Decoding the Dynamics that Guide Selection of TAR RNA Conformation

By

Jeremy David Kratz

A thesis submitted for partial fulfillment of Bachelor of Science Honors Degree.

(Cellular and Molecular Biology Department)

University of Michigan

2011

Thesis Committee:

Professor Hashim Al-Hashimi

Professor James Bardwell

Professor Andrzej Wierzbicki

Acknowledgements

To my grandfather, Gerald B. Kratz, and all educators

with innate enthusiasm for their pupils.

Abstract

The information encoded in ribonucleic acids (RNA) is as an important modulator of cellular function through structure and distinct conformational changes that facilitate the regulation of transcription, cellular signaling, and catalysis[1]. The Human Immunodeficiency Virus (HIV) is a retrovirus that infects the immune system's T cells, leading to the Autoimmune Deficiency Syndrome (AIDS) and ultimately death. Currently approved HIV therapeutics target proteins important in fusion of viral particles with the host cell, the reverse transcription of the viral RNA genome into DNA, and enzymes necessary to cleave proteins into mature functionally active forms[2]. It has long been known that the 5' Long Terminal Repeat (LTR) Trans-Activation Response (TAR) RNA element is vital to the lifecycle of HIV as a regulator of viral transcription[3]. Therapeutic advances have been developed to inhibit the interaction of TAR RNA with the regulatory protein Tat, preventing efficient transcription of new viral transcripts[4]. No therapeutics that target TAR RNA have advanced to clinical fruition, likely due to limitations in an understanding of the structural recognition of TAR, a flexible target that samples an expansive structural landscape. In this thesis, it is shown that characterizing the structural landscape of TAR using both a computational and experimental methods provide a deep understanding of the sequence dependent dynamics of TAR that guide Tat recognition and affords the discovery of novel therapeutics that target unique conformations over the entire structural landscape.

The native dynamics of the highly flexible TAR RNA was studied using point mutation of a single base pair. Using a simple A-U to G-C mutation, the local and global dynamic characteristics of the Trans-Activation Response element RNA was rationally altered over timescales extending up to milliseconds. This mutant (TAR^{GC}) was characterized using Nuclear

Magnetic Resonance (NMR) spectroscopy of isotopically labeled RNA. Experimental parameters used to understand the altered dynamics of TAR RNA included Residual Dipolar Coupling (RDC) measurements, in combination with chemical shift and normalized intensity measurements (peak height) for both base and sugar residues.

It was found that this single mutation biased the constructs conformation towards the ligand-bound state of arganinamide (ARG), a small molecule mimic of arginine residues in the cognate protein Tat[5], upon initial investigation of peak chemical shift. Upon further investigation it was found that the residual dipolar couplings from TAR^{GC} in the free-state correlated with TAR+ARG (R = 0.92) versus free TAR (R = 0.82). In addition, it was found that both constructs converge to a similar ligand-bound state upon saturation with argininamide and there was a corresponding decrease in free energy required to overcome the thermodynamic barrier to reach the ligand-bound state for TAR^{GC5}.The mutant, biased towards the bound conformation was found to bind ARG with slightly enhanced affinity using a conformation indistinguishable from the wild-type sequence (R = 0.99). By understanding the importance of sequence junctions, in addition to recent work in defining the topology of RNA[6], a new framework emerges in which the dynamics of RNA molecules can be designed with the distinct properties necessary for performing functionally important roles.

A recent study has refined a pool of some 80,000 structures from computational molecular dynamics simulations using experimental Nuclear Magnetic Resonance data that probe motions on timescales up to milliseconds[7]. Using the top 20 structures from this pool, this structural ensemble was validated using computational docking. These structures were found to recapitulating the free binding energies of all molecules previously known to bind TAR RNA. The quantitative prediction of binding energies for small molecules that bind entirely different

RNA conformations was found. By docking small molecules onto a RNA dynamic ensemble constructed by joint application of experimental domain-elongation NMR spectroscopy[8] and computational molecular dynamics (MD)[7], over 50,000 small molecules were virtually screened that target the entire structural landscape of TAR RNA[9] from the human immunodeficiency virus.

By targeting this structural ensemble of RNA conformations virtually, six novel molecules were discovered; each validated using *in vitro* assays. Unique features of the docking predicted models from computation match experimental data are apparent when monitoring heteronuclear chemical shift changes upon site-specific binding involving site specific residues using NMR spectroscopy (chemical shift). The *de novo* molecules bind TAR with near record affinity (K_d as low as 55 nM), while also inhibiting the interaction between HIV-1 TAR and Tat peptide *in vitro* with inhibition constants (K_is) ranging between 710 nM–169 μM. Netilmicin, the only molecule to show therapeutic promise in inhibiting HIV viral replication, displayed unique selectivity for HIV-1 TAR versus other stem-loop RNA construct including the prokaryotic 16S A-Site RNA, the Rev Response Element, TAR derived from HIV-2, and an additional TAR construct lacking all bulge residues. In comparison to sisomicin, a closely related aminoglycoside that differs from netilmicin by a single ethylene group, the later specifically targets TAR RNA.

Furthermore it was found using an *in vivo* reporter gene assays that netilmicin specifically inhibits Tat-mediated activation of the HIV-1 long terminal repeat by 81% in T cell lines and HIV replication in an HIV-1 indicator cell lines (IC_{50} of ~23.1 μM). In addition, Netilmicin failed to inhibit the expression of HIV-2 long terminal repeat, showing unique selectivity for HIV-1 TAR. Using this method, a new strategy for selectively targeting highly

flexible RNAs using an *In Silico* pipeline emerges. This will advance the discovery of novel therapeutics using a dynamics-based approach that can be extended to larger chemical libraries and other flexible biomolecules.

References

1 Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Curr. Opin. Struc. Biol.* **18**, 321-329 (2008).

2 Greene, W. The brightening future of HIV therapeutics. *Nat. Immunol.* **5**, 867-871 (2004).

3 Bannwarth, S. & Gatignol, A. HIV-1 TAR RNA: The target of molecular interactions between the virus and its host. *Curr. HIV Res.* **3**, 61-71 (2005).

4 Richter, S. N. & Palu, G. Inhibitors of HIV-1 Tat-mediated transactivation. *Curr. Med. Chem.* **13**, 1305-1315 (2006).

5 Stelzer, A. C., Kratz, J. D., Zhang, Q. & Al-Hashimi, H. M. RNA dynamics by design: biasing ensembles towards the ligand-bound state. *Angew. Chem. Int. Edit.* **49**, 5731-5733 (2010).

6 Bailor, M. H., Sun, X. Y. & Al-Hashimi, H. M. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **327**, 202-206 (2010).

7 Frank, A. T., Stelzer, A. C., Al-Hashimi, H. M. & Andricioaei, I. 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 (2009).

8 Zhang, Q., Sun, X. Y., Watt, E. D. & Al-Hashimi, H. M. Resolving the motional modes that code for RNA adaptation. *Science* **311**, 653-656 (2006).

9 Puglisi, J. D., Tan, R. Y., Calnan, B. J., Frankel, A. D. & Williamson, J. R. Conformation of the TAR RNA-arginine complex by NMR-spectroscopy. *Science* **257**, 76-80 (1992).

Table of Contents

Chapter 1

Utilizing Sequence Mutation to Design the Conformation of TAR RNA

An outstanding challenge in the field of structural biology and biophysics is the characterization at the atomic level of biomolecules, as this information yields insight into the structure and function of rationally design biomolecules. In addition to distinct, energetically favorable conformations, many biomolecules possess dynamic properties that allow them to access transient conformational sub-states that facilitate functionally important roles in folding, catalysis, recognition, and signal transduction[1-3]. Much work has looked at designing proteins to fold into predicted 3D conformations[1], while a recent study has explored the links between RNA secondary structure to an understanding of global tertiary structure[2]. The challenges of rationally designing a biomolecule's dynamics are great given the consideration of broad energy landscapes and the complex spatial and temporal dependence that dynamics play on sequence specified structure, particularly for flexible molecules such as ribonucleic acids[4-5].

Using Nuclear Magnetic Resonance (NMR) spectroscopy[6-7] it was recently reported[8] that the transactivation response element (TAR) RNA from HIV-1 codes for a dynamic structure ensemble in which various ligands bind distinct conformers within a broad ensemble by the model of 'conformational selection'. In the free-state of TAR, the two helices are on average highly bent in relation to one another (~28^{o}) and undergo large amplitude global motions about a flexible bulge pivot point, consisting of a three residue UCU bulge and an A22-U40 base-pair junction, which lacks predicted Watson-Crick canonical alignment (**Figure 1**)[6-7, 9]. Upon binding of argininamide (ARG), a ligand widely used to mimic the TAR cognate viral protein and transactivator Tat, TAR undergoes a conformational to coaxial stacked helices, stable A22-

U40 Watson-Crick hydrogen bonding, a base triple (U23•A27-U38), and increased local flexibility of the remaining two bulge residues adopting a looped out conformation (**Figure 1**) [6-8,10].

Our work redesigned the junction of the TAR sequence to bias its structural dynamics towards the more linear ARG bound state without hindering the distinct ability of TAR to interact with ARG. We introduce mutations that played no role in ARG binding but that promoted conformations uniquely observed in the TAR-ARG state. We replaced the flexible A22-U40 base-pair, a confluence point for dynamics both local, with stronger hydrogen bond formation in a GC base-pair (**Figure 1**). We aimed at promoting both a canonical base-pair and furthermore the co-axial helical stacking due to improved (by ~1.2kcal/mol) GC/GC stacking without mutating any residues known to interact with ARG[11].

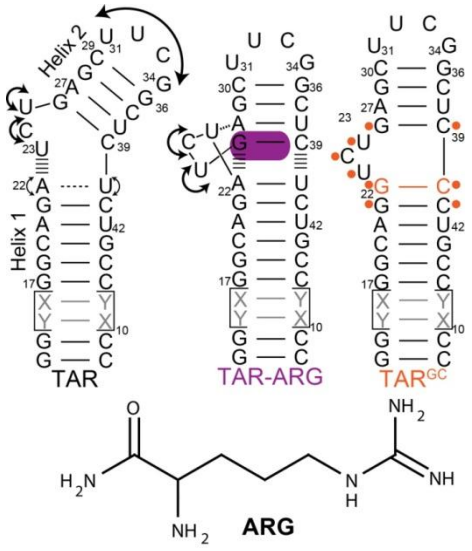


Figure 1 | Biasing the TAR dynamic structure ensemble towards the ligand bound state. The design of TAR^{GC} based on the structural dynamics of free TAR and TAR-ARG bound complex. 'X-Y' base-pairs denote alternating base-pairs used in the elongation strategy[6-7]. Residues undergoing large chemical shift perturbations (CSPs) due to mutation are highlighted by orange circles on the TAR^{GC} structure.

The structural dynamics of the TAR mutant (TAR^{GC}), where the A22-U40 base pair was replaced with G22-C40 (**Figure 1**), were characterized using NMR spectroscopy. The imino spectrum immediately revealed an additional resonance peak that was unambiguously assigned to G22 (**Figure 11**). This finding confirmed that unlike unbound TAR's A22-U40, the G22-C40 mutation forms the expected Watson-Crick canonical alignment in TAR^{GC}. The mutation also resulted in gross changes of the 2D C-H Heteronuclear Single Quantum Coherence (HSQC) spectra of TAR, indicating that mutation affects the structural dynamics of additional residues besides the localized site of mutation (**Figure 2 and 11**). The largest differences were observed for many residues that experience large chemical shift perturbations upon addition of ARG, including the three nucleotide bulge residues and the G26-C39 base-pair directly above the bulge (**Figure 2 and 11**). Further analysis revealed that many of the perturbations induced by the GC mutation directionally similar to those observed upon ARG binding (**Figure 2 and 11**). For example, U23 carbon chemical shift perturbations downfield are consistent with both unstacking of base resonances and coaxial stacking of neighboring helices.

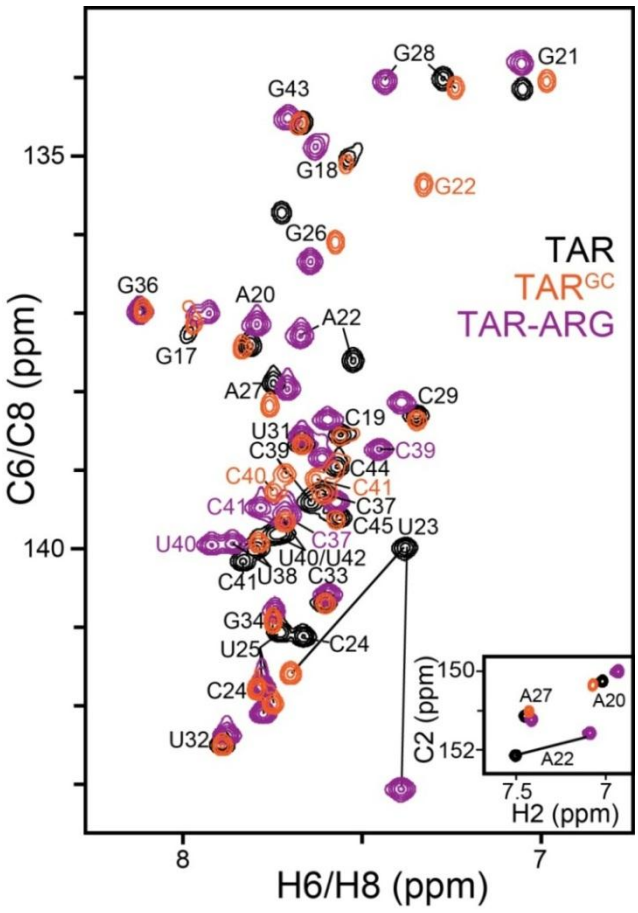


Figure 2 | 2D CH HSQC C6H6 and C8H8 spectra of TAR (black), TAR^{GC} (orange) and TAR-ARG (purple). Inset: 2D CH HSQC C5H5 spectra.

A further investigation into the structural dynamics of TAR^{GC} was conducted in which we measured residual dipolar couplings (RDCs)[12,13] between directly bonded C-H and N-H bonds. These values were then compared to values previously reported for free TAR[12] and ARG bound TAR-ARG[10] (**Figure 3 and Table 1**). RDCs probe both orientation and dynamics of bond vector relative to a molecule-fixed order tensor frame. Their value is sensitive to internal motions occurring over a broad range of timescales (<milliseconds). At many sites the RDCs measured in TAR^{GC} differed significantly from those measured in free TAR (R = 0.82) but strikingly were in much better agreement with those measured in TAR-ARG (R = 0.92) (**Figure 3**). Thus, a single base pair mutation significantly altered the TAR structural dynamics; specifically biasing it on a site specific basis towards the ARG bound conformation.

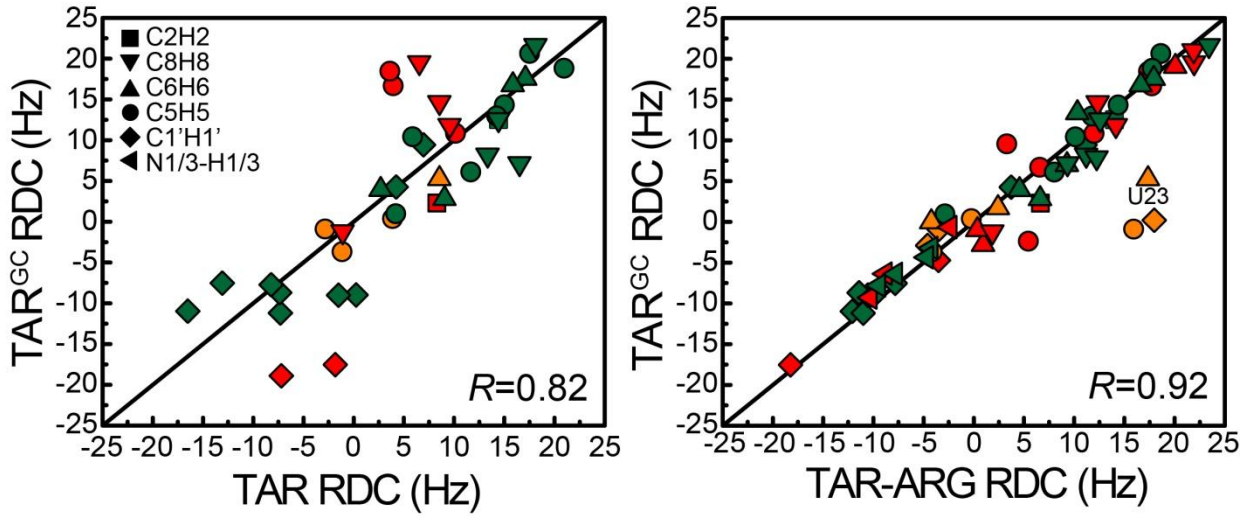


Figure 3 | Correlation plots between RDC values measured in non-elongated TAR, TAR^{GC}, and TAR-bound complex. Residues in helix I, helix II, and the bulge, are colored red, green, and orange respectively with correlation coefficients (R) shown. Also shown are labels specific to base/sugar residue assignment.

