefficient (13). The chemical structures are converted into bit strings (a string of 0 and 1) that contain the patterns of the substructures in a molecule. The following equation (eq. 3) is used to calculate the Tanimoto coefficient where N is the number of bit strings in A, B, or A&B.

$$\text{Similarity} = \frac{N_{A\&B}}{N_A + N_B - N_{A\&B}} \qquad \text{(eq. 3)}$$

3.3 Results and Discussion

High Throughput Screening Targeting HIV-1 TAR

The high throughput screening of 103,146 compounds were performed using TAR-Tat displacement assay(1) in 384-format (Figure 3.3). The 2812 compounds that showed activity above three standard deviations of the negative controls were measured in triplicate to confirm their activity. Out of 2812 compounds, 267 compounds were confirmed to be active and the dose response curves with the concentration range from 1 μM to 100 μM were measured. In addition, the fluorescence intensities in the presence of Tat and the confirmed compounds only were recorded in order to discriminate the compounds that interfere with the fluorescence signals from Tat.

The majority of the confirmed hits (>75%) interfered with the fluorescence signals of Tat and they were excluded from further analysis. Finally, 22 hits were confirmed by dose response curves with pIC_{50} between 4 and 5. These 22 hits were purchased as fresh powder and re-tested for reproducibility. Unfortunately, 5 hits were no longer available from the commercial sources and excluded from analysis. The retesting of the remaining 17 hits yielded 9 compounds of which their activities were reproducible (Figure 3.4). The 8 compounds that failed to reproduce the inhibitory activity of TAR-Tat interaction could be due to the contaminations of the CCG stock compounds. These compounds could also react with DMSO over a long period of time.

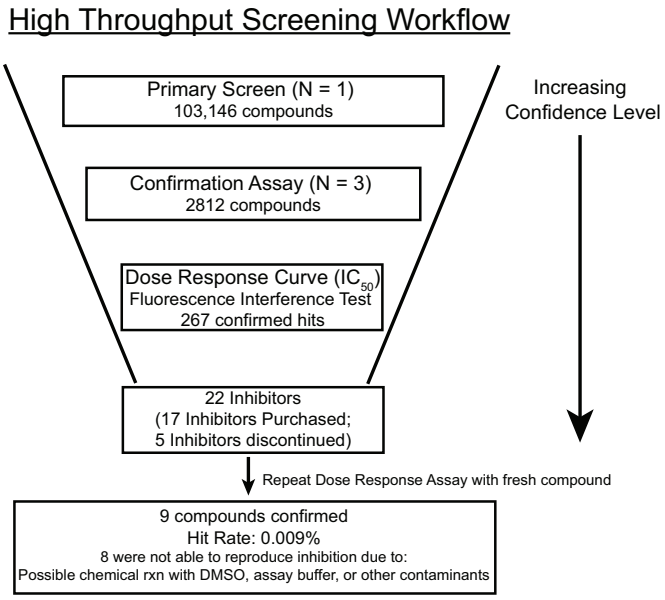

Figure 3.3 Workflow of high throughput screening

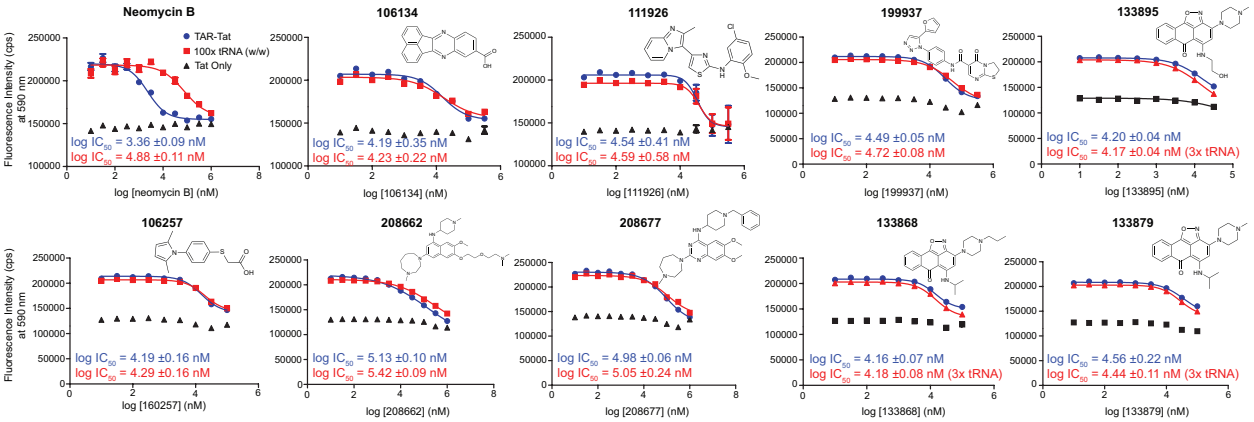

Figure 3.4 IC50 of 9 hits from HTS. The IC50s were measured in the absence (blue) and the presence of excess tRNA (red)

The dose response curves of the 9 compounds were measured in the absence and the presence of 100-fold excess tRNA (Baker's Yeast) in order to test for the specificity of the inhibitory activities (Figure 3.4). First, the average log IC50 of all 9 hits is low micromolar range (~31 μM). Strikingly, all of the 9 hits showed high specificity to TAR against the excessive tRNA (3-fold and 100-fold excess), whereas a neomycin control which is known to bind both TAR and tRNA strongly, showed deteriorated IC50 by an order of magnitude, from log IC50 of 3.36 in the absence of tRNA to 4.88 in the presence of 100x tRNA.

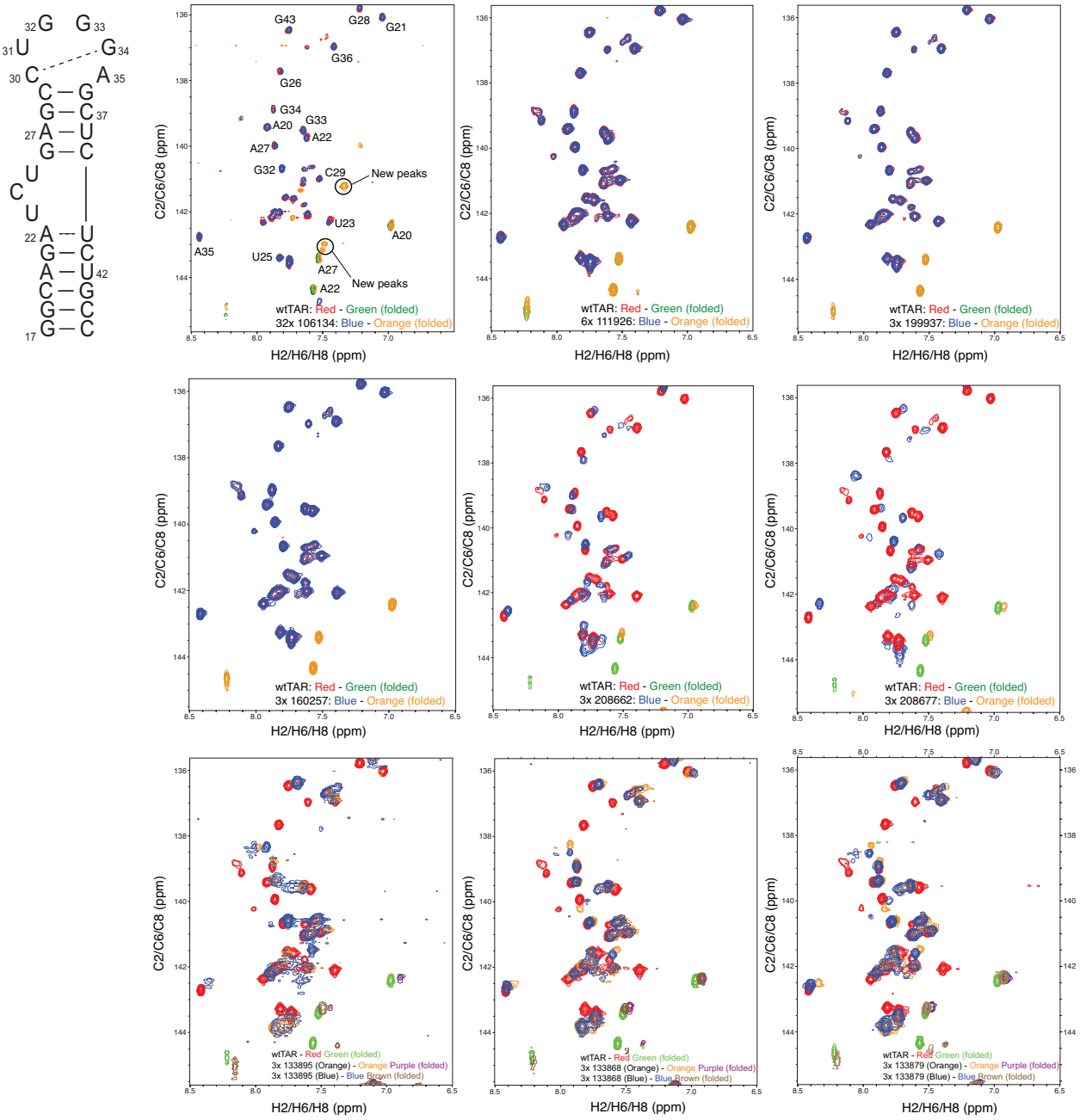

Figure 3.5 NMR titration spectra of 9 hits from HTS. Shown are the aromatic SOFAST-HMQC spectra of wtTAR in presence of each 9 hits.

Next, we subjected the 9 hits to NMR spectroscopy to further validate their binding to TAR. We measured the SOFAST-HMQC(6) on aromatic carbons to probe the

structural change of TAR upon binding of compounds (Figure 3.5). The compounds (208662, 208677, 133868, 133879, and 133895) exhibited strong chemical shift perturbations of TAR resonances, showing that the compounds bind to TAR tightly. Unfortunately, the other 4 compounds (106134, 111926, 160257, and 199937) did not induce the chemical shift perturbations of TAR resonances due to compound precipitating in the NMR tube. The precipitations were due to low concentration (< 1% by volume) of DMSO in the tube. Increasing the concentration of compound 106134 to achieve 32:1 compound-TAR molar ratio, thereby increasing the DMSO concentration to 3% (v/v), seemed to partially solubilize the compound 106134, as new resonances appeared in the NMR spectrum possibly from the free and solubilized compound 106134. Yet, there were no significant chemical shift perturbations on TAR resonances. This might imply that the poor solvation of compound 106134 might have affected binding to TAR.

During the dose-response curve measurements using the fluorescence assay, the compounds 133868, 133879, 133895 showed changes in color over time when exposed to light. The color changed from orange to blue over the course of an hour in the presence of light. When the compounds were orange, the fluorescence assay could not determine the dose-response curves due to fluorescence interference from the compounds with fluorescein and TAMRA. However, when the compounds changed to blue, the fluorescence assays yielded the dose-response curves (Figure 3.4). The NMR spectra of these compounds bound to TAR showed that the chemical shift perturbations were the same regardless of the color change. This showed that the compound binding to TAR was not affected by the color change.

The compounds 133868, 133879, 133895 are anthraquinone binders. A possible mechanism of their color change has been proposed in a previous study(14) (Figure 3.6). In that mechanism, DMSO reacts with the anthraquinone to form DMSO-anthraquinone. This complex exhibits an orange color, consistent with our observation of the compounds in 100% DMSO. When this DMSO-anthraquinone is exposed to water, the DMSO dissociates and the color of the anthraquinone becomes blue, consistent with what we have observed in the fluorescence and the NMR experiments. It is still unclear, however, how an exposure to light catalyzed the color changing reactions with DMSO.

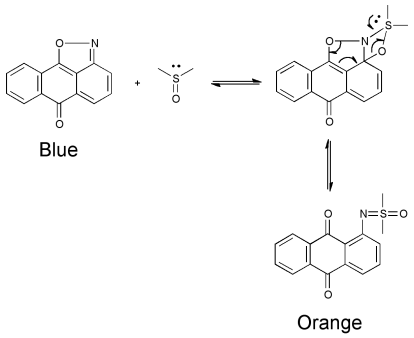

Figure 3.6 Reaction of anthraquinone with DMSO

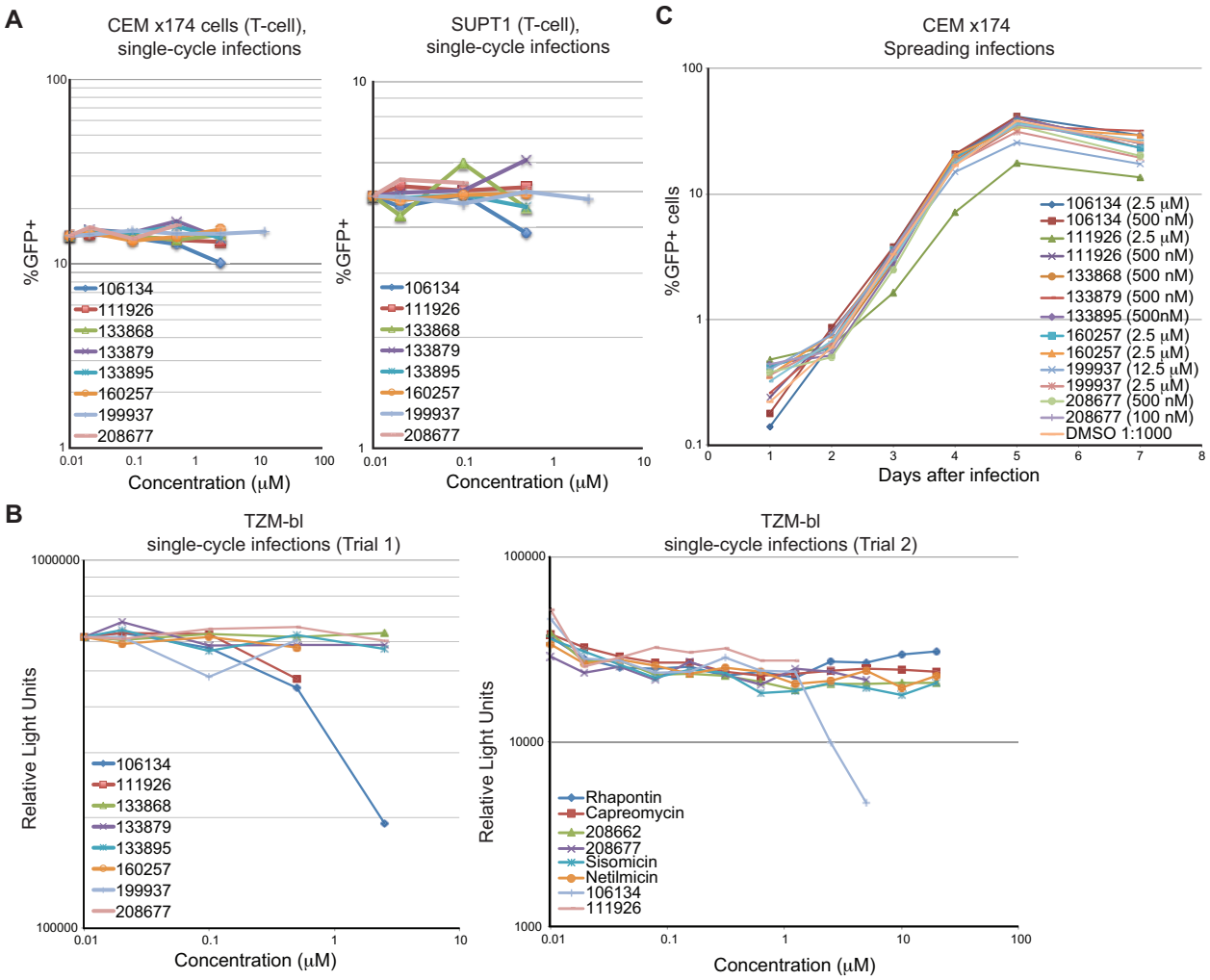

Figure 3.7 Cell Assays of HTS hits. A. Single-cycle infections on T-cell lines **B.** Single-cycle infections in TZM-bl cells. Sisomicin and netilmicin were controls. **C.** HIV-1 spreading infections in CEM x174 (T-cell).

