Characterizing HIV-1 Transactivation Response Element Dynamics that Govern Ligand Recognition: Direct Applications to Drug Discovery

by

Andrew Corbet Stelzer

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in The University of Michigan
2010

Doctoral Committee:

Professor Hashim M. Al-Hashimi, Chair
Professor Charles L. Brooks III
Professor Gary D. Glick
Professor Richard R. Neubig
Professor Nils G. Walter

2010

To my loving family and friends. Without all of you I would not have come this far.

Acknowledgements

First, I would like to thank my advisor Hashim M. Al-Hashimi for reassuring my belief that nothing is impossible. His support and drive kept me motivated to continue tackling difficult projects. There is no doubt that my time in the Al-Hashimi lab has been a giant step toward meeting my career goals. Second, I would like to thank all of the current and past members of the Al-Hashimi lab. Forging ahead in this journey would not have been as enjoyable without all of them. I would like to especially thank my colleague and close friend Aaron T. Frank. Working with him has been exhilarating and there is no doubt we will have many successes together in the future. Most importantly, I would like to thank my family and loving girlfriend Nina. At times graduate school can get the best of you and your loving support buoyed my spirits when I felt like I could not continue.

Table of Contents

List of Figures

Figure

List of Tables

Table

List of Appendices

Appendix

Abstract

RNAs regulate numerous cellular processes making them highly sought therapeutic targets. One mechanism to inhibit functional RNAs is to alter their structura-dynamics using small molecule binders. While structure based drug design is often used to discover small molecule inhibitors, difficulties arise because, unlike proteins, RNAs undergo large conformational changes between the free and ligand-bound states that cannot be determined *a priori*. The spatial and temporal complexity of these conformational changes precludes accurate characterization that would allow one to visualize the conformational changes. While no method currently exists that can fully interrogate complex RNA dynamics, nuclear magnetic resonance (NMR) is a prime candidate because of its ability to site-specifically report on dynamics over a wide range of timescales (picoseconds-minutes). However, the number of measurable observables by NMR pales in comparison to the number of parameters needed to fully describe complex dynamic motions. Molecular dynamics (MD) can provide an all-atom description of RNA dynamics, but the force-fields governing the theoretical description are inaccurate and simulations remain limited to ~100 nanosecond timescales.

Using NMR and MD we aim to uncover the biophysical principles governing HIV-1 transactivation response element RNA (TAR) mediated ligand recognition and discover new TAR binding small molecules. First, we present an adapted Sample and Select (SAS) approach, which combines NMR residual dipolar couplings (RDCs) and MD to provide an accurate all-atom description of RNA dynamics over sub-millisecond timescales. RDCs measured on elongated TAR molecules are used to separate internal and overall motions and impose a helix-anchored reference frame. Using the SAS method, refined RNA ensembles that re-capitulate experimental RDCs are generated from an MD trajectory. Specific snapshots of the ensemble closely agree with previously determined ligand-bound TAR structures suggesting that the bound state conformations are sampled in the absence of ligand. In a second study we investigate the sequence dependence of TAR dynamics and show that a modest mutation greatly perturbs global and local dynamics giving rise to changes in small molecule binding affinity while still forming the same bound state RNA conformation. This suggests that RNA dynamics are sequence encoded and lead to pre-tuned functional dynamic ensembles. Lastly, the SAS ensemble structures are used in RNA structure-based drug discovery. Computational docking simulations are used to discover 11 TAR-binding small molecules, 8 of which have never before been shown to bind TAR and 2 never before been shown to bind RNA. NMR chemical shift perturbations and fluorescence polarization measurements verify that the small molecules bind TAR and inhibit the TAR-Tat interaction with inhibition constants ranging 0.627-300 μM. RDCs measured on TAR bound to the small molecule netilmicin suggest that docking against the SAS structures accurately re-capitulates the bound-state. Remarkably, netilmicin is also shown to inhibit TAR-mediated HIV-1 LTR expression and HIV-1 replication in an HIV-1 indicator cell line, TZM-bl, with an IC_{50} of 23.1 μM.

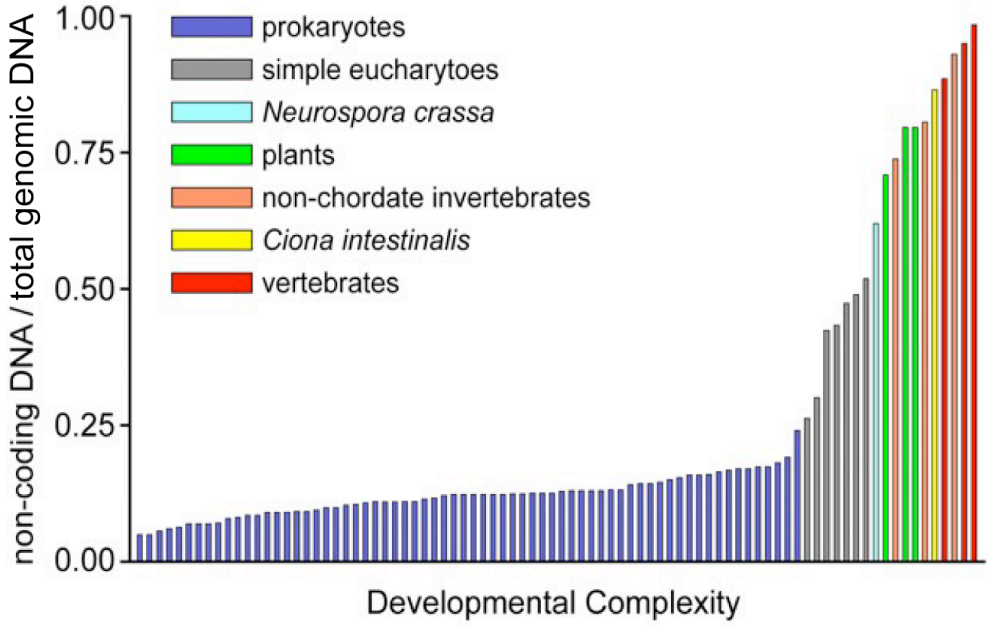
Chapter 1

Background and Introduction

1.1 Central Role of RNA in Biology

Watson and Crick's discovery of the DNA structure more than 50 years ago(1) led to the formulation of the central dogma of biology. While it was long thought that DNA molecules hold the library of genetic codes that are converted into "single-tasked" mRNAs specifying the protein sequences responsible for carrying out cellular functions, an increasing number of studies have revealed that RNA plays a much more important regulatory role. Discovery of the self-splicing pre-ribosomal catalytic RNA in the early 1980s(2,3) triggered an explosion in the number of studies aimed at understanding the role non-coding RNAs (ncRNAs) play in regulating and affecting cellular processes. Examples of ncRNAs include: catalytic RNAs that play key roles in protein synthesis, tRNA maturation, self-splicing intron removal, and viroid replication; small nuclear RNAs (snRNAs) that comprise the intron removal pre-mRNA splicing machinery; telomerase RNA required for chromosome end replication; guide RNAs critical to RNA editing; the signal recognition particle RNA necessary for protein translocation; small nucleolar

RNAs responsible for ribosomal RNA (rRNA) modification; interfering and micro RNAs (RNAi and miRNA) that underlie an entirely new genetic network for regulating gene expression(4); and riboswitches that regulate transcription(5), translation initiation(6), self-cleavage(7,8), degradation(9,10), and antisense mRNA production(11). While this list does not cover the entire breadth of regulatory RNA molecules, it illustrates the broadening of RNA functions beyond the traditional roles ascribed to mRNA, tRNA, and rRNA in the central molecular dogma



Adapted from Taft, R. J. & Mattick, J. S. Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. *arXiv Preprint Archive* [online], <http://www.arxiv.org/abs/q-bio.GN/0401020> (2003).

Figure 1.1. Cellular ncRNA increases with developmental complexity. As organisms exhibit more complex biology the ratio of transcribed ncDNA to coding DNA increases.

Recent studies indicate that the amount of transcribed ncRNA increases significantly with organismal complexity, and that as much as 96% of gene products in humans could correspond to ncRNAs whose functions remain to be elucidated. In comparison, only 2% is translated into proteins, leaving as much as 96% ncRNA available to perform diverse tasks (Figure 1.1)(12-14). Given the increasing importance

of ncRNAs, it is easy to appreciate the increasing scientific focus devoted to understanding the roles RNA plays in biology and how they can be exploited as drug targets.

1.2 RNA as a Therapeutic Target

A growing number of RNAs and RNA-mediated processes are emerging as targets for antibacterial and antiviral therapies. Targeting rRNA is already a paradigm for developing antibiotics. For example, preferential binding of aminoglycoside molecules to the bacterial A-site in the 30S ribosomal subunit leads to mistranslation of protein genes and provides one avenue to combat bacterial infections (other avenues for targeting rRNA for antibacterial treatment are reviewed in(15,16)). Many RNA structural elements located in the untranslated region (UTR) of retroviral genomes are considered opportune anti-viral targets. In the case of the human immunodeficiency virus (HIV), this includes the transactivation response element (TAR)(17-19), which regulates HIV transcriptional fidelity by forming a catalytic ribonucleoprotein complex that hyperphosphorylates RNA polymerase II; the rev response element (RRE)(20), which serves as the recognition site for the Rev protein responsible for exporting the transcribed HIV genome outside the nucleus; and the dimerization initiation site (DIS)(21-23), which is a key recognition site for HIV genome dimerization prior to capsid formation and viral maturation.

A growing number of genetic disorders are being associated with RNA dependent processes. Aberrant mRNA function(s) has been linked to the genetic diseases Fragile X syndrome, Huntington's disease, and Autism to name a few (an extensive list of such diseases can be found in the review(24)). In these genetic disorders it is known that specific mutations lead to defective RNA-protein recognition that disrupt normal alternative splicing(24). Cancer therapies aimed at targeting RNA have gained traction recently due to the discovery that cellular levels of miRNA are

directly related to cancerous cell phenotypes and therefore could potentially be targeted for anti-cancer therapies(25-28). Also, the ability to target telomerase RNA with small molecules and oligonucleotides has been explored as a potential avenue for anti-cancer development(29,30). In fact, the Geron corporation has developed a human telomerase binding oligonucleotide that has proven successful in phase III clinical trials(30).

Targeting RNA for drug development could alleviate some of the difficulties in inhibiting protein function(s). While the percentage of the proteome that is druggable is a matter of debate, a recent study indicated that all FDA approved small molecule drugs target only 207 proteins, 50% of which are G-protein coupled receptors(31-38). This number is strikingly small compared to the 1620 known proteins that are directly linked to genetic disease and the hundreds of thousands of proteins translated in human cells(35). A recent computational study suggests that out of the entire known proteome, only 15% contains a druggable binding pocket(33), and the disruption of protein-protein interactions, which feature large interaction surfaces, has proven quite difficult(39). Thus, while proteins have been and will continue to be the central focus for most drug discovery efforts in the near future, the increasing number of potential RNA targets provides an alternative route for therapeutic development.

The potential of RNA as an alternative drug target has spurred numerous studies aimed at understanding the molecular principles that underlie RNA-small molecule recognition. Figure 1.2 shows a timeline of important discoveries that have advanced the field of RNA-targeted drug discovery. While studies in the 1960s by Weinstein I. B. and Finkelstein I. B. showed the drug proflavine binds tRNA(40,41), appreciation for RNA as a *bona fide* drug target did not come to fruition until the validation that many known antibacterials including aminoglycosides(42-44), macrolides(45), tetracyclines(46), and

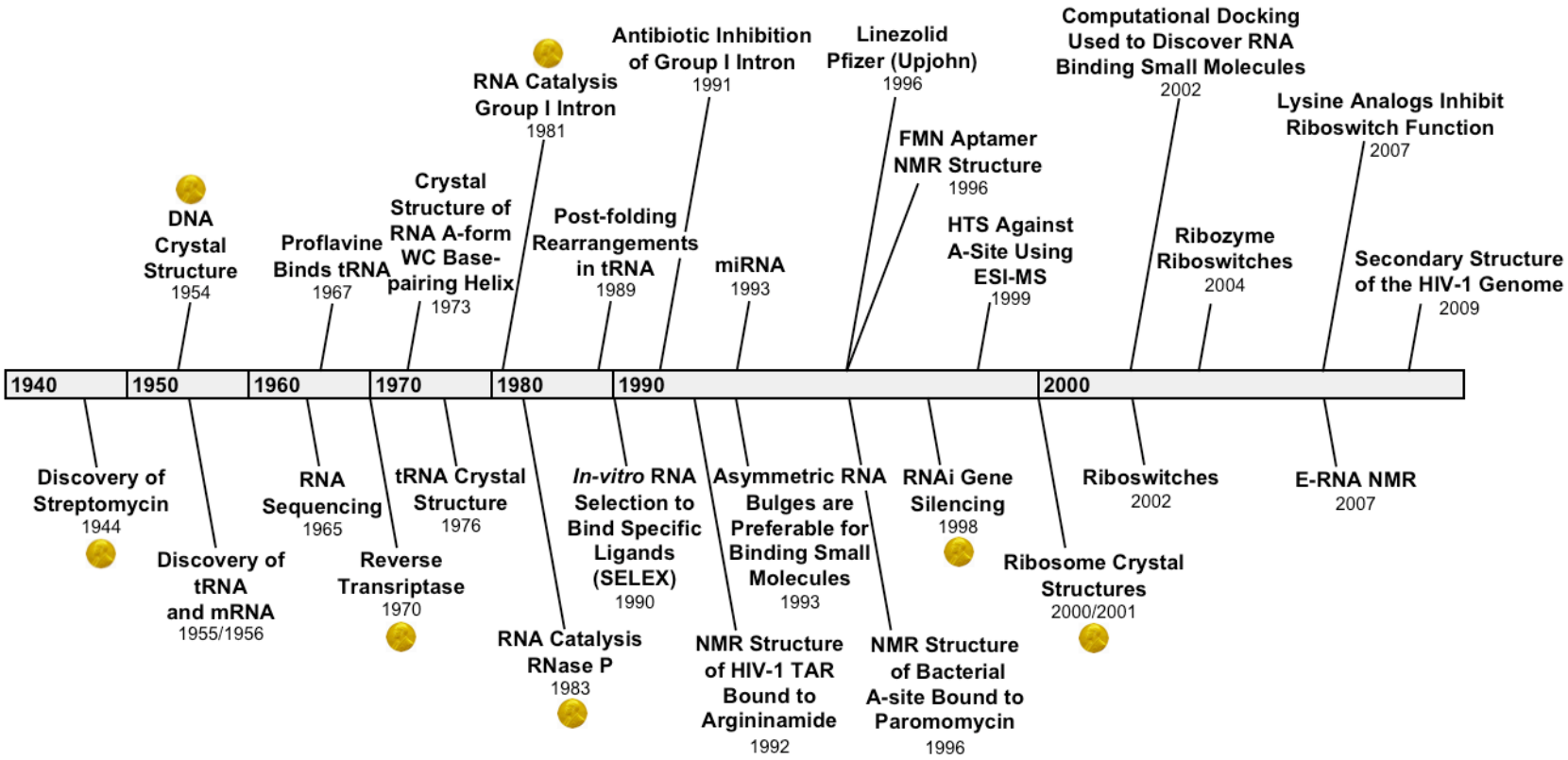


Figure 1.2. Timeline of important discoveries in RNA-targeted drug discovery. Nobel Prize medals indicate discoveries that were awarded a Nobel Prize. (2,3,6,7,40-42,47-75)

oxazolidinones(76) exert their function by binding rRNA. The mechanism of action has been extensively studied for aminoglycosides, where reports have shown that many bind the 30S ribosomal A-site and disrupt the translational fidelity resulting in the incorporation of the incorrect amino acid and thus the synthesis of aberrant proteins and ultimately cell death. These findings are encouraging for the field of RNA-targeted drug discovery, but they are also an example of the ideal situation since rRNA is catalytically active and abundant in cells. The "holy-grail" in RNA-targeted drug discovery is to specifically target a catalytically inactive low abundance RNA, which has yet to be realized.

Table 1.1. Examples of RNA therapeutic targets.(16,24,77-83)

Disease	RNA Type	RNA	Function
Bacterial Infection	rRNA	A-site	translation
Bacterial Infection	rRNA	Peptidyl-transferase center	translation
Bacterial Infection	mRNA	Riboswitch	transcription, translation, splicing, mRNA degradation
Bacterial Infection	ncRNA	RNase P	RNA cleavage
HIV	Viral RNA	Trans-activation response element	transcription
HIV	Viral RNA	Dimerization initiation site	genome packaging
HIV	Viral RNA	Rev-response element	genome export
Hepatitis C Virus	mRNA	Internal ribosome entry site	translation
Spinal Muscular Atrophy	ncRNA	SMN2 pre-mRNA	splicing
Amyotrophic Lateral Sclerosis	ncRNA	TARDBP	RNA splicing, transport
Wolcott-Rallison Syndrome	mRNA	EIF2AK3	translation
Prostate Cancer	mRNA	SNHG5	ribosome biogenesis
Cancer	ncRNA	miR-372	RNAi
Cancer	ncRNA	RBM5	RNA splicing

1.3 Current RNA Drug Targets

1.3.1 Antibiotics

Ribosome-targeting antibiotics remain the only validated success story in RNA-targeted drug discovery. To date, all of the FDA approved RNA-targeting drugs bind

rRNA. In fact, nine years ago the FDA approved the use of the gram-positive targeting small molecule linezolid (Pfizer), which is an oxazolidinone that is thought to prevent initiator tRNA binding to the ribosome and thus inhibit protein synthesis(62,84,85). Currently there are 1604 open clinical trials investigating antibiotics, some of which likely employ rRNA-targeting small molecules(clinicaltrials.gov).

Since the discovery of streptomycin in 1944(47,48) numerous studies have focused on developing novel rRNA-binding small molecule antibiotics. The advances in X-ray crystallography by Steitz T. A., Ramakrishnan V., and Yonath A. D. and other groups allowed the atomic resolution visualization of the ribosome in the apo-state and bound to various anitbacterials(67,69-71,86-95). Differences between the apo and small molecule bound ribosome structures answered many questions about rRNA-targeting antibacterial mechanism(s) and agreed with many of the previously determined rRNA-targeting small molecule antibiotic mechanisms, including the pioneering NMR work by Puglisi and co-workers(42,96-98). This Nobel Prize winning accomplishment is providing the basis for further structure-based design of new antibiotics. The realization that proposed models for antibacterial aminoglycosides targeting the A-site matched those revealed by the ribosome crystal structures (discussion of the inhibitory mechanism of aminoglycosides that target the bacterial A-site can be found in section 1.4) also highlights an important development for RNA-targeting drug discovery -- the modularity of RNA allows for extensive biochemical and biophysical studies of RNA targets outside of more biologically relevant contexts. This motivates consideration of other RNAs as targets for antibiotic development.

Many RNAs outside the ribosome are targets for small molecule antibiotics including tRNA(99-103), non-rRNA catalytic RNAs(78,104), and mRNA elements including riboswitches. Some studies are also focused on binding RNAs in the ribosome assembly pathway(105-110). Inhibiting aminoacyl-tRNA synthetase enzymes, which are

responsible for coupling the correct amino acid to the corresponding tRNA, is considered an attractive target for antibacterial development(99). Rather than targeting the enzyme, specific binding to the $tRNA^{phe}$ variable loop has been shown for neomycin, potentially indicating specificity to inhibit tRNA-amino acid charging. The catalytic nature

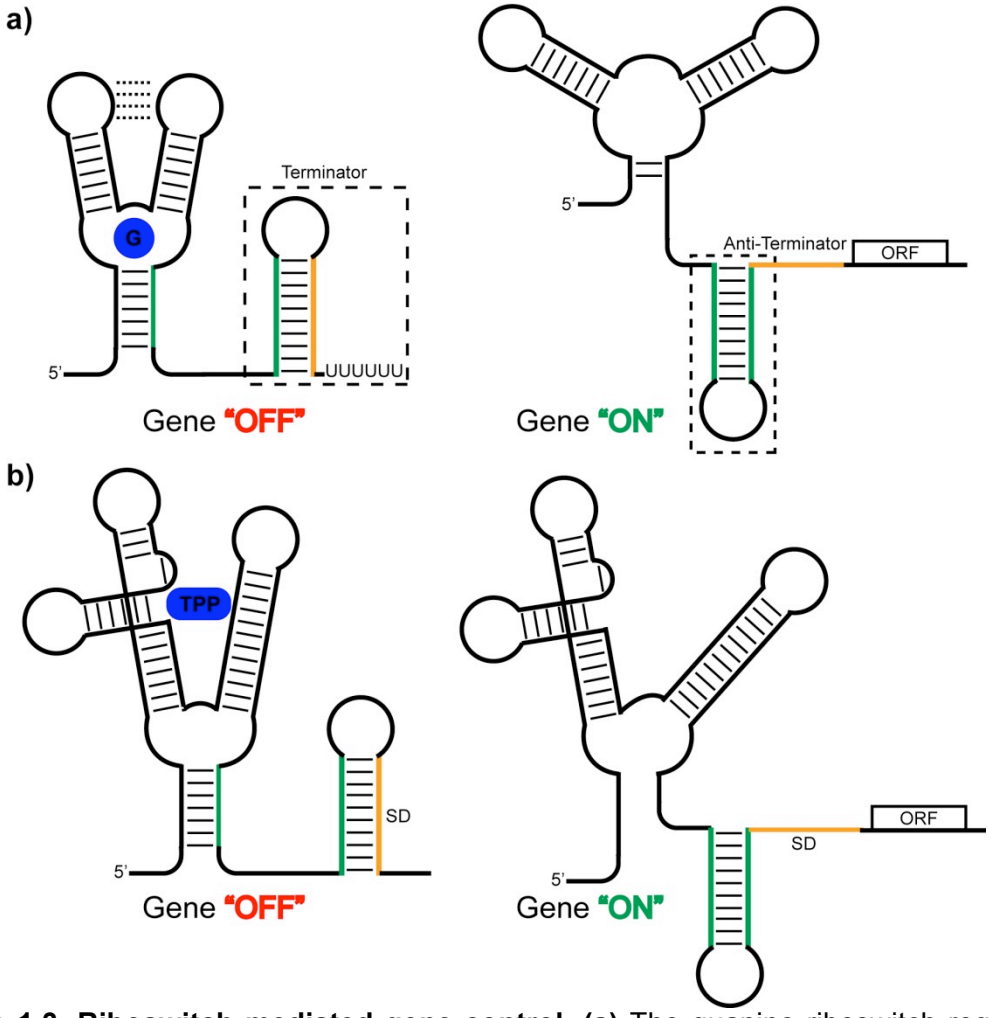


Figure 1.3. Riboswitch mediated gene control. (a) The guanine riboswitch regulates gene transcription by binding the guanine nucleobase resulting in the formation of a terminator hairpin that prevents transcription of the open reading frame(5) and **(b)** the TPP riboswitch regulates translation of the ORF by sequestering the Shine-Dalgarno (SD) sequence on binding the TPP metabolite(6).

of ribozymes affords easy readout for monitoring small molecule mediated inhibition. This combined with the multiple X-ray structures of ribozymes provides the basis for structure based drug design and an excellent case study for understanding RNA-ligand

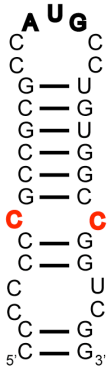
interactions. Small molecule inhibition of ribozymes has been extensively studied since 1991 with the first report of aminoglycoside inhibition of the group I intron(58). Since this initial report many studies have investigated antibiotic ribozyme particularly in the case of the hammerhead ribozymes(104) and the RNase P ribozyme(78).

Along with RNAs involved in antisense-mediated translational control(111), the regulatory functions controlled by riboswitches serve as a very attractive mRNA antibacterial target. Riboswitches are generally found in the 5' UTR of bacterial mRNA transcripts and bind specific cellular metabolites (e.g. guanine, adenine, S-adenosyl methionine, and lysine) to regulate gene expression at the transcriptional or translational level(112). The genes that are downstream of the riboswitch are usually involved in transport or synthesis of the cognate metabolite and thus are direct regulators of cellular metabolite concentrations. Most often if the metabolite is not present the gene is "on" and transcription or translation will persist, however when the metabolite is in sufficiently high concentration then the riboswitch will bind the metabolite and turn the gene "off" by forming an RNA structure that will suppress further transcription or translation(Figure 1.3). The structure that forms on binding the cognate metabolite is often in the form of a terminator hairpin (GC rich helix followed by a stretch of uridines) (Figure 1.3a) when inhibiting transcription, or results in sequestering the Shine-Dalgarno sequence (Figure 1.3b) to prevent ribosome recognition and thus translation of the mRNA transcript. A viable antibacterial drug-discovery approach would be to target these riboswitches thereby inducing a desirable change in metabolite regulation to kill infectious bacteria in the host. In a recent study, Breaker R. R. and co-workers found five lysine analogs that bind the lysine riboswitch within 40 fold of the native lysine binding affinity (360nM) (77). This study further showed that three of the compounds inhibit bacterial growth with minimum inhibitory concentrations ranging from 26-938μM. Mutating the riboswitch confirmed that these lysine analogs bind the aptamer domain of the riboswitch to inhibit

cell growth, which significantly increased the minimum inhibitory concentrations (1.56-5mM) but did not affect lysine binding. This also suggested that the binding mode of these small molecules was different than that of lysine. Thus, it is plausible that riboswitch mediated regulation is a kinetically and not necessarily a thermodynamically controlled process, explaining why relatively weak binders can inhibit bacterial growth likely through formation of kinetically trapped alternative conformations. This is also supported by other reports(113-117). Thus, while still an infant field for antibacterial therapeutic development, the ability and desire to target riboswitches will likely gain much more traction in the near future.

1.3.2 Anti-cancer

RNA-targets provide an alternate route for anti-cancer therapies. Various cellular processes are being targeted for anti-cancer small molecule development including RNA splicing (118), translocation(119), miRNA regulation(25-28), ribosome assembly(120), and transcription(121-124). For example, an attractive anti-cancer mRNA drug target that exists in mammalian cells is the thymidylate synthase (TS) RNA. The TS protein is involved in regulating the generation of thymidine monophosphate that becomes further phosphorylated to form thymidine triphosphate used in DNA synthesis and repair. Inhibiting the TS mechanism has been at the forefront of many pharmaceutical research companies for years as a target for anti-cancer therapy(125,126). While the TS protein can be targeted in cells, certain classes of TS inhibitors make the TS protein unable to regulate its own translation, resulting in unregulated expression. Thus, as the



Molecule	K_d(μM)
Neomcyin B	0.9
Paromomycin	2.3
Gentamicin	2.0
Tobramycin	2.2
Kanamycin B	2.7

Figure 1.4. TS RNA is an anti-cancer target. Aminoglycosides have been shown to bind the 1X1 internal loop (in red) with nM-μM affinity.

concentration of TS increases in cells, these inhibitors exhibit reduced efficacy(121-123). Rather than targeting the enzyme, an alternative approach would be to target the TS RNA. Briefly explained, the TS protein binds to the TS mRNA at two distinct sites, the first of which is a stem-loop RNA with two symmetric internal loops and a seven-nucleotide apical loop (Figure 1.4) (127,128). Located within this first TS protein binding site RNA is the AUG start codon(127,128). Thus, the likely transcription-regulation mechanism for this protein is to bind its own RNA, thereby sequestering the start codon as a negative feedback for its own production(127). The first study targeting the TS RNA with small molecules to stabilize the TS RNA hairpin showed varying affinities for a library of aminoglycosides (K_d0.9-2.7 μM) and mutational studies indicate that the aminoglycosides bind the 1X1 internal loop and not the 2X2 or apical loops (Figure 1.4)(124).

1.3.3 Anti-viral

Some of the classically studied RNAs are directly involved in the life-cycles of viruses and for many years have been sought as anti-viral therapeutic targets. As previously mentioned, three such RNAs are the TAR, RRE and DIS in the HIV genome. Briefly, TAR is located in the 5' LTR sequence of the HIV genome and forms a stem-loop structure that serves as a scaffold for binding regulatory proteins, namely Tat, Cyclin T1, and Cdk9, that together ensure the fidelity of gene expression by hyper-phosphorylating RNA polymerase II(Figure 1.5)(129,130). Anti-HIV drug discovery efforts have focused on binding TAR or Tat and thus preventing their association leading to decreased viral efficacy. Tat forms specific and tight interactions (K_d low-sub nM) with TAR through its arginine rich motif (69,131-135), and numerous reports have focused on understanding the TAR-Tat recognition mechanism and the mode of action for many known inhibitors(59,136-147) (see further discussion in section 1.8). While there is debate over

the action of many TAR-binding small molecule inhibitors, it is likely that many small molecules allosterically inhibit TAR-Tat binding by locking TAR into alternate conformations that prevent Tat binding(148).

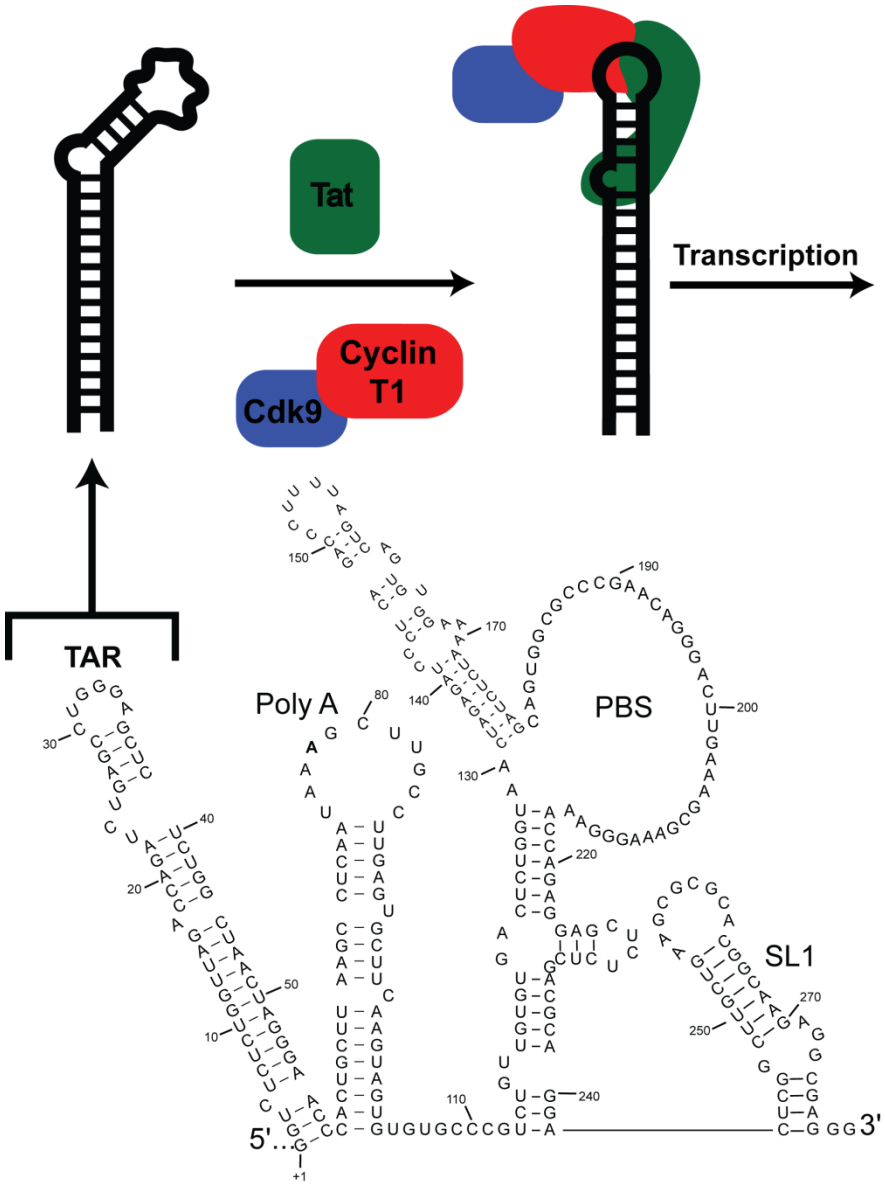


Figure 1.5. TAR regulates HIV transcription fidelity. TAR is located in the 5' LTR of the HIV genome (on bottom) and regulates HIV Transcription through the formation of a ribonucleoprotein complex. In the absence of TAR truncated transcripts are produced, making it a highly sought anti-HIV drug target.

1.4 RNA Structural Adaptation on Small Molecule Binding

Structure-based drug design is one of the most ambitious approaches for drug discovery and is based on rationally using knowledge of a molecule's 3D structure in identifying small molecules capable of binding that structure. Highly flexible RNA structures pose a unique challenge in this regard because even related ligands can bind very different RNA conformations that cannot, in general, be determined *a priori*. The earliest observations of post-folding rearrangements in RNA structure were made in the 1980s by Steitz T. A., Soll D. and co-workers who showed that tRNA undergoes changes in the alignment of its helices on binding to glutaminyl-tRNA synthetase(55). In subsequent NMR studies, Williamson J. R. and Puglisi J. D. showed that even a small ligand, such as argininamide, can induce large changes in the conformation of TAR (TAR) involving large changes (~40º) in the orientation of A-form helices coupled with local formation of a base-triple (59). In an elegant series of NMR studies, Puglisi and co-workers demonstrated that small molecule aminoglycoside antibiotics exert their activity by inducing conformational changes in the A-site RNA, underscoring the importance of structural transitions in RNA targeting (42,96,98,149). Briefly, the A-site (Figure 1.6) is located in the small 30S subunit of the ribosome and is responsible for ensuring the transcript codon matches the incoming tRNA anticodon to ensure correct amino acid incorporation. If a match is present then the two critical internal loop adenine bases (A1492 and A1493) flip out and permit incorporation of the amino acid, which is attached to the tRNA, into the polypeptide chain. When a match is not found then the adenine bases remain looped inside thus rejecting the tRNA and another matching event occurs. The A-site is the target for many antibiotics (e.g. paromomycin – K_d~200nM)(150), which exert their activity by binding the internal loop and stabilizing a looped out conformation for the two universally conserved adenine residues A1492 and A1493, thereby allowing incorporation of incorrect amino acids leading to the production of aberrant proteins and

eventually cell death. Remarkably, a single substitution of internal looped adenine A1408 with a guanine residue, which is found in humans and mutated bacterial strains, confers antibiotic resistance through mechanisms that are not fully understood.

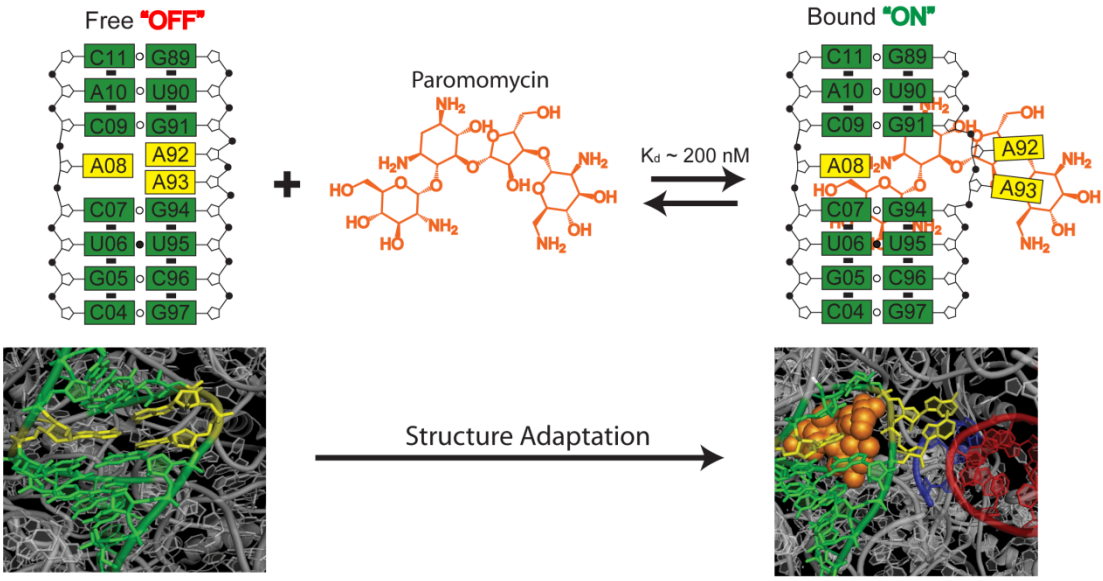


Figure 1.6. A-site structural adaptation on small molecule binding. Translation regulation by the A-site is inhibited by the aminoglycoside paromomycin (orange). On binding paromomycin the internal loop adenines (yellow) loop outside the helix and interact with the tRNA (red) and the mRNA transcript (blue) resulting in incorporation of aberrant amino acids in the polypeptide. Note that A-site base numbering is truncated for clarity (e.g. 1492 is shown as 92).

An extreme example of structural rearrangements on interaction with a small molecule is observed in RNA aptamers (Figure 1.7). Non-biologically native RNA aptamers are developed to specifically bind a small molecule ligand using the SELEX protocol(56,57). Largely unstructured in the free state, these aptamers fold on contact with the ligand to form complex tertiary structures, with multiple specific RNA-small molecule interactions that confer tight binding affinities (nM-μM)(151). Often times the small molecule ligand is almost completely buried inside the RNA. Analogously, large structural rearrangements are seen in the naturally occurring aptamer domains of riboswitches. As previously mentioned, when the concentration of the cognate metabolite is sufficiently high the riboswitch aptamer binds the ligand via numerous

specific interactions. On binding the metabolite, the aptamer induces formation of an RNA structure that impedes transcription or translation(112). However, in the absence of the metabolite an alternate aptamer conformation is adopted and transcription or translation proceeds (see Figure 1.3 for an example of the riboswitch mechanism)(5,112,152).

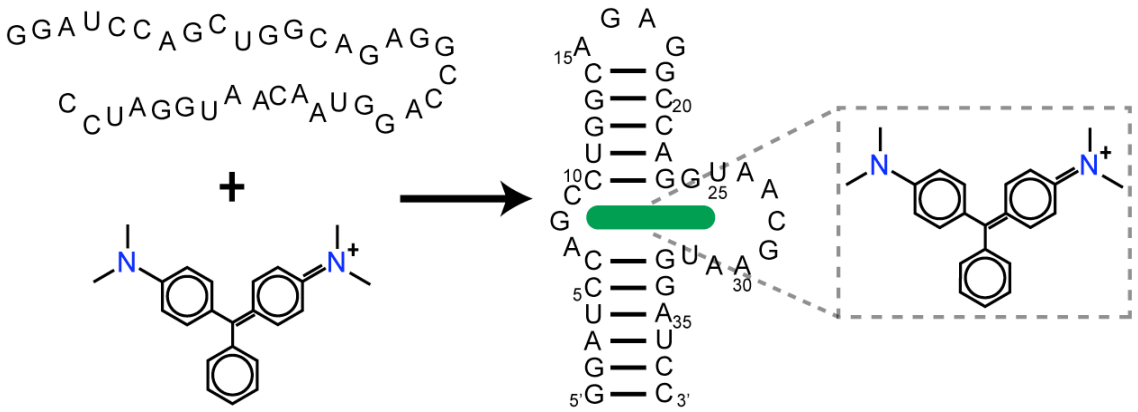


Figure 1.7. Structural rearrangement of the malachite green aptamer. On interaction with the small molecule malachite green, the unstructured free RNA folds into a complex tertiary structure that exhibits specific stacking and hydrogen bonding interactions(154).

Given that many small molecules exhibit their activity non-competitively by stabilizing inactive RNA conformations(155), identifying RNA-targeting small molecules presents a difficult challenge because high affinity does not guarantee high activity. For example, fluorescence studies provide strong evidence that the aforementioned antibacterial activity of aminoglycoside antibiotics that bind the ribosomal decoding-site, some of which exhibit nM binding affinity, is not determined by their binding affinity, but by the dynamics of the internal loop adenines in the bound RNA structure(156). Likewise, studies have shown that neomycin B inhibits the interaction between TAR and Tat by binding to TAR and preventing necessary conformational changes from taking place(146). Thus, the development of methodology for identifying allosteric small molecules is of great importance to advance the field of RNA-targeted drug discovery.

Numerous studies over the past two decades have focused on understanding the physical principles guiding RNA structural changes that take place on ligand binding(139,151,157,158). A deeper understanding of these transitions will not only provide a more thorough description of the complex mechanisms that guide RNA-mediated biological processes, but will also serve to advance the field of RNA-targeted structure based drug design. Two mechanisms have been used to describe these structural adaptations(139,158,159). The “induced-fit” model, first described by Koshland D. E. in the 1950s to explain the decreased rates of protein enzyme catalysis observed when using substrate analogs(160), was used to explain many of the initial discoveries of RNA structural changes on ligand binding. Using the induced-fit model it was assumed that the bound state RNA conformation is not largely populated in the unbound state because it is thermodynamically unstable and/or separated by very high kinetic barriers; factors that are overcome via favorable interactions with ligands. However, there was also evidence that ligands do not induce the change in RNA structure but instead capture existing RNA conformations. An early example of “conformational selection” or “tertiary capture”, first introduced by Foote J. and Milstein C.(161) to explain kinetic data on a set of antigen-antibody interactions, was the observation that free-state conformations of RNA stem-loop structures are stabilized on binding the ColE1 Rom protein(162). Subsequently, Weeks K. M. and Cech T. R. also used conformational selection to explain adaptive recognition in the assembly of the CBP2-group I intron ribonucleoprotein complex(163,164). More recently, Al-Hashimi H. M. and co-workers used Elongated RNA (E-RNA) NMR to provide compelling evidence that the conformational selection model best describes TAR small molecule binding events(74). It is likely that many RNA-small molecule binding events are governed by a conformational selection mechanism. Thus, knowledge of free-state RNA dynamics is critical to discovering novel RNA-targeting small molecules.

1.5 Structural Basis for RNA-Small Molecule Recognition

In the early 1990s, NMR structures were reported for RNA in complex with small molecules that began to reveal the structural basis for RNA-small molecule recognition. Importantly, these studies allowed for interrogation of small molecule properties that lead to specific interactions with RNA. Structures of TAR in complex with argininamide revealed large global and local conformational rearrangements on binding that give rise to a network of specific interactions. Unbound TAR exhibits an inter-helical bend angle of ~47^{o}, however on binding argininamide rearranges into a coaxially stacked state that is nearly linear with a bend angle of ~11^{o} (59,74,142,157,165). The global rearrangement is accompanied by a conformational change in the inter-helical bulge where the residue U23 moves from being stacked inside the bulge with A22, to being looped into the major groove of the upper helix forming a base triple with bases A27 and U38. These RNA structural rearrangements result in specific interactions that confer binding specificity including direct stacking interactions between U23 and argininamide, hydrogen bonding between the guanidinium group of argininamide and G26, and hydrogen bonding between argininamide and the phosphate backbone of A22 and U23(59). Subsequently, NMR structures of the A-site rRNA and other RNA elements in complex with aminoglycosides revealed the importance of hydrogen bonding, electrostatic interactions, and the complementarity between small molecules and the RNA surface(42,96,98,149,166). Studies of artificial and more recently natural RNA aptamers, which bind to a variety of ligands, also revealed complex RNA conformations giving rise to deep binding pockets and a multitude of specific small-molecule RNA interactions resulting in high affinity (nM-μM) binding(151,167,168).

Much like proteins, RNAs have many structural features that make it a suitable target for small molecules. DNA and RNA are made up of only four chemically similar building blocks that tend to form regularized structures such as helices. However, the

DNA B-form helix differs substantially from the RNA A-form helix. While the DNA B-form helix has a shallow and widened major groove making it amenable to small molecule

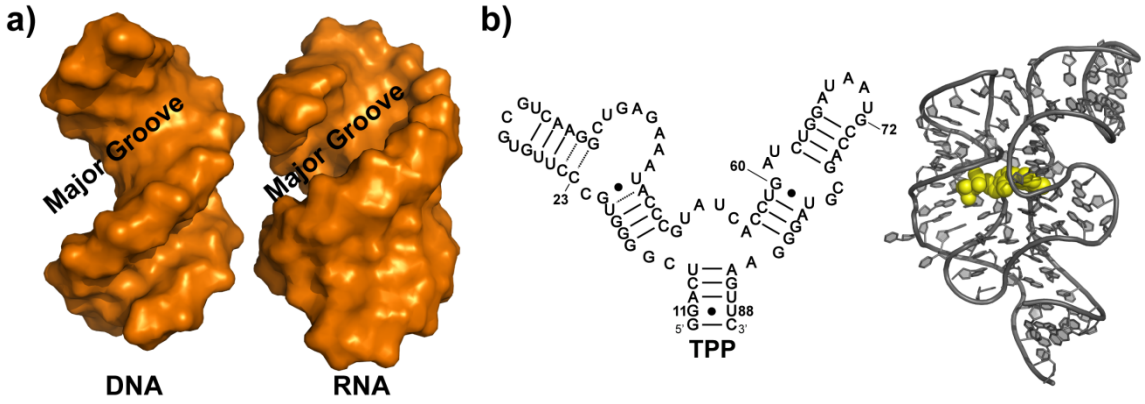


Figure 1.8. The higher order structure of RNA differs from DNA and provides more complex folding architectures. (a) The shallow and wide major groove of the B-form DNA helix makes it more amenable to small molecule binding compared to the deep and narrow A-form RNA major groove. **(b)** The TPP riboswitch exemplifies how junctions in RNA elements give rise to complex tertiary structures and provide highly specific small molecule binding pockets(169).

binding, the RNA A-form helix has a narrower and deeper major groove that is less accessible for small molecule binding(Figure 1.8a)(170-172). However, unlike DNA, the interruption of RNA A-form helices by bulges, internal loops, and junctions results in more complex folds with deeper and more accessible binding pockets capable of forming tight interactions with other RNAs(173,174), proteins(175), and small molecules(176) (Figure 1.8b). These tertiary structures give rise to complex electronegative surface potentials that are more similar to those of proteins compared to DNA(177), albeit the contours are predominantly negative potentials in RNA whereas proteins exhibit both positive and negative potentials. That said, the varying electronegativity seen across the known RNA tertiary structures could give rise to the necessary specificity needed to develop high-affinity RNA-targeting drugs.

Many RNA-binding small molecules resemble nucleic acid structures in that they contain pyrimidine-like moieties, purine-like moieties, phosphates, or sugars (reviewed in

(81)). These moieties specify the types of interactions known to confer specific RNA binding. Hydrogen bonding likely gives rise to the majority of specificity in RNA-small molecule complexes. The large amount of energy held within hydrogen bonds (~1-3 kcal/mol), which are often formed with exposed nucleobases and/or the phosphate backbone, largely determines the small molecule binding conformation(s). An example of such specificity is seen in the guanine and adenine riboswitches where out of nine potential hydrogen-bond donors and acceptors in the nucleobase ligand, eight are occupied by hydrogen bond interactions with the RNA, three of which are in exact Watson-Crick (WC) geometry between the ligand and the RNA(Figure 1.9a)(152,153,178). Batey R. and co-workers show that mutating the C74 residue, which forms WC base-pairing with the guanine ligand, to U74 changes the specificity of the guanine riboswitch to bind adenine through the formation of an adenine-U74 WC base-pair(113,179). They also show that the guanine analog benzimidazole, which has only two of the nine hydrogen bond donors and acceptors, does not bind the riboswitch, likely due to the lack of specificity(179). Aminoglycoside recognition by the A-site also highlights the importance of hydrogen bonding in small molecule recognition where ring-1 of the aminoglycoside paromomycin forms two hydrogen bonds with the WC face of A1408 and on mutation of A1408 to a guanine eliminates paromomycin binding(180-182). It should be mentioned that water mediated hydrogen bonds are also highly critical to recognition of RNA-binding small molecules especially in solvent-exposed complexes, which are often seen in the aminoglycoside-A-site crystal structures(150).

Electrostatic interactions also play a critical role in small molecule-RNA recognition. Cationic functional groups in small molecules often compete for metal ion binding sites. For example, studies show that tobramycin competes for Mg^{2+} binding to $tRNA^{Asp}$(103), and neomycin and pentamidine compete for Mg^{2+} or Ca^{2+} binding to the group I intron(183,184). As expected, the addition of metal ions significantly reduces the

binding affinity of these small molecules. A-site recognition of aminoglycosides is another example of a highly electrostatically dependent binding mechanism, where most of the amino groups that are charged at physiologically pH(185) are directly involved in dipole-dipole or charge-charge interactions with the RNA backbone phosphates and water molecules(68,166,180-182). Electrostatic interactions via metal ion coordination also plays a major role in RNA-small molecule recognition as seen in the TPP riboswitch, where the ligand phosphates are recognized by small molecule-RNA bridging Mg^{2+} ions(6).

Stacking (π-π) interactions that are critical for the formation of DNA and RNA helices and complex tertiary structures are often observed in RNA-small molecule complexes(1,49,53,186-190). In fact, all riboswitches that bind nucleic-acid-like small molecules exhibit significant RNA-ligand π-π stacking interactions. The guanine and adenine riboswitches, which recognize their respective cognate ligands via hydrogen bonding and extensive π-π stacking on both sides of the purine ring (Figure 1.9b), result

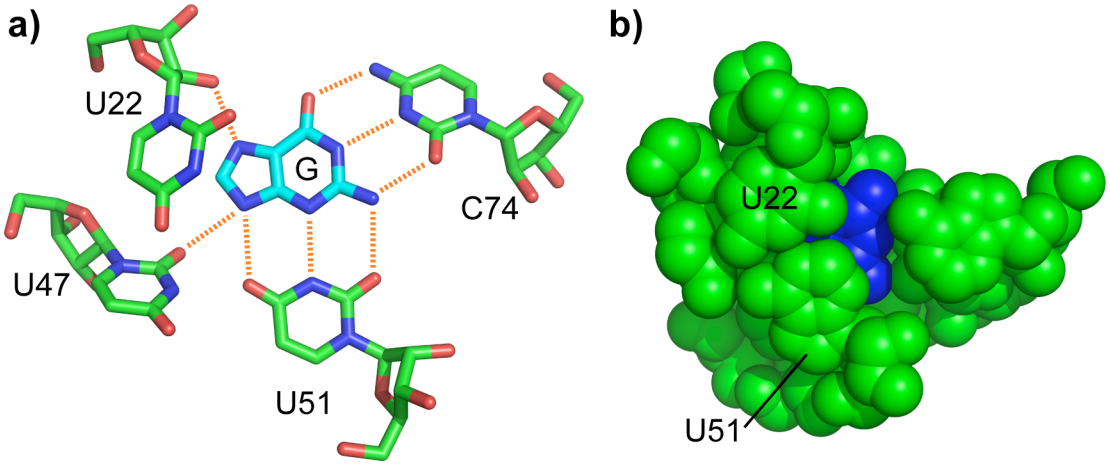


Figure 1.9. Hydrogen bonding and π-π stacking interactions are critical for conferring small molecule-RNA binding specificity. (a) Eight out of nine possible hydrogen bond donors and acceptors specify the guanine orientation in the guanine riboswitch binding pocket. **(b)** Significant π-π stacking interactions are also observed in the guanine riboswitch, depicted by the space-filling representation, resulting in ~98% burial of guanine(153).

in ~98% burial of the small molecule(153). Significant π-π stacking is also observed in the binding of acetylpromazine by TAR where the anthracene-like moiety stacks with U25, G26, and U40 in the bulge binding pocket(145). Other somewhat ignored stacking mechanisms are the CH-π and cation-π interactions which have been observed between ring I of aminoglycosides on binding the A-site(166) and $GlcN_6P$ and a guanine in the GlmS ribozyme(191-193), and in TAR recognition of argininamide(194,195).

With "drug-like" guidelines in place, the questions remain: do these parameters permit the design of RNA-binding drugs given the types of known interactions listed in this section, and can a small molecule be designed to bind a specific target RNA? Using a library of known clinically tested small molecules, Lipinski laid out the classic Rule of 5(196) (<5 hydrogen bond donors, <10 hydrogen bond acceptors, molecular weight<500, LogP<5), which describes necessary small molecule "drug-like" properties. While the recently FDA approved oxazolidinone linezolid(197) meets the "drug-likeness" criteria, many of the known RNA-binding small molecules violate the Lipinski rules. Difficulties in specifically binding a target RNA arise from the somewhat ubiquitous nature of many RNA interaction surfaces and the abundance of tRNA, estimated to comprise ~15% of cellular RNA(198). A major challenge in RNA-targeted therapeutic development will be to advance the design of small molecules with the necessary functionalities that meet the pharmacological criteria developed over the years of protein-targeted drug discovery that can bypass off-target binding events.

1.6 Techniques for Discovering RNA-binding Small Molecules

Contemporary tools for drug discovery have been designed to target proteins and are often inadequate for targeting RNAs, which have unique structural and functional properties (reviewed in(111)). Most high throughput drug screening methodology relies on measuring enzymatic activity, which is absent for most RNA drug targets. While a

variety of assays can be used to probe binding of small molecules to RNA targets, most cannot be implemented in a high throughput manner particularly for flexible and structurally delicate RNAs. To date, very few studies have been reported that successfully utilize modern high-throughput screening technology to identify small molecules. The reported high-throughput screens have been focused on well characterized RNA targets such as TAR(199,200) and the bacterial A-site(66,201), and implementation of a *de novo* high-throughput screen has not been reported. Initial attempts to use microarrays to detect RNA-small molecule binding in a high throughput manner failed to reproduce the rank-order of small molecules that bind the A-site RNA in solution likely because the glass surface, linker and linker position substantially affect the RNA binding properties of small molecules, again underscoring the structural delicateness of highly dynamic RNA systems(201). The inability to use high-throughput screening techniques against RNA targets explains the largely unexplored chemical space of RNA-binding small molecules.

The most common biochemical assays aimed at identifying RNA-specific small molecule binders utilize fluorescence. Often a fluorescent tag is inserted into an RNA sequence or appended to the terminus with the premise that conformational rearrangements on small molecule binding can be monitored by the changes in fluorescence intensity. Due to the chemical similarity of 2-aminopurine to pyrimidines, the fluorophore is often substituted in the RNA sequence giving site-specificity and has been used to monitor ligand binding in numerous studies(202-206). In the event that the binding site is not known for a particular RNA system a bulkier fluorescent tag, such as fluorescein or pyrene, is often attached to the terminus and in some cases to the 2' position of the ribose(73,207-211). Although these assays are relatively straightforward and provide an accurate measurement of small molecule binding, a major flaw in both of these fluorescence-based assay examples is the need to modify the RNA construct in

order to monitor small molecule binding. An alternative to labeling the RNA with a fluorescent tag is to fluorescently label a native binding molecule and monitor its fluorescence on competition with an inhibitor(212). While fluorescently labeling a native binding molecule can overcome the need to chemically modify the target RNA, not all systems are amenable to such methods for *de novo* drug discovery. Other techniques for measuring binding, including surface plasmon resonance, isothermal titration calorimetry, NMR, and footprinting, cannot be implemented in a high throughput manner and are more appropriately used as validation rather than lead discovery tools.

Computational docking provides the means to rapidly screen millions of small molecules against an RNA target, however prior knowledge of the 3-D RNA structure is required. Using a 3-D static RNA structure as a small molecule receptor, computational docking simulates the binding of a small molecule to the target by optimizing the interactions based on a force-field that describes intermolecular interactions (e.g. hydrogen bonding, electrostatic interactions, Van der Waals contacts) and small molecule properties (e.g. configurational entropy and solvent dissociation on binding) to determine the optimal small molecule binding pose. Although computational docking has proven quite successful in protein docking studies, the much larger conformational rearrangements that occur on RNA-ligand binding events prevent widespread success.

Numerous successes have been realized from protein docking efforts(213,214). For example, small molecule inhibitors against DNA gyrase(215), tyrosine phosphatase(216), dihydrodipicolinate reductase(217), and HIV integrase(218,219), which led to the FDA approved drug Isentress (Merck & Co.), were discovered using computational docking. In a more recent study(220), Jorgensen W. L. and co-workers docked 41,000 small molecules into the macrophage migration inhibitory factor, which is a cytokine that binds the cell surface receptor CD74 and plays key roles in inflammatory diseases and cancer(221,222). Out of the 23 commercially available small molecules

found 11 were confirmed to inhibit protein-protein binding with IC_{50} values <5 μM. The ability to predict protein-binding small molecules based on crystal structure or homology model receptors is likely in part due to smaller dynamic motions in protein binding pocket(s) and the use of protein-centric small molecule virtual libraries.

Despite the limitations in accurately accounting for RNA structural adaptation on small molecule binding, a few success stories have been reported. James T. L. and co-workers have reported two such studies, one targeting TAR(72,145,223) and a second where they developed a docking software named MORDOR(224). In the former study a static structure of TAR was used to screen against 181,000 small molecules in which they identified acetylpromazine that binds TAR with a K_d of 270 μM and inhibits TAR-Tat mediated HIV transcription *in-vitro*. In the latter report James T. L. and co-workers implement a "fully flexible" docking scheme that first holds the RNA static while docking the small molecules, and then implements an all atom docking simulation that allows for RNA and small molecule fluctuations simultaneously with restricted movement of the RNA atoms in an attempt to account for structural adaptation on binding(30). Using this protocol they screened 5750 small molecules against the human telomerase RNA and found 48 small molecules that bind the RNA, which were confirmed using saturation-transfer difference NMR experiments (no binding affinities were reported). Other studies have focused on developing RNA specific docking software by modifying the existing Autodock code with knowledge based potentials(225) and another using empirical potentials that include RNA-specific interaction terms (e.g. π-π stacking)(226). The most recent study reported modifications to the DOCK6 software that implements high-level solvent approximations and RNA flexibility as a subsequent docking step analogous to that reported by James T. L. to account for local adaptation on small molecule binding(227). In this report, they find that the accuracy of predicting the correct docking pose of a small molecule compared to the experimentally determined crystal or NMR

structure strongly depends on the number of small molecule flexible bonds, with the most accurate predictions for rigid molecules. This was also discovered in the Al-Hashimi lab and is discussed in Chapter 4.