In order to quantitatively characterize the structural dynamics upon mutation in TAR^{GC}, we utilized the technique of domain-elongation[6,7] (**Figure 12**) to broaden the timescales sensitive to spin relaxation measurements. This technique reduced couplings between internal motions from overall reorientation, a key feature for quantitatively interpreting RDCs. Ignoring chemical exchange, the resonance intensities from 2D CH HSQC spectra of elongated RNAs report the dynamics of a given site relative to an applied magnetic field occurring at fast picosecond-nanosecond timescales. In unbound helix I elongated TAR (EI-TAR), elevated relative intensities were observed for helix II as a result of collective inter-helical motions, with residues A22, U40, C24 and U25 due to elevated local base/sugar mobility (**Figure 13**)[7]. The binding of ARG arrests the inter-helical motions and local fluctuations involving A22 and U40. In addition, it lead to a dramatic increase in the picosecond to nanosecond dynamics of C24 and U25, which now adopting a distinct looped out conformation (**Figure 13**)[7]. EI-TAR^{GC} resonance intensities differed significantly from those of EI-TAR (R=0.73), while again they appear to be in very good agreement with those of EI-TAR-ARG (R=0.94) (**Figure 4 and 13**). The outlying U23 was found to be more flexible in TAR^{GC}. While U23 adopts a looped out conformation in free TAR^{GC}, it fails to form the U23•A27-U38 base-triple upon ARG binding TAR. Thus, the picosecond-nanosecond dynamic properties of TAR^{GC} bias the local and global properties of TAR towards the ARG bound state.

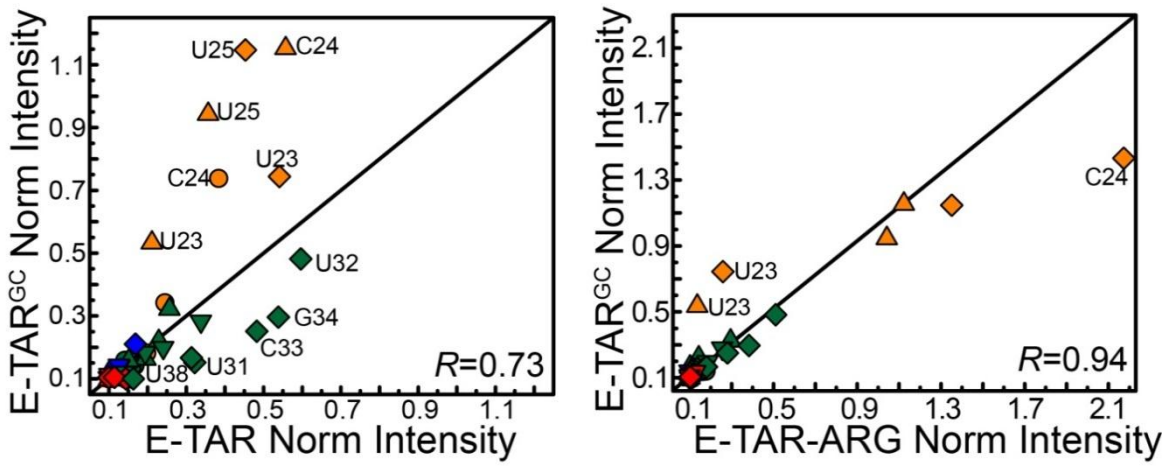


Figure 4 | Correlation between resonance intensities measured in Elongated Helix I constructs of TAR, TAR^{GC} and TAR-ARG.

The TAR^{GC} dynamic structure ensemble was characterized over longer timescales by measuring RDCs in EI-TAR^{GC}, followed by model-free order tensor analysis for values measured in each helical domain (**Table S2**), as described previously for free TAR and ARG-TAR-ARG elongated by helix I[6]. This analysis uses RDCs from an idealized A-form helix geometry to determine five order tensor elements that describe the ordering of each a helix relative to elongated helix axis. The RDCs fit excellently to the idealized A-form geometry obtained for each helix (**Table S2**), confirming a standard canonical Watson-Crick geometry as reported previously for free TAR[6]. The analysis helped define ϑint where $\vartheta_{short}/\vartheta_{elongated}$ ranges between 0 and 1 for maximum and minimum motions. The mutant TAR^{GC} features a global inter-helical conformation in which very limited inter-helical motions were observed ϑint = ~ 1.02±0.1. In addition, TAR^{GC} was found to be in a nearly coaxially stacked (interhelical bend angle ~12±7°) conformation (**Figure 5**). This conformation differs considerably from the highly bent (~28°±3) and flexible (ϑint ~ 0.45±0.1) free TAR conformation as previously reported[6],

but agrees with the coaxial helical alignment (~8°±4) and globally rigid (ϑint ~ 1.09±0.1) TAR-ARG complex (**Figure 5**). In addition, the EI-TAR^{GC} bulge RDCs are near zero consistent with both looped out and highly flexible conformation (**Table S1**). This result suggests that despite having reduced and distinct structural dynamics, TAR^{GC} dynamically samples the TAR-ARG bound state (**Figure 5**).

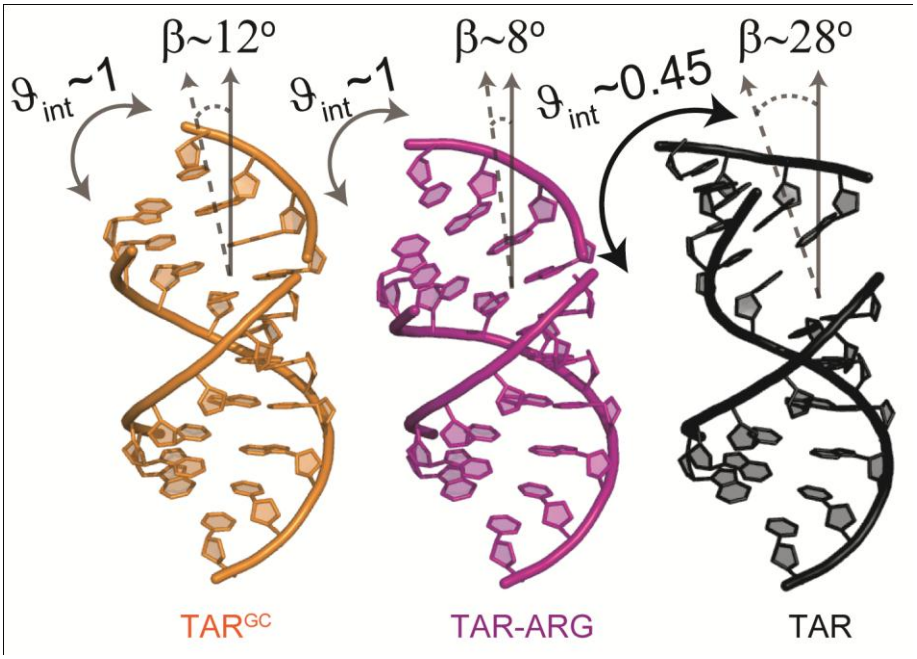


Figure 5 | Inter-helical structural dynamics of elongated helix I TAR, TAR^{GC} and TAR-ARG from order tensor analysis of RDCs. Note: there are no constraints on rotations around the elongated axis.

Next, we examined if TAR^{GC} can bind ARG given its similar global and dynamic properties. Strikingly, both the RDC measurements **(Figure 6)** as well as the NMR spectra (**Figure 7**) measured in TAR^{GC} bound to ARG were in good agreement with counterparts measured in the non-mutated TAR-ARG complex. Despite the large differences in their unbound

structural dynamics, both constructs were found to converge to a common bound structure. The RDCs values reveal that ARG binding induces minor conformational changes in TAR^{GC} localized on residue U23, most likely involved in the formation of the U23•A27-U38 base triple (**Figure 6**). Lastly, it was found that ARG binds TAR^{GC} with a slightly higher (3-fold) affinity as compared to TAR (**Figure 8)**. It is possible that the free energy cost associated with changes in RNA conformational is reduced in the pre-biased TAR^{GC} structural ensemble. As a control experiment, the binding of the aminoglycoside Neomycin B (NEOB) was monitored, as NEOB binds a more bent (~30º) TAR conformation[2],[14] that is likely less accessible to TAR^{GC}. Unlike ARG, NEOB was found to induce entirely different chemical shift perturbations in the two constructs, indicating different conformational binding modes for the different constructs (**Figure 10**).

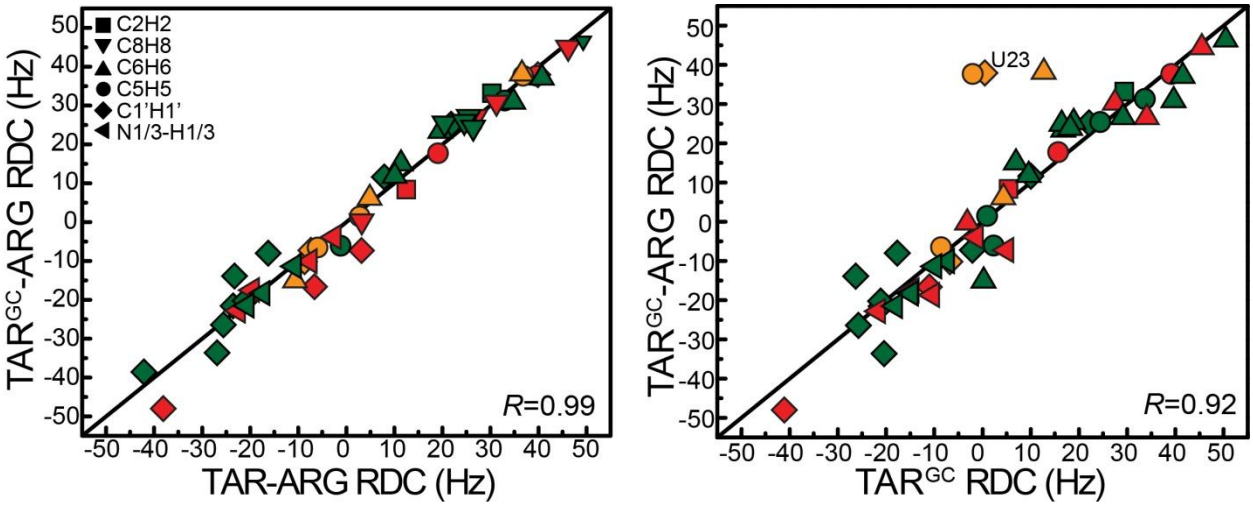


Figure 6 | Correlation plots between RDC measurements in TAR^{GC}, TAR^{GC} bound to ARG, and TAR bound to ARG colored as in Fig 3.

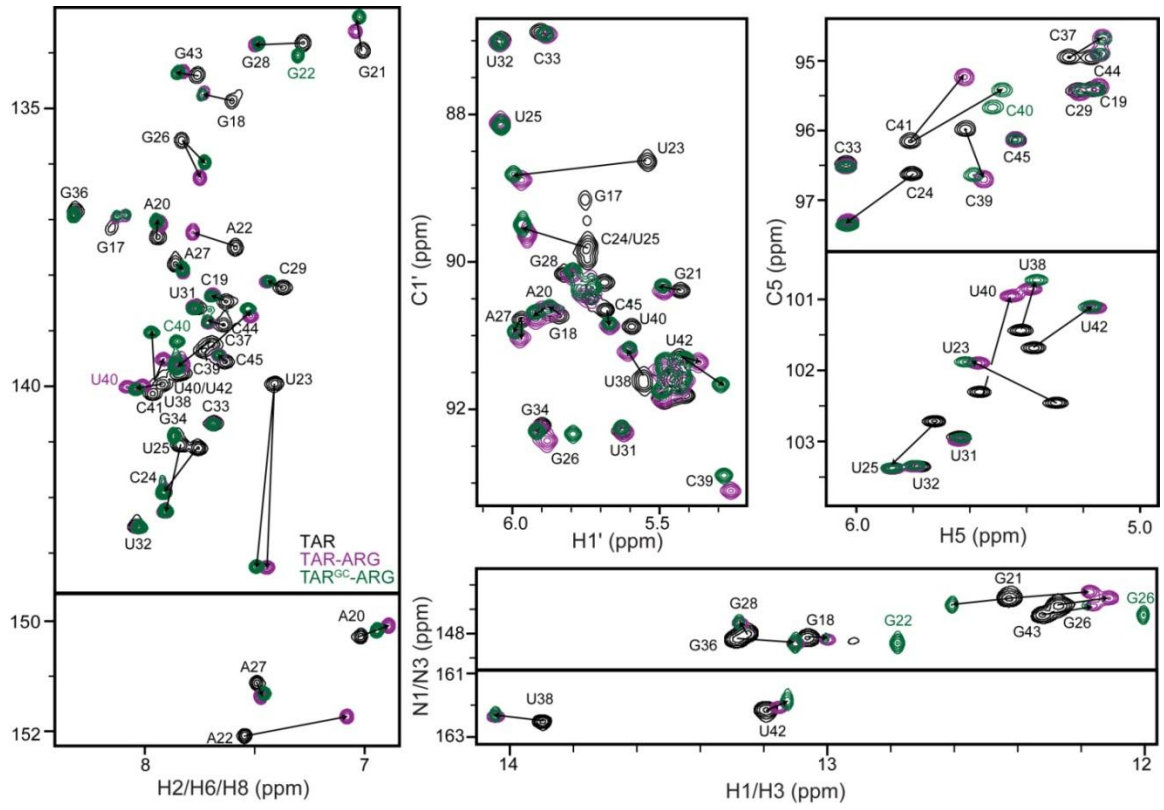


Figure 7 | TAR and TAR^{GC} converge to a common bound conformation. 2D CH and NH HSQC spectra of TAR (black), TAR^{GC} bound to ARG (green) and TAR bound ARG (purple).

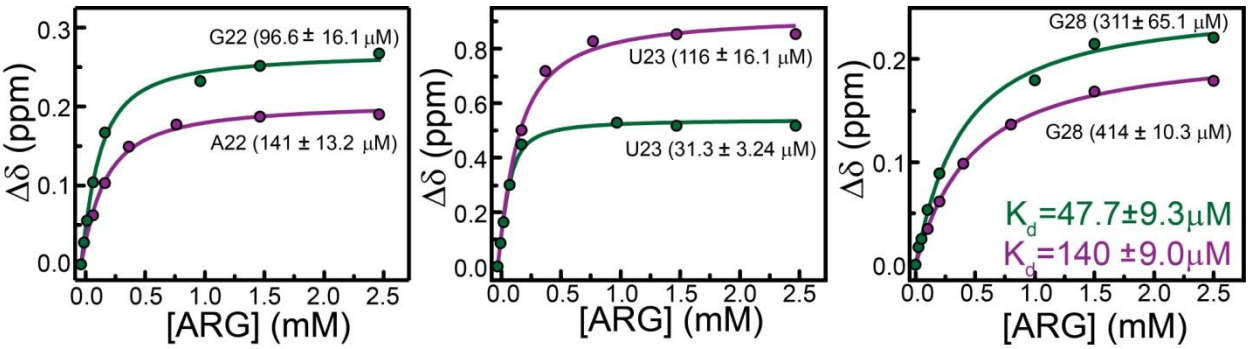


Figure 8 | Representative ARG titration curves for TAR (purple) and TAR^{GC} (green) as a function of total ARG concentration with Kd values obtained from individual and global fitting of data shown.