Next, the anti-HIV activities of the 9 hits were evaluated with single-cycle infections with T-cell lines (CEM x174 and SUPT1), and a reporter cell line (TZM-bl), in collaboration with the Bieniasz lab at Rockfeller University (Figure 3.7A and Figure 3.7B, respectively). The HIV-1 strain in T-cell lines is recombinant strain with GFP in replacement of HIV-1 Nef protein. The activity of small molecules in GFP (HIV-1 Nef)

production is monitored by the GFP fluorescence. The TZM-bl has HIV-1 LTR with luciferase gene. The effect of small molecules in viral transcription is monitored by the luminescence from the luciferase. Out of 9 compounds, one compound 106134, showed promising anti-HIV activity. In T-cell lines, the compound 106134 showed 20 – 30% changes in GFP fluorescence signal at 2.5 μM concentrations in CEM x174 cells and at 500 nM concentrations in SUPT1 cells (Figure 3.7A). In TZM-bl cell line, the compound 106134 showed 70%-80% decrease in luciferase signal at 2.5 μM concentrations (Figure 3.7B). The compound 111926 was weaker, showing 23% decrease in luciferase signal at 500 nM concentrations in the first trial, but the activity was not reproducible in the second trial. It is possible that the compound 111926 is inactive or very weak inhibitor. Interestingly, the compound 111926 showed a small inhibitory activity in HIV-1 replication over a week (Figure 3.7C), even though it was inactive or very weak inhibitor in the single-cycle infection experiments. It might suggest that the compound 111926 has off-targets that inhibit the HIV-1 replication.

Virtual Screening Targeting HIV-1 TAR

The same small molecule library as the HTS (103,146 compounds) was used to screen small molecules against the dynamic ensemble of 20 TAR conformers(12). In order to compare the docking with the HTS without a bias, the workflow of the virtual screening (VS) was similar to the HTS (Figure 3.8). After the primary screening, the top 2812 compounds were subjected to the confirmation stage where they were docked in triplicates. The top 267 compounds that showed docking scores with standard deviation less than 3 were subject to the rigorous docking by increasing the sampling time. This rigorous docking is equivalent to the dose-response curve measurement in HTS. The

top 17 commercially available compounds were purchased and tested with the TAR-Tat displacement assay.

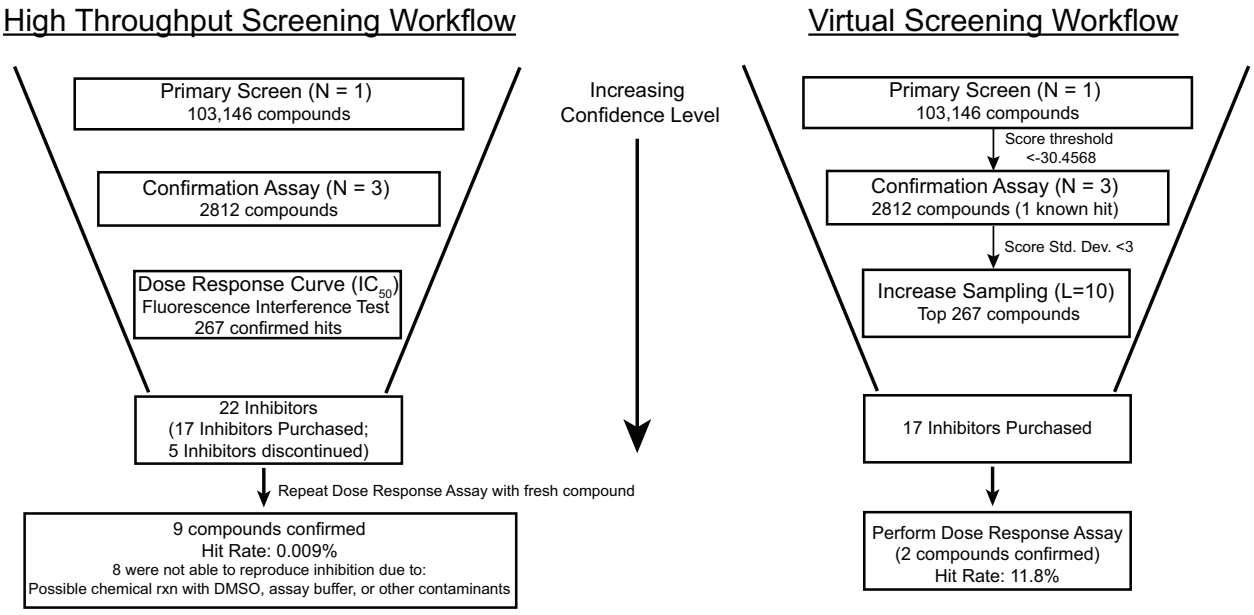

Figure 3.8 Comparison of screening workflow between HTS and VS

Among the 17 final hits from VS, two compounds, rhapontin and capreomycin, showed activity with IC50 of 178 μM and 31 μM, respectively (Figure 3.9). The hit rate of the virtual screening was ~12%, which was comparable to the previous virtual screening (12). This results in the enrichment rate of >1300 by VS when only accounting for experimentally tested compounds. The predicted bound pose of the rhapontin showed that it would bind to the apical loop of TAR, while the bound pose of the capreomycin predicted that it would interact with both bulge and the apical loop (Figure 3.10). Indeed, NMR titration of capreomycin in 6:1 molar ratio with TAR resulted in significant chemical shift perturbations around the bulge, suggesting that it may bind to the bulge. The binding of rhapontin in 6:1 molar ratio did not perturb the

chemical shifts of TAR, possibly due to weak binding affinity. Unfortunately, capreomycin and rhapontin did not show activities in cells (Figure 3.7B).

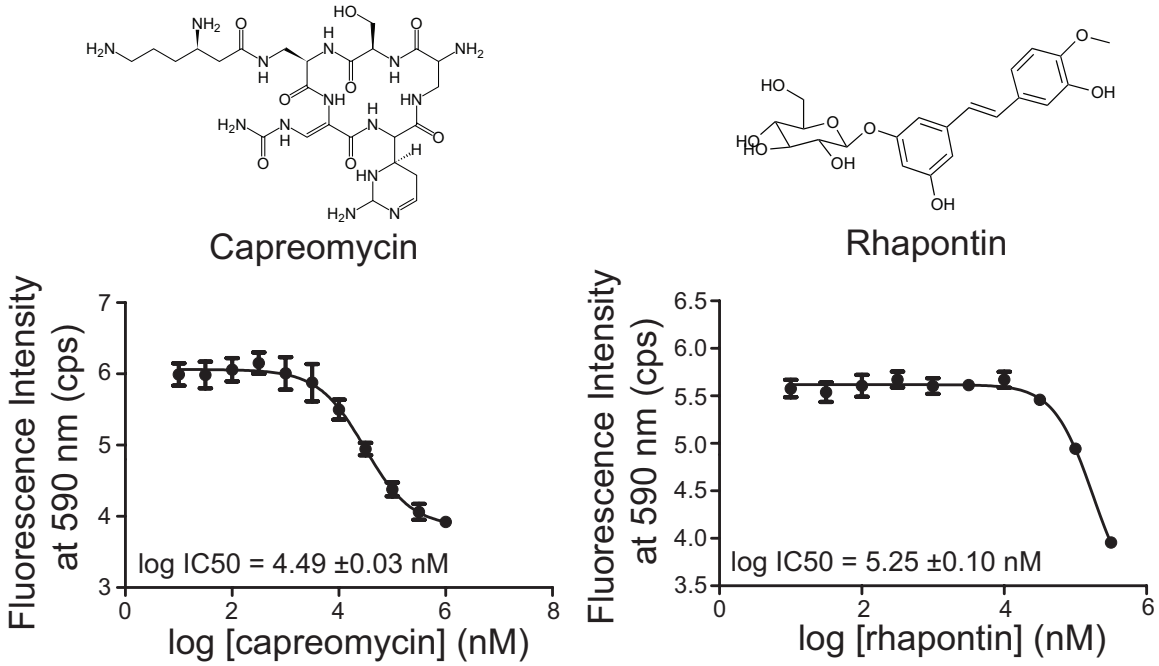

Figure 3.9 Dose-response curves of hits from virtual screening

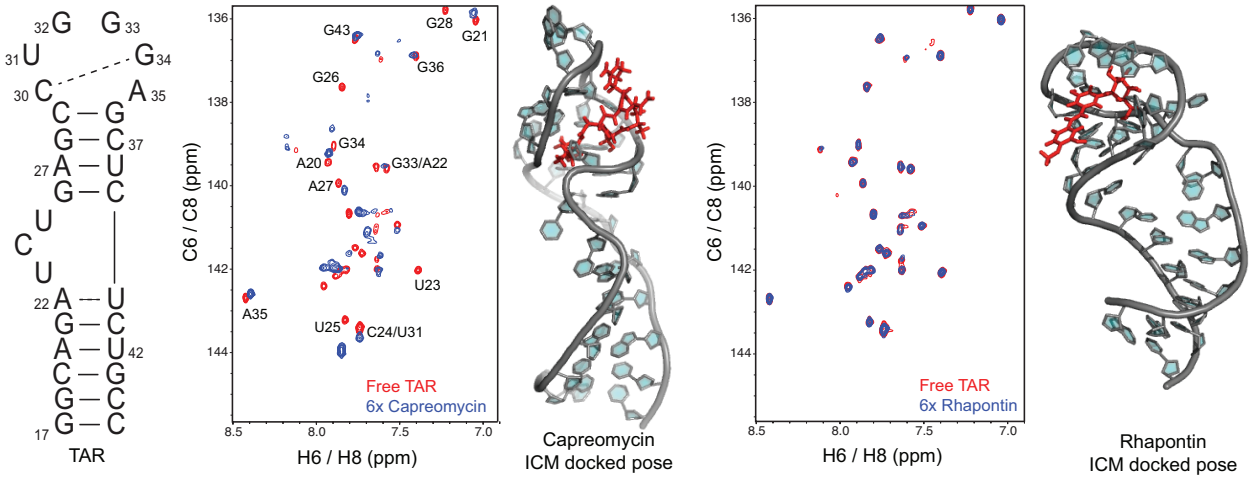

Figure 3.10 NMR chemical shift perturbations and ICM predicted bound pose of hits from virtual screening

Comparison of Virtual Screening and High Throughput Screening

The structures of the small molecule hits from HTS and the hits from VS were very different. None of the 9 hits from the HTS were selected as hits by the virtual screening, and likewise capreomycin and rhapontin were not selected as hits by the HTS. It is possible that HTS did not select capreomycin in the screening, because capreomycin was not soluble in DMSO. However, the ICM should have selected the 9 hits from the HTS. Looking at the histogram of ICM scores of the entire compound library (Figure 3.11), the 9 hits were all scored near the mean of the score distribution (ICM score ~ -21 kcal/mol). The best-scored compound was ranked at 2,089th, and the least scored compound was ranked at 89,440th. Interestingly, the compound 106134, which showed anti-HIV activity in cells, were ranked 89,440th.

It is questionable whether the ICM scoring function was indeed the problem for the false negatives. In order to address this question, we have selected two best-scored compounds from the dose-response measurements in HTS. These compounds were tested by NMR due to fluorescence interference (Figure 3.12). The first compound 135485, which was ranked 14th by ICM, was not available commercially and excluded from analysis. The second compound, Rutin, was ranked 33rd by ICM, but it did not show binding to TAR at 6:1 molar ratio by NMR. The third compound, compound 133905, was ranked 639th by ICM, and it did bind to TAR strongly at 6:1 molar ratio by NMR. Looking at the molecular structure of the compound 133905, it shares the same scaffold with the compound 133868, 133879, and 133895, which were ranked 55,369th, 7,008th, and 5,485th by ICM, respectively. These results suggested that the poor enrichment of active compounds by ICM could be due to many high-scored inactive compounds that constitute the great portion of the chemical library.

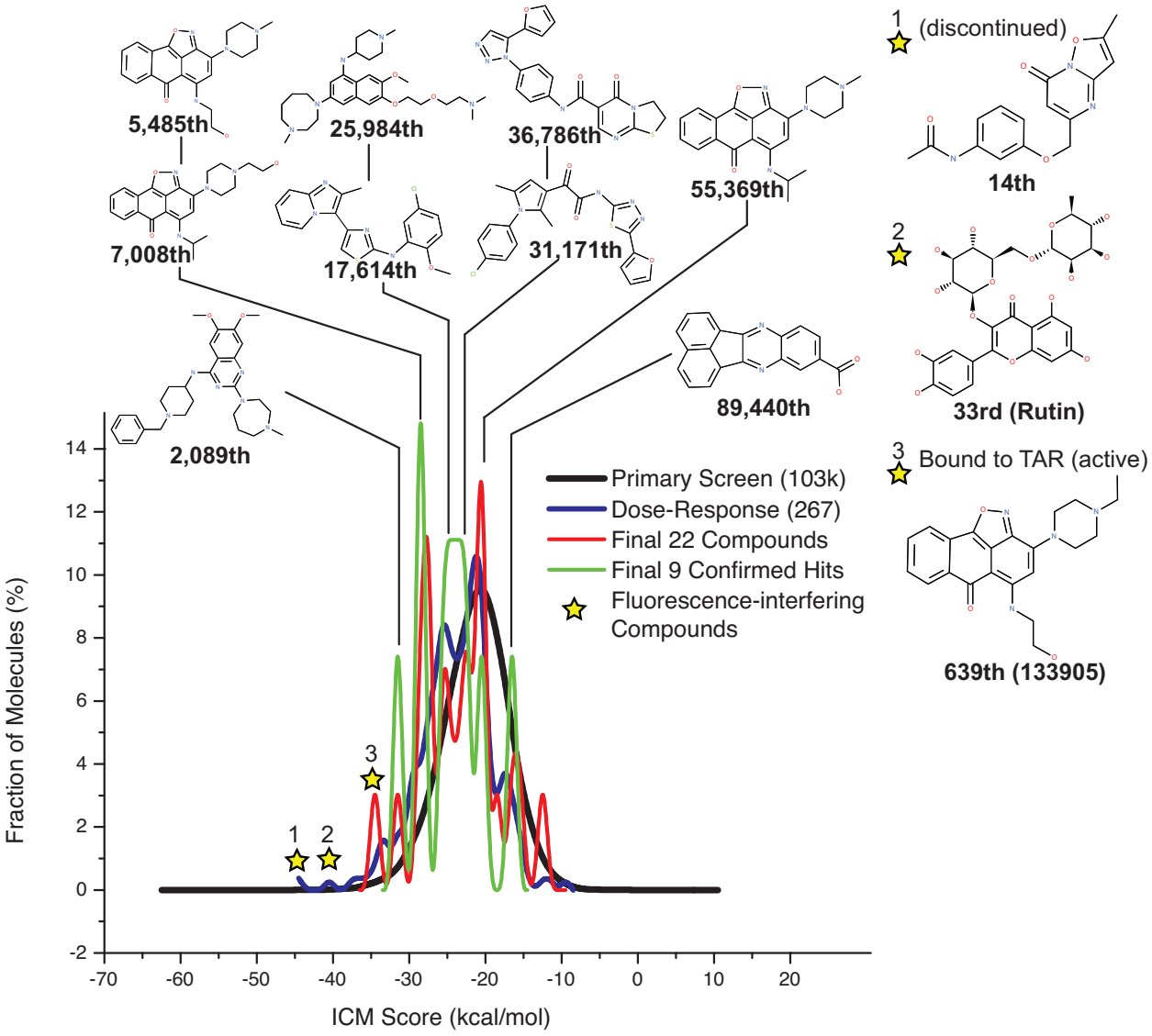

Figure 3.11 Histogram of ICM scores and distribution of 9 hits from HTS

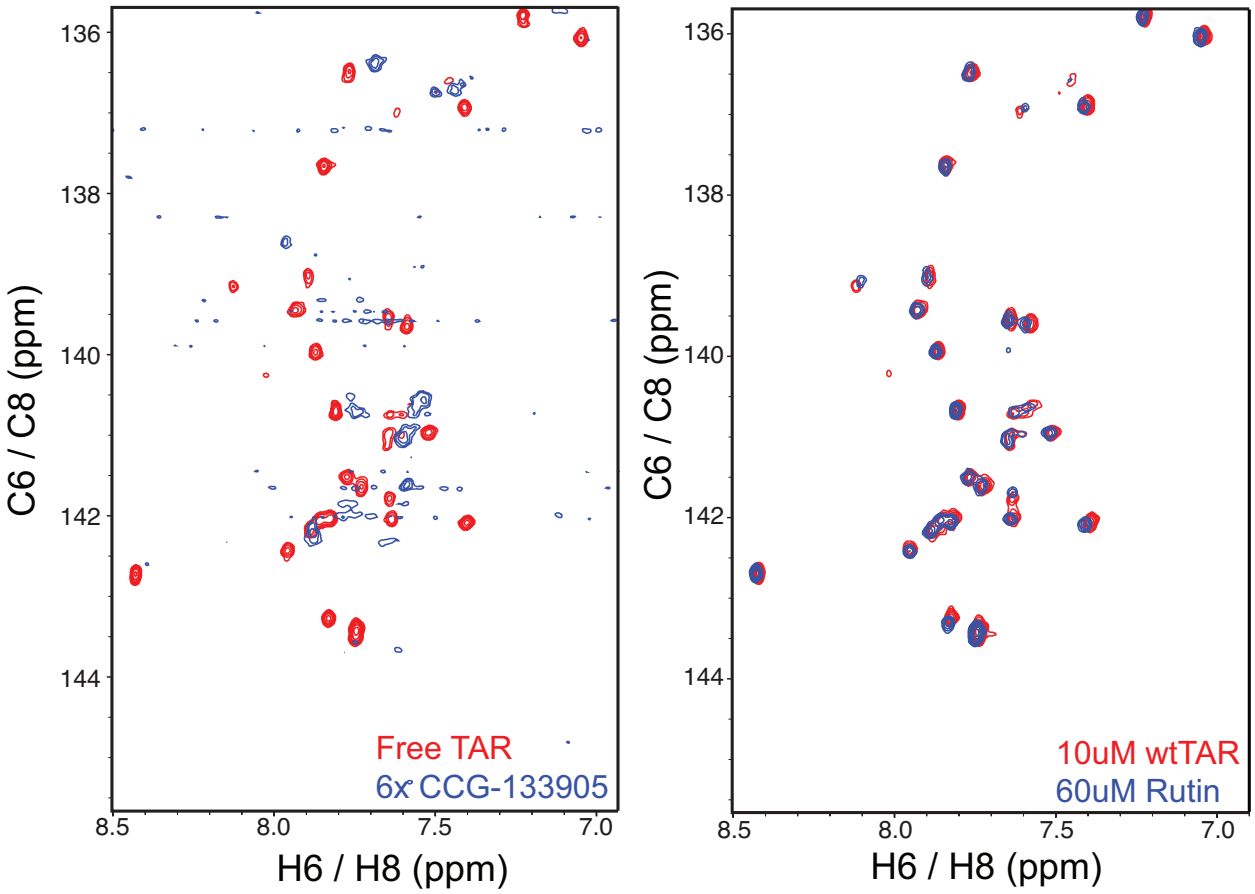

Figure 3.12 NMR titrations of compound 133905 and Rutin

Given the broad distribution of compounds that have similar scaffolds, we hypothesized that by clustering the compounds by chemical similarity, the enrichment rate of ICM would increase, because putatively inactive compounds will be clustered, reducing the number of inactive compounds ranked among the actives. Here, we assume that the chemically similar compounds also share the similar activity. Using the sphere exclusion clustering, we clustered the entire small molecule library with chemical similarity threshold 80% or more (Figure 3.13). This means that each cluster will contain 80% or more similar compounds to each other. The 80% threshold was chosen to optimize the similarity within each cluster. The entire library of 103,146

compounds was reduced to 25,446 representative clusters (4-fold decrease). There are 7 representative clusters for the hits from HTS, because the chemical structures of compound 133895, 133879, and 133868 were clustered.