While recent developments make it possible to accommodate limited flexibility in proteins(228) with little sacrifice in computational speed, no method exists to accommodate the inherently much greater RNA flexibility. Indeed, the current inability to predict how RNA structures adapt following small molecule binding is a major obstacle for identifying RNA-binding small molecules using computational docking. Recent developments in the Al-Hashimi lab have resulted in the generation of the first ever experimentally generated structural ensemble of TAR providing the necessary global and local conformational details to accurately predict small molecule-RNA binding affinities and is the focus of this dissertation.

1.7 Structure, Dynamics, and Intermolecular Interactions Using NMR

The site-specific resolution afforded by NMR is a unique advantage over other techniques and provides the ability to critically investigate interactions between proteins or nucleic acids and ligands. Chemical shifts are exquisitely sensitive to their electronic environment and perturbations can easily be monitored using NMR. They are often the first step to identify critical regions of macromolecules involved in ligand binding. Experiments to more accurately identify direct RNA-ligand or protein-ligand interactions include NOEs, saturation transfer, transfer-NOE, and cross relaxation experiments(229-231). In Appendix 1 we present a study on the oligomycin sensitivity conferral protein where binding to the small molecule inhibitor Bz-423 is investigated by monitoring chemical shift perturbations and cross-relaxation. Using these two techniques we are able to identify the most likely small molecule binding site and provide insight into the inhibitory mechanism.

The ability to monitor structural and dynamic fluctuations from ps-m timescales makes NMR one of the most powerful techniques to study the behavior of biomolecules. Classic spin relaxation measurements monitor the nuclear relaxation rates as a result of thermal redistribution (R_1), inhomogeneity of the local magnetic field (R_2), and the inter-nuclear dependence of saturation on relaxation rates between bonded species (NOE)(232). These techniques have been widely used to probe the ps-ns timescale motions of biomolecules, and information regarding site-specific internal motions can be obtained using these relaxation rates in a model-free analysis(233,234). Relaxation dispersion measurements, which utilize an external radiofrequency field to spin-lock magnetization in the rotating frame ($R1\rho$) (235,236) or implement a train of 180^{o} pulses with precisely chosen delay times (CPMG)(237,238), are sensitive to exchange processes occurring on μs-ms timescales. Other NMR techniques used to probe slower motions include: ZZ exchange, which monitors the exchange of longitudinal magnetization between multiple populated states on ms-s timescales(239); lineshape analysis, which relies on monitoring changes in lineshapes as a result of external perturbations (e.g. temperature, denaturant, ligand) and can be used to probe ms-s timescale processes(240); and hydrogen-deuterium exchange methods, which probe s-m timescales and have been widely used in monitoring the global stability and local dynamic fluctuations in proteins(241,242). Residual dipolar couplings (RDCs) and residual chemical shift anisotropies (RCSAs) can be used to probe motions on the sub-ms timescales(243-246). Unlike the aforementioned methods, RDC and RCSA measurements report on the orientational distribution of bond vectors and a common reference frame, often the external magnetic field, and thus relative to one another.

1.7.1 Residual Dipolar Couplings

RDCs arise due to incomplete averaging of the dipolar interaction in partially aligned molecules(243,245,246). The time-averaged Hamiltonian between two spins (I and S) separated by a distance *r* is given by,

$$\overline{H}_{DD}^{hetero} = -\left(\frac{\mu_o}{4\pi}\right)\frac{\gamma_I\gamma_S \mathrm{h}}{2\pi r^3}\left\langle 3\cos^2\theta - 1\right\rangle \vec{I}_z \cdot \vec{S}_z, \qquad [1.1]$$

where ***I*** and ***S*** are the spin angular momentum operators, γ_I and γ_s are the gyromagnetic ratios for the individual spins, μ_o is the permittivity in a vacuum, h is Planck's constant, *r* is the inter-nuclear distance, and θ is the angle between the inter-nuclear vector and the external magnetic field (Figure 1.10a). This gives rise to the expression for the dipolar coupling (D_{ij}) between two directly bonded spin ½ nuclei (*i* and *j*),

$$D_{ij} = -\frac{\mu_o\gamma_i\gamma_j \mathrm{h}}{4\pi^2 r_{ij,eff}^3}\left\langle \frac{3\cos^2\theta - 1}{2}\right\rangle, \qquad [1.2]$$

where r_{ij} is the inter-nuclear distance between the spins. The angular term in Equation [1.2] is the second rank Legendre polynomial, $P_2(\cos\theta_{ij})$, and is a function of the angle θ between the inter-nuclear vector and the external magnetic field (Figure 1.10a). The angular bracket denotes a time average over all angles sampled by the inter-nuclear vector due to both overall motions and internal motions occurring at timescales shorter than the inverse of the dipolar interaction (<ms). An effective bond length, $r_{ij,\ eff}$, subsumes the effects of distance averaging.

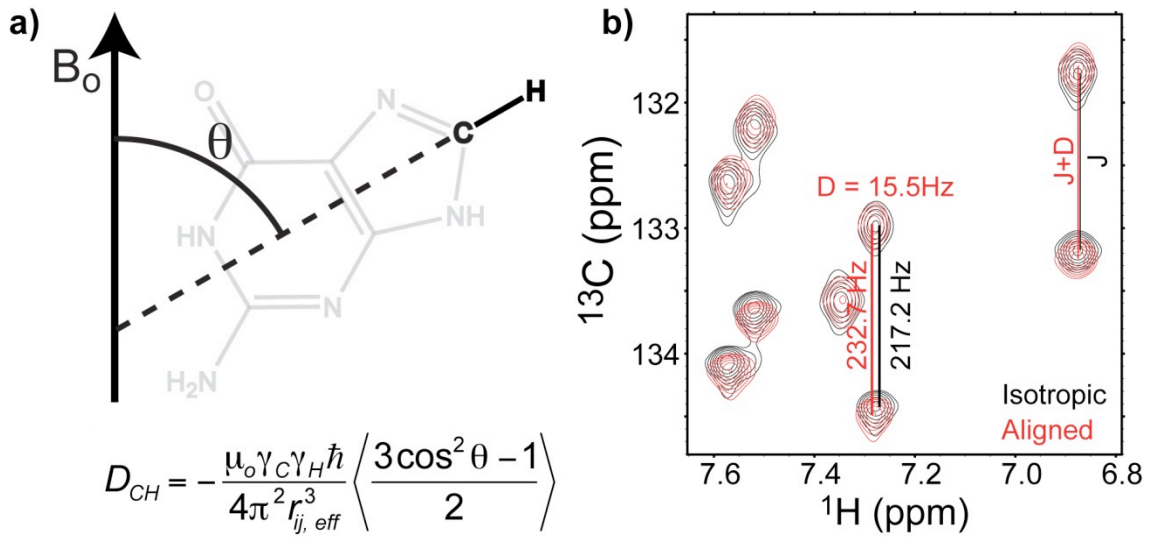


Figure 1.10. Measurement of RDCs in partially aligned molecules. (a) Residual dipolar couplings between spins *i* and *j* provide long-range constraints on the average orientation (θ) of the inter-nuclear bond vector relative to the external magnetic field (B_o). **(b)** Measurement of residual dipolar couplings (D) arise as new contributions to the splittings of resonances (J) observed upon partial molecular alignment.

Motional averaging will generally reduce the value of the angular term $\left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle$ and thus the magnitude of observed dipolar couplings (which can be on the order of kHz for non-reorienting directly bonded spins). When overall tumbling is isotropic, the angular term averages to zero, and dipolar couplings are not observed under solution conditions. However, if the molecule can be aligned in the magnetic field, the angular term will no longer average to zero. The greater the degree of alignment the greater the value of the angular term and magnitude of observed dipolar couplings. As is the case for through-bond scalar couplings (J), through-space dipolar couplings (D) effectively increase or decrease the average magnetic field at a given nucleus, resulting in splitting of resonances. Dipolar couplings are therefore often measured as new contributions to scalar couplings (J) that are observed under conditions of molecular alignment (J+D)(Figure 1.10b).

The measurement of RDCs in solution NMR relies on inducing an appropriate degree of alignment(247). Imparting a degree of alignment where approximately 1 in 10,000 molecules are on average completely aligned ($\leq 10^{-5}$) lead to RDCs that are too small compared to NMR line widths and do not allow measurements at a useful level of precision. At much higher degrees of alignment ($\geq 10^{-2}$), extensive dipolar couplings render RDCs too large and compromise the spectral resolution needed to analyze large biomolecules. In general, an optimum balance is achieved for when 1 in 1000 ($\sim 10^{-3}$) molecules are completely aligned. Under these conditions, many RDCs can be measured with optimal magnitude, precision, and with minimal sacrifice in spectral resolution. Alignment on the order of 10^{-4} can allow measurements of a smaller subset of RDCs with suboptimal magnitude/precision ratios.

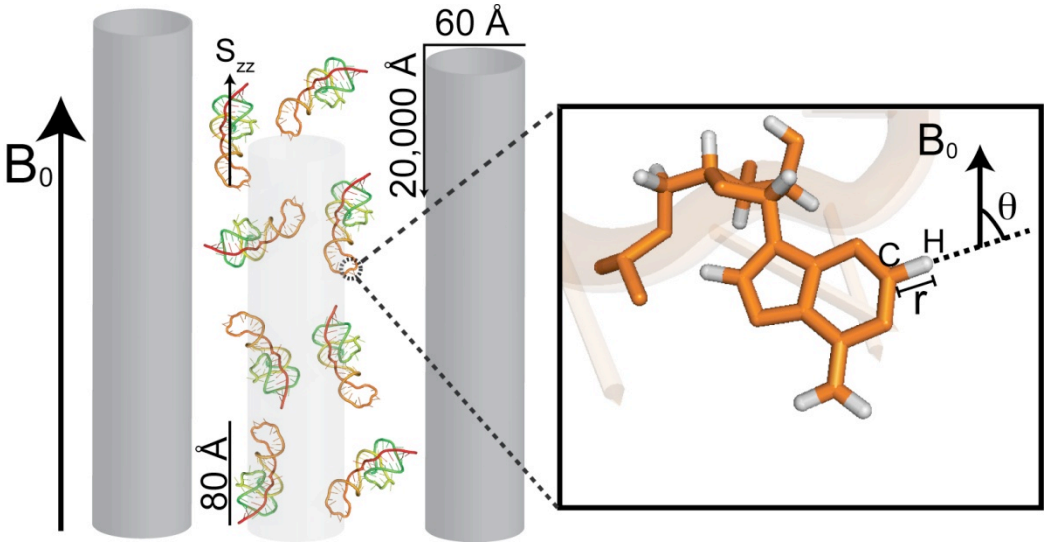


Figure 1.11. Partial alignment of nucleic acids using ordering media. Pf1 phage (gray) transmits order through steric and electrostatic mechanisms resulting the principal direction of order (S_{zz}) being oriented along the long axis of the molecule. Dimensions of phage and a representative dimension of an RNA are shown.

Most commonly, partial alignment on the order of 10^{-3} is introduced by dissolving the solute in an alignment medium (for reviews see(248,249)). Ordered media can transmit some of their order to solute molecules through mechanisms that are believed to involve a combination of steric obstruction and charge-charge interactions. This was

first demonstrated for liquid crystalline disc-shaped phospholipids called “bicelles”(245), which were originally used as a mimic of membrane bilayers in studies of membrane associated biomolecules(250,251). While the original neutral bicelle medium has been employed in studies of nucleic acids, other media have become more popular. In general, media with high tolerance to ionic strength are desired for nucleic acid applications. Also, positively charged ordering media are likely not useful as they may partake in unfavorable interactions. The most widely used medium to study nucleic acids is the filamentous bacteriophage (Pf1) (Figure 1.11)(252-254). Pf1 phage is negatively charged reducing the possibility for adverse interactions with nucleic acids and induces alignment through electrostatic and steric mechanisms(255,256). Owing to the uniform distribution of charge in polyanionic nucleic acids, the steric and electrostatic forces are believed to have a similar functional form(255,256). Consequently, nucleic acids generally align in ordering media with the principal direction of order (S_{zz}) oriented along the long axis of the molecule(Figure 1.11). In general, one expects positive alignment (S_{zz}> 0) with the S_{zz} direction being, on average, oriented parallel to the magnetic field. Experimentally, RDCs are computed from the difference in splittings measured in the absence and presence of ~20-25 or ~6-8 mg/ml Pf-1 phage for ~30 residue and ~80 residue RNAs, respectively (Figure 1.10b).

In nucleic acids, two strategies have been developed and applied for interpreting RDCs in terms of the structure and dynamics of RNA. The most widely used approach involves a variant of the simulated annealing approach in which individual helical domains are allowed to have independent order and/or alignment tensors. Here, RDCs and other experimental and non-experimental restraints are combined to simultaneously determine the local structure of molecular fragments as well as their order tensors(257,258). This approach can be generally applied to a variety of RNA fragments provided that a sufficient number of experimental restraints (including RDCs) are

measured and care has been taken to exclude RDCs that may have been attenuated by local motions. A second approach uses the idealized A-form helix geometry to model contiguous stretches of non-terminal Watson-Crick (WC) base-pairs(165,259). The validity of this approach was recently supported by a statistical survey(260) of 421 WC base-pairs in 40 unbound and bound RNA X-ray structures (solved with < 3Å resolution) and the 2.4 Å X-ray structure of the ribosome(261). This study showed that the local conformation of two or more non-terminal contiguous WC base-pairs can, for the purpose of determining order tensors using RDC accurately, be modeled *a priori* using a standard idealized A-form helix geometry(262,263). These WC base-pairs can be experimentally identified/verified using NOESY connectivity and *trans*-hydrogen bond J_{NN}-COSY type NMR experiments for directly detecting N-H--N hydrogen bonds(264,265). The study also developed approaches for taking into account structural noise in the A-form geometry in the determination of order tensors and provided evidence that local motions in such helical fragments will not compromise the accuracy of derived order tensors. By eliminating the need to solve the local helix structure, a number of applications are possible using the idealized A-form helix geometry. First, one can determine the relative orientation and dynamics of helices for RNAs that may be too large for complete high-resolution structure determination. Second, the order tensor analysis of RDCs can be conducted with high efficiency making possible systematic studies of how RNA's global conformational dynamics varies in response to changes in environmental conditions(165,266-270). The latter method is subsequently described.

The time-averaged angular term in Equation [1.2] can be expressed in terms of the time-independent orientation of an inter-nuclear vector relative to an arbitrary frame (α) and the five order tensor elements (S_{kl})(271,272),

$$\left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle = \sum_{kl=xyz} S_{kl} \cos(\alpha_k)\cos(\alpha_l), \qquad [1.3]$$

where α_n is the angle between the inter-nuclear vector and the n^{th} axis of the arbitrary frame. Assuming A-form geometries for a helix-bulge-helix RNA motif, the measurement of more than five independent RDCs for each helix allows determination of five elements of an order or alignment tensor(272,273). The order tensor consists of five independent parameters and describes the average alignment of the two helices relative to the external magnetic field. Three Euler angles specify a helix-fixed order tensor frame (S_{xx}, S_{yy}, S_{zz}) that depict the average orientation of helices relative to the external magnetic field. The average relative fragment orientations can be obtained by superimposing their order tensor frames, assuming that the helical fragments share a common view of the magnetic field direction when properly assembled (see refs. (165,271,274)). Two principal order parameters(274) define the degree (GDO = $\sqrt{\frac{2}{3}\left(S_{xx}^2 + S_{yy}^2 + S_{zz}^2\right)}$; $|S_{zz}| \geq |S_{yy}| \geq |S_{xx}|$) and asymmetry ($\eta = \frac{|S_{yy} - S_{xx}|}{S_{zz}}$) of molecular alignment and can be used to obtain information on sub-millisecond timescale inter-helical motions.

1.7.2 Domain-Elongation to Decouple Internal and Overall Motions

Use of RDCs in constructing RNA structural ensembles requires the ability to predict RDCs for a given proposed structural ensemble in which conformers interconvert at timescales shorter than the inverse of the dipolar interaction (<ms). This task is considerably simplified when one can (i) assume that internal motions are not correlated to overall alignment (the so-called “decoupling approximation”) and (ii) determine the overall order or alignment tensor governing partial alignment of the molecule. Studies

have shown that the decoupling approximation does not always hold for highly flexible and extended nucleic acid structures(275-280). In particular, changes in the relative orientation of helical domains can result in large changes in the overall RNA shape, which can in turn alter overall alignment of the molecule relative to the magnetic field (Figure 1.12a). The ensuing breakdown in the decoupling approximation can make it difficult if not impossible to compute RDCs for a proposed structural ensemble.

One approach developed in the Al-Hashimi H. M. laboratory to overcome this problem is to elongate a target helix such that the overall RNA shape and hence its overall alignment is dominated by the elongated helix and is no longer as sensitive to internal motions occurring elsewhere in the molecule (Figure 1.12b)(74,279). The elongation also serves to predefine the overall order tensor to be axially symmetric with principal direction oriented nearly parallel to the axis of the elongated helix. The RDCs measured in the elongated helix can be used to experimentally determine the overall order tensor (S_{kl}). Thus, the time-averaged angular term in Equation [1.4] can be expressed for the n^{th} conformer in terms of

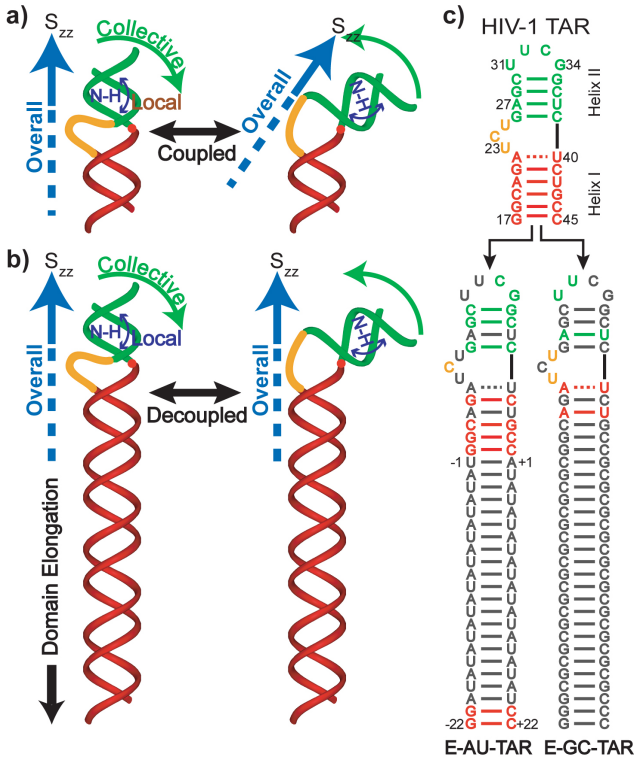


Figure 1.12. Decoupling internal and overall motions in RNA by domain-elongation. (a) Collective helix motions lead to coupled changes in overall alignment as described by the principal axis of the order tensor (S_{zz}). **(b)** Decoupling collective motional modes by domain elongation effectively renders S_{zz} parallel to the long-axis of the RNA. **(c)** NMR invisible elongation of TAR RNA is accomplished using isotopically unlabelled residues (gray). Two terminal GC base-pairs are added to helix-I in EI-TAR to maximize transcription yields.

the time-independent orientation of an inter-nuclear vector relative to an arbitrary frame (α_i) and the five order tensor elements (S_{kl}) describing overall alignment(271,272). For a structural ensemble consisting of *N* conformers that interconvert at timescales shorter than the inverse of the dipolar interaction (<ms), the observed RDCs will be a population-weighted average over the ensemble

$$\left\langle \mathrm{D}_{ij} \right\rangle = \sum_{n}^{N} p_n D_{ij}^{n} \quad . \qquad [1.4]$$

The elongation of an RNA target is typically preformed by extending the length of a terminal helix using a stretch of Watson-Crick base-pairs(281). To avoid increasing spectral overlap due to elongation residues, an isotopic labeling strategy is used to render elongation residues “NMR invisible”(281). Here, two constructs are prepared in which stretches of alternating unlabelled A–U/U-A (E-AU-RNA) and G–C/C-G (E-GC-RNA) base-pairs are employed for elongation using otherwise uniformly $^{13}C/^{15}N$ labeled G–C and A–U nucleotides, respectively (Figure 1.12c)(281). Two G-C base pairs are added to the terminal end of E-AU-RNA to facilitate RNA synthesis by *in vitro* transcription. The two constructs thus allow acquisition of NMR data over the entire RNA target while keeping elongation residues “NMR invisible”.

Care must be taken to ensure that the elongation does not perturb the structural and functional integrity of the RNA. This can be conveniently accomplished by comparing spectra of elongated and non-elongated RNA samples. In general, elongation of terminal helices is not expected to give rise to significant RNA structural perturbations. However, depending on the RNA context, elongation of other non-terminal helices, which can allow the measurement of independents set of RDCs (and RCSAs) that can

be applied to generate structural dynamics with enhanced spatial resolution(74), may cause unwanted perturbations and should be carefully analyzed.

The degree of helix elongation needed to sufficiently decouple internal and overall motions will vary depending on the RNA target. In general, the elongation should render a target helix at least 4-5-fold longer than other helices in the RNA. If a structure for the RNA target is available, one can perform simulations using programs such as PALES(282), using idealized A-form helices to model the elongated helix(74), to examine the degree of motional coupling. In these simulations, one computes the variance in the predicted overall alignment tensor relative to the elongated helix as a function of varying the relative orientation of other shorter helices in the RNA. In general, the principal direction of order (S_{zz}) should not vary more than ±7 degrees about the elongated axis and the asymmetry (η) should always be <0.15. The secondary structure of an E-RNA construct should be verified using a secondary structure prediction programs such as mfold 3.3 to make sure that no alternative secondary structures can form(283).

1.8 HIV-1 TAR: A Paradigm for Targeting RNA

TAR is an RNA stem-loop that is located at the 5' end of the HIV-1 genome that plays essential roles in the transcription of viral RNA and is a major target for developing anti-HIV therapeutics. As previously mentioned, the minimum TAR RNA consists of two helices that are linked by a trinucleotide bulge and capped by a hexanucleotide apical loop (Figure 1.5). The bulge and apical loop form two distinct protein recognition centers that can be targeted in the development of anti-HIV therapeutics(284). The viral transactivator protein Tat(129,285) binds residues in and around the TAR bulge (A22, U23, G26 and A27) and apical loop (G34) and enhances viral transcription by recruiting the positive transcription elongation factor b, which consists of Cyclin T1 (CycT1) and

Cdk9(286,287)(Figure 1.5). Tat can bind TAR in the absence of other cellular co-factors(288) whereas CycT1 does not bind TAR in the absence of Tat and the two proteins bind TAR cooperatively by mechanisms that remain poorly understood(129). Biochemical studies suggest that CycT1 interacts with the U31 region of the TAR apical loop(289) and forms other interactions with Tat. NMR structures of TAR bound to Tat-derived peptides and ligands indicate that Tat binds residues in and around the highly conserved bulge and that recognition leads to large structural rearrangements involving reorientation of the two helical stems accompanied by formation of a U23-A27-U38 base-triple(136-140). There are no structures for HIV-1 Tat (86-101 residues) in complex with TAR and there is controversy regarding the unbound structure of Tat, which are complicated by use of different conditions and constructs.

The discovery that small molecules, such as aminoglycosides, can inhibit TAR-Tat complex formation(19) and significantly reduce HIV-1 viral replication rates(17,18) spurred numerous studies that continue to date and are directed towards the development of agents that can bind TAR and disrupt its interaction with Tat (reviewed in (17,79)). The collective effort of many investigators over the course of eighteen years has resulted in 38 publications reporting TAR-binding small molecules. These studies identified 205 small molecules, which bind TAR with K_ds ranging between 5nM-5mM (80μM average), and 159 small molecules comprising 14 different motifs/scaffolds, which bind TAR and inhibit the TAR-Tat interaction with K_is ranging between 39nM-4mM or IC_{50}s ranging 20nM-500μM. Although these studies strongly suggest that TAR is a druggable target, only two of these compounds have been sufficiently potent to proceed to pre-clinical trials showing intraperitoneal acute toxicity LD_{50}s of 320 and 104 mg/kg in mice(290).

TAR exemplifies the major challenges encountered when targeting RNA. In the case of TAR, the difficulties in high-throughput screening can be appreciated by noting

that none of the TAR binding compounds were discovered by modern high-throughout experimental screens. Rather, most TAR binding small molecules have been identified by painstaking efforts that are focused on derivatizing specific classes of compounds, such as aminoglycosides, using various mimetics of the TAR protein target Tat, rationally designing small-molecule scaffolds that have certain electrostatic and 3D features, and in a more recent effort using split pool synthesis(199).

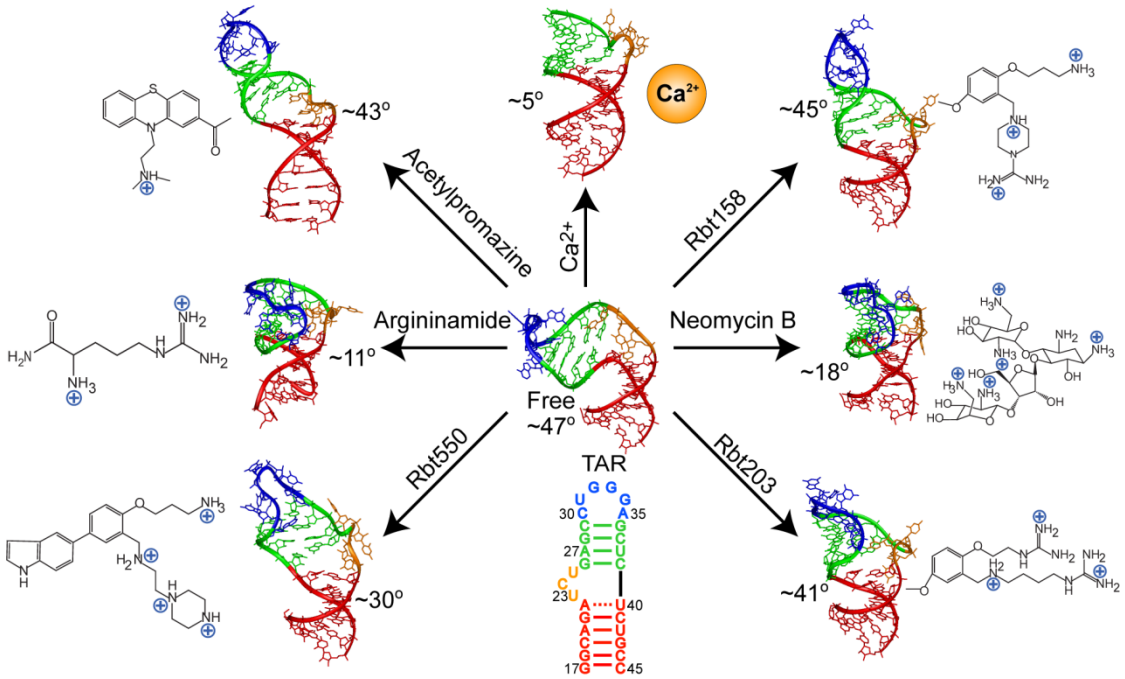


Figure 1.13. HIV-1 TAR structural adaptation. The solved TAR structures vary significantly in RNA conformation (inter-helical bend angle ranges from ~5° to ~47°) and small molecule identity(59,144-147,194,291). The large structural rearrangements observed for TAR are likely governed by a conformational selection mechanism.

A major obstacle in employing structure based drug design to develop specific TAR binders is the lack of knowledge of the free state beyond the calcium bound crystal structure(291,292). Numerous studies have shown that TAR undergoes dramatic conformational rearrangements involving large rigid-body movements of its two helical domains (averaging ~47° and as large as 71°)(165) that allow it to bind diverse targets in and around the bulge, including peptide derivatives of its cognate protein Tat(59,141,142), divalent ions(143), and five different small molecules that inhibit the

TAR-Tat interaction(144-147) (Figure 1.13). While the structures for these states have been solved, the physical principles underlying the structural transitions and the effects on the small molecule binding pocket(s) remains poorly understood.

We recently published the first ever experimentally NMR based ensemble of an RNA using the elongation strategy developed in the Al-Hashimi H. M. laboratory(74) (Stelzer A. C. was second author on the publication in the journal *Nature*). We transformed the basic NMR experiment by anchoring frames of reference onto individual RNA helices and thereby measuring their dynamics directly as motions of one helix relative to the other using multiple sets of RDCs. The frames were anchored using the method described in section 1.7.2, which involves elongating helices(281) so that they dominate overall alignment of the elongated RNA (E-RNA) in ordering media, with the elongated axis being on average orientated parallel to the magnetic field (Figure 1.14). This effectively disentangles helix motions from overall alignment and renders RDCs dependent on the angle (Ω) between bond vectors and the internal elongated axis, and not a detached external magnetic field. By anchoring the frame of reference onto different helices, the same helix motions can be measured from different helix-centered perspectives, opening a new avenue for extending the achievable spatial sensitivity with which

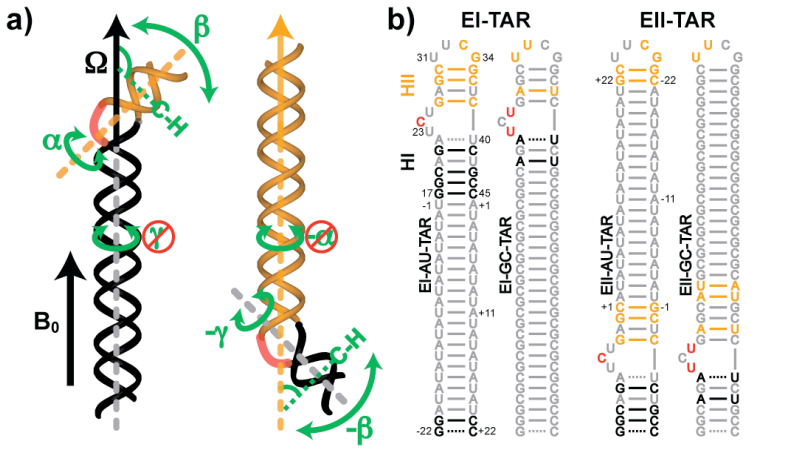


Figure 1.14. Measurement of RNA helix motions in 3D using helix-anchored frames and RDCs. (a) NMR reference frames are anchored by elongating helices so that the RNA aligns with the elongated axis being on average oriented parallel to the magnetic field (B_0). **(b)** Independent elongation of HI and HII in TAR using a strategy that renders elongation residues NMR invisible. $^{13}C/^{15}N$ labeled and unlabelled nucleotides are shown in color and gray, respectively.

motions can be characterized. For example, while RDCs probe inter-helical bending (β) and twisting (α) motions of the short helix, they are insensitive to twisting motions (γ) around the axially symmetric elongated helix (Figure 1.14a). This vanishing sensitivity can be resurrected, and helix motions thereby measured with complete 3D rotational sensitivity, by inverting which helix is elongated and measuring RDCs that probe the angles β and γ, but not α (Figure 1.14a). By measuring motions in this relativistic manner, the helix-anchored frames also allow spatial correlations between motions of two or more helices to be directly established.

Using the two sets of elongated RDCs we were able to visualize the spatially non-random inter-helical motions, by performing a search over ensembles(293) with up to three (N = 3) equally populated inter-helical conformers that can reproduce the nine independent parameters afforded by the RDCs. The conformers within the ensemble define key points along the trajectory that serve to capture its essential 3D spatial features even if the trajectory involves many more conformations. Loose steric constraints were also implemented in the conformational search. Both the search over ensembles with N=1 and 2 yielded a very poor RDC fit, confirming the existence of an inter-helix motional trajectory that is more complex than a simple two-state jump. In contrast, a very good fit was obtained for N=3 with very slight improvements obtained with N=4. Strikingly the three state ensemble, which also yields good agreement with magnetic field induced RDCs measured in non-elongated TAR(279), gives rise to a motional trajectory in which HI and HII twist by ~53° and ~110°, respectively in a highly synchronized manner (R=0.97) while simultaneously (R=0.81-0.92) bending by ~94° (Figures 1.15a-b), thus the bending motions between the two TAR helices are accompanied by a correlated twisting motions about the respective helix axes.

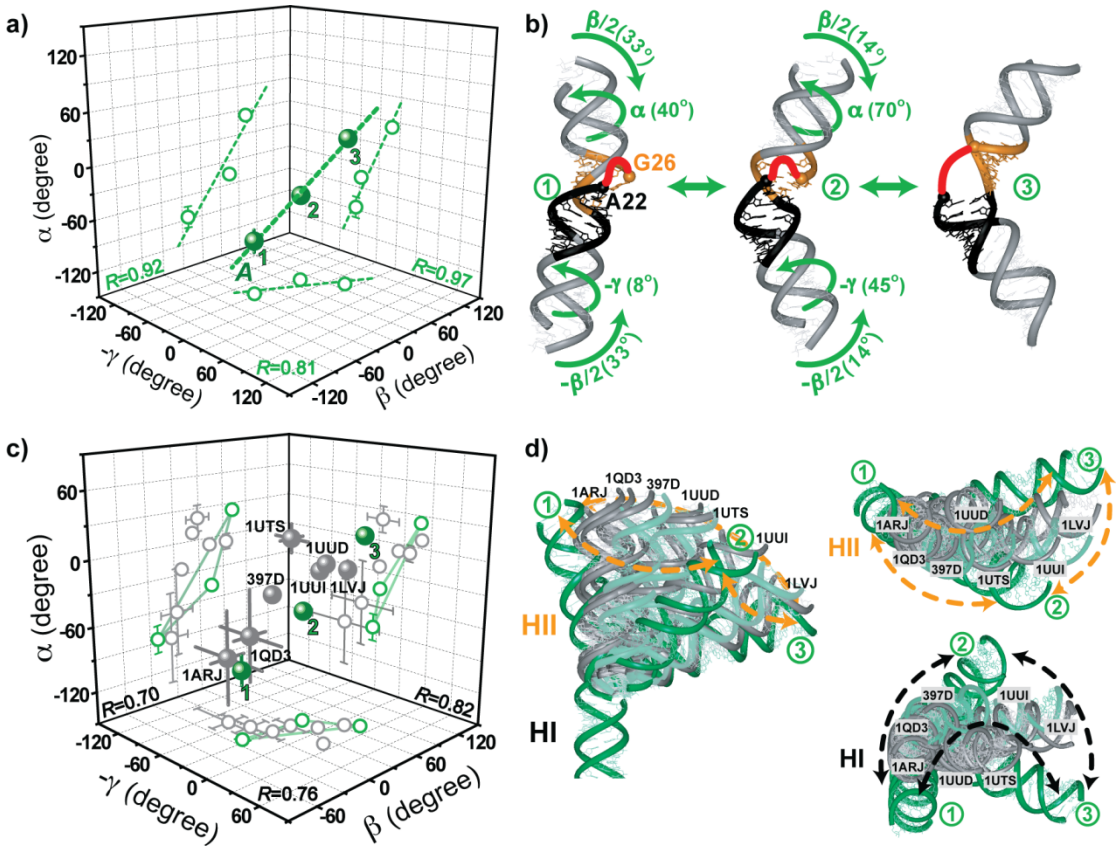


Figure 1.15. Correlated TAR dynamics guide ligand-induced transitions. (a) The best-fit TAR ensemble (in green) in which each conformer (1, 2 and 3) is specified by three Euler angles defining twisting around HI (-γ), HII (α) and inter-helical bending (β). The 3D best-fitted line is shown together with its 2D projections along each plane and the associated correlation coefficient (*R*). **(b)** The three TAR conformers in the best-fit TAR ensemble. Helices are elongated for clarity. **(c)** The three TAR dynamical conformers (in green) and the TAR conformation (in gray) bound to peptide derivatives of Tat (1ARJ)(144,194), divalent ions (397D) (145), and five different small molecules (1QD3, 1LVJ, 1UUD, 1UUI, 1UTS) (59,146,147,291). Shown on each 2D plane is the correlation coefficient (*R*) between angles for the ligand bound conformations. For bound structures, the s.d. is over the family of reported NMR models. Only one model was reported in structures 397D, 1UUI, and 1UUD, thus no s.d. is reported. **(d)** Comparison of the three TAR dynamical conformers (in green) and ligand bound TAR conformations (in gray). Sub-conformers along the linear pathway linking conformers 1→2, 2→3, and 3→1 are shown in light green and the direction of the trajectory shown using arrows. Horizontal view following superposition of HI (top right), vertical view down and up the helix axis of HI (bottom right) and HII (bottom left) following superposition of HI and HII, respectively.

Visualization of the TAR inter-helical trajectory provided us the means to interrogate the role TAR plays in directing ligand-induced structural transitions. Specifically, it allowed us to test the hypothesis that TAR samples all of the bound states

in the absence of ligand and thus recognition is governed by a conformational selection mechanism(157,158). Remarkably, we found that the seven bound TAR conformations fall along various positions of the dynamical trajectory (Figure 1.15c). The bound conformers also trace out a similar linear trajectory in the 3D inter-helix Euler space (R=0.70-0.82), confirming that correlated dynamics is an intrinsic property of the TAR structure (Figure 1.15c). The dynamic space traced out by the three TAR conformers encapsulates nearly all of the ligand bound conformations (Figure 1.15d), indicating that ligands can induce the TAR structural transitions by capturing existing conformers along various positions of the free RNA dynamical trajectory. Thus, a highly flexible RNA can be spatially tuned to undergo structured motions that direct functional transitions along specific pathways.

While the experimental NMR based three state ensemble provided strong evidence that TAR, and possibly many other RNAs, bind small molecules via a conformational selection mechanism, currently only rigid body “movies” of A-form helices can be generated using this method. One of the objectives of this Dissertation is to better understand the physical principles guiding RNA structural transitions on small molecule binding and expand the methodology to generate “movies” with the global and local conformational details necessary to improve the accuracy of computational docking and establish a high-throughput method of RNA-targeted drug design.

1.9 References

1. Watson, J.D. and Crick, F.H.C. (1953) Molecular Structure of Nucleic Acids-Structure for Deoxyribose Nucleic Acid *Nature*, 171, 737-738.
2. Guerriertakada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA Moiety of Ribonuclease-P is the Catalytic Subunit of the Enzyme. *Cell*, 35, 849-857.
3. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA-Auto-excision and Auto-cyclization of the Ribosomal-RNA Intervening Sequence of Tetrahymena. *Cell*, 31, 147-157.
4. He, L. and Hannon, G.J. (2004) Micrornas: Small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, 5, 522-531.
5. Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C. and Breaker, R.R. (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell*, 113, 577-586.
6. Winkler, W., Nahvi, A. and Breaker, R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, 419, 952-956.
7. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A. and Breaker, R.R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428, 281-286.
8. Collins, J.A., Irnov, I., Baker, S. and Winkler, W.C. (2007) Mechanism of mRNA destabilization by the glmS ribozyme. *Genes & Development*, 21, 3356-3368.
9. Cromie, M.J., Shi, Y.X., Latifi, T. and Groisman, E.A. (2006) An RNA sensor for intracellular Mg2+. *Cell*, 125, 71-84.
10. Spinelli, S.V., Pontel, L.B., Vescovi, G. and Soncini, F.C. (2008) Regulation of magnesium homeostasis in Salmonella: Mg2+ targets the mgtA transcript for degradation by RNase E. *Fems Microbiology Letters*, 280, 226-234.
11. Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. and Gelfand, M.S. (2004) Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory systems. *Nucleic Acids Research*, 32, 3340-3353.
12. Frith, M.C., Pheasant, M. and Mattick, J.S. (2005) The amazing complexity of the human transcriptome. *European Journal of Human Genetics*, 13, 894-897.
13. Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *Embo Reports*, 2, 986-991.
14. Mattick, J.S. (2004) RNA regulation: a new genetics? *Nature Reviews Genetics*, 5, 316-323.
15. Tenson, T. and Mankin, A. (2006) Antibiotics and the ribosome. *Molecular Microbiology*, 59, 1664-1677.
16. Hermann, T. (2005) Drugs targeting the ribosome. *Current Opinion in Structural Biology*, 15, 355-366.
17. Froeyen, M. and Herdewijn, P. (2002) RNA as a target for drug design, the example of Tat-TAR interaction. *Curr Top Med Chem*, 2, 1123-1145.
18. Krebs, A., Ludwig, V., Boden, O. and Gobel, M.W. (2003) Targeting the HIV trans-activation responsive region--approaches towards RNA-binding drugs. *Chembiochem*, 4, 972-978.
19. Mei, H.-Y., Galan, A.A., Halim, N.S., Mack, D.P., Moreland, D.W., Sanders, K.B., Truong, H.N. and Czarnik, A.W. (1995) Inhibition of an HIV-I Tat-derived peptide binding to TAR RNA by aminoglycoside antibiotics. *Bioorganic & Medicinal Chemistry Letters*, 5, 2755-2760.

20. Heaphy, S., Finch, J.T., Gait, M.J., Karn, J. and Singh, M. (1991) Human-Immunodeficiency-Virus Type-1 Regulator of Virion Expression, Rev, Forms Nucleoprotein Filaments after Binding to a Purine-Rich Bubble Located within the Rev-Responsive Region of Viral Messenger-Rnas. *Proceedings of the National Academy of Sciences of the United States of America*, 88, 7366-7370.
21. Laughrea, M. and Jette, L. (1994) A 19-Nucleotide Sequence Upstream of the 5' Major Splice Donor Is Part of the Dimerization Domain of Human-Immunodeficiency-Virus-1 Genomic Rna. *Biochemistry*, 33, 13464-13474.
22. Paillart, J.C., Marquet, R., Skripkin, E., Ehresmann, C. and Ehresmann, B. (1996) Dimerization of retroviral genomic RNAs: Structural and functional implications. *Biochimie*, 78, 639-653.
23. Skripkin, E., Paillart, J.C., Marquet, R., Ehresmann, B. and Ehresmann, C. (1994) Identification of the Primary Site of the Human-Immunodeficiency-Virus Type-1 Rna Dimerization in-Vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 4945-4949.
24. Cooper, T.A., Wan, L. and Dreyfuss, G. (2009) RNA and disease. *Cell*, 136, 777-793.
25. Sassen, S., Miska, E.A. and Caldas, C. (2008) MicroRNA - implications for cancer. *Virchows Archiv*, 452, 1-10.
26. Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D. and Slack, F.J. (2005) RAS is regulated by the let-7 MicroRNA family. *Cell*, 120, 635-647.
27. Takamizawa, J., Konishi, H., Yanagisawa, K., Tomida, S., Osada, H., Endoh, H., Harano, T., Yatabe, Y., Nagino, M., Nimura, Y. *et al.* (2004) Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Research*, 64, 3753-3756.
28. Meltzer, P.S. (2005) Cancer genomics - Small RNAs with big impacts. *Nature*, 435, 745-746.
29. Herbert, B.S., Gellert, G.C., Hochreiter, A., Pongracz, K., Wright, W.E., Zielinska, D., Chin, A.C., Harley, C.B., Shay, J.W. and Gryaznov, S.M. (2005) Lipid modification of GRN163, an N3 ' -> P5 ' thio-phosphoramidate oligonucleotide, enhances the potency of telomerase inhibition. *Oncogene*, 24, 5262-5268.
30. Pinto, I.G., Guilbert, C., Ulyanov, N.B., Stearns, J. and James, T.L. (2008) Discovery of Ligands for a Novel Target, the Human Telomerase RNA, Based on Flexible-Target Virtual Screening and NMR. *Journal of Medicinal Chemistry*, 51, 7205-7215.
31. Drews, J. (1996) Genomic sciences and the medicine of tomorrow. *Nature Biotechnology*, 14, 1516-&.
32. Drews, J. and Ryser, S. (1997) The role of innovation in drug development. *Nature Biotechnology*, 15, 1318-1319.
33. Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nature Reviews Drug Discovery*, 1, 727-730.
34. Imming, P., Sinning, C. and Meyer, A. (2006) Opinion - Drugs, their targets and the nature and number of drug targets. *Nature Reviews Drug Discovery*, 5, 821-834.
35. Overington, J.P., Al-Lazikani, B. and Hopkins, A.L. (2006) Opinion - How many drug targets are there? *Nature Reviews Drug Discovery*, 5, 993-996.
36. Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34, D668-D672.

37. Zheng, C.J., Han, L.Y., Yap, C.W., Ji, Z.L., Cao, Z.W. and Chen, Y.Z. (2006) Therapeutic targets: Progress of their exploration and investigation of their characteristics. *Pharmacological Reviews*, 58, 259-279.
38. Zheng, C.J., Han, L.Y., Yap, C.W., Xie, B. and Chen, Y.Z. (2006) Progress and problems in the exploration of therapeutic targets. *Drug Discovery Today*, 11, 412-420.
39. Fletcher, S. and Hamilton, A.D. (2005) Protein surface recognition and proteomimetics: mimics of protein surface structure and function. *Current Opinion in Chemical Biology*, 9, 632-638.
40. Finkelstein, T. and Weinstein, I.B. (1967) Proflavine Binding to Transfer Ribonucleic Acid Synthesic Ribonucleic Acids and Deoxyribonucleic Acids *Journal of Biological Chemistry*, 242, 3763-&.
41. Weinstein, I.B. and Finkelstein, I.B. (1967) Proflavine Inhibition of Protein Synthesis *Journal of Biological Chemistry*, 242, 3757-&.
42. Fourmy, D., Recht, M.I., Blanchard, S.C. and Puglisi, J.D. (1996) Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science*, 274, 1367-1371.
43. Tor, Y. (2006) The ribosomal A-site as an inspiration for the design of RNA binders. *Biochimie*, 88, 1045-1051.
44. Vicens, Q. and Westhof, E. (2003) RNA as a drug target: The case of aminoglycosides. *Chembiochem*, 4, 1018-1023.
45. Brisson-Noel, A., Trieu-Cout, P. and Courvalin, P.J. (1988) Macrolides. *Journal of Antimicrobial Chemotherapy*, 22.
46. Brodersen, D.E., Clemons, W.M., Carter, A.P., Morgan-Warren, R.J., Wimberly, B.T. and Ramakrishnan, V. (2000) The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell*, 103, 1143-1154.
47. Jones, D., Metzger, H.J., Schatz, A. and Waksman, S.A. (1944) Control of gram-negative bacteria in experimental animals by streptomycin. *Science*, 100, 103-105.
48. Schatz, A., Bugie, E. and Waksman, S.A. (1944) Streptomycin, a substance exhibiting antibiotic activity against gram positive and gram-negative bacteria. *Proceedings of the Society for Experimental Biology and Medicine*, 55, 66-69.
49. Watson, J.D. and Crick, F.H.C. (1953) Genetical Implications of the Strucrture of Deoxyribonucleic Acid. *Nature*, 171, 964-967.
50. Vincent, W.S. (1957) Heterogeneity of Nuclear Ribonucleic Acid. *Science*, 126, 306-307.
51. Nishimur, S., Jacob, T.M. and Khorana, H.G. (1964) Synthetic Deoxyribopolynucleotides as Templates for Ribonucleic Acid Polymerase-Formation and Characterization of a Ribopolynucleotide with a Repeating Trinucleotide Sequence *Proceedings of the National Academy of Sciences of the United States of America*, 52, 1494-&.
52. Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquise.M, Merrill, S.H., Penswick, J.R. and Zamir, A. (1965) Structure of a Ribonucleic Acid. *Science*, 147, 1462-&.
53. Rosenberger, J.M., Seeman, N.C., Japarkki.J, Suddath, F.L., Nicholas, H.B. and Rich, A. (1973) Double Helix at Atomic Resolution *Nature*, 243, 150-154.
54. Sussman, J.L. and Kim, S.H. (1976) 3-Dimensional Structure of a Transfer-RNA in 2 Crystal Forms. *Science*, 192, 853-858.

55. Rould, M.A., Perona, J.J., Soll, D. and Steitz, T.A. (1989) Structure of Escherichia-Coli Glutaminyl-Transfer Rna-Synthetase Complexed with Transfer Rnagln and Atp at 2.8-a Resolution. *Science*, 246, 1135-1142.
56. Ellington, A.D. and Szostak, J.W. (1990) Invitro Selection of Rna Molecules That Bind Specific Ligands. *Nature*, 346, 818-822.
57. Tuerk, C. and Gold, L. (1990) Systematic Evolution of Ligands by Exponential Enrichment - Rna Ligands to Bacteriophage-T4 DNA-Polymerase. *Science*, 249, 505-510.
58. Vonahsen, U., Davies, J. and Schroeder, R. (1991) Antibiotic Inhibition of Group-I Ribozyme Function. *Nature*, 353, 368-370.
59. Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D. and Williamson, J.R. (1992) Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science*, 257, 76-80.
60. Weeks, K.M. and Crothers, D.M. (1993) Major Groove Accessibility of RNA *Science*, 261, 1574-1577.
61. Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional Regulation of the Heterochronic Gene LIN-14 Mediates Temporal Pattern-Formation in C-Elegans *Cell*, 75, 855-862.
62. Brickner, S.J., Hutchinson, D.K., Barbachyn, M.R., Manninen, P.R., Ulanowicz, D.A., Garmon, S.A., Grega, K.C., Hendges, S.K., Toops, D.S., Ford, C.W. *et al.* (1996) Synthesis and antibacterial activity of U-100592 and U-100766, two oxazolidinone antibacterial agents for the potential treatment of multidrug-resistant Gram-positive bacterial infections. *Journal of Medicinal Chemistry*, 39, 673-679.
63. Fan, P., Suri, A.K., Fiala, R., Live, D. and Patel, D.J. (1996) Molecular recognition in the FMN-RNA aptamer complex. *Journal of Molecular Biology*, 258, 480-500.
64. Fire, A., Xu, S.Q., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391, 806-811.
65. Griffey, R.H., Hofstadler, S.A., Sannes-Lowery, K.A., Ecker, D.J. and Crooke, S.T. (1999) Determinants of aminoglycoside-binding specificity for rRNA by using mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 10129-10133.
66. Hofstadler, S.A., Sannes-Lowery, K.A., Crooke, S.T., Ecker, D.J., Sasmor, H., Manalili, S. and Griffey, R.H. (1999) Multiplexed screening of neutral mass-tagged RNA targets against ligand libraries with electrospray ionization FTICR MS: A paradigm for high-throughout affinity screening. *Analytical Chemistry*, 71, 3436-3440.
67. Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution. *Science*, 289, 905-920.
68. Carter, A.P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Wimberly, B.T. and Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, 407, 340-348.
69. Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. *et al.* (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstrom resolution. *Cell*, 102, 615-623.

70. Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T. and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature*, 407, 327-339.
71. Ogle, J.M., Brodersen, D.E., Clemons, W.M., Tarry, M.J., Carter, A.P. and Ramakrishnan, V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science*, 292, 897-902.
72. Lind, K.E., Du, Z.H., Fujinaga, K., Peterlin, B.M. and James, T.L. (2002) Structure-based computational database screening, in vitro assay, and NMR assessment of compounds that target TAR RNA. *Chemistry & Biology*, 9, 185-193.
73. Blount, K.F., Zhao, F., Hermann, T. and Tor, Y. (2005) Conformational constraint as a means for understanding RNA-aminoglycoside specificity. *Journal of the American Chemical Society*, 127, 9818-9829.
74. Zhang, Q., Stelzer, A.C., Fisher, C.K. and Al-Hashimi, H.M. (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature*, 450, 1263-U1214.
75. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Swanstrom, R., Burch, C.L. and Weeks, K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, 460, 711-U787.
76. Bozdogan, B. and Appelbaum, P.C. (2004) Oxazolidinones: activity, mode of action, and mechanism of resistance. *International Journal of Antimicrobial Agents*, 23, 113-119.
77. Blount, K.F., Wang, J.X., Lim, J., Sudarsan, N. and Breaker, R.R. (2007) Antibacterial lysine analogs that target lysine riboswitches. *Nature Chemical Biology*, 3, 44-49.
78. Mikkelsen, N.E., Brannvall, M., Virtanen, A. and Kirsebom, L.A. (1999) Inhibition of RNase P RNA cleavage by aminoglycosides. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 6155-6160.
79. Richter, S.N. and Palu, G. (2006) Inhibitors of HIV-1 Tat-mediated transactivation. *Current Medicinal Chemistry*, 13, 1305-1315.
80. Zapp, M.L., Stern, S. and Green, M.R. (1993) Small Molecules That Selectively Block Rna-Binding of Hiv-1 Rev Protein Inhibit Rev Function and Viral Production. *Cell*, 74, 969-978.
81. Vicens, Q. (2009) RNA's coming of age as a drug target. *Journal of Inclusion Phenomena and Macrocyclic Chemistry*, 65, 171-188.
82. Spahn, C.M.T., Kieft, J.S., Grassucci, R.A., Penczek, P.A., Zhou, K.H., Doudna, J.A. and Frank, J. (2001) Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science*, 291, 1959-1962.
83. Kieft, J.S., Zhou, K.H., Jubin, R. and Doudna, J.A. (2001) Mechanism of ribosome recruitment by hepatitis CIRES RNA. *Rna-a Publication of the Rna Society*, 7, 194-206.
84. Barbachyn, M.R., Hutchinson, D.K., Brickner, S.J., Cynamon, M.H., Kilburn, J.O., Klemens, S.P., Glickman, S.E., Grega, K.C., Hendges, S.K., Toops, D.S. *et al.* (1996) Identification of a novel oxazolidinone (U-100480) with potent antimycobacterial activity. *Journal of Medicinal Chemistry*, 39, 680-685.
85. Bobkova, E.V., Yan, Y.P., Jordan, D.B., Kurilla, M.G. and Pompliano, D.L. (2003) Catalytic properties of mutant 23 S ribosomes resistant to oxazolidinones. *Journal of Biological Chemistry*, 278, 9802-9807.
86. Clemons, W.M., Brodersen, D.E., McCutcheon, J.P., May, J.L.C., Carter, A.P., Morgan-Warren, R.J., Wimberly, B.T. and Ramakrishnan, V. (2001) Crystal structure of the 30 S ribosomal subunit from Thermus thermophilus: Purification,

crystallization and structure determination. *Journal of Molecular Biology*, 310, 827-843.
87. Ogle, J.M., Murphy, F.V., Tarry, M.J. and Ramakrishnan, V. (2002) Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, 111, 721-732.
88. Vila-Sanjurjo, A., Ridgeway, W.K., Seymaner, V., Zhang, W., Santoso, S., Yu, K. and Cate, J.H.D. (2003) X-ray crystal structures of the WT and a hyper-accurate ribosome from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 8682-8687.
89. Murray, J.B., Meroueh, S.O., Russell, R.J.M., Lentzen, G., Haddad, J. and Mobashery, S. (2006) Interactions of designer antibiotics and the bacterial ribosomal aminoacyl-tRNA site. *Chemistry & Biology*, 13, 129-138.
90. Selmer, M., Dunham, C.M., Murphy, F.V., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, 313, 1935-1942.
91. Borovinskaya, M.A., Pai, R.D., Zhang, W., Schuwirth, B.S., Holton, J.M., Hirokawa, G., Kaji, H., Kaji, A. and Cate, J.H.D. (2007) Structural basis for aminoglycoside inhibition of bacterial ribosome recycling. *Nature Structural & Molecular Biology*, 14, 727-732.
92. Borovinskaya, M.A., Shoji, S., Holton, J.M., Fredric, K. and Cate, J.H.D. (2007) A steric block in translation caused by the antibiotic spectinomycin. *Acs Chemical Biology*, 2, 545-552.
93. Weixlbaumer, A., Murphy, F.V., Dziergowska, A., Malkiewicz, A., Vendeix, F.A.P., Agris, P.F. and Ramakrishnan, V. (2007) Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nature Structural & Molecular Biology*, 14, 498-502.
94. Borovinskaya, M.A., Shoji, S., Fredrick, K. and Cate, J.H.D. (2008) Structural basis for hygromycin B inhibition of protein biosynthesis. *Rna-a Publication of the Rna Society*, 14, 1590-1599.
95. Voorhees, R.M., Weixlbaumer, A., Loakes, D., Kelley, A.C. and Ramakrishnan, V. (2009) Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nature Structural & Molecular Biology*, 16, 528-533.
96. Fourmy, D., Recht, M.I. and Puglisi, J.D. (1998) Binding of neomycin-class aminoglycoside antibiotics to the A-site of 16 S rRNA. *Journal of Molecular Biology*, 277, 347-362.
97. Lynch, S.R., Gonzalez, R.L. and Puglisi, J.D. (2003) Comparison of x-ray crystal structure of the 30S subunit-antibiotic complex with NMR structure of decoding site oligonucleotide-paromomycin complex. *Structure*, 11, 43-53.
98. Lynch, S.R. and Puglisi, J.D. (2001) Structural origins of aminoglycoside specificity for prokaryotic ribosomes. *Journal of Molecular Biology*, 306, 1037-1058.
99. Jarvest, R.L., Berge, J.M., Berry, V., Boyd, H.F., Brown, M.J., Elder, J.S., Forrest, A.K., Fosberry, A.P., Gentry, D.R., Hibbs, M.J. *et al.* (2002) Nanomolar inhibitors of Staphylococcus aureus methionyl tRNA synthetase with potent antibacterial activity against gram-positive pathogens. *Journal of Medicinal Chemistry*, 45, 1959-1962.
100. Kirillov, S., Vitali, L.A., Goldstein, B.P., Monti, F., Semenkov, Y., Makhno, V., Ripa, S., Pon, C.L. and Gualerzi, C.O. (1997) Purpuromycin: An antibiotic inhibiting tRNA aminoacylation. *Rna-a Publication of the Rna Society*, 3, 905-913.