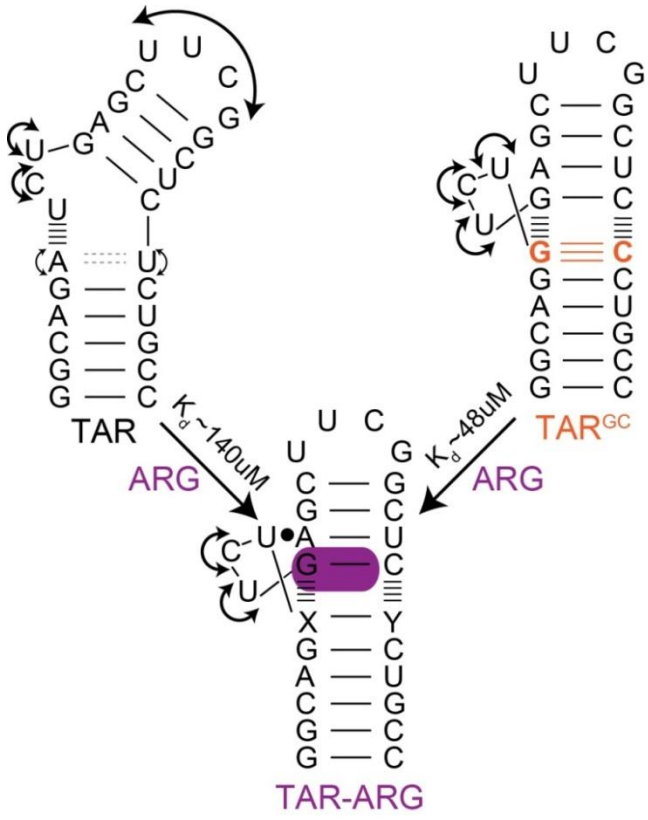


Figure 9 | TAR and TAR^{GC} adopt distinct dynamic structure ensembles in the free-state, converging to a common bound-state.

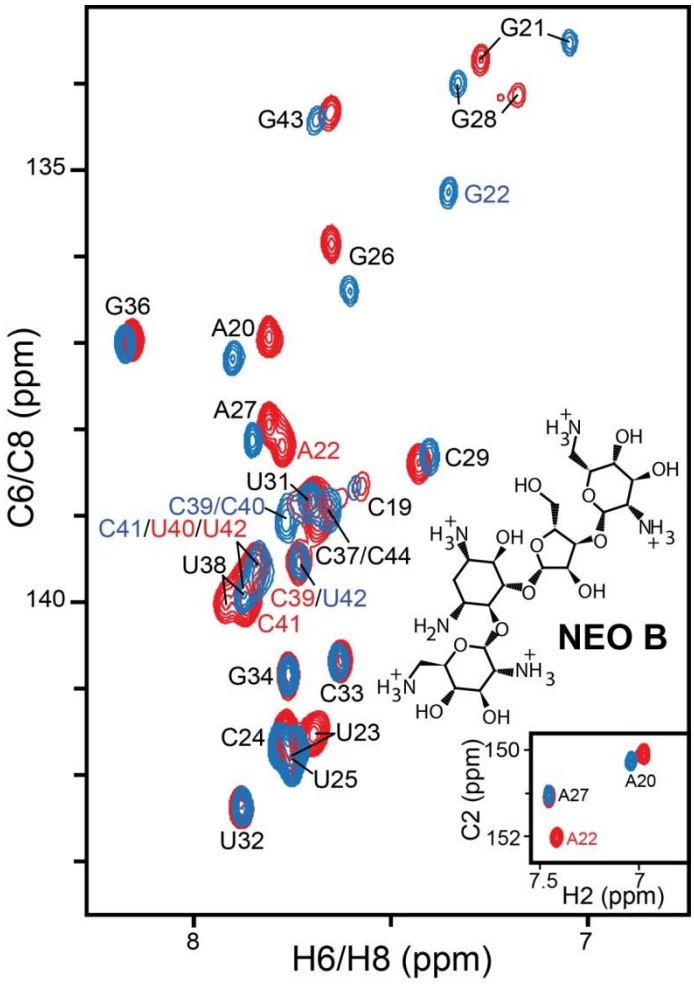


Figure 10 | Representative 2D CH HSQC C2H2, C6H6, and C8H8 spectra of TAR (red) and TAR^{GC} (blue) when in the NEOB bound state. Shown is the chemical structure of NEOB (inset).

In conclusion, the dynamics of an unbound RNA can be used to tune ligand binding affinities independent of ligand bound RNA conformations. Our results suggest that Watson-Crick base-pairs flanking inter-helical junctions are important in determining of local and global RNA dynamics. Furthermore, these effects can be predicted in certain cases based upon simple thermodynamic considerations. This finding, together with sequence-independent topological constraints imposed by length and junction asymmetry[2],[15] provide a framework for rationally designing RNA structures with specific dynamic and functional properties.

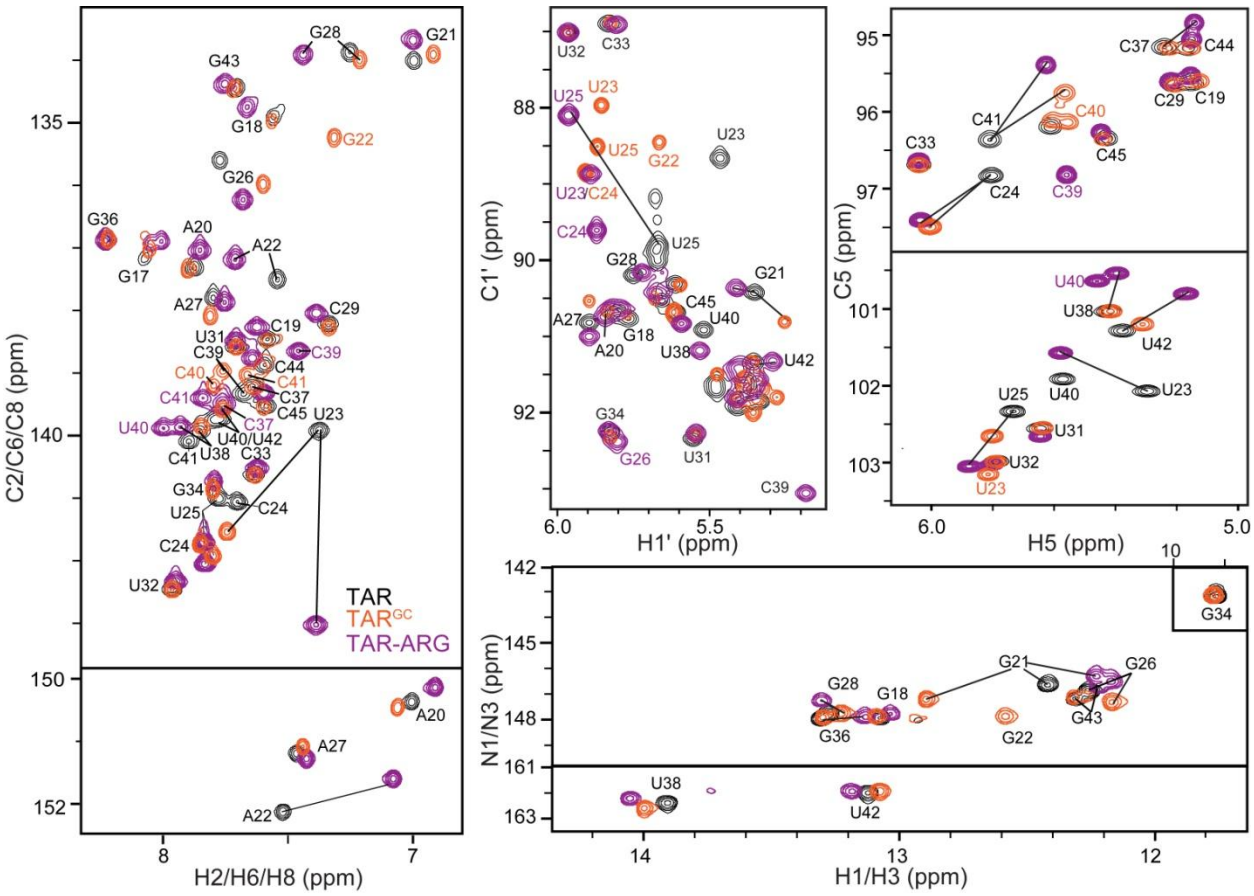


Figure 11 | 2D CH and NH HSQC spectra of free TAR (black), TAR^{GC} (orange) and TAR-ARG.

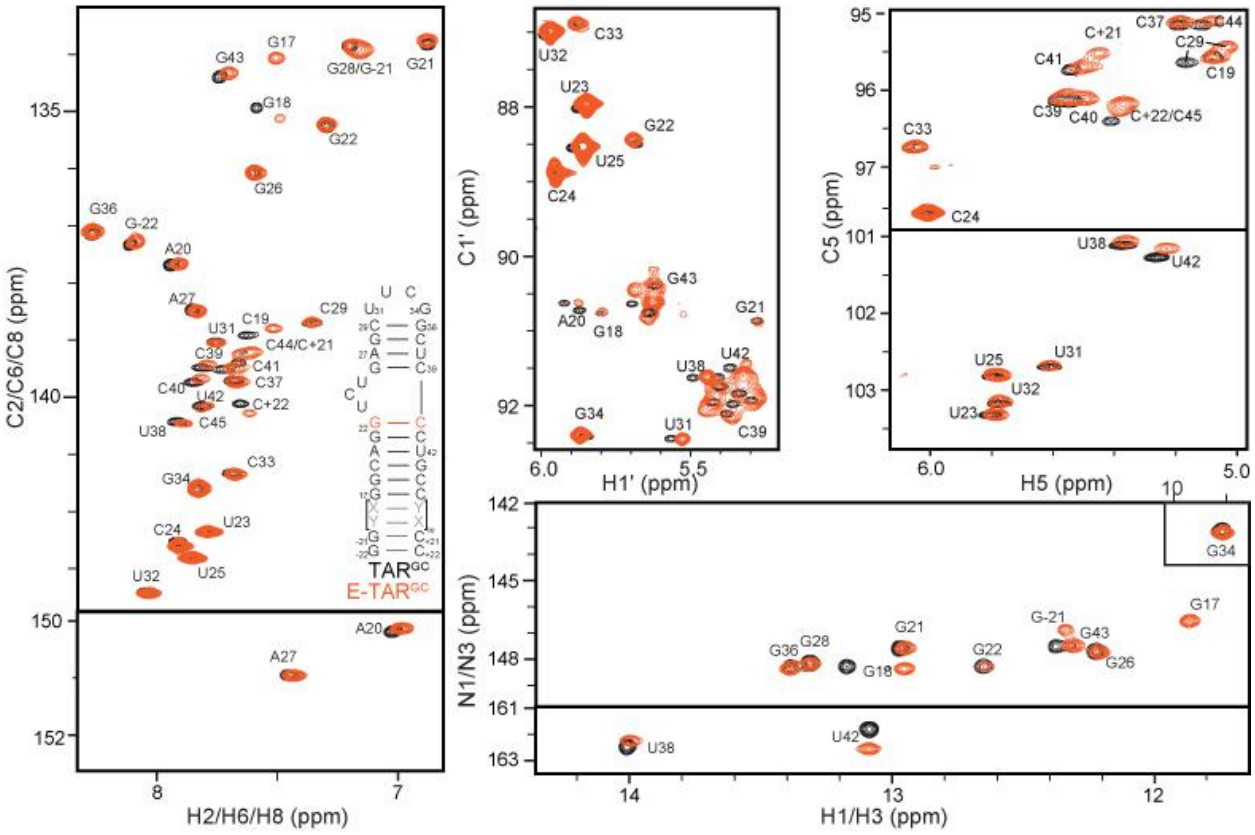


Figure 12 | 2D CH and NH HSQC spectra of TAR^{GC} (black) and Elongated TAR^{GC} (red) demonstrating that elongation does not influence the structural integrity of TAR^{GC}. Significant perturbations are localized to residues near the site of elongation.

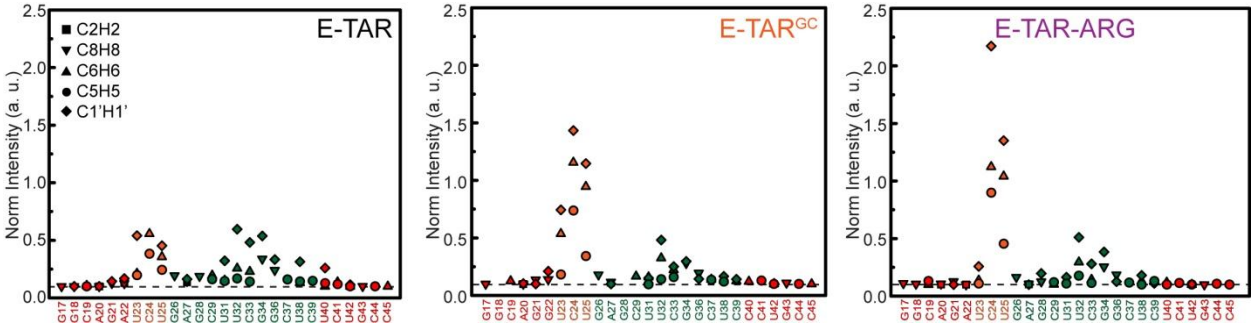


Figure 13 | Normalized resonance intensities measured in 2D HSQC spectra of elongated TAR, TAR bound to ARG, and TAR^{GC}. Residues in helix I, helix II, and the bulge are colored-coded red, green, and orange respectively.

Methods

Sample Preparation and Resonance Assignments

Uniformly $^{13}C/^{15}N$ labeled TAR^{GC} was prepared by *in-vitro* transcription using synthetic double-stranded DNA containing the T7 promoter and downstream RNA sequence of interest (*Integrated DNA Technologies*). Both elongated and non-elongated TAR^{GC} constructs were purified by 15% (w/v), and 20% (w/v) denaturing polyacrylamide gel electrophoresis containing 1x TBE and 8M Urea, respectively. Samples were purified using electroelution in 20 mM Tris pH 8 buffer and standard EtOH precipitation methods. The resultant RNA pellet was dissolved and later exchanged into NMR buffer (15 mM sodium phosphate, 0.1 mM EDTA, and 25 mM NaCl at pH ~6.4) using a Centricon concentrator to a final concentration of ~0.5-1.0 mM (Millipore Corp.). All NMR samples contained 10% D_2O and experiments were conducted in NMR buffer (above) at 298 K on an Avance Bruker 600 MHz NMR spectrometer featuring a triple-resonance 5 mm cryogenic probe. TAR^{GC} NMR spectra were assigned using the following experiments: an exchangeable 3D ^{1}H-^{15}N NOESY-HSQC, a non-exchangeable 2D ^{1}H-^{13}C NOESY-HSQC, 2D HCN, and a 2D IP-COSY. Argininamide (Sigma Aldrich) and Neomycin B (MP Biomedicals) titrations were performed by sequentially adding small volumes of concentrated ARG and NEOB samples to 0.1 mM TAR^{GC} and TAR, up to final ligand concentrations of 2.5 mM and 0.8 mM, respectively followed by 2D HSQC spectra collection for each titration point.