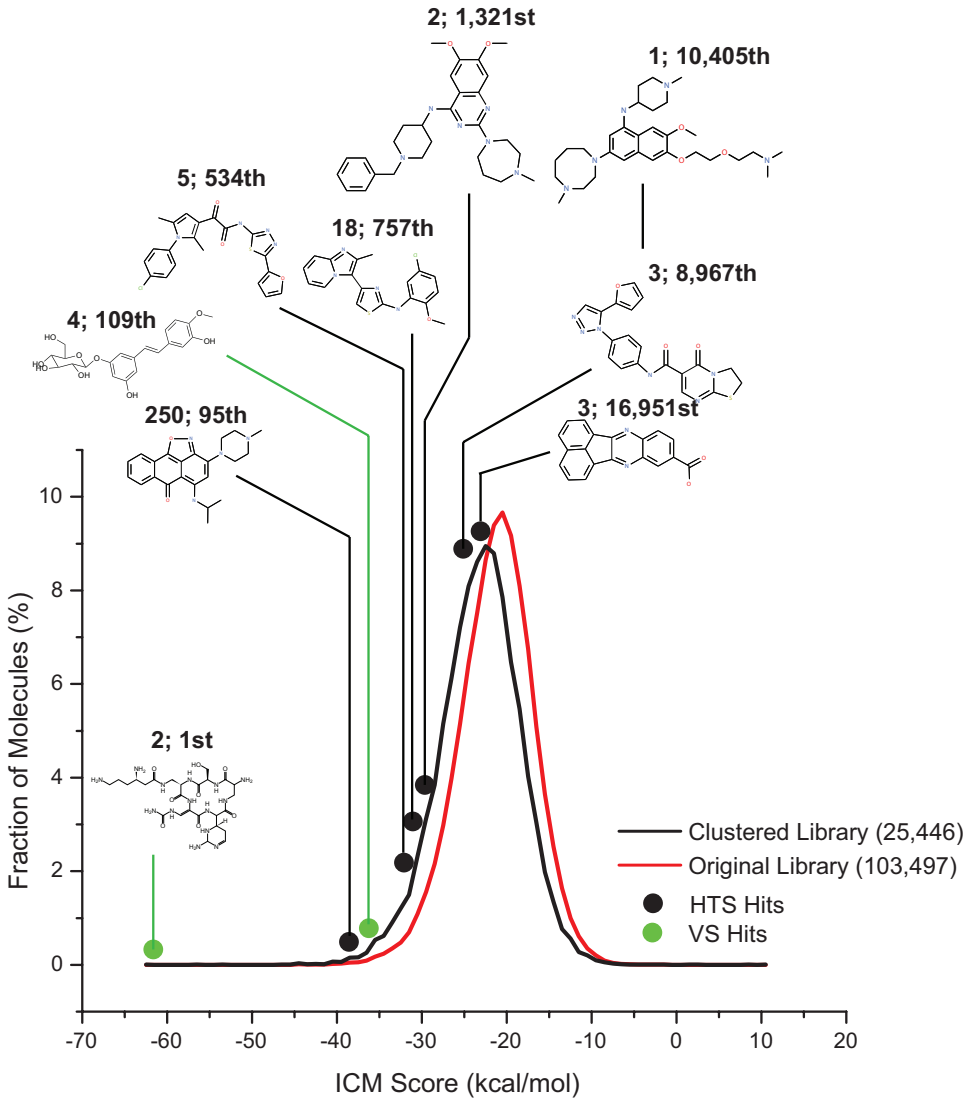

Figure 3.13 ICM score histogram of the reduced chemical library by clustering. The numbers above the molecule show the number of similar compounds in a cluster and the ICM score ranking of a cluster.

The histogram of best ICM score from each cluster was compared with the original ICM score distribution. The clustering improved the mean ICM score of the histograms marginally from ICM score of -21 kcal/mol to -22 kcal/mol, but there were notable changes in distributions of active compounds from HTS. For example, the ICM score of compound 106134 improved from -16 kcal/mol to -22 kcal/mol and the ICM score of the compound 133868 improved from -20 kcal/mol to -39 kcal/mol. The improvements on ICM scores were more significant for a cluster of similar compounds that had wide range of ICM scores as shown in the case of anthraquinones (compound 133895, 133879, and 133868). The enrichment rate calculated by the receiver operating characteristic curve (ROC)(15) shows that the clustering improved the enrichment by ~20% based on the areas under the curves (Figure 3.14).

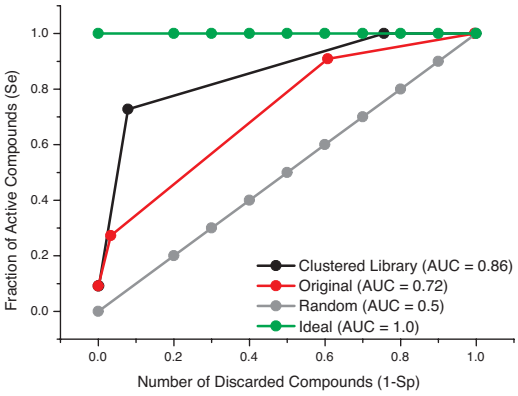

Figure 3.14 Improvement on ICM enrichment rate by clustering.

3.4 Conclusion

In the present study, we have identified 11 small molecule (IC50 ~ 30 – 170μM) inhibitors of TAR-Tat interaction from HTS and VS. One of the compounds showed small but promising anti-HIV activity despite high cytotoxicity. The hits from HTS and VS did not overlap at all, suggesting that the HTS and VS are complimentary to each other. While the VS is more efficient with enrichment rate larger than 1300 compared to HTS, ICM poorly enriched the hits from HTS. Clustering the chemical library improved the enrichment rate by ICM, suggesting that there are many false positives (inactive compounds that are highly scored) that share the similar chemical scaffolds. Clustering

can also reduce the number of small molecules that the HTS needs to screen. Therefore, combining HTS and VS will boost the efficiency of the drug discovery.

3.5 References

1. Matsumoto C, Hamasaki K, Mihara H, & Ueno A (2000) a high-throughput screening utilizing intramolecular fluorescence resonance energy transfer for the discovery of the molecules that bind hiv-1 tar rna specifically. *Bioorganic & Medicinal Chemistry Letters* 10(16):1857-1861.
2. Calnan BJ, Biancalana S, Hudson D, & Frankel AD (1991) Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. *Genes & Development* 5(2):201-210.
3. Shojania S & O'Neil JD (2006) HIV-1 Tat Is a Natively Unfolded Protein: THE SOLUTION CONFORMATION AND DYNAMICS OF REDUCED HIV-1 Tat-(1,Äì72) BY NMR SPECTROSCOPY. *Journal of Biological Chemistry* 281(13):8347-8356.
4. Loret EP, Georgel P, Johnson WC, & Ho PS (1992) Circular dichroism and molecular modeling yield a structure for the complex of human immunodeficiency virus type 1 trans-activation response RNA and the binding region of Tat, the trans-acting transcriptional activator. *Proceedings of the National Academy of Sciences* 89(20):9734-9738.
5. Zhang J-H, Chung TDY, & Oldenburg KR (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *Journal of Biomolecular Screening* 4(2):67-73.
6. Sathyamoorthy B, Lee J, Kimsey I, & Al-Hashimi HM (2014) Development and application of aromatic 13C-1H SOFAST-HMQC NMR experiment for nucleic acids. *In Preparation*.
7. Delaglio F, *et al.* (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. (Translated from English) *J Biomol NMR* 6(3):277-293 (in English).
8. Goddard TD & Kneller DG (SPARKY 3 (University of California, San Francisco).
9. Abagyan R, Totrov M, & Kuznetsov D (1994) ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry* 15(5):488-506.
10. Abagyan R & Totrov M (1994) Biased Probability Monte Carlo Conformational Searches and Electrostatic Calculations for Peptides and Proteins. *Journal of Molecular Biology* 235(3):983-1002.

11. Frank AT, Stelzer AC, Al-Hashimi HM, & Andricioaei I (2009) Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. *Nucleic Acids Research* 37(11):3670-3679.
12. Stelzer AC, *et al.* (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* 7(8):553-559.
13. Rogers DJ & Tanimoto TT (1960) A Computer Program for Classifying Plants. *Science* 132(3434):1115-1118.
14. Sutter P & Weis CD (1982) Ring opening reactions of 6H-anthra[1,9-cd]isoxazol-6-ones and related compounds. *Journal of Heterocyclic Chemistry* 19(5):997-1011.
15. Triballeau N, Acher F, Brabet I, Pin J-P, & Bertrand H-O (2005) Virtual Screening Workflow Development Guided by the ‚ÄúReceiver Operating Characteristic‚Äù Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *Journal of Medicinal Chemistry* 48(7):2534-2547.

Chapter 4

A Sparsely Populated Transient State Gives Rise to Long-Range Correlated Changes in RNA Secondary Structure

4.1 Introduction

Many ribonucleic acids (RNA) do not fold into a single static conformation but rather can populate a wide range of conformation along a rugged free energy landscape(1, 2). Cellular cues ranging from changes in physiochemical parameters such as temperature and pH to the binding of proteins, RNAs, and ligands have been shown to preferentially stabilize different RNA conformations along this free energy landscape, resulting in dynamic change in structure that play crucial roles in the catalytic activity of ribozymes(3), regulatory activity of riboswitches(4) and other RNA-based switches(5-8), and in the hierarchical assembly and disassembly of ribonucleic acid- protein complexes(9-11).

One of the most ubiquitous and fundamental mode of RNA functional dynamics involves large-scale changes in secondary structure, typically triggered by external co-factors(12) or with the aid of ATP-dependent chaperones(13, 14), sometimes during co-transcriptional folding(7, 15), that serve to expose or sequester key regulatory elements. Recently, NMR R1ρ relaxation dispersion experiments(16-19) in concert with mutagenesis(20) have helped uncover a new mode of RNA secondary structure dynamics involving much more rapid and localized transitions in secondary structure representing excursions away from the energetically favorable ground state (GS)

towards low populated (typically populations <15 %) and short-lived (lifetime < milliseconds) species often referred to as excited states (ES). These transitions feature subtle changes in base-pairing in and around non-canonical motifs such as bulges, internal loops, and apical loops(20). As for larger-scale secondary structural transition, transitions towards RNA ESs can sequester functionally important residues via base-pairing that otherwise have to be available in order to interact with cellular factors or promote ATP-independent large-scale changes in secondary structure(18, 19). Because these localized changes in secondary structure occur at rates that are two to four orders of magnitude faster than conventional RNA secondary structural transitions, they can potentially meet unique demands in RNA-based regulatory functions (20).

Using NMR R1ρ relaxation dispersion experiment(21-23), we recently reported(20) an excited state for the apical loop of the transactivation response element (TAR) from the human immunodeficiency virus type-1 (HIV-1) RNA(24) that has a population of ~13% and exceptionally short lifetime of ~45 microsecond. TAR is a regulatory viral RNA located at the 5′-end of the HIV retroviral genome that plays many roles in HIV viral replication, including transcription elongation(25-28), viral translation(29, 30), dimerization(31, 32), packaging(32, 33), and viral latency(34, 35). TAR contains two motifs that play essential roles in protein recognition; a trinucleotide bulge and a hexanucleotide apical loop that are separated by a four base-pair helix(36). The TAR bulge and apical loop also serve as binding sites for small molecules that are being developed as anti-HIV therapeutics(37-40). In the ground state, the apical loop forms a conformation that exposes residues C30, U31, G32, and A35 so that they are exposed and available to interact with binding partners(36, 41-43). The ES (which we will refer to as ES1) features a more compact apical loop structure in which C30, U31, G34, and A35 are sequestered through formation of A^+-C and U31-G34 trans-wobble

base-pairs(20). Importantly, mutations that stabilize the ES1(20) are correlated to the reduced binding affinity to Tat/cyclin T1 proteins as well as the reduced efficiencies of transcriptional activation(44), implying a potential regulatory role.

During our studies, we also obtained evidence for a slower process directed toward a second distinct excited state (which we refer to as ES2) involving apical loop residues G33 and A35. This ES has a much lower population (p_B) of ~0.4% and longer lifetime (τ) of ~2 milliseconds. It is remarkable that A35 simultaneously senses the two ESs; with the sugar C1′ sensing the fast process toward ES1 and base C8 sensing the slower process toward ES2. Here, we perform a broader set of NMR relaxation dispersion experiments aimed at characterizing the second transient state (ES2) in HIV-1 TAR. Our results reveal a rearrangement in secondary structure that results in correlated changes in secondary structure from the tip of the apical loop toward the lower stem of HIV-1 TAR. This ES can possibly explain functional data hinting at allosteric communication between the TAR bulge and apical loop. These results provide fundamental new insights into transient structure landscape of RNA and suggest a new mechanism for RNA based allostery involving transient states.

4.2 Materials and Methods

Preparation and NMR resonance assignment of labeled and unlabeled RNA

RNA samples (wt TAR, TAR-UUCG, TAR-G28U, and TAR-Δbulge) were prepared by *in vitro* transcription using T7 RNA polymerase (Takara Mirus Bio, Inc.), uniformly $^{13}C/^{15}N$-labeled nucleotide triphosphates (ISOTEC, Inc., Cambridge Isotope Labs) and synthetic DNA templates (Integrated DNA Technologies, Inc.) containing the T7 promoter and sequence of interest. All RNAs were purified by 20 % (w/v) denaturing polyacrylamide gel electrophoresis, using 8M urea and 1x TBE. The RNA was electro-

eluted from the gel in 20 mM Tris buffer at pH 8 followed by ethanol precipitation. The RNA pellet was dissolved in water, annealed by heating to 95 °C for 5 minutes and rapid cooling on ice and exchanged into NMR buffer (15 mM sodium phosphate, 0.1 mM EDTA, and 25 mM NaCl at pH 6.4) by multiple times using a Centricon Ultracel YM-3 concentrator (Millipore Corp.). Unlabeled RNA samples (TAR-G28U) were purchased from Dharmacon (Thermo Fisher Scientific) and dissolved in water for annealing as described above and exchanged into NMR buffer (15 mM sodium phosphate, 0.1 mM EDTA, 25 mM NaCl at pH 6.4). For NMR experiments in ~100% D_2O, the sample in the NMR buffer was lyophilized for 24 hours first, followed by adding 99.99% D_2O (Cambridge Isotope Labs) and lyophilized again for 24 hours. The lyophilized pellet was finally dissolved in the same volume of D_2O as the NMR buffer. For the argininamide (ARG) studies, 1 μL of the concentrated ARG solution in the NMR buffer at pH 6.4 was added to the wt TAR to achieve final 16:1 ARG:wt TAR molar ratio.