101. Kirk, S.R. and Tor, Y. (1999) tRNA(Phe) binds aminoglycoside antibiotics. *Bioorganic & Medicinal Chemistry*, 7, 1979-1991.
102. Mikkelsen, N.E., Johansson, K., Virtanen, A. and Kirsebom, L.A. (2001) Aminoglycoside binding displaces a divalent metal ion in a tRNA-neomycin B complex. *Nature Structural Biology*, 8, 510-514.
103. Walter, F., Putz, J., Giege, R. and Westhof, E. (2002) Bindin of tobramycin leads to conformational changes in yeast tRNA(Asp) and inhibition of aminoacylation. *Embo Journal*, 21, 760-768.
104. Stage, T.K., Hertel, K.J. and Uhlenbeck, O.C. (1995) Inhibition of the Hammerhead Ribozyme by Neomycin. *Rna-a Publication of the Rna Society*, 1, 95-101.
105. Champney, W.S. (2001) Bacterial Ribosomal Subunit Synthesis A Novel Antibiotic Target. - 1, - 36.
106. Champney, W.S. (2003) Bacterial ribosomal subunit assembly is an antibiotic target. *Current Topics in Medicinal Chemistry*, 3, 929-947.
107. Champney, W.S. (2006) The Other Target for Ribosomal Antibiotics: Inhibition of Bacterial Ribosomal Subunit Formation. - 6, - 390.
108. Champney, W.S. and Burdine, R. (1998) Macrolide antibiotic inhibition of translation and 50S ribosomal subunit assembly in methicillin-resistant Staphylococcus aureus cells. *Microbial Drug Resistance-Mechanisms Epidemiology and Disease*, 4, 169-174.
109. Silvers, J.A. and Champney, W.S. (2005) Accumulation and turnover of 23S ribosomal RNA in azithromycin-inhibited ribonuclease mutant strains of Escherichia coli. *Archives of Microbiology*, 184, 66-77.
110. Usary, J. and Champney, W.S. (2001) Erythromycin inhibition of 50S ribosomal subunit formation in Escherichia coli cells. *Molecular Microbiology*, 40, 951-962.
111. Thomas, J.R. and Hergenrother, P.J. (2008) Targeting RNA with small molecules. *Chemical Reviews*, 108, 1171-1224.
112. Roth, A. and Breaker, R.R. (2009) The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annual Review of Biochemistry*, 78, 305-334.
113. Gilbert, S.D., Mediatore, S.J. and Batey, R.T. (2006) Modified pyrimidines specifically bind the purine riboswitch. *Journal of the American Chemical Society*, 128, 14214-14215.
114. Wickiser, J.K., Cheah, M.T., Breaker, R.R. and Crothers, D.M. (2005) The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry*, 44, 13404-13414.
115. Buck, J., Furtig, B., Noeske, J., Wohnert, J. and Schwalbe, H. (2007) Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 15699-15704.
116. Rieder, R., Lang, K., Graber, D. and Micura, R. (2007) Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *Chembiochem*, 8, 896-902.
117. Geis, M., Flamm, C., Wolfinger, M.T., Tanzer, A., Hofacker, I.L., Middendorf, M., Mandl, C., Stadler, P.F. and Thurner, C. (2008) Folding kinetics of large RNAs. *Journal of Molecular Biology*, 379, 160-173.
118. Ghigna, C., Valacca, C. and Biamonti, G. (2008) Alternative Splicing and Tumor Progression. *Current Genomics*, 9, 556-570.
119. Sucheck, S.J., Greenberg, W.A., Tolbert, T.J. and Wong, C.H. (2000) Design of small molecules that recognize RNA: Development of aminoglycosides as

potential antitumor agents that target oncogenic RNA sequences. *Angewandte Chemie-International Edition*, 39, 1080-+.
120. Strunk, B.S. and Karbstein, K. (2009) Powering through ribosome assembly. *Rna-a Publication of the Rna Society*, 15, 2083-2104.
121. Chu, E., Koeller, D.M., Johnston, P.G., Zinn, S. and Allegra, C.J. (1993) Regulation of Thymidylate Synthase in Human Colon Cancer-Cells Treated with 5-Fluorouracil and Interferon-Gamma. *Molecular Pharmacology*, 43, 527-533.
122. Vanderwilt, C.L., Pinedo, H.M., Smid, K. and Peters, G.J. (1992) Elevation of Thymidylate Synthase Following 5-Fluorouracil Treatment Is Prevented by the Addition of Leucovorin in Murine Colon Tumors. *Cancer Research*, 52, 4922-4928.
123. Keyomarsi, K., Samet, J., Molnar, G. and Pardee, A.B. (1993) The Thymidylate Synthase Inhibitor, Ici-D1694, Overcomes Translational Detainment of the Enzyme. *Journal of Biological Chemistry*, 268, 15142-15149.
124. Tok, J.B.H., Cho, J.H. and Rando, R.R. (1999) Aminoglycoside antibiotics are able to specifically bind the 5 '-untranslated region of thymidylate synthase messenger RNA. *Biochemistry*, 38, 199-206.
125. Chu, E., Callender, M.A., Farrell, M.P. and Schmitz, J.C. (2003) Thymidylate synthase inhibitors as anticancer agents: from bench to bedside. *Cancer Chemotherapy and Pharmacology*, 52, S80-S89.
126. Gmeiner, W.H. (2005) Novel chemical strategies for thymidylate synthase inhibition. *Current Medicinal Chemistry*, 12, 191-202.
127. Chu, E., Voeller, D., Koeller, D.M., Drake, J.C., Takimoto, C.H., Maley, G.F., Maley, F. and Allegra, C.J. (1993) Identification of an Rna-Binding Site for Human Thymidylate Synthase. *Proceedings of the National Academy of Sciences of the United States of America*, 90, 517-521.
128. Chu, E., Koeller, D.M., Casey, J.L., Drake, J.C., Chabner, B.A., Elwood, P.C., Zinn, S. and Allegra, C.J. (1991) Autoregulation of Human Thymidylate Synthase Messenger-Rna Translation by Thymidylate Synthase. *Proceedings of the National Academy of Sciences of the United States of America*, 88, 8977-8981.
129. Richter, S., Ping, Y.H. and Rana, T.M. (2002) TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. *Proc Natl Acad Sci U S A*, 99, 7928-7933.
130. Frankel, A.D. and Young, J.A.T. (1998) HIV-1: Fifteen proteins and an RNA. *Annual Review of Biochemistry*, 67, 1-25.
131. Roy, S., Delling, U., Chen, C.H., Rosen, C.A. and Sonenberg, N. (1990) A Bulge Structure in HIV-1 TAR RNA is Required for Tat Binding and Tat-mediated Transactivation *Genes & Development*, 4, 1365-1373.
132. Delling, U., Roy, S., Sumnersmith, M., Barnett, R., Reid, L., Rosen, C.A. and Sonenberg, N. (1991) The Number of Positively Charged Amino-acids in the Basic Domain of Tat is Critical for Transactivation and Complex-formation with TAR RNA *Proceedings of the National Academy of Sciences of the United States of America*, 88, 6234-6238.
133. Weeks, K.M. and Crothers, D.M. (1991) RNA Recognition by Tat-derived Peptides- Interaction in the Major Groove *Cell*, 66, 577-588.
134. Weeks, K.M. and Crothers, D.M. (1992) RNA-binding Assays for Tat-derived Peptides-Implications for Specificity. *Biochemistry*, 31, 10281-10287.
135. Churcher, M.J., Lamont, C., Hamy, F., Dingwall, C., Green, S.M., Lowe, A.D., Butler, P.J.G., Gait, M.J. and Karn, J. (1993) Higi-affinity Binding of TAR RNA by the Human-Immunodeficiency-Virus Type-1 Tat Protein Requires Base-pairs in

the RNA Stem and Amino-acid-residues Flanking the Basic Region *Journal of Molecular Biology*, 230, 90-110.
136. Long, K.S. and Crothers, D.M. (1999) Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. *Biochemistry*, 38, 10059-10069.
137. Nifosi, R., Reyes, C.M. and Kollman, P.A. (2000) Molecular Dynamics Studies of the HIV-1 TAR and its Complex with Argininamide. *Nucleic Acids Research*, 4944-4955.
138. Perez-Canadillas, J.M. and Varani, G. (2001) Recent advances in RNA-protein recognition. *Current Opinion in Structural Biology*, 11, 53-58.
139. Williamson, J.R. (2000) Induced fit in RNA-protein recognition. *Nature Structural Biology*, 7, 834-837.
140. Zacharias, M. and Hagerman, P.J. (1995) The Bend in RNA Created By the Transactivation Response Element Bulge of Human-Immunodeficiency-Virus Is Straightened By Arginine and By Tat-Derived Peptide. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 6052-6056.
141. Aboul-ela, F., Karn, J. and Varani, G. (1995) The Structure of the Human-Immunodeficiency-Virus Type-1 Tar RNA Reveals Principles of RNA Recognition By Tat Protein. *Journal of Molecular Biology*, 253, 313-332.
142. Aboul-ela, F., Karn, J. and Varani, G. (1996) Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Research*, 24, 3974-3981.
143. Ippolito, J.A. and Steitz, T.A. (1998) A 1.3-angstrom resolution crystal structure of the HIV-1 trans- activation response region RNA stem reveals a metal ion-dependent bulge conformation. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 9819-9824.
144. Davis, B., Afshar, M., Varani, G., Murchie, A.I., Karn, J., Lentzen, G., Drysdale, M., Bower, J., Potter, A.J., Starkey, I.D. *et al.* (2004) Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *J Mol Biol*, 336, 343-356.
145. Du, Z.H., Lind, K.E. and James, T.L. (2002) Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. *Chemistry & Biology*, 9, 707-712.
146. Faber, C., Sticht, H., Schweimer, K. and Rosch, P. (2000) Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J Biol Chem*, 275, 20660-20666.
147. Murchie, A.I., Davis, B., Isel, C., Afshar, M., Drysdale, M.J., Bower, J., Potter, A.J., Starkey, I.D., Swarbrick, T.M., Mirza, S. *et al.* (2004) Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. *J Mol Biol*, 336, 625-638.
148. Mei, H.Y., Galan, A.A., Halim, N.S., Mack, D.P., Moreland, D.W., Sanders, K.B., Truong, H.N. and Czarnik, A.W. (1995) Inhibition of an Hiv-1 Tat-Derived Peptide Binding to Tar Rna by Aminoglycoside Antibiotics. *Bioorganic & Medicinal Chemistry Letters*, 5, 2755-2760.
149. Yoshizawa, S., Fourmy, D. and Puglisi, J.D. (1998) Structural origins of gentamicin antibiotic action. *Embo Journal*, 17, 6437-6448.
150. Vicens, Q. and Westhof, E. (2001) Crystal structure of paromomycin docked into the eubacterial ribosomal decoding a site. *Structure*, 9, 647-658.
151. Hermann, T. and Patel, D.J. (2000) Biochemistry - Adaptive recognition by nucleic acid aptamers. *Science*, 287, 820-825.

152. Mandal, M. and Breaker, R.R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Structural & Molecular Biology*, 11, 29-35.
153. Serganov, A., Yuan, Y.R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A.T., Hobartner, C., Micura, R., Breaker, R.R. and Patel, D.J. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chemistry & Biology*, 11, 1729-1741.
154. Baugh, C., Grate, D. and Wilson, C. (2000) 2.8 angstrom crystal structure of the malachite green aptamer. *Journal of Molecular Biology*, 301, 117-128.
155. Hermann, T. (2002) Rational ligand design for RNA: the role of static structure and conformational flexibility in target recognition. *Biochimie*, 84, 869-875.
156. Kaul, M., Barbieri, C.M. and Pilch, D.S. (2006) Aminoglycoside-induced reduction in nucleotide mobility at the ribosomal RNA A-site as a potentially key determinant of antibacterial activity. *Journal of the American Chemical Society*, 128, 1261-1271.
157. Al-Hashimi, H.M. (2005) Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *Chembiochem*, 6, 1506-1519.
158. Leulliot, N. and Varani, G. (2001) Current topics in RNA-protein recognition: Control of specificity and biological function through induced fit and conformational capture. *Biochemistry*, 40, 7947-7956.
159. Frankel, A.D. and Smith, C.A. (1998) Induced folding in RNA-protein recognition: More than a simple molecular handshake. *Cell*, 92, 149-151.
160. Koshland, D.E. (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 44, 98-104.
161. Foote, J. and Milstein, C. (1994) Conformational Isomerism and the Diversity of Antibodies. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 10370-10374.
162. Eguchi, Y. and Tomizawa, J. (1990) Complex Formed by Complementary Rna Stem-Loops and Its Stabilization by a Protein - Function of Cole1 Rom Protein. *Cell*, 60, 199-209.
163. Weeks, K.M. and Cech, T.R. (1995) Protein Facilitation of Group-I Intron Splicing by Assembly of the Catalytic Core and the 5'-Splice-Site Domain. *Cell*, 82, 221-230.
164. Weeks, K.M. and Cech, T.R. (1996) Assembly of a ribonucleoprotein catalyst by tertiary structure capture. *Science*, 271, 345-348.
165. Al-Hashimi, H.M., Gosser, Y., Gorin, A., Hu, W.D., Majumdar, A. and Patel, D.J. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *Journal of Molecular Biology*, 315, 95-102.
166. Vicens, Q. and Westhof, E. (2003) Crystal structure of geneticin bound to a bacterial 16 S ribosomal RNA A site oligonucleotide. *Journal of Molecular Biology*, 326, 1175-1188.
167. Penchovsky, R. and Breaker, R.R. (2005) Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes. *Nature Biotechnology*, 23, 1424-1433.
168. Tucker, B.J. and Breaker, R.R. (2005) Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology*, 15, 342-348.
169. Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R. and Patel, D.J. (2006) Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature*, 441, 1167-1171.

170. Arya, D.P., Xue, L. and Willis, B. (2003) Aminoglycoside (neomycin) preference is for A-form nucleic acids, not just RNA: Results from a competition dialysis study. *Journal of the American Chemical Society*, 125, 10148-10149.
171. Chen, Q., Shafer, R.H. and Kuntz, I.D. (1997) Structure-based discovery of ligands targeted to the RNA double helix. *Biochemistry*, 36, 11402-11407.
172. Jin, E., Katritch, V., Olson, W.K., Kharatisvili, M., Abagyan, R. and Pilch, D.S. (2000) Aminoglycoside binding in the major groove of duplex RNA: The thermodynamic and electrostatic forces that govern recognition. *Journal of Molecular Biology*, 298, 95-110.
173. Sevignani, C., Calin, G.A., Siracusa, L.D. and Croce, C.M. (2006) Mammalian microRNAs: a small world for fine-tuning gene expression. *Mammalian Genome*, 17, 189-202.
174. Wagner, E.G.H., Altuvia, S. and Romby, P. (2002), *Homology Effects*, Vol. 46, pp. 361-398.
175. Dejong, E.S., Luy, B. and Marino, J.P. (2002) RNA and RNA-Protein Complexes as Targets for Therapeutic Intervention. - 2, - 302.
176. Winkler, W.C. and Breaker, R.R. (2003) Genetic control by metabolite-binding riboswitches. *Chembiochem*, 4, 1024-1032.
177. Chin, K., Sharp, K.A., Honig, B. and Pyle, A.M. (1999) Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nature Structural Biology*, 6, 1055-1061.
178. Batey, R.T., Gilbert, S.D. and Montange, R.K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, 432, 411-415.
179. Gilbert, S.D., Stoddard, C.D., Wise, S.J. and Batey, R.T. (2006) Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *Journal of Molecular Biology*, 359, 754-768.
180. Bottger, E.C., Springer, B., Prammananan, T., Kidan, Y. and Sander, P. (2001) Structural basis for selectivity and toxicity of ribosomal antibiotics. *Embo Reports*, 2, 318-323.
181. Prammananan, T., Sander, P., Brown, B.A., Frischkorn, K., Onyi, G.O., Zhang, Y.S., Bottger, E.C. and Wallace, R.J. (1998) Single 16S ribosomal RNA substitution is responsible for resistance to amikacin and other 2-deoxystreptamine aminoglycosides in Mycobacterium abscessus and Mycobacterium chelonae. *Journal of Infectious Diseases*, 177, 1573-1581.
182. Taniguchi, H., Chang, B., Abe, C., Nikaido, Y., Mizuguchi, Y. and Yoshida, S. (1997) Molecular analysis of kanamycin and viomycin resistance in Mycobacterium smegmatis by use of the conjugation system. *Journal of Bacteriology*, 179, 4795-4801.
183. Hoch, I., Berens, C., Westhof, E. and 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, 557-569.
184. Zhang, Y., Li, Z.J., Pilch, D.S. and Leibowitz, M.J. (2002) Pentamidine inhibits catalytic activity of group I intron Ca.LSU by altering RNA folding. *Nucleic Acids Research*, 30, 2961-2971.
185. Kaul, M., Barbieri, C.M., Kerrigan, J.E. and Pilch, D.S. (2003) Coupling of drug protonation to the specific binding of aminoglycosides to the A site of 16 S rRNA: Elucidation of the number of drug amino groups involved and their identities. *Journal of Molecular Biology*, 326, 1373-1387.
186. Day, R.O., Seeman, N.C., Rosenber.Jm and Rich, A. (1973) Crystalline Fragment of Double Helix-Structure of Dinucleoside Phosphate Guanylyl-3'. 5'-

cytidine *Proceedings of the National Academy of Sciences of the United States of America*, 70, 849-853.
187. Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E. (1981) Structure of a B-DNA Dodecamer-Conformation and Dynamics .1 *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 78, 2179-2183.
188. Rich, A. and Davies, D.R. (1956) A New 2 Stranded Helical Structure-Polyadenylic Acid and Polyuridylic Acid *Journal of the American Chemical Society*, 78, 3548-3549.
189. Wilkins, M.H.F., Stokes, A.R. and Wilson, H.R. (1953) Molecular Structure of Deoxypentose Nucleic Acids *Nature*, 171, 738-740.
190. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E. (1980) Crystal-structure Analysis of a Complete Turn of B-DNA *Nature*, 287, 755-758.
191. Cochrane, J.C., Lipchock, S.V. and Strobel, S.A. (2007) Structural investigation of the GlmS ribozyme bound to its catalytic cofactor. *Chemistry & Biology*, 14, 97-105.
192. Klein, D.J. and Ferre-D'Amare, A.R. (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*, 313, 1752-1756.
193. Klein, D.J., Wilkinson, S.R., Been, M.D. and Ferre-D'Amare, A.R. (2007) Requirement of helix p2.2 and nucleotide g1 for positioning the cleavage site and cofactor of the glmS ribozyme. *Journal of Molecular Biology*, 373, 178-189.
194. Aboulela, F., Karn, J. and Varani, G. (1995) The Structure of the Human-Immunodeficiency-Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein *Journal of Molecular Biology*, 253, 313-332.
195. Brodsky, A.S. and Williamson, J.R. (1997) Solution structure of the HIV-2 TAR-argininamide complex. *Journal of Molecular Biology*, 267, 624-639.
196. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23, 3-25.
197. Hermann, T. (2003) Chemical and functional diversity of small molecule ligands for RNA. *Biopolymers*, 70, 4-18.
198. Tor, Y. (2003) Targeting RNA with small molecules. *Chembiochem*, 4, 998-1007.
199. Hwang, S., Tamilarasu, N., Kibler, K., Cao, H., Ali, A., Ping, Y.H., Jeang, K.T. and Rana, T.M. (2003) Discovery of a small molecule Tat-trans-activation-responsive RNA antagonist that potently inhibits human immunodeficiency virus-1 replication. *Journal of Biological Chemistry*, 278, 39092-39103.
200. Mei, H.Y., Mack, D.P., Galan, A.A., Halim, N.S., Heldsinger, A., Loo, J.A., Moreland, D.W., SannesLowery, K.A., Sharmeen, L., Truong, H.N. *et al.* (1997) Discovery of selective, small-molecule inhibitors of RNA complexes .1. The Tat protein TAR RNA complexes required for HIV-1 transcription. *Bioorganic & Medicinal Chemistry*, 5, 1173-1184.
201. Disney, M.D. and Seeberger, P.H. (2004) Aminoglycoside microarrays to explore interactions of antibiotics with RNAs and proteins. *Chemistry-a European Journal*, 10, 3308-3314.
202. Bradrick, T.D. and Marino, J.P. (2004) Ligand-induced changes in 2-aminopurine fluorescence as a probe for small molecule binding to HIV-1 TAR RNA. *Rna-a Publication of the Rna Society*, 10, 1459-1468.

203. Kirk, S.R., Luedtke, N.W. and Tor, Y. (2001) 2-aminopurine as a real-time probe of enzymatic cleavage and inhibition of hammerhead ribozymes. *Bioorganic & Medicinal Chemistry*, 9, 2295-2301.
204. Lacourciere, K.A., Stivers, J.T. and Marino, J.P. (2000) Mechanism of neomycin and Rev peptide binding to the Rev responsive element of HIV-1 as determined by fluorescence and NMR spectroscopy. *Biochemistry*, 39, 5630-5641.
205. Shandrick, S., Zhao, Q., Han, Q., Ayida, B.K., Takahashi, M., Winters, G.C., Simonsen, K.B., Vourloumis, D. and Hermann, T. (2004) Monitoring molecular recognition of the ribosomal decoding site. *Angewandte Chemie-International Edition*, 43, 3177-3182.
206. Yan, Z.H. and Baranger, A.M. (2004) Binding of an aminoacridine derivative to a GAAA RNA tetraloop. *Bioorganic & Medicinal Chemistry Letters*, 14, 5889-5893.
207. DeNap, J.C.B., Thomas, J.R., Musk, D.J. and Hergenrother, P.J. (2004) Combating drug-resistant bacteria: Small molecule mimics of plasmid incompatibility as antiplasmid compounds. *Journal of the American Chemical Society*, 126, 15402-15404.
208. Thomas, J.R., DeNap, J.C.B., Wong, M.L. and Hergenrother, P.J. (2005) The relationship between arninoglycosides' RNA binding proclivity and their antiplasmid effect on an IncB plasmid. *Biochemistry*, 44, 6800-6808.
209. Liu, X.J., Thomas, J.R. and Hergenrother, P.J. (2004) Deoxystreptamine dimers bind to RNA hairpin loops. *Journal of the American Chemical Society*, 126, 9196-9197.
210. Thomas, J.R., Liu, X.J. and Hergenrother, P.J. (2005) Size-specific ligands for RNA hairpin loops. *Journal of the American Chemical Society*, 127, 12434-12435.
211. Tok, J.B.H., Bi, L.R. and Saenz, M. (2005) Specific recognition of napthyridine-based ligands toward guanine-containing bulges in RNA duplexes and RNA-DNA heteroduplexes. *Bioorganic & Medicinal Chemistry Letters*, 15, 827-831.
212. Matsumoto, C., Hamasaki, K., Mihara, H. and 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, 1857-1861.
213. Shoichet, B.K., McGovern, S.L., Wei, B.Q. and Irwin, J.J. (2002) Lead discovery using molecular docking. *Current Opinion in Chemical Biology*, 6, 439-446.
214. Blake, J.F. and Laird, E.R. (2003) Recent advances in virtual ligand screening. *Annual Reports in Medicinal Chemistry, Vol 38*, 38, 305-314.
215. Boehm, H.J., Boehringer, M., Bur, D., Gmuender, H., Huber, W., Klaus, W., Kostrewa, D., Kuehne, H., Luebbers, T. and Meunier-Keller, N. (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. *Journal of Medicinal Chemistry*, 43, 2664-2674.
216. Doman, T.N., McGovern, S.L., Witherbee, B.J., Kasten, T.P., Kurumbail, R., Stallings, W.C., Connolly, D.T. and Shoichet, B.K. (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *Journal of Medicinal Chemistry*, 45, 2213-2221.
217. Paiva, A.M., Vanderwall, D.E., Blanchard, J.S., Kozarich, J.W., Williamson, J.M. and Kelly, T.M. (2001) Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology*, 1545, 67-77.

218. Schames, J.R., Henchman, R.H., Siegel, J.S., Sotriffer, C.A., Ni, H.H. and McCammon, J.A. (2004) Discovery of a novel binding trench in HIV integrase. *Journal of Medicinal Chemistry*, 47, 1879-1881.
219. Sotriffer, C.A., Ni, H.H. and McCammon, J.A. (2000) Active site binding modes of HIV-1 integrase inhibitors. *Journal of Medicinal Chemistry*, 43, 4109-4117.
220. Cournia, Z., Leng, L., Gandavadi, S., Du, X., Bucala, R. and Jorgensen, W.L. (2009) Discovery of Human Macrophage Migration Inhibitory Factor (MIF)-CD74 Antagonists via Virtual Screening. *Journal of Medicinal Chemistry*, 52, 416-424.
221. Morand, E.F., Leech, M. and Jurgen, B. (2006) MIF: a new cytokine link between rheumatoid arthritis and atherosclerosis. *Nature Reviews Drug Discovery*, 5, 399-410.
222. Hagemann, T., Robinson, S.C., Thompson, R.G., Charles, K., Kulbe, H. and Balkwill, F.R. (2007) Ovarian cancer cell-derived migration inhibitory factor enhances tumor growth, progression, and angiogenesis. *Molecular Cancer Therapeutics*, 6, 1993-2002.
223. Filikov, A.V., Mohan, V., Vickers, T.A., Griffey, R.H., Cook, P.D., Abagyan, R.A. and James, T.L. (2000) Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *Journal of Computer-Aided Molecular Design*, 14, 593-610.
224. Guilbert, C. and James, T.L. (2008) Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *Journal of Chemical Information and Modeling*, 48, 1257-1268.
225. Pfeffer, P. and Gohlke, H. (2007) DrugScore(RNA) - Knowledge-based scoring function to predict RNA-ligand interactions. *Journal of Chemical Information and Modeling*, 47, 1868-1876.
226. Morley, S.D. and Afshar, M. (2004) Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock (R). *Journal of Computer-Aided Molecular Design*, 18, 189-208.
227. Lang, P.T., Brozell, S.R., Mukherjee, S., Pettersen, E.F., Meng, E.C., Thomas, V., Rizzo, R.C., Case, D.A., James, T.L. and Kuntz, I.D. (2009) DOCK 6: Combining techniques to model RNA-small molecule complexes. *Rna-a Publication of the Rna Society*, 15, 1219-1230.
228. Cozzini, P., Kellogg, G.E., Spyrakis, F., Abraham, D.J., Costantino, G., Emerson, A., Fanelli, F., Gohlke, H., Kuhn, L.A., Morris, G.M. *et al.* (2008) Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *Journal of Medicinal Chemistry*, 51, 6237-6255.
229. Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. and Shimada, I. (2000) A novel NMR method for determining the interfaces of large protein-protein complexes. *Nature Structural Biology*, 7, 220-223.
230. Mayer, M. and Meyer, B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie-International Edition*, 38, 1784-1788.
231. Meyer, B., Weimar, T. and Peters, T. (1997) Screening mixtures for biological activity by NMR. *European Journal of Biochemistry*, 246, 705-709.
232. Abragam, A. (1961) *Principles of Nuclear Magnetism*. Clarendon Press, Oxford.
233. Lipari, G. and Szabo, A. (1981) Nuclear magnetic resonance relaxation in nucleic acid fragments: models for internal motion. *Biochemistry*, 20, 6250-6256.
234. Lipari, G. and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *Journal of the American Chemical Society*, 104, 4546-4559.

235. Davis, D.G., Perlman, M.E. and London, R.E. (1994) Direct Measurements of the Dissociation-Rate constant for Inhibitor-Enzyme Complexes via the T-1-Rho and T-2 (CPMG) Methods. *Journal of Magnetic Resonance Series B*, 104, 266-275.
236. Deverell, C., Morgan, R.E. and Strange, J.H. (1970) Studies of Chemical Exchange by Nuclear Magnetic Relaxation in Rotating Frame. *Molecular Physics*, 18, 553-&.
237. Carr, H.Y. and Purcell, E.M. (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Physical Review*, 94, 630-638.
238. Meiboom, S. and Gill, D. (1958) Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments*, 29, 688-691.
239. Led, J.J., Gesmar, H. and Abildgaard, F. (1989) Applicability of Magnetization Transfer Nuclear Magnetic-Resonance to Study Chemical-Exchange Reactions. *Methods in Enzymology*, 176, 311-329.
240. Rao, B.D.N. (1989) Nuclear Magnetic-Resonance Line-Shape Analysis and Determination of Exchange-Rates. *Methods in Enzymology*, 176, 279-311.
241. Legge, G.B., Kriwacki, R.W., Chung, J., Hommel, U., Ramage, P., Case, D.A., Dyson, H.J. and Wright, P.E. (2000) NMR solution structure of the inserted domain of human leukocyte function associated antigen-1. *Journal of Molecular Biology*, 295, 1251-1264.
242. Hernandez, G., Jenney, F.E., Adams, M.W.W. and LeMaster, D.M. (2000) Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 3166-3170.
243. Bothner-By, A.A. (1995) In Grant, D. M. and Harris, R. K. (eds.), *Encyclopedia of Nuclear Magnetic Resonance*. Wiley, Chichester, pp. 2932-2938.
244. Hansen, A.L. and Al-Hashimi, H.M. (2006) Insight into the CSA tensors of nucleobase carbons in RNA polynucleotides from solution measurements of residual CSA: towards new long-range orientational constraints. *J Magn Reson*, 179, 299-307.
245. Tjandra, N. and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, 278, 1111-1114.
246. Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information For Structure Determination in Solution. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 9279-9283.
247. Tjandra, N. (1999) Establishing a degree of order: obtaining high-resolution NMR structures from molecular alignment. *Struct. Fold. Des.*, 7, R205-R211.
248. Prestegard, J.H. and Kishore, A.I. (2001) Partial alignment of biomolecules: an aid to NMR characterization. *Curr Opin Chem Biol*, 5, 584-590.
249. Tolman, J.R. and Al-Hashimi, H.M. (2003) In Webb, G. A. (ed.), *Annual Reports on NMR Spectroscopy*. Academic Press, Vol. 51, pp. 105-166.
250. Ram, P. and Prestegard, J.H. (1988) Magnetic-Field Induced Ordering of Bile-Salt Phospholipid Micelles - New Media For Nmr Structural Investigations. *Biochim. Biophys. Acta.*, 940, 289-294.
251. Sanders, C.R., Hare, B.J., Howard, K.P. and Prestegard, J.H. (1994) Magnetically-Oriented Phospholipid Micelles As a Tool For the Study of Membrane-Associated Molecules. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 26, 421-444.
252. Clore, G.M., Starich, M.R. and Gronenborn, A.M. (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a

colloidal suspension of rod- shaped viruses. *Journal of the American Chemical Society*, 120, 10571-10572.
253. Hansen, M.R., Hanson, P. and Pardi, A. (2000) Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods in Enzymology*, 317, 220-240.
254. Hansen, M.R., Mueller, L. and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Structural Biology*, 5, 1065-1074.
255. Wu, B., Petersen, M., Girard, F., Tessari, M. and Wijmenga, S.S. (2006) Prediction of molecular alignment of nucleic acids in aligned media. *Journal Of Biomolecular Nmr*, 35, 103-115.
256. Zweckstetter, M., Hummer, G. and Bax, A. (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophys J*, 86, 3444-3460.
257. McCallum, S.A. and Pardi, A. (2003) Refined solution structure of the iron-responsive element RNA using residual dipolar couplings. *J Mol Biol*, 326, 1037-1050.
258. Sibille, N., Pardi, A., Simorre, J.P. and Blackledge, M. (2001) Refinement of local and long-range structural order in theophylline-binding RNA using C-13-H-1 residual dipolar couplings and restrained molecular dynamics. *Journal of the American Chemical Society*, 123, 12135-12146.
259. Mollova, E.T., Hansen, M.R. and Pardi, A. (2000) Global structure of RNA determined with residual dipolar couplings. *Journal of the American Chemical Society*, 122, 11561-11562.
260. Musselman, C., Pitt, S.W., Gulati, K., Foster, L.L., Andricioaei, I. and Al-Hashimi, H.M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J Biomol NMR*, 36, 235-249.
261. Klein, D.J., Schmeing, T.M., Moore, P.B. and Steitz, T.A. (2001) The kink-turn: a new RNA secondary structure motif. *Embo J*, 20, 4214-4221.
262. Neidle, S. (1999) *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press, New York.
263. Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X., Neidle, S., Sakked, Z., Sklenar, H. *et al.* (2001) A Standard Refernce Frame for the Description of Nucleic Acid Base-pair Geometry. *Journal of Molecular Biology*, 313, 229-237.
264. Dingley, A.J. and Grzesiek, S. (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (2)J(NN) couplings. *Journal of the American Chemical Society*, 120, 8293-8297.
265. Pervushin, K., Ono, A., Fernandez, C., Szyperski, T., Kainosho, M. and Wuthrich, K. (1998) NMR scaler couplings across Watson-Crick base pair hydrogen bonds in DNA observed by transverse relaxation optimized spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 14147-14151.
266. Al-Hashimi, H.M., Pitt, S.W., Majumdar, A., Xu, W. and Patel, D.J. (2003) Mg2+-induced variations in the conformation and dynamics of HIV-1 TAR RNA probed using NMR residual dipolar couplings. *J Mol Biol*, 329, 867-873.
267. Pitt, S.W., Majumdar, A., Serganov, A., Patel, D.J. and Al-Hashimi, H.M. (2004) Argininamide binding arrests global motions in HIV-1 TAR RNA: Comparison with Mg2+-induced conformational stabilization. *Journal of Molecular Biology*, 338, 7-16.

268. Pitt, S.W., Zhang, Q., Patel, D.J. and Al-Hashimi, H.M. (2005) Evidence that electrostatic interactions dictate the ligand-induced arrest of RNA global flexibility. *Angew Chem Int Ed Engl*, 44, 3412-3415.
269. Reiter, N.J., Blad, H., Abildgaard, F. and Butcher, S.E. (2004) Dynamics in the U6 RNA intramolecular stem-loop: a base flipping conformational change. *Biochemistry*, 43, 13739-13747.
270. Reiter, N.J., Nikstad, L.J., Allmann, A.M., Johnson, R.J. and Butcher, S.E. (2003) Structure of the U6 RNA intramolecular stem-loop harboring an S(P)-phosphorothioate modification. *Rna*, 9, 533-542.
271. Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of Magnetic Resonance*, 138, 334-342.
272. Saupe, A. (1968) Recent results in the field of liquid crystals. *Angew. Chem., Int. Ed. Engl.*, 7, 97-112.
273. Tjandra, N. and Bax, A. (1997) Measurement of dipolar contributions to (1)J(CH) splittings from magnetic-field dependence of J modulation in two- dimensional NMR spectra. *Journal of Magnetic Resonance*, 124, 512-515.
274. Tolman, J.R., Al-Hashimi, H.M., Kay, L.E. and Prestegard, J.H. (2001) Structural and dynamic analysis of residual dipolar coupling data for proteins. *Journal of the American Chemical Society*, 123, 1416-1424.
275. Getz, M.M., Andrews, A.J., Fierke, C.A. and Al-Hashimi, H.M. (2006) Structural plasticity and Mg2+ binding properties of RNase P P4 from combined analysis of NMR residual dipolar couplings and motionally decoupled spin relaxation. *Rna*.
276. Musselman, C., Al-Hashimi, H.M. and Andricioaei, I. (2007) iRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. *Biophys J*, 93, 411-422.
277. Showalter, S.A., Baker, N.A., Tang, C.G. and Hall, K. (2005) Iron responsive element RNA flexibility described by NMR and isotropic reorientational eigenmode dynamics. *Journal of Biomolecular NMR*, 32, 179-193.
278. Sun, X., Zhang, Q. and Al-Hashimi, H.M. (2007) Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg2+ binding: role in the kissing-duplex structural transition. *Nucleic Acids Res*, 35, 1698-1713.
279. Zhang, Q. and Al-Hashimi, H.M. (2008) Extending the NMR spatial resolution limit for RNA by motional couplings. *Nat Methods*, 5, 243-245.
280. Zhang, Q., Throolin, R., Pitt, S.W., Serganov, A. and Al-Hashimi, H.M. (2003) Probing motions between equivalent RNA domains using magnetic field induced residual dipolar couplings: accounting for correlations between motions and alignment. *J Am Chem Soc*, 125, 10530-10531.
281. Zhang, Q., Sun, X., Watt, E.D. and Al-Hashimi, H.M. (2006) Resolving the motional modes that code for RNA adaptation. *Science*, 311, 653-656.
282. Zweckstetter, M. and Bax, A. (2000) Predicition of sterically induced alignment in a dilute liquid crystalline phase; aid to protein structure determination by NMR. *J . Am. Chem. Soc.*, 122, 3791-3792.
283. Zuker, M. (2000) Calculating nucleic acid secondary structure. *Current Opinion in Structural Biology*, 10, 303-310.
284. Dethoff, E.A., Hansen, A.L., Musselman, C., Watt, E.D., Andricioaei, I. and Al-Hashimi, H.M. (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, 3906-3915.
285. Karn, J. (1999) Tackling Tat. *Journal of Molecular Biology*, 293, 235-254.

286. Jones, K.A. (1997) Taking a new TAK on tat transactivation. *Genes Dev*, 11, 2593-2599.
287. Majello, B., Napolitano, G., Giordano, A. and Lania, L. (1999) Transcriptional regulation by targeted recruitment of cyclin-dependent CDK9 kinase in vivo. *Oncogene*, 18, 4598-4605.
288. Tao, J. and Frankel, A.D. (1992) Specific binding of arginine to TAR RNA. *Proc Natl Acad Sci U S A*, 89, 2723-2726.
289. Richter, S., Cao, H. and Rana, T.M. (2002) Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. *Biochemistry*, 41, 6391-6397.
290. Yu, X.L., Lin, W., Pang, R.F. and Yang, M. (2005) Design, synthesis and bioactivities of TAR RNA targeting beta-carboline derivatives based on Tat-TAR interaction. *European Journal of Medicinal Chemistry*, 40, 831-839.
291. Ippolito, J.A. and Steitz, T.A. (1998) A 1.3-angstrom resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 9819-9824.
292. AboulEla, F., Karn, J. and Varani, G. (1996) Structure of HIV-1 TAB RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Research*, 24, 3974-3981.
293. Clore, G.M. and Schwieters, C.D. (2004) Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. *Biochemistry*, 43, 10678-10691.

Chapter 2

Filtering MD Trajectories Using Motionally Decoupled NMR RDCs Reveals New Insights Into TAR Dynamics and Adaptive Ligand Recognition

2.1 Introduction

Characterization of RNA at atomic resolution has been a major focus of structural biology and biophysics for many years and ongoing research is aimed at developing new techniques to probe the structural architecture, however insights into mechanistic RNA structure-function questions have been impeded by the lack of biophysical techniques that afford 3D visualization of intrinsic RNA dynamics over biologically relevant timescales. NMR spectroscopy is one of the most powerful techniques for characterizing structural-dynamics of biomolecules providing comprehensive information regarding the amplitude, timescale, and - in favorable cases - direction of motions with site-specific resolution(1-3). However, even with the abundant number of measurements that can be made using NMR, the total number of parameters that can be directly determined from NMR data are much fewer than the total number of observables. While Molecular Dynamics (MD) simulations provide an all-atom description of dynamics, the force fields that govern the theoretical description of time-resolved dynamic transitions remain to be thoroughly validated particularly for nucleic acids and simulation timescales are limited to ~100 ns(4,5).

Because they are complementary on the spatial and temporal scales, the limitations inherent to NMR and MD could in principle be overcome by combining the two techniques; MD can fill the shortage of information available through NMR data by providing an all-atom time-resolved model, and NMR can provide a means for validating and potentially correcting force fields and accelerate MD conformational sampling to millisecond timescales. Several studies have emerged in which MD and NMR are used in concert in studies of nucleic acid dynamics(6-10). The two techniques can also be integrated to yield a unified view of structural dynamics. While such combined NMR/MD approaches have successfully been used in studies of protein dynamics(11-14), extension to RNA can prove very difficult because, unlike globular proteins, overall motions in highly flexible RNAs can be strongly coupled to internal motions making it difficult, if not impossible, to predict NMR data from an MD trajectory(9,15,16).

Recently, a domain-elongation strategy was introduced for decoupling internal and overall motions in RNA(17,18). As discussed in Chapter 1, a target helix is elongated so that the overall motion is slowed down relative to internal motions and rendered less sensitive to internal fluctuations in other parts of the molecule. The elongation also simplifies analysis of NMR spin relaxation(19) and residual dipolar coupling (RDC)(20,21) data because it predefines the overall diffusion or alignment tensor to be axially symmetric with principal axis oriented nearly parallel to the elongated helix axis. This makes elongated RNAs ideally suited for computing NMR observables from a given MD trajectory of a corresponding non-elongated RNA in which snapshots are aligned by superimposing the reference elongated helix.

In this chapter we describe the implementation of a general strategy that combines MD simulations and NMR RDCs measured in elongated RNAs for constructing atomic resolution ensembles of RNA structures with timescale sensitivity extending up to milliseconds. The ensembles of HIV-1 TAR (TAR) and HIV-2 TAR

constructed in this manner provide new insights into the bulge-length dependence of RNA dynamics and the mechanism of conformational adaptation on target recognition.

2.2 RDC-based Sample and Select Method

Simulations of wild-type TAR and HIV-2 TAR were generated using the CHARMM package(22) with force field parameter set 27(23). Model 3 of the unbound TAR NMR-determined structure (PDB ID# 1ANR)(24) was used as the starting coordinates for the simulations. For HIV-2 TAR, starting coordinates were obtained by removing argininamide from model 1 of the argininamide-bound NMR-determined structure (PDB ID# 1AKX)(25). The RNAs were charge-neutralized using sodium counter ions and solvated in a 35 Å and 33 Å sphere of TIP3 water, for TAR and HIV-2 TAR, respectively(26). A spherical boundary potential was applied to maintain the density of water around the RNA site(27). The system was minimized and heated to 300K while harmonically constraining the heavy atoms of the RNA with a force constant of 62 kcal/mol/Å for 100 ps, after which constraints were removed and the system equilibrated for 1 ns. A Nosé-Hoover thermostat(28,29) was used to maintain a constant temperature of 300 K throughout the simulation, with a 1 fs time-step and a coupling constant of 50 ps^{-1}. 50 distinct trajectories were initiated from this equilibrated structure, by assigning different initial velocities. It has previously been demonstrated that this technique can be used to enhance conformational sampling relative to a single trajectory of the same total duration(30,31). The first 0.5 ns of each trajectory was discarded and next 1.6 ns used for analysis. Conformations from each of the 50, 1.6 ns, trajectories were pooled to give a total effective simulation time of 80 ns. These 80,000 conformations were used as a structural pool for the “selection” phase of the select-and-sample strategy (SAS, see below).

In the original implementation, Chen and coworkers used S^2 NH spin relaxation order parameters to select structural ensembles for a series of protein systems(11). We adapted this "sample and select" (SAS) approach to allow use of RDC data which probe the orientational dynamics of individual bond vectors over timescales extending up to milliseconds. In the SAS approach, one generates a set of conformations for the system of interest, in which we use MD to generate such structures. One then searches for an *N*-membered subset of structures that minimizes a cost function. Specifically, an *N*-membered subset of structures is randomly selected from a total pool of *M* structures and an initial χ^2 value is evaluated using Equation [2.1]. Next, one of the *N*-membered structures is randomly chosen and replaced, by a random structure from the remaining *M-N* conformational pool. The "move" from step *k* to *k*+1 is then accepted if $\chi^2_{(k+1)} < \chi^2_{(k)}$. If $\chi^2_{(k+1)} > \chi^2_{(k)}$, the move is accepted with a probability $p = \exp((\chi^2_{(k)} - \chi^2_{(k+1)})/T_{eff})$, where T_{eff} is an effective temperature that is linearly decreased in a simulated annealing scheme. The cost function is,

$$\chi^2 = L^{-1}\sum_{i}^{L}\left(D_{ij}^{cal} - D_{ij}^{meas}\right), \qquad [2.1]$$

where D_{ij}^{cal} and D_{ij}^{exp} are the calculated and measured RDCs, respectively, *L* is the total number bond vectors, and D_{ij}^{cal} is calculated using,

$$D_{ij}^{cal} = \frac{\mu_0 \gamma_i \gamma_j \hbar}{8\pi^3 \left\langle r_{ij}^3 \right\rangle} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle , \qquad [2.2]$$

where $\gamma_{i \text{ or } j}$ is the gyromagnetic ratio of the i^{th} *or* j^{th} nucleus, r_{ij} is the bond length, θ is the angle between the internuclear bond vector and the external magnetic field and the angular brackets denote a time-average over all sampled orientations. As described in Chapter 1, the time-averaged angular term can be expressed in terms of the time-independent orientation of an internuclear vector relative to an arbitrary frame and the five order tensor elements (S_{kl})(32,33),

$$\left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle = \sum_{kl=xyz} S_{kl} \cos(\alpha_k)\cos(\alpha_l) \quad [2.3]$$

,

where $\langle_n$ is the angle between the internuclear vector and the n^{th} axis of the arbitrary frame. Equation [2.3] assumes that internal motions do not affect overall alignment of the molecule. This assumption can break down in RNA because collective motions of helical domains can lead to large changes in the overall shape and thus overall alignment of the molecule(15,16,34). This can make it impossible to accurately compute D_{ij}^{cal} for a given ensemble and thus to use RDCs in selecting conformers from an MD trajectory. Domain-elongation allows one to overcome this problem by rendering the overall shape of the molecule far less sensitive to internal motions(17,18). Elongation also predefines the overall order tensor to be axially determined with principal direction oriented approximately parallel to axis of the elongated helix(17). This overall order tensor can be conveniently determined experimentally using RDCs measured in the elongated helix(17).

The SAS RDC approach was implemented using C++ programs written by Frank A. T. F. The previously reported(17) 47 and 35 TAR one-bond base and sugar C-H RDCs measured in elongated helix-I TAR (EI-TAR) and elongated helix-II TAR (EII-

TAR), respectively, were used independently or in combination to select an TAR structural ensemble from a pool of 80,000 conformers derived from the combined 80 ns MD trajectory. For HIV-2 TAR, 35 RDCs measured in helix-1 elongated HIV-2 TAR (HIV-2 EI-TAR) were used. Note that although RDCs were measured on a TAR construct in which the apical loop was replaced by a UUCG loop, a detailed NMR study recently showed that this apical loop replacement does not affect interhelical motions or local motions at the TAR bulge(35). When used in combination, both the EI-TAR and EII-TAR sets of RDCs were used (L =47+35=82) in computing the χ^2 in Eq. [1]. In all cases, the overall alignment tensor was assumed to be axially symmetric (η=0) with principal direction (S_{zz}) oriented parallel to the elongated helix axis as computed using the program CURVES(36,37). The magnitude of S_{zz} was obtained from an order tensor analysis as described previously.

Each RDC-SAS run was initiated from N randomly selected conformers. A Monte Carlo (MC) simulated annealing scheme was then used to minimize the cost function in Eq. [1] as described above. Simulations were started at a high effective temperature where the MC acceptance probability was high (~0.99), and slowly decreased until the MC acceptance probability was ~10^{-5}. At a given effective temperature 10^5 MC steps were carried out. The effective temperature was then decreased, with $T_{i+1} = 0.9T_i$. The same protocol was used for HIV-2 TAR, except that only 35 RDCs measured in the helix-I elongated sample were used in the SAS analysis. Here, there was less motivation to acquire an additional set of helix II elongated RDCs given the near linear alignment of the two helices which renders the two sets of data degenerate.

The base angles, buckle (κ), opening (σ), propeller twist (ω) and twist (Ω), were calculated for eight non-terminal base pairs for each of the SAS selected TAR conformers using the program 3DNA(38). The interhelical Euler angles $\langle_h$, β_h, and γ_h

were also computed for each member of the ensemble, using the lower helix as a reference, as previously described(39). Note that degenerate sets of Euler angles reflect the same interhelical orientation(39) and the set of Euler angles were selected to minimize the equation $\sqrt{\alpha_h^2 + \beta_h^2 + \gamma_h^2}$.

2.3 Selection of Distinct Conformers from a Large Conformational Pool

The ability to select discrete conformers to capture distinct features of a large conformational pool was investigated using 20 ns of the 80 ns TAR MD trajectory, resulting in a total conformational pool of 20,000 structures. For thee simulations, all 20,000 conformers were considered as the TAR ensemble. RDCs for bond vectors that were experimentally measured (47 from EI-TAR and 35 from EII-TAR) were back-calculated from each conformer assuming helix-1 and helix-2 alignment tensors from measured EI-TAR and EII-TAR RDCs, respectively(17). The 20,000 sets of RDCs were then averaged independently for each elongation to give the final input RDCs. Using all 20,000 conformers as the selection pool, we performed a series of SAS

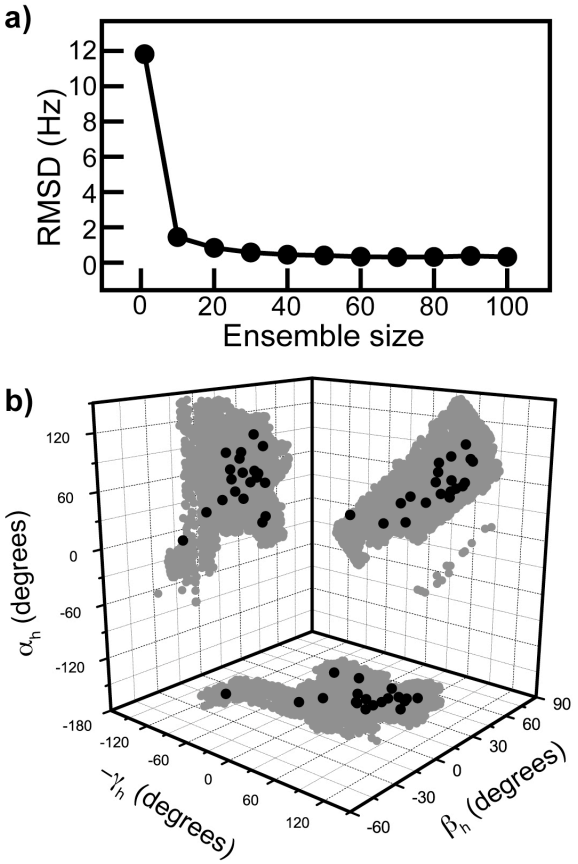


Figure 2.1. SAS selection of 20 structures to model a structurally diverse ensemble. **(a)** A selected ensemble size of *N*=20 minimized the cost function below 1 Hz RMSD. **(b)** The 20 selected conformers span the range of Euler angles and are more densely populated in regions that are often sampled in the MD trajectory.

simulations with various ensemble sizes of *N*=1, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100, and calculated the root mean square derivation (RMSD) between measured and predicted RDCs. An ensemble size of *N*=20 minimized the cost function below 1 Hz with minimal decrease on increasing the ensemble size and was used as the final set of selected conformers (Figure 2.1a). Euler angles were used to compare the 20 selected conformers to the MD trajectory. The relative orientation of two helices *i* and *j* can be defined using three interhelical Euler angles that describe the twist angle $\langle_h$ about helix *i*, the interhelix bend angle β_h, and the twist angle γ_h about helix *j*(17,39). Indeed, the Euler angles of the 20 discrete conformers largely span the range of conformations sampled in the trajectory and are more densely populated in regions that are often sampled in the MD trajectory (Figure 2.1b). Thus, using SAS, a discrete set of structures can be used to model an ensemble of many conformations.

The ability to select a particular set of 20 from the pool of 20,000 was examined by using the first 20 conformers of the trajectory to represent the "known" theoretical ensemble. RDCs for the 20 structures were calculated as previously described. The 20 sets of RDCs were then averaged independently for each elongation to give the final input RDCs. All 20,000 conformers were used as the selection pool and SAS simulations run using input RDCs from theoretical EI-TAR, EII-TAR, and both EI-TAR and EII-TAR RDCs. For all three SAS simulations, the RMSD between input and selected RDCs was <1 Hz. Only 8 of the 20 conformers were selected using EI-TAR RDCs, and the remaining 12 selected conformers are within 3Å backbone heavy-atom RMSD to the unselected structures (Figure 2.2a). When only considering EII-TAR RDCs, 3 of the 20 conformers are selected and the remaining 17 conformers exhibit a backbone heavy-atom RMSD < 3Å to the remaining structures (Figure 2.2b). Remarkably, when EI-TAR and EII-TAR RDCs are used all 20 conformers are selected (Figure 2.2c). The ability to more accurately select the correct structures using EI-TAR and EII-TAR RDCs is most

likely a result of the increased sensitivity to helical motions that is realized when combining both elongations(17). To simulate experimental error, a 3 Hz error was added to the theoretical RDCs. When using EI-TAR and EII-TAR theoretical RDCs with simulated 3 Hz experimental error, 16 of the 20 structures are selected and the remaining 4 are within 4Å backbone heavy-atom RMSD to the unselected theoretical structures (Figure 2.2d). Thus, experimental error does not preclude selecting most all of the 20-members of the ensemble using SAS.

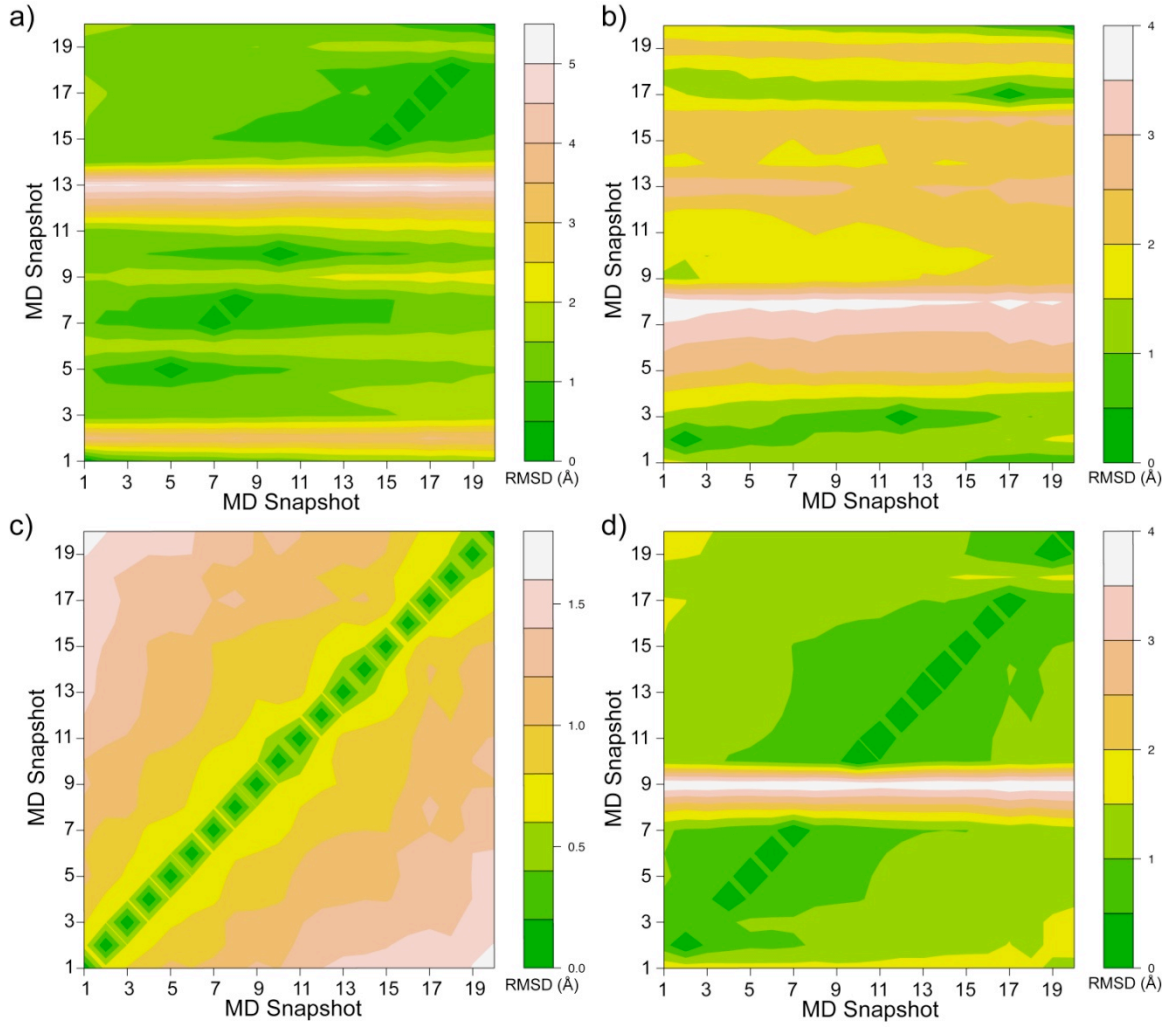


Figure 2.2. Using SAS to select a 20-member ensemble. Back-calculated RDCs from the first 20 MD conformers represent the "known" TAR ensemble and are used as input for the SAS selection. Results from SAS simulations using **(a)** EI-TAR RDCs, **(b)** EII-TAR RDCs, **(c)** EI-TAR and EII-TAR RDCs, and **(d)** EI-TAR and EII-TAR RDCs including a simulated 3 Hz experimental error are shown.