RDC Measurements and Order Tensor Analysis

RDCs were measured in TAR^{GC} and elongated TAR^{GC} as previously described[10,6] using concentrations of ~7mg/ml and ~23 mg/ml of Pf1 phage order medium,[16] respectively (**Table**

S1). RDCs measured in the two helices of TAR^{GC} were independently subjected to order tensor analysis[17-18] (**Table S2**). The RDCs measured in E-AU-TAR^{GC} and E-GC-TAR^{GC} were normalized (L=0.66) as previously described[6] taking into account differences in the degree of alignment arising from slightly different Pf1 phage concentrations. The normalized RDCs were combined in the order tensor analysis for each independent helix. Due to deviations from Watson-Crick geometry, the A22-U40 and terminal G17-C45 base-pairs were not included in the analysis. AFORM-RDC[18] was used to estimate errors in the order tensor arising as a result of "structural noise" and RDC measurement uncertainty. A-form helices were made using Insight II (Molecular Simulations, Inc), noting a correction in the propeller twist angles from +15° to the standard A-form value of -15°[18].

Measurements of Resonance Intensities and K_ds

Resonance intensities were measured then normalized to a baseline value of 0.1 as described previously for Elongated TAR and Elongated TAR bound to ARG[7]. TAR-ARG dissociation constants were calculated from changes in weighted average chemical shift for each titration point using the following equation[19]:

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

in which $[ARG]_T$ was the total ARG concentration at each point, $[RNA]_T$ is the RNA concentration based on UV absorbance at 260 nm, $\Delta\delta_T$ is the total difference in chemical shifts between the free and ligand-associated states (in ppm), δ_{obs} is the observed chemical shift

perturbation (in ppm), and δ_{free} is the free state chemical shift (in ppm). The data was fit using Origin software (OriginLab Corporation) allowing both $\Delta\delta_T$ and K_d to float during the fitting.

T**able 1** | RDCs measured in Elongated TAR^{GC}, TAR^{GC}-ARG, and TAR^{GC}

Residue	**Bond**	**EI-TAR^{GC}**	**TAR^{GC}-ARG**	**TAR^{GC}**
G17	(C8H8)	18.0	NA	-1.3
G17	(N1H1)	-14.6	-10.1	0.7
G17	(C1'H1')	NA	NA	NA
G18	(C8H8)	NA	0.0	NA
G18	(C1'H1')	NA	NA	NA
G18	(N1H1)	-18.5	-3.9	NA
C19	(C5H5)	NA	NA	-2.4
C19	(C6H6)	11.3	NA	NA
A20	(C2H2)	25.2	8.4	2.3
A20	(C8H8)	21.0	30.5	11.7
A20	(C1'H1')	NA	NA	-19.2
G21	(C8H8)	24.5	44.7	19.4
G21	(C1'H1')	NA	-48.0	-17.2
G21	(N1H1)	-20.2	-22.8	9.3
G22	(C2H2)	NA	NA	NA

G22	(C8H8)	23.8	NA	20.9
G22	(C1'H1')	2.2	NA	8.9
G22	(N1H1)	-16.5	-18.7	4.5
U23	(C5H5)	NA	37.7	-0.9
U23	(C6H6)	3.6	38.4	5.4
U23	(C1'H1')	-4.5	37.9	0.2
C24	(C5H5)	-3.8	-6.5	-3.7
C24	(C6H6)	0.9	6.4	1.9
C24	(C1'H1')	-5.2	-10.2	-2.9
U25	(C5H5)	-0.5	1.5	0.4
U25	(C6H6)	-1.8	-14.8	0.1
U25	(C1'H1')	-2.1	-7.3	-1.3
G26	(C8H8)	23.2	26.9	12.4
G26	(C1'H1')	NA	NA	-10.9
G26	(N1H1)	NA	NA	3.2
A27	(C2H2)	20.7	33.2	12.5
A27	(C8H8)	20.7	25.2	7.0

A27	(C1'H1')	NA	-8.0	-6.5
G28	(C8H8)	NA	25.6	8.1
G28	(C1'H1')	NA	-13.9	-12.0
G28	(N1H1)	NA	-18.3	6.4
C29	(C5H5)	NA	NA	6.1
C29	(C6H6)	12.6	37.4	17.7
C29	(C1'H1')	NA	NA	NA
U31	(C5H5)	18.0	31.3	14.3
U31	(C6H6)	19.9	31.2	16.9
U31	(C1'H1')	-13.7	-21.5	-8.8
U32	(C5H5)	NA	25.4	10.4
U32	(C6H6)	12.1	15.3	2.9
U32	(C1'H1')	19.2	25.2	9.0
C33	(C5H5)	6.0	-6.1	1.0
C33	(C6H6)	13.2	12.1	4.1
C33	(C1'H1')	1.0	-33.6	-8.5
G34	(C8H8)	22.6	24.1	7.7

G34	(C1'H1')	20.1	11.6	5.2
G34	(N1H1)	-24.9	NA	7.1
G36	(C8H8)	25.7	46.6	21.5
G36	(C1'H1')	-11.2	NA	NA
G36	(N1H1)	NA	-21.5	7.8
C37	(C5H5)	20.8	NA	20.7
C37	(C6H6)	14.0	NA	NA
C37	(C1'H1')	NA	NA	-8.0
U38	(C5H5)	23.2	NA	18.8
U38	(C6H6)	NA	23.7	7.1
U38	(C1'H1')	-13.2	-20.2	-9.3
U38	(N3H3)	-11.1	-11.4	4.3
C39	(C5H5)	NA	NA	13.0
C39	(C6H6)	14.6	NA	13.5
C39	(C1'H1')	NA	-38.5	NA
C40	C6H6	15.9	NA	19.2
C40	(C5H5)	NA	NA	9.6

C40	(C1'H1')	NA	NA	-13.1
C41	(C5H5)	22.6	NA	10.9
C41	(C6H6)	NA	NA	NA
U42	(C5H5)	22.4	37.7	16.7
U42	(C1'H1')	NA	-7.3	NA
U42	(N3H3)	-14.5	-17.5	6.4
G43	(C8H8)	15.1	26.9	14.5
G43	(C1'H1')	NA	NA	NA
G43	(N1H1)	NA	-7.2	-2.0
C44	(C6H6)	NA	NA	-2.7
C44	(C5H5)	NA	NA	18.5
C45	(C6H6)	NA	NA	-0.9
C45	(C5H5)	NA	17.8	6.7
C45	(C1'H1')	NA	-16.6	-6.7
G-21	(C8H8)	19.6	NA	NA
C+21	(C5H5)	21.8	NA	NA

Table 2 | Statistics for the order tensor RDC Analysis measured in the helices of Elongated TAR^{GC} and TAR^{GC} bound to ARG. Shown are the number of RDCs (*N*) used in the tensor determination, the root-mean-square deviation (RMSD), the correlation coefficient (*R*) between measured and back-predicted values, the asymmetry parameter (η), the generalized degree of order (ϑ) for each helix, internal generalized degree of order (ϑ_{int}), and inter-helical bend angle (β) produced using the program RAMAH. Errors in the parameters of ϑ, ϑ_{int}, η, and β were obtained using the program AFORM-RDC taking into account both RDC measurement uncertainty and local structural noise from the idealized A-form helix geometry [18].

	Helix	*N*	RMSD (Hz)	*R*	η	$\vartheta(x10^{-3})$	$\vartheta_{\iota\nu\tau}$	$\beta°$
E-TAR^{GC}	I	19	4.1	0.98	0.19±0.08	1.04±0.04	1.02±0.1	11.9±6.5
	II	12	3.2	0.98	0.16±0.10	1.06±0.10		
E-TAR^{GC} - ARG	I	13	1.9	0.99	0.36±0.21	1.89±0.21	0.88±0.14	17.7±6.0
	II	14	4	0.99	0.26±0.17	2.15±0.20		

References

1 Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77**, 363-382 (2008).

2 Bailor, M. H., Sun, X. & Al-Hashimi, H. M. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **327**, 202-206 (2010).

3 Schwalbe, H., Buck, J., Furtig, B., Noeske, J. & Wohnert, J. Structures of RNA switches: insight into molecular recognition and tertiary structure. *Angew. Chem. Int. Ed. Engl.* **46**, 1212-1219 (2007).

4 Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Curr. Opin. Struc. Biol.* **18**, 321-329 (2008).

5 Furtig, B. *et al.* Time-resolved NMR studies of RNA folding. *Biopolymers* **86**, 360-383 (2007).

6 Zhang, Q., Stelzer, A. C., Fisher, C. K. & Al-Hashimi, H. M. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**, 1263-1267 (2007).

7 Zhang, Q., Sun, X., Watt, E. D. & Al-Hashimi, H. M. Resolving the motional modes that code for RNA adaptation. *Science* **311**, 653-656 (2006).

8 Puglisi, J. D., Tan, R. Y., Calnan, B. J., Frankel, A. D. & Williamson, J. R. Conformation of the TAR RNA-arginine complex by NMR-spectroscopy. *Science* **257**, 76-80 (1992).

9 Aboul-ela, F., Karn, J. & Varani, G. 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 (1996).

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

11 Turner, D. H., Sugimoto, N. & Freier, S. M. RNA structure prediction. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 167-192 (1988).

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

13 Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 9279-9283 (1995).

14 Faber, C., Sticht, H., Schweimer, K. & Rosch, P. Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.* **275**, 20660-20666 (2000).

15 Chu, V. B. *et al.* Do conformational biases of simple helical junctions influence RNA folding stability and specificity? *RNA* **15**, 2195-2205 (2009).

16 Hansen, M. R., Hanson, P. & Pardi, A. Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol.* **317**, 220-240 (2000).

17 Losonczi, J. A., Andrec, M., Fischer, M. W. F. & Prestegard, J. H. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* **138**, 334-342 (1999).

18 Musselman, C. *et al.* 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 (2006).

19 Zuiderweg, E. R. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1-7 (2002).

This work was published in:

Stelzer[1], A. C., Kratz[1], J. D., Zhang, Q. & Al-Hashimi, H. M. RNA dynamics by design: biasing ensembles towards the ligand-bound state. *Angew. Chem. Int. Edit.* **49**, 5731-5733 (2010).

Chapter 2

From *In Silico* to *In Vivo*: Targeting an RNA Dynamic Ensemble with Small Molecules

Recent work has highlighted the role of small molecules in triggering RNA conformational change by selecting unique RNA conformations from pre-existing dynamic ensembles[1-5]. The Al-Hashimi research group recently introduced a new approach[6] for visualizing at atomic resolution RNA dynamic ensembles. By extending the timescales (<milliseconds) needed to broadly sample the entire RNA structural landscape, new ensembles could be generated with the enhanced sensitivity required for computational screening. By using sets of NMR residual dipolar coupling (RDC) data that report at the atomic level dynamics relative to an elongated RNA helices,[3] a large pool of possible conformers generated using MD[6] can be refined with experimentally guided selection[7]. By determining a minimum number of conformers satisfying all RDC data[8], a compact ensemble is constructed that samples both unique and dominant conformations across the entire structural landscape. This combination of experiment and theory helps define new RNA ensembles given the challenges of vast conformational space of flexible receptors, and limitations in current force fields underdeveloped for RNA. This new ensemble of twenty HIV-1 TAR conformers (TAR^{NMR-MD}) was shown to span known small molecule bound conformations while providing a basis for targeting the entire structural landscape[6] (**Figure 1**).

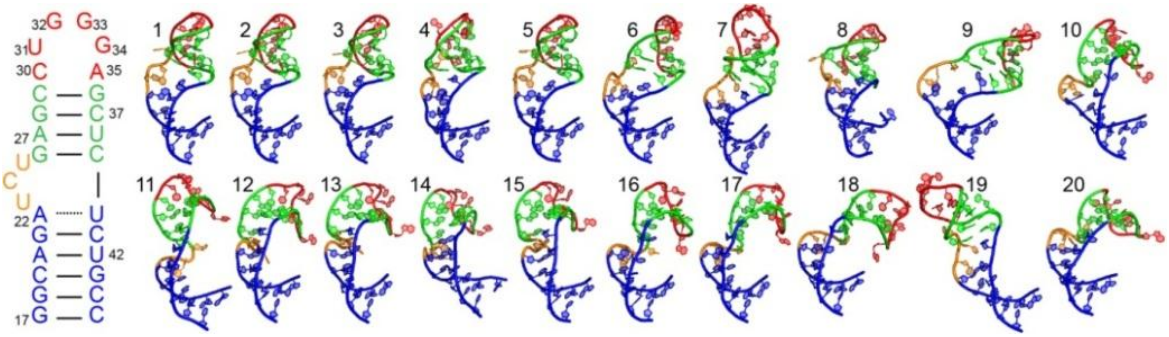


Figure 1 | Secondary structure of HIV-1 TAR RNA and the NMR-refined ensemble[6].

Computational docking small molecules against individual conformers of NMR-MD ensembles rather than a single static conformation, provides a new approach for targeting RNA conformational adaptation during an *In Silico* screen. This approach assumes the fidelity of computational docking for predicting RNA-small molecule interactions, to sufficient accuracy, when the bound structure is known. The Internal Coordinate Mechanics (ICM, Molsoft LLC)[9] docking program was benchmarked for an ideal case where the bound RNA structure was previously known. A diverse set of 96 bound RNA structures, 48 of which had corresponding experimental K_ds, were validated to correctly predict changes in free energy upon binding using ICM. Structures in which the small molecule was highly flexible (defined as N_{flex} >20 where N_{flex} is the number of flexible torsions in a small molecule) were excluded due to problems of sampling the correct ligand conformation[10]. For each docking simulation, the small molecule was removed from the binding pocket, energy minimized, and then re-docked onto the target RNA structure. The RNA binding pocket was defined as all heavy atoms within 5 Å of the small molecule. The lowest docking score obtained from sampling different small molecule conformations was recorded. The binding energies were predicted (R = 0.71) (**Figure 2**) at a level comparable to extensive work in the docking of protein structures[11]. In addition, greater than half (53%) of the predicted binding poses were within a heavy atom RMSD cut-off of 2.5 Å when compared to the X-ray/NMR solved structure (**Figure 2**, inset). This compares well with variability the different bundles of NMR structures, which themselves result in an average RMSD of around 1.8Å[12]. This finding suggests that the accuracy of docking predictions is not limited in the ability to accurately predict the proper bound state pose or conformation or the ability of the scoring function to properly score binding interactions.

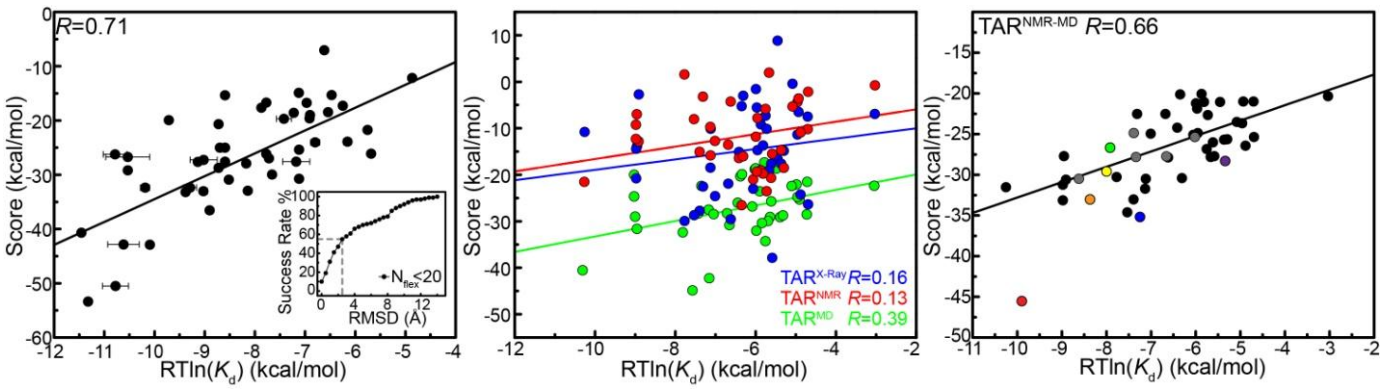


Figure 2 | Correlations between experimental changes in free energy $\Delta G = RT\ln(K_d)$ (including reported error) and docking scores. Correlation coefficients (R) are shown in each case. The plots are arranged as: 48 known RNA-small molecule complexes with known structure (inset: percentage ligand poses falling within a heavy-atom RMSD relative to solved structure), 38 TAR binders when docking small molecules against X-ray (blue), NMR (red) structures of free-TAR, and a random MD-derived ensemble of 20 snapshots from a 80 ns MD simulation (green), and the 20 TAR conformers from the $TAR^{NMR\text{-}MD}$ ensemble, respectively.