All NMR assignment experiments were performed on Bruker Avance 600 MHz NMR spectrometer or Agilent 600 MHz NMR spectrometer, both of which are equipped with a 5 mm triple-resonance cryogenic probe. The NMR assignments of wt TAR, TAR-UUCG, TAR-Δbulge were published previously. The resonances of TAR-G28U were assigned using 2D HCCNH-TOCSY(45) that correlates imino protons of G and U to their C8 and C6, respectively, 2D HCN(46), 2D ^{15}N-edited NOESY (150ms mixing time), and 2D NOESY experiments(47) (200ms and 300ms mixing times). The NOESY experiments were also performed with TAR-G28U in ~100% D_2O. Except for ^{15}N-edited NOESY which was performed at 10.5^{o}C, all experiments were performed at 25^{o}C.

Carbon and nitrogen R1ρ relaxation dispersion

All NMR relaxation dispersion experiments were performed on a Bruker Avance 600 MHz NMR spectrometer equipped with a 5 mm triple-resonance cryogenic probe. Experiments were performed at 25°C, and 35°C using uniformly $^{13}C/^{15}N$ labeled TAR samples. Rotating frame carbon(21) and nitrogen(48) R1ρ relaxation dispersion data was measured using a 1D acquisition scheme that extends the sensitivity to chemical exchange into millisecond timescales relative to conventional 2D relaxation dispersion methods. On- and off-resonance relaxation dispersion data was recorded at various spinlock and at various offset frequencies and spinlock powers, respectively, listed in Table S1. For each nucleus, the maximum relaxation delay was determined by the signal intensity being 30% of the initial signal without the delay. When our technical limit of the relaxation delays for carbon (55ms) and nitrogen (80ms) were reached without reducing the signal intensity to 30%, these limits became the maximum relaxation delays regardless of the signal intensity. Between the 0 ms and the maximum relaxation delay, 6-10 relaxation delays were used to characterize the mono-exponential decay curve. Two duplicate delays were also included for error analysis. Data were processed using nmrPipe(49) and the R1ρ values were computed by fitting the resonance intensities with mono-exponential decays using Mathematica 6.0 script(50) (Wolfram Research, Inc., Champaign, IL). Errors in R1ρ were determined using the standard Monte Carlo simulations with 500 iterations. Measured on- and off-resonance relaxation dispersion data was fitted to the two-state Laguerre(51) equation in Origin 8.5.1 (OriginLab).

Laguerre:

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta + \frac{\sin^2\theta p_{GS} p_{ES} \Delta\omega^2 k_{ex}}{\omega_{GS}^2\omega_{ES}^2/\omega_{eff}^2 + k_{ex}^2 - \sin^2\theta p_{GS} p_{ES}\Delta\omega^2\left(1 + \frac{2k_{ex}^2(p_{GS}\omega_{GS}^2 + p_{ES}\omega_{ES}^2)}{\omega_{GS}^2\omega_{ES}^2 + \omega_{eff}^2 k_{ex}^2}\right)} \quad \text{(Eq. 1)}$$

$$\omega_{eff}^2 = \Delta\Omega^2 + \omega_1^2,\ \omega_{GS}^2 = (\Omega_{GS} - \omega_{rf})^2 + \omega_1^2,\ \omega_{ES}^2 = (\Omega_{ES} - \omega_{rf})^2 + \omega_1^2,$$

where R_1 and R_2 are the intrinsic longitudinal and transverse relaxation rates, respectively, (assumed to be identical for GS and ES species), Ω is the resonance offset from the spinlock carrier, ω_{eff} is the spinlock strength; $\Omega = \tan(\omega_{eff}/\Omega_{ave})$, $\Delta\omega = \Omega_{ES} - \Omega_{GS}$, $\Omega_{obs} = p_{GS}\Omega_{GS} - p_{ES}\Omega_{ES}$, where p_{GS} (p_{ES}) is the ground (excited) state fractional population ($p_{GS} + p_{ES} = 1$); $k_{ex} = k_1 + k_{-1}$ is the exchange rate constant for a two-state equilibrium, where $k_1 = p_{ES}*k_{ex}$ and $k_{-1} = p_{GS}*k_{ex}$ are the forward and reverse rate constants, respectively. Note that for p_{ES} <1%, $\Omega_{obs} \sim \Omega_{GS}$.

Model selection was carried out using F-test, which uses chi-square (χ^2 – applying the Levenberg –Marquardt minimization algorithm) to determine the feasibility of a model (e.g. 2-state Laguerre equation) versus a more complex model (i.e. 3-state equation) expanded from the first model.

Note that G28-C8 also showed a sign of relaxation dispersion, but it was excluded from our analysis because we were not able to retrieve the p_B and the exchange rate constant (k_{ex}) reasonably from fitting with the two-state Laguerre equation (Figure 4.1).

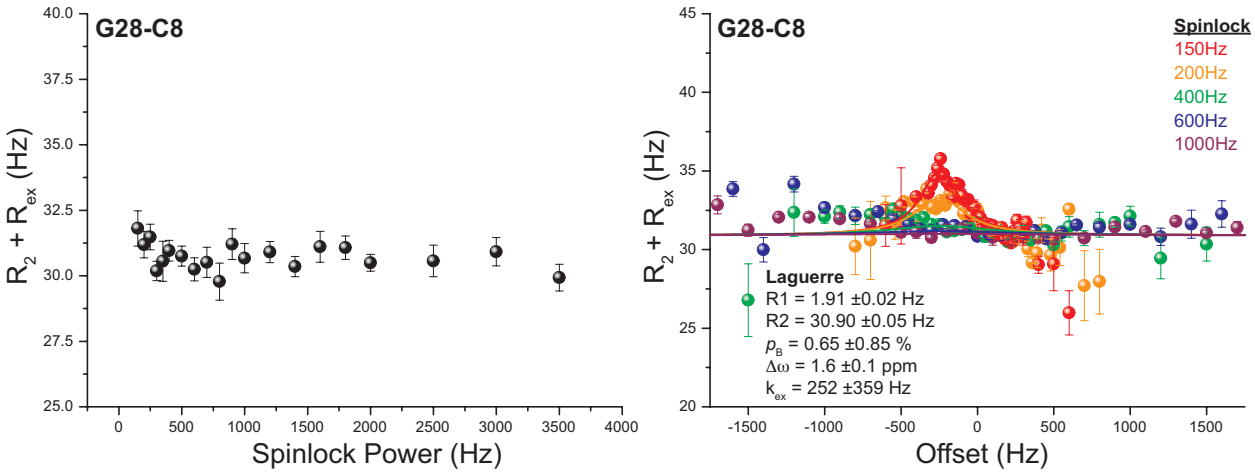

Figure 4.1 R1ρ relaxation dispersion profiles of G28-C8 in wt TAR

Thermodynamic analysis

The free energy difference between the GS and ES was calculated by the modified van't Hoff equation.(52)

$$\ln\left(\frac{k_i(T)}{T}\right) = \ln\left(\frac{k_B \kappa}{h}\right) - \frac{\Delta G_i^T(T_{hm})}{RT_{hm}} - \frac{\Delta H_i^T}{R}\left(\frac{1}{T} - \frac{1}{T_{hm}}\right) \qquad \text{(Eq. 2)}$$

where k_i (i = 1, -1) is the forward and backward rate constant of the two-state equilibrium, ΔG_i^T and ΔH_i^T are the free energy and the enthalpy of activation, respectively. R and T are the gas constant and the temperature in Kelvin, respectively. k_B and h are Boltzmann's constant and Planck's constant, respectively. κ is the transmission coefficient which is assumed to be 1 in the pre-exponential factor of Eyring's theory. T_{hm} is the harmonic mean of the experimental temperatures calculated by $T_{hm} = n/\sum_{i=1}^{n}(\frac{1}{T_i})$.

MC-fold predictions of RNA secondary structure

All RNA secondary structures except the full-length TAR sequences were predicted based on sequence using the program MC-Fold(53) with standard input options. The full-length TAR sequences were predicted using the same program but with the best 500 structures among the 30% sub-optimal structures.

H-Factor analysis of NMR chemical shifts

The H-factor (manuscript in preparation) analysis was performed with ^{13}C chemical shifts of wt TAR and the chemical shifts from the relaxation dispersion (Δω), using the in-house program.

4.3 Results and Discussion

Observation of chemical exchange involving the HIV-1 TAR bulge and apical loop.

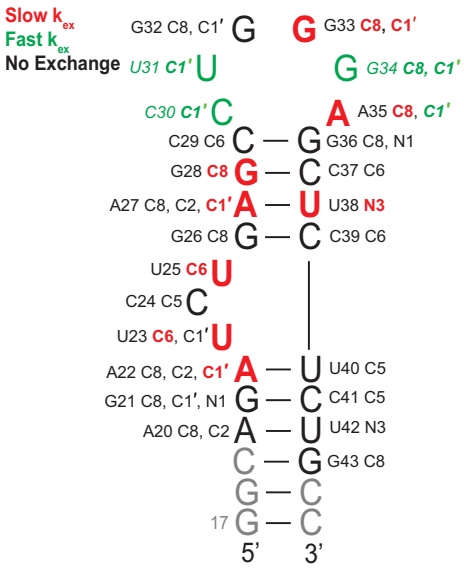

Figure 4.2 Secondary structure of HIV-1 TAR. Sites showing slow, fast, and no chemical exchange are highlighted in red, green, and black respectively.

We previously showed that G33-C8, G33-C1′ and A35-C8 experience slow chemical exchange towards an excited state (ES2, p_B = ~0.4% and τ = ~2ms) that is distinct from the ES1 (p_B = ~13% and τ = ~45μs), which is sensed by virtually all other apical loop residues(20) (Figure 4.2). To obtain further insights into ES2, we performed a broader set of carbon and nitrogen NMR R1ρ relaxation dispersion experiments targeting residues throughput HIV-1 TAR. Unexpectedly, we found evidence for slow chemical exchange in several TAR resonances located in the upper stem (A27-C1′ and U38-N3), bulge (U23-C6 and U25-C6), and lower stem (A22-C1′) (Figure 4.2).

Several residues showed no evidence for chemical exchange, including A20, G21, U42, and G43 in the lower stem and G26, C29, G36, C37, and C39 in upper stem (Figure 4.3). The two-state analysis ($A \underset{k_B}{\overset{k_A}{\rightleftharpoons}} B$) of these newly measured R1ρ relaxation dispersion data (Figure 4.4) using the Laguerre equation(51) revealed an exchange process directed to distinct ES that has a smaller population (p_B ~ 0.26-0.66%) and longer lifetime (τ = 1.2 – 3.1 ms) and that is very similar to that observed for the slow exchanging apical loop residues (G33-C1′, G33-C8, and A35-C8) where p_B = ~0.4 % and τ = ~2 ms. Indeed, based on the global fit (Table 4.1), the R1ρ data collected for all of the above residues could be combined into a single two-state fit with p_B=0.40 ±0.05 %, lifetime of τ = 2.1 ±0.3 ms.

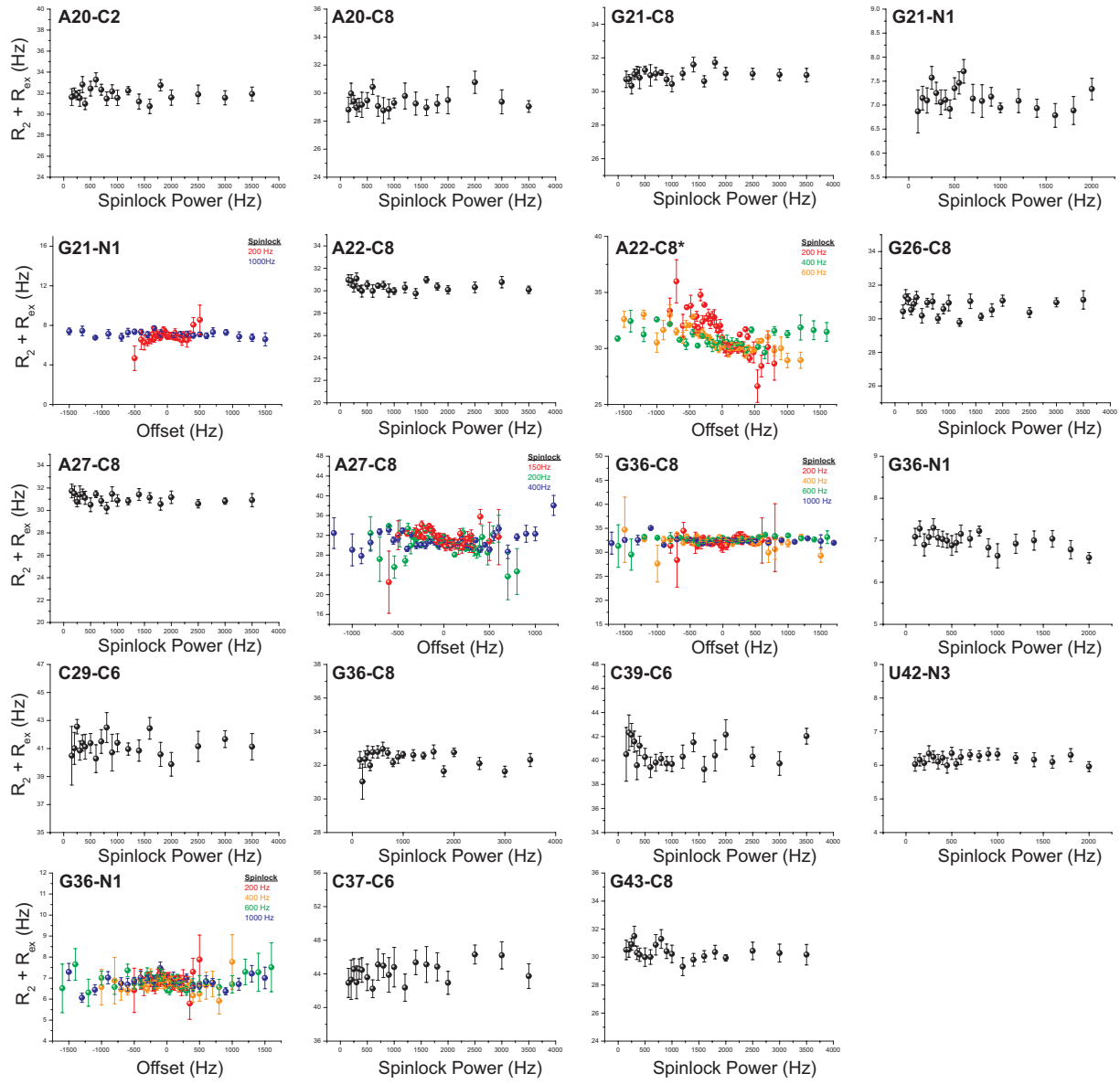

Figure 4.3 R1ρ relaxation dispersion profiles of wt TAR that showed no dispersion at 25 °C. The plot with the asterisk is the off-resonance profile of A22-C8 that was poorly fitted to the two-state Laguerre equation (Eq.1) and likely to be a noise in R1ρ values. Error bars represent experimental uncertainty (one s.d.) as determined from propagation of errors obtained from mono-exponential fitting of duplicate sets of R1ρ data and analysis of signal-to-noise.

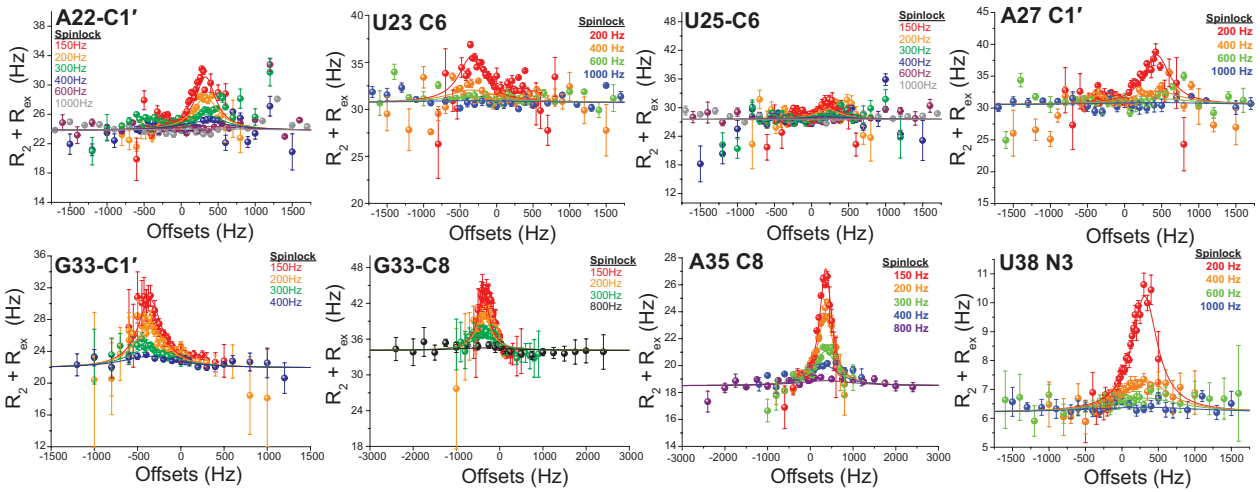

Figure 4.4 Off-resonance profiles of R1ρ relaxation dispersion in slow exchanging nuclei in wt TAR. Shown are dependence of $R_2 + R_{ex}$ on the offsets and the spinlock power. The solid lines denote the global fits to the two-state Laguerre exchange model (Eq. 1). The error analysis was done same as shown in Figure 4.3.