The SAS methodology can be used to accurately select 20 conformers provided that enough input experimental RDCs are measured in different reference frames. However, the first 20 conformers of MD trajectory are quite similar and cover relatively limited conformational space(Figure 2.3a). For example, the interhelical bend angles range only 24.3-36.7°. To test the ability to select 20 conformers that cover a wide range of conformational space, 20 structures were randomly chosen from the MD trajectory. Interhelical angles for the 20 random conformers range 12.4-60.9° (Figure 2.3a). Simulated EI-TAR and EII-TAR RDCs were calculated as stated above and combined to give the input RDCs. To simulate experimental conditions, a second set of "known"

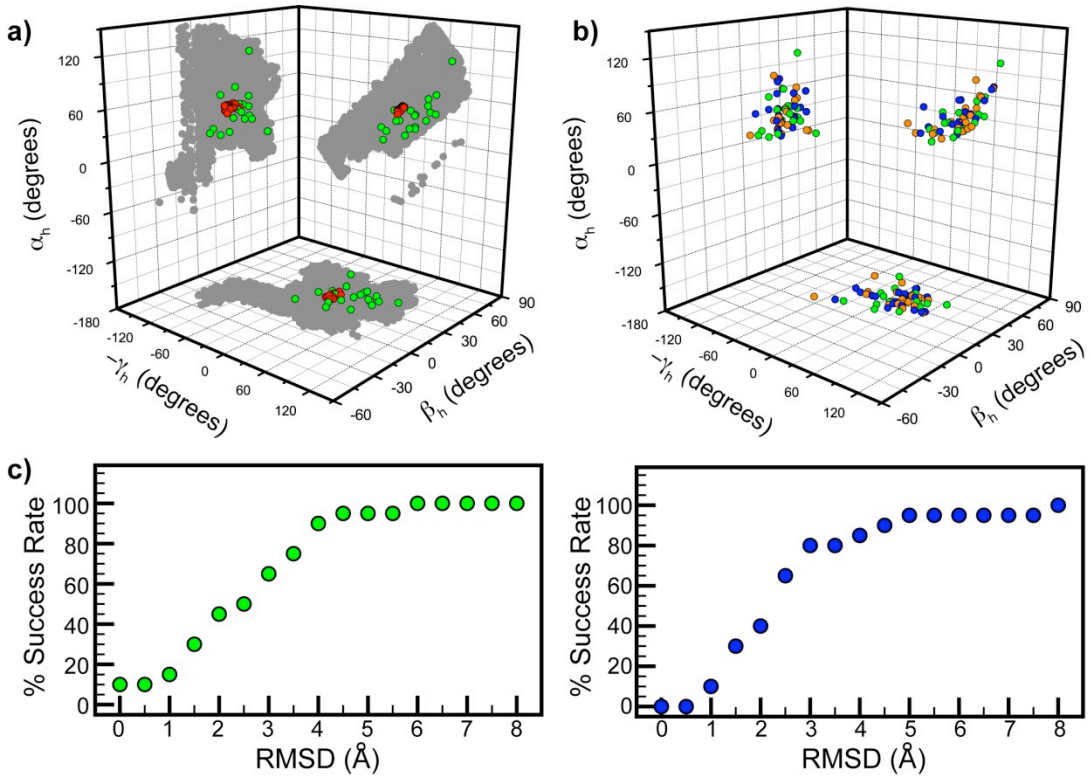


Figure 2.3. SAS selection of structurally diverse conformers. (a) The first 20 snapshots from the MD trajectory (red) cover a small conformational space as shown by the Euler angles, whereas the 20 randomly selected conformers (green) span more of the Euler angle space in the MD trajectory (gray). **(b)** SAS-selected 20-member ensembles selected assuming no experimental error (orange) and a 3 Hz error (blue) exhibit Euler angles in close agreement with the 20 randomly selected snapshots. **(c)** Most of the 20 randomly selected conformers are selected within 3Å backbone heavy-atom root mean square deviation (RMSD).

RDCs was generated by adding 3 Hz to each RDC. Although the exact 20 conformers are not selected using either set of "known" RDCs, the RMSD between the input and selected RDCs for both data sets is <1 Hz. Thus, for structurally diverse ensembles, there exist degenerate solutions that minimize the cost function.

Success in selecting the input structures was verified by comparing Euler angles and backbone heavy-atom RMSD for the input and selected structures(Figure 2.3). Euler angles for the input and selected conformers are in close agreement (Figure 2.3b). The average $\langle_h$, β_h, and γ_h angles for the input structures (26.0°, 35.1°, 10.6°) are in close agreement with those selected using the input RDCs assuming no experimental error (26.0°, 34.7°, 6.2°) and an error of 3 Hz (29.0°, 34.0°, 6.2°). Standard deviations for the $\langle_h$, β_h, and γ_h are also similar for the input structures (24.9°, 11.9°, 28.4°), structures selected using RDCs with no experimental error (24.1°, 10.8°, 33.7°), and the structures selected assuming a 3 Hz error (20.8°, 10.6°, 30.4°). While the exact structures are not selected using either set of input data, small backbone heavy-atom RMSDs are calculated between the input and selected conformers. We define the success rate as the percentage of structures selected within a defined heavy atom RMSD cutoff. Using a cutoff of 3 Å, 13 of the 20 conformers are selected using input RDCs assuming no experimental error, and 16 of the 20 conformers are selected when a 3 Hz experimental error is introduced (Figure 2.3c). All 20 conformers are selected at an RMSD of 6 Å assuming no experimental error and 8 Å assuming an error of 3 Hz (Figure 2.3c). Thus, the SAS methodology can be used to select 20 conformers that sufficiently re-capitulate the input data, and it is likely that more RDC data will minimize the differences between the input and selected conformers and eliminate the degenerate solutions for structurally diverse ensembles. Note that in all experiments discussed in section 2.3, similar results were obtained using all 80,000 conformers from the TAR MD trajectory, but only 20,000 were used to ensure computational efficiency.

2.4 Analysis of E-TAR SAS Ensemble

In Figure 2.4, we compare the RDCs measured previously(26) in EI-TAR and EII-TAR with those predicted based on the combined 50 x 1.6 = 80 ns MD trajectory. Though some correlation is observed between the measured and predicted RDCs for both EI-TAR (Figure 2.4b) and EII-TAR (Figure 2.4c), the deviations (RMSD ranging between 13-16 Hz) remains substantially larger than the estimated RDC measurement uncertainty (~3-4 Hz). The MD trajectory does not systematically under or overestimate the RDCs measured throughout the RNA, indicating that it does not significantly over or underestimate the amplitude of motions present. This is the case even though the RDC timescale sensitivity to motions (<ms) is greater than that of MD (~80 ns). Thus, it appears that the motions in TAR saturate at nanosecond timescales consistent with relaxation dispersion studies that provide no evidence for ⌠s-ms motions in the helix-bulge-helix element of TAR(35).

We examined if SAS could be used to pull out sub-ensembles from the MD trajectory that satisfy the measured RDCs. Similar to the validation simulations in section 2.3, we first conducted a series of SAS runs with various ensemble sizes of *N*=5, 10, 20, 30, 40, 50 and 100. The RMSD between measured and predicted RDCs when combining the EI-TAR and EII-TAR RDCs is shown in Figure 2.4d as a function of *N*. Increasing the ensemble size beyond *N*=20 did not lead to significant improvements in the fit and in fact a deterioration was observed for *N* >20. This is likely due to sampling problems during the Monte Carlo simulated annealing minimization of the cost function as the number of possible combinations increases steeply with *N* and the cost function exhibits, due to the frustration stemming from the underdetermined nature of the problem, a vast number local minima. An ensemble size of *N*=20 was used in all subsequent SAS runs.

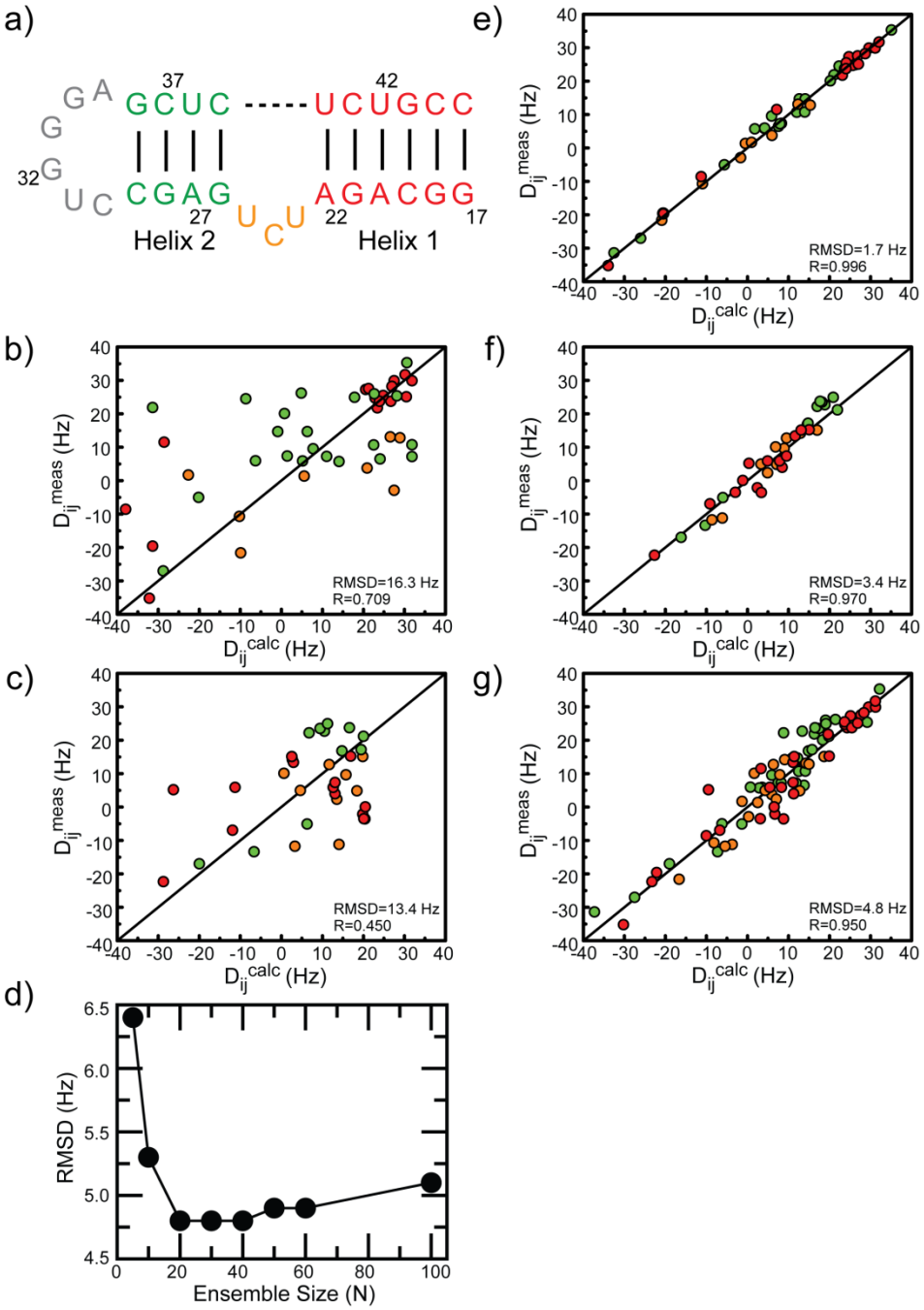


Figure 2.4. SAS selection of E-TAR RDCs. (a) Secondary structure of TAR with helix-I highlighted in red, helix-II in green and tri-nucleotide bulge in orange. HIV-2 TAR lacks bulge residue C24. **(b-c)** Plots of experimental RDCs versus values computed from the 80 ns MD trajectory for **(b)** EI-TAR and **(c)** EII-TAR. Data for helix-I, helix-II, and bulge, are shown in red, green and orange, respectively. Also shown is the root mean square deviation (RMSD) and correlation coefficient (*R*). **(d)** RMSD (Hz) between calculated and experimental RDCs as a function of *N*, following a SAS analysis using both EI-TAR and EII-TAR RDCs. **(e-g)** Plots of experimental RDCs versus values calculated from the *N*=20 SAS ensemble using **(e)** EI-TAR **(f)** EII-TAR, and **(g)** EI-TAR and EII-TAR RDCs.

The RDCs calculated using a 20-member ensemble selected using the SAS approach exhibit a markedly improved fit to the EI-TAR (Figure 2.4e), EII-TAR (Figure 2.4f) and EI-TAR+EII-TAR (Figure 2.4g) RDCs as compared to those calculated from the entire MD trajectory. Importantly, sub-ensembles can be determined that simultaneously reproduce the EI-TAR and EII-TAR RDCs with an RMSD (4.8 Hz) that is comparable to the estimated experimental RDC uncertainty (~3-4 Hz) (Figure 2.4g). The SAS selected conformers also did not lead to any steric collisions with the elongated helices for both EI-TAR and EII-TAR despite the fact that the helices were not actually elongated in the MD simulations (data not shown).

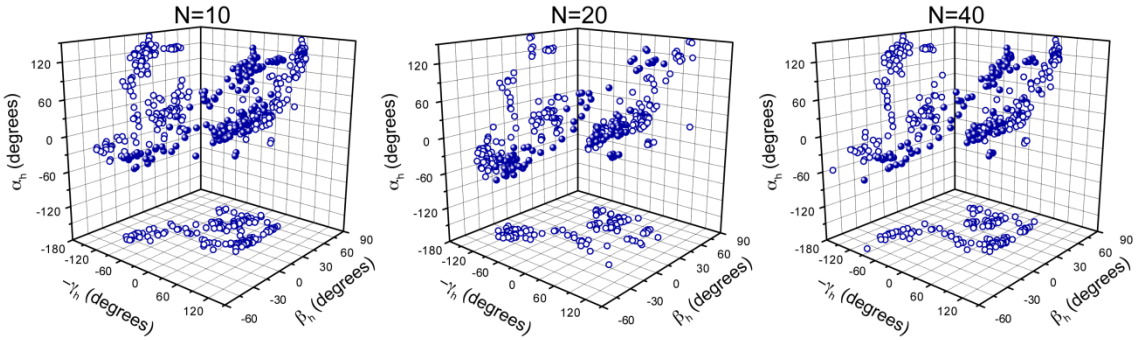


Figure 2.5. Interhelical Euler angles of SAS selected TAR conformers with *N* = 10, 20, and 40.

As shown in section 2.3, degenerate solutions exist for structurally diverse ensembles with N=20. To examine the uniqueness of the selected TAR ensemble, we carried out 100 independent *N* = 20 SAS runs. Here, a maximum of 2000 unique snapshots can be selected. However, repeated runs resulted in selection of a narrow set of similar conformations. The total number of unique snapshots selected by the RDC-SAS optimization algorithm was 422, 48, and 121 for EI-TAR, EII-TAR, and EI-TAR+EII-TAR RDCs, respectively. Thus, the RDC data favors selection of specific conformations from the available pool. As expected, the conformers selected did vary when changing the value *N*. However, the overall distribution of conformations remained similar as shown for example for the interhelical orientation in Figure 2.5.

2.5 SAS Ensemble Provides Details of Local and Collective Motions

In Figure 2.6a, we plot (in black) the average and standard deviation for various base angles computed for the 121 unique TAR conformers obtained from 100 *N*=20 SAS runs. For comparison, the mean value and standard deviation for idealized A-form helical geometry computed from a statistical comparison of high-resolution X-ray structures is shown in red(40). For the majority of the residues, very good agreement is observed between the SAS ensemble and canonical values. Large deviations are however observed for the junctional A22-U40 base-pair, which in the MD simulation frequently deviates from a hydrogen bonded alignment. This is in excellent agreement with previous NMR data showing that while the junctional G26-C39 base-pair forms a detectable hydrogen bond, the A22-U40 base-pair is flexible and does not form the expected base-pair in TAR(41-43).

The bulge residues exhibit different levels of motions. As shown in Figure 2.6b, the highly conserved U23 bulge stacks onto A22 in the majority of the conformations, consistent with observation of NOE connectivity between A22 and U23. These two residues undergo limited motions consistent with previous ^{13}C relaxation studies of dynamics in elongated TAR(42). Interestingly, select conformations exist in which U23 adopts a looped out conformation as observed in several ligand bound TAR structures (PDB ID# 1QD3, 1UTS, and 397D) In contrast, residues C24 and U25 predominantly exist in a looped out conformation and are significantly more flexible, again in agreement with the previous ^{13}C relaxation studies(42).

We computed the three interhelical Euler angles for the 121 SAS selected TAR conformers and compared them to angles obtained for the entire MD trajectory. As shown in Figure 2.7a, the MD trajectory spans a large range of interhelical angles. The SAS conformers are widely distributed across the MD trajectory (Figure 2.7a). For both the MD trajectory and SAS ensembles, significant correlation is observed between the

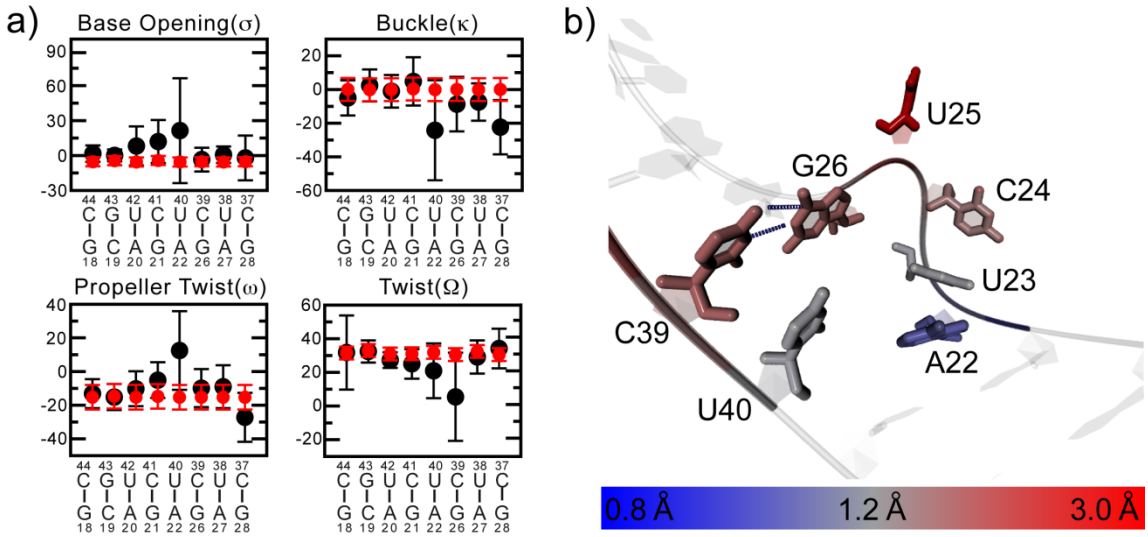


Figure 2.6. Local motions in the TAR dynamical ensemble. (a) Shown in black are the mean values for the base opening (∫), buckle (|), propeller twist (⌉) and twist (∧) angles and their standard deviation calculated over 121 TAR conformers obtained from multiple *N*=20 SAS runs. For comparison, shown in red are corresponding values for an idealized A-form helix as obtained from a statistical survey of high resolution X-ray structures(40). **(b)** Average conformation of the TAR bulge and neighboring base-pairs calculated from the 121-membered SAS ensemble. The bases of the bulge and flanking base pairs are color coded based on the root mean square fluctuations (RMSF) calculated over the ensemble.

three interhelical angles, particularly between the twist angles $\langle_h$ and γ_h. Such spatial correlations were reported previously based on an a three-state rigid-body refinement of the TAR interhelical orientation(17). As shown in Figure 2.7b, the conformations obtained by SAS sample orientations that are in very good agreement with those obtained by a previous 3-state analysis of the RDCs(17). The only significant deviations are observed in the $\langle_h$-γ_h plane for the near coaxial conformer (β_h~-21°). This is not surprising given that for this near coaxial conformer, the RDCs measured in both EI-TAR and EII-TAR are highly insensitive to the twist angles $\langle_h$ and γ_h. We also examined if there exists any correlation between the geometry of base-pairs at the junction and the interhelical bend angles. The only significant observation was anti-correlation (*R* ~-0.70) between the base-pair step angle (Ω) at the G26-C39 junctional base-pair and $\langle_h$ the

twist angle about helix II. Thus, the unraveling of the G26-C39 base-pair may give rise to twisting motions around the axis of helix II.

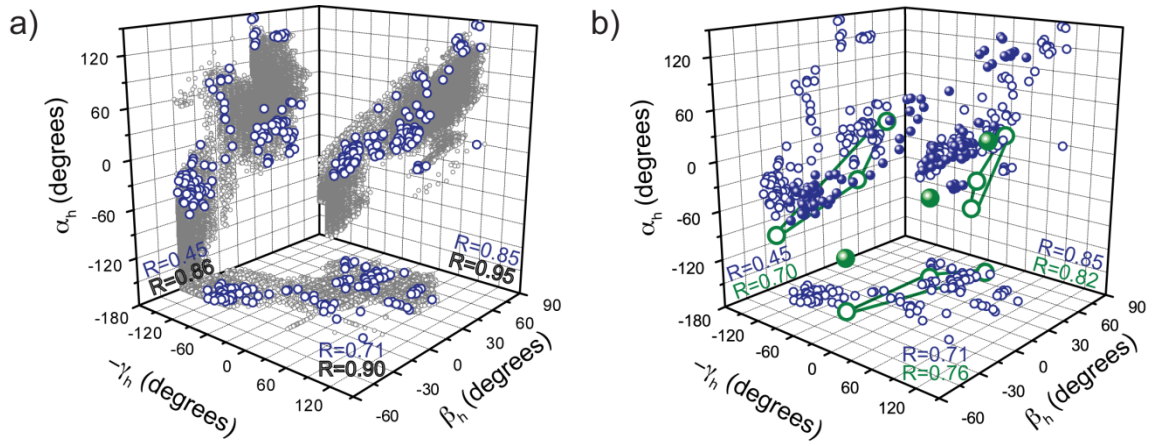


Figure 2.7. Global interhelical dynamics in the TAR dynamical ensemble. (a) Shown in gray are the interhelical twist ($\langle_h$ and γ_h) and bend (β_h) angles for 80,000 TAR conformers derived from an 80ns MD trajectory (conformer selected every 1 ps). In blue are the corresponding SAS selected conformers. The correlation coefficient (*R*) is shown on individual planes. **(b)** Comparison of the SAS selected interhelical angles and those derived previously(26) based on a 3-state rigid body ensemble analysis of E-TAR RDCs.

2.6 Comparison of TAR and HIV-2 TAR SAS Ensembles

We used the SAS approach to analyze RDCs previously measured in HIV-2 EI-TAR(**26**) in which the bulge residue C24 is omitted. Poor agreement (RMSD = 15.1 Hz) was again observed between the measured EI-TAR RDCs and values computed using entire MD trajectory (Figure 2.8a). By using SAS, we were able to find an *N*=20 sub-ensemble that yields an RMSD of 1.7 Hz (Figure 2.8b). Compared to TAR, repeated SAS runs resulted in selection of a larger number (276) of unique conformers for HIV-2 TAR. This could be attributed to a smaller RDC sensitivity to twisting motions both because the HIV-2 TAR structure is more linear and because only RDCs measured in EI-TAR were available for analysis.

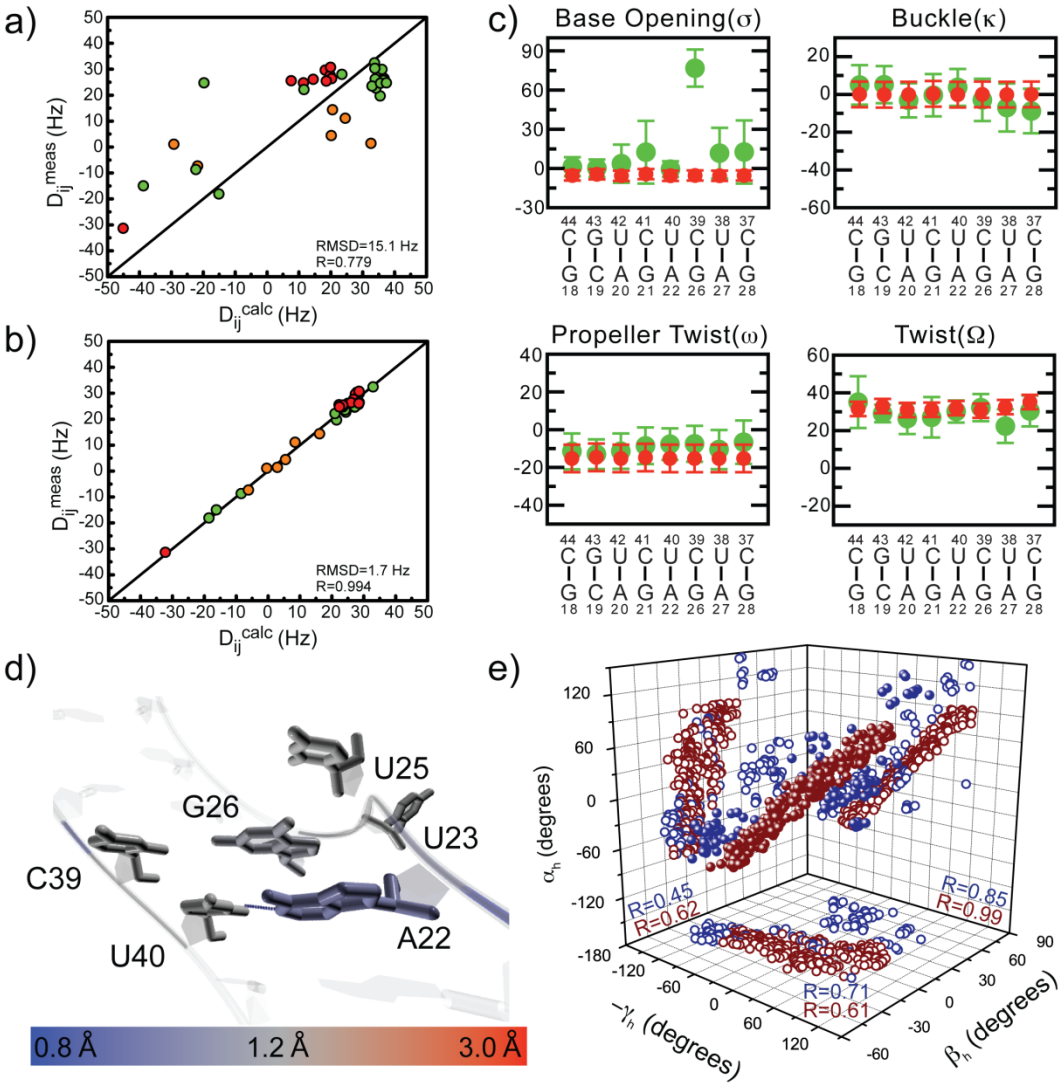


Figure 2.8. HIV-2 TAR dynamical ensemble. Plot of experimental HIV-2 EI-TAR RDCs versus values calculated from **(a)** an 80 ns MD trajectory of HIV-2 TAR and **(b)** following SAS selection of *N*=20 conformers from the MD trajectory based on HIV-2 EI-TAR RDCs. Coloring scheme is same as in Figure 2.4. **(c)** Shown in black are the mean base-pair angles and their standard deviation calculated over a 279-membered HIV-2 TAR SAS ensemble. For comparison shown in green are corresponding angles for an idealized A-form helix as obtained from a statistical survey of high resolution X-ray structures(48). **(d)** Average conformation of the bulge of HIV-2 TAR calculated from the 279-membered HIV-2 TAR SAS ensemble. Bases of the bulge and flanking base pairs are color coded based on the root mean square fluctuations (RMSF) calculated over the ensemble. **(e)** Comparison of SAS selected TAR (blue) and HIV-2 TAR (red) interhelical conformations. The correlation coefficient (*R*) is shown on individual planes.

Comparison of the SAS selected HIV-2 TAR conformers with those obtained for TAR revealed that reducing the length of the bulge in HIV-2 TAR led to a marked

decrease in the local motions in the junctional A22-U40 base-pair (Figure 2.8c). In TAR the standard deviations are approximately 30°, 45°, 13°, 16° for the base angles κ, σ, ω and Ω respectively. In HIV-2 they reduce to 10°, 6°, 8°, and 6° respectively. In contrast, we observe significant static deviations in the opening angle for the G26-C39 junctional base-pair. Likewise, a significant reduction is observed in the local dynamics of bulge residues U23 and C25 (Figure 2.8d). U23 is less flexible and forms more stable stacking interactions on an also less flexible A22. The root mean square fluctuations (RMSF) of the atomic positions of U23 and U25 decrease from 1.89 Å and 3.49 Å respectively in TAR to 1.45 Å and 1.49 Å respectively, in HIV-2 TAR.

The reduction in the local motions in and around the bulge linker is as expected, accompanied by a reduction in the interhelical motional amplitudes, as shown in Figure 2.8e. Such a reduction is clearly observed for the interhelical bending which decreases in standard deviation from ~33° to ~12°. This is in agreement with an order tensor analysis of RDCs, which reported a reduction in the GDO_{int} (which ranges between 0 and 1 for maximum and minimum interhelical motional amplitudes, respectively) from 0.45±0.10 to 0.77±0.04(17). As expected, the HIV-2 conformers cluster more tightly around more linear (β_h~0°) conformations

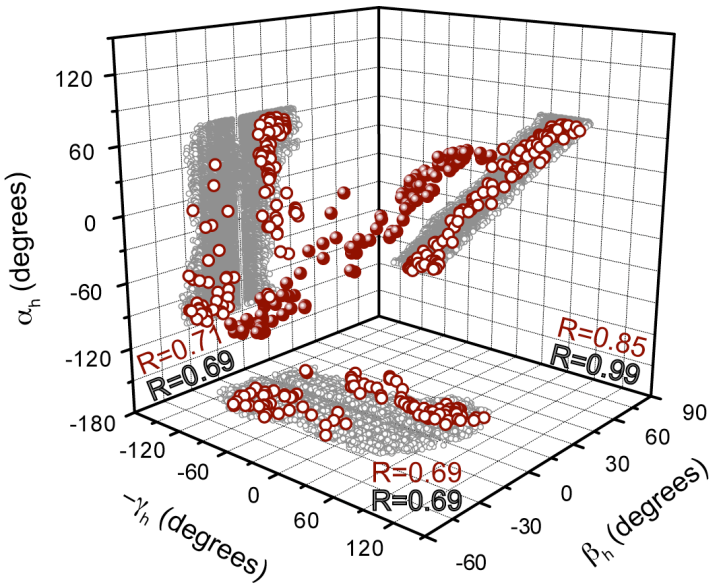


Figure 2.9. Interhelical Euler angles of the HIV-2 TAR dynamical ensemble. Euler angles obtained from an 80 ns MD trajectory are shown in gray and the SAS selected structures shown in red. The correlation coefficient (*R*) is shown on individual planes.

(Figure 2.9). While we do not observe a significant reduction in the amplitude of twisting motions about the two helices (α_h and γ_h) in HIV-2, the RDC sensitivity to these angles is diminished in HIV-2 TAR both because the structure is on average more linear and because RDCs were only measured on the helix-I elongated construct (Figure 2.9). Nevertheless, we observe a stronger correlation between the twisting motions indicating that the $\langle_h$ and γ_h correlations originates in part from the steric drag one helix exerts on the other.

2.7 Insights into TAR-Ligand Recognition by Conformational Selection

Numerous studies have shown that TAR undergoes large conformational rearrangements that allow binding of diverse targets in and around the bulge, including peptide derivatives of its cognate protein Tat(44-46), divalent ions(47), and five chemically distinct small molecules(48-51). The three interhelical conformers obtained by a 3-state ensemble analysis of RDCs measured in unbound TAR revealed a global interhelical motional trajectory that encapsulated many of the ligand bound conformations(17). As shown in Figure 2.10a, the SAS selected interhelical conformers trace orientations similar to the ligand-bound TAR conformations, supporting the notion that unbound TAR can dynamically access its ligand-bound global conformations. This can also be seen in Figure 2.10b in which we compare the best matching ligand-bound TAR conformers and SAS-selected conformers as determined by superimposing all heavy atoms excluding the flexible terminal base-pairs (G17-C45) and the apical loop.

The SAS dynamical ensembles also allowed us to examine to what extent are local features of the ligand binding pocket dynamically preformed in the absence of ligands. In Figure 2.10c, we compare the best matching ligand-bound TAR conformers and SAS-selected conformers as determined by superimposing heavy atoms in the bulge and immediately neighboring base-pairs. In general, the largest deviations are

seen for the highly flexible residues C24 and U25, which are also known to be flexible in many of the ligand bound TAR conformations(17,52,53). Overall, these results suggest that local dynamics in and around the TAR bulge likely facilitate formation of the ligand binding pocket.

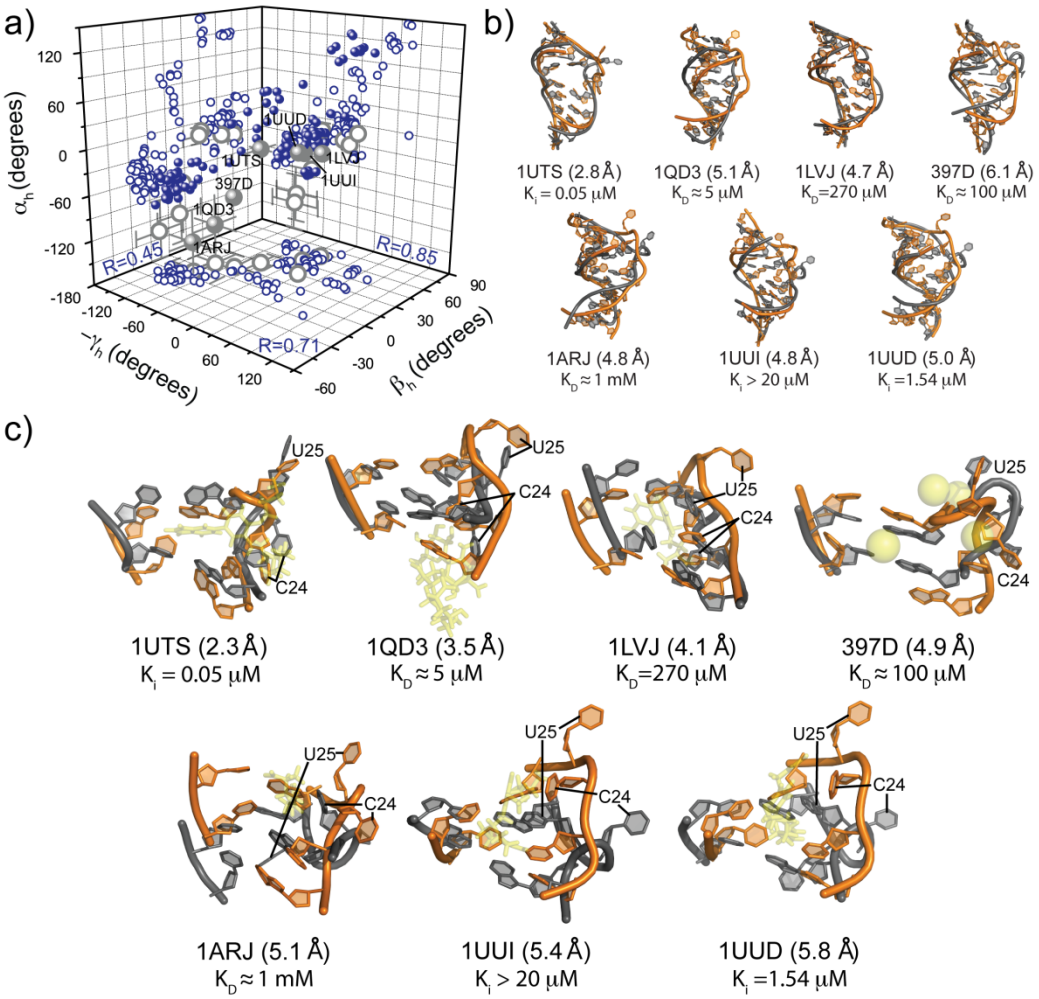


Figure 2.10. Comparison of the SAS derived TAR dynamical ensemble and ligand-bound TAR conformations. (a) Shown in blue are the SAS selected angles and seven distinct ligand-bound TAR structures in gray. **(b, c)** Shown is the comparison of the **(b)** global and **(c)** local structure of SAS TAR conformers and seven distinct ligand-bound TAR conformations (PDB ID#s 1QD3, 1UUI, 1UTS, 1UUD, 1ARJ, 1LVJ, and 397D). Shown are the pairs yielding the lowest RMSD fit when superimposing (b), all heavy atoms excluding terminal base-pairs G17-C45 and the apical loop and (c), all heavy atoms in the bulge and immediately adjacent base-pairs. Every model in the ligand bound NMR ensembles was used in the superposition. The corresponding ligand is colored yellow.

2.8 Conclusion

RNA elongation provides a reliable approach for computing time-averaged RDCs in highly flexible RNAs providing a basis for integration with MD as we showed here using the SAS approach. That the measured RDCs in both TAR and HIV-2 TAR can be satisfied using conformers selected from the MD trajectory suggests that neither TAR nor HIV-2 TAR undergo significant ⌠s-ms motions that are not sampled in the MD trajectory. This is consistent with relaxation dispersion NMR studies of TAR, which provide no evidence for ⌠s-ms motions in and around the bulge(35).

However, it should also be noted that the SAS conformers represent a “discrete” approximation to what is more likely a continuous complex distribution of many more conformations. The SAS conformers can be thought of as discrete points along the configuration space that may help define salient features of the motional trajectory(54). It should be kept in mind that averaging over such a discrete number of conformers leads to efficient averaging of the RDCs. A correspondingly larger continuous distribution of many more conformations will likely be required to accomplish the same level of motional averaging. It is therefore not surprising that the MD trajectory evolves outside the envelope defined by the SAS conformers and that the SAS envelope increases with the size of the ensemble (Figure 2.7). Thus, the most likely source of discrepancy between the measured RDCs and the MD simulation is the assignment of the relative weights to the various conformers. RDC studies of partially unfolded proteins have emphasized the exquisite sensitivity of RDCs to the underlying conformational distribution(55).

The comparison of the dynamical ensembles generated for TAR and HIV-2 TAR provide fundamental new insights into the dependence of RNA dynamics on the bulge length. As would be expected, reducing the length of the TAR trinucleotide bulge by a single nucleotide led to a significant reduction in the local motions in and around the

bulge, as well as global interhelical motions. In both TAR and HIV-2 TAR, we observe spatial correlations between twisting motions about individual helices, and to a lesser extent between twisting and bending as first reported based on the 3-state ensemble analysis of the RDCs(17). Importantly, the spatial correlations between the twisting motions increase significantly in HIV-2 TAR (Figure 2.8e). This is expected if one were to consider limiting cases for the bulge length; at the limit of not having a bulge linker, the twisting dynamics becomes perfectly correlated, whereas for an infinitely long bulge, one would expect little correlation.

The TAR dynamical ensemble allowed us to directly examine if unbound TAR dynamically samples diverse ligand-bound conformations that have been reported to date. Our results suggest that many of the key features of the ligand bound TAR conformations, including the global interhelical orientation and local aspects of the ligand-binding pocket appear to be dynamically preformed in unbound TAR. However, a detailed comparison of the SAS conformations and the ligand bound TAR structures remains complicated by a number of factors. As mentioned above, the SAS conformers only represent an approximate discrete state ensemble to what is likely a more continuous conformational distribution. One also has to consider the uncertainty in ligand bound TAR structures, both due to experimental imprecision and because the ligand bound states may be flexible themselves. Previous NMR studies reveal significant mobility in the TAR bulge especially residues C24 and U25 when in complex with argininamide(43,56) and Mg^{2+}(52,57). The ACP bound TAR structure also exhibits a large degree of interhelical motions(53). Notwithstanding these complications, our results suggest that local dynamics in and around the bulge together with global motions of helices can drive much of the conformational adaptation required to bind to different ligand targets.

Our results also suggest that some of the TAR conformational changes require ligand binding to occur efficiently. For example, we do not observe the key U38-A27•U23 base-triple in either of the TAR or HIV-2 TAR SAS conformers that is known to form in the TAR-argininamide complex(46). In this regard, it is interesting to note that larger deviations between the SAS and ligand bound conformations are generally observed for the weaker binding ligands and *vice versa*. It is possible that the weaker binding ligands expend a greater fraction of the binding energy changing the TAR conformation. Additional studies are needed to shed light on these key energetic questions.

This work has been published in the journals Nucleic Acids Research (58) and Methods (59). The idea was conceived by Stelzer A. C. and Frank A. T. under the guidance Al-Hashimi H. M. and Andricoaei I. MD simulations were generated and SAS program written and implemented by Frank A. T. Frank A. T. analyzed the A-form nature of the SAS structures. Stelzer A. C. analyzed the Euler angles and SAS ensemble in comparison to previously determined structures.

2.9 References

1. Furtig, B., Buck, J., Manoharan, V., Bermel, W., Jaschke, A., Wenter, P., Pitsch, S. and Schwalbe, H. (2007) Time-resolved NMR studies of RNA folding. *Biopolymers*, 86, 360-383.
2. Henzler-Wildman, K. and Kern, D. (2007) Dynamic personalities of proteins. *Nature*, 450, 964-972.
3. Mittermaier, A. and Kay, L.E. (2006) New tools provide new insights in NMR studies of protein dynamics. *Science*, 312, 224-228.
4. Mackerell, A.D., Jr. and Nilsson, L. (2008) Molecular dynamics simulations of nucleic acid-protein complexes. *Curr Opin Struct Biol*, 18, 194-199.
5. Orozco, M., Noy, A. and Perez, A. (2008) Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr Opin Struct Biol*, 18, 185-193.
6. Duchardt, E., Nilsson, L. and Schleucher, J. (2008) Cytosine ribose flexibility in DNA: a combined NMR 13C spin relaxation and molecular dynamics simulation study. *Nucleic Acids Res*, 36, 4211-4219.
7. Ferner, J., Villa, A., Duchardt, E., Widjajakusuma, E., Wohnert, J., Stock, G. and Schwalbe, H. (2008) NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops. *Nucleic Acids Res*, 36, 1928-1940.
8. Hall, K.B. (2008) RNA in motion. *Curr Opin Chem Biol*.
9. Showalter, S.A. and Hall, K.B. (2005) Isotropic reorientational eigenmode dynamics complements NMR relaxation measurements for RNA. *Methods Enzymol*, 394, 465-480.
10. Trantirek, L., Caha, E., Kaderavek, P. and Fiala, R. (2007) NMR (13)C-relaxation study of base and sugar dynamics in GCAA RNA hairpin tetraloop. *J Biomol Struct Dyn*, 25, 243-252.
11. Chen, Y., Campbell, S.L. and Dokholyan, N.V. (2007) Deciphering protein dynamics from NMR data using explicit structure sampling and selection. *Biophys J*, 93, 2300-2306.
12. Lindorff-Larsen, K., Best, R.B., Depristo, M.A., Dobson, C.M. and Vendruscolo, M. (2005) Simultaneous determination of protein structure and dynamics. *Nature*, 433, 128-132.
13. Markwick, P.R., Bouvignies, G. and Blackledge, M. (2007) Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy. *J Am Chem Soc*, 129, 4724-4730.
14. Showalter, S.A. and Bruschweiler, R. (2007) Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. *J Am Chem Soc*, 129, 4158-4159.
15. Musselman, C., Al-Hashimi, H.M. and Andricioaei, I. (2007) iRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. *Biophys J*, 93, 411-422.
16. Zhang, Q. and Al-Hashimi, H.M. (2008) Extending the NMR spatial resolution limit for RNA by motional couplings. *Nat Methods*, 5, 243-245.
17. Zhang, Q., Stelzer, A.C., Fisher, C.K. and Al-Hashimi, H.M. (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature*, 450, 1263-1267.
18. Zhang, Q., Sun, X., Watt, E.D. and Al-Hashimi, H.M. (2006) Resolving the motional modes that code for RNA adaptation. *Science*, 311, 653-656.
19. Lipari, G. and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *Journal of the American Chemical Society*, 104, 4546-4559.

20. Tjandra, N. and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, 278, 1111-1114.
21. Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information For Structure Determination in Solution. *Proc. Natl. Acad. Sci. U. S. A.*, 92, 9279-9283.
22. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) CHARMM - A PROGRAM FOR MACROMOLECULAR ENERGY, MINIMIZATION, AND DYNAMICS CALCULATIONS. *Journal of Computational Chemistry*, 4, 187-217.
23. MacKerell, A.D., Banavali, N. and Foloppe, N. (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56, 257-265.
24. Aboul-ela, F., Karn, J. and 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, 3974-3981.
25. Brodsky, A.S. and Williamson, J.R. (1997) Solution structure of the HIV-2 TAR-argininamide complex. *J. Mol. Biol.*, 267, 624-639.
26. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics*, 79, 926-935.
27. Brooks III, C.L. and Karplus, M. (1983) Deformable stochastic boundary in molecular dynamics. *Journal of Chemical Physics*, 79, 6312.
28. Hoover, W.G. (1985) Canonical Dynamics - Equilibrium Phase-Space Distributions. *Physical Review A*, 31, 1695-1697.
29. Nose, S. (1984) A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods. *Journal of Chemical Physics*, 81, 511-519.
30. Auffinger, P. and Westhof, E. (1996) H-bond stability in the tRNA(Asp) anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophys J*, 71, 940-954.
31. Caves, L.S.D., Evanseck, J.D. and Karplus, M. (1998) Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science*, 7, 649-666.
32. Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.*, 138, 334-342.
33. Saupe, A. (1968) Recent results in the field of liquid crystals. *Angew. Chem., Int. Ed. Engl.*, 7, 97-112.
34. Zhang, Q., Throolin, R., Pitt, S.W., Serganov, A. and Al-Hashimi, H.M. (2003) Probing motions between equivalent RNA domains using magnetic field induced residual dipolar couplings: accounting for correlations between motions and alignment. *J Am Chem Soc*, 125, 10530-10531.
35. Dethoff, E.A., Hansen, A.L., Musselman, C., Watt, E.D., Andricioaei, I. and Al-Hashimi, H.M. (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, 3906-3915.
36. Lavery, R. and Sklenar, H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn*, 6, 63-91.
37. Lavery, R. and Sklenar, H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J Biomol Struct Dyn*, 6, 655-667.

38. Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*, 31, 5108-5121.
39. Bailor, M.H., Musselman, C., Hansen, A.L., Gulati, K., Patel, D.J. and Al-Hashimi, H.M. (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat Protoc*, 2, 1536-1546.
40. Musselman, C., Pitt, S.W., Gulati, K., Foster, L.L., Andricioaei, I. and Al-Hashimi, H.M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J Biomol NMR*, 36, 235-249.
41. Al-Hashimi, H.M., Gosser, Y., Gorin, A., Hu, W., Majumdar, A. and Patel, D.J. (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, 95-102.
42. Hansen, A.L. and Al-Hashimi, H.M. (2007) Dynamics of large elongated RNA by NMR carbon relaxation. *J Am Chem Soc*, 129, 16072-16082.
43. Pitt, S.W., Majumdar, A., Serganov, A., Patel, D.J. and Al-Hashimi, H.M. (2004) Argininamide binding arrests global motions in HIV-1 TAR RNA: comparison with Mg2+-induced conformational stabilization. *J Mol Biol*, 338, 7-16.
44. Aboul-ela, F., Karn, J. and 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, 313-332.
45. Long, K.S. and Crothers, D.M. (1999) Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. *Biochemistry*, 38, 10059-10069.
46. Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D. and Williamson, J.R. (1992) Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science*, 257, 76-80.
47. Ippolito, J.A. and Steitz, T.A. (1998) A 1.3-angstrom 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.
48. Davis, B., Afshar, M., Varani, G., Murchie, A.I., Karn, J., Lentzen, G., Drysdale, M., Bower, J., Potter, A.J., Starkey, I.D. *et al.* (2004) Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *J Mol Biol*, 336, 343-356.
49. Du, Z., Lind, K.E. and James, T.L. (2002) Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. *Chem Biol*, 9, 707-712.
50. Faber, C., Sticht, H., Schweimer, K. and Rosch, P. (2000) Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J Biol Chem*, 275, 20660-20666.
51. Murchie, A.I.H., Davis, B., Isel, C., Afshar, M., Drysdale, M.J., Bower, J., Potter, A.J., Starkey, I.D., Swarbrick, T.M., Mirza, S. *et al.* (2004) Structure-based drug design targeting an inactive RNA conformation: Exploiting the flexibility of HIV-1 TAR RNA. *J. Mol. Biol.*, 336, 625-638.
52. Al-Hashimi, H.M., Pitt, S.W., Majumdar, A., Xu, W. and Patel, D.J. (2003) Mg2+-induced variations in the conformation and dynamics of HIV-1 TAR RNA probed using NMR residual dipolar couplings. *J Mol Biol*, 329, 867-873.
53. Pitt, S.W., Zhang, Q., Patel, D.J. and Al-Hashimi, H.M. (2005) Evidence that electrostatic interactions dictate the ligand-induced arrest of RNA global flexibility. *Angew Chem Int Ed Engl*, 44, 3412-3415.

54. Fisher, C.K., Zhang, Q., Stelzer, A. and Al-Hashimi, H.M. (2008) Ultrahigh Resolution Characterization of Domain Motions and Correlations by Multialignment and Multireference Residual Dipolar Coupling NMR. *Journal of Physical Chemistry B*, 112, 16815-16822.
55. Meier, S., Blackledge, M. and Grzesiek, S. (2008) Conformational distributions of unfolded polypeptides from novel NMR techniques. *J Chem Phys*, 128, 052204.
56. Dayie, K.T., Brodsky, A.S. and Williamson, J.R. (2002) Base Flexibility in HIV-2 TAR RNA Mapped by Solution (15)N, (13)C NMR Relaxation. *J Mol Biol*, 317, 263-278.
57. Casiano-Negroni, A., Sun, X. and Al-Hashimi, H.M. (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, 6525-6535.
58. Frank, A.T., Stelzer, A.C., Al-Hashimi, H.M. and 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, 3670-3679.
59. Stelzer, A.C., Frank, A.T., Bailor, M.H., Andricioaei, I. and Al-Hashimi, H.M. (2009) Constructing Atomic-Resolution RNA Structural Ensembles Using MD and Motionally Decoupled NMR RDCs. *Methods*, 49, 167-173.

Chapter 3

A Conservative Sequence Mutation in HIV-1 TAR Pre-tunes the Free-State Dynamics Towards the Argininamide-Bound Conformation

3.1 Introduction

Previous studies comparing the dynamics of HIV-1 TAR (TAR) and HIV-2 TAR show that the internal motions are strongly dependent on the RNA secondary structure(1). TAR, which contains a trinucleotide bulge, exhibits greater interhelical motions and distinct local dynamics in and around the bulge in comparison to HIV-2 TAR, which contains a dinucleotide bulge. In this chapter, we investigate whether changing the TAR sequence without significantly perturbing the secondary structure leads to changes in dynamical behavior. In particular, the local flexibility at the pivotal A22-U40 base-pair below the bulge, which does not form a typical Watson-Crick base-pair, may be important for activating collective helical motions and also affect local motions within the bulge. Here, the dynamic properties of a mutant TAR construct, in which the labile A22-U40 base pair is replaced with a G22-C40 base-pair (ΔGCTAR), are investigated. We test the hypothesis that the stronger GC hydrogen bonds leads to base-pairing and attendant changes in the local and global helical dynamics. Note that the GC mutant is not strictly a mutant since it is observed in the HIV subtype-O isolate(2,3). The effects of perturbing the free-state dynamical properties on ligand binding are also investigated by titrating two known TAR-binding small molecules,

argininamide (ARG), which stabilizes a linear TAR structure, and neomycin B, which stabilizes a bent TAR conformation, on TAR and ΔGCTAR.

3.2 Materials and Methods

3.2.1 Sample Preparation and Assignments

Uniformly $^{13}C/^{15}N$ labeled RNAs were prepared by run-off *in-vitro* transcription using synthetic double-stranded DNAs containing the T7 promoter and RNA sequence of interest (*Integrated DNA Technologies*). Elongated RNA constructs were purified by 15% (w/v) denaturing polyacrylamide gel electrophoresis containing 8M urea and 1x TBE followed by electroeultion in 20 mM Tris pH 8 buffer and EtOH precipitation. The resultant RNA pellet was dissolved and exchanged into NMR buffer (15 mM sodium phosphate, 0.1 mM EDTA, and 25 mM NaCl at pH ~6.4) using a Centricon Ultracel YM-3 concentrator to a final concentration of ~0.5mM (Millipore Corp.). All NMR samples contained 10% D_2O. The ΔGCTAR NMR spectra were assigned using conventional NMR methods such as 3D exchangeable $^{1}H-^{15}N$ NOESY-HSQC, 2D non-exchangeable $^{1}H-^{13}C$ NOESY-HSQC, 2D HCN, 2D IP-COSY (correlated H5/H6 resonances) and spectral overlays with similar RNA constructs.

3.2.2 Normalized Intensity and Chemical Shift Analysis

Resonance intensities are normalized to a baseline value of 0.1 by dividing all resolved resonance intensities by 10x the lowest resonance intensity for a similar residue and spin type with known Watson-Crick A-form geometry. For example, the C8H8 of adenines were all normalized to A20 and the C6H6 uridine resonances to U42.

Weighted average chemical shifts were calculated using the equation,

$$\Delta\delta_{av} = \sqrt{\left(\Delta\delta_H\right)^2 + \left(\frac{\gamma_C}{\gamma_H}\left(\Delta\delta_C\right)\right)^2}\ . \qquad [3.1]$$

3.2.3 RDC Measurements and Order Tensor Analysis

One bond RDCs for ΔGCTAR were measured by computing the difference in splittings measured in NMR buffer (J) and after suspending the RNA sample in ~7mg/ml of pf1 bacteriophage (J+D) as described in(4). Because we implement the elongation strategy(1, 2) to decouple internal and overall motions, TROSY detection was employed in the pulse sequence to enhance sensitivity(5). CH splittings in the nucleobases (C2H2/C8H8 of adenosine and guanosine bases, C5H5/C6H6 of uridine and cytosine, and C1'H1' of the ribose) were measured as the difference between the upfield and downfield components of the ^{1}H –^{13}C doublet along the ^{1}H dimension using the narrow TROSY component in the ^{13}C dimension. NH splittings in the nucleobases (N1H1 and N3H3 of the guanine and uracil bases, respectively) were measured using a coupled HSQC experiment.

The relative orientation and dynamic amplitudes for the helices in ΔGCTAR were determined using order tensor analysis as reported previously(6-8). RDCs measured for each elongated construct were combined by normalizing adenine and uridine RDCs by a coefficient L that minimizes the quality factor Q(1,9). Order tensors describing the partial alignment of each helix relative to the magnetic field were determined by fitting measured RDCs to idealized A-form helices(6,8). Due to deviations from Watson-Crick geometry, terminal base-pairs were excluded from the analysis. The program AFORM-

RDC was used to estimate errors in the order tensor arising from “structural noise” and RDC measurement uncertainty(8).

3.2.4 Calculation of Dissociation Constants

Dissociation constants were calculated from the change in weighted average chemical shift for each titration point using the equation(10),

$$\delta_{obs} = \delta_{free} + (\Delta\delta_T) \times \left(\left([Arg]_T + [RNA]_T + K_d \right) - \frac{\sqrt{\left([Arg]_T + [RNA]_T + K_d^2 \right) + \left(4[Arg]_T [RNA]_T \right)}}{2[RNA]_T} \right), \quad [3.2]$$

where $[Arg]_T$ is the total argininamide concentration, $[RNA]_T$ is the RNA concentration based on UV absorbance at 260 nm, $\Delta\delta_T$ is the difference in chemical shifts between the free and ligand-associated states (in ppm), δ_{obs} is the observed chemical shift (in ppm), and δ_{Free} is the chemical shift in the free state (in ppm). The data was fit using the Origin software (OriginLab Corporation) in which $\Delta\delta_T$ and K_d were allowed to float during the fit.

3.3 Impact of the G22-C40 Mutation on TAR Structural Dynamics

Previous studies have shown that while the A22-U40 base-pair exists in an A-form geometry, there is significant flexibility and no direct evidence for Watson-Crick hydrogen bonding(11,12). In particular, the U40 imino resonance has only been observed in HIV-2 TAR at low temperatures (13) or for TAR in the ARG bound state(TAR-ARG)(11,14). In stark contrast, we were able to directly observe the G22 resonance in ΔGCTAR, which was assigned using a 3D ^{1}H-^{15}N NOESY-HSQC experiment. The observation of this imino proton resonance together with sequential

NOEs between G21 and G22 strongly supports Watson-Crick base-pairing for G22-C40 in ΔGCTAR (Figure 3.1).