A more general and potentially severe problem with computational docking of RNA is that the correct small molecule bound RNA structure is generally unknown or may deviate significantly from one small molecule to the next. This uncertainty results in failure of computational docking protocols to sample the correct conformations necessary to procure unique details of binding pockets, particularly for flexible RNA receptors that undergo large structural changes upon binding small molecules. The impact of this uncertainty relative to docking an RNA ensemble has never been quantified in current docking protocols. As an initial test, we conducted docking simulations against free structures of TAR solved using available X-ray[13] and NMR[14] structures to quantify the current accuracy predicting experimental binding energies for 38 TAR-binding compounds. It was found that the quality of the docking predictions deteriorate abruptly (R<0.13) when conducted against these static structures, so as to become

completely uninformative in predicting the binding of lead compounds (**Figure 2**). In addition, docking against a computational (TAR^{MD}) ensemble of 20 random snapshots from an 80 ns MD simulation of free-TAR[6] resulted in minor improvement (R = 0.39), nowhere near the accuracy attained when the bound RNA structure is available (**Figure 2**, R = 0.71). These simulations were conducted by independently docking each small molecule onto each conformer and then recoding the lowest overall score, corresponding to the dominant bound state. These findings suggest the accuracy of docking predictions is limited by the uncertainty in the final RNA bound structure.

We next examined if we could recover the accuracy of docking against proper structures by docking the test set of 38 small molecules against the 20 conformers from the $TAR^{NMR\text{-}MD}$ ensemble. This NMR-refined ensemble was shown previously[6] to sample many known ligand bound TAR conformations. Surprisingly, the binding energies were found to be predicted with an accuracy (R=0.66) (**Figure 2**) comparable to that when the bound RNA structure is known (R = 0.71) (**Figure 2**). These results reinforce the notion that ligands do not 'induce' new conformations in TAR, rather they 'capture' unique conformations from an ensemble of populated states in the $TAR^{NMR\text{-}MD}$ conformational pool.

The interaction between TAR and the protein Tat has been targeted for inhibition of HIV replication, as the interaction plays an important role in regulating the transcription of new viral particles, but has failed to result in clinical therapeutic development[15]. Our ensemble-targeted approach was used to identify novel TAR-targeting compounds. The 20 conformers in the $TAR^{NMR\text{-}MD}$ ensemble was screened i*n silico* against a library of ~51,000 small molecules. From this list, 57 commercially available hits predicted to bind with high affinity (<10μM) were tested

for activity using fluorescence-based assays to probe (i) binding to TAR containing a fluorescent probe at residue U25 validated using known TAR binders (**Figure 3**)[16] and (ii) inhibition of the interaction between TAR and N-terminus-labelled-fluorescein peptide containing the arginine rich motif of Tat known to interact with TAR[17] (**Figures 4 and 5**).

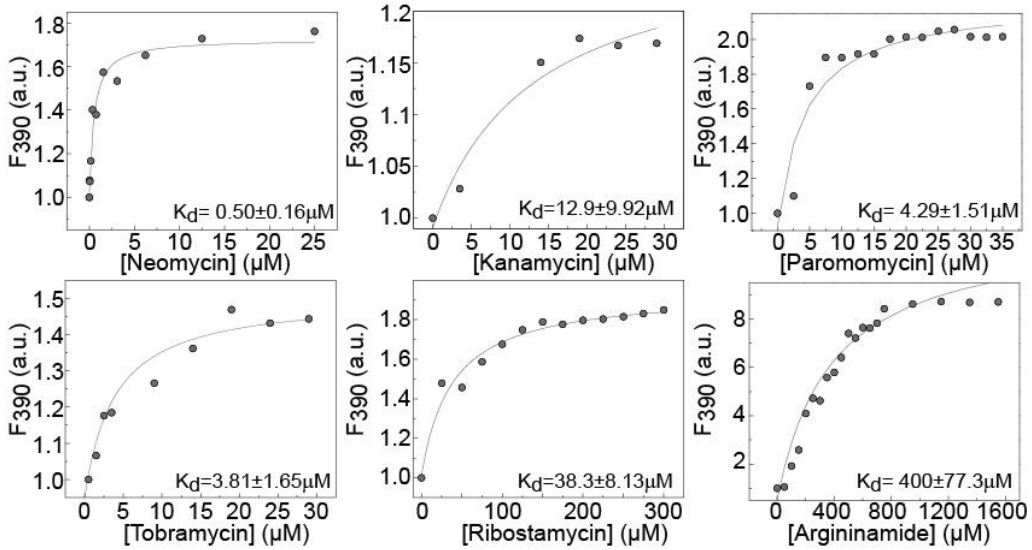


Figure 3 | Validating the fl-TAR binding assay using known TAR binders neomycin[18], kanamycin[19], paromomycin[19], tobramycin[19], ribostamycin[19], and argininamide[18].

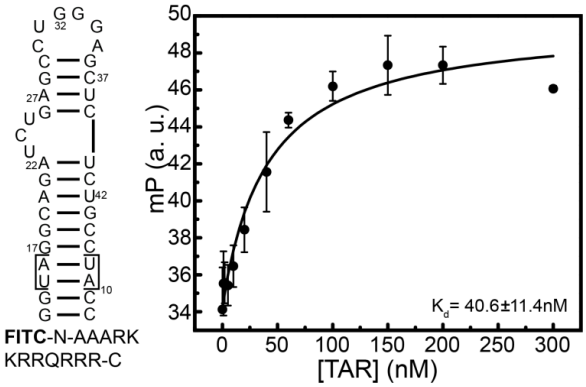


Figure 4 | Binding of F-Tat peptide to E-TAR using Fluorescence Polarization (FP). FP values were monitored upon addition of E-TAR to 10 nM F-Tat. Average values with standard deviations over triplicate runs are reported.

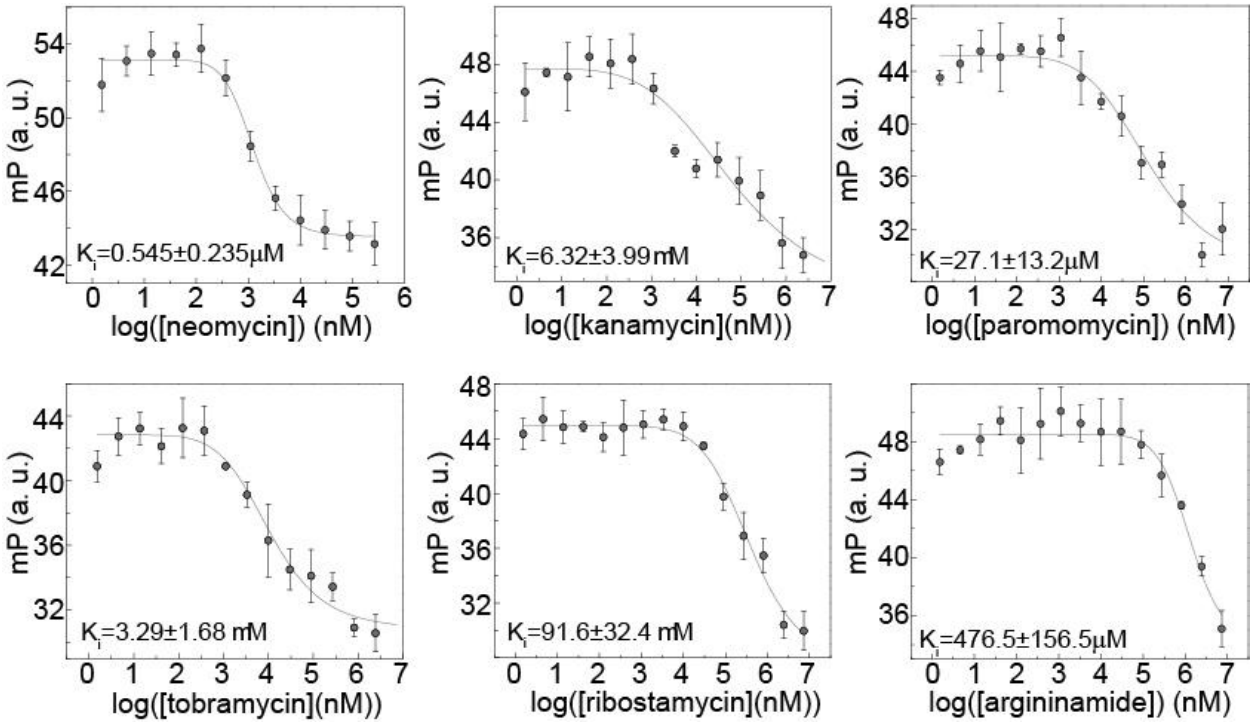


Figure 5 | The validation of the FP displacement assay was completed using six known inhibitors of the TAR-Tat interaction. K_i values obtained for neomycin (0.545 ± 0.235 μM) and argininamide (476.5 ± 156.5 μM) are in close agreement with those reported previously (1 μM[18] and 1 mM[20] respectively).

Six novel small molecules (**Figure 6**) were experimentally validated from the 57 commercially available molecules that were found to bind TAR (K_d = 55 nM–122 μM) (**Figure 7**) and inhibit its interaction with F-Tat peptide (K_i = 710 nM–169 μM) (**Figure 8**). Together with spermine (see **Appendix I)**, the hit rate was found to be 12% with a hit rate as high as 50% when focusing on water-soluble compounds. It is of note that the docking simulations were conducted in without DMSO, a common solvent used to solubilize highly hydrophobic molecules for experimental assays. This is far improved from a hit rate of 0% when screening a random set of 57 small molecules from the initial library (**Table 1**). The virtual screen also

identified several additional small molecules, including two aminoglycosides that were correctly predicted to not interact with TAR despite their multiple positive charges, again verified using fluorescence-based binding assays (**Figures 9**).

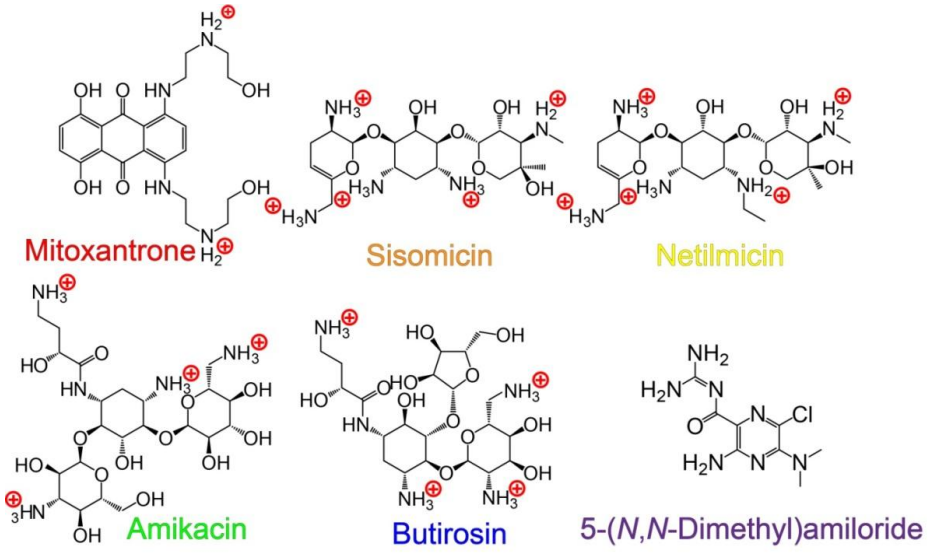


Figure 6 | Chemical structure of six novel small molecules discovered *In Silico*, validated *In Vitro*, and arranged descending binding affinity.

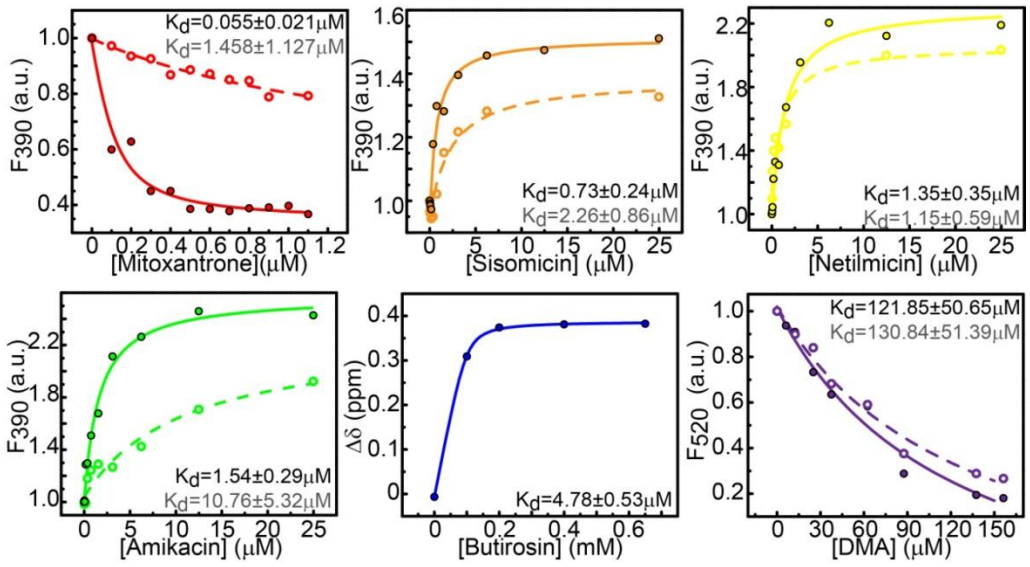


Figure 7 | Binding to HIV-1 TAR using normalized fluorescence intensity measurements or NMR chemical shift perturbations (butirosin) in the absence (filled circles) and presence (open circles) of 100-fold excess unlabeled *S. Cerevisiae* tRNA (relative to f-TAR). Best-fit dissociation constants (K_d) are shown in each case.

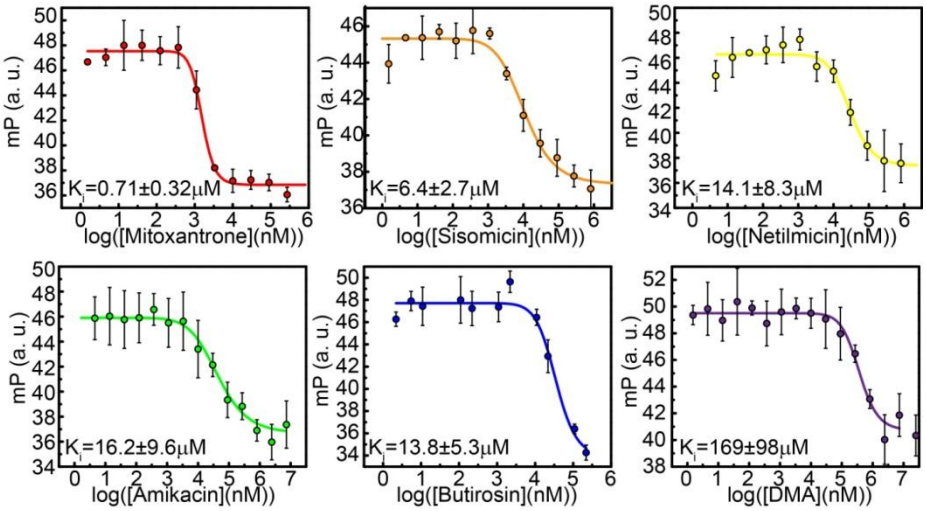


Figure 8 | Displacement of N-terminus-labeled-fluorescein Tat peptide from E-TAR as measured by fluorescence polarization as a function of small molecule concentration. Computed inhibition constants (K_i) are shown with corresponding standard deviations.