Nuclei	R1 (Hz)	R2 (Hz)	pB (%)	k_{ex} (Hz) τ (ms)	Δω (ppm)
U23-C6	2.50 ±0.04	30.7 ±0.1	0.40 ±0.05	474 ±69 2.1 ±0.3	2.3 ±0.1
U25-C6	3.00 ±0.04	27.5 ±0.1			-1.5 ±0.1
A22-C1′	1.70 ±0.02	23.8 ±0.05			-2.2 ±0.05
A27-C1′	1.61 ±0.05	30.5 ±0.1			-2.9 ±0.1
U38-N3	1.40 ±0.03	6.2 ±0.03			-5.2 ±0.2
G33-C1′	1.93 ±0.07	21.9 ±0.1			2.4 ±0.1
G33-C8	2.56 ±0.10	34.1 ±0.1			2.5 ±0.1
A35-C8	2.13 ±0.02	18.5 ±0.04			-2.4 ±0.04

Table 4.1 Globally fitted parameters from R1ρ relaxation dispersion of wt TAR at 25 °C

The similar slow exchange parameters observed for residues in the apical loop, upper stem, bulge, and lower stem could potentially reflect a common transition toward an excited state that requires simultaneous and concerted changes in the structure of the bulge, upper stem, and apical loop. To test this hypothesis, we examined the impact of omitting the bulge on relaxation dispersion measured in the apical loop (TAR-Δbulge, Figure 4.5A). As expected, omitting the bulge had no effect on the fast exchange process in the apical loop, since ES1 involve localized changes in base-pairing within the apical loop itself. However, omitting the bulge completely quenched the slow exchange process at G33-C8, G33-C1′, and A35-C8, which are four base-pairs away (Figure 4.5B). Thus, omitting the bulge specifically quenches the slow processes at G33-C8, G33-C1′ and A35-C8 without affecting the fast process at A35-C1′ in the same residue (Figure 4.5B). Therefore, the slow exchange observed in the apical loop is dependent on the TAR bulge. Inversely, replacing the wild-type (wt) TAR hexanucleotide apical loop with a UUCG tetraloop (TAR-UUCG, Figure 4.5A) completely quenched the relaxation dispersion observed in the bulge and the upper stem (A22-C1′, U23-C6, A27-C1′ and U38-N3, Figure 4.5C) indicating that the slow exchange observed in the bulge is strongly dependent upon having the wt TAR apical loop.

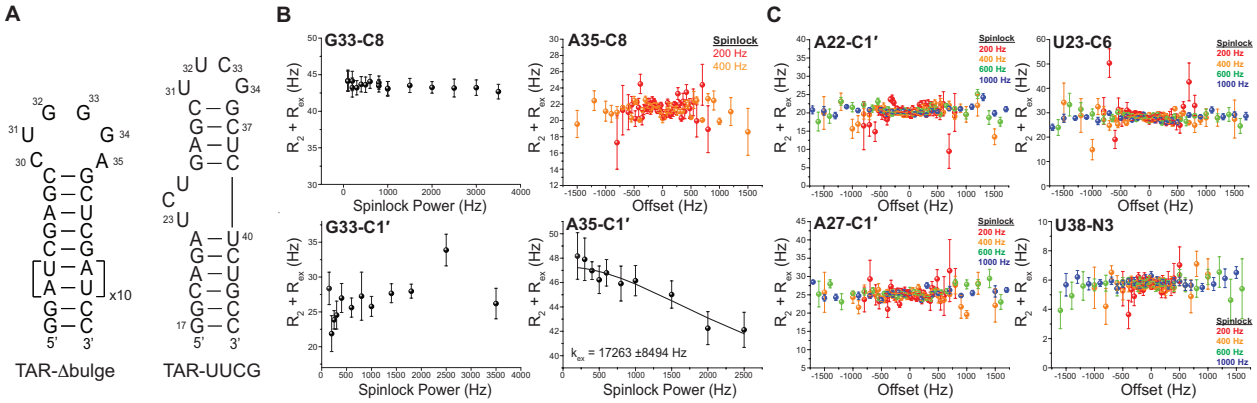

Figure 4.5 Co-dependence of slow relaxation dispersion in the TAR bulge and apical loop. **A.** Constructs used to independently delete the TAR bulge (TAR-Δbulge) and replace the wt apical loop (TAR-UUCG). **B.** Deleting the TAR bulge selectively quenches slow exchange in the TAR apical loop without affecting fast exchange. Error bars represent experimental uncertainty (one s.d.) as determined from propagation of errors obtained from mono-exponential fitting of duplicate sets of R1ρ data and analysis of signal-to-noise. **C.** Replacement of the wt TAR apical loop with a UUCG quenches slow relaxation dispersion in the bulge and the upper stem. Error bars were determined as mentioned in B.

Structure of HIV-1 TAR transient state by secondary structure prediction and mutate-and-chemical-shift-fingerprinting

The two-state analysis of the off-resonance relaxation dispersion data yields chemical shifts data for the transient state, which carry important clues about the ES2 structure. The ES2 carbon chemical shifts can be converted into recently formulated H-factors(54) which measure deviations away from canonical Watson-Crick geometry toward non-canonical geometry and which can be used to aid analysis and visualization of excited state carbon chemical shifts. The H-factor ranges between 0 - 68 for residues that are the most to the least similar to a WC geometry respectively (Figure 4.6). A few of the ES2 chemical shifts from the R1ρ data (Δω) and the corresponding H-factors were very informative. The very significantly upfield shifted (by 5.2 ppm) U38-N3 resonance

strongly suggests the absence of WC hydrogen bonding to U38-N3H3 in ES2 possibly due to formation of a non-canonical base-pair with distinct h-bonding or loss of any h-bonding to U38-N3(48) (Figure 4.7A). The upfield shifted base resonances for U25-C6 (1.5 ppm) and A35-C8 (2.4 ppm) and correspondingly lower H-factors suggest increased base-stacking(55) in ES2, while the downfield shifted U23-C6 (2.3 ppm) and increased H-factor suggest a base-flipped-out conformation(55) (Figure 4.7A). The upfield shifted A22-C1′ (2.2 ppm) and A27-C1′ (2.9 ppm) suggest sugar repuckering toward the C2′-endo conformation. Finally, the down-field shifted G33-C1′ (2.4 ppm) and G33-C8 (2.5 ppm) and increased H-factor strongly suggest a *syn* G33 base(56), and is consistent with a *syn* G involved in a trans-wobble G-U base-pair, as observed in closing base-pairs of apical loops such as UUCG(57) and the UGGG apical loop(20) observed in the TAR ES1 state.

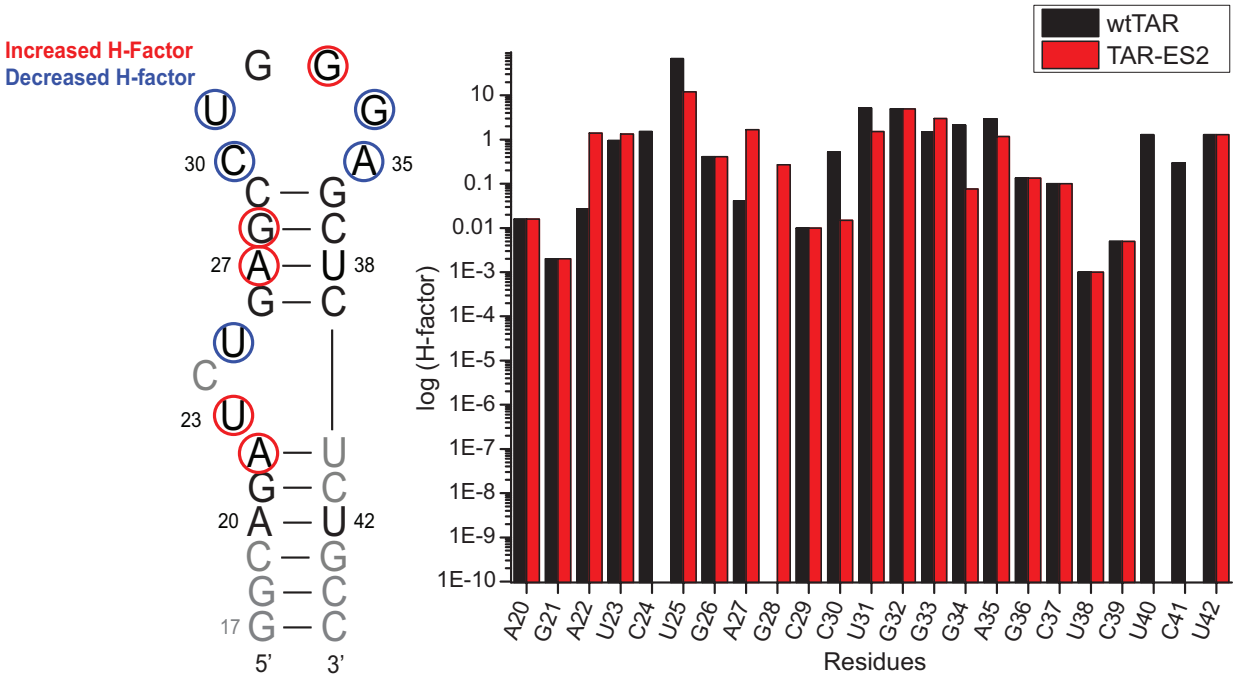

Figure 4.6 H-factor analysis of wt TAR chemical shifts and ES2 chemical shifts from R1ρ relaxation dispersion. Shown are the secondary structure of wt TAR with colored

circles on the residues showing increased (red) or decreased (blue) H-factor upon transition into ES2. The bar graph of H-factor on each residue is also shown. Note that except for 3 base-pairs at the terminus, gray residues (C24, U40, C41) were excluded from analysis due to unavailability of relaxation dispersion data.

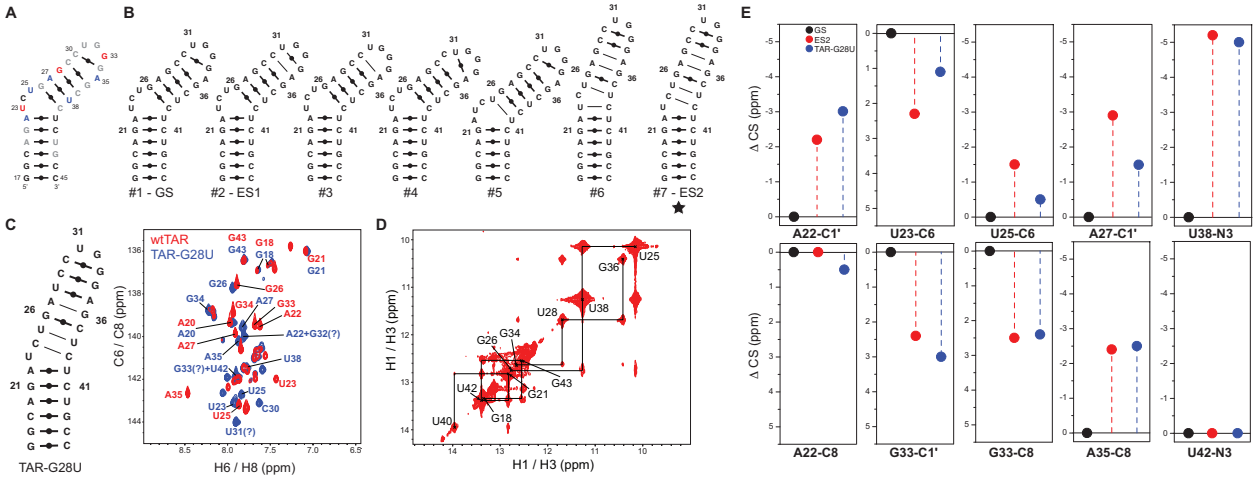

Figure 4.7 Trapping a candidate structure for the HIV-1 transient state. A. Ground state secondary structure of HIV-1 TAR. Chemical shift fingerprints indicating increased stacking and/or *anti* glycosidic angles, loss of hydrogen bonds, and C2'-endo sugar pucker are in blue; decreased stacking and/or *syn* glycosidic angles, and C3'-endo sugar pucker are in red. Sites with little to no or too fast exchange are in grey. The sites at which the exchange cannot be measured due to spectral overlap are in black. **B.** HIV-1 TAR secondary structures predicted using MC-Fold. The structure that best agrees with the R1ρ data is highlighted in star. **C.** Comparison of 2D C6/C8-H6/H8 HSQC spectra of TAR-G28U and wt TAR. The assignments with question marks are tentative assignments. **D.** Example imino proton NOE-connectivity that support the proposed secondary structure for TAR-G28U. **E.** Chemical shift fingerprinting. Shown are the differences in chemical shifts determined using relaxation dispersion between the ground and transient state (in red) with the corresponding differences in chemical shifts obtained from comparing resonances in TAR-G28U (in blue) and the ground state of wt TAR (in black).

We previously showed that secondary structure prediction programs such as MC-Fold can help identify RNA excited states(20). MC-Fold correctly predicts the HIV-1 TAR GS and ES1 (structures #1 and #2, Figure 4.7B) as the first and second most energetically favorable secondary structures. The next energetically favorable structures #3-#5 involve changes in either the apical loop (#3 and #4) or bulge (#5) without affecting the upper stem. Strikingly, the very similar structures #6 and #7 feature simultaneous changes in the bulge, upper stem, and apical loop without affecting the lower stem, which show little signs of relaxation dispersion. This is accomplished by a global change in the register of bulge residues C24 and U25, which pair up with residues in the upper stem and result in reshuffling of base-pairs that propagates to the apical loop through creation of non-canonical base-pairs.

Between the two structures, structure #7 better explains the decreased stacking for U23 and increased stacking for U25 observed in ES2 (Figure 4.7B). The non-canonical U25-U38 base-pair can help explain the unusual upfield shifted U38-N3 as well as increased base-stacking of U25 in ES2 (Figure 4.7B). G26 and A27 retain base-pairing in ES2 as in GS, consistent with the lack of significant relaxation dispersion measured at these sites. Formation of the G28-A35 mismatch can help explain the increased base-stacking at A35 in ES2. Importantly, both structures #6 and #7 position G33 at the apical loop closing base-pair where it can form the G(*syn*)- C trans-wobble base pair and explaining the strongly downfield shifted G33-C8 and G33-C1′. A transition toward structures #6 or #7 would minimally require the opening of a stable canonical A27-U38 base-pair, consistent with the measured activation free energy (16.8 kcal/mol; See Thermodynamic analysis in Methods), which is comparable to the energy needed to open RNA WC base-pairs(58) (13-16 kcal/mol). This helps explain why the

transition (k_{ex} ~ 474 Hz) is slower than that observed for ES1 (k_{ex} ~ 25 kHz), which requires disruption of the weaker cross stranded G34-C30 base-pair.

Testing the proposed HIV-1 TAR Transient State using Mutate-and-chemical-shift-fingerprinting

We used a 'Mutate-and-Chemical Shift-Fingerprint' (MCSF) strategy(20) to test the proposed HIV-1 TAR RNA transient sate. In particular, we introduced a G28U point mutation targeting the center of the upper stem that is predicted by MC-Fold to stabilize structures #6 and #7 by replacement of a non-canonical G28-A35 base-pair with a more stable canonical U28-A35 base-pair (TAR-G28U, Figure 4.7C). This mutant was chosen among many options because it minimizes the stabilization of other potentially competing low energy states and because it avoids changing key residues that are used as chemical shift reporters.