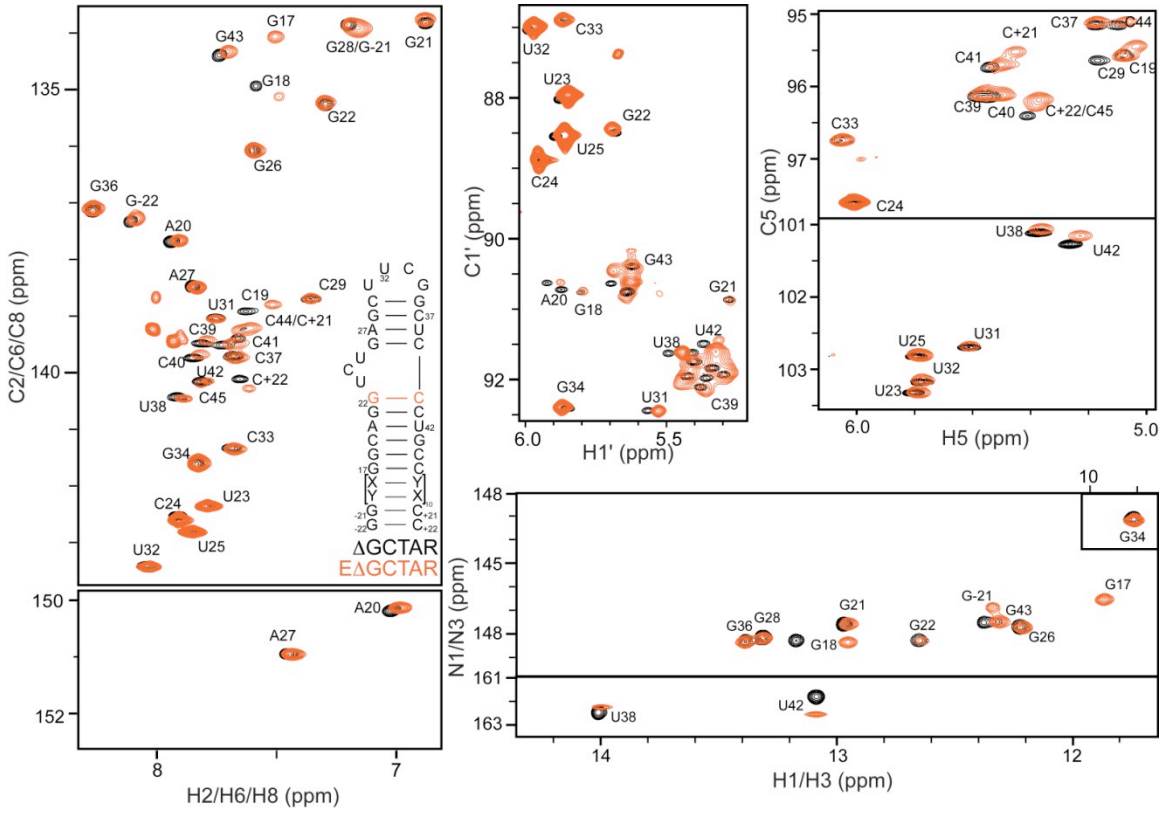


Figure 3.1. Overlay of short and elongated ΔGCTAR. The ΔGCTAR construct is shown in the left-most panel, where ΔGCTAR contains residues 17 to 45 and EΔGCTAR contains residues -22 to +22 where X-Ys refer to sequential AU or GC base pairs used in the elongation strategy as previously described(1,15). Similar resonance positions for the both RNAs indicate that the elongation does not cause structural perturbations.

3.3.1 Domain-Elongation Does Not Impact the Integrity of ΔGCTAR

One of the most difficult problems to address when analyzing dynamic motions of biomolecules is the coupling between internal and overall motions. Al-Hashimi H. M. and co-workers showed that fast-timescale (ps-ns) local and collective motions in TAR are masked by the diffusion timescale dependence of internal and interhelical motions, which remain unresolved in transverse and longitudinal relaxation rates(15). To overcome this problem, they elongated helix-1 to slow overall diffusion thus resurrect previously un-detectable local and collective interhelical motions. More recently we

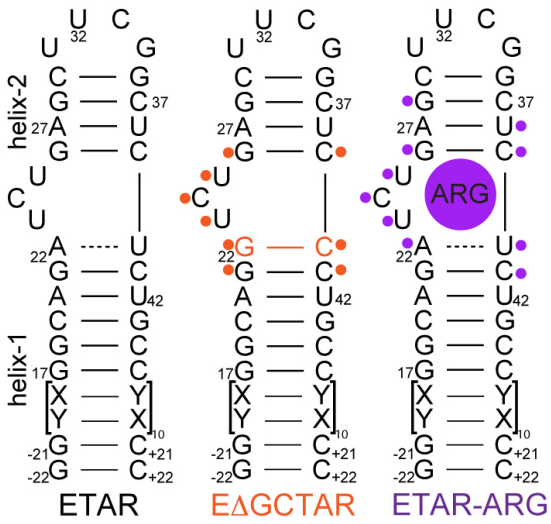


Figure 3.2. ETAR constructs. Residues exhibiting significant changes in weighted average chemical shift are indicated with a filled circle and the sequential X-Ys refer to sequential AU or GC base pairs used in the elongation strategy as previously described(1,15).

showed that the elongation technique can also be used in RDC measurements(1). Interhelical motions in TAR affect the overall alignment of the molecule in the presence of alignment media preventing accurate definition of the order tensor for each helix. Elongating either helix leads to well-defined alignment along the axial direction of the elongated helix and affords accurate calculation of the order tensor. Thus, the dynamics of TAR are more precisely described. In this study, helix-1 of ΔGCTAR was elongated (EΔGCTAR) by 22 base-pairs as previously described(1,2). Aside from resonances belonging to terminal residues the close agreement between the short and elongated ΔGCTAR spectra indicate no significant structural perturbations result from the elongation (Figure 3.1).

3.3.2 Analysis of Chemical Shift Perturbations

To analyze the effects of mutating A22-U40 to G22-C40, we compared 2D CH HSQC spectra of EΔGCTAR and ETAR and computed chemical shift perturbations (CSPs) that reflect changes in chemical shift due to introduction of the mutation. Interestingly, the largest CSPs were observed for residues that have been shown to be critical for ARG binding (Figure 3.2)(16-18). A large downfield shift is observed for the U23 C6H6 resonance in EΔGCTAR along both the ^{1}H and ^{13}C dimension(Figure 3.3).

This is consistent with the looping out of the U23 base as observed in the Mg^{2+} bound state (19). Large CSPs are observed for the G26 C8H8 resonance, which moves towards a position similar to that observed for HIV-2 TAR and TAR bound to Mg^{2+} (Figure 3.3) suggesting increased coaxial stacking. Likewise, large CSPs are observed for C6H6 resonances of C24 and U25 in ΔGCTAR that move toward positions that closely agree with those observed for ETAR-ARG (Figure 3.3)(18,20). Indeed, with the exception of U23, all of the CSPs observed in EΔGCTAR are along directions that are similar to those induced by ARG binding and some fall at a position that is a midway between free and ARG-bound ETAR, such as the C5H5 resonance of U25. Thus, the GC mutation seems to have the same effects on the conformation of TAR as does ARG binding and to a lesser extent Mg^{2+} binding (Figure 3.3).

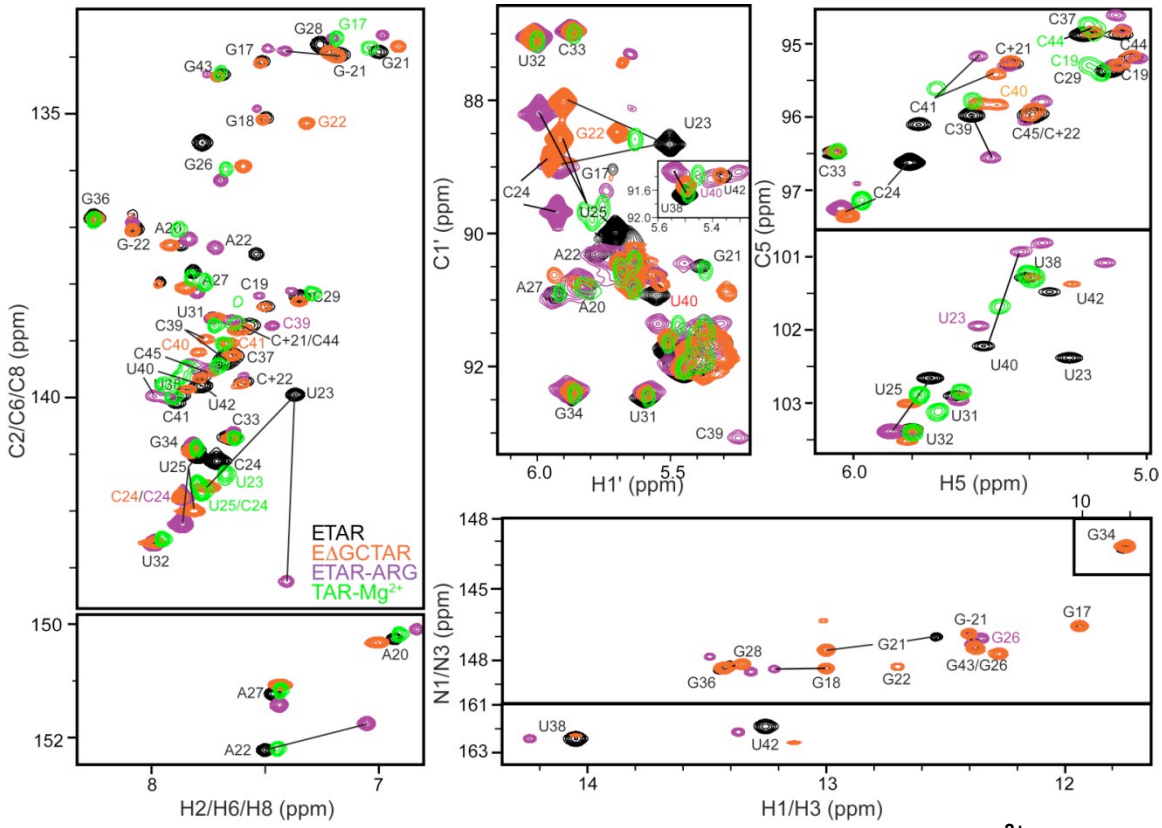


Figure 3.3. 2D CH HSQC ETAR, EΔGCTAR, ETAR-ARG, and TAR-Mg^{2+} spectra. EΔGCTAR, ETAR-ARG, and TAR-Mg^{2+} resonances that do not closely overlay are shown in orange, purple, and green respectively.

3.3.3 Sub-nanosecond Motions from Motional Narrowing of Resonances and Domain-Elongation

Comparison of the normalized resonance intensities observed for ETAR, EΔGCTAR, and ETAR-ARG provide additional insights into the effects on fast sub-nanosecond internal motions. In unbound ETAR, we observe elevated intensities for helix-II compared to helix-I due to collective motions and elevated intensities for residues A22, U40, C24 and U25 that reflect local motions. Binding of ARG arrests collective motions and local motions involving A22, U40 but leads to increased dynamics of C24 and U25, which are looped out in the ARG bound state. Remarkably, the corresponding

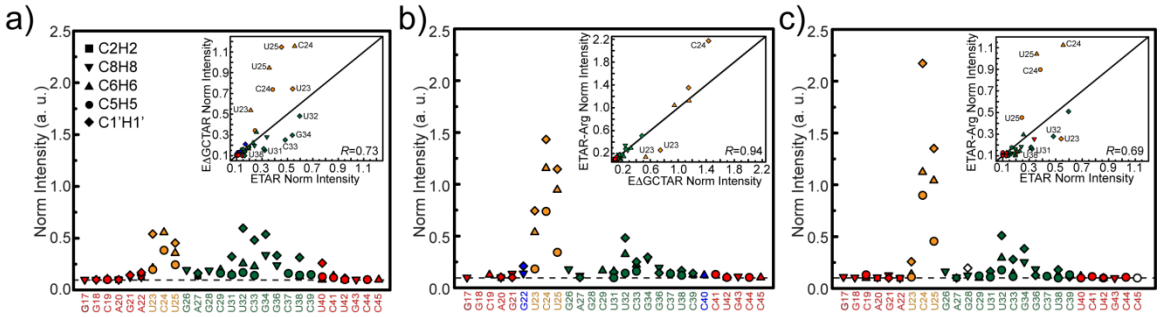


Figure 3.4. Normalized resonance intensities. Normalized resonance intensities for **(a)** ETAR **(b)** EΔGCTAR and **(c)** ETAR-ARG show similarities between ΔGCTAR and TAR-ARG. Helix-1 is colored red, the trinucleotide bulge orange, and helix-2 green. Open symbols are resonances that could not be measured in all three RNAs. Inset correlation plots show a greater correlation between EΔGCTAR and ETAR-ARG normalized intensities.

resonance intensities observed in the EΔGCTAR mutant are far more similar to those of ETAR-ARG (R=0.94) compared to free ETAR (R=0.73). This is in excellent agreement with the above CSP results. The most significant differences between the resonance intensities include residues that are involved in the U23-A27-U38 base-triple, which forms in TAR upon binding to ARG. For example, much higher intensities are observed for U23 C6H6 and C1'H1' resonances in the EΔGCTAR mutant compared to ETAR-ARG. The high intensity and local mobility observed for U23 suggests that the U23-A27-U38 base-triple is not formed in EΔGCTAR as observed for ETAR-ARG. In general,

higher intensities are observed for bulge resonances C24 and U25 in ETAR-ARG and likely reflect more complete coaxial stacking of the helices and extrusion of the nucleobases. Interestingly, much lower intensities are observed for U38 C1'H1' in EΔGCTAR and ETAR-ARG compared to ETAR (Figure 3.4). This likely reflects increased co-axial stacking and possibly transient and/or incomplete formation of the base-triple. Taken together, the CSP and resonance intensity data strongly suggest that that the GC mutation has the same effect on the TAR conformation as does ARG binding (Figure 3.3 and 3.4).

3.3.4 Analysis of ^{15}N Relaxation Data

In order to quantitatively analyze of the global and local dynamic effects of the G22-C40 mutation, we measured imino ^{15}N longitudinal (R1) and transverse (R2) rates for all resolved ^{15}NH resonances in EΔGCTAR and compared findings with previous results reported for ETAR and ETAR-ARG(15). Shown in Figure 3.5 are the R2/R1 values, which are insensitive to fast local motions occurring at the picosecond timescale and report on slower motions occurring at nanosecond timescales including collective motions of helical domains(15). Here, elongation of helix-1 is critical for decoupling collective helix motions from overall rotational diffusion(15).

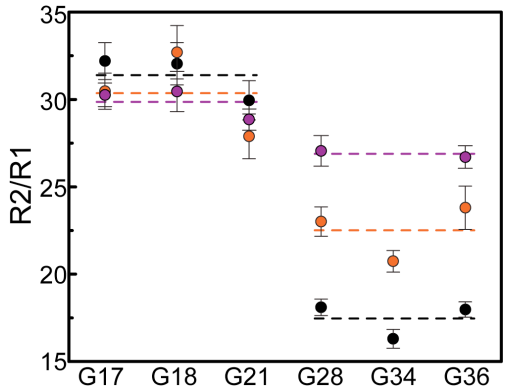


Figure 3.5. ^{15}N Relaxation. ^{15}N R2/R1 relaxation data obtained for ETAR (black), EΔGCTAR (orange), and ETAR-ARG (purple). The Average R2/R1 values for each helix are indicated by dotted lines.

The R2/R1 values observed in ETAR are in very good agreement with hydrodynamic predictions(15). The similar values observed across different residues is consistent with having similar local motions

at nanosecond timescales. The R2/R1 values measured in helix-2 are uniformly attenuated relative to values in helix-1 indicating that helix-2 residues experience an additional internal dynamic process at nanosecond timescales. The uniformity of the attenuation across different residues is consistent with having a collective dynamic process that leads to similar levels of reorientation for distinct residues. The addition of ARG results in near total arrest of these collective motions and the helix-2 R2/R1 values become more similar to the helix-1 counterparts. The small changes observed in the helix-1 R2/R1 values likely reflect changes in the overall structure of ETAR-ARG and thus changes in its overall rotational diffusion. Interestingly, while the R2/R1 values observed for helix-2 in EΔGCTAR are attenuated relative to values in helix-1, the level of attenuation is diminished by a factor of 2.5 compared to ETAR-ARG. Thus, the collective helical motions are significantly reduced in EΔGCTAR compared to ETAR but not to the extent on ARG binding. This again most likely originates from formation of the G22-C40 base-pair and stronger preference for co-axial stacking of the two helices. The ^{15}N relaxation data therefore supports the notion that the GC mutation leads to a perturbation towards a TAR conformation similar to that observed in the ARG-bound state.

3.3.5 Structural Dynamics at Sub-millisecond Timescales from Residual Dipolar Couplings

We further characterized the impact of the GC mutation by measuring RDCs in EΔGCTAR . RDCs probe sub-millisecond motions and thus provide a larger window to investigate the G22-C40 mutation effects. In ETAR we observe a systematic reduction in helix-2 RDCs suggesting more dynamics compared to helix-1(Figure 3.6). The bulge RDCs exhibit a large range of RDCs indicating dynamic variability, which is not unexpected since U23 is known to be looped inside the bulge and stacked on A22(11).

On binding ARG, the average RDCs for both helices are nearly identical suggesting attenuated interhelical dynamics. As expected, on binding ARG the U23 RDCs are significantly increased due to the formation of the U23-A27-U38 base-triple, and the C24 and U25 RDCs are closer to zero indicative of being looped out. Remarkably, EΔGCTAR exhibits similar features as ETAR-ARG where the average RDCs for both helices are

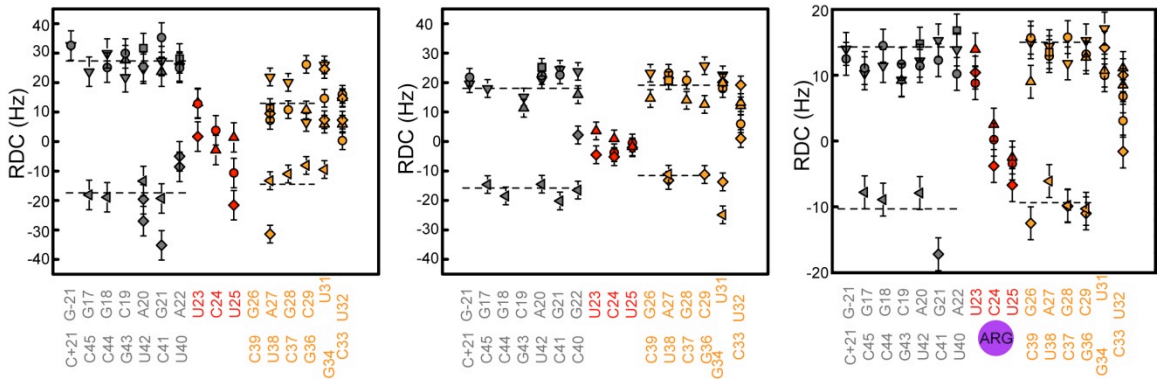


Figure 3.6. RDCs measured for ETAR, EΔGCTAR, and ETAR-ARG. RDCs measured in EΔGCTAR and ETAR-ARG indicate attenuated interhelical dynamics and similar bulge dynamics compared to ETAR.

nearly identical. Interestingly, except for U23 in ETAR-ARG, the bulge RDCs for EΔGCTAR and ETAR-ARG all have magnitudes near zero indicative of highly isotropic motions likely resulting from the looped out bulge conformation (Figure 3.6). The larger RDC magnitudes for U23 in ETAR-ARG are due to the formation of the U23-A27-U38 base triple. In the case of EΔGCTAR, the G22 C1'H1' RDC is 2.2 Hz. While it can not be ruled out that the orientation of the G22 C1'H1' bond vector gives rise to the small RDC value, it is in agreement with the increased A22 C1'H1' resonance intensity in the more coaxially stacked HIV-2 TAR RNA (1) potentially indicating unexpected increases in local dynamics on forming a more coaxially stacked state. Unfortunately, the A22 C1'H1' RDC could not be measured for ETAR-ARG due to spectral overlap and line broadening on addition of pf1 phage precluding further comparison.

To further characterize the structural dynamics of EΔGCTAR, we subjected the RDCs measured in A-form helices to an order tensor analysis. Assuming an idealized A-form helical geometry, the RDCs are used to determine five parameters of an order tensor describing the overall alignment relative to the magnetic field. By superimposing the order tensor frames obtained for each helix, we were able to compute an average interhelical bend angle. In addition, by considering the generalized degree of order (GDO) for each helix, which describes dynamic averaging resulting from overall alignment and internal motions independent of the average orientation of a molecule fragment and the magnetic field, the internal GDO (GDO_{int}) is computed to separate contributions from overall alignment and internal motions resulting in a measure of interhelical dynamics, where a GDO_{int} of 1 corresponds to interhelical rigidity and a GDO_{int} of 0 reflects isotropic interhelical motions.

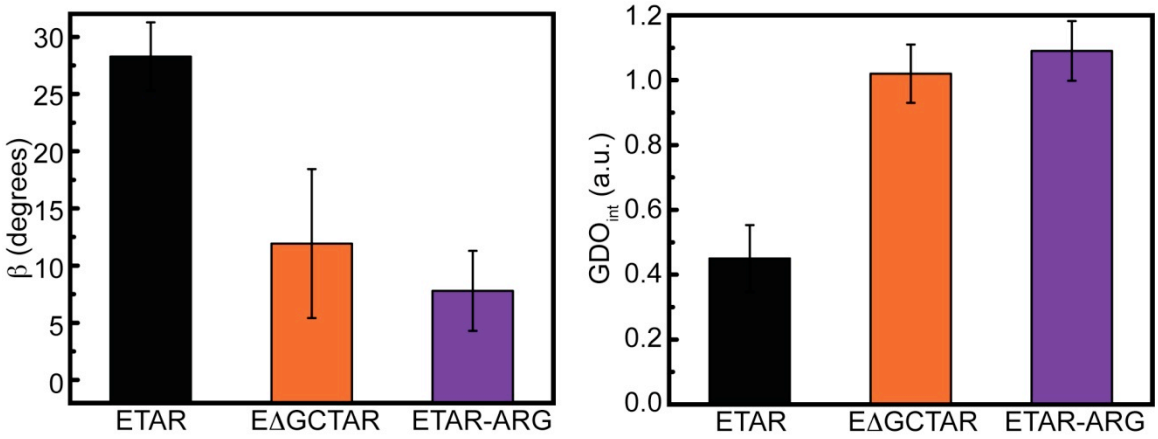


Figure 3.7. Order tensor analysis derived parameters describing the structure and dynamics of ETAR, EΔGCTAR, and ETAR-ARG. The interhelical bend angle (β) describes the angle between the principal axis of alignment and helix-2, and the GDO_{int} describes the interhelical motional amplitude where 1 corresponds to complete interhelical rigidity and 0 to fully isotropic motions.

Previous order tensor analysis of RDCs measured in ETAR and ETAR-ARG showed that ETAR exhibits significant interhelical motions (GDO_{int}=0.45±0.10) with an average interhelical bend angle (β) of 28.3±3.0°. These motions are significantly reduced

on binging ARG resulting in a GDO_{int} of 1.09±0.09 and β of 7.8±3.5° consistent with a coaxially stacked and rigid conformation. SVD order tensor analysis was used to probe the interhelical dynamics of EΔGCTAR. The measured and back-calculated RDCs closely correlate for all three RNAs as indicated by the small RMSD values and Q factors reported in Table 3.1. Decreased interhelical bend angle (β=11.9±6.5°) and increased GDO_{int} (1.02±0.09) (Figure 3.7 and Table 3.1) indicate that the G22-C40 mutation significantly reduces interhelical motions. In fact, the β and GDO_{int} values are in much closer agreement with ETAR-ARG compared to ETAR (Figure 3.7).

Table 3.1. Structural parameters derived from SVD order tensor analysis of ETAR, EΔGCTAR, and ETAR-ARG. N is the number of RDCs included in the analysis for each helix, RMSD is the root mean square deviation between measured and back-calculated RDCs, CN# is the condition number, Q is the quality factor between the measured and back-calculated RDCs, η describes the asymmetry of the order tensor (where 0 indicates a fully symmetric tensor), R is the correlation between measured and back-calculated RDCs after the order tensor analysis, GDO, is the generalized degree of order for each order tensor, GDO_{int} describes the magnitude of interhelical motions with 1 describing interhelical rigidity and 0 fully isotropic motions, and β° is the interhelical bend angle, Full descriptions of these parameters can be found in reference (6).

Structure	Helix	N	RMSD (Hz)	CN#	Q (%)	*R*	η	GDO(x10^{-3})	GDO_{int}	β°
ETAR	I	23	4.0	3.43	10.9	0.99	0.08	1.83±0.05	0.45±0.10	28.3±3.0
	II	21	2.5	3.26	15.2	0.99	0.41	0.83±0.09		
ETAR-ARG	I	22	2.8	2.99	18.8	0.99	0.11	0.75±0.05	1.09±0.09	7.8±3.5
	II	21	2.0	3.58	12.3	0.99	0.19	0.82±0.03		
EΔGCTAR	I	19	4.1	3.80	20.8	0.98	0.19	1.04±0.04	1.02±0.09	11.9±6.5
	II	12	3.2	4.93	14.5	0.98	0.16	1.06±0.08		

3.4 ARG Captures Common Conformations from the Distinct ΔGCTAR and TAR Dynamic Ensembles

Our results suggest that the GC mutation perturbs the TAR structural dynamics towards the ARG bound state. Thus, the ARG bound state is still, and perhaps even

more, dynamically accessible in ΔGCTAR. Given that A22 and U40 are not involved in any sequence specific interactions with ARG, we can expect that ΔGCTAR binds to ARG using similar binding modes. To investigate this hypothesis, we performed CSP NMR studies with ARG and ΔGCTAR. Titration of ARG led to significant CSPs for many residues that are know to interact with ARG including U23, G26, A27, U38, and C39(14,21). Remarkably, despite large differences between the starting spectra of ΔGCTAR and TAR, the resultant spectra for ΔGCTAR-ARG and TAR-ARG are in excellent agreement (Figure 3.8). These results strongly suggest that ARG captures similar coaxial conformations from the distinct ΔGCTAR and TAR dynamic ensembles.

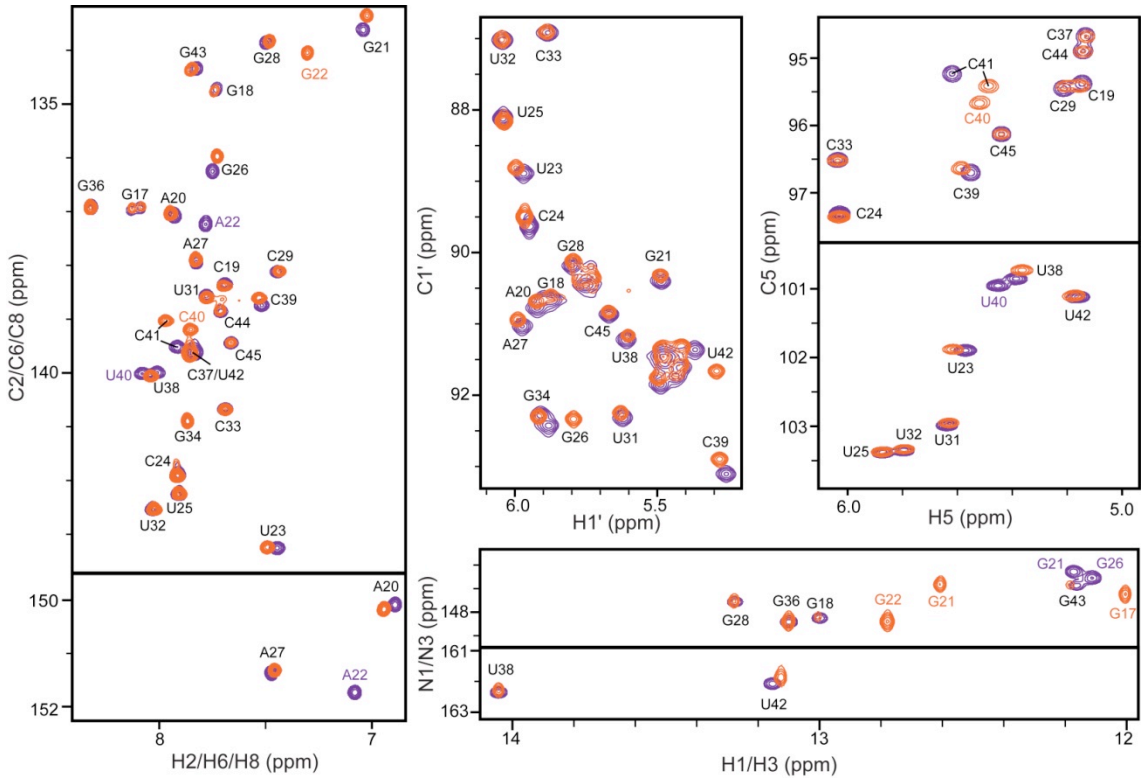


Figure 3.8. Comparison of TAR-ARG and ΔGCTAR-ARG spectra. Saturated TAR-ARG and ΔGCTAR-ARG spectra are overlaid. Resonances that do not closely overlay are shown in purple and orange for TAR and ΔGCTAR respectively. The largest differences between chemical shifts are observed for residues immediately neighboring the A22-U40 to G22-C40 mutation.

In order to assess the changes in binding affinity between ΔGCTAR and TAR-ARG, dissociation constants were independently calculated for residues exhibiting large ARG-induced CSPs in ΔGCTAR and TAR (Figures 3.9). Comparison of the binding affinities for the residues A22/G22, U23, and G28 show a 2-4 fold tighter ARG binding for ΔGCTAR compared to TAR. Fitting the data globally to determine an overall dissociation constant results in ~3 fold increase in binding affinity for ΔGCTAR (47.7±9.3 μM) compared to TAR (140±9.0 μM). It is very likely that the GC mutation increases the ARG binding affinity by preferentially stabilizing the ARG bound conformation; specifically coaxial stacking of the two helices and looping out of U23. The binding affinity of ARG to TAR matches that calculated by Gdaniec Z. and co-workers who report a 300 ±100 μM K_d using fluorine NMR (22).

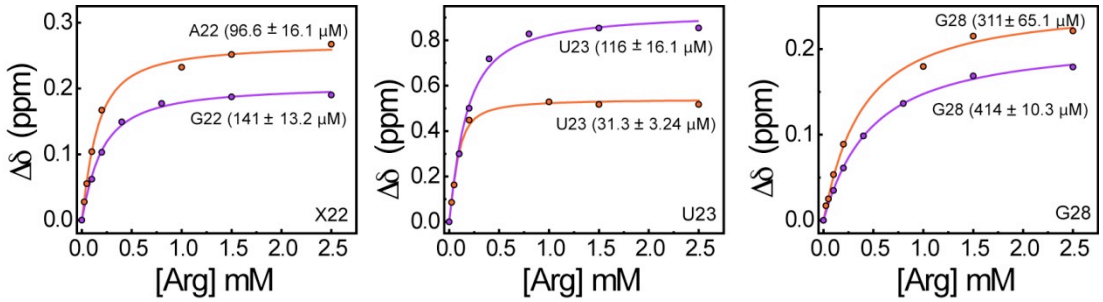


Figure 3.9. Enhanced binding affinity for ΔGCTAR-ARG. Dissociation constants for similar resolved resonances that shift greater than the weighted average chemical shift are shown, indicating that the modest A22-U40 to G22-C40 mutation results in a ~2-4 fold increase in binding affinity.

3.5 Neomycin B Stabilizes Conformations in TAR that are Inaccessible in ΔGCTAR

As a second test the known TAR-binding small molecule neomycin B was independently titrated onto ΔGCTAR and TAR. Neomycin B has been shown to stabilize a bent TAR conformation that is flexible in the bound state(23). Surprisingly, neomycin B does not appear to bind ΔGCTAR (Figure 3.10). In fact, many resonances follow linear CSP trajectories at a 2:1 neomycin B:ΔGCTAR ratio and then begin to shift in other directions on further addition of neomycin B (Figure 3.10). Also, the CSPs for ΔGCTAR

compared to TAR are not in the same directions as was observed in the ARG titrations indicating different binding modes. These data most likely indicates non-specific binding and precludes the calculation of dissociation constants. Thus, it is likely that the G22-C40 mutation does not allow access to the bent conformation necessary to bind neomycin B.

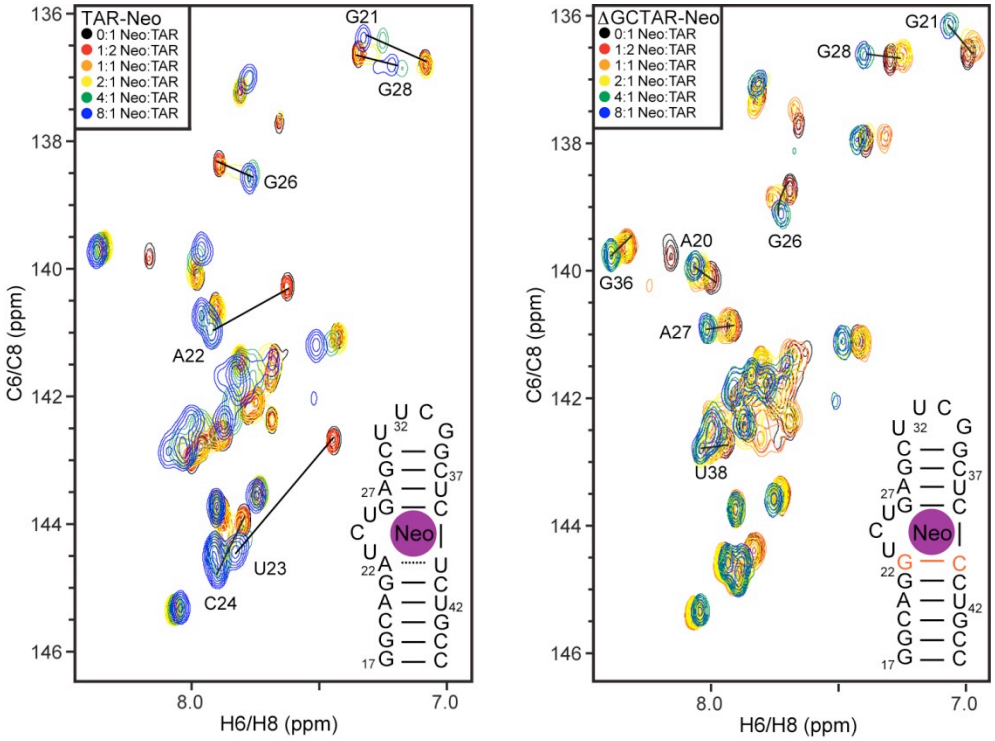


Figure 3.10. Neomycin B does not specifically bind ΔGCTAR. 2D C6/C8 HSQC spectra from titration of neomycin B onto TAR and ΔGCTAR are shown. Initially linear changes in chemical shift are observed for ΔGCTAR, but deviate at >2:1 neomycin B:ΔGCTAR ratios. Residues that shift significantly on titration of neomycin B are labeled.

3.6 Conclusion

This study highlights that small sequence mutations can lead to dramatic changes in the average structure and dynamics. Reduction in the average R2/R1 values, decreased β, and increased GDO_{int} for EΔGCTAR indicate that stabilizing the labile A22-U40 base-pair leads to a drastic reduction in collective interhelical motions that closely resemble ETAR-ARG. Sequence effects are also observed for local motions where hydrogen bonding of the G22-C40 base-pair results in significant reduction of the G22-

C40 resonance intensities that are accompanied by formation of a U23 looped out conformation giving rise to the increased U23 resonance intensities and reduced RDCs. Thus, modest sequence mutations can lead to dramatic perturbations of functionally important junctions that result in very different local and global motions. The results presented here also suggest that the dynamic perturbations observed for ΔGCTAR pre-tunes the dynamic equilibrium toward a coaxially stacked state. This is analogous to a previous study showing that on binding Mg^{2+} the dynamic equilibrium of TAR is biased toward the coaxially stacked state and the population of bent TAR conformations is significantly reduced(19). On saturation with argininamide, TAR and ΔGCTAR form the same bound-state conformation as indicated by nearly identical chemical shifts, however, on titration with neomycin B very different chemical shifts are observed for TAR and ΔGCTAR suggesting different binding modes. Thus, biasing the conformational ensemble toward the coaxial state increases the affinity for argininamide, which stabilizes a linear TAR conformation, but does not allow access to the bent state necessary to bind neomycin B.

Due to the ever-increasing evidence that RNA is critical in regulating cellular processes making it a highly sought drug target, many studies have aimed to characterize the physical basis for RNA structural plasticity. One question that remains unanswered is the effects of sequence mutations on dynamics and intermolecular interactions. Using TAR as a model system to interrogate the effects of sequence mutations, we show that by modestly changing one base-pair, the free-state dynamics are modified from a flexible and bent state toward a more coaxially-stacked and rigid conformation. This leads to an ~3 fold increase in binding affinity for ARG, which could indicate that transition to the bound state involves first pre-forming the coaxially aligned and looped out bulge conformation prior to binding ARG. Overall, these results show that conservative RNA sequence mutations can have dramatic effects on the free-state

dynamics and structurally bias RNA-ligand interactions, which is critically important in many biological processes.

The idea was conceived by Al-Hashimi H. M. and Stelzer A. C. Stelzer A. C. and Kratz J. D. synthesized RNA oligonucleotides and collected and analyzed NMR data.

3.7 References

1. Zhang, Q., Stelzer, A.C., Fisher, C.K. and Al-Hashimi, H.M. (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature*, 450, 1263-U1214.
2. Berkhout, B. (1992) Structural Features in Tar Rna of Human and Simian Immunodeficiency Viruses - a Phylogenetic Analysis. *Nucleic Acids Research*, 20, 27-31.
3. Chakrabarti, L.A., Luckay, A. and Marx, P.A. (2001) A divergent simian immunodeficiency virus from sooty mangabey with an atypical Tat-TAR structure. *Aids Research and Human Retroviruses*, 17, 1155-1165.
4. Getz, M., Sun, X.Y., Casiano-Negroni, A., Zhang, Q. and Al-Hashimi, H.M. (2007) NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers*, 86, 384-402.
5. Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1997) Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 12366-12371.
6. Bailor, M.H., Musselman, C., Hansen, A.L., Gulati, K., Patel, D.J. and Al-Hashimi, H.M. (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nature Protocols*, 2, 1536-1546.
7. Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of Magnetic Resonance*, 138, 334-342.
8. Musselman, C., Pitt, S.W., Gulati, K., Foster, L.L., Andricioaei, I. and Al-Hashimi, H.M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *Journal of Biomolecular Nmr*, 36, 235-249.
9. Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *Journal of the American Chemical Society*, 120, 6836-6837.
10. Gonzalez, R.L. and Tinoco, I. (1999) Solution structure and thermodynamics of a divalent metal ion binding site in an RNA pseudoknot. *Journal of Molecular Biology*, 289, 1267-1282.
11. AboulEla, F., Karn, J. and Varani, G. (1996) Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Research*, 24, 3974-3981.
12. Al-Hashimi, H.M., Gosser, Y., Gorin, A., Hu, W.D., Majumdar, A. and Patel, D.J. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *Journal of Molecular Biology*, 315, 95-102.
13. Long, K.S. and Crothers, D.M. (1999) Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. *Biochemistry*, 38, 10059-10069.
14. Aboulela, F., Karn, J. and Varani, G. (1995) The Structure of the Human-Immunodeficiency-Virus-Type 1 Reveals Principles of RNA recognition By Tat Protein. *Journal of Molecular Biology*, 253, 313-332.
15. Zhang, Q., Sun, X.Y., Watt, E.D. and Al-Hashimi, H.M. (2006) Resolving the motional modes that code for RNA adaptation. *Science*, 311, 653-656.

16. Hamy, F., Asseline, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G., Butler, P.J.G., Karn, J. and Gait, M.J. (1993) Hydrogen-bonding Contacts in the Major Groove are Required for Human-Immuno-Deficiency-Virys Type 1 Tat Protein Recognition of TAR RNA. *Journal of Molecular Biology*, 230, 111-123.
17. Pritchard, C.E., Grasby, J.A., Hamy, F., Zacharek, A.M., Singh, M., Karn, J. and Gait, M.J. (1994) Methyl Phosphonate Mapping of Phosphate Contacts Critical for RNA Recognition by the Human Immuno-Deficiency-Visrus Tat and Rev Proteins. *Nucleic Acids Research*, 22, 2592-2600.
18. Pitt, S.W., Majumdar, A., Serganov, A., Patel, D.J. and Al-Hashimi, H.M. (2004) Argininamide binding arrests global motions in HIV-1 TAR RNA: Comparison with Mg2+-induced conformational stabilization. *Journal of Molecular Biology*, 338, 7-16.
19. Casiano-Negroni, A., Sun, X.Y. and Al-Hashimi, H.M. (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, 6525-6535.
20. Dayie, K.T., Brodsky, A.S. and Williamson, J.R. (2002) Base flexibility in HIV-2 TAR RNA mapped by solution N-15,C-13 NMR relaxation. *Journal of Molecular Biology*, 317, 263-278.
21. Puglisi, J.D., Tan, R.Y., Calnan, B.J., Frankel, A.D. and Williamson, J.R. (1992) Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy *Science*, 257, 76-80.
22. Olejniczak, M., Gdaniec, Z., Fischer, A., Grabarkiewicz, T., Bielecki, L. and Adamiak, R.W. (2002) The bulge region of HIV-1 TAR RNA binds metal ions in solution. *Nucleic Acids Research*, 30, 4241-4249.
23. Pitt, S.W., Zhang, Q., Patel, D.J. and Al-Hashimi, H.M. (2005) Evidence that electrostatic interactions dictate the ligand-induced arrest of RNA global flexibility. *Angewandte Chemie-International Edition*, 44, 3412-3415.

Chapter 4

Dynamics-Based Screening Affords Efficient Discovery of Novel HIV-1 TAR-Binding Small Molecules

4.1 Introduction

Structure-based drug design is one of the most ambitious approaches for drug discovery and is based on rationally using knowledge of a molecule's 3D structure to identify small molecule binders. The flexible nature of RNA presents a unique challenge in this regard because even related ligands can bind very different RNA conformations that cannot, in general, be determined *a priori*. Characterizing the large conformational rearrangements that occur in RNA on ligand binding has been impeded by the lack of experimental techniques capable of reporting on motions at biologically important timescales. Thus, RNA structure-based predictions of RNA-ligand interactions have largely not proven successful. In Chapter 2 we showed that NMR and MD can be used in concert to characterize RNA dynamics up to millisecond timescales provided that internal and overall motions are decoupled. Using a Sample and Select (SAS) approach we were able to generate ensembles comprised of 20 HIV-1 TAR (TAR) conformers that precisely re-capitulate the experimental RDCs and suggest that bound-state TAR conformations are present in the absence of ligands(1). Here, we test the hypothesis that computational docking can be used to calculate a docking score, which reports on

the RNA-ligand interaction energy, that correlates with the experimentally determined change in binding free energy (ΔG_{exp}) provided an accurate description of the ligand-bound RNA conformation is known. Virtual screening simulations conducted using the ensemble of 20 SAS-generated TAR structures result in the identification of the small molecule netilmicin, which shows the ability to inhibit HIV-1 replication *in-vitro*.

4.2 Materials and Methods

4.2.1 ICM Computational Docking

ICM (Molsoft LLC) employs internal coordinate mechanics (based on bond angles, bond lengths, torsion angles, and phase angles) in grid-based docking simulations to efficiently determine interaction energies and optimal small molecule binding conformations in receptor-ligand complexes(2). Since the energy function is expressed analytically in internal coordinates, ICM uses a Monte Carlo gradient minimization algorithm during docking simulations, which has been shown to greatly increases the search area from which the global minimum can be found and also significantly reduce computational time(2). In order to rank the interactions of ligands with macromolecules, ICM implements a scoring function containing eight terms,

$$Score = \alpha_1 N_{at} + \alpha_2 N_{flex} + \alpha_3 \Delta E_{HB} + \alpha_4 \Delta E_{vdW} + \alpha_5 \Delta E_{Intl} + \alpha_6 \Delta E_{DeSolv} + \alpha_7 \Delta E_{SolEl} + \alpha_8 \Delta E_{Hphob}, \quad [4.1]$$

where N_{at} is a correction term to prevent biasing toward larger small molecules, N_{flex} represents the change in energy due to entropy loss of the ligand on binding, ΔE_{HB} is an angle and distance dependent hydrogen bonding term, ΔE_{VdW} characterizes Van der Waals interactions between the ligand and receptor accounting for the quality of steric fit, ΔE_{Intl} is the internal energy of the ligand calculated using the ECEPP/3 force field,

ΔE_{Desolv} takes into account the disruption of hydrogen bonds between the ligand and solvent, ΔE_{SolEl} is the solvation electrostatics energy on binding that is calculated using a distance dependent dielectric constant, and ΔE_{Hphob} reports on the hydrophobic free energy gain according to the amount of buried surface area upon binding. Each term is weighted using a coefficient, α_n, that affords optimization of the scoring function to agree with experimental data.

ICM docking simulations were conducted using rigid RNA receptors and fully flexible small molecule ligands. In all simulations the RNA binding pocket was pre-defined in order to efficiently target probable binding sites and minimize the size of the calculated grid maps. Small molecule-bound RNA structures used to validate the ability of ICM to accurately dock RNA targets were downloaded from the Protein Data Bank (PDB) (www.pdb.org). The small molecule was separated from the RNA, the RNA was converted into an ICM object, and the RNA binding pocket was defined as all atoms within 5 Å of the bound small molecule. For structures with multiple binding sites, each site was defined as a receptor, independently docked, and the lowest score recorded as the final docking result. Docking simulations against TAR employed individual SAS-ensemble structures. These RNA structures were first converted into ICM objects and probable binding pockets were predicted using the ICM PocketFinder module, which calculates the surface area and volume of cavities on the receptor surface. All binding pockets that fall within the "druggable" range (volume=150-500 $Å^3$ and surface area=200-550 $Å^2$), as determined by ICM(3), were set as receptor centroids and the final receptor defined as all atoms within 5 Å of the predicted binding site. In the event that multiple "druggable" binding pockets were identified for one structure, all RNA binding pockets were docked independently.

All small molecules used to dock against the RNA receptors were converted to SDF files. In the validation simulations the small molecule protonation states were assigned according to reported studies of the RNA-small molecule complex. In the event that the protonation state was not known, all protonation states for pH 5.4-9.4 were calculated using the majormicrospecies module of ChemAxon© (www.chemaxon.com) and docked against the RNA receptor. When multiple protonation states were found, each small molecule protonation state was docked and the minimum score recorded. Predicted small molecule protonation states were also used in the docking simulations against the SAS-ensemble structures.

4.2.2 Sample Preparation and Assignments

All RNAs (both $^{13}C/^{15}N$ and non-isotopically labeled) were prepared as described in section *3.2.1*. The same NMR data were collected as described in section *3.2.1* including an additional ^{13}C-edited-NOESY-HSQC experiment to help determine the resonance assignments of TAR bound to netilmicin. N-terminal-labeled-fluorescein Tat peptide (FTat), was purchased from Genscript Corp.

4.2.3 Weighted Average Chemical Shift Perturbations, RDC Measurements, and Order Tensor Analysis

Methods to calculate weighted average chemical shift perturbations are described in section *3.2.2*, and methods used to measure RDCs and implement order tensor analysis are described in section *3.2.3*.

4.2.4 Fluorescence Polarization

Fluorescence polarization (FP) measurements were collected using 384 well plates read with a PHERAstar Plus plate reader (BMG LABTECH) and a 485 nm

excitation wavelength and 520 nm detection wavelength optic module. All measurements were run in triplicate and averaged to give the final data. Elongated AU HIV-1 TAR (EAUTAR) was used to increase the dynamic range of the FP measurements. The FP buffer consisted of 50mM Tris (pH 7.4), 100mM NaCl, and 0.01% nonidet-P40. If soluble in aqueous solution, small molecules were dissolved in the FP buffer. Otherwise, they were initially suspended in DMSO followed by dilution into the FP buffer. Determination of the TAR-Tat K_d was performed by adding sequentially higher concentrations of TAR to 10nM Tat with a final volume of 32 μL. The K_d was fit using the equation,

$$P_{obs} = P_{free} + \left(P_{bound} - P_{free}\right) \times \left(\left([L]_T + [RNA] + K_d \right) - \frac{\sqrt{\left([L]_T + [RNA] + K_d \right)^2 + \left(4[L]_T [RNA] \right)}}{2[L]_T} \right) \quad , \qquad [4.2]$$

where P_{bound} and P_{free} are the measured fluorescence polarization values of the bound and unbound FTat peptide, respectively, $[L]_T$ is the total concentration of FTat, [RNA] is the EAU-TAR concentration, and K_d is the dissociation constant. The data was fit using the Origin software (OriginLab Corporation) in which P_{bound}, P_{free}, and K_d were allowed to float during the fit. The data was collected for 30 minutes at 5 minute intervals to ensure no change in the K_d over time. After 10 minutes no change in the K_d was found.

Small molecules were tested to inhibit the TAR-Tat interactions using 10 nM FTat, 60 nM EAU-TAR, and varying concentrations of small molecule in each well. The small molecules were first incubated with TAR for 10 minutes, followed by addition of FTat and a second 10 minute incubation period after which the FP was collected for 20 minutes at 5 minute intervals to ensure stability. IC_{50} values were calculated using the equation,

$$y = P_{free} + \frac{\left(P_{bound} - P_{free}\right)}{1 + \left([I]/IC_{50}\right)^{m}} \quad , \qquad [4.3]$$

where P_{bound} and P_{free} are the measured fluorescence polarization values of the RNA-bound and RNA-unbound FTat peptide, respectively, *[I]* is the log of the concentration of the small molecule inhibitor, IC_{50} is the 50% small molecule inhibitory concentration, and *m* is the slope of the linear portion of the sigmoidal curve. The data was fit using the Prizm software (GraphPad Software Inc.). Inhibition constants (K_i) were calculated with the Prizm software (GraphPad Software Inc.) using the equation,

$$\log\left(IC_{50}\right) = \log\left(10^{\log(K_i)} \times \left(1 + \frac{[FTat]}{K_d}\right)\right), \qquad [4.4]$$

where *[FTat]* is the concentration of FTat and K_d is the dissociation constant of the FTat-TAR complex. The K_is reported in this manuscript are the exact K_is calculated from the IC_{50} and a range of K_is based on the 95% confidence interval. On identification of a FTAR-Tat inhibitor, the UV absorption spectrum was recorded to ensure no spectral overlap with fluorescein.

4.3 Validation of ICM to Computationally Dock RNA Targets

A database of 107 small molecule-bound RNA crystal and NMR structures, which is the largest set of small molecule-bound RNA structures interrogated by docking to date, was constructed by downloading previously determined RNA-small molecule complexes from the PDB (www.pdb.org). These structures included a number of bacterial/human A-site, riboswitch, ribozyme, and aptamer structures that provide a

representative example of diverse RNA-ligand interactions. 56 of the 107 structures had reported K_ds and were used to test the accuracy of the ICM scoring function to predict the experimental change in binding free energy (ΔG_{exp}=RTLn(K_d)), and all 107 structures were used to assess the accuracy of predicting the bound-state small molecule conformation. It should be noted that structures with metal-mediated RNA-small molecule interactions were omitted from testing the correlation between the docking score and ΔG_{exp} because ICM fails to accurately score these interactions.

All small molecule-bound RNA crystal and NMR structures were "self-docked". For each structure, the small molecule was removed and the protonation states assigned as described in section *4.2.1*. The small molecule was subsequently docked onto the RNA using the suggested "maximum" number of Monte-Carlo iterations according to the ICM manual (thoroughness=10), and the lowest scoring binding pose recorded. A larger number of Monte-Carlo iterations were tested without improvement in the scoring predictions (data not shown). For the A-site structures with two internal loop binding sites, each site was defined as a receptor, independently docked, and the lowest score was recorded as the final docking result. For the 56 structures tested, ICM predicts the binding scores relatively well resulting in a correlation of *R*=0.62 between the docking score and ΔG_{exp} (Figure 4.1a). The two outliers in Figure 4.1a (shown in open circles) are (i) the FMN aptamer, for which the solved RNA structure is different than that used to determine the K_d(4), and (ii) the 2,4,6-triaminopyrimidine bound purine riboswitch, which exhibits a substantially different K_d (>1 order of magnitude) in the reported study when using different biochemical methods(5). Interestingly, the agreement between docking score and ΔG_{exp} continuously improves as the small molecule flexibility decreases (defined by the number of flexible torsion angles; N_{flex}), most likely because conformational sampling and entropic contributions become less

problematic. Very good agreement (R = 0.73) is observed for 43 of the 56 small molecules that have N_{flex} <15 (Figure 4.1b), and 25 of the 56 small molecules with N_{flex}<10 (Figure 4.1c). Thus, even without further MD refinement of initial or final poses, the binding affinity can be predicted at a useful level of accuracy provided the bound RNA structure is well defined.

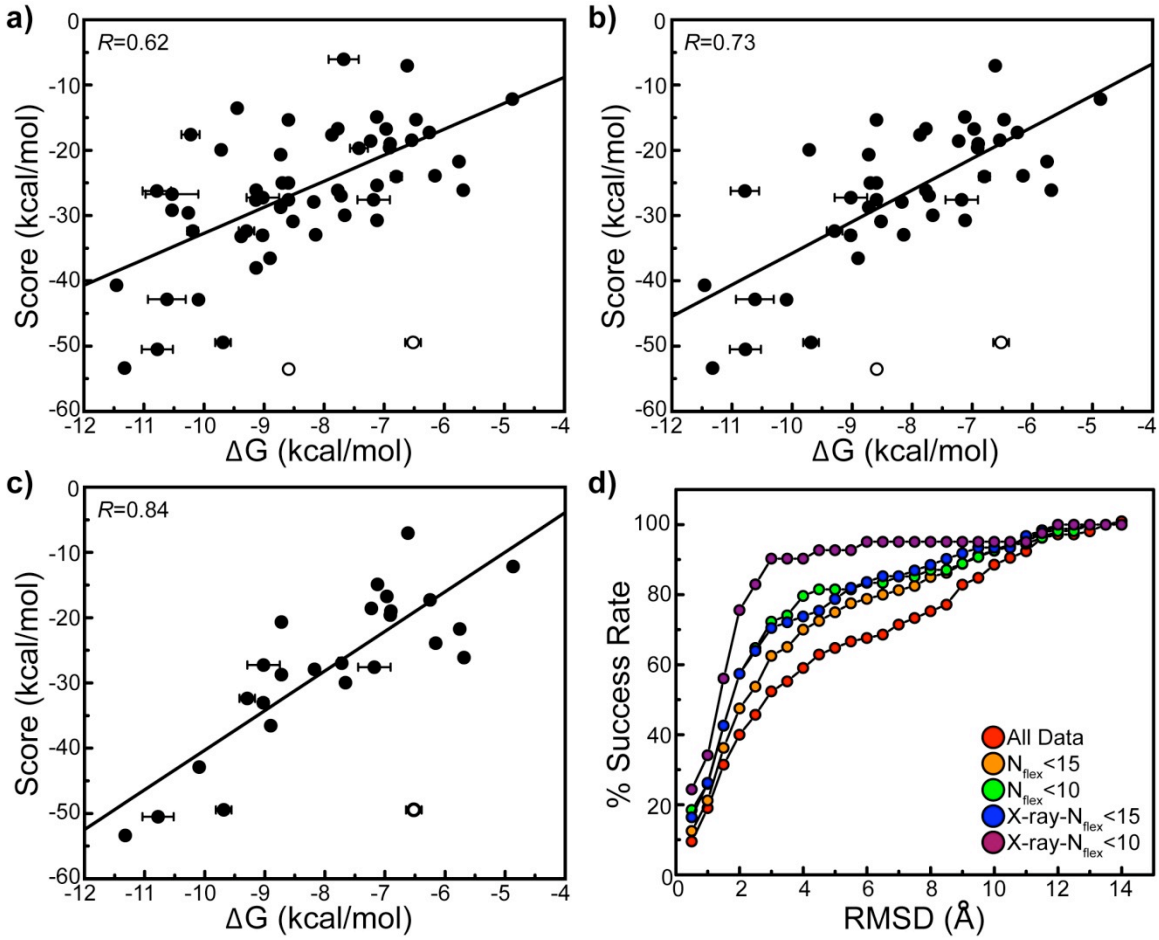


Figure 4.1. Crystal and NMR docking validation. Small molecules were "self-docked" to their respective crystal and NMR structures and correlated to ΔG_{exp}. The correlation increases with decreasing N_{flex}. All 56 structures are plotted in **(a)**, structures with N_{flex}<15 are plotted in **(b)** and N_{flex}<10 plotted in **(c)**. The success rate of predicting the bound small molecule conformation for all 107 RNA-small molecule structures is plotted in **(d)**. Error bars in **(a-c)** are obtained from experimental studies reporting the K_d or ΔG.

The success rate is a second metric used to determine the accuracy of ICM to dock RNA targets. The success rate is defined as the percentage of small molecules whose predicted bound conformation matches the crystal or NMR structure to within a

defined heavy-atom root mean square deviation (RMSD) cutoff. An RMSD threshold of 2.5 Å was used, which is the same cutoff reported in recent RNA-docking studies(6,7). The lowest scoring small molecule pose was used as the final docking result for comparison to the solved structure. We calculate a success rate of ~45% when considering all 107 structures. While this is disappointingly low, the success rate rapidly increases with decreasing N_{flex}, where the success rate is ~53% for N_{flex}<15 and ~65% for N_{flex}<10 (Figure 4.1d). Thus, it is likely that the increased success rate on decreasing N_{flex} is also a result of conformational sampling and entropic contributions.