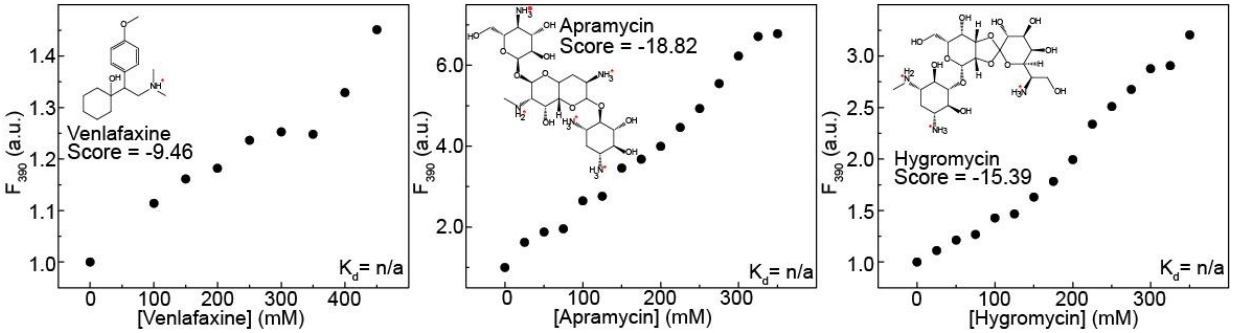


Figure 9 | Aminoglycosides and positively charged small molecules, correctly predicted to bind TAR with weak affinity verified using fluorescence intensity binding assays.

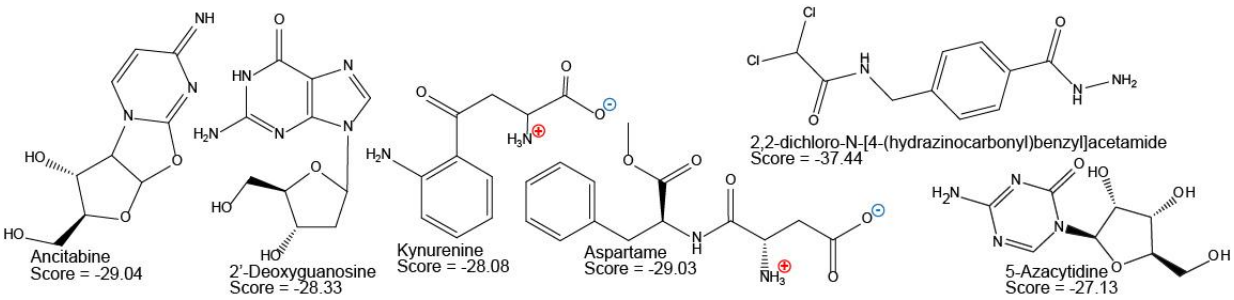


Figure 10 | Example molecules predicted to bind TAR that exhibited no activity *In Vitro*. Docking score is indicated next to each small molecule.

The six small molecules add to the diversity of known TAR-binding small molecules (**Figure 6**), in addition to recently developed cyclic peptides that bind TAR with both high affinity and specificity[21,22]. For example, mitoxantrone (**Figure 6**) a known RNA binder[23], binds TAR with an affinity (K_d ~ 55 nM), which among non-neamine derivatives is second only to one other small molecule, WM5 (K_d ~ 19 nM) [24] identified over the two decades TAR targeted for therapeutic development. 5-(N,N-dimethyl)amiloride (DMA) (K_d ~ 122 μM) lacks any cationic groups at biologically relevant pH (5.4 to 9.4), contains a newly discovered RNA binding scaffold of a 5-chloropyrazin-2-amine core (**Figure 6**), and targets a unique pocket within the TAR apical loop (see **Appendix I**). DMA represents a rare example of a small molecule binding exclusively to the apical loop of TAR RNA. The molecules include four semi-synthetic aminoglycosides (**Figure 6**), amikacin (K_d ~ 1.5 μM), butirosin (K_d ~ 4.8 μM), netilimcin (K_d ~ 1.4 μM) and sisomicin (K_d ~ 0.73 μM) which have never been shown to interact with TAR RNA.

Using a competition experiments, in which we re-measured the K_d following the addition of 100-fold excess unlabelled tRNA[19], we were able to assay different levels of selectivity for each compound (**Figure 7**). Using this assay, we observed significant deterioration in the binding affinities of mitoxatrone, amikacin, and sisomicin, (reduction in K_ds by factors of 27, 7, and 3 respectively), consistent with non-specific binding to tRNA, while observing little to no changes in the K_ds for netilmicin and DMA. This was indicative of a high level of specificity for HIV-1 TAR (**Figure 7)**. Interestingly, the addition of a single ethyl group significantly (3-fold) improved the binding specificity of netilmicin in comparison to sisomicin in this assay.

For a more extensive test of specificity, fluorescence-based assays were used to measure directly the binding affinity of the small molecules (excluding butirosin, due to limitations in

commercially availability) to a panel of three additional RNAs that more closely resembled TAR's hairpin structure. These included a HIV-2 TAR variant (HIV-2) differing from HIV-1 TAR by a single bulge residue deletion, the insertion of G-U, and a swap of a G-C base-pair in the upper stem; the highly characterized aminoglycoside-binding prokaryotic ribosomal A-site hairpin (A-site); and the rev response element, another hairpin regulatory RNA element in HIV-1 (RRE)[25,26] (**Figure 11**). These assays (**Figure 12**) yielded specific binding profiles for the various compounds that were mirrored to those observed with tRNA (**Figure 7**). Netilmicin was shown to have the highest levels of selectivity. It bound the closely related HIV-2 TAR with negligible affinity and the ribosomal A-site and RRE RNA with 35 and 86 fold reduced affinities, respectively (**Figure 12**). For a more stringent test of selectivity, we measured the affinity of netilmicin to a TAR mutant (HIV-2*) that only featured a single cytosine bulge deletion versus HIV-1 TAR, resulting in a 16-fold reduction in binding affinity (**Figure 13**). In contrast, sisomicin displayed a marked reduction in specificity compared to netilmicin, binding RRE with a comparable affinity to that observed in HIV-1 TAR (**Figure 12**). Interestingly, DMA which displayed unique chemical shift perturbation the apical loop of TAR (see **Appendix I**), showed strong selectivity against A-site and RRE but not HIV-2 TAR (**Figure 12**). Besides netilmicin, all other hit molecules were found to bind at least one other RNA with an affinity comparable to its affinity for HIV-1 TAR (**Figure 13**). This specificity was further confirmed using competition assays analogous to those described for tRNA in which these unlabelled RNA constructs were employed as competitors (**Figure 14)**.

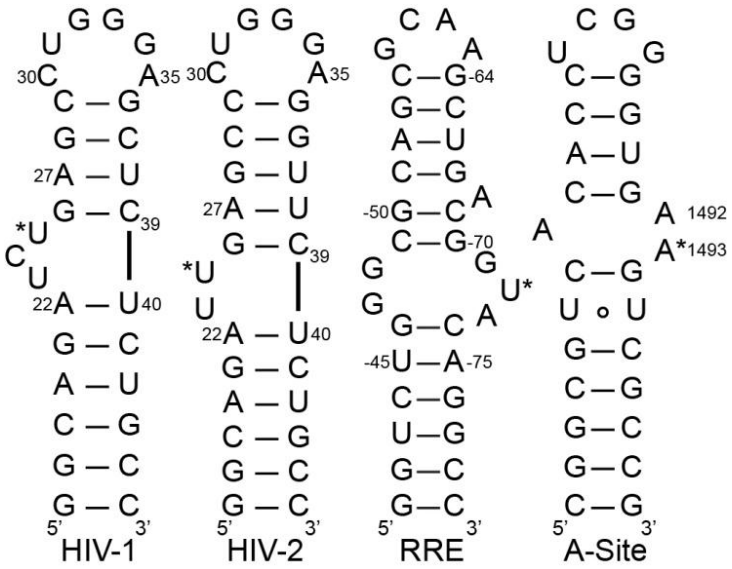


Figure 11 | Shown are the secondary structure of three related RNA hairpins (HIV-2, RRE, and A-site) used in binding specificity analysis. Sites of 2'-fluorescein and 2-aminopurine base fluorophore substitutions are indicated (*).

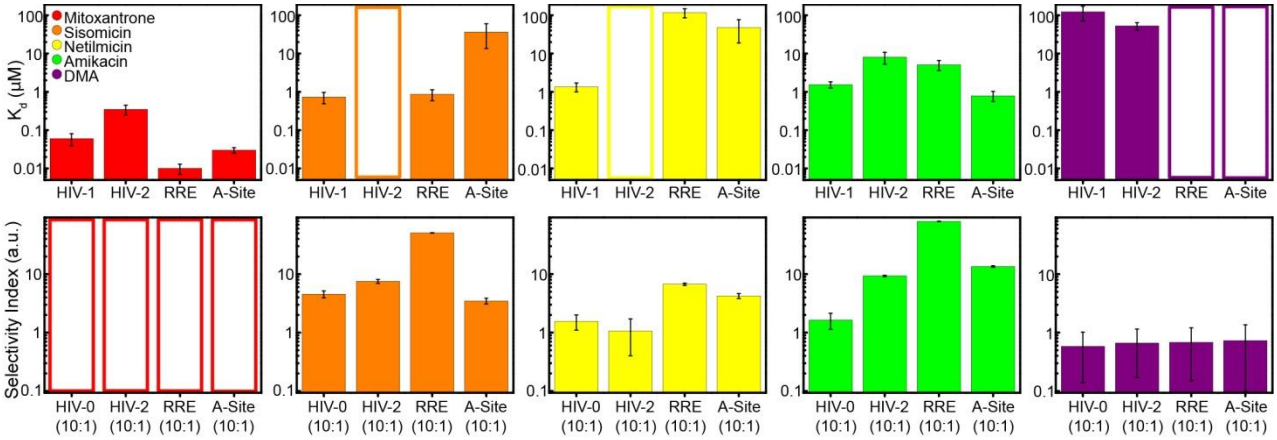


Figure 12 | Dissociation constants (K_d) for hit molecules against each of four RNA constructs determined using fluorescence intensity measurements (Note: plotted on log base-10 scale). Open symbols indicate failure to accurately fit K_d due to weak affinity of small molecule to fluorescently labeled RNA. Selectivity Indices (SI) defined as the ratio of the K_d measured for a small molecule to HIV-1 TAR in the presence and absence of the noted competitor RNA are shown. Open symbols again indicate unreliable K_d fit due to strong competition from unlabeled competitor RNA resulting in incomplete F-TAR saturation.

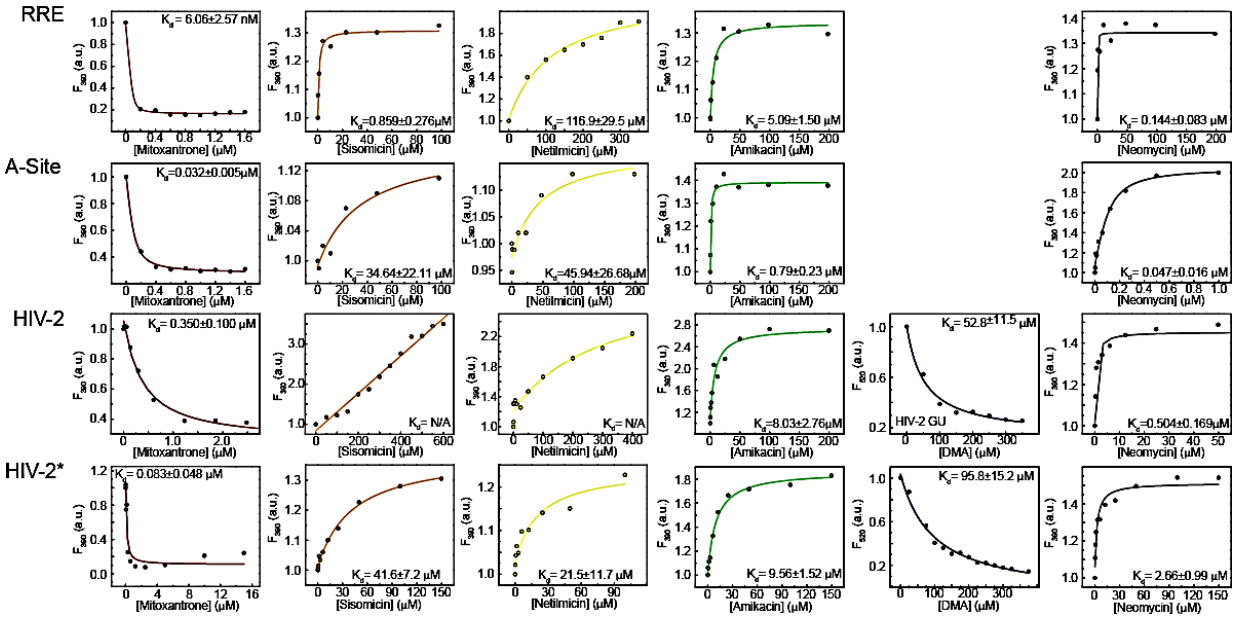


Figure 13 | Binding isotherms for hit molecules (color coded as in **Figure 6**) to RRE, A-site, HIV-2, and HIV-2* using fluorescence intensity measurements**.**

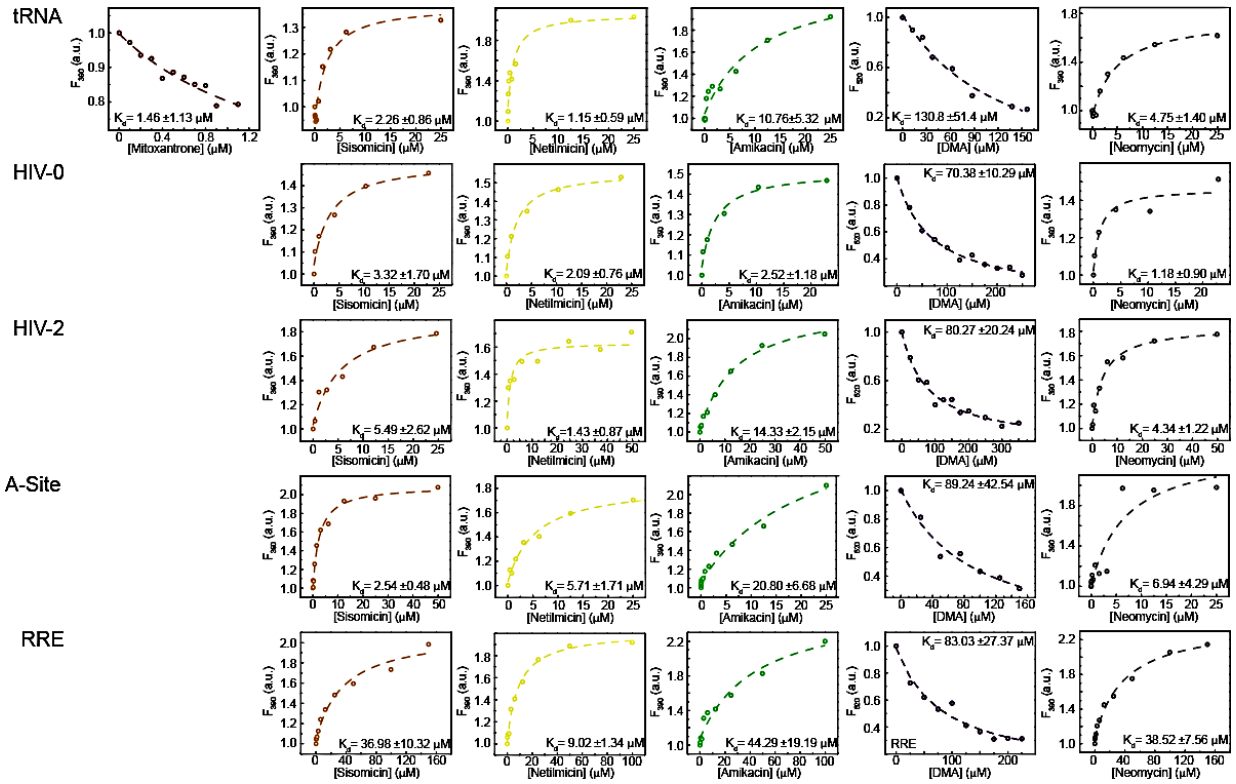


Figure 14 | Binding isotherms of hit molecules binding HIV-1 TAR in the presence of unlabeled: 100x tRNA, 10x HIV-0 TAR (an AU-elongated TAR construct lacking the UCU bulge), 10x wildtype HIV-2 TAR, 10x prokaryotic A-Site RNA, and 10x RRE. Mitoxantrone binding was undetectable due to non-specific interactions with unlabeled competitor.