The G28U point mutation resulted in large changes in TAR chemical shifts including residues in the lower stem (A22), bulge (U23, U25), upper stem (A27, U38) and apical loop (G33, G34, A35), consistent with stabilization of an alternative conformation (Figure 4.7C and Figure 4.8). NMR spectra confirmed that the TAR-G28U adopts the secondary structure predicted for structure #7. For example, we observe unique imino resonances and NOE connectivity that establish formation of the non-canonical U25-U38 base pair and G36-A27 mismatch (Figure 4.7D). We also observe the expected NOE connectivity involving non-exchangeable protons (G17-H8 to A22-H8/H1′) that are interrupted at U23, consistent with U23 rather than C24 forming the bulge, as predicted by structure #7. We also observe a strong H8/H1′ NOE that confirms

G33 adopts a *syn* conformation along with the very strongly upfield shifted G34-H1′ resonance (Figure 4.8) which is characteristic of residues 3′-adjacent to *syn* G at tip of apical loops engaged in trans-wobble base-pairs (e.g. G(*syn*)-U in cUUCGg tetraloop). This suggests that G33 likely forms a trans-wobble with C30. These results suggest that the G28U mutation slightly favor the structure #7 over structure #6 and strongly favor structures #6 and #7 over all other structures (#1 - #5). As expected, the mutation causes minimal perturbations to chemical shifts of the lower stem residues (G17, G18, A20, G21, U40, U42, G43, and C45, Figure 4.7C and Figure 4.8).

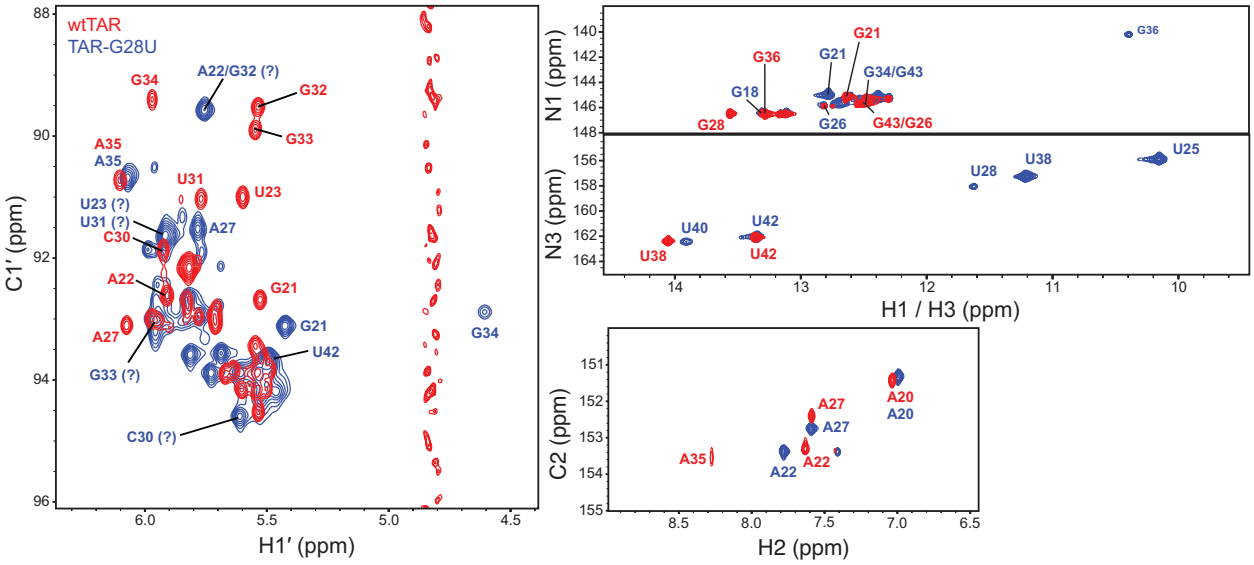

Figure 4.8 C1′-H1′, imino N-H, and C2-H2 HSQC spectra of wtTAR (red) and TAR-G28U (blue). The assignments with question marks are tentative assignments.

For all eight resonances showing signs of slow exchange (A22-C1′, U23-C6, U25-C6, A27-C1′, G33-C8, G33-C1′, A35-C8, and U38 N3), we observe very strong agreement between the carbon and nitrogen chemical shifts measured in TAR-G28U and the ES2 chemical shifts (Δω) deduced by relaxation dispersion (Figure 4.7E). The slightly poorer agreement (~1.5 ppm) observed for the magnitude (but not direction) of perturbations

at A22-C1′, U23-C6, U25-C6, and A27-C1′ can also be explained by proximity to the mutated site (A27-C1′) and susceptibility to exchange contributions (A22-C1′, U23-C6 and U25-C6) with structure #6 that may be absent or modified in TAR-G28U as compared to wt TAR. Note that while we were not able to obtain specific assignments for G33-C8, and G33-C1′ due to severe spectral overlap in TAR-G28U, we were able to tentatively assign G33 using NOE walk and HCN from G34. The G33 resonances fall downfield (by 2.4ppm, and 3.0ppm, respectively) of wt TAR resonances consistent with the ES2 carbon chemical shifts (2.5 ppm and 2.4 ppm, respectively) for these resonances obtained by relaxation dispersion (Figure 4.7E).

Importantly, residues that show little or no sign of relaxation dispersion also show small differences in carbon or nitrogen chemical shifts (typically <0.5 ppm) when comparing wt TAR and TAR-G28U, including G21-C1′, A22-C8, U23-C1′, A27-C2, G32-C1′, and U42-N3 (Figure 4.7E and Figure 4.9).

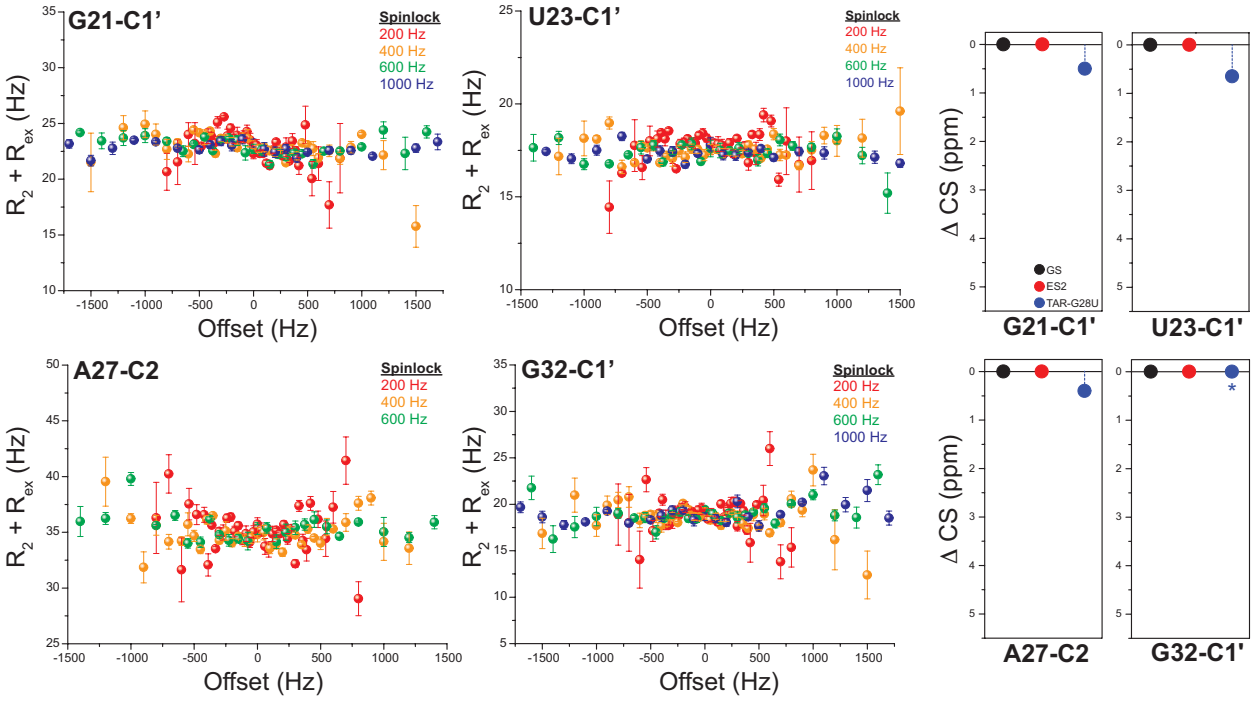

Figure 4.9 Off-resonance profiles of R1r relaxation dispersion and the chemical shift fingerprints of the residues that did not show relaxation dispersion.

Deconvoluting ES1 and ES2 relaxation dispersion contributions through temperature dependencies

Given that transitions to ES1 and ES2 entail very different timescales, one might wonder why certain apical loop residues do not sense both ES1 and ES2, and therefore necessitate a 3-state rather than 2-state fit to the dispersion data. For many resonances, the differences in chemical shifts between GS and ES ($\Delta\omega$) are expected to be significant only for one of the two ESs. For example, based on comparison of spectra for wt TAR and trapped mutants of ES1 and ES2, the differences in chemical shifts for A35-C1′ are expected to be 1.3 ppm and 0 ppm for ES1 and ES2 (Figure 4.8), respectively. Likewise, for G34-C8, the differences in chemical shifts are expected to be 1.8 and 0 ppm for ES1

and ES2 (Figure 4.7C), respectively. For other sites in the apical loop (C30-C1′, U31-C1′, and G34-C1′), it is possible that any relaxation dispersion contributions due to ES2 is masked by greater relaxation dispersion contributions due to ES1. Indeed, the simulations of relaxation dispersion on these resonances using 3-state model (ES1 ←→ GS ←→ ES2) show that the effective transverse relaxation rate constant with chemical exchange ($R_2 + R_{ex}$) from the low populated ES2 is completely masked by the effective transverse relaxation rate constant from ES1 which has higher population and faster exchange rate constant (data not shown). The relaxation rate constant due to chemical exchange of ES1 (R_{ex} ~ 17 - 21 Hz) is approximately 2-3 times higher than that of ES2 (R_{ex} ~ 5 - 10 Hz) and if the chemical shift differences (Δω) of these resonances in ES1 and ES2 from the ground state are also similar in direction, the relaxation dispersion profile of ES2 is highly likely masked by that of ES1.

To detect slow exchange to ES2 that may be masked by fast exchange to ES1, we performed relaxation dispersion experiments at a higher temperature of 35ºC, where we expect to push the already very fast process involving ES1 outside the limit of detection by the R1ρ relaxation dispersion experiment. As expected, all residues showing slow exchange due to ES2 at 25ºC also exhibited slow exchange at 35ºC (Figure 4.10 and Table 4.2) yielding similar population (p_B = ~0.2 %), ES2 chemical shifts, and as expected increased exchange rate constants (k_{ex} = ~2100 Hz) compared to those measured at 25ºC. Among the apical loop residues showing signs of fast exchange, G34-C8 and A35-C1′ no longer showed signs of relaxation dispersion at 35ºC (Figure 4.11A) indicating that the fast exchange process is now outside the limit of detection and that any changes in chemical shifts accompanying formation of ES2 are small – this was later confirmed with the chemical shifts of TAR-G28U (see below).

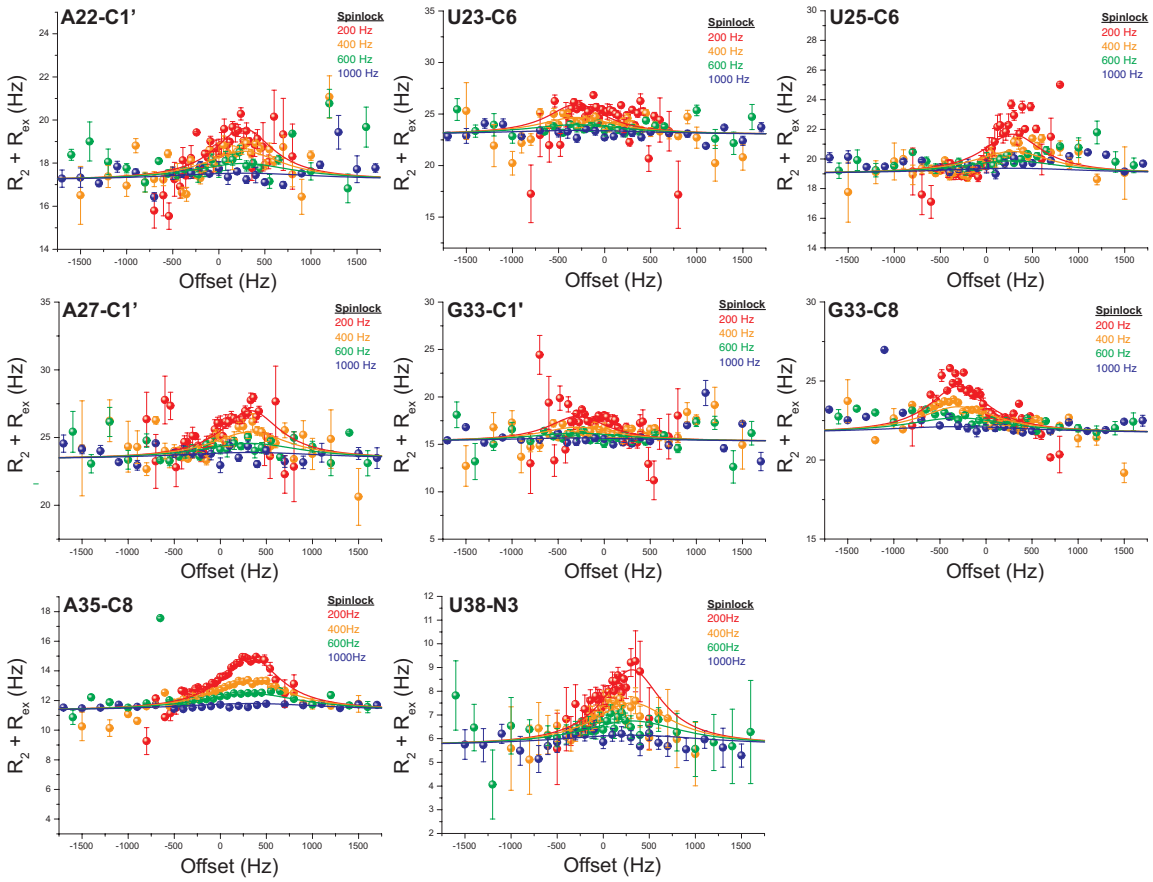

Figure 4.10 Off-resonance profiles of R1ρ relaxation dispersion at 35 °C

Nuclei	R1 (Hz)	R2 (Hz)	pB (%)	k_{ex} (Hz) τ (ms)	$\Delta\omega$ (ppm)
U23-C6	2.82 ±0.05	23.0 ±0.1			2.0 ±0.1
U25-C6	2.97 ±0.03	19.0 ±0.1			-1.9 ±0.1
A22-C1′	2.02 ±0.03	17.2 ±0.1			-1.8 ±0.1
A27-C1′	1.58 ±0.04	23.4 ±0.1			-2.2 ±0.1
U38-N3	1.90 ±0.08	5.7 ±0.1			-5.2 ±0.3
G33-C1′	1.98 ±0.06	15.3 ±0.1	0.24 ±0.01	2131 ±96 0.47 ±0.02	1.9 ±0.1
G33-C8	2.26 ±0.02	21.7 ±0.1			2.0 ±0.1
A35-C8	2.06 ±0.02	11.3 ±0.04			-2.2 ±0.1
C30-C1′	2.11 ±0.03	25.6 ±0.1			1.9 ±0.1
G34-C1′	2.15 ±0.04	22.6 ±0.1			2.9 ±0.1
U31-C1′	2.34 ±0.03	18.7 ±0.1			1.3 ±0.1

Table 4.2 Globally fitted parameters from R1ρ relaxation dispersion of wt TAR at 35 °C

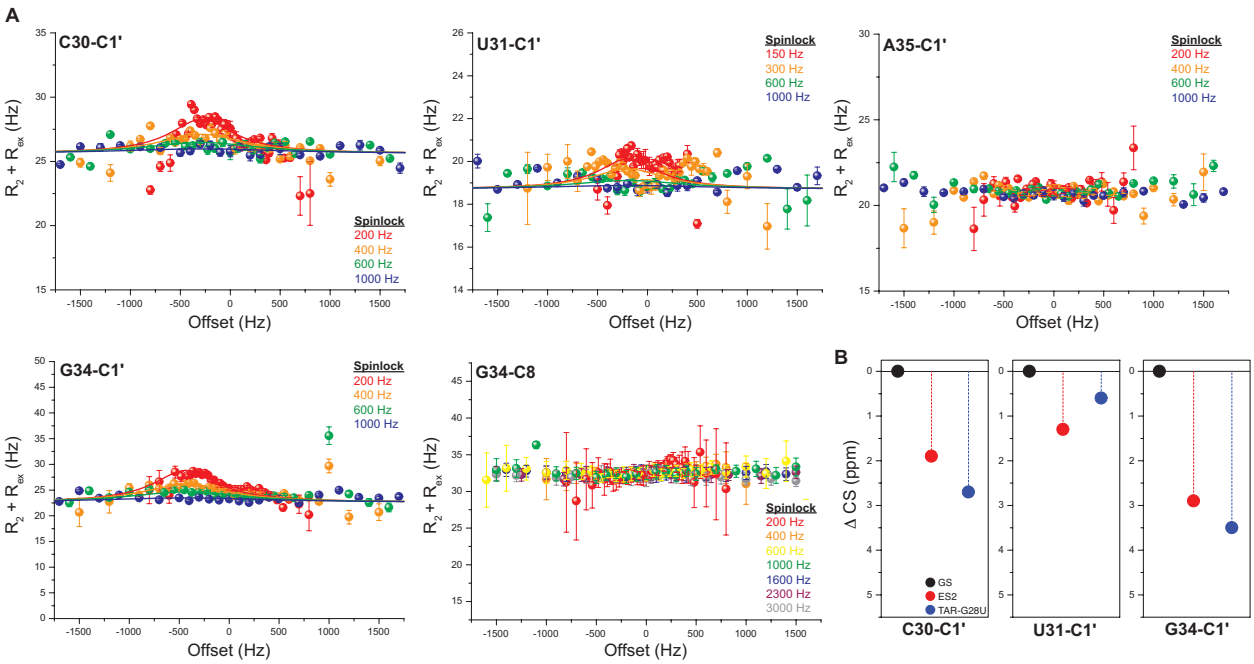

Figure 4.11 R1ρ relaxation dispersion of apical loop ES1 nuclei at 35 °C. A. Off-resonance profiles of C30-C1′, U31-C1′, A35-C1′, G34-C1′, and G34-C8. Shown are global fits (solid line) to a two-state Laguerre exchange model (Eq. 1). Error bars represent experimental uncertainty (one s.d.) as determined from propagation of errors obtained from mono-exponential fitting of duplicate sets of R1ρ data and analysis of signal-to-noise. **B.** Chemical shift fingerprinting of nuclei showing dispersion at 35 °C. The color scheme is same as in Figure 4.9.