Potential errors in the success rate could result from the dynamics of RNA-small molecule complexes(8). Often, small molecules bound to RNA are solvent-exposed and in some cases regions of the small molecules are not directly interacting with the RNA. Since it is likely that the regions of the small molecule not in contact with the RNA remain flexible in the bound state, which is not accurately represented in crystal or NMR structures, prediction of these small molecule conformations cannot be expected. The variability in NMR structure determination results in an average RMSD of 1.8Å between different models for RNA-ligand structures in the PDB and in some cases can exhibit RMSDs >3Å(7). Each member of an NMR ensemble represents a solution to the observed NMR and the dynamics of the RNA-small molecule complex will influence the variability in the ensemble structures. When considering only RNA-small molecule structures solved by X-ray crystallography, we calculate a dramatic increase in success rate, where ~65% of small molecules with N_{flex}<15 and ~83% with N_{flex}<10 are accurately predicted(Figure 4.1d). This is likely a result of more accurately defined RNA and ligand conformations. Also, it is likely that more strongly interacting complexes will exhibit less dynamics in the bound state. When only considering small molecules from NMR and X-ray structures with a K_d<1 μM, the success rates increase to ~60% for N_{flex}<15 and ~70% for N_{flex}<10. Thus, using a static representation of an otherwise flexible system

precludes accurate prediction of the bound state small molecule conformation. However, the flexibility of the bound state could explain why the docking scores correlate with ΔG_{exp} while the correct small molecule binding conformation is not always predicted.

Table 4.1. Small molecule-bound RNA structures used to validate ICM. A (**) indicates that the structure was used to correlate docking scores with ΔG_{exp}.

PDB	Receptor	Ligand	Res. (Å)	N_{flex}	RMSD (Å)	Score (kcal/mol)	
1F1T	Malachite green aptamer	Tetramethyl rosamine	2.80	1	0.98	-42.89	**
1F27	Biotin aptamer	Biotin	1.30	5	1.77	-14.91	**
1I9V	tRNA	Neomycin	2.60	17	11.30	-30.15	
1J7T	rRNA A-site	Paromomycin	2.50	17	3.16	-27.62	**
1LC4	rRNA A-site	Tobramycin	2.54	11	2.76	-16.69	**
1MWL	rRNA A-site	Genticin	2.40	12	10.95	-21.10	**
1NTA	Streptromycin aptamer	Streptomycin	2.90	16	4.43	-5.15	
1NTB	Streptromycin aptamer	Streptomycin	2.90	16	2.05	-29.21	
1O9M	rRNA A-site	Antibiotic	2.40	23	6.77	-6.05	**
1U8D	Guanine riboswitch	Hypoxanthine	1.95	0	0.16	-29.98	**
1Y26	Adenine riboswitch	Adenine	2.10	0	2.41	-36.56	**
1Y27	Guanine riboswitch	Hypoxanthine	2.40	0	0.20	-53.38	**
1YRJ	rRNA A-site	Apramycin	2.70	12	9.04	-26.12	**
1ZZ5	rRNA A-site	Restricted neomycin	3.00	10	4.64	-25.02	**
2A04	rRNA A-site	Neomycin	2.95	16	4.49	-29.19	**
2B57	Guanine riboswitch Mutant	2,6-diamino purine	2.15	0	0.19	-50.50	**

2BE0	rRNA A-site	JS5	2.63	22	11.01	-25.07	
2BEE	rRNA A-site	JS4	2.60	23	13.70	-30.97	
2CKY	TPP riboswitch	TPP	2.90	8	2.66	-24.44	
2EES	Guanine riboswitch Mutant	Hypoxanthine	1.75	0	0.12	-32.35	**
2EET	Guanine riboswitch Mutant	Hypoxanthine	1.95	0	0.20	-33.02	**
2EEU	Guanine riboswitch Mutant	Hypoxanthine	1.95	0	0.16	-23.91	**
2EEV	Guanine riboswitch Mutant	Hypoxanthine	1.95	0	0.18	-27.26	**
2EEW	Guanine riboswitch Mutant	Hypoxanthine	2.25	0	0.30	-27.57	**
2ESI	rRNA A-site	Kanamycin A	3.00	13	6.36	-15.29	**
2ESJ	rRNA A-site	Lividomycin A	2.20	22	9.98	-29.56	**
2ET3	rRNA A-site	Gentamicin C1A	2.80	9	0.33	-18.56	**
2ET4	rRNA A-site	Neomycin B	2.40	16	7.31	-26.71	**
2ET5	rRNA A-site	Ribostamycin	2.20	12	10.61	-18.48	**
2ET8	rRNA A-site	Neamine	2.50	7	11.23	-16.79	**
2F4S	rRNA A-site	Neamine	2.80	7	11.85	-17.26	**
2F4T	rRNA A-site	Antibiotic	3.00	6	2.35	-26.99	**
2F4U	rRNA A-site	Antibiotic	2.60	19	10.04	-19.67	**
2FCX	DIS	Neamine	2.00	7	1.37	-19.00	**
2FCY	DIS	Neomycin B	2.20	16	8.54	-32.40	**
2FCZ	DIS	Ribostamycin	2.01	12	1.70	-21.29	**
2FD0	DIS	Lividomycin A	1.80	22	8.23	-17.59	**

2G5K	rRNA A-site	Apramycin	2.80	12	7.03	-15.32	**
2G5Q	rRNA A-site	Amikacin	2.70	19	13.30	-14.39	
2G9C	Purine riboswitch	2,4,6-Triaminopyrimidine	1.70	0	0.22	-49.44	**
2GDI	TPP riboswitch	TPP	2.05	8	1.92	-15.84	
2GIS	SAM riboswitch	S-adenosylmethionine	2.90	10	1.24	-40.69	**
2HOJ	TPP riboswitch	TPP	2.50	8	1.52	-25.45	
2HOK	TPP riboswitch	TPP	3.20	8	1.76	-41.78	
2HOL	TPP riboswitch	TPP	2.90	8	1.79	-26.79	
2HOM	TPP riboswitch	TMP	2.89	6	1.61	-14.90	
2HOO	TPP riboswitch	Benfotiamine	3.00	11	5.04	-39.90	
2HOP	TPP riboswitch	Pyrithiamine	3.30	5	5.78	-19.60	**
2NZ4	GlmS ribozyme	GL6P	2.50	7	2.52	-0.98	
2O3V	rRNA A-site	Paromamine derivative	2.80	15	8.94	-19.03	
2O3W	rRNA A-site	Paromomycin	2.80	17	9.60	-23.99	
2O3X	rRNA A-site	Paromamine derivative	2.90	12	8.69	-20.18	
2OE5	RNA fragment	Apramycin	1.51	12	3.56	-27.58	**
2OE8	RNA fragment	Apramycin	1.80	12	4.64	-25.02	**
2PWT	rRNA A-site	LHABA aminoglycoside	1.80	28	9.82	-13.53	**
2QWY	SAM II riboswitch	S-adenosylmethionine	2.80	10	0.87	-32.96	**
3C44	DIS	Paromomycin	2.00	18	8.80	-39.84	
3D0U	Lysine riboswtich	Lysine	2.80	7	1.30	-34.17	

3D2G	TPP riboswitch	TPP	2.25	8	2.14	-16.22	
3D2V	TPP riboswitch	Pyrithiamine Pyrophosphate	2.00	8	4.30	-27.38	
3D2X	TPP riboswitch	Oxythiamine Pyrophosphate	2.50	8	2.60	3.10	
3DIG	Lysine riboswtich	S-(2-aminoethyl)-L-cystine	2.80	7	1.00	-55.72	
3DIL	Lysine riboswtich	Lysine	1.90	7	1.30	-48.54	
3DIM	Lysine riboswtich	Lysine	2.90	7	1.40	-41.29	
3DIO	Lysine riboswtich	Lysine	2.40	7	1.30	-48.03	
3DIQ	Lysine riboswtich	Homoarginine	2.70	7	0.80	-51.31	
3DIR	Lysine riboswtich	N6-1-iminoethyl-L-Lysine	2.90	7	1.60	-34.81	
3DIX	Lysine riboswtich	Lysine	2.90	7	1.30	-53.32	
3DIY	Lysine riboswtich	Lysine	2.71	7	1.30	-46.84	
3DIZ	Lysine riboswtich	Lysine	2.85	7	1.40	-54.19	
3DJ0	Lysine riboswtich	L-4-oxalysine	2.50	7	0.80	-43.80	
3DJ2	Lysine riboswtich	Lysine	2.50	7	1.70	-45.62	
3DVV	DIS	Ribostamycin(U267OMe)	2.00	12	2.30	-24.99	
3E5C	SAM III riboswitch	S-adenosylmethionine	2.25	10	3.00	-33.92	**
3E5E	SAM III riboswitch	S-adenosylmethionine	2.90	10	8.00	-27.25	
3E5F	SAM III riboswitch	S-adenosylmethionine	2.70	10	5.30	-25.48	
1FUF	Mismatched Duplex	Spermine	1.70	13	0.96	-39.95	
1YKV	Ribozyme	N-pentylmaleamide	3.30	6	0.67	-17.98	

1XPF	DIS	Spermine	2.30	13	7.91	-25.46	
2TOB	Aptamer	Tobramycin	NA	11	0.70	-42.85	**
1TOB	Aptamer	Tobramycin	NA	11	1.36	-26.23	**
1BYJ	rRNA A-site	Gentamicin C1A	NA	13	1.44	-30.75	**
1UTS	TAR	Rbt550	NA	8	9.03	-25.62	
1Q8N	Aptamer	Malachite green A	NA	1	7.52	-27.92	**
2AU4	Aptamer	GTP	NA	11	4.35	-19.93	**
1LVJ	TAR	Acetylpromazine	NA	4	6.83	-12.16	**
1NEM	Aptamer	Neomycin B	NA	17	0.54	-33.17	**
1PBR	rRNA A-site	Paromomycin	NA	22	0.74	-26.11	**
1EHT	Aptamer	Theophylline	NA	0	3.53	-28.71	**
1O15	Aptamer	Theophylline	NA	0	3.30	-20.65	**
1FMN	Aptamer	Flavin Mononucleotide	NA	11	2.61	-53.56	**
1EI2	Tau exon 10	Neomycin	NA	16	6.97	-18.06	
1UUD	TAR	Rbt203	NA	10	2.81	-45.79	
1AM0	Aptamer	AMP	NA	7	3.63	-47.76	
1QD3	TAR	Neomycin B	NA	22	13.38	-25.39	**
1FYP	rRNA A-site	Paromomycin	NA	22	1.31	-38.03	**
1KOC	Aptamer	Arginineamide	NA	8	9.74	-21.70	**
1KOD	Aptamer	Citrulline	NA	6	10.28	-26.09	**
1AJU	HIV-2 TAR	Arginineamide	NA	8	2.74	-25.26	

1RAW	Aptamer	AMP	NA	7	12.78	-7.04	**
1UUI	TAR	Rbt158	NA	6	8.89	-22.26	
1NBK	Aptamer	Arginineamide	NA	8	2.38	-22.98	
3C3Z	DIS Duplex	Ribostamycin	NA	12	-21.81	14.39	
3C7R	DIS Duplex	Lividomycin A	NA	22	-0.1829	0.8	
3C5D	DIS Duplex	Neomycin B	NA	16	-28.69	8.59	
1AKX	HIV-2 TAR	Arginineamide	NA	5	-29.4	3.56	
2JUK	HIV-1 Frameshift RNA	Guanidino-Neomycin B	NA	22	-30.15	11.2	

4.4 Docking to the TAR SAS Ensemble Recapitulates ΔG_{exp}

Results from the validation simulations suggest that ICM can be used to accurately calculate a score that correlates with ΔG_{exp} provided that the bound-state RNA conformation is known. Structures in the SAS-generated TAR ensemble were shown to recapitulate the experimental RDCs and closely resemble previously determined ligand-bound TAR conformations, suggesting that the bound-state TAR conformations are present in the absence of ligand (see Chapter 2). Thus, using the 20 SAS ensemble TAR structures as docking receptors should afford docking scores that re-capitulate ΔG_{exp} for known TAR-binding small molecules. To test this hypothesis, known 33 TAR-binding small molecules with reported K_ds and N_{flex}<15 were docked against the SAS ensemble structures. Remarkably, docking scores are calculated that correlate with ΔG_{exp} (R=0.69) (Figure 4.2a). In fact, the correlation coefficient is only slightly lower than the results of the solved RNA structure validation simulations. The error bars in Figure 4.2 are calculated from the average residual error in the crystal and NMR structural validation simulations for small molecules with N_{flex}<15. Within the

residual error, only two small molecules do not agree with the best-fit line. Interestingly, the best-fit lines are different for the validation and SAS ensemble simulations. This likely arises from the increased accuracy of the crystal and NMR structures, which provide details of local RNA rearrangements on small molecule binding that is not accounted for when docking small molecules to the SAS structures.

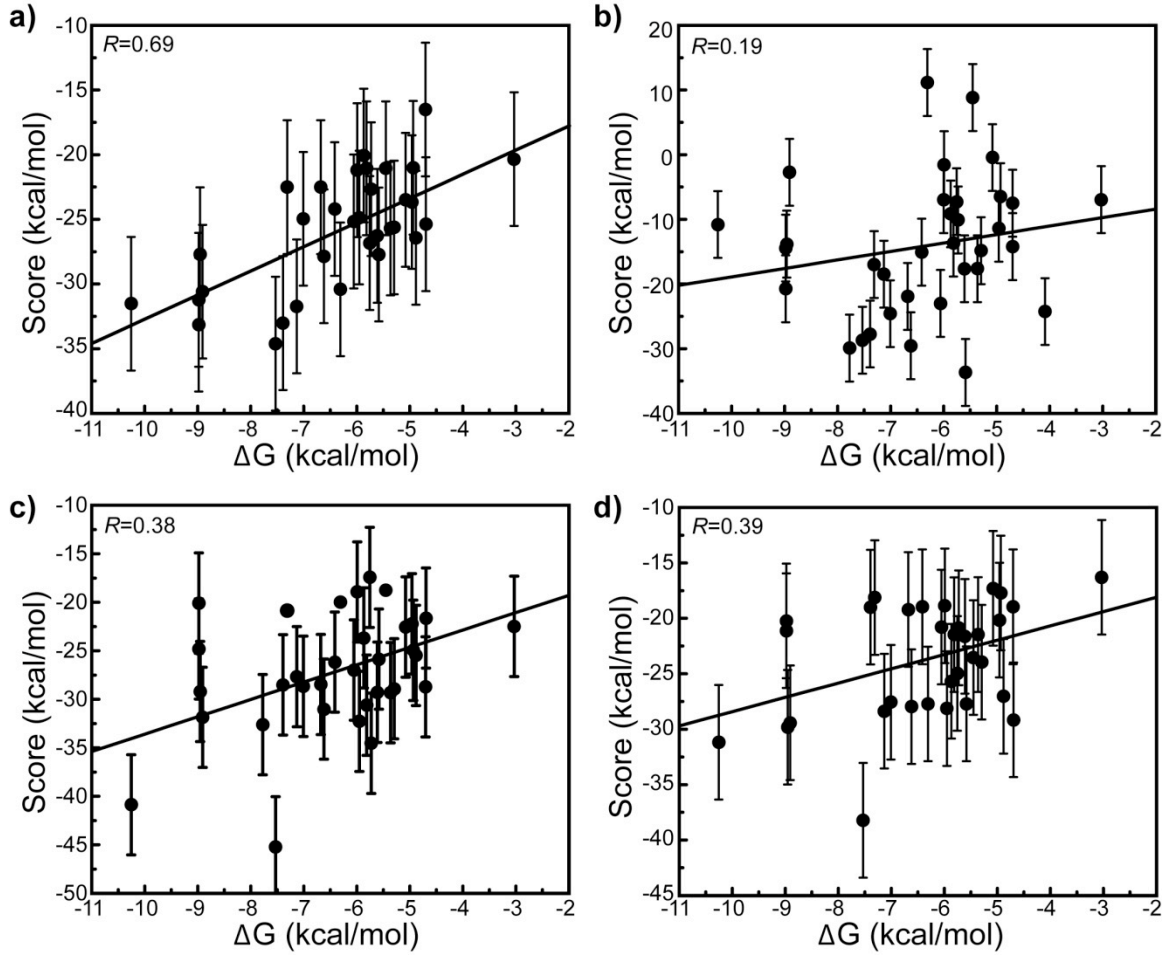


Figure 4.2. Docking known TAR binders to the SAS structures affords scores that correlate with ΔG_{exp}. (a) Docking scores of 33 known TAR binders with N_{flex}<15 onto the 20 SAS ensemble structures strongly correlate with ΔG_{exp}. Control simulations using **(b)** the apo-TAR crystal structure(9), **(c)** all 20 models of the apo-TAR structure solve by NMR(10), and **(d)** 20 random MD structures as docking receptors do not afford a strong correlation between docking score and ΔG_{exp}.

Control simulations were run using the Mg^{2+}-bound TAR structure (PDB ID# 397D)(9) (Figure 4.2b), the 20 ensemble structures from the apo-TAR structure determined using

NMR (PDB ID# 1ANR)(10) (Figure 4.2c), and 20 random snapshots from the MD simulation (Figure 4.2d). In all three cases the docking scores failed to correlate with ΔG_{exp}. The crystal structure fails to generate docking scores that correlate with ΔG_{exp} likely because TAR is highly dynamic and a static representation does not accurately represent the TAR structure. The apo-TAR structure solved by NMR and the random MD conformers afford an increased correlation between the docking score and ΔG_{exp} compared to the apo-TAR crystal structure, but the correlation coefficient is still much lower than when using the SAS-generated TAR structures. These results suggest that the longer timescale (sub ms) and angular sensitivity of RDCs afford a more robust ensemble compared to NOEs (ps-ns) used to determine the apo-TAR structure by NMR, and force-field errors present in the random MD conformers give rise to artificial conformations that are filtered using the SAS approach.

4.5 Virtual Screening Using TAR SAS Ensemble Structures

The SAS-generated TAR ensemble structures were virtually screened against two small molecule libraries: (i) ~60,000 small molecules used at the Center for Chemical Genomics (CCG) at the University of Michigan used in high-throughput screening assays and (ii) ~2500 small molecule library that was built in-house and represents published RNA-binding small molecules. The former collection includes various small molecules obtained from commercial suppliers such as Chembridge, ChemDiv, and Maybridge, and also includes collections from the National Cancer Institute as well as FDA approved drugs. The in-house library includes RNA-binding small molecule scaffolds including macrocyclic antibiotics, aminoglycosides, and various combinations of aromatic and charged moieties. Only small molecules with $N_{flex}<15$ were docked against the 20 SAS structures. Of the ~60,000 small molecules in the CCG library and ~2500 in the in-house library, 49,166 and 2045 small molecules have an

N_{flex}<15. Two rounds of docking were used with the initial round using less Monte-Carlo iterations (thoroughness=1) and the top 20% scoring small molecule re-docked with the maximum number of Monte-Carlo iterations (thoroughness=10). The lowest score for each small molecule was recorded.

Virtual screening results from both libraries yield Gaussian distributions. The distribution of scores for the CCG library gives an average of -17.3 kcal/mol and a standard deviation of 4.17 (Figure 4.3a). Interestingly results from docking the in-house library yields two Gaussian distributions, one giving an average of -19.5 kcal/mol and a standard deviation of 5.97, and the second an average of 4.47 kcal/mol and a standard deviation of 5.90 (Figure 4.3b). The latter of the two histograms calculated for the in-

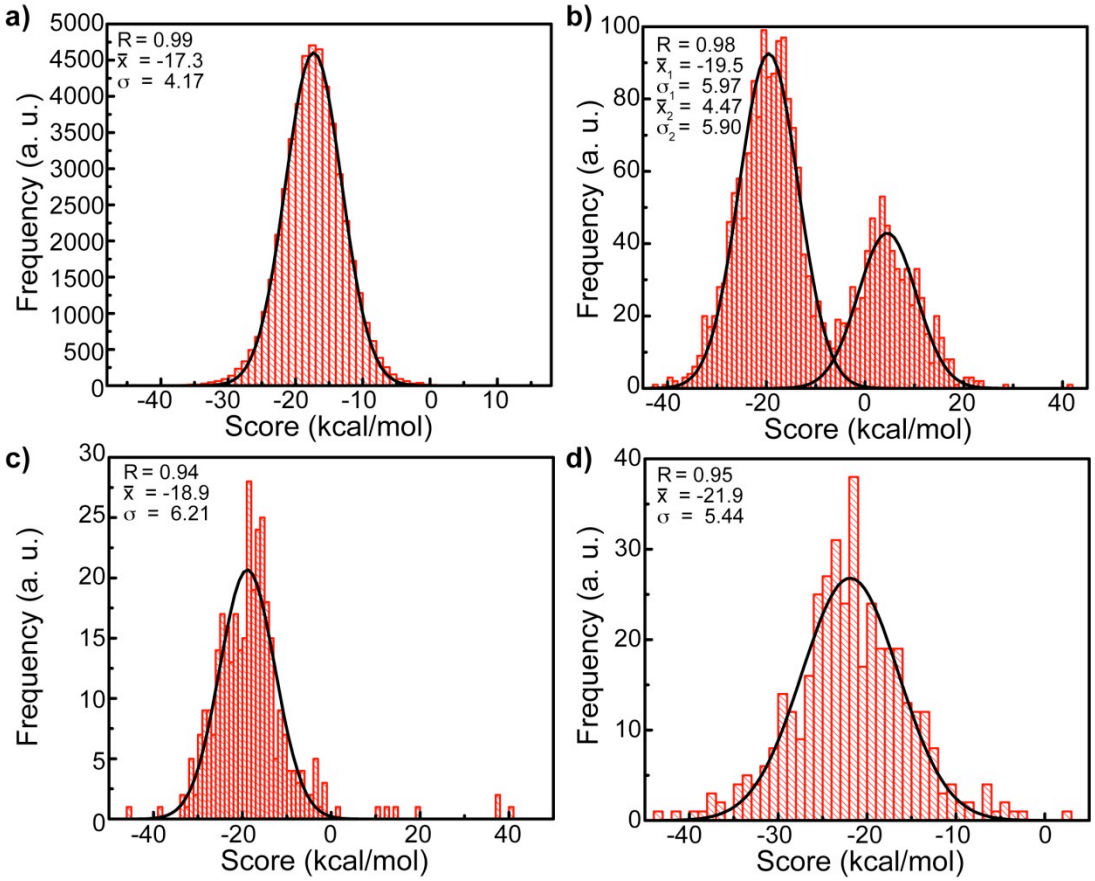


Figure 4.3. Minimum scores of the CCG and in-house small molecule libraries against the SAS ensemble structures. (a) and **(b)** are the histograms of the minimum scores for the CCG and in-house libraries for small molecules with N_{flex}<15, respectively. **(c)** and **(d)** Score distributions of small molecules with Log(P)<-1 for the CCG and in-house libraries, respectively.

house library is largely comprised of the erythromycin-like ribosome-binding small molecules that are too large to form favorable interactions with TAR. Small molecules with predicted Log(P)<-1, where P is the octanol:water ratio of small molecule concentration in the two phases of a water-octanol mixture, also yield Gaussian distributions with slightly more negative average scores (Figure 4.3c-d). The significance of the soluble distributions will be discussed below. A "good" score is defined here as <-20 kcal/mol.

Top scoring small molecules from the virtual screening simulations were validated to bind TAR using fluorescence polarization (FP) measurements and NMR chemical shift perturbations (CSPs). In the FP assay, a N-terminal-labeled-fluorescein Tat peptide (FTat), comprised of 9 residues of the Tat arginine rich motif(11), was bound to EAU-TAR and displaced by small molecules. To confirm that a small molecule inhibitor binds TAR and not the Tat peptide, CSPs were monitored on titration of the small molecule to TAR. Initially, the K_d was measured for the TAR-FTat binding (40.6±11.4 nM) (Figure 4.4), which agrees with previous studies using similar Tat peptides(12,13). Displacement of FTat from TAR was tested using known TAR-binding small molecules with nM-mM dissociation constants (K_d)(Figure 4.5).

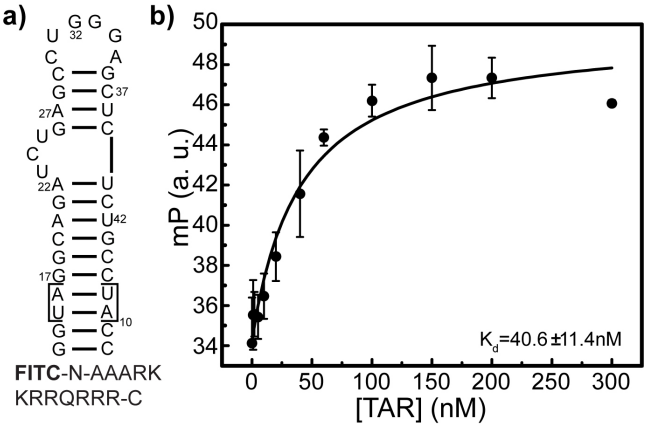


Figure 4.4. FTAR-Tat K_d determined by FP. (a) The EAU-TAR and N-labeled-fluorescein Tat peptides using the FP assay. **(b)** Binding isotherm for titrating EAU-TAR onto FTat.

Many small molecules from the CCG library can be purchased from commercial vendors, while most of the in-house library was derived from literature and thus could not be easily obtained. Only three small molecules from the in-house library were tested due

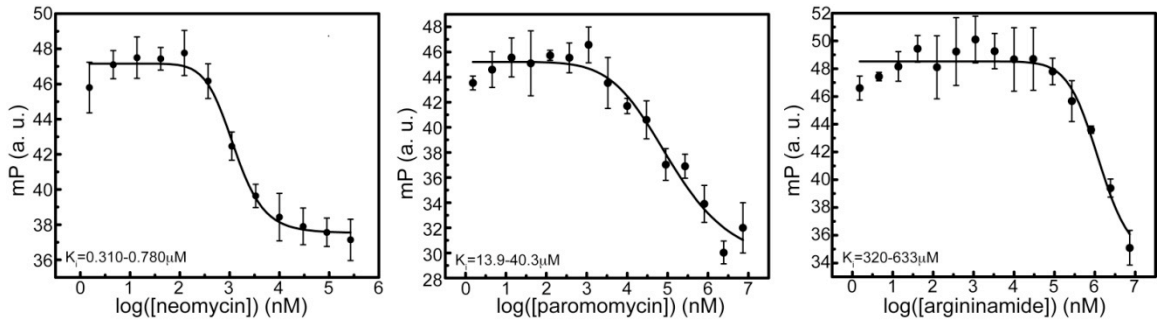


Figure 4.5. Known TAR-binding small molecules displace the FTat peptide. Three known TAR-binders were tested to displace FTat peptide.

to commercial unavailability. Initially we obtained the top 39 commercially available hits from the CCG library with scores ranging from -45.53 to -33.22 kcal/mol, and the three commercially available small molecules from the in-house library. These 42 small molecules were tested to inhibit the TAR-Tat interaction. The three aqueous-soluble small molecules (10 mM) from the in-house library, butirosin (score= -35.18 kcal/mol), netilmicin (score= -29.58 kcal/mol) and amikacin (score= -25.65 kcal/mol), inhibited TAR-Tat binding interaction with inhibition constants (K_is) of 12.7 μM, 11.5 μM and 13.1 μM, respectively, and were confirmed to bind TAR by monitoring CSPs (Figure 4.6a-c). However, only three of the 39 small molecules from the CCG library were soluble in aqueous solution at 10 mM, and of these only mitoxantrone displaced FTat (K_i=1.54 μM) and was confirmed to interact with TAR using CSPs (Figure 4.6e). The other 36 aqueous-insoluble small molecules were dissolved in DMSO at 10mM and serially diluted into aqueous buffer and failed to displace the Tat peptide. To ensure that suspending these 36 small molecule in DMSO followed by dilution into aqueous buffer did not prevent detection of small molecule binding, DMSO was titrated onto the TAR-Tat complex up to 20% DMSO. No perturbations that indicate interference from DMSO were observed (data not shown). We hypothesize that the source of scoring inaccuracies are a result of the hydrophobic and aromatic small molecules in the CCG library not being amenable to binding the solvent exposed grooves of TAR. ICM uses a

distant dependent dielectric to simulate an aqueous environment, which is not as accurate as using a higher-level approximation such as the Generalized-Born, Poisson-Boltzmann, or explicit solvent models, which could be a source of error in the ICM scoring function. Also, the ΔG_{Hphob} term in the ICM scoring could be over-approximated for RNA docking resulting in artificially low docking scores for hydrophobic small molecules.

To experimentally test the hypothesis that soluble molecules will give rise to more accurate docking predictions, only hits with a Log(P)<-1 were considered. Out of the initial 49,166 small molecules with N_{flex}<15, only 446 have a predictedLog(P)<-1. Of these 446 small molecules, the top 38 commercially available hits were obtained from commercial vendors. Interestingly, three of these small molecules are previously known TAR binders kanamycin (score= -27.72 kcal/mol), ribostamycin (score= -25.37 kcal/mol), and tobramycin (score= -24.86 kcal/mol), which were tested by FP and shown to inhibit the FTAR-Tat interaction with K_is of 4.89 μM, 85.6 μM, and 2.83 μM, respectively. Of the remaining 35 small molecules, 32 were soluble in aqueous solution and subjected to the FP displacement assay. Remarkably, four small molecules inhibit the TAR-FTat interaction. In fact, this is the first evidence that the small molecules amiloride (score= -23.73 kcal/mol) and 5-(N,N)-diemthyl amiloride (14) (score= -28.31 kcal/mol) bind RNA. In total, of the 35 water-soluble small molecules tested, seven more inhibitors were discovered with K_is ranging from 5.85-300 μM (Figure 4.6). The four hits other than the previously known TAR binders were also shown to bind TAR using CSPs (Figure 4.6e-4.7h). Thus, only considering the CCG library, the hit rate increases from 2.6% to 25% when pre-filtering the docking scores according to the predicted Log(P)<-1, and overall the SAS ensemble affords the discovery of 11 TAR binders resulting in a 31% hit rate. Table 4.2 summarizes the Tat displacement assay results.

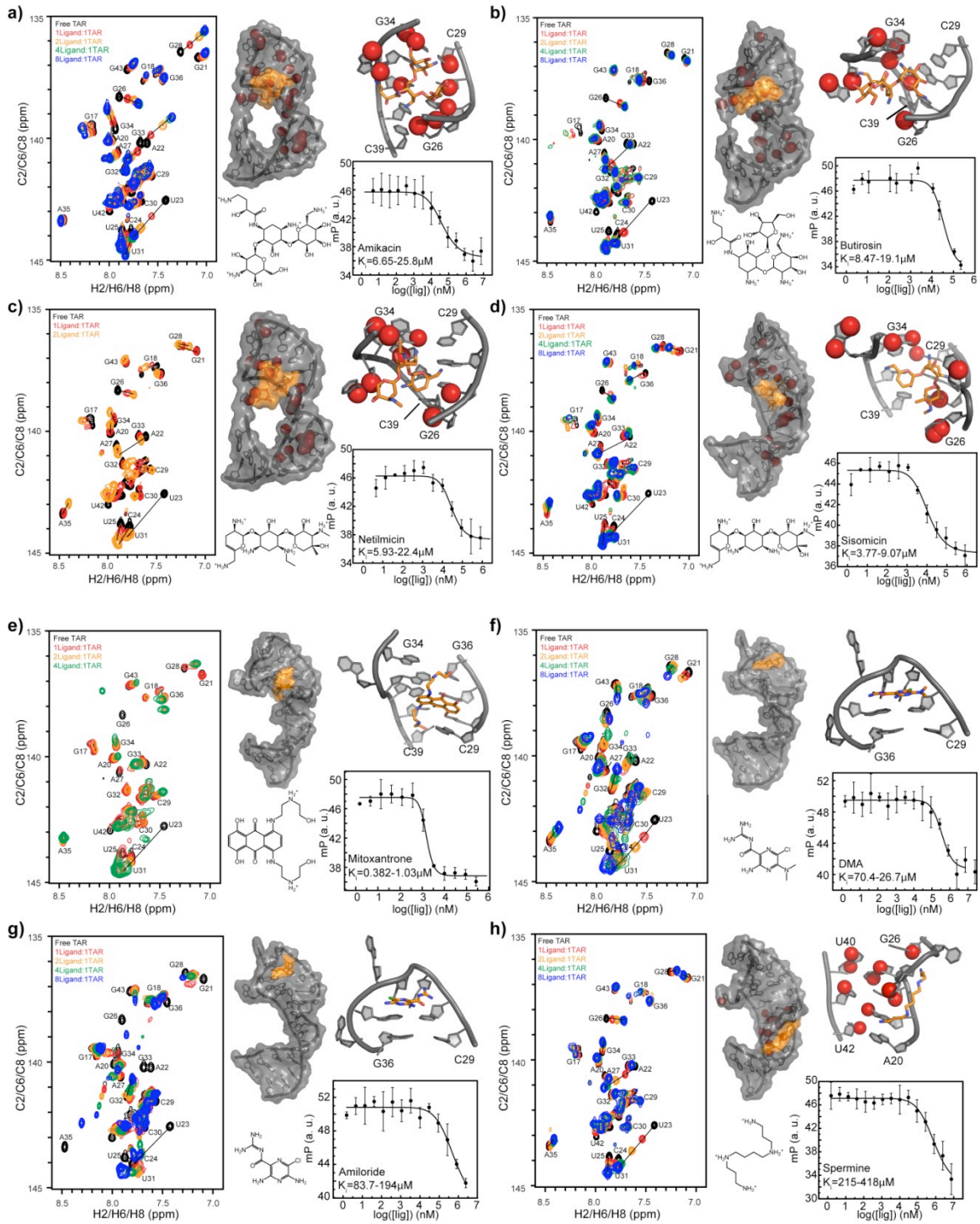


Figure 4.6. Eight newly discovered TAR binders. FP and CSP data validate that the eight newly discovered small molecules **(a)** amikacin, **(b)** butirosin, **(c)** netilmicin, **(d)** sisomicin, **(e)** mitoxantrone, **(f)** amiloride, **(g)** DMA, and **(h)** spermine bind TAR. Each panel displays the C2/C6/C8 NMR spectra from titration of the small molecule (left), the Tat-displacement FP data (bottom right), and the docking model (top right). Atoms depicted by red spheres are resonances that shift >0.1 ppm. Red spheres are not shows for mitoxantrone, amiloride and DMA because they intercalate with TAR and cause perturbations throughout the entire RNA.

Table 4.2. Summary of FTat displacement assay results.

Molecule	Score (kcal/mol)	IC_{50} (μM)	K_i (μM)
Neomycin[1, 2]	-30.49	1.3	0.492 (0.310-0.780)
Paromomycin[1, 2]	-27.81	68.4	23.6 (13.9-40.3)
Argininamide[1, 2]	-26.43	2260	450 (320-633)
Mitoxantrone	-45.53	1.54	0.627 (0.382-1.03)
Tobramycin[2]	-24.86	8.73	2.83 (1.61-4.96)
Sisomicin	-33.04	14	5.85 (3.77-9.07)
Netilmicin	-29.58	28.7	11.5 (5.93-22.4)
Butirosin	-35.18	36.8	12.7 (8.47-19.1)
Amikacin	-25.65	43.6	13.1 (6.65-25.8)
Kanamycin[2]	-27.72	48.1	4.89 (2.33-10.3)
5-(N,N-dimethyl)-amiloride	-28.31	398	137 (70.4-267)
Ribostamycin[2]	-25.37	435	85.6 (59.2-124)
Amiloride	-23.73	517	128 (83.7-194)
Spermine	-29.25	2340	300 (215-418)

[1]Small molecules used to test FP assay
[2]Previously known TAR binders

4.6 Docking Model Accuracy

4.6.1 Chemical Shift Perturbations

While changes in chemical shift of RNA resonances can arise for many different reasons, such as interaction with a small molecule or a conformational change, they do provide a means to qualitatively validate the accuracy of the docking predicted models. All small molecules were titrated in the same TAR:ligand ratio except for netilmicin and mitoxantrone which precipitate TAR at ratios of >2:1 and >4:1, respectively. All C8H8/C6H6/C2H2, C5H5, C1'H1', and NH spectra are shown at the end of this section. On titration with mitoxantrone, amiloride and DMA many TAR resonances become progressively weaker and coalesce, indicating that all three likely intercalate with TAR. The docking models agree with this mode of action for amiloride and DMA where the small molecules stack on top of the C30-G34 apical loop base-pair and insert into the

apical loop (Figure 4.6f-g). However, the model for mitoxantrone only predicts stacking with G26 suggesting intercalation is less likely (Figure 4.6e). This could be explained by the larger size of mitoxantrone compared to amiloride and DMA and is a potential source of error in the docking predictions. Resonances that exhibit significant CSPs are not indicated on the structure models due to the large change in chemical for shift for many resonances. This highlights one disadvantage of being unable to simulate binding induced RNA conformational changes. CSPs are observed in many regions of TAR on binding the four newly discovered aminoglycosides, amikacin, butirosin, netilmicin, and sisomicin. The

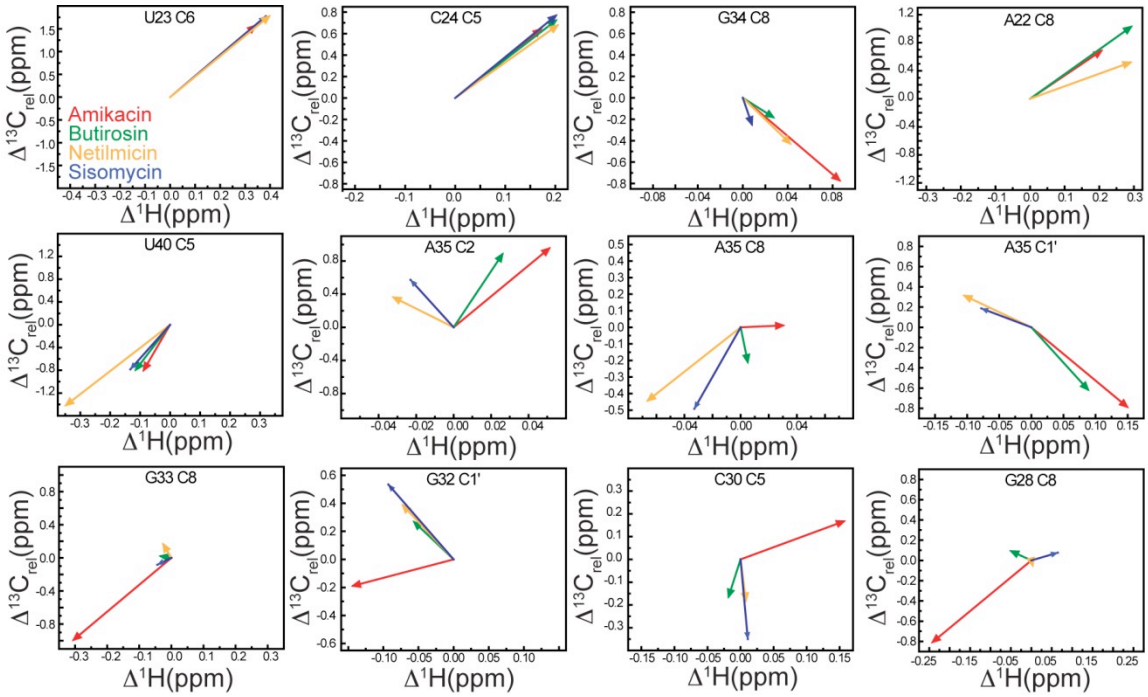


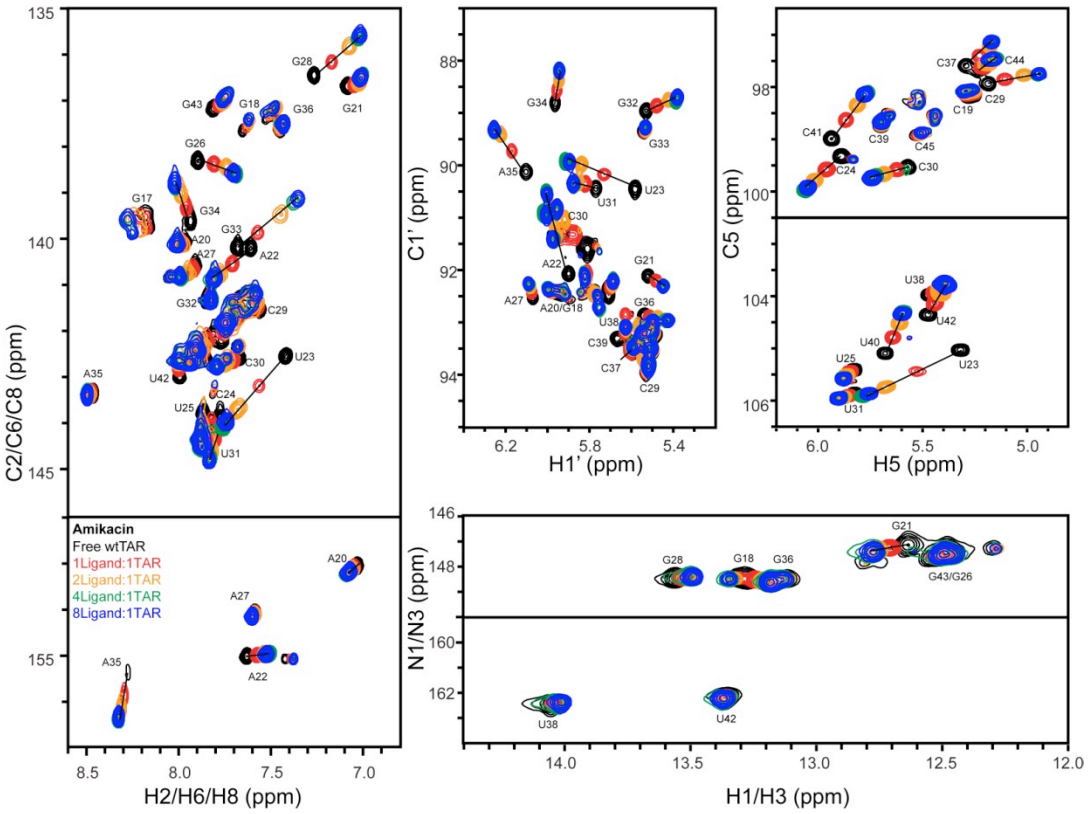
Figure 4.7. Weighted average CSPs indicate differentiating binding modes for the four newly discovered aminoglycosides. Examples of interesting weighted average CSPs for amikacin (red), butirosin (green), netilmicin (yellow), and sisomycin (blue) are shown. Weighted average CSPs are calculated between the saturated and free TAR.

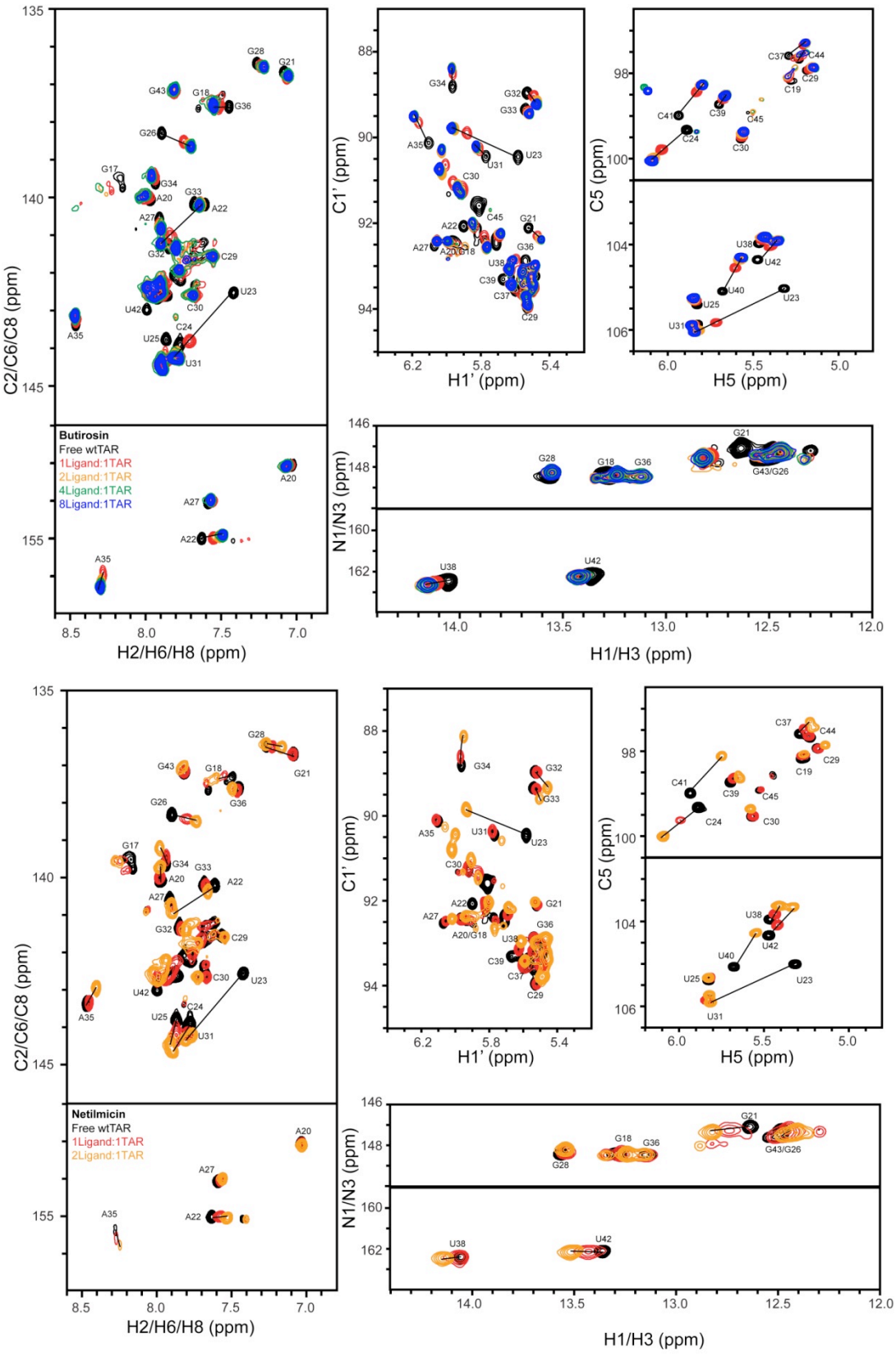
docking predictions for all four aminoglycosides indicate a helix-2 binding mode, which largely agrees with the helix-2 CSPs(Figure 4.6a-d). However, the CSPs in the bulge region and helix-1 cannot be explained by the lowest scoring binding modes.

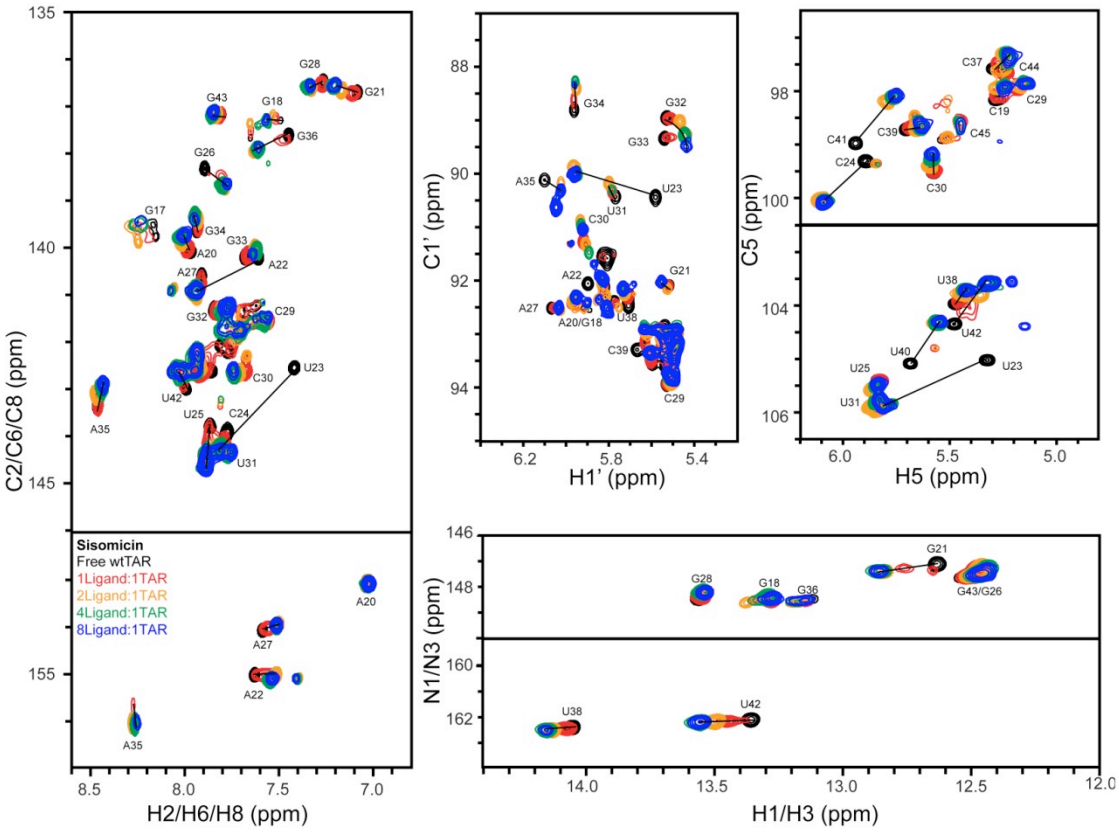
Perturbations in these regions could result from >1 stoichiometry, conformational changes on binding, and/or incorrect docking predictions. More positive scoring poses are predicted to bind the bulge, loop, and the major groove of helix-1 suggesting that alternative small molecule conformations or >1 binding stoichiometry is possible. Some CSPs also agree with spermine docking model, but the small size and linearity of spermine affords predictions of largely buried poses, whereas it is more likely spermine binds the major groove similar to argininamide(15).

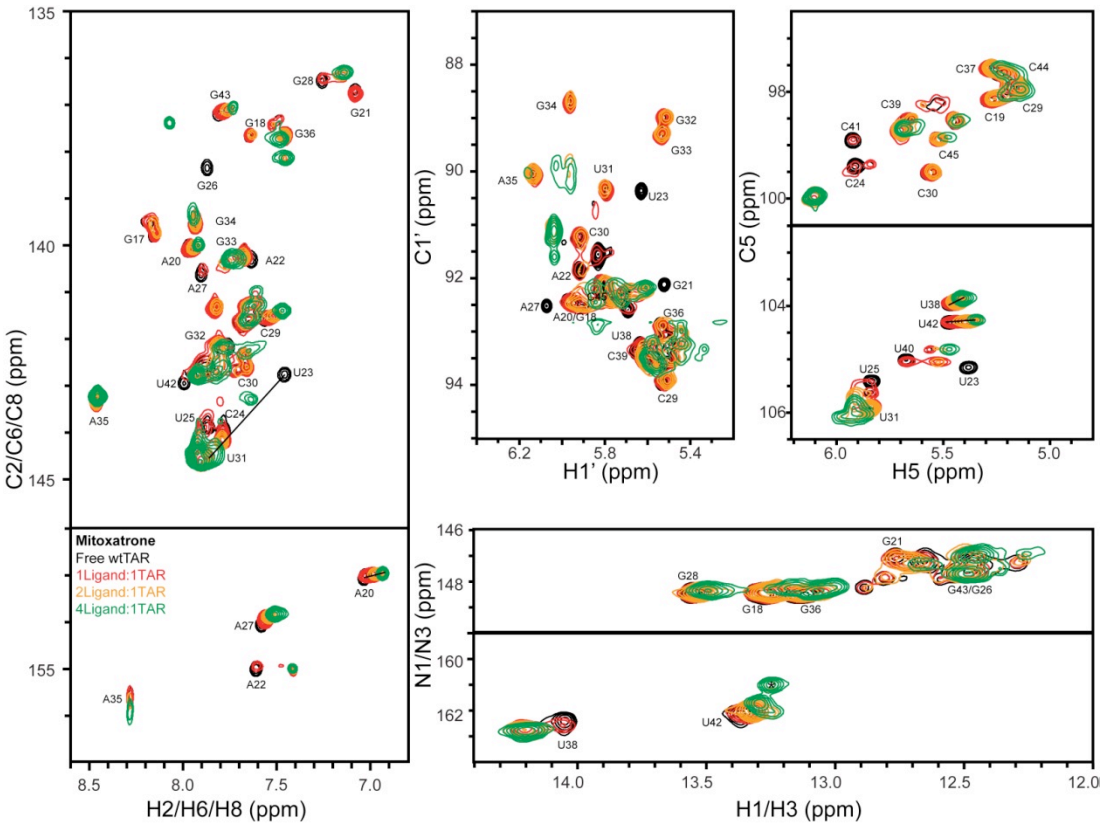
Comparison of the CSPs on titration of the four newly discovered aminoglycosides also reveals interesting trends that in some cases can be explained by the docking predicted binding conformations. The difference in weighted average chemical shift for the endpoints of amikacin, butirosin, netilmicin, and sisomicin were computed. As expected, similar shifts are observed for many resonances such as the C6H6 resonance of U23 (Figure 4.7), which shifts downfield in both dimensions indicative of a looped out conformation similar to TAR binding to Mg^{2+}(16,17). Also, the C5H5 resonance of C24, the C8H8 resonance of G34, the C8H8 resonance of A22, and the C5H5 resonance of U40 exhibit similar CSP directions, albeit different magnitudes, likely indicating similar bound state conformations (Figure 4.7). Interestingly, there are differentiating chemical shifts between the four aminoglycosides that group together according to the chemically similar amikacin and butirosin, and netilmicin and sisomicin. Some of the perturbations can be explained by the different predicted bound-state conformations. For example, the C8H8, C2H2, and C1'H1' resonances of A35 shift in two general directions with similar shifts for amikacin and butirosin, and netilmicin and sisomicin (Figure 4.7). These varying shifts for the two groups of aminoglycosides can be explained by the different docking-predicted TAR-aminoglycoside interactions. Amikacin and butirosin both have the L-haba substituent that interacts with the RNA backbone below A35 likely giving rise to the similar chemical shifts, whereas only

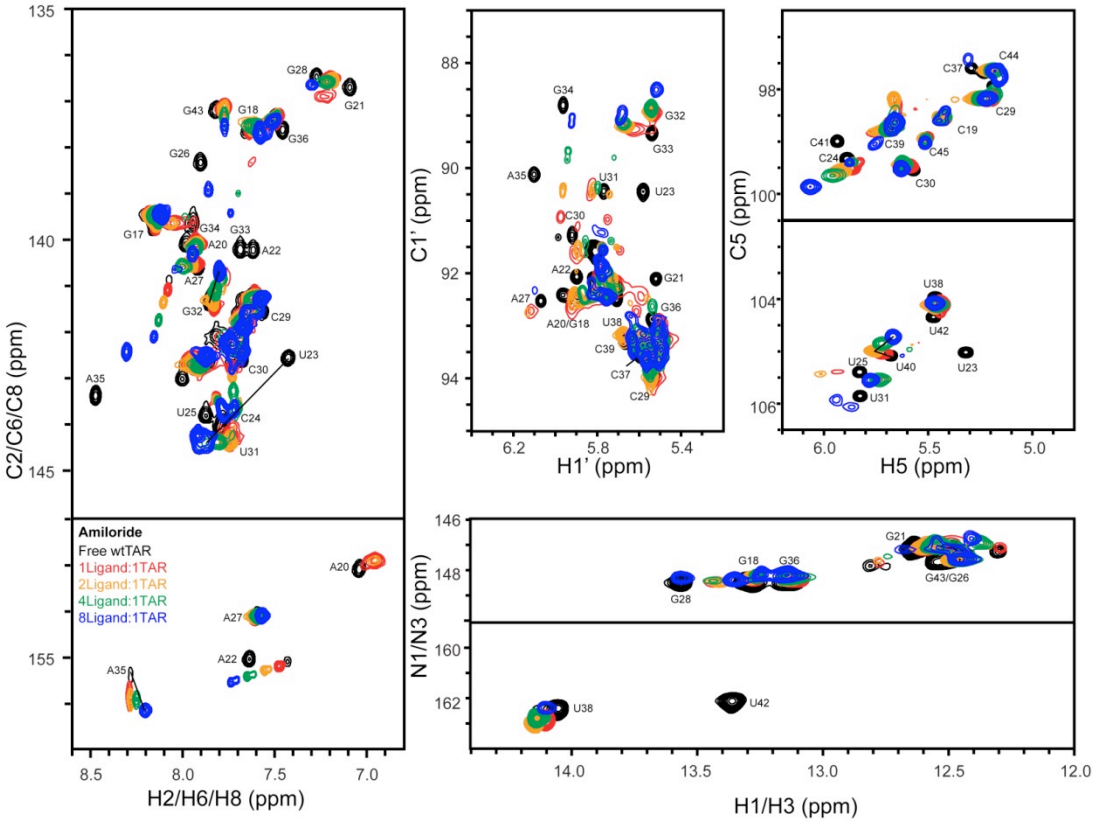
netilmicin interacts with the A35 nucleobase and neither netilmicin or sisomicin interact with the A35 backbone region. While not clearly re-capitulated in the docking models, other data such as the C8H8 resonance of G33, the C1'H1' resonance of G32, the C5H5 resonance of C30, and the C8H8 resonance of G28 indicate differential binding for amikacin compared to the other three aminoglycosides (Figure 4.7). While these data suggest that the four aminoglycosides likely stabilize alternative RNA conformations, the subtle differences cannot be adequately characterized with the current methodology. Further improvements to the SAS and docking methods will likely improve these predictions.