Before conducting cellular assays, we compared with the atomic resolution afforded by NMR chemical shift mapping experimental binding interactions with predicted binding interactions output from the top scoring poses of the hit molecules binding TAR. Many of the hit molecules were predicted to bind conformers within a common region (conformers 12-15) of the TAR^{NMR-MD} structural landscape (**Figure 15**), regions characterized by near coaxial alignment of the helices (inter-helical bend angle <12°). These global confirmations were also observed for TAR when bound to Tat mimics[27]. Correspondingly, all of the hit molecules caused chemical shift perturbations characteristic of coaxial helix alignment and stacking interactions previously observed when bound to Tat peptides and divalent ions[27] (e.g. U23 and C24, **Figure 16** and **Appendix I**).

Interestingly, netilmicin which shows unique selectivity for binding HIV-1 TAR *in vitro*, was also predicted to bind a single conformer within the TAR ensemble with the highest local specificity. One conformer (18) accounted for a population of 66% over the entire TAR^{NMR-MD} ensemble. The hit molecules were predicted to bind TAR using distinct modes and to contact various combinations of residues in the bulge, upper stem, and apical loop (**Figure 17**) which are noted for forming critical interactions with Tat, thus providing a structural basis for the activity of TAR-Tat inhibition. Correspondingly, the six hit molecules were shown to induce distinct chemical shift perturbations, particularly at residues within distinct binding pocket (**Figures 16 and 17** and **Appendix I**). Significant chemical shift perturbations (>0.1 H^1 ppm) were observed for 87% of residues predicted to be within 5Å of the small molecule (**Figure 17**, red spheres). Perturbations outside this cut-off (**Figure 17**, green spheres) that displayed significant chemical shift perturbations typically corresponded to nearby flexible residues, likely to undergo conformational change upon on binding the small molecule.

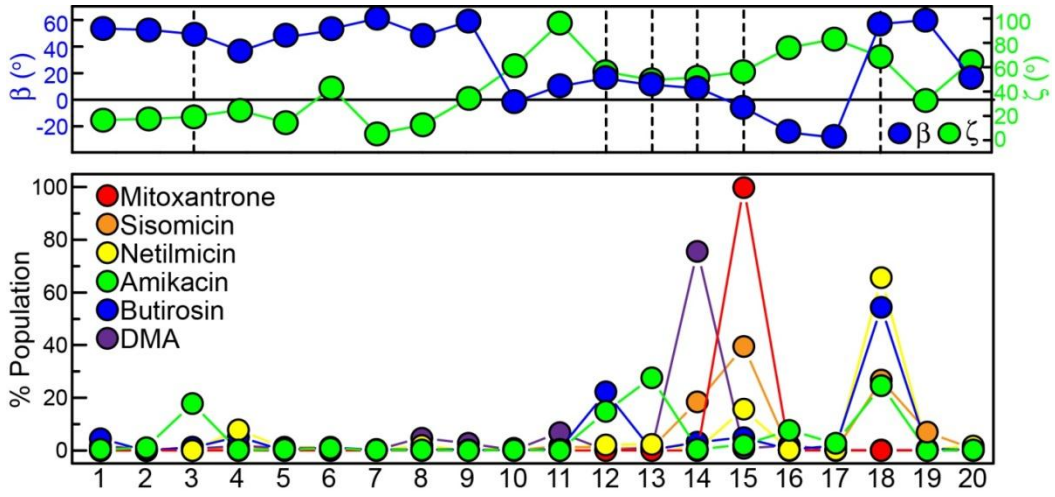


Figure 15 | Shown are the global preferences for hit molecules to interact with distinct TAR conformers in the TAR^{NMR-MD} ensemble. The inter-helical bend (β) and twist (ζ) angle is shown for each TAR conformer in addition to the predicted fraction of the entire docking predicted population for each of the 20 TAR conformers assuming an ICM score-derived population.

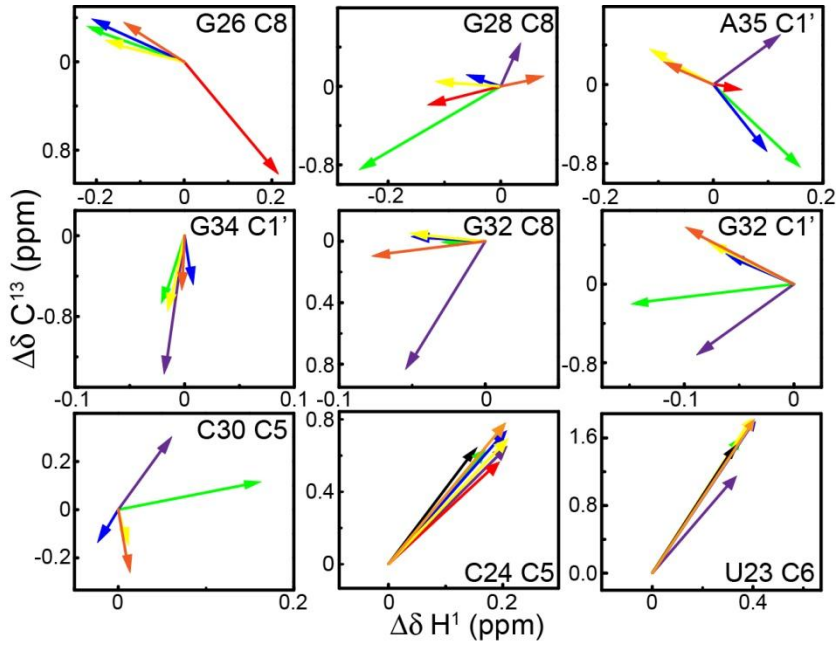


Figure 16 | Representative chemical shift perturbations for unique residues (colored as **Figure 6**) showing resonance position perturbations from free TAR to >97% bound TAR. Mg^{2+} perturbations are shown in black. NMR peaks are not shown for clarity, instead they are provided in **Appendix I.**

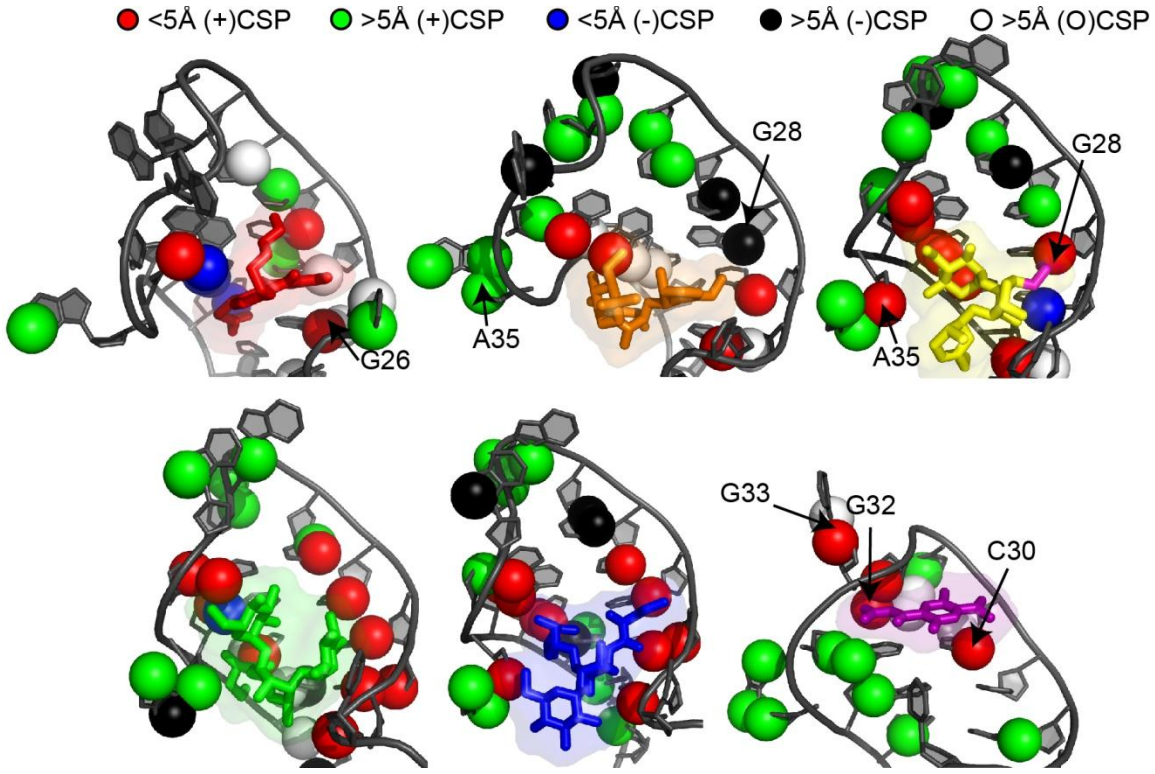


Figure 17 | Mapping experimental NMR chemical shift perturbation onto models of docking predicted TAR-bound structures. Individual bond heavy atom residues are colored according to a 5Å distance cut-off from any small molecule atom and significant chemical shift perturbations (+CSP) or not (-CSP) (where significance is defined as >0.1 ppm), or limitations in assignment due to spectral overlap (O). The ethylene substituent of netilmicin is shown in purple.

It was found that even detailed aspects of the predicted binding modes were supported in certain cases by the NMR chemical shift data (**Figures 16 and 17**). One example was a unique stacking of mitoxantrone on base residue G26 (C8H8) that resulted in distinct binding modes further supported by and directionally unique perturbations. In addition, the correctly predicted binding of DMA to a unique pocket within the TAR apical loop was again confirmed experimentally. This binding mode was confirmed by observations of extensive, directionally

unique perturbations at many apical loop residues namely (G32(C8/C1′), G33(C1′), G34(C1′) and A35(C1′)). While other aminoglycosides have been shown to bind the TAR bulge and upper/lower stems[28-30], netilmicin was shown to interact uniquely with TAR's apical loop in addition residues in the bulge and upper stem. In addition, a comparison between netilmicin and sisomicin showed unique contacts involving netilmicin's ethylene group. For example, netilmicin was shown to induce significant perturbations at G28 and A35 while not found upon addition of sisomicin. Conversely, the binding of butirosin and amikacin to similar pockets of TAR is supported by observation of significant perturbations at many common sites. Thus, irrespective of the limitations in interpreting chemical shift perturbations using quantitative measurement, the data support unique global aspects of the docking predictions, particularly the suggesting unique differences in the binding modes of the six small molecules.

Of the five compounds tested (butirosin was excluded as it became commercially unavailable at the onset of these experiments), netilmicin, the molecule shown to bind TAR with the highest specificity *in vitro* (**Figure 12**), was also found to inhibit Tat-mediated activation of the HIV-1 promoter by ~81% versus a control (**Figure 18**) when assayed in live human T cells containing a luciferase reporter gene. The additional compounds failed to show activity in this assay most likely due to either weaker binding specificity although other effects could include differences between full length Tat and Tat peptides used in the *in vitro* fluorescence polarization displacement assay. As a control, the measurements were repeated upon addition of phorbol 12-myristate 13-acetate (PMA), which activates the HIV-1 LTR in a Tat-independent manner. If netilmicin works specifically to block Tat-mediated activation of the HIV-1 promoter through an interaction with TAR, then no inhibition would be observed upon PMA-mediated activation.

Indeed, this finding that netilmicin did not inhibit the PMA-mediated stimulation of the HIV-1 promoter, strongly suggests the possibility of off-target effects (**Figure 18**).

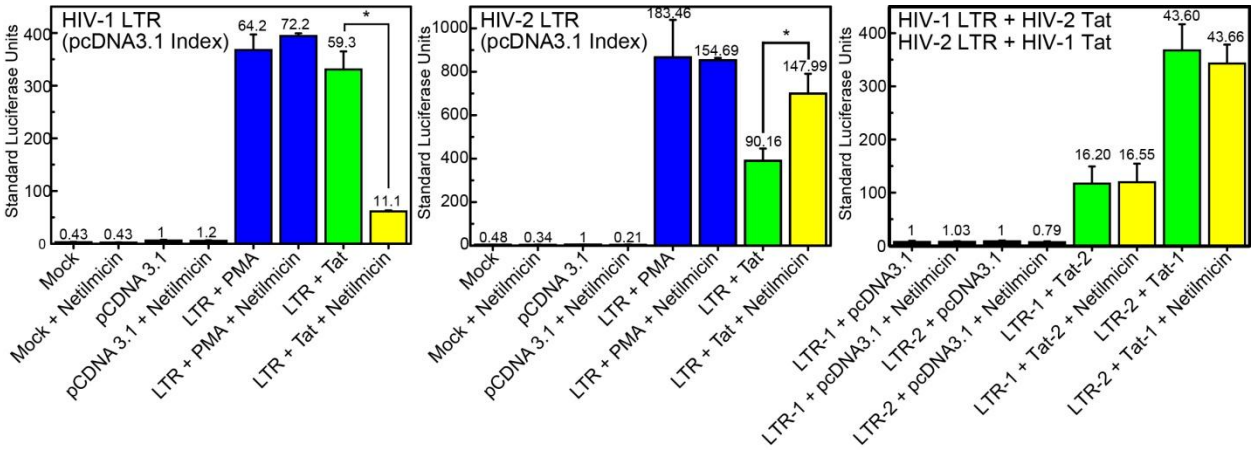


Figure 18 | Netilmicin activity in inhibiting Tat-mediated transactivation. **a,** Jurkat T cells treated for 24 hours with 100 μM netilmicin, or vehicle, were immediately transfected with the HIV-1 LTR-luc (LTR) and the HIV-1 Tat (Tat) constructs or a corresponding empty vector (pcDNA3.1). "Mock" refers to a non-transfected control. After transfection, netilmicin or vehicle was added to the media with luciferase activity measured 24 hours later. For control, cells were transfected with only the HIV-1 LTR-luc and, 4 hours prior to the 24 hour completion point, were treated with 10 ng/mL PMA in both the presence and absence of netilmicin. Significance was calculated with *$p<0.02$. All experiments were also repeated using HIV-2 LTR with HIV-2 Tat, HIV-2 LTR with HIV-1 Tat, and HIV-1 LTR with HIV-2 Tat to rule out off target effects and targeting of netilmicin to HIV-1 Tat.

To further assay the specificity of netilmicin in a cellular context, inhibitory activity was measured when using a HIV-2 TAR promoter containing an identical HIV-2 TAR sequence used in *In vitro* binding assays (**Figure 12**). I*n vitro* binding data (**Figure 12**) suggested that netilmicin bound HIV-2 TAR with negligible affinity. If the mode of action of netilmicin was specific to binding of TAR, one would predict it would in turn be far less effective at inhibiting

the stimulation of the HIV-2 promoter. Indeed, netilmicin was actually found to minimally stimulate the HIV-2 promoter (**Figure 18**). In addition, inhibition was not observed when using HIV-1 Tat to stimulate the HIV-2 transcriptional promoter, again indicating that netilmicin did not bind and inhibit HIV-1 Tat. Instead, it suggested that netilmicin affects Tat-mediated transactivation through its direct interaction with HIV-1 TAR (**Figure 18**). Thus, the *in vitro* activity of netilmicin is specific to inhibiting the activation of HIV-1 but not the closely related HIV-2 promoter when assayed in a cellular context.