By contrast, resonances belonging to C30-C1′, U31-C1′, and G34-C1′ which exhibit signs of fast exchange at 25ºC unexpectedly start to exhibit slow exchange (k_{ex} = 1552 – 2666 Hz) at higher temperature of 35ºC (Figure 4.11A) that are directed toward a lower populated (p_B ~ 0.2 %) transient state. These exchange rates and populations are in excellent agreement with those attributed to ES2 at 35ºC (Table 4.2). For example, G34-C1′ of TAR-G28U shows 3.5ppm downfield shift from that of wt TAR, which agrees well with the ES2 carbon chemical shift (2.9 ppm) from the dispersion data at 35ºC, which is consistent with G34-C1′ being adjacent to G33(*syn*)-C30 trans-wobble

base pair (Figure 4.11B). Likewise, G34-C8 and A35-C1′, which no longer shows the relaxation dispersion at 35ºC, have identical carbon chemical shifts between TAR-G28U and wt TAR (Figure 4.7C and Figure 4.8). Again, we were not able to obtain specific assignments for C30-C1′ and U31-C1′ due to severe spectral overlap in TAR-G28U but are tentatively assigned by HCN and TAR-UUCG spectra(59) which has similar tetraloop to the TAR-G28U. The C30-C1′ and U31-C1′ shift downfield (by 2.7 ppm and 0.6 ppm, respectively) of wt TAR resonances consistent with the ES2 carbon chemical shifts for these resonances obtained by relaxation dispersion at 35ºC, all of which shift downfield relative to the GS by 1.9ppm and 1.3ppm, respectively (Figure 4.11B). Although the data is more noisy than that measured at 25ºC possibly due to lower ES2 population at 35ºC (Figure 4.10 and Figure 4.11A), global fitting of the dispersion data measured at 35ºC (Table 4.2) allowed us to deduce the ES2 chemical shifts belonging to C30-C1′, U31-C1′, and G34-C1′ which are otherwise masked by the fast exchange to ES1 at 25ºC.

Functional significance of TAR ES2

Correlated exchange at the apical loop and bulge provides a basis for long-range allosteric communication in HIV-1 TAR. To test this possibility, we examined how binding of argininamide (ARG) - a ligand mimic of TAR's cognate protein target Tat - to the TAR bulge affects dispersion measured in the apical loop. Previous NMR studies(36, 60, 61) showed that ARG binds HIV-1 TAR in and around the bulge stabilizing a coaxial conformation in which bulge residues U23 forms reverse Hoogsteen with the A27-U38 base-pair in the upper stem. If our proposed transient state is correct, one would expect that ARG binding would quench chemical exchange due to ES2 because by locking the bulge conformation, the ARG bound state is incapable of forming ES2 conformation.

Indeed, addition of 16-fold ARG (11 mM) to TAR quenched the relaxation dispersion at the slow exchanging sites in the apical loop (G33-C8 and A35-C8, Figure 4.12). Interestingly, ARG binding also quenched relaxation dispersion at sites associated with fast exchange to ES1 (U31-C1′ and G34-C8). Prior studies have shown that ARG binding affects the chemical exchange at TAR apical loop resonances undergoing fast exchange and this was attributed to a secondary ARG binding site(62). Evidence for two or possibly more ARG binding sites in TAR was also previously obtained using the surface plasmon resonance experiments(63). Although we cannot rule out that binding of ARG to the apical loop is the cause of quenching dispersion at slow exchanging sites, other data showing that omission of the TAR bulge quenches dispersion at the apical loop strongly suggests that binding to the bulge allosterically quenches slow millisecond exchange at the apical loop, providing a new mechanism for allosteric regulation.

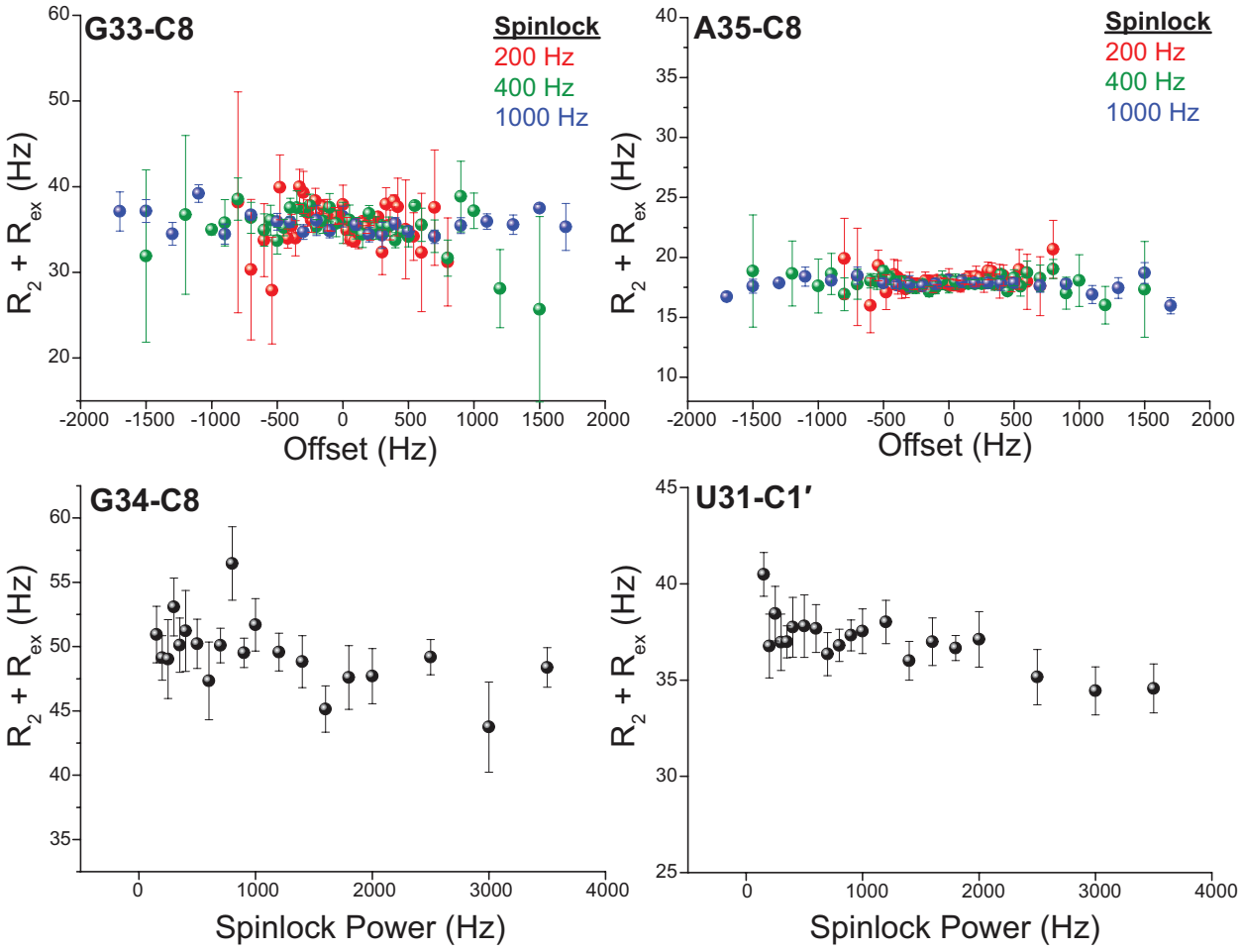

Figure 4.12 R1ρ relaxation dispersion profiles of wt TAR bound to ARG. The off-resonance profiles depict the relaxation dispersion of ES2 nuclei (G33-C8 and A35-C8) and the on-resonance profiles depict the relaxation dispersion of ES1 nuclei (G34-C8 and U31-C1′). Error bars represent experimental uncertainty (one s.d.) as determined from propagation of errors obtained from mono-exponential fitting of duplicate sets of R1ρ data and analysis of signal-to-noise.

One of the unique features of TAR is that it is capable of binding to wide variety of proteins that are thought to interact with the TAR bulge and/or apical loop. For example, the protein, Tat, is thought to primarily bind to the TAR bulge(64) whereas the proteins TAR RNA binding protein (TRBP)(65) and possibly RNA-activated protein

kinase (PKR)(66) are thought to also interact with apical loop. It is possible that some of these proteins recognize ES2 or ES1 forms of TAR. The ES2 may also help explain prior data showing evidence for long-range functional communication between the TAR bulge and apical loop. This includes the cooperative binding of cyclin T1-Tat complex to TAR bulge and the apical loop(44), elusive correlated motions between U31 of the apical loop and the U23 and C24 of the bulge(67), and a small molecule achieving high affinity and specificity to TAR by targeting the bulge and the apical loop simultaneously(37). Indeed, among 3166 full-length TAR sequences obtained from HIV sequence database (http://www.hiv.lanl.gov/), 850 non-redundant sequences were subject to the secondary structure prediction by MC-fold. Among them, 627 sequences were predicted to support ES2 (data not shown). Therefore, the ES2 seemed to be highly conserved (74%) among the HIV-1 strains, implying its significant involvement in the viral life cycle.

4.4 Conclusion

In the present study, we have characterized the slowly exchanging state, ES2, (τ = 2.1 ms; p_B = 0.4%) of TAR using R1ρ relaxation dispersion. Unlike the ES1 structure that involved local base-pair arrangements restricted in the apical loop, the ES2 involved the bulge, upper stem, and the apical loop residues rearranged to form a more helical conformation. We have also shown that by increasing temperature, the slowly exchanging state can be isolated from the fast exchanging state which otherwise hid the slow exchange process. Based on the R1ρ measurements at high temperature, the ES1 and ES2 may be independent of each other, but we could not rule out the possibility of the exchange process between ES1 and ES2 (Figure 4.13). It is also unclear how ES2 is involved in the HIV-1 life cycle, but we showed that the ES2 is highly conserved in the

HIV-1 strains and hinted that the ES2 likely involve with an allosteric regulatory mechanism. Nonetheless, our study experimentally characterized the long-range correlation between the bulge and the apical loop of TAR and may provide further insights into the transient structures within RNA.

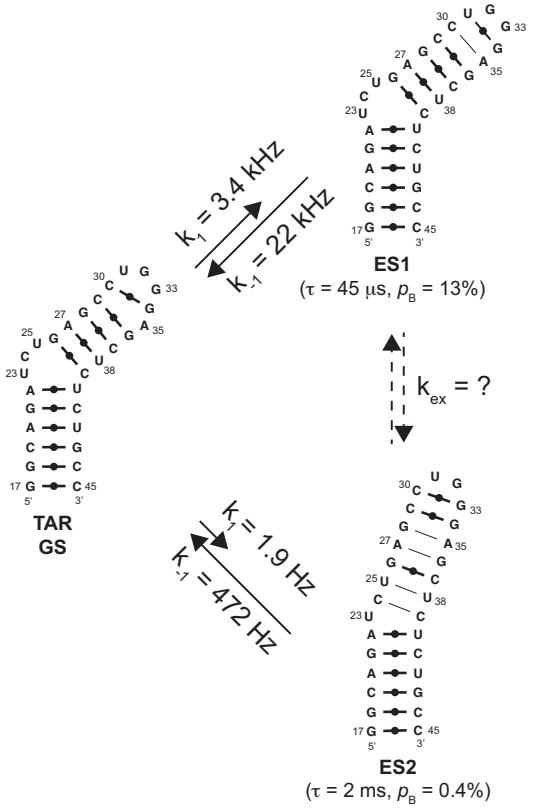

Figure 4.13 Proposed μs – ms exchange model in TAR. Shown are the secondary structures of ES1 and ES2 with lifetime (τ), forward rate constants (k_1) and backward rate constants (k_{-1}) calculated from their exchange rate constants ($k_{ex} = k_1 + k_{-1}$) and population (p_B).

4.5 References

1. Dethoff EA, Chugh J, Mustoe AM, & Al-Hashimi HM (2012) Functional complexity and regulation through RNA dynamics. *Nature* 482:322-330.
2. Boehr DD, Nussinov R, & Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5(11):789-796.
3. Hoogstraten CG & Sumita M (2007) Structure–function relationships in RNA and RNP enzymes: Recent advances. *Biopolymers* 87(5-6):317-328.
4. Haller A, Soulière MF, & Micura R (2011) The Dynamic Nature of RNA as Key to Understanding Riboswitch Mechanisms. *Acc. Chem. Res.* 44(12):1339-1348.
5. Krueger BJ, Varzavand K, Cooper JJ, & Price DH (2010) 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(8):e12335.
6. Ray PS, *et al.* (2009) A stress-responsive RNA switch regulates VEGFA expression. *Nature* 457(7231):915-919.
7. Grundy FJ, Winkler WC, & Henkin TM (2002) tRNA-mediated transcription antitermination in vitro: Codon-anticodon pairing independent of the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 99(17):11121-11126.
8. Zhang Q, Stelzer AC, Fisher CK, & Al-Hashimi HM (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* 450(7173):1263-1267.
9. Mulder AM, *et al.* (2010) Visualizing Ribosome Biogenesis: Parallel Assembly Pathways for the 30S Subunit. *Science* 330(6004):673-677.
10. Menichelli E, Isel C, Oubridge C, & Nagai K (2007) Protein-induced Conformational Changes of RNA During the Assembly of Human Signal Recognition Particle. *J. Mol. Biol.* 367(1):187-203.
11. Stone MD, *et al.* (2007) Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* 446(7134):458-461.
12. Roth A & Breaker RR (2009) The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annu. Rev. Biochem.* 78(1):305-334.
13. Herschlag D, Khosla M, Tsuchihashi Z, & Karpel RL (1994) An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO J.* 13(12):2913-2924.
14. Treiber DK & Williamson JR (2001) Beyond kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* 11(3):309-314.

15. Babitzke P & Yanofsky C (1993) Reconstitution of Bacillus subtilis trp attenuation in vitro with TRAP, the trp RNA-binding attenuation protein. *Proc. Natl. Acad. Sci. U.S.A.* 90(1):133-137.
16. Sekhar A & Kay LE (2013) NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci. U.S.A.* 110(32):12867-12874.
17. Palmer Iii AG (2014) Chemical exchange in biomacromolecules: Past, present, and future. *J. Magn. Reson.* 241(0):3-17.
18. Hoogstraten CG, Wank JR, & Pardi A (2000) Active Site Dynamics in the Lead-Dependent Ribozyme. *Biochemistry* 39(32):9951-9958.
19. Blad H, Reiter NJ, Abildgaard F, Markley JL, & Butcher SE (2005) Dynamics and Metal Ion Binding in the U6 RNA Intramolecular Stem-Loop as Analyzed by NMR. *J. Mol. Biol.* 353(3):540-555.
20. Dethoff EA, Petzold K, Chugh J, Casiano-Negroni A, & Al-Hashimi HM (2012) Visualizing transient low-populated structures of RNA. *Nature* 491(7426):724-728.
21. Hansen AL, Nikolova EN, Casiano-Negroni A, & Al-Hashimi HM (2009) 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(11):3818-3819.
22. Massi F, Johnson E, Wang C, Rance M, & Palmer AG (2004) NMR R1 rho Rotating-Frame Relaxation with Weak Radio Frequency Fields. *J. Am. Chem. Soc.* 126(7):2247-2256.
23. Korzhnev DM, Orekhov VY, & Kay LE (2004) 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(2):713-721.
24. Bannwarth S & Gatignol A (2005) HIV-1 TAR RNA: the target of molecular interactions between the virus and its host. *Curr. HIV Res.* 3(1):61-71.
25. Bieniasz PD, Grdina TA, Bogerd HP, & Cullen BR (1999) Recruitment of cyclin T1/P-TEFb to an HIV type 1 long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. *Proc. Natl. Acad. Sci. U.S.A.* 96(14):7791-7796.
26. Roy S, Delling U, Chen CH, Rosen CA, & Sonenberg N (1990) A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev.* 4:1365-1373.
27. Selby MJ, Bain ES, Luciw PA, & Peterlin BM (1989) Structure, sequence, and position of the stem-loop in tar determine transcriptional elongation by tat through the HIV-1 long terminal repeat. *Genes Dev.* 3:547-558.