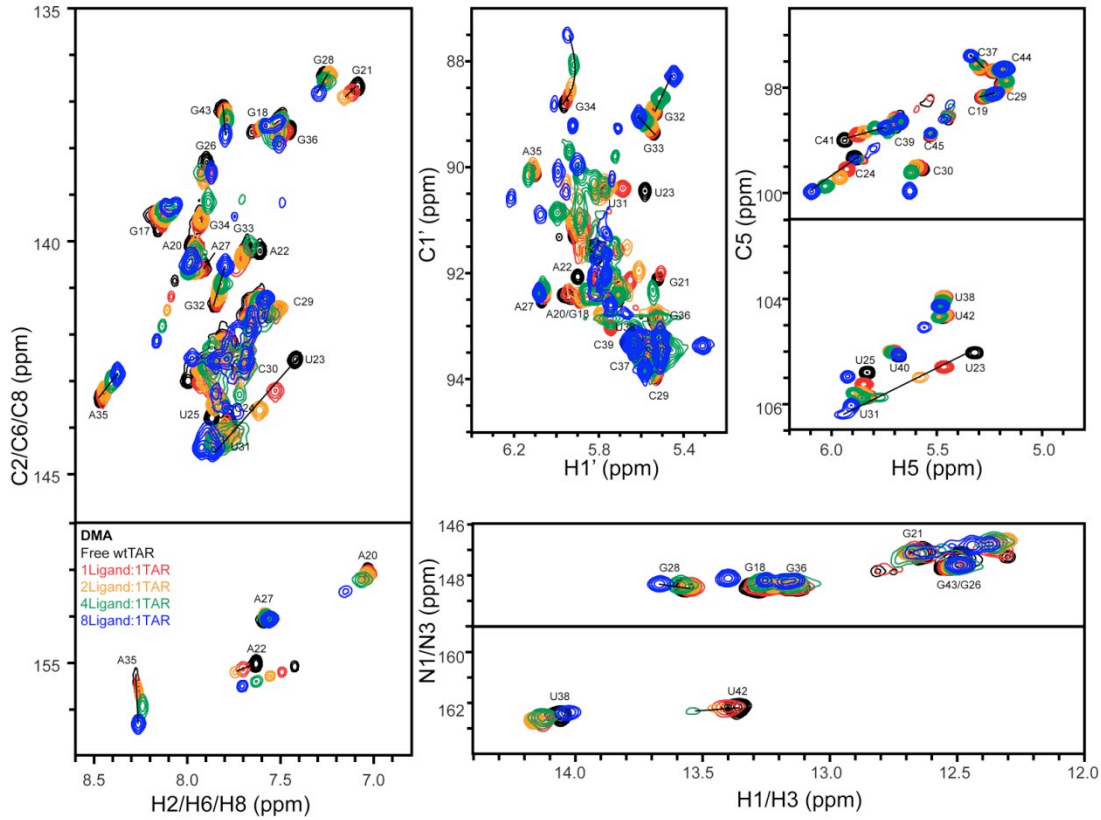



Butirosin
Free wtTAR
1Ligand:1TAR
2Ligand:1TAR
4Ligand:1TAR
8Ligand:1TAR
C2/C6/C8 (ppm)
H2/H6/H8 (ppm)
C1' (ppm)
H1' (ppm)
C5 (ppm)
H5 (ppm)
N1/N3 (ppm)
H1/H3 (ppm)
Netilmicin
Free wtTAR
1Ligand:1TAR
2Ligand:1TAR


Sisomicin
Free wtTAR
1Ligand:1TAR
2Ligand:1TAR
4Ligand:1TAR
8Ligand:1TAR
C2/C6/C8 (ppm)
H2/H6/H8 (ppm)
C1' (ppm)
H1' (ppm)
C5 (ppm)
H5 (ppm)
N1/N3 (ppm)
H1/H3 (ppm)


Mitoxatrone
Free wtTAR
1Ligand:1TAR
2Ligand:1TAR
4Ligand:1TAR
C2/C6/C8 (ppm)
H2/H6/H8 (ppm)
C1' (ppm)
H1' (ppm)
C5 (ppm)
H5 (ppm)
N1/N3 (ppm)
H1/H3 (ppm)


Amiloride
Free wtTAR
1Ligand:1TAR
2Ligand:1TAR
4Ligand:1TAR
8Ligand:1TAR
C2/C6/C8 (ppm)
H2/H6/H8 (ppm)
C1' (ppm)
H1' (ppm)
C5 (ppm)
H5 (ppm)
N1/N3 (ppm)
H1/H3 (ppm)


DMA
Free wtTAR
1Ligand:1TAR
2Ligand:1TAR
4Ligand:1TAR
8Ligand:1TAR
C2/C6/C8 (ppm)
H2/H6/H8 (ppm)
C1' (ppm)
H1' (ppm)
C5 (ppm)
H5 (ppm)
N1/N3 (ppm)
H1/H3 (ppm)

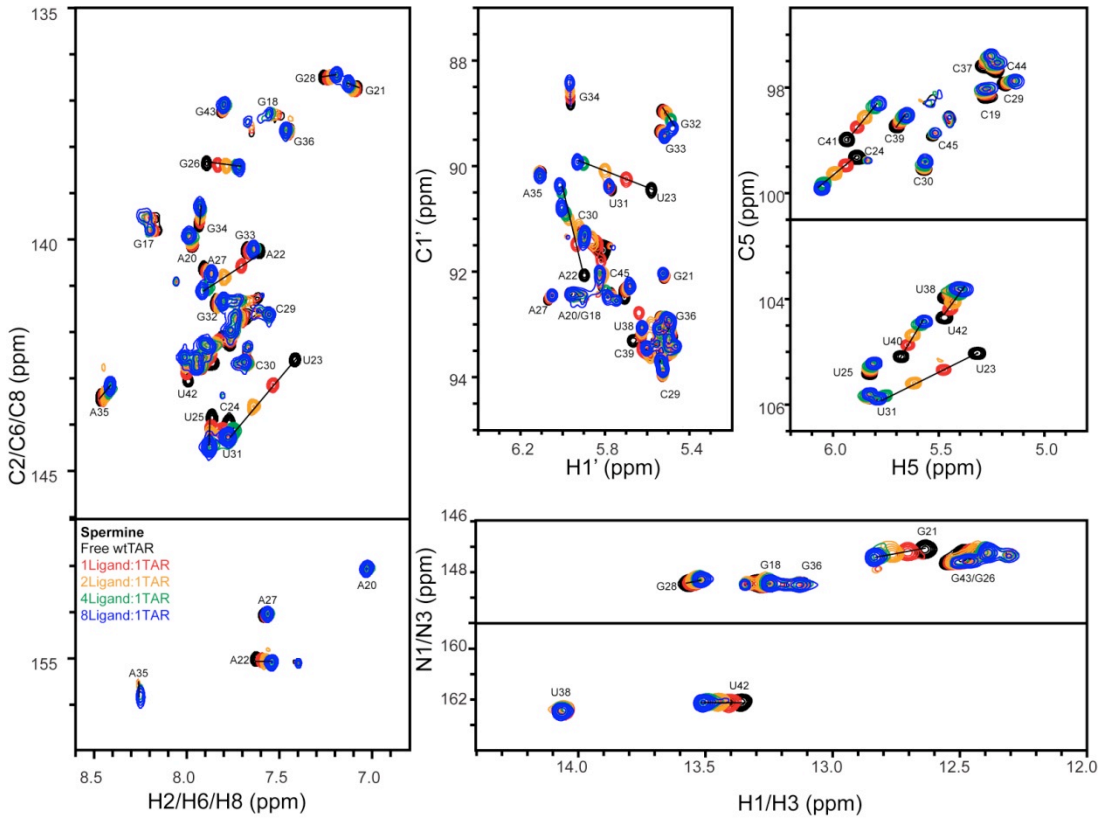


Figure 4.8. NMR spectra of TAR on titration with the eight docking-predicted hits.

4.6.2 RDC Validation of the Netilmicin-TAR Docking Model

To quantitatively test the ability of docking to accurately predict the small molecule bound RNA structure, RDCs measured for TAR bound to netilmicin were compared to predicted RDCs based on the docked SAS ensemble structures. SVD order tensor analysis was performed to determine the time-averaged netilmicin-bound TAR conformation. Results from the order tensor analysis show that the back-calculated and measured RDCs closely match to within experimental uncertainty (~2 Hz) (Figure 4.9). Euler angles calculated from the RDC determined structure indicate that on binding netilmicin, TAR exhibits a β of 22.92±2.7°, an α of 51.21°, and a γ of -56.17° (Figure 4.9) (A description of the interhelical Euler angles can be found in section 2.3). These angles are in relatively close agreement with the docking predicted top-scoring SAS structure,

which exhibits an α of 114.9°, a β of 56.94°, and a γ of -46.19°. Interestingly, the interhelical generalized degree of order (GDO_{int}) describing the interhelical dynamic amplitude, where 1 corresponds to complete interhelical rigidity and 0 to fully isotropic motions, is 0.68±0.09 indicating that TAR remains dynamic when bound to netilmicin.

Helix	N	RMSD (Hz)	CN#	Q (%)	*R*	η	GDO(x10^{-3})	GDO_{int}	α^o, β^o, γ^o
I	11	1.74	3.71	7.35	0.99	0.201	1.20±0.05	0.68±0.09	51.2, 22.9±2.7, -56.2
II	11	1.99	5.11	5.78	0.98	0.212	1.76±0.11		

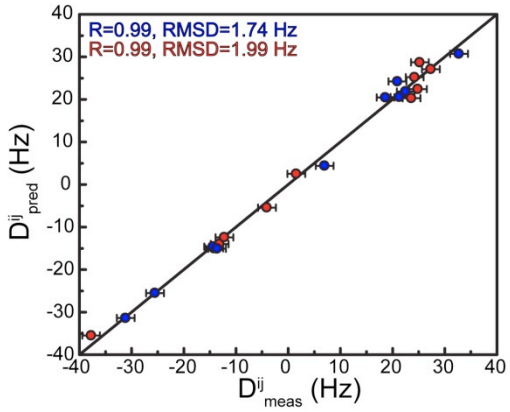


Figure 4.9. SVD order tensor analysis of TAR bound to netilmicin. Results from SVD order tensor analysis of TAR bound to netilmicin measured RDCs. Helix-1 RDCs are shown in blue and helix-2 in red. N is the number of RDCs included in the analysis for each helix, RMSD is the root mean square deviation between measured and back-calculated RDCs, CN# is the condition number, Q is the quality factor between the measured and back-calculated RDCs, η describes the asymmetry of the order tensor (where 0 indicates a fully symmetric tensor), R is the correlation between measured and back-calculated RDCs after the order tensor analysis, GDO, is the generalized degree of order for each order tensor, GDO_{int} describes the magnitude of interhelical motions with 1 describing interhelical rigidity and 0 fully isotropic motions, and α^o, β^o, and γ^o are the interhelical Euler angles. Full descriptions of these parameters can be found in reference(18).

To further probe the accuracy of netilmicin-bound TAR model, RDCs were predicted based on the SAS structure structures docked to netilmicin and compared to experimental values. A previous study by Al-Hashimi H. M. and co-workers showed that on average all RNA structures in the PDB exhibit A-form helices(19), thus A-form helices were used in place of the SAS ensemble helices. Since the experimentally determined

principal axis of the order tensor (S_{zz}) for TAR bound to netilmicin is greater for helix-2 ($1.75x10^{-3}$) compared to helix-1 ($1.20x10^{-3}$), we assumed that helix-2 aligned collinear with the z-axis and all 20 SAS structures were superimposed onto the experimentally determined helix-2 principal axis system (PAS). Next, A-form helices comprising residues 17-22 and 40-45 (Figure 4.10) were superimposed onto helix-1 of the SAS structures. The resulting structure for each SAS snapshot consisted of an idealized A-form helix-1 and helix-2, and the trinucleotide bulge and apical loop from the SAS structures. These structures were then used to calculate theoretical RDCs. Since TAR remains dynamic when bound to netilmicin, RDCs calculated for each SAS structure were weighted using a partition function. The netilmicin-bound docking score for each SAS structure was converted into a ΔG using the best-fit line equation from the $N_{flex}<15$ crystal and NMR structure validation data. These free energies were then used to calculate a probability (P) using a partition function,

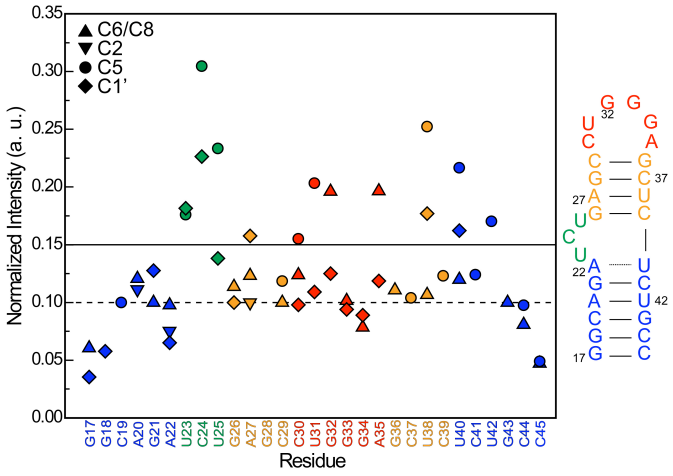


Figure 4.10. Normalized resonance intensities for TAR bound to netilmicin.

$$P = \frac{e^{\Delta G_i / RT}}{\sum_{i=1}^{i} e^{\Delta G_i / RT}}. \qquad [4.5]$$

P was used to weight contribution from each SAS structure to the predicted RDCs. RDCs for atoms that exhibit large normalized intensities (>0.15) were eliminated from comparison to experimental RDCs because these nuclei likely exhibit fast local motions that cannot be accurately determined in this analysis (Figure 4.10). Since apical loop RDCs were not used in the

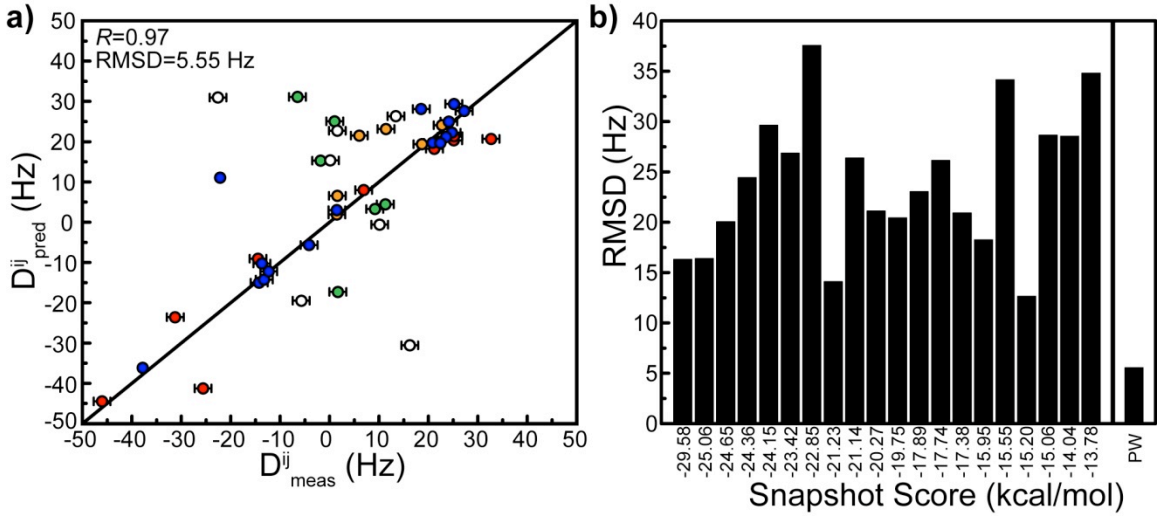


Figure 4.11. Netilmicin-bound SAS structures predict experimental RDCs. (a) The experimental RDCs are plotted against the back-calculated RDCs that are partition weighted according to the docking score for each structure. Helix-1 RDCs are shown in blue, helix-2 RDCs in red, bulge RDCs in orange, and apical loop RDCs in green. RDCs that were excluded based on high resonance intensity are shown as open circles. **(b)** RMSDs for each SAS structure generally increase as the docking score becomes more positive and all are significantly larger than the partition weighted (PW) RMSDs.

generation of the SAS ensemble, only atoms for residues that contact the small molecule should be stabilized on binding netilmicin and were used in the predictions. Remarkably, the resulting theoretical and experimental RDCs exhibit an RMSD of 5.55 Hz (Figure 4.11a). Although the trend is not uniform, increasing RMSDs are calculated for poorer scoring snapshots, and the lowest RMSD is calculated for the partition weighted RDCs from all 20 structures (Figure 4.11b). These data suggest that docking against the 20 SAS structures affords an accurate prediction of the experimental RDCs

that cannot be accounted for by one structure. This is consistent with TAR being flexible in the netilmicin-bound state.

4.7 Netilmicin Inhibits HIV-1 Replication

Results from the FP assay indicate that netilmicin binds TAR and inhibits binding of the Tat peptide. To determine if netilmicin inhibits the TAR-Tat interaction in a more biological setting, we tested the ability to inhibit Tat-mediated activation of the HIV-1 LTR using a luciferase reporter assay. Mammalian cells were transiently transfected with an HIV-1 LTR luciferase reporter, a reporter plasmid for normalization of transfection efficiency (Renilla luciferase), and either an expression vector expressing HIV-1 Tat (Tat101) or the vector alone (Tat0). The LTR contains the same TAR sequence used in the NMR, FP, and computational studies. The cells were then incubated with netilmicin or PBS and then harvested for measuring luciferase activity. As can be seen in Figure 4.12a, addition of 100 μM netilmicin inhibits the Tat-mediated activation of the HIV-1 promoter by ~71% when compared to the PBS-treated control. Addition of phorbol 12-myristate 13-acetate (PMA) to the cells can activate the HIV-1 LTR in the absence of Tat. Since this activation is also independent of TAR, we would expect no inhibition of PMA-mediated HIV-1 LTR activation by netilmicin, and this was indeed the case. Mock treatments of cells not containing the LTR or Tat expression vector with netilmicin further ensures that netilmicin decreases luciferase expression by binding TAR.

Having shown that netilmicin inhibits Tat-mediated activation of the HIV-1 LTR in reporter gene assays, we next asked whether this would translate into inhibition of HIV-1 replication by netilmicin. Sisomicin, which differs from netilmicin by an ethylene group, was also tested in these experiments (Figure 4.12b). An HIV-1 indicator cell line, TZM-bl, was used in this experiment. The cells contain the luciferase and β-galactosidase genes under the control of the HIV-1 promoter, as well as the HIV-1 receptors CD4, CXCR4,

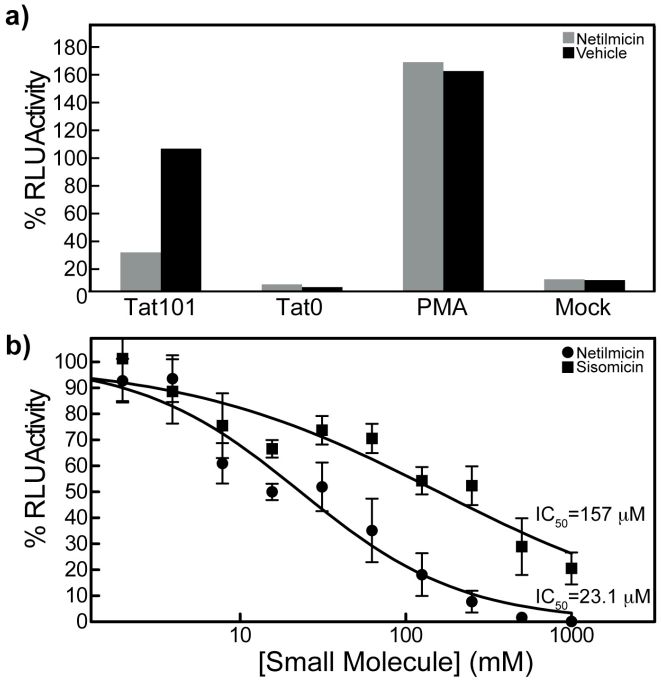


Figure 4.12. Netilmicin inhibits Tat mediated HIV-1 transcription and HIV-1 replication. (a) Netilmicin reduces HIV-1 LTR-controlled luciferase expression. Tat101 indicates a cell line transfected with a Tat expression vector, Tat0 cells contain the LTR but no Tat expression vector, PMA indicates Tat-independent activation of the LTR, and Mock indicates cells were not transfected but were treated with **(b)** Netilmicin and sisomicin inhibit HIV-1 replication in live HIV-1 mimic cell lines (TMZ-bl).

and CCR5. Thus, when they are infected with HIV, the replication genes are activated, allowing for assessment of HIV-1 infection by measuring reporter gene expression. The virus used was the HIV-1 isolate NL4-3 virus containing the same TAR sequence used NMR, FP, and computational studies. When netilmicin was added to the cells prior to infection, we observed decreased HIV-1 LTR activation, demonstrating a decrease in infection (Figure 4.12b). Using non-linear regression, we calculated the IC_{50} value of netilmicin to be 23.1 μM, which is nearly identical to the value measured using the Tat displacement assay. The IC_{50} value for sisomicin (157 μM) is approximately 100X greater than the value measured in the FP assay. These results corroborate the results of the transient transfection experiment and show that netilmicin can inhibit HIV-1 replication. Therefore, netilmicin, a compound selected through the use of NMR and computational modeling, can block TAR-Tat interaction in vitro, specifically inhibit Tat-mediated activation of the HIV promoter, and interfere with HIV replication in vivo.

4.8 Conclusion

Computational docking provides the ability to rapidly screen small molecules against an RNA provided the structure is known *a-priori*. Validation of docking to accurately predict ΔG_{exp} is shown here to depend on the accuracy of the RNA structure and the conformational flexibility of the small molecule. When the structure of the RNA-small molecule complex is well characterized using X-ray crystallography or NMR, ICM can be used to accurately calculate docking scores that correlate strongly with the ΔG_{exp}. Also, the small molecule conformation is more accurately predicted for less flexible small molecules, X-ray structures, and small molecules with tight binding affinities likely due to less flexibility in the bound-state. The ability to use the SAS ensemble structures as docking receptors was investigated by docking 33 known TAR-binding small molecules with N_{flex}<15 and reported ΔG_{exp}. Remarkably, ICM is able to predict docking scores that correlate strongly with ΔG_{exp} (R=0.69). Virtually screening ~60,000 small molecules from the CCG and ~2500 from an in-house RNA-binding small molecule library resulted in the identification of 11 TAR binding small molecules that inhibited the TAR-Tat interaction with K_is ranging from 5.85-300 μM, resulting in a 31% hit rate. The eight small molecules were confirmed to bind TAR using NMR CSPs. The CSPs strongly suggest that mitoxantrone, amiloride, and DMA intercalate with TAR, and all four aminoglycosides exhibit similar helix-2 binding profiles. Support for the docking predicted models of the eight newly discovered TAR-binders is reinforced by the CSP data, however there are outlying data that suggest higher stoichiomety binding or alternative binding conformations, which could be due to the inaccuracy of the SAS and docking methods. These potential errors may be alleviated with more input NMR data. Validation that the SAS ensemble predicts the correct netilmicin-bound TAR conformation suggests that the SAS ensemble is able to capture dynamical features of TAR that afford accurate docking

predictions and together recapitulate the experimental RDC data. Overall, the SAS ensembles can be used to generate RNA conformational ensembles that are beneficial as a pre-processing docking method and provides more accurate predictions of small molecule docking scores, efficient identification of lead compounds, and reliably predict experimental small molecule-bound RNA structures.

The idea was conceived by Al-Hashimi H. M. and Stelzer A. C. Stelzer A. C. and Kratz J. K. synthesized RNA oligonucleotides and collected and analyzed NMR and FP data. Docking simulations were conducted and analyzed by Stelzer A. C. The in-cell experiments were conducted in the Markovitz D. laboratory by Gonzalez-Hernandez M. (gene reporter assays) and Swanson M. (TMZ-bl assays).

4.9 References

1. Frank, A.T., Stelzer, A.C., Al-Hashimi, H.M. and 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, 3670-3679.
2. Abagyan, R., Totrov, M. and 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, 488-506.
3. An, J., Totrov, M. and Abagyan, R. (2004) Comprehensive identification of "druggable" protein ligand binding sites. *Genome Inform*, 15, 31-41.
4. Fan, P., Suri, A.K., Fiala, R., Live, D. and Patel, D.J. (1996) Molecular recognition in the FMN-RNA aptamer complex. *Journal of Molecular Biology*, 258, 480-500.
5. Gilbert, S.D., Mediatore, S.J. and Batey, R.T. (2006) Modified pyrimidines specifically bind the purine riboswitch. *Journal of the American Chemical Society*, 128, 14214-14215.
6. Detering, C. and Varani, G. (2004) Validation of automated docking programs for docking and database screening against RNA drug targets. *Journal of Medicinal Chemistry*, 47, 4188-4201.
7. Guilbert, C. and James, T.L. (2008) Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *Journal of Chemical Information and Modeling*, 48, 1257-1268.
8. Pitt, S.W., Zhang, Q., Patel, D.J. and Al-Hashimi, H.M. (2005) Evidence that electrostatic interactions dictate the ligand-induced arrest of RNA global flexibility. *Angew Chem Int Ed Engl*, 44, 3412-3415.
9. Ippolito, J.A. and Steitz, T.A. (1998) A 1.3-angstrom resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 9819-9824.
10. Aboul-ela, F., Karn, J. and Varani, G. (1996) Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Research*, 24, 3974-3981.
11. Long, K.S. and Crothers, D.M. (1995) Interaction of Human-Immunodeficiency-Virus Type-1 Tat-Derived Peptides with Tar Rna. *Biochemistry*, 34, 8885-8895.
12. Matsumoto, C., Hamasaki, K., Mihara, H. and 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, 1857-1861.
13. Hamasaki, K. and Ueno, A. (2001) Aminoglycoside antibiotics, neamine and its derivatives as potent inhibitors for the RNA-protein interactions derived from HIV-1 activators. *Bioorganic & Medicinal Chemistry Letters*, 11, 591-594.
14. Henzler-Wildman, K. and Kern, D. (2007) Dynamic personalities of proteins. *Nature*, 450, 964-972.
15. Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D. and Williamson, J.R. (1992) Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science*, 257, 76-80.
16. Casiano-Negroni, A., Sun, X.Y. and Al-Hashimi, H.M. (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, 6525-6535.
17. Al-Hashimi, H.M., Pitt, S.W., Majumdar, A., Xu, W.J. and Patel, D.J. (2003) Mg2+-induced variations in the conformation and dynamics of HIV-1 TAR RNA probed using NMR residual dipolar couplings. *Journal of Molecular Biology*, 329, 867-873.
18. Bailor, M.H., Musselman, C., Hansen, A.L., Gulati, K., Patel, D.J. and Al-Hashimi, H.M. (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nature Protocols*, 2, 1536-1546.
19. Musselman, C., Pitt, S.W., Gulati, K., Foster, L.L., Andricioaei, I. and Al-Hashimi, H.M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *Journal of Biomolecular Nmr*, 36, 235-249.

Conclusion

5.1 Conclusion

The role of RNA dynamics in biological processes has been extensively studied over the past few decades(*1, 2*). Recent discoveries continue to highlight the need to extensively characterize the change in global and local RNA conformation on recognition of an effector molecule(*1, 2*). The realization that RNA exhibits complex tertiary structure, allosterically mediates biological processes, and can be targeted by small molecules birthed the field of RNA-targeted drug discovery. However, specifically targeting cellular RNA molecules has been largely unsuccessful. While some of the classic principles learned from protein-targeted drug-discovery can be applied when binding RNAs, many are not suited to tackle their highly dynamic nature. A complete description of the biophysical principles governing the large global and local conformational changes that take place between the unbound and ligand-bound state(s), would greatly increase our understanding of RNA-mediated recognition processes and advance the field of RNA-targeted drug discovery.

While no technique can probe the vast range of timescales and large amplitude motions necessary to fully characterize RNA motions, employing a combination of

experimental (i.e. NMR) and theoretical (i.e. Molecular Dynamics) techniques can overcome the shortages of information garnered from each technique and afford the necessary temporal and spatial accuracy. The RDC SAS approach is a large step forward in this regard. However, SAS currently remains limited by (i) NMR data density (ii) RDC timescale sensitivity (<ms). A recent theoretical study by Al-Hashimi H. M. and co-workers shows that all 25 elements of the Wigner matrix, which fully describes the order-tensor, and in this case RNA helix, orientation and dynamics, can be obtained with five independent sets of RDCs(*3*). However, we are currently limited to two, obtained by elongating either helix of a helix-bulge-helix RNA. To obtain all 25 Wigner matrix elements, alternative RNA alignment frames need to be designed. Using the RDC data from five linearly independent RDC sets with the SAS method would afford more accurate refinement of the MD trajectory. Incorporation of NOEs and RCSAs into the SAS methodology will also aid in a higher-definition ensemble. New developments by Al-Hashimi H. M. and co-workers show that lowly populated (~1%) nucleic acid conformations can be characterized using NMR relaxation dispersion measurements(*4*). While these methods do not give angular description of bond vectors like RDCs, they can be used to bias MD simulations toward RNA ensembles that agree with experimental data. Incorporation of these data will entail using novel MD methods (e.g. Replica Exchange, steered MD, and variable temperature MD) to theoretically generate these lowly populated species.

5.2 References

1. H. M. Al-Hashimi, *Chembiochem* 6, 1506 (Sep, 2005).
2. H. M. Al-Hashimi, N. G. Walter, *Current Opinion in Structural Biology* 18, 321 (Jun, 2008).
3. C. K. Fisher, Q. Zhang, A. Stelzer, H. M. Al-Hashimi, *Journal of Physical Chemistry B* 112, 16815 (Dec 25, 2008).
4. A. L. Hansen, E. N. Nikolova, A. Casiano-Negroni, H. M. Al-Hashimi, *Journal of the American Chemical Society* 131, 3818 (Mar 25, 2009).

Appendix 1

NMR Study of an Immunomodulatory Benzodiazepine Binding to its Molecular Target on the Mitochondrial F_1F_0-ATPase

A1.1 Introduction

Bz-423 is a 1,4-benzodiazepine that potently suppresses disease in autoimmune mice by selectively killing pathogenic lymphocytes(1,2). Affinity-based screening of a phage-display human cDNA expression library identified the oligomycin-sensitivity conferring protein (OSCP), a component of the mitochondrial F_1F_0-ATPase, as the molecular target of Bz-423(3). Binding of Bz-423 to the OSCP in the context of intact enzyme inhibits both synthesis and hydrolysis of ATP(4). Consistent with inhibition of the F_1F_0-ATPase, Bz-423 increases the generation of superoxide from the mitochondrial respiratory chain and this reactive oxygen species is the signal initiating apoptosis (as opposed to changes in ATP concentration)(5,6).

The OSCP is a 213 amino-acid long protein (including the 23 amino acid mitochondrial leader sequence) that is conserved among mammals and is not present in other ATPases(7). The OSCP along with subunits b, d, and F_6 form the peripheral stalk in mammalian F_1F_0-ATPases(8). The stalk links the integral membrane F_0 component of the enzyme with its soluble catalytic F_1 domain, which is located in the mitochondrial matrix. The peripheral stalk is believed to act as a stator, holding the F_1 $\alpha_3\beta_3$ hexamer

static while the central stalk ($\gamma\delta\varepsilon c_{10}$) rotates(9,10). In order to function properly, the stalk subunits must act in concert with one another; while the peripheral stalk subunits do not need to move in relation to one another during catalysis, disrupting the connections between the subunits disrupts coupling between F_0 and F_1(11). In addition, the stalk must be anchored at each end for proper function: the transmembrane domain of subunit b holds the stalk in the mitochondrial membrane, while the N-terminal end of the OSCP sits on top of the $\alpha_3\beta_3$ hexamer with the C-terminus protruding almost 100 Å along the surface of F_1 towards F_0(12). The N-terminal tails of the α-subunits are critical for binding to OSCP(13).

The structure of a 134-amino acid long N-terminal fragment of the δ-subunit from E. coli, the equivalent of the bovine OSCP, has been studied by NMR spectroscopy(14). The protein adopts a 6-helix bundle with a disordered C-terminus. NMR studies using the N-terminal domain of the bovine OSCP (OSCP-NT, residues 1–120) reveals a similar fold(15,16). Binding experiments with peptide fragments from the N-termini of F_1 α-subunits suggests that the interaction site on OSCP-NT comprises a hydrophobic groove between helices I and V. Hence, this interface, which is essential for the rotary mechanism of the enzyme, probably consists of helix–helix interactions.

Since Bz-423 does not bind at the active site of enzyme, we hypothesized that it may bind at, or near the F_1-OSCP interface where it can perturb one or more of the conformational transitions associated with the rotary (binding-change) mechanism of catalysis(4). To gain support for this hypothesis, we studied the binding of water soluble Bz-423 analogs (Figure 1) with OSCP-NT and a construct representing amino acids 1-145 by NMR spectroscopy using chemical shift perturbation and cross-relaxation experiments to localize the binding site. Our data identified a recognition site near the F_1-OSCP interface and a conformational change in the protein upon drug binding.

Collectively, these data suggest that Bz-423 like inhibitors may function in an allosteric manner inducing a conformational change that hinders catalysis.

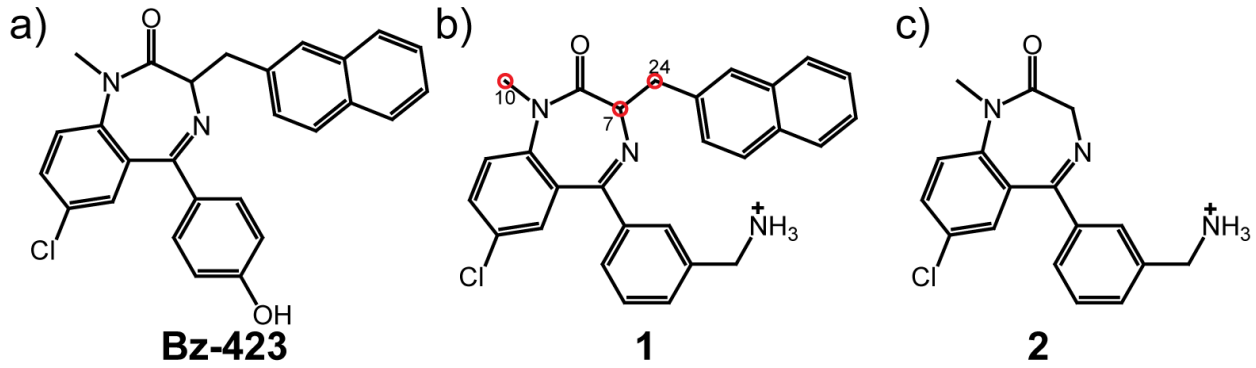


Figure 1. Chemical structure of (a) Bz-423 and the molecules used in this study: (b) 5-(3-(aminomethyl)phenyl)-7-chloro-1-methyl-3-(naphthalen-2-ylmethyl)-1*H*-benzo[*e*][1,4]diazepin-2(3*H*)-one (**1**) and (c) (aminomethyl)phenyl)-7-chloro-1-methyl-1*H*-benzo[*e*][1,4]diazepin-2(3*H*)-one (**2**). Protons of **1** that were saturated in cross-relaxation experiments are labeled. Based on prior structure-reactivity studies as a guide, **1** was designed to replace the critical phenolic proton with an ammoninum group to enhance aqueous solubility. The activity of **1** in enzyme and cellular assays is comparable. Removing the naphthyl group in **2** abolishes all activity against the enzyme.

A1.2 Materials and Methods

A1.2.1 Synthesis of Benzodiazepines

Benzodiazepines **1** and **2** were synthesized and characterized as previously described(15). Samples of both compounds used in the NMR titration experiments were >98% pure.

A1.2.2 OSCP Isolation

A 360-base-pair long insert encoding a truncated bovine OSCP, containing amino acids 1-120, was prepared via PCR using forward (5'CACCATGTTTGCCAAGCTTGTGAGGCC3') and reverse (5'CTAAACTGTGCATGGTACTTCTCC3') oligonucleotide primers (25 μM each) in the presence of dNTPs (0.5 mM), pOSCP (1 ng), PFU Turbo DNA polymerase (3-5 U;

Stratagene, La Jolla, CA), and 1X PFU Turbo DNA polymerase buffer. PCR conditions consisted of 30 s at 94 ^{o}C, 30 s annealing at 55 ^{o}C, and a 1 min extension at 72 ^{o}C for 31 cycles(17) A 435-base-pair long insert encoding a truncated bovine OSCP containing amino acids 1-145 was prepared in a similar manner using the same forward oligonucleotide primer as above together with a reverse (5'CTGGCCTTTACTTAGGAAGCTCTTCAGG3') oligonucleotide primer. The PCR products were purified using a Qiagen (Valencia, CA) PCR Clean Up kit following the manufacturer's instructions. Both inserts were cloned into expression vectors provided in the TOPO 10 kit following the manufacture's instructions (Invitrogen, Carlsbad, CA). The OSCP 1-120 insert was ligated into pCRT7/NT-TOPO which contains both an Express^{TM} Epitope coding sequence and a hexa-his tag, placing the OSCP 1-120 coding region C-terminal to these sequences. The OSCP 1-145 insert was ligated into pCRT7/CT-TOPO, which contains both a V5 Epitope coding sequence and a hexa-his tag, placing the OSCP 1-145 coding region N-terminal to these sequences. Thus, OSCP120 contains an N-terminal hexa-his tag connected by the linker sequence GMASMTGGQQMGRDLYDDDDKDPTL, while the OSCP145 construct contains a C-terminal hexa-his tag connected by the linker sequence KGNSKLEGKPIPNPLLGLDSTRTG. Primary structures for all OSCP constructs are shown in Figure A1.2. Ligation products were transformed into TOP-10 cells and selected on LB-ampicillin (50 μg/mL) plates. Colonies were screened for inserts via PCR using mini-prep plasmid DNA and both the forward and reverse T7 primers. The presence of the correct inserts was confirmed by automated DNA sequencing (Sequencing Core Facility, University of Michigan, Ann Arbor, MI).

NT-OSCP120 →
1 10 20 30 40 50
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPTLFAKLVRPPVQIYGIE

51 60 70 80 90 100
GRYATALYSAASKQNKLEQVEKELLRVGQILKEPKMAASLLNPYVKRSVK

101 110 120 130 140 150
VKSLSDMTAKEKFSPLTSNLINLLAENGRLTNTPAVISAFSTMMSVHRGE

CT-OSCP145 →
1 10 20 30 40 50
FAKLVRPPVQIYGIEGRYATALYSAASKQNKLEQVEKELLRVGQILKEPK

51 60 70 80 90 100
MAASLLNPYVKRSVKVKSLSDMTAKEKFSPLTSNLINLLAENGRLTNTPA

101 110 120 130 140 150
VISAFSTMMSVHRGEVPCTVTTASALDETTLTELKTVLKSFLSKGQKGNS

151 160 170
KLEGKPIPNPLLGLDSTRTGHHHHHH

OSCP190 →
1 10 20 30 40 50
FAKLVRPPVQIYGIEGRYATALYSAASKQNKLEQVEKELLRVGQILKEPK

51 60 70 80 90 100
MAASLLNPYVKRSVKVKSLSDMTAKEKFSPLTSNLINLLAENGRLTNTPA

101 110 120 130 140 150
VISAFSTMMSVHRGEVPCTVTTASALDETTLTELKTVLKSFLSKGQVLKL

151 160 170 180 190
EVKIDPSIMGGMIVRIGEKYVDMSAKTKIQKLSRAMREIL

Figure A1.2. Sequences of OSCP120, OSCP145, and OSCP190. Hexa-his tag and linker sequences are shown in blue.

OSCP constructs pCRT7/NT- OSCP120 and pCRT7/CT-OSCP145 were transformed into One Shot BL21(DE3)pLysS (Invitrogen, Carlsbad, CA) using the manufacturer's procedure. Individual colonies were picked immediately to inoculate 1-10 mL LB overnight cultures supplemented with ampicillin (200 μg/mL). Cultures were incubated at 37 ^{o}C with rotary shaking at 250 RPM. One-liter cultures containing either Spectra 9-N (>98% ^{15}N), Spectra 9-CN (>98% ^{15}N, >98% ^{13}C, or Spectra 9-dCN (>97% D_2O, >98% ^{15}N, >98% ^{13}C) media (Spectra Stable Isotopes, Columbia, MD), were

supplemented with ampicillin (200 μg/mL) and inoculated with the overnight culture. Cultures were grown at 37 °C with rotary shaking at 250 RPM until A_{595} reached 0.6-1 (Spectra 9-dCN required double the growth time compared to the other media) at which point cultures were removed to 4 °C while the rotary shaker temperature was brought down to 20 °C using a heat exchanger over a period of 1 h. IPTG was added to a final concentration of 0.02 mM. Growth and expression continued for 20 h at 20 °C. Cell pastes were harvested via centrifugation in a Beckman JLA8.1 at 6,000 X *g* for 10 min at 4 °C. Cell pastes were re-suspended in ice-cold Nickel NTA buffer (25 mL; 50 mM Tris-HCl pH 8, 300 mM NaCl, 0.001% phenylmethanesulfonyl fluoride, and Roche Complete protease inhibitors w/o EDTA as per manufacturers instruction). The suspension was transferred to a 50 mL Falcon tube on ice and stored at -80 °C.

Cell paste was thawed at room temperature and placed immediately on ice. The suspension was transferred to a 50 mL beaker on ice and sonicated using the standard horn (set at output = 8.5) for 6 20-s long intervals between which was a 2-min long rest on ice to dissipate heat. After sonication, the mixture was transferred to two Beckman JA35.5 centrifuge tubes and centrifuged at 21,000 RPM at 4 °C for 45 min. The soluble extract was loaded (0.5 mL/min) onto a Ni-NTA affinity column (1 cm diameter X 5 cm long) equilibrated in Nickel NTA Buffer at 4 °C. The column was washed with Nickel NTA Buffer-25 mM imidazole (25 mL) and protein was eluted with Nickel NTA Buffer-250 mM imidazole (25 mL). Fractions of about 1 mL were collected and those from fractions 3-14 were direct placed into snakeskin dialysis tubing (7,000 MWCO, Pierce, Rockville IL) immersed in of OSCP Buffer (4 L; 50 mM Tris-Cl pH 8, 30 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol). Complete Protease inhibitors (with or without EDTA) were added following the manufacturer's instructions, directly to the fractions in the dialysis bag. Following overnight dialysis at 4 °C with stirring, the dialysate was cleared by centrifugation using a Beckman JA 25.5 rotor for 20 min at 21,000 RPM and 4 °C. The

clarified extract was loaded into a ÄKTA prime super loop (50 mL) at 4 ^{o}C and injected at 1.3 mL/min onto a 5 mL HiTrap SpHP column (Pharmacia) equilibrated in at least 5 volumes of modified OSCP Buffer (50 mM Tris-Cl pH 8, 30 mM NaCl, 1 mM EDTA, 0.001% phenylmethanesulfonyl fluoride, Roche Complete protease inhibitors w/o EDTA, 5 mM β-mercaptoethanol; MOB) at 4 ^{o}C. The column was washed with two bed volumes of the same buffer and then eluted with a 75 mL linear gradient (MOB and MOB + 500 mM NaCl) at 1 mL/min while collecting 1 mL fractions. OSCP 120NT eluted in a sharp band centered around 250 mM NaCl while OSCP 145CT eluted in a similar manner at 280 mM NaCl.

A1.2.4 NMR Spectroscopy

All NMR experiments were performed at 25 °C unless indicated otherwise using an Avance Bruker 600 MHz spectrometer equipped with a 5 mm triple-resonance cryogenic probe. NMR spectra were processed and analyzed using NMRPipe and SPARKY 3(18,19) The NMR buffer consisted of 90/10% H_2O/D_2O containing 50 mM Tris, 5 mM KCl, 5 mM β-mercaptoethanol, 0.001% PMSF and protease inhibitor cocktail at pH ~7. For cross-relaxation experiments, a $^{2}H/^{15}N$ labeled OSCP120 (0.3 mM) sample was used in an NMR buffer consisting of 90/10% D_2O/H_2O containing deuterated TRIS (50 mM) and deuterated β-mercaptoethanol (5 mM) at pH ~7 without PMSF and protease inhibitor cocktail to minimize spectral overlap with **1**. 93% of the OSCP120 backbone amides could be assigned based on a previous NMR study (BMRB entry 6564) (Figure A1.3).(20) The OSCP120 resonances in OSCP145 were assigned by overlaying spectra and using standard triple resonance experiments on a doubly labeled ($^{13}C/^{15}N$) OSCP145 sample (0.5 mM).

Two-dimensional ^{1}H-^{15}N HSQC spectra of ^{15}N labeled OSCP were recorded following incremental addition of **1** from a stock solution (20 mM) in NMR buffer for

protein:ligand ratios of 1:1, 1:2, 1:4, and 1:8. Weighted average chemical shift perturbations were calculated using,

$$\Delta_{av}NH = \sqrt{\frac{(\delta H)^2 + (\delta N/5)^2}{2}},$$

[A1.1]

where δH and δN are the proton and nitrogen amide chemical shift values in ppm. Weighted average chemical shifts >0.01 ppm were considered significant.

Cross-relaxation experiments were performed at 14 ^{o}C to improve cross-relaxation efficiency by modifying a 2D ^{1}H-^{15}N HSQC experiment from the Bruker pulse program library (hsqcfpf3gpphwg)(21). A 1.8 s adiabatic Wurst pulse was applied following a pre-

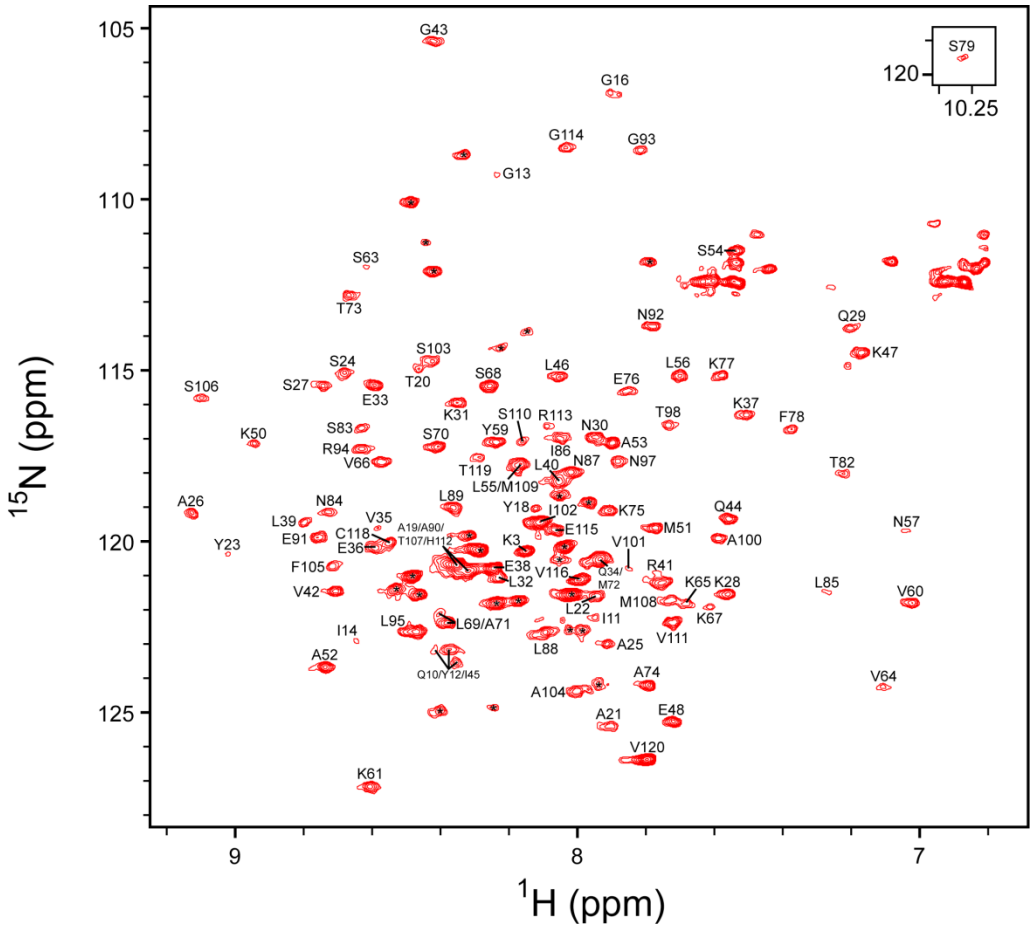


Figure A1.3. Spectrum of free OSCP120 with resonance assignments taken from the Biological Magnetic Resonance Databank (BMRB Entry 6564). A (*) indicates a resonance resulting from the hexa-his tag linker sequence.

delay of 1.4 s for saturating one of three different ligand proton resonances (protons 7, 10, and 24 see Figure 1) with a total saturation bandwidth of ~30 Hz(22). An additional reference experiment was also recorded with off-resonance (-50 ppm) saturation. Each experiment required ~23 h of acquisition time. Peak intensity errors were calculated using NMRPipe and ranged between 4-19%(18,19). Cross-relaxation intensity ratios were calculated by taking the ratio of peak intensities measured with on-resonance (I_{sat}) and off-resonance (I_o) saturation. Only well resolved resonances with a signal:noise ratio >10 were analyzed. Chemical shift perturbations and cross-relaxation results were mapped onto OSCP structures and visualized using Pymol.(23)

A1.3 Bz-423 Specifically Binds to the Shoulder Region of OSCP

In a first group of studies, we attempted chemical shift perturbation experiments using full-length, ^{15}N labeled OSCP (OSCP190) (Figure A1.2). However, the 2D HSQC spectra of OSCP190 were intractable due to aggregation identified by uniform reduction in NMR signal intensities(16,20). We subsequently prepared two truncated constructs for binding measurements. The first protein comprised residues 1-120 (OSCP120) and has been studied previously by NMR, and the second construct contained residues 1-145 (OSCP145). Titration of **1** into ^{15}N labeled OSCP120 at protein:ligand ratios of 1:1, 1:2, 1:4, and 1:8 led to chemical shift perturbations that suggest rapid exchange on the NMR timescale consistent with μM affinity, which agreed with previous data for Bz-423 (Figures 1 and 2 and Table A1.1)(4). To assess the significance of the OSCP chemical shift perturbations, a second set of titration experiments were conducted using an analog lacking the naphthalene substituent (**2**). This analog possesses no activity in the F_1F_0-ATPase or cell based assays (data not shown) and therefore we hypothesized should not to bind to the protein. Indeed, chemical shift perturbations were not observed in titration experiments with **2** (Figure A1.4), indicating that the perturbations observed with **1** reflect specific binding interaction(s) that presumably are related to its inhibitory activity.

The chemical shift perturbations induced by **1** were distributed at different sites within the OSCP120 construct (Figure A1.4). This observation suggests that **1** either binds at multiple sites and/or binding at one site causes conformational changes elsewhere in the protein. The range of intensities observed in OSCP120 suggests that the protein is highly flexible and potentially prone to ligand-induced (allosteric) conformational changes. In addition, previous studies of Bz-423 suggest a 1:1 stoichiometry upon binding to the F_1F_0-ATPase(4). Collectively, these observations are consistent with ligand-induced conformational changes away from the binging site.

Table A1.1: Chemical shift perturbations for OSCP120 and OSCP145 where (--) represents unmeasurable resonances, (NS) represents no significant chemical shift, and (S) indicates a significant shift greater than the threshold cutoff. A (*) indicates residues that had similar free resonance positions in free OSCP120 and OSCP145, which were used to compare titration data. Cross-relaxation data for OSCP120 is also shown where (--) represents unmeasurable resonances, (NR) represents no significant intensity reduction, and (R) indicates a reduction in intensity greater than the threshold cutoff of 0.8.

Residue	120 CSP	145 CSP	120 CR
2ALA	--	--	--
3LYS	**NS**	--	**NR**
4LEU	--	**NS**	--
5VAL	--	--	--
6ARG	--	--	--
9VAL	--	--	--
10GLN	--	--	--
11ILE	--	--	--
12TYR	--	**S**	--
13GLY	--	**S**	--
14ILE	--	--	**NR**
15GLU	--	--	--
16GLY	**NS**	**S**	--
17ARG	--	--	--
18TYR	**NS**	--	--
19ALA	--	--	--
20THR	**S**	**S**	--
21ALA	--	**NS**	--
22LEU	**S**	--	--

23TYR	--	**S**	--	
24SER	**NS**	**NS**	--	
25ALA	**S**	**S**	--	
26ALA	**NS**	**NS**	--	
27SER	**NS**	**S**	--	
28LYS	**NS**	**NS**	**R**	
29GLN	**NS**	**NS**	--	*
30ASN	**NS**	**NS**	--	*
31LYS	**NS**	**S**	**R**	*
32LEU	--	**NS**	--	*
33GLU	**NS**	**NS**	**NR**	*
34GLN	--	--	--	
35VAL	--	--	--	
36GLU	**S**	**S**	**NR**	
37LYS	**NS**	**NS**	--	*
38GLU	**NS**	--	--	
39LEU	**NS**	**NS**	--	
40LEU	**S**	**S**	--	
41ARG	**NS**	**S**	**R**	
42VAL	**NS**	**NS**	--	*
43GLY	**NS**	**NS**	--	*
44GLN	**NS**	**NS**	--	*
45ILE	--	--	--	
46LEU	**NS**	**S**	--	*

47LYS	**S**	**S**	**R**	*
48GLU	**NS**	**S**	--	*
50LYS	**NS**	**NS**	**R**	*
51MET	**S**	**S**	**NR**	*
52ALA	**NS**	**S**	**R**	*
53ALA	**NS**	**NS**	**NR**	*
54SER	**NS**	**NS**	--	*
55LEU	--	**S**	--	
56LEU	**S**	**S**	--	*
57ASN	--	**NS**	--	*
59TYR	**NS**	**NS**	**NR**	*
60VAL	**NS**	**NS**	**R**	*
61LYS	**NS**	**NS**	**NR**	*
62ARG	--	--	--	
63SER	**NS**	**NS**	--	
64VAL	--	**S**	--	*
65LYS	**S**	**S**	--	*
66VAL	**S**	**S**	**R**	*
67LYS	**NS**	**S**	--	*
68SER	**NS**	**NS**	**NR**	*
69LEU	--	--	--	
70SER	**NS**	**NS**	**NR**	*
71ASP	--	--	--	
72MET	--	--	--	

73THR	**NS**	**S**	--	*
74ALA	**NS**	**NS**	--	*
75LYS	**S**	**S**	--	*
76GLU	**NS**	**NS**	--	*
77LYS	**S**	**NS**	**NR**	*
78PHE	**NS**	**NS**	--	
79SER	**NS**	**S**	**R**	
81ILE	--	--	--	
82THR	**S**	**NS**	--	
83SER	**S**	**S**	**NR**	
84ASN	**NS**	--	--	
85LEU	--	--	--	
86ILE	**NS**	--	**R**	
87ASN	**NS**	--	**R**	
88LEU	**S**	**NS**	**NR**	
89LEU	**NS**	**NS**	--	
90ALA	--	--	--	
91GLU	**NS**	**S**	--	
92ASN	**S**	**S**	--	
93GLY	**NS**	**NS**	**NR**	
94ARG	**NS**	**S**	--	
95LEU	**NS**	**NS**	**NR**	*
97ASN	**NS**	**NS**	--	
98THR	**NS**	**S**	--	

100ALA	**NS**	**NS**	--	
101VAL	--	**S**	--	
102ILE	**S**	--	--	*
103SER	**NS**	**NS**	--	*
104ALA	**NS**	**S**	--	*
105PHE	**S**	**S**	--	
106SER	**NS**	**NS**	--	*
107THR	--	--	--	
108MET	**S**	--	--	
109MET	--	**NS**	--	
110SER	**S**	**S**	**NR**	
111VAL	**S**	--	**R**	
112HIS	--	**NS**	--	
113ARG	**S**	--	**NR**	
114GLY	**NS**	**NS**	**NR**	
115GLU	**NS**	--	**NR**	
116VAL	**S**	**NS**	--	
118CYS	**S**	**NS**	--	
119THR	**S**	**S**	--	
120VAL	**NS**	**S**	**NR**	

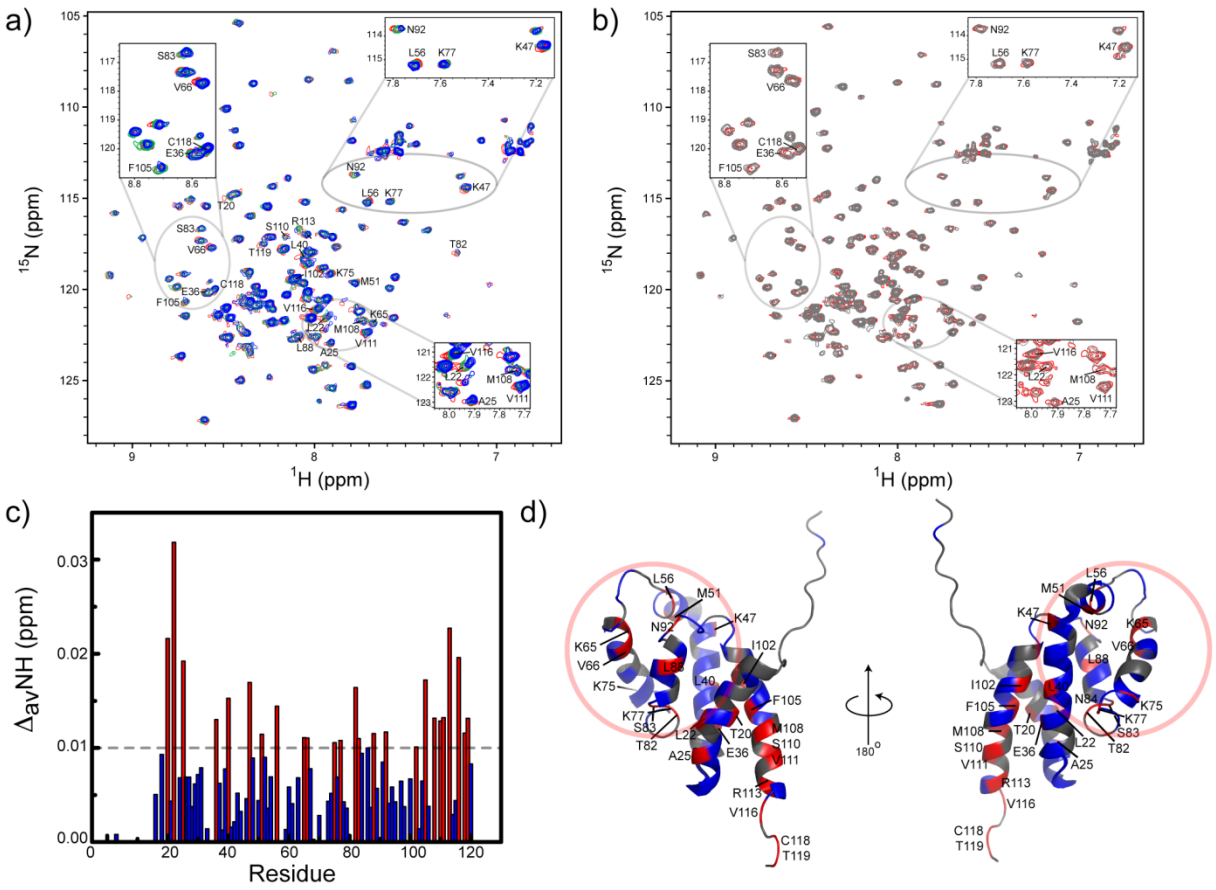


Figure A1.4. Chemical shift perturbations from titrations of **1** and **2** onto OSCP120 are shown in (a) and (b), respectively. Free OSCP120 (red), 1:2 OSCP120:**1** (24), 1:4 OSCP120:**1** (blue), and 1:4 OSCP120:**2** (gray) spectra are shown. Only resonances that exhibited significant chemical shift perturbations are labeled in (a), and peaks labeled in (b) are for comparison purposes. Weighted average of chemical shift perturbations (c) are color coded to match coloring on the OSCP120 structure (d) with blue representing resonances that showed no significant perturbation and red showing those resonances with shifts greater than the threshold of 0.01 ppm. Residues that could not be monitored are colored in gray and the shoulder region is circled in (d). Protein precipitation was observed at 1:8 OSCP:ligand ratios, which precluded further increasing the ligand concentration.

The chemical shift perturbations caused by **1** fall in three general regions of OSCP120; one is located between helices III, IV, and V ("shoulder") and includes residues M51, L56, K65, V66, K75, K77, T82, S83, and N92; the second is located at the C-terminal tails of helices I and VI ("tail region") and includes residues V111, R113, V116, C118, and T119; a third potential locus is located between the "tail" and "shoulder"

regions (Figure A1.5). Interestingly, the latter region includes residues A35 and L88, which were previously shown to interact directly with the α peptide mimicking the F_1 domain of the F_1F_0-ATPase(16,20).

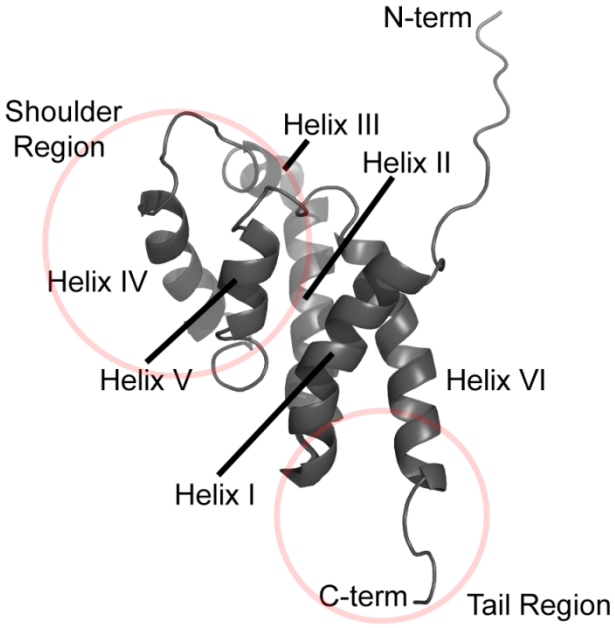


Figure A1.5. The structure of OSCP120 (PDB ID# 2BO5) is shown with helices, termini, shoulder, and tail regions labeled.

To confirm the data obtained with OSCP120, we used a second OSCP construct comprising 145 amino acids (OSCP145). Unlike the OSCP190, this truncated protein did not aggregate and was amenable to chemical shift perturbation experiments with **1**. Comparing the OSCP120 and OSCP145 spectra reveals that the majority of the OSCP120 resonances are not altered due to the additional 25 amino acids in the OSCP145 construct (Figure A1.6). Not surprisingly, residues that showed significant differences between the two constructs were primarily located at the N- and C- termini. However, chemical shift differences were observed at helix V, which interacts with the F_1 peptide(16). This region, which also undergoes perturbations with **1**, may be subject to

allosteric conformational perturbations that could be important in mediating OSCP interactions with the F_1F_0-ATPase F_1 domain(4).

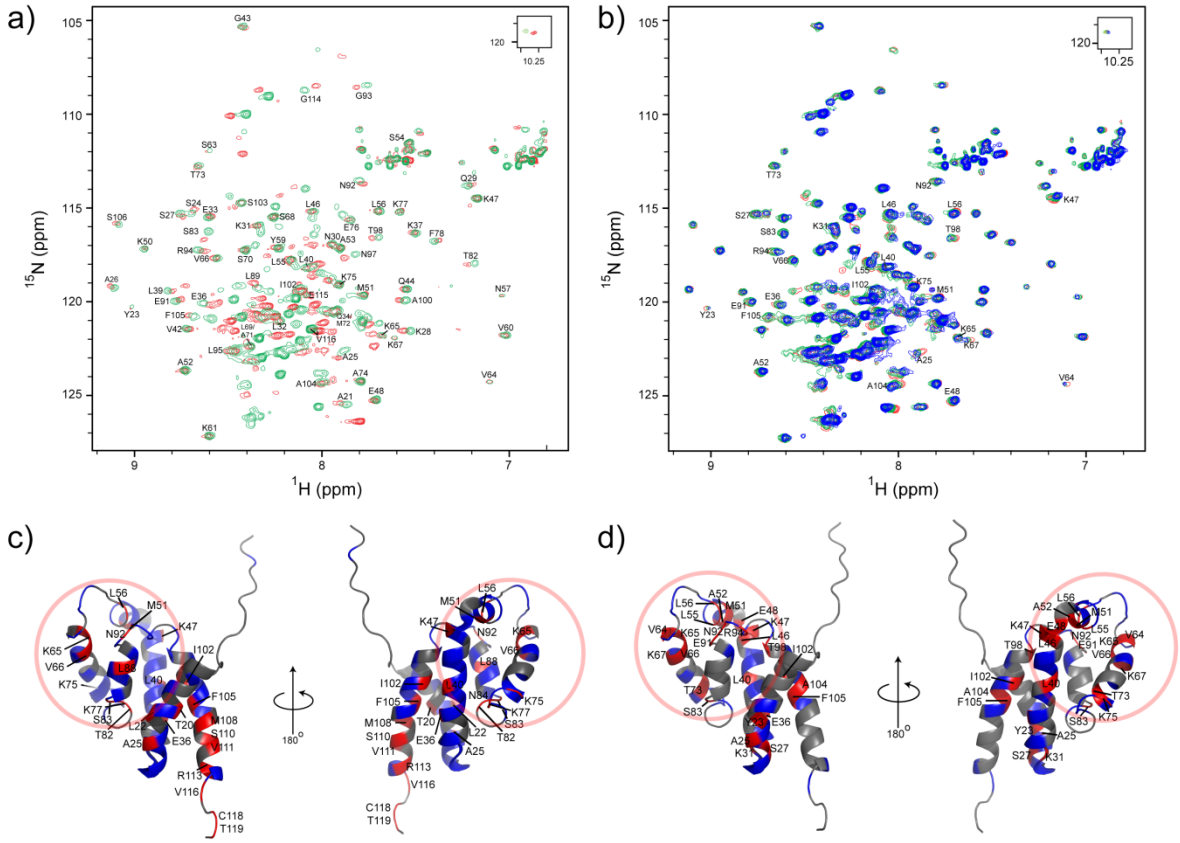


Figure A1.6. The overlay of OSCP120 (red) and OSCP145 (24) is shown in (a) with overlapping residues labeled. Spectra from the titration of **1** onto OSCP145 (b) show chemical shift perturbations (labeled) of similar residues between OSCP120 and OSCP145. Chemical shift perturbations from the titration of **1** on OSCP120 (c) and OSCP145 (d) show that the shoulder region is common in both sets of data and thus is likely the biologically relevant binding site. In (c) and (d), residues with weighted average chemical shift perturbations greater than the threshold cutoff of 0.01 are shown in red, and amino acids showing no significant weighted average chemical shift perturbations are colored in blue. Residues that could not be measured are colored in gray and the shoulder region is circled in (c) and (d).

Titration with **1** induced chemical shift perturbations in OSCP145 whereas **2** did not (Figure A1.7). Several resonances from the shoulder (M51, L56, K65, V66, K75, and N92) and middle region of the protein (A35, E36, L40, K47, I102, and F105) that overlaid in OSCP120 and OSCP145 showed similar perturbations with **1,** including A35, which

interacts with the F_1 helix (Figure A1.6). Many residues in the shoulder region (E48, A52, K67, T73, S79, S83, E91, and R94) also showed larger perturbations with **1** in OSCP145 compared to OSCP120. By contrast, some of the perturbations observed in the "tail" (residues V116 and C118), middle (L88), and shoulder (T82) regions were less pronounced in OSCP145. Taken together, these data indicate that the shoulder region is the most probable binding site for **1** on OSCP.

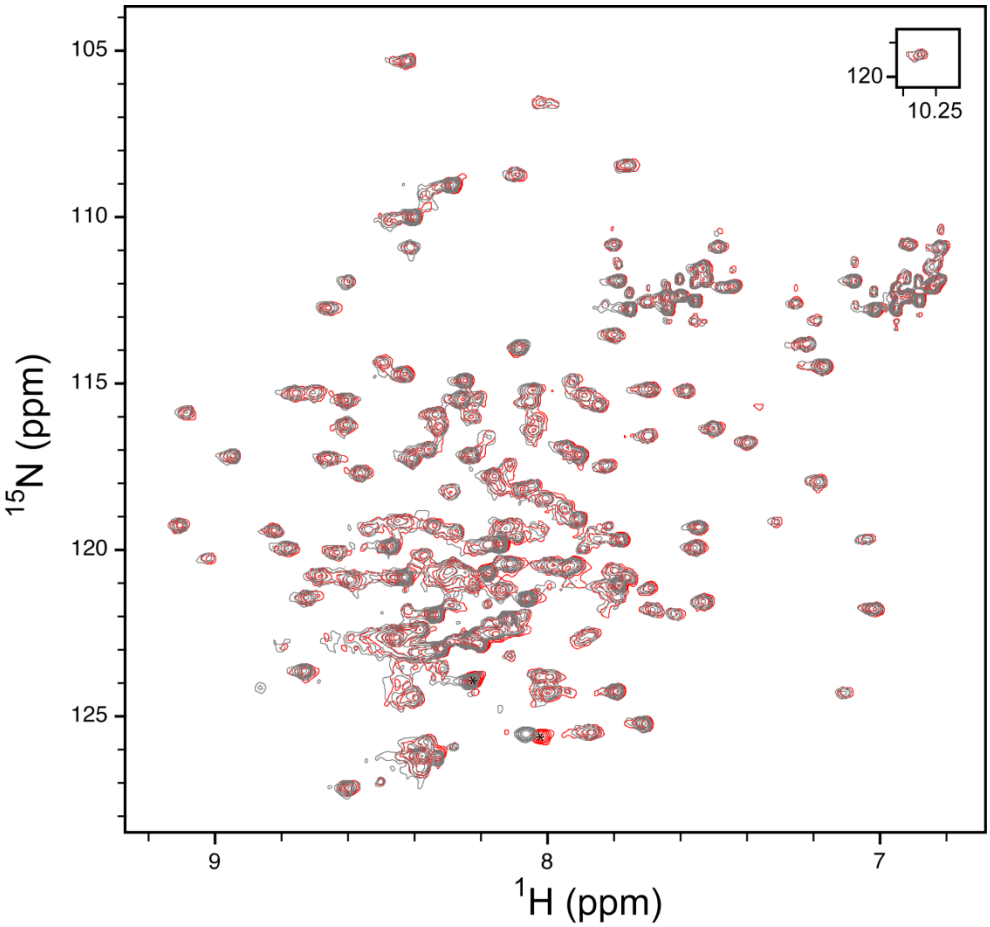


Figure A1.7. Overlay of free OSCP145 (red) and titration of **2** at a protein:ligand ratio of 1:4 (gray). The magnitude of chemical shift perturbations is null compared to titrations with **1**. A (*) indicates peaks that were identified as C-terminal hexa-his tag resonances that perturb upon titration with **2** but are not relevant in the biologically native form of OSCP.

Since chemical shift perturbations can arise from binding and/or conformational changes, cross-relaxation experiments were employed to localize the binding site of **1** on

OSCP120(25) For these studies, 2D HSQC spectra of OSCP following saturation of specific ^{1}H resonances in **1**, and an equivalent reference spectrum with off-resonance (-50 ppm) saturation were measured. A reduction in OSCP resonance intensities in the on-resonance (I_{sat}) versus off-resonance (I_o) experiment indicates cross-relaxation processes due to proximity of ligand protons to OSCP amide protons resulting from a specific binding interaction.

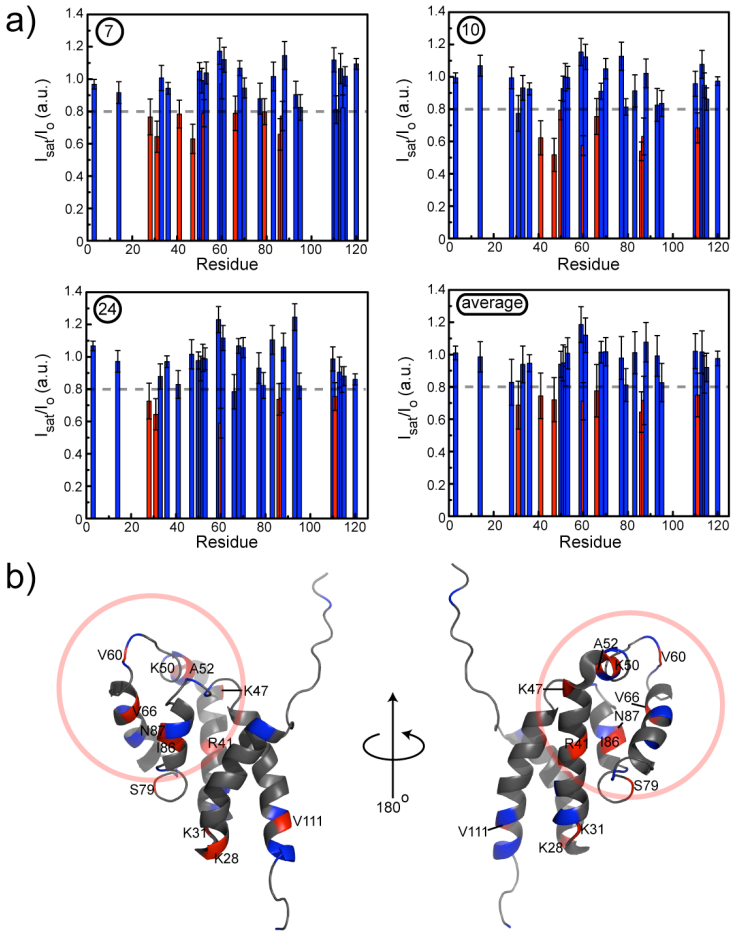


Figure A1.8. Intensity ratios (I_{sat}/I_o) from cross-relaxation experiments saturating protons at positions (7), (10), and (24) and their average is shown in (a). The 32 residues that were measured are color coded according to intensity reductions with those showing no significant reduction in intensity colored in blue and resonances showing intensity reductions greater than the threshold cutoff of 0.8 colored in red. The same color scheme is used in (b) with labeled residues being those that showed significant intensity reductions in any of the three sets of cross-relaxation data. Residues that could not be monitored are colored in gray and the shoulder region is circled in (b).

Although the cross-relaxation experiment is inherently insensitive since it recorded in 90% D_2O to avoid saturation of the water resonance, and despite the limited solubility of OSCP, which further limited sensitivity, 32 well-resolved resonances with sufficient signal:noise were observed in the OSCP120•**1** complex. As shown in Figure A1.8, significant intensity reductions (>20%) are observed for a number of residues. The majority of these residues (R41, K50, A52, V60, V66, S79, I86, and N87) fall in the "shoulder" region. Cross-relaxation to residues in the tail region (K28, K31 and V111) was also observed and V111 corresponds to a residue for which significant chemical shift perturbations upon titration of **1** were measured. However, this site is most likely cryptic and only present in the truncated OSCP constructs since the perturbations are diminished or otherwise significantly altered in OSCP145. Taken together, the chemical shift perturbation and cross relaxation studies are consistent with a unique binding site for **1** on the OSCP.

A1.4 Conclusion

Inhibitors of the mitochondrial F_1F_0-ATPase are powerful tools for probing the structure and function of the enzyme and like Bz-423, some have therapeutic potential. A diverse group of molecules inhibit the enzyme by binding within the F_1 domain(26-28). Representative compounds here include the antibiotics aurovetrin and efrapeptin, phytochemicals like resveratrol, and the naturally occurring peptide inhibitor, IF1. Aurovetrin binds to β_{TP} and β_E states of the enzyme and is thought to function by preventing closure of the interfaces necessary for catalytic cycling between subunit conformational states(4,29-32). Efrapeptin interacts with the γ and β_E subunits where it can block re-charging after catalysis(29). Similarly, resveratrol inhibits the F_1F_0-ATPase by binding between the γ and β_{TP} subunits where it can block the rotation of the γ subunit so that the catalytic cycle cannot progress(33). IF1 inhibits ATP hydrolysis when the

ability of the enzyme to synthesize ATP is compromised. IF1 is thought to function by hindering the closure of the α_{DP}-β_{DP} catalytic interface thereby blocking ATP hydrolysis(34). Several molecules such as the macrolide oligomycin inhibit the F_1F_0-ATPase by binding to the embedded F_0 domain(8). Biochemical studies suggest that oligomycin binding blocks the flow of protons through the c-subunit.

Here we used NMR spectroscopy to investigate binding of a water-soluble Bz-423 analog to OSCP constructs of varying length in an effort to define the binding site for these inhibitory benzodiazepines on the protein. The chemical shift perturbation data on both OSCP120 and OSCP145 together with cross-relaxation data for OSCP120 suggest that **1** most likely binds to the shoulder region in a pocket defined by residues M51, L56, K65, V66, K75, K77, and N92. The peak intensities of unbound OSCP120 suggest that both the shoulder and tail regions are more flexible relative to the rest of the protein, potentially allowing for local rearrangements to occur upon ligand binding (data not shown). While we cannot rule out a possible secondary binding site in OSCP120 involving the tail region, this interaction was diminished in OSCP145, indicating that interactions at this site are not likely relevant and probably result as a consequence of the truncation only present in the 120-amino acid long OSCP construct.

Several residues show chemical shift perturbations in the titration experiments but no significant cross-relaxation, which is most consistent with conformational changes on binding to **1**. These residues include A35 and L88, which are thought to directly interact with the $F_{1\ (16,35)}$. L88 along with other residues in its vicinity also shift when comparing free OSCP120 and OSCP145, even though the 25 amino acids are added to the C-terminal end of OSCP145. Based on the NMR structure of OSCP120, L88 is distant from the additional 25 amino acids. These data further support the hypothesis that this region of the OSCP is susceptible to conformational changes, which may be

important for the function of the protein in the fully assembled enzyme and the activity of Bz-423 like inhibitors.

Based on our data, we propose an allosteric model for the inhibitory action of **1**. Binding of **1** to the shoulder region of OSCP results in conformational rearrangements at a site that contacts the F_1 domain of F_1F_0-ATPase, which interferes with the rotary mechanism of catalysis (Figure A1.9). Allosteric communication between **1** and F_1 binding sites may be achieved by relative twisting motions of helices I and VI. Such a structural rearrangement could explain the widespread perturbations observed in the middle region linking **1** and the F_1 binding site.

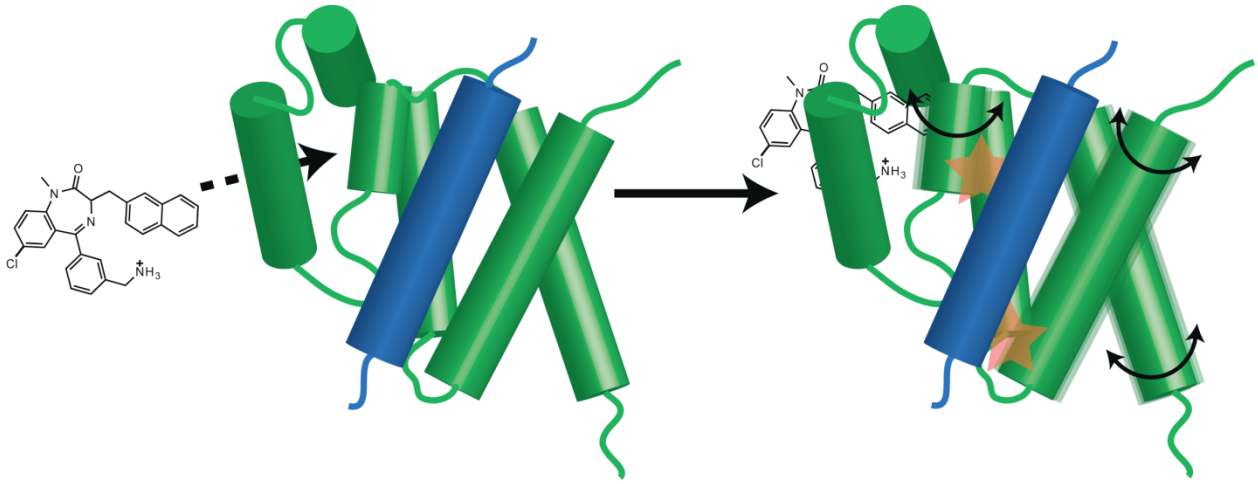


Figure A1.9. Proposed allosteric model where binding of **1** causes conformational rearrangements (at helices I, V, and VI) in OSCP thus altering OSCP-F_1 interactions. Stars indicate site for the OSCP-F_1 peptide interaction that were perturbed during titration experiments with **1**.

Another group of benzodiazepines have been reported that are structurally similar to Bz-423(24). Unlike Bz-423, these compounds selectively inhibit ATP hydrolysis catalyzed by the mitochondrial F_1F_0-ATPase and may have use as anti-ischemia drugs. Preliminary studies suggest that these benzodiazepines also function through binding to the OSCP(36). Together, these data highlight the important role the OSCP plays in regulating both the synthetic and hydrolytic the function of the

mitochondrial F_1F_0-ATPase. A better understanding how these compounds interact with the OSCP and inhibit the enzyme should assist in further elucidating the function of the OSCP and may also provide additional opportunities for drug discovery.

This work was published in the journal Biopolymers(37). Experimental Design was conceived by Al-Hashimi H. M. and Stelzer A. C. OSCP was synthesized by Frazee R. Stelzer A. C. collected and analyzed the NMR data.

A1. 5 References

1. Bednarski, J.J., Warner, R.E., Rao, T., Leonetti, F., Yung, R., Richardson, B.C., Johnson, K.J., Ellman, J.A., Opipari, A.W. and Glick, G.D. (2003) Attenuation of autoimmune disease in Fas-deficient mice by treatment with a cytotoxic benzodiazepine. *Arthritis and Rheumatism*, 48, 757-766.
2. Blatt, N.B., Bednarski, J.J., Warner, R.E., Leonetti, F., Johnson, K.M., Boitano, A., Yung, R., Richardson, B.C., Johnson, K.J., Ellman, J.A. *et al.* (2002) Benzodiazepine-induced superoxide signals B cell apoptosis: mechanistic insight and potential therapeutic utility. *Journal of Clinical Investigation*, 110, 1123-1132.
3. Johnson, K.M., Chen, X.N., Boitano, A., Swenson, L., Opipari, A.W. and Glick, G.D. (2005) Identification and validation of the mitochondrial F1F0-ATPase as the molecular target of the immunomodulatory benzodiazepine Bz-423. *Chemistry & Biology*, 12, 485-496.
4. Johnson, K.M., Cleary, J., Fierke, C.A., Opipari, A.W. and Glick, G.D. (2006) Mechanistic basis for therapeutic targeting of the mitochondrial F1F0-ATPase. *Acs Chemical Biology*, 1, 304-308.
5. Blatt, N.B., Boitano, A.E., Lyssiotis, C.A., Opipari, A.W. and Glick, G.D. (2008) Bz-423 superoxide signals apoptosis via selective activation of JNK, Bak, and Bax. *Free Radical Biology and Medicine*, 45, 1232-1242.
6. Blatt, N.B., Boitano, A.E., Lyssiotis, C.A., Opipari, A.W., Jr. and Glick, G.D. (2009) Bz-423 superoxide signals B cell apoptosis via Mcl-1, Bak, and Bax. *Biochem Pharmacol*.
7. De Milito, A., Iessi, E., Logozzi, M., Lozupone, F., Spada, M., Marino, M.L., Federici, C., Perdicchio, M., Matarrese, P., Lugini, L. *et al.* (2007) Proton pump inhibitors induce apoptosis of human B-cell tumors through a caspase-independent mechanism involving reactive oxygen species. *Cancer Research*, 67, 5408-5417.
8. Devenish, R.J., Prescott, M., Boyle, G.M. and Nagley, P. (2000) The oligomycin axis of mitochondrial ATP synthase: OSCP and the proton channel. *Journal of Bioenergetics and Biomembranes*, 32, 507-515.

9. Pedersen, P.L., Ko, Y.H. and Hong, S.J. (2000) ATP syntheses in the year 2000: Evolving views about the structures of these remarkable enzyme complexes. *Journal of Bioenergetics and Biomembranes*, 32, 325-332.
10. Senior, A.E., Nadanaciva, S. and Weber, J. (2002) The molecular mechanism of ATP synthesis by F1F0-ATP synthase. *Biochimica Et Biophysica Acta-Bioenergetics*, 1553, 188-211.
11. Walker, J.E. and Collinson, I.R. (1994) The Role of the Stalk in the Coupling Mechanism of F1f0-Atpases. *Febs Letters*, 346, 39-43.
12. Rubinstein, J.L. and Walker, J.E. (2002) ATP synthase from Saccharomyces cerevisiae: Location of the OSCP subunit in the peripheral stalk region. *Journal of Molecular Biology*, 321, 613-619.
13. Walker, J.E., Fearnley, I.M., Gay, N.J., Gibson, B.W., Northrop, F.D., Powell, S.J., Runswick, M.J., Saraste, M. and Tybulewicz, V.L.J. (1985) Primary Structure and Subunit Stoichiometry of F1-Atpase from Bovine Mitochondria. *Journal of Molecular Biology*, 184, 677-701.
14. Wilkens, S., Dunn, S.D., Chandler, J., Dahlquist, F.W. and Capaldi, R.A. (1997) Solution structure of the N-terminal domain of the delta subunit of the E-coli ATPsynthase. *Nature Structural Biology*, 4, 198-201.
15. Boitano, A., Emal, C.D., Leonetti, F., Blatt, N.B., Dineen, T.A., Ellman, J.A., Roush, W.R., Opipari, A.W. and Glick, G.D. (2003) Structure activity studies of a novel cytotoxic benzodiazepine. *Bioorganic & Medicinal Chemistry Letters*, 13, 3327-3330.
16. Carbajo, R.J., Kellas, F.A., Yang, J.C., Runswick, M.J., Montgomery, M.G., Walker, J.E. and Neuhaus, D. (2007) How the N-terminal domain of the OSCP subunit of bovine F1F0-ATP synthase interacts with the N-terminal region of an alpha subunit. *Journal of Molecular Biology*, 368, 310-318.
17. Walker, J.E., Gay, N.J., Powell, S.J., Kostina, M. and Dyer, M.R. (1987) ATP Synthase from Bovine Mitochondria - Sequences of Imported Precursors of Oligomycin Sensitivity Conferral Protein, Factor-6, and Adenosine-Triphosphatase Inhibitor Protein. *Biochemistry*, 26, 8613-8619.
18. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of Biomolecular Nmr*, 6, 277-293.
19. Goddard, T.D. and Kneller, D.G. *SPARKY 3*. University of California, San Franscisco.
20. Carbajo, R.J., Kellas, F.A., Runswick, M.J., Montgomery, M.G., Walker, J.E. and Neuhaus, D. (2005) Structure of the F-1-binding domain of the stator of bovine F1Fo-ATPase and how it binds an alpha-subunit. *Journal of Molecular Biology*, 351, 824-838.
21. Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-Tailored Excitation for Single-Quantum Nmr-Spectroscopy of Aqueous-Solutions. *Journal of Biomolecular Nmr*, 2, 661-665.
22. Kupce, E. and Wagner, G. (1995) Wideband homonuclear decoupling in protein spectra. *Journal of Magnetic Resonance Series B*, 109, 329-333.
23. Delano, L. The Pymol Molecular Graphics System. *Delano Scientific, San Carlos, CA, USA*.
24. Hamann, L.G., Ding, C.Z., Miller, A.V., Madsen, C.S., Wang, P., Stein, P.D., Pudzianowski, A.T., Green, D.W., Monshizadegan, H. and Atwal, K.S. (2004) Benzodiazepine-based selective inhibitors of mitochondrial F1F0 ATP hydrolase. *Bioorg Med Chem Lett*, 14, 1031-1034.

25. Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. and Shimada, I. (2000) A novel NMR method for determining the interfaces of large protein-protein complexes. *Nature Structural Biology*, 7, 220-223.
26. Gledhill, J.R. and Walker, J.E. (2005) Inhibition sites in F-1-ATPase from bovine heart mitochondria. *Biochemical Journal*, 386, 591-598.
27. Hong, S. and Pedersen, P.L. (2008) ATP Synthase and the Actions of Inhibitors Utilized To Study Its Roles in Human Health, Disease, and Other Scientific Areas. *Microbiology and Molecular Biology Reviews*, 72, 590-641.
28. Toogood, P.L. (2008) Mitochondrial drugs. *Current Opinion in Chemical Biology*, 12, 457-463.
29. Abrahams, J.P., Buchanan, S.K., vanRaaij, M.J., Fearnley, I.M., Leslie, A.G.W. and Walker, J.E. (1996) The structure of bovine F-1-ATPase complexed with the peptide antibiotic efrapeptin. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 9420-9424.
30. Duser, M.G., Zarrabi, N., Cipriano, D.J., Ernst, S., Glick, G.D., Dunn, S.D. and Borsch, M. (2009) 36 degrees step size of proton-driven c-ring rotation in F(o)F(1)-ATP synthase. *EMBO J*.
31. vanRaaij, M.J., Abrahams, J.P., Leslie, A.G.W. and Walker, J.E. (1996) The structure of bovine F-1-ATPase complexed with the antibiotic inhibitor aurovertin B. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 6913-6917.
32. Johnson, K.M., Swenson, L., Opipari, A.W., Jr., Reuter, R., Zarrabi, N., Fierke, C.A., Borsch, M. and Glick, G.D. (2009) Mechanistic basis for differential inhibition of the F(1)F(o)-ATPase by aurovertin. *Biopolymers*, 91, 830-840.
33. Gledhill, J.R., Montgomery, M.G., Leslie, A.G.W. and Walker, J.E. (2007) Mechanism of inhibition of bovine F-1-ATPase by resveratrol and related polyphenols. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 13632-13637.
34. Campanella, M., Casswell, E., Chong, S., Farah, Z., Wieckowski, M.R., Abramov, A.Y., Tinker, A. and Duchen, M.R. (2008) Regulation of mitochondrial structure and function by the F1Fo-ATPase inhibitor protein, IF1. *Cell Metab*, 8, 13-25.
35. Boyer, P.D. (1997) The ATP synthase - A splendid molecular machine. *Annual Review of Biochemistry*, 66, 717-749.
36. Johnson, K.M. *PhD Thesis, University of Michigan, 2005.*
37. Stelzer, A.C., Frazee, R.W., Van Huis, C., Cleary, J., Opipari, A.W., Glick, G.D. and Al-Hashimi, H.M. NMR studies of an immunomodulatory benzodiazepine binding to its molecular target on the mitochondrial F(1)F(0)-ATPase. *Biopolymers*, 93, 85-92.

Appendix 2

SAS Methodology

A2.1 Design of "NMR invisible" elongated RNA

The elongation of an RNA target is typically preformed by extending the length of a terminal helix using a stretch of Watson-Crick base-pairs(1). To avoid increasing spectral overlap due to elongation residues, an isotopic labeling strategy is used to render elongation residues "NMR invisible"(1). Here, two constructs are prepared in which stretches of alternating unlabelled A–U/U-A (E-AU-RNA) and G–C/C-G (E-GC-RNA) base-pairs are employed for elongation using otherwise uniformly $^{13}C/^{15}N$ labeled G–C and A–U nucleotides, respectively(1). Two G-C base pairs are added to the terminal end of E-AU-RNA to facilitate RNA synthesis by *in vitro* transcription. Structure prediction software such as mfold 3.3 are used to ensure that no alternate structures form as the result of elongation (34). The two constructs thus allow acquisition of NMR data over the entire RNA target while keeping elongation residues "NMR invisible".

One must ensure that the elongation does not perturb the structural and functional integrity of the RNA. This can be conveniently done by comparing spectra of elongated and non-elongated RNA samples. In general, elongation of terminal helices is not expected to give rise to significant RNA structural perturbations. However, depending on the RNA context, elongation of other non-terminal helices, which can allow the

measurement of independents set of RDCs (and RCSAs) that can be applied to generate structural dynamics with enhanced spatial resolution (2), may cause unwanted perturbations and should be carefully analyzed.

The degree of helix elongation needed to sufficiently decouple internal and overall motions will vary depending on the RNA target. In general, the elongation should render a target helix at least 4-5-fold longer than other helices in the RNA. If a structure for the RNA target is available, one can perform simulations using programs such as PALES(3), using idealized A-form helices to model the elongated helix(2), to examine the degree of motional coupling. In these simulations, one computes the variance in the predicted overall alignment tensor relative to the elongated helix as a function of varying the relative orientation of other shorter helices in the RNA. In general, the principal direction of order (S_{zz}) should not vary more than ±7 degrees about the elongated axis and the asymmetry (η) should always be <0.15. The secondary structure of an E-RNA construct should be verified using a secondary structure prediction programs such as mfold 3.3 to make sure that no alternative secondary structures can form(4). The E-RNA NMR samples (> 0.2 mM) are prepared using standard *in vitro* transcription reactions employing the appropriately $^{13}C/^{15}N$ labeled nucleotides(5,6). Formation of the elongated helix can be verified using NMR(1). First, one expects to observe an intense ^{1}H signal corresponding to the chemically degenerate Watson-Crick hydrogen bonded imino protons of guanines and uridines used in the elongation in 1D ^{1}H spectra. Second, one expects to observe signals that are characteristic of sequential $^{13}C/^{15}N$ enriched terminal GC base-pairs in the elongated helix, which are included to enhance transcription efficiency.

A2.2 Partial alignment of E-RNA

The measurement of RDCs in solution NMR is contingent upon inducing an appropriate degree of alignment typically on the order of 10^{-3}(7). These levels of alignment can now be routinely achieved by dissolving biomolecules in inert ordered media (for reviews see(8,9)) that transmit some of their order to solute molecules through mechanisms that are believed to involve a combination of steric obstruction and charge-charge interactions. The most popular and commercially available ordering medium is Pf1 phage(10,11). Relative to other media, Pf1 phage is tolerant to the high salt concentrations used in nucleic acid samples and is negatively charged thus reducing the possibility for adverse inter-molecular interactions. Pf1 phage is available commercially or can be prepared using the methods described in reference(10). Typically, a Pf1 solution is exchanged into NMR buffer by repeated (at least three) rounds of ultracentrifugation (1hr in a Beckman TLA-100.3 rotor at 95,000 rpm, or 3-6 hrs in a VTi50 rotor at 40,000 rpm) followed by re-suspension of the pellet into the NMR buffer. Alternatively, one can dialyze Pf1 phage into the desired buffer. After completing the aligned experiments, the same ultracentrifugation procedure can in principle be used to recover the nucleic acid (supernatant) from the phage solution (pellet). Note that it will generally be difficult to achieve perfect separation of the nucleic acid sample from the phage medium.

Owing to their large structural anisotropy, the concentration of Pf1 phage needed to align E-RNA samples (~6-8 mg/ml) is usually considerably smaller than for non-elongated RNA (~20-25 mg/ml). If a model structure for the RNA is available, programs for predicting steric alignment such as PALES(3) can be used to assess relative levels of order and the Pf1 phage concentration be adjusted accordingly. The aligned RNA sample is typically prepared by adding a pre-concentrated RNA solution (~0.5-1.5 mM) in NMR buffer to a desired volume of Pf1 phage (50 mg/mL) in NMR buffer in an Eppendorf tube and the sample gently transferred into the NMR tube avoiding formation

of bubbles. It is important to verify that the ordering medium does not interfere with the RNA conformation by comparing chemical shifts obtained in the unaligned and aligned samples. Note that small variations in the chemical shifts of nucleobase carbons and nitrogens are expected between unaligned and aligned samples due to incomplete averaging of Residual Chemical Shift Anisotropies (RCSAs) (12-14). These RCSA contributions scale linearly with the magnetic field and degree of order.

It should be noted that an alternative approach for aligning nucleic acids involves spontaneous alignment due to interactions with the magnetic field itself. Although the degree of field induced alignment remains smaller than optimal, there are reasons to believe that optimal levels will inevitably be reached as larger nucleic acids are investigated and magnetic field strengths continue to rise. This approach is not described here and the reader can consult reviews on this topic.

A2.3 Measurement of RDCs in E-RNA

A large number of pulse sequences have been reported for the measurement of a wide variety of RDCs in nucleic acids. These experiments are not described in detail here as they have been reviewed elsewhere.(52) The RDCs are generally computed from the difference in splittings observed in the absence (J) and presence of alignment media (J+D). For large E-RNA, it is important to employ TROSY schemes for measuring RDCs in the nucleobases(15). For example, CH splittings in the nucleobases are measured as the difference between the upfield and downfield components of the ^{1}H – ^{13}C doublet along the ^{1}H dimension using the narrow TROSY component in the ^{13}C dimension. For E-RNA, the most commonly targeted RDCs are those between directly bonded C-H and N-H nuclei (e.g. C2H2/C8H8 of the adenine and guanine bases, C5H5/C6H6 of the uracil and cytosine bases, N1H1 and N3H3 of the guanine and uracil bases, and C1'H1' of the ribose), which yield the largest RDC magnitudes. While

additional one, two, and three bond RDCs can also be measured, these RDCs are smaller and may prove difficult to measure in larger E-RNAs (>60 nt).

When using frequency domain experiments to measure splittings, phase distortions due to improper calibration of timing delays and/or imperfections in shaped pulses can yield splitting measurement errors that are larger than theoretical limits (approximately given by 0.7*Linewidth*(1/Signal:Noise))(16). To avoid differential contributions from magnetic field induced RDCs and relaxation interference effects, splittings in unaligned and aligned samples should be measured at the same magnetic field strength. It is advisable to estimate the experimental RDC uncertainty from the standard deviation in duplicate measurements. Resonances exhibiting significant differences (>3σ) as a result of considerable broadening, overlap, presence/absence of unresolved multiplets should not be used in subsequent data analysis.

A2.4 Normalizing RDCs measured in distinct E-RNA samples

The RDCs measured in the different E-AU and E-GC samples need to be normalized to take into account possible differences in the degree of alignment before the data can be pooled together and analyzed in constructing atomic-resolution ensembles(2). The normalization is carried out using RDCs measured in contiguous Watson-Crick base-pairs, which can be modeled assuming an idealized A-form geometry (see below). The errors introduced due to A-form structural noise can be estimated as described previously(53). The normalization is accomplished by repeatedly fitting the total pool of RDCs to individual idealized A-form helices following uniform scaling of the RDCs measured in one sample (typically the sample with the lesser number of measured RDCs) by a normalization factor *L*. The *L* value that minimizes the Quality factor (Q)(17) is computed using(2):

$$Q = \sqrt{\frac{\sum w_i \left(D_i^{calc} - D_i^{\exp}\right)^2}{\sum w_i \left(D_i^{\exp}\right)^2}} \quad (4)$$

Fitting of the RDCs to the A-form helices can be accomplished using various programs including ORDERTEN-SVD(18), REDCAT(19), PALES(3), iDC(20), CONFORMIST(21) and RAMAH(12).

A2.5 Determining the overall order tensor

The interpretation of E-RNA RDCs is greatly simplified by determining the overall order tensor governing alignment. The overall order tensor can be determined using RDCs measured in the reference elongated helix using a procedure that has been described in detail elsewhere(22,23). The procedure is briefly outlined below:

1. Build idealized A-form helices (PDB files) corresponding to the sequence of the reference helix. For example, to build an A-form helix using the Biopolymer module of Insight II 2000.1 (Molecular Simulations, Inc): click on the module icon in the upper left corner and select *append* from the nucleotide menu. In the pop-up box, select "A_RNA_Duplex". Input a name for the molecule into the text field. Next, select the appropriate Watson-Crick base-pair in the *Nucleotide* text field. Continue to append base-pairs – following along in sequence from 5' to 3' – until you have completed building the desired helix. Click cancel and then select the *File* menu and choose the desired export option for the helix coordinates. The idealized A-form helices should conform to published parameters(23-25). If building helices using INSIGHT II 2000.1 (Molecular Simulations, Inc.), care

needs to be taken to correct the propeller twist angles to the proper value of -14.5^0. We have a program named HPmod to correct for the improper propeller twist (available from Author upon request). Programs such as Curves 5.1(26), FreeHelix98(27), 3DNA(25,28) , SCHNAaP(29), NUPARM and NUCGEN(30) can be used to compute relevant helix parameters.

2. Compute five order tensor elements for each A-form helix by fitting the RDCs to the A-form PDB coordinates. Several programs are available to carry out such calculations including ORDERTEN-SVD(18), REDCAT(19), PALES(3), iDC(20), CONFORMIST(21) and RAMAH(12). Note, non-ideal WC base-pairs are excluded from this analysis. In our lab RAMAH is used to determine the five order tensor elements.
3. Examine the correlation between measured and back-calculated RDCs. Major outliers should be interrogated for possible measurement errors.
4. Use AFORM-RDC(23) or other approaches(31) to estimate the order tensor error due to structural noise and RDC measurement uncertainty.

Owing to the uniform distribution of charge in polyanionic nucleic acids, the steric and electrostatic forces are believed to have a similar functional form(32,33). Consequently, E-RNAs are expected to align in ordering media with the principal direction of order (S_{zz}) oriented along the elongated axis. In general, one expects positive alignment (S_{zz}> 0) with the S_{zz} direction being, on average, oriented parallel to the magnetic field(2). The asymmetry of alignment is also expected to be nearly axially symmetric (η~0)(2).

A2.6 Molecular Dynamics (MD) simulations of RNA

MD simulations are used in this case to generate a large conformational pool that can be filtered using experimental RDCs so as to generate a more accurate ensemble of

RNA conformers. A variety of MD simulation packages and force-fields can be employed to simulate nucleic acids. Here we describe a protocol for simulating RNA using the CHARMM MD package(34) with force-field parameter set 27 for nucleic acids(35).

A) System Preparation using MMTSB Toolkit: To prepare the system for simulation with CHARMM, the MMTSB Toolset (http://blue11.bch.msu.edu/mmtsb/Main_Page)(36) is used. The Toolset consists of a set of perl scripts that can be used to prepare, initiate and analyze a system for MD simulation. Here they are only described in terms of preparing a system for an MD simulation:

   i. Obtain coordinates for the RNA from the Protein Data Bank (PDB) (http://www.pdb.org) or the Nucleic Acid Database (NDB) (http://ndbserver.rutgers.edu/). Use the MMTSB toolset script *convpdb.pl* to add solvent and ions to the system using the command: *convpdb.pl -solvate -cutoff* 15 *-cubic -ions* SOD:27 hivtar-0.pdb > hivtar-0-solvated-ions.pdb. In this example -solvate flag indicates that pdb solvent molecules will be added,
   -cutoff 15 specifies the distance (15 Å) from the RNA to the edge of the solvent cube, -cubic specifies a solvent shape (alternatively a octahedron water box could be specified with -octahedron flag), -ions SOD:27 specifies that 27 sodium ions be added to make the system charge neutral, and hivtar-0.pdb indicates the input pdb file.

   ii. Use the MMTSB toolset script *genPSF* to generate protein structure file (PSF) and CHARMM formatted coordinate file using the command: *genPSF.pl -par* nodeoxy *-crdout* hivtar-0-solvated-ions.cor hivtar-0-solvated-ions.pdb > hivtar-0-solvated-ions.psf. Here the *-par* nodeoxy flag specifies that the system is an RNA molecule, *-crdout* and hivtar-0-solvated-ions.cor specifies the filename for the output CHARMM formatted coordinate, and is hivtar-0-solvated-ions.psf is the output PSF file.

B) Equilibration using CHARMM

i) Read in parameter and topology file set 27 for nucleic acids

ii) Read in generate PSF and coordinates files

iii) Do 500 steps Steepest Descent (SD) minimization to remove bad contacts using the CHARMM MINI SD module

iv) Place harmonic constraints on heavy atoms of the RNA using *CONS HARM* (see CHARMM's cons.doc manual) and perform 1000 steps of SD minimization, followed by 2000 steps of Adopted Basis Newton-Raphson (ABNR) minimization or until energy converges (see CHARMM's minimiz.doc manual). This process allows solvent and counter-ions to relax around the RNA.

v) The energy minimized structure is heated up to 300 K, by carrying out a series of constant temperature simulations starting at 0 K for ~20 ps at each temperature. The temperature is increased by 25 K each incremental time-step using coordinates from the previous run as input for the next 20 ps simulation. We usually use the Nosé Hoover thermostat with a coupling constant of 100 ps^{-1} together with the Velocity Verlet (VVER) integrator with a integration time-step of 1 fs. Periodic boundary conditions are used, with electrostatics calculated using particle-mesh Ewald and Lenard-Jones interaction truncated at 12 Å, with a switching function applied between 10 to 12 Å.

vi) At 300 K the harmonic constraint is gradually removed by decreasing the force constant.

vii) Equilibration is continued at 300 K and without any harmonic constraints for an additional 500 ps.

viii) Generation of trajectory: Generation of structural ensembles from an MD derived pool of conformations hinges on the ability to adequately sample the relevant regions of conformational space. It has been shown that using multiple short trajectories may enhance the rate of conformational sampling(37,38) when compared to a single long trajectory of same effective length, which is advantageous to our methods since we use experimental data to filter unrealistic conformations. In this protocol we will describe the use of multiple short trajectories to generate such selection pools.

ix) Initiate a set of independent MD constant temperature replica trajectories starting from the same equilibrated structure generated in part-A. We have found 50 replicas to be sufficient. The thermostat, integrator and non-bonded energy calculation options are same as stated above.

x) Assign each replica a different set of initial velocities by specifying a different seed value for the random number generator used to assign initial velocities using the ISEED input flag into DYNA.

xi) Generate trajectories of about ~ 3 ns, while saving conformations each 1 ps.

xii) Discard the first 1 ns and pool the remaining ~ 2 ns trajectories for each replica to create a selection pool. In this case we generated ~100,000 conformations. Use this pool for selection of structural ensembles as described below.

A2.7 Combine NMR RDCs and MD to generate a structural ensemble – Sample and Select

he Sample and Select method was originally implemented as a tool to generate structural ensembles of proteins using a combination of MD and NMR derived NH spin relaxation order parameters (S^2)(39). The basic idea is to sample the relevant

conformational space and then select a sub-ensemble that is most consistent with the experimental data. We adapted this method to use RDCs measured on E-RNA to select structural ensembles of RNA from a selection pool derived from an MD simulation(40). The SAS selection is preformed following the recently described procedure(40) in which one minimizes a cost function utilizing a Monte-Carlo simulated annealing approach:

i) Randomly select an N-membered subset of structures from the total pool of M structures and calculate χ^2 using $\chi^2 = L^{-1}\sum_{i}^{L}\left(D_{ij}^{cal} - D_{ij}^{\exp}\right)^2$, where D_{ij}^{cal} and $D_{ij}^{\exp}$ are the calculated and measured RDCs, respectively, and *L* is the total number of bond vectors.

ii) Randomly replace one of the N-membered structures with a randomly chosen structure from the remaining M-N conformational pool. Accept the move for step *k* to *k*+1 if $\chi^2(k+1) < \chi^2(k)$. If $\chi^2(k+1) > \chi^2(k)$, accept the move with a probability $P = e^{\left((\chi^2(k)-\chi^2(k+1))/T_i\right)}$, where T_i is the effective temperature. We typically use a linear cooling schedule, specifically $T_i+1 = 0.9T_i$, where the index i runs over the temperature increments.

iii) Continue Monte-Carlo annealing simulations until χ^2 converges. In our simulations with M = 80000, N=20 and L=82, convergence was achieved within 100 temperature increments with each consisting of 100,000 MC steps.

Once an ensemble is constructed, it is important to have the means to establish its validity. This can be done using experimental data that is not included in the selection process. In addition to RDCs, future studies can also include RCSAs(43, 45), NOEs, as well as spin relaxation order parameters(69, 70), or data from other experimental techniques such as SHAPE foot printing(71) data and fluorescence(72). The constructed

ensembles can also be compared with expected parameters. For example, average and standard deviations for various base angles have been reported for Watson-Crick base-pairs that are flanked by other Watson-Crick base-pairs in A-form helices(53). Last but not least, the ensemble can be interpreted in light of known mechanistic aspects of the RNA function.

A2.8 References

1. Zhang, Q., Sun, X., Watt, E.D. and Al-Hashimi, H.M. (2006) Resolving the motional modes that code for RNA adaptation. *Science*, **311**, 653-656.
2. Zhang, Q., Stelzer, A.C., Fisher, C.K. and Al-Hashimi, H.M. (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature*, **450**, 1263-1267.
3. Zweckstetter, M. and Bax, A. (2000) Predicition of sterically induced alignment in a dilute liquid crystalline phase; aid to protein structure determination by NMR. *J . Am. Chem. Soc.*, **122**, 3791-3792.
4. Zuker, M. (2000) Calculating nucleic acid secondary structure. *Current Opinion in Structural Biology*, **10**, 303-310.
5. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide Synthesis Using T7 RNA-Polymerase and Synthetis DNA Templates. *Nucleic Acids Research*, **15**, 8783-8798.
6. Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of Small RNAs using T7-RNA- Polymerase. *Methods in Enzymology*, **180**, 51-62.

7. Tjandra, N. (1999) Establishing a degree of order: obtaining high-resolution NMR structures from molecular alignment. *Structure With Folding & Design*, **7**, R205-R211.
8. Prestegard, J.H. and Kishore, A.I. (2001) Partial alignment of biomolecules: an aid to NMR characterization. *Curr Opin Chem Biol*, **5**, 584-590.
9. Tolman, J.R. and Al-Hashimi, H.M. (2003) In Webb, G. A. (ed.), *Annual Reports on NMR Spectroscopy*. Academic Press, Vol. 51, pp. 105-166.
10. Hansen, M.R., Mueller, L. and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Structural Biology*, **5**, 1065-1074.
11. Clore, G.M., Starich, M.R. and Gronenborn, A.M. (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod- shaped viruses. *Journal of the American Chemical Society*, **120**, 10571-10572.
12. Hansen, A.L. and Al-Hashimi, H.M. (2006) Insight into the CSA tensors of nucleobase carbons in RNA polynucleotides from solution measurements of residual CSA: towards new long-range orientational constraints. *J Magn Reson*, **179**, 299-307.
13. Ottiger, M., Tjandra, N. and Bax, A. (1997) Magnetic field dependent amide N-15 chemical shifts in a protein-DNA complex resulting from magnetic ordering in solution. *Journal of the American Chemical Society*, **119**, 9825-9830.
14. Ying, J., Grishaev, A. and Bax, A. (2006) Carbon-13 chemical shift anisotropy in DNA bases from field dependence of solution NMR relaxation rates. *Magn Reson Chem*, **44**, 302-310.
15. Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1997) Attenuated T-2 relaxation by mutual cancellation of dipole- dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 12366-12371.
16. Kontaxis, G., Clore, G.M. and Bax, A. (2000) Evaluation of cross-correlation effects and measurement of one- bond couplings in proteins with short transverse relaxation times. *Journal of Magnetic Resonance*, **143**, 184-196.
17. Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquide crystalline phase. *J. Am. Chem. Soc.*, **120**, 6836-6837.
18. Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of Magnetic Resonance*, **138**, 334-342.
19. Valafar, H. and Prestegard, J.H. (2004) REDCAT: a residual dipolar coupling analysis tool. *J Magn Reson*, **167**, 228-241.
20. Wei, Y.F. and Werner, M.H. (2006) iDC: A comprehensive toolkit for the analysis of residual dipolar couplings for macromolecular structure determination. *Journal of Biomolecular Nmr*, **35**, 17-25.
21. Skrynnikov, N.R., Goto, N.K., Yang, D., Choy, W.Y., Tolman, J.R., Mueller, G.A. and Kay, L.E. (2000) Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. *J Mol Biol*, **295**, 1265-1273.
22. Bailor, M.H., Musselman, C., Hansen, A.L., Gulati, K., Patel, D.J. and Al-Hashimi, H.M. (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat Protoc*, **2**, 1536-1546.

23. Musselman, C., Pitt, S.W., Gulati, K., Foster, L.L., Andricioaei, I. and Al-Hashimi, H.M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J Biomol NMR*, **36**, 235-249.
24. Neidle, S. (1999) *Oxford Handbook of Nucleic Acid Structur*e. Oxford University Press, New York.
25. Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X., Neidle, S., Sakked, Z., Sklenar, H. *et al.* (2001) A Standard Refernce Frame for the Description of Nucleic Acid Base-pair Geometry. *Journal of Molecular Biology*, **313**, 229-237.
26. Ravishanker, G., Swaminathan, S., Beveridge, D.L., Lavery, R. and Sklenar, H. (1989) Conformational and Helicoidal Analysis of 30 ps of Molecular-Dynamics on the D(CGCGAATTCGCG) Double Helix - Curves, Dials and Windows. *Journal of Biomolecular Structure & Dynamics*, **6**, 669-699.
27. Dickerson, R.E. (1998) DNA bending: The prevalence of kinkiness and the virtues of normality. *Nucleic Acids Research*, **26**, 1906-1926.
28. Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*, **31**, 5108-5121.
29. Lu, J. (1997) Collectins: collectors of microorganisms for the innate immune system. *BioEssays*, **19**, 509-518.
30. Bansal, M., Bhattacharyya, D. and Ravi, B. (1995) NUPARM AND NUCGEN - Software for Analysis and Generation of Sequence-Dependent Nucleic-Acid Structures. *Computer Applications in the Biosciences*, **11**, 281-287.
31. Zweckstetter, M. and Bax, A. (2002) Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *J Biomol NMR*, **23**, 127-137.
32. Wu, B., Petersen, M., Girard, F., Tessari, M. and Wijmenga, S.S. (2006) Prediction of molecular alignment of nucleic acids in aligned media. *Journal Of Biomolecular Nmr*, **35**, 103-115.
33. Zweckstetter, M., Hummer, G. and Bax, A. (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophys J*, **86**, 3444-3460.
34. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) CHARMM - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry*, **4**, 187-217.
35. MacKerell, A.D., Banavali, N. and Foloppe, N. (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, **56**, 257-265.
36. Feig, M., Karanicolas, J. and Brooks, C.L. (2002), *223rd National Meeting of the American-Chemical-Society*, Orlando, Fl, pp. 377-395.
37. Auffinger, P. and Westhof, E. (1996) H-bond stability in the tRNA(Asp) anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophys J*, **71**, 940-954.
38. Caves, L.S., Evanseck, J.D. and Karplus, M. (1998) Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci*, **7**, 649-666.
39. Chen, Y., Campbell, S.L. and Dokholyan, N.V. (2007) Deciphering protein dynamics from NMR data using explicit structure sampling and selection. *Biophys J*, **93**, 2300-2306.
40. Frank A.T., Stelzer A.C., Al-Hashimi H.M., and Andricoaei I. (2009) Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR

RDCs: new insights into RNA dynamics and adaptive ligand recognition. *Nuc Acids Res*, 37, 3670-3679.