28. Laspia MF, Rice AP, & Mathews MB (1989) HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation. *Cell* 59(2):283-292.
29. Benkirane M, *et al.* (1997) Oncogenic potential of TAR RNA binding protein TRBP and its regulatory interaction with RNA-dependent protein kinase PKR. *EMBO J.* 16(3):611-624.
30. Carpick BW, *et al.* (1997) Characterization of the Solution Complex between the Interferon-induced, Double-stranded RNA-activated Protein Kinase and HIV-I Trans-activating Region RNA. *J. Biol. Chem.* 272(14):9510-9516.
31. Jalalirad M, Saadatmand J, & Laughrea M (2012) Dominant Role of the 5' TAR Bulge in Dimerization of HIV-1 Genomic RNA, but No Evidence of TAR-TAR Kissing during in Vivo Virus Assembly. *Biochemistry* 51(18):3744-3758.
32. Das A, Vrolijk M, Harwig A, & Berkhout B (2012) Opening of the TAR hairpin in the HIV-1 genome causes aberrant RNA dimerization and packaging. *Retrovirology* 9(1):59.
33. Helga-Maria C, Hammarskjöld M-L, & Rekosh D (1999) An Intact TAR Element and Cytoplasmic Localization Are Necessary for Efficient Packaging of Human Immunodeficiency Virus Type 1 Genomic RNA. *J. Virol.* 73(5):4127-4135.
34. Klase Z, *et al.* (2007) HIV-1 TAR element is processed by Dicer to yield a viral micro-RNA involved in chromatin remodeling of the viral LTR. *BMC Mol. Biol.* 8(1):63.
35. Wagschal A, *et al.* (2012) Microprocessor, Setx, Xrn2, and Rrp6 Co-operate to Induce Premature Termination of Transcription by RNAPII. *Cell* 150(6):1147-1157.
36. Puglisi JD, Tan R, Calnan BJ, Frankel AD, & Williamson JR (1992) Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy. *Science* 257(5066):76-80.
37. Davidson A, *et al.* (2009) Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. *Proc. Natl. Acad. Sci. U.S.A.* 106(29):11931-11936.
38. Ludwig V, *et al.* (2007) Tripeptides from Synthetic Amino Acids Block the Tat–TAR Association and Slow Down HIV Spread in Cell Cultures. *ChemBioChem* 8(15):1850-1856.
39. Stelzer AC, *et al.* (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* 7(8):553-559.
40. Stevens M, De Clercq E, & Balzarini J (2006) The regulation of HIV-1 transcription: Molecular targets for chemotherapeutic intervention. *Medicinal Research Reviews* 26(5):595-625.

41. Jaeger JA & Tinoco I (1993) An NMR study of the HIV-1 TAR element hairpin. *Biochemistry* 32(46):12522-12530.
42. Kulinski T, *et al.* (2003) The Apical Loop of the HIV-1 TAR RNA Hairpin Is Stabilized by a Cross-loop Base Pair. *J. Biol. Chem.* 278(40):38892-38901.
43. Aboul-ela F, Karn J, & Varani G (1996) Structure of HIV-1 TAR RNA in the Absence of Ligands Reveals a Novel Conformation of the Trinucleotide Bulge. *Nucleic Acids Res* 24(20):3974-3981.
44. Richter S, Cao H, & Rana TM (2002) Specific HIV-1 TAR RNA Loop Sequence and Functional Groups Are Required for Human Cyclin T1-Tat-TAR Ternary Complex Formation. *Biochemistry* 41(20):6391-6397.
45. Wöhnert J, Görlach M, & Schwalbe H (2003) 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. (Translated from English) *J Biomol NMR* 26(1):79-83 (in English).
46. Sklenár V, Peterson R, Rejante M, & Feigon J (1993) Two-and three-dimensional HCN experiments for correlating base and sugar resonances in 15N, 13C-labeled RNA oligonucleotides. (Translated from English) *J Biomol NMR* 3(6):721-727 (in English).
47. Fürtig B, Richter C, Wöhnert J, & Schwalbe H (2003) NMR Spectroscopy of RNA. *ChemBioChem* 4(10):936-962.
48. Nikolova EN, Gottardo FL, & Al-Hashimi HM (2012) Probing Transient Hoogsteen Hydrogen Bonds in Canonical Duplex DNA Using NMR Relaxation Dispersion and Single-Atom Substitution. *J. Am. Chem. Soc.* 134(8):3667-3670.
49. Delaglio F, *et al.* (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. (Translated from English) *J Biomol NMR* 6(3):277-293 (in English).
50. Spyracopoulos L (2006) A suite of Mathematica notebooks for the analysis of protein main chain 15N NMR relaxation data. (Translated from English) *J Biomol NMR* 36(4):215-224 (in English).
51. Miloushev VZ & Palmer III AG (2005) R(1rho) relaxation for two-site chemical exchange: General approximations and some exact solutions. *J. Magn. Reson.* 177(2):221-227.
52. Coman D & Russu IM (2005) A Nuclear Magnetic Resonance Investigation of the Energetics of Basepair Opening Pathways in DNA. *Biophys. J.* 89(5):3285-3292.
53. Parisien M & Major F (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452(7183):51-55.

54. McBrairty M, Xue Y, Petzold K, & Al-Hashimi HM (The H-factor: A New Parameter that Relates RNA Chemical Shifts and Secondary Structure. *In Preparation*.
55. Xu X-P & Au-Yeung SCF (2000) Investigation of Chemical Shift and Structure Relationships in Nucleic Acids Using NMR and Density Functional Theory Methods. *J. Phys. Chem. B* 104(23):5641-5650.
56. Ghose R, Marino JP, Wiberg KB, & Prestegard JH (1994) Dependence of 13C Chemical Shifts on Glycosidic Torsional Angles in Ribonucleic Acids. *J. Am. Chem. Soc.* 116(19):8827-8828.
57. Senada N, Boris Fr, Hendrik RAJ, Christian R, & Harald S (2010) High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res* 38(2):683-694.
58. Snoussi K & Leroy JL (2001) Imino Proton Exchange and Base-Pair Kinetics in RNA Duplexes. *Biochemistry* 40(30):8898-8904.
59. Zhang Q, Sun X, Watt ED, & Al-Hashimi HM (2006) Resolving the Motional Modes That Code for RNA Adaptation. *Science* 311(5761):653-656.
60. Puglisi JD, Chen L, Frankel AD, & Williamson JR (1993) Role of RNA structure in arginine recognition of TAR RNA. *Proc. Natl. Acad. Sci. U.S.A.* 90(8):3680-3684.
61. Pitt SW, Majumdar A, Serganov A, Patel DJ, & Al-Hashimi HM (2004) Argininamide Binding Arrests Global Motions in HIV-1 TAR RNA: Comparison with Mg2+-induced Conformational Stabilization. *J. Mol. Biol.* 338(1):7-16.
62. Bardaro MF, Shajani Z, Patora-Komisarska K, Robinson JA, & Varani G (2009) How binding of small molecule and peptide ligands to HIV-1 TAR alters the RNA motional landscape. *Nucleic Acids Res* 37(5):1529-1540.
63. Davis B, *et al.* (2004) Rational Design of Inhibitors of HIV-1 TAR RNA through the Stabilisation of Electrostatic "Hot Spots". *J. Mol. Biol.* 336(2):343-356.
64. Dingwall C, *et al.* (1990) HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. *EMBO J.* 9(12):4145-4153.
65. Erard M, Barker DG, Amalric Fo, Jeang K-T, & Gatignol A (1998) An Arg/Lys-rich core peptide mimics TRBP binding to the HIV-1 TAR RNA upper-stem/loop. *J. Mol. Biol.* 279(5):1085-1099.
66. Kim I, Liu CW, & Puglisi JD (2006) Specific Recognition of HIV TAR RNA by the dsRNA Binding Domains (dsRBD1-dsRBD2) of PKR. *J. Mol. Biol.* 358(2):430-442.
67. Musselman C, Al-Hashimi HM, & Andricioaei I (2007) iRED Analysis of TAR RNA Reveals Motional Coupling, Long-Range Correlations, and a Dynamical Hinge. *Biophys. J.* 93(2):411-422.

Chapter 5

Conclusions and Future Directions

5.1 Conclusions and Future Directions

Despite non-coding RNAs being attractive therapeutic targets and much effort directed toward targeting RNA with small molecules to modulate the regulatory functions, there has not been novel drugs that have been discovered that target RNA other than the well-known aminoglycoside antibiotics that target ribosomes. This dissertation examined the feasibility of using experimental and virtual screening approaches to identify small molecules that bind RNA.

In **Chapter 2**, we investigated the influence of dimethyl sulfoxide (DMSO) on RNA structure, dynamics, and ligand binding using fluorescence binding assays and NMR spectroscopy. We showed that DMSO (up to 10% DMSO by volume) locally melts the junctions that connect RNA helices such as bulge in HIV-1 TAR and internal loop in 16S ribosomal A-site. The binding affinities of positively charged compounds such as argininamide, and kanamycin, did not show significant changes in DMSO, but mitoxantrone which binds TAR and A-site via intercalation notably showed 3 – 4-fold reduced binding affinity. While the effect of DMSO on ligand binding was not significant compared to prior studies on protein-ligand complexes (1), the RNA structural changes induced by DMSO deserve careful attention before proceeding with ligand binding assays and high throughput screening campaigns.

Future studies should examine how DMSO affects the structure, dynamics and folding behavior of RNA containing deep binding pockets such as the aptamer domains of RNA riboswitches. The influence of DMSO may differ from that observed here for hairpin structures with more solvent-exposed binding pockets. Future studies should also examine how other organic solvents affect RNA behavior to more fully understand the origin of the DMSO effect.

In **Chapter 3**, we embarked on the high throughput screening (HTS) and the ensemble-based virtual screening (VS) of ~100,000 small molecules targeting HIV-1 TAR in order to find novel inhibitors as well as compare the performance of HTS and the VS. This provided the first opportunity to comprehensively assess the performance of RNA-targeted VS. We showed that the two techniques are complimentary to one another, yielding different classes of small molecule hits. Furthermore, we showed the importance of working with an unbiased small molecule library when assessing the VS approach. With this approach we were able to identify total 11 novel small molecule hits, one of which showed activity in the cell-based assays using T-cell lines and TZM-bl reporter cell lines.

Future studies should examine how the quality of ensembles affect the performance of the VS. In particular, the current ensemble used in the VS contained RDCs targeting the helices and bulge; a new ensemble should be tested in which RDCs are also measured targeting the apical loop. Improvements in the ensemble could also be obtained by using other experimental constraints in the ensemble determination such as chemical shifts. A key question is whether having the proper representation of the unbound solution ensemble leads better predictions in VS and whether small molecules broadly target conformers within the ensemble or prefer a subset of conformations that

may have unique binding pockets. There is also a need to assess various docking programs and not only ICM tested here to ascertain what the main weaknesses in scoring functions are. Of particular importance is the effects of conformational entropy which are not readily captured in scoring functions and which can be experimentally evaluated using ITC experiments. Likewise, there is a need to critically assess pose predictions obtained form docking. This can be done efficiently with the use of NMR chemical shift mapping experiments. On the experimental side, the compound 106134 showed promising activity. Future follow up studies should examine variants of this small molecule to see if the activity can be further improved. One possibility would be to link the small molecule with dimethyl amiloride (DMA), which uniquely targets the TAR apical loop. In addition, determination of high resolution structures for the compound bound to TAR by NMR or X-ray crystallography can help directing future rational modifications of the small molecule.

In **Chapter 4**, we have characterized the second transient state of HIV-1 TAR using NMR R1ρ relaxation dispersion and mutagenesis. This transient state involves the re-orientation of bases and base-pairs in bulge, upper helix, and apical loop of TAR, suggesting a long-range correlation between the bulge and the apical loop. Even though this transient state is lowly populated (0.4%) with lifetime of 2 milliseconds, the occurrence of the transient state in HIV-1 TAR sequences seems highly conserved, implying that this transient state is biologically important. Future studies should examine the exact topology linking the TAR ground state and the two distinct transient states that have been uncovered so far (2). This could be accomplished by carrying out dispersion measurements as a function of pH so to perturb one equilibrium while not affecting the second. In addition, studies should be carried out to examine the functional significance of the HIV transient state uncovered here. In particular, cell-

based assays can be used to examine how TAR mutants that favor the ground state or the transient states affect Tat binding and transcriptional activation. More broader viral studies may make it possible to examine the role of these transient states in other steps of viral replication. Of particular interest will be to examine whether the mutants that trap the transient states support transcriptional activation of the viral genome. Our secondary structure suggests that the transient states do not have the requirements for Tat binding and consequently, should disfavor transcriptional activation. If this proves to be the case, studies can be carried out for finding small molecules that target specifically transient states of RNA.

The work presented in this thesis also paves the way for a new mode of RNA-screening. In particular, transient states of RNA have sufficiently altered secondary structure that they may have vastly different functions. One possibility is therefore to find small molecules that specifically target transient states of RNA. This can be accomplished using two distinct strategies. The first involves using mutants to trap the transient state and then subject the mutants to conventional HTS and VS. A second approach would involve using the wild-type construct, but using NMR chemical shift mapping to identify small molecules that specifically trap a transient state as assessed based on chemical shift fingerprinting.

There are many unsolved mysteries about how small molecules bind RNA. The most daunting challenge in RNA-targeted drug discovery is to find small molecules that are active *in vivo* since many small molecules that showed activities in *in vitro* biochemical assays failed to show activities in *in vivo* assays. One hypothesis is that the small molecules must achieve high specificity to the target RNA in order to avoid being sequestered by other RNAs *in vivo*. The most widely used RNA for testing the

specificity of small molecules is transfer RNA (tRNA) which is one of the most abundant RNA in cells. However, the known antibiotics such as aminoglycosides are promiscuous binders to a variety of RNA elements including tRNA, but yet they exhibit antibiotic activities in cells by binding to a specific 16S ribosomal decoding RNA. In addition, we have discovered 9 inhibitors of TAR-Tat interaction in the presented dissertation that showed exquisite specificity against excess tRNA, but only 1 small molecule showed a promising activity in cells. It is still unclear whether the specificity to the target RNA governs the *in vivo* activities or there are other mechanisms that are active in cells.

Another question that needs to be addressed is what are the properties of small molecules that favor RNA binding and how do they differ form protein-binding small molecules. Because of differences in the biophysical and structural properties of RNA and proteins, we can expect that the properties of RNA-binding small molecules will differ from those of protein-binding small molecules. Indeed, many small molecules that bind to RNA including aminoglycosidic antibiotics are positively charged or contain a planar aromatic ring that can be stacked with nucleobases or intercalated between the nucleobases. However, the presented dissertation showed 10 new TAR-Tat inhibitors that are not positively charged and rarely contain aromatic rings, suggesting that there are still many chemotypes of RNA-binding small molecules that have not yet been discovered. There is, therefore, a need to increase the scale of RNA-targeted screening to more efficiently explore new regions of small molecule space and a variety of RNA targets.

Targeting RNA with small molecules is still challenging but not impossible. Currently, we need more small molecules that are unique and bind RNA specifically in

order to address the aforementioned problems and questions. This dissertation showed two approaches that can facilitate the search for potential RNA-binding small molecules. These approaches have targeted specifically HIV-1 TAR, but they can be generalized to target other therapeutic RNAs. In the next decade, we can expect an expansion in the chemical space of RNA-binding small molecules and our understanding of RNA-small molecule interactions and how small molecules achieve their activities *in vivo*. With 20 classes of non-coding RNAs that form unique secondary and tertiary structures, a universe of RNA-binding small molecules is expected.

5.2 References

1. Cubrilovic D & Zenobi R (2013) Influence of Dimehylsulfoxide on Protein-Ligand Binding Affinities. *Anal. Chem.* 85(5):2724-2730.
2. Dethoff EA, Petzold K, Chugh J, Casiano-Negroni A, & Al-Hashimi HM (2012) Visualizing transient low-populated structures of RNA. *Nature* 491(7426):724-728.