Remarkably, netilmicin was not only found to inhibit Tat activation, but also HIV-1 viral replication, as assayed using the HIV-1 indicator cell line, TZM-bl and HIV-1 NL4-3 isolate, containing the same HIV-1 TAR sequence employed for *in vitro* studies. The addition of netilmicin to cells prior and during infection resulted in decrease of HIV replication, yielding an IC_{50} value of ~23.1 μM (**Figure 19**), strikingly similar to the K_i value of 14.1 μM measured for the *in vitro* in the Tat displacement assay (**Figure 8**). The similarity of such inhibition constants further supports netilmicin's specificity to inhibit HIV replication by targeting TAR. These results were further corroborated in an *in vivo* assay using HUT-78 T-cell lines infected with HIV-1 NL4-3 in the presence of 100 μM netilmicin. HIV-1 replication was assessed at three day intervals by measuring the levels of p24 antigen in culture supernatants. A decrease in p24 was observed for samples treated with netilmicin in comparison to vehicle alone, with the most significant difference observed at day 9 (**Figure 20**). Netilmicin toxicity was ruled out using trypan blue staining for cell viability, displaying no significant difference between the vehicle and netilmicin treated cells. Lastly, our *in vitro* and cellular (gene reporter) assays showed again that despite the strong chemical similarity of netilmicin and sisomicin, the later was found to be less effective at inhibition of Tat-mediated activation of the HIV-1 promoter, likely due to

reduction in TAR specificity. If netimilicin inhibits HIV-1 replication by inhibiting the TAR-Tat interaction, we would netilmicin to outperform sisomicin in activity as an inhibitor. Indeed, we observed an IC_{50} value (~157.1 mM) for sisomicin, some 3 to 16-fold higher than that measured for netilmicin (**Figure 20**). Given the IC_{50} typically observed with resistant viruses are on the order of two-fold these differences are significant. These data provide additional support that netilmicin in fact inhibits HIV replication through selective inhibition of the the TAR-Tat interaction.

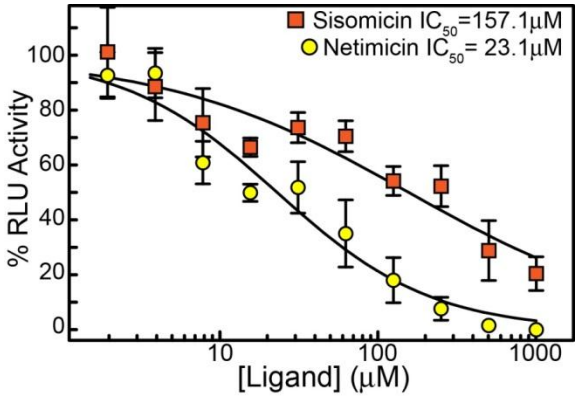


Figure 19 | Netilmicin's activity in inhibiting HIV-1 replication was monitored in the HIV-1 indicator cell line TZM-bl, expressing luciferase upon HIV-1 infection. The error bars in this plot were calculated from duplicate measurements.

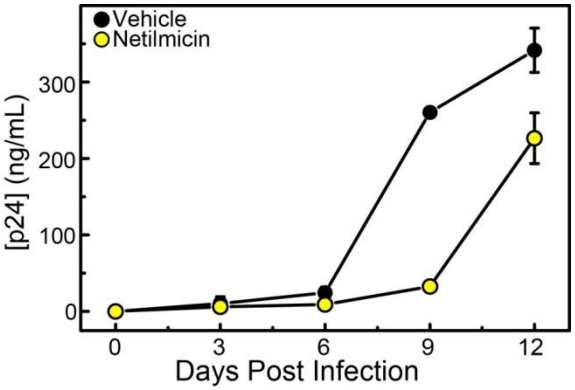


Figure 20 | Netilmicin inhibits HIV-1 replication *in* vivo in the T-cell line Hut78. Cells were first infected with HIV-1 isolate NL4-3, monitoring HIV-1 replication at three-day intervals by p24 ELISA. Error bars were obtained from duplicate measurements.

Netilmicin is the first experimentally validated compound designed and monitored for *in vivo* activity that was identified using a virtual screen that takes into account the dynamic excursions of RNA. As a molecule, netilmicin exhibits exquisite selectivity in binding to HIV-1 TAR, important in determining its activity. This high level of specificity both *In Vitro* and *In Vivo* is achieved by a single ethyl substituent on a key cationic amine group, outperforming related molecule sisomicin (**Figure 6**). When comparing the binding modes of netilmicin and sisomicin it was revealed that this modification alone significantly perturbs the binding of netilmicin to tRNA (**Figure 7**) and RRE (**Figure 12**) without altering its binding affinity for HIV-1 TAR. This alkyl group may play a role in stereo-chemically blocking access to the cationic group in the more rigid and less adaptable RNA binding pockets of tRNA and RRE versus the more flexible and malleable HIV-1 TAR. Consistent with this notion, it was observed that reducing the flexibility of TAR flexibility by deleting a single cytosine residue, which was not observed to make direct contacts with netilmicin, results in a 16-fold reduction in the binding affinity of netilmicin (**Figure 13**). Previous studies have shown that the flexibility of TAR enhances its ability to accommodate small molecules with conformational restrain as compared to the less malleable A-site[31], supporting the notion that stereochemical crowding of key cationic groups on small molecules may prove to be a general strategy for improving the selectivity of small molecules for highly flexible RNAs.

This method for targeting highly flexible RNA receptors can be implemented to additional highly flexible targets such as intrinsically unfolded proteins implied in neurodegenerative diseases as traditional structure and assay-based approaches have thus far proved ineffective, and for which computational dynamic ensembles are beginning to emerge guided by NMR experimentation[32]. Utilizing an experimentally informed atomic level

understanding of dynamics, a new method for the rational discovery of therapeutics that target their flexible structures can emerge.

Methods

All docking simulations were performed using the program Internal Coordinate Mechanics (ICM, Molsoft LLC. La Jolla, CA)[9] using a rigid RNA receptors with fully flexible ligand. TAR binding pockets were defined using ICM' PocketFinder module while the small molecules were docked at protonation states of 5.4, 7.4, and 9.4 as predicted using ChemAxon© (www.chemaxon.com). All purchased small molecules were obtained from Maybridge, Chembridge, LKT Labs, and Sigma. All were guaranteed to be ≥ 95% pure with the exception of sisomicin (purity ≥ 80%). All fluorescence intensity (FI) measurements were collected in triplicate by a Fluoromax-2 Fluorimeter. Fluorescently-labeled TAR was prepared commercially (Integrated DNA Technologies) and diluted to 0.1 μM for experimentation with or without a 100-fold excess (10 μM) of *S. Cerevisiae* tRNA (Sigma-Aldrich). All RNA was reannealed immediately before performing the experiment by heating on 95^{o} C for 5 minutes and immediately cooling on ice for an hour. Fluorescence polarization (FP) measurements were collected in triplicate using 384 well plate reader using a PHERAstar Plus plate reader (BMG LABTECH) with a 485 nm excitation wavelength and corresponding 520 nm detection wavelength optic module. *In vitro* synthesized elongated TAR (60 nM) was used to increase the dynamic range of the FP measurements while commercially prepared N-labeled fluorescein Tat peptide was directly monitored for polarization(10 nM) (Genscript). The IC_{50} and K_i values were calculated using Prizm software (GraphPad Software Inc.) with K_i values and errors reported in **Figure 8** within the 95% confidence interval. UV absorption spectra were recorded for all identified inhibitors to ensure no spectral overlap with both fluorescein and 2-aminopurine. NMR experiments using an Avance Bruker 600 MHz spectrometer equipped with 5 mm triple-

resonance cryogenic probe were performed at 25 °C. All NMR spectra was processed and analyzed using NMRPipe[33] and SPARKY 3[34].

Transfection assays included the pretreatment of cells with either small molecule, vehicle, or water as a control for 24 hours prior to transfection. The data shown in **Figure 18** was normalized to relative Renilla luciferase activity, representing the average of three independent transfections. Student's T-test was used in comparison of Tat and netilmicin treatment (Tat + Net) to Tat and vehicle treatment (Tat) to obtain the p-values in **Figure 18**. A TZM-bl HIV-1 indicator cell line which expresses luciferase upon HIV-1 infection, was plated in 96-well plates and correspondingly treated with either netilmicin or control (water), 24 hours prior to infection. The cells were then infected with HIV-1 strain NL4-3. 48 hours after infection, luciferase activity was quantified using relative light units (RLU). Values were normalized then using non-linear regression an IC_{50} value was computed using Prism. The T-cell line Hut78 was infected with HIV-1 isolate NL4-3 and viral replication was assessed using p24 ELISA. Half media (v/v) was harvested then replaced with uninfected Hut78 cells with or without 100 μM netilmicin at three day intervals.

Docking

Docking simulations were performed using the ICM (Molsoft LLC) docking program[9]. The program employs internal coordinate mechanics using bond angles, bond lengths, torsion angles, and phase angles for a grid-based docking simulation to both efficiently determine interaction energies while also optimizing the small molecule binding conformations in receptor-ligand complexes. For all cases, the RNA structures were converted into ICM objects. The binding pockets were predefined for known RNA-small molecule structures and identified using ICM PocketFinder for unknown binding sites. For structures with multiple binding sites each

pocket was docked independently recording only the minimum score for each small molecule. Small molecules either hand drawn using ChemDraw (CambridgeSoft Corp.) were converted into SDF format for docking simulations from the CCG (University of Michigan) or from an in-house library of over 2,000. OpenBabel software was used for .sdf conversation. The major microspecies module of ChemAxon© (www.chemaxon.com) was used to define protonation states over the pH range 5.4, 7.4, and 9.4. Each small molecule protonation state was docked independently recording only the lowest score corresponding to the most favorable binding interaction. For the validation simulations of known RNA-small molecule structure (**Figure 2**), protonation states were assigned as reported in literature.

Benchmark of Docking using Known RNA Structures

The structures of all RNA-small molecule complexes (excluding ribosome) as of January, 2009 were downloaded from the Protein Data Bank (www.pdb.org). This included 96 original structures including riboswitches, aptamers, and various groove-binding RNA-small molecules defined with $N_{flex} < 20$. 48 RNA-small molecule structures (22 with metal-mediated interactions and 26 without a reported K_d) were excluded from correlation. The 22 metal-mediated contact structures were excluded due to limitations in the ICM force-field that does not take into account the local interaction energies. For all simulations, RNA was separated from the small molecule, converted into an ICM object, with the receptor binding pocket defined as all receptor atoms within 5 Å of the bound small molecule. A maximum number of Monte-Carlo iterations (thoroughness of 10) were used in the docking simulations.

Docking Known TAR binders

20 conformers in the TAR^{NMR-MD} ensemble were generated as previously described[6] from one round of SAS selection. The X-ray (TAR^{X-ray}) (397D) [13] and NMR (TAR^{NMR}) (1ANR) [14]

structures of apo-TAR were downloaded from the PDB. Chemical structure of 38 known TAR-binding compounds was obtained from literature[16,19,31,35-42]. The TAR^{MD} ensemble was constructed at random from an 80 ns simulation of apo-TAR. The binding pockets were predicted using ICM's PocketFinder module, calculating the surface area and volume of cavities on the receptor surface. Binding pockets were included if they were in the "druggable" range (volume=150-500 $Å^3$ and surface area=200-550 $Å^2$), as determined using ICM. These pockets were set as the receptor centroids and the binding pocket consisting of all atoms with 5 Å around this centroid.

Virtual screening of the TAR^{NMR-MD} ensemble

The 20 TAR^{NMR-MD} conformers were docked against 51,000 small molecules as described above by initially docking all small molecules with $N_{flex}<20$ using a thoroughness of 1. Next a second set of simulations in which the top ~10% of all scoring small molecules were re-docked using a thoroughness of 10. The library itself consisted of ~49,000 small molecules from the Center for Chemical Genomics at the University of Michigan in high throughput screening with ~2,000 small molecules reported to bind RNA generated in-house.

Fluorescence-based binding assay

A fluorescence-based assay was employed a TAR construct labeled with 2-aminopurine at bulge residue U25 (TAR-ap25) or fluorescein attached to bulge residue U25 (TAR-fl25) as was the case for the binding of DMA due to spectral overlap[16]. RNAs were purchased commercially from Integrated DNA Technologies. RNA was annealed and by diluted (100 nM) into working buffer (10 mM phosphate, 20 mM NaCl, 0.1 mM EDTA, pH 6.8). Samples were pre-equilibrated for 5 minutes following addition of small molecule before measurement. An

excitation wavelength of 320 nm and emission wavelength of 390 nM was used for for TAR-ap25, while an excitation wavelength of 485 nm and an emission wavelength of 520 nm used for TAR-fl25. The TAR-fl25 binding assay recapitulated within experimental error the K_d for neomycin (0.52 ± 0.08 μM) with the TAR-ap25 assay (0.50 ± 0.16 μM). Fluorescence intensity measurements were recorded in triplicate following addition of small molecules with values normalized to unbound-TAR. All K_ds were calculated using non-linear regression analysis by fitting the measured fluorescence intensities (F) to the following equation[16] with the Origin software (OriginLab Corporation),

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

$[RNA]_T$ is the total TAR concentration (100 nM), $[L]_T$ is the total ligand concentration, A and B are weighting factors allowed to float during curve fitting. The assay was validated by measuring K_ds for known TAR binders neomycin B[18], paromomycin[19], kanamycin[19], ribostamycin[19], tobramycin[19], and argininamide[18] (**Figure 3**).

Similar binding assays were used to measure binding of the small molecules to HIV-2 TAR[16,43], prokaryotic A-site[26] , RRE[25], and HIV-2* TAR (**Figures 11 and 13**). Competition binding assays, analogous to those described above using tRNA were performed with unlabelled A-site, RRE, and HIV-2 RNA constructs (**Figure 14**).

Fluorescence Polarization TAR-Tat Inhibition Assay

A fluorescence polarization (FP) displacement assay was used to measure the inhibition of TAR-Tat binding by the small molecules. The assay employed a N-terminus fluorescein labeled Tat peptide (F-Tat) of sequence N-AAARKKRRQRRR-C, (**Figure 4**) (Genscript Corp),

containing the arginine rich motif (RKKRRQRRR) necessary for interaction with TAR[36]. Elongated HIV-1 TAR (E-TAR) (**Figure 4**) was prepared by *in vitro* transcription, purified by PAGE, and employed to increase the dynamic range of FP measurement. The Fluorescence Polarization buffer consisted of 50 mM Tris, 100 mM NaCl, and 0.01% nonidet-P40 with pH = 7.4. The binding of F-Tat to E-TAR was first assayed by monitoring FP of F-Tat (10 nM) as a function of increasing E-TAR concentrations. The apparent K_d was calculated by fitting observed FP (FP_{obs}) to the following equation[44] again using the Origin software (OriginLab Corporation),

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

where FP_{free} and FP_{bound} were measured for free and bound F-Tat, respectively, $[L]_T$ is the concentration of F-Tat, and $[RNA]_T$ is the total E-TAR concentration. The parameters P_{bound}, P_{free}, and K_d were allowed to float during the fit. The K_d (~40 nM) obtained was is in very good agreement previously reported values in the literature (25 nM) [36].

Fluorescence polarization measurements upon increasing concentrations of small molecules were performed by first incubating E-TAR (60 nM) with increasing concentrations of small molecules for 10 mins followed by the addition of F-Tat (10 nM) and incubation for 10 mins. For each small molecule, data was collected for 30 minutes at 5 minute intervals to ensure full equilibration. FP data was collected in triplicate with IC_{50} values calculated after fitting the observed FP (FP_{obs}) to the following equation from the Prizm software (GraphPad Software Inc.),

$$FP_{obs} = FP_{free} + \frac{(FP_{bound} - FP_{free})}{1 + (\frac{[I]}{IC_{50}})^{m}},$$

where FP_{free} and FP_{bound} were the FP values measured for free and E-TAR bound F-Tat peptide, respectively, [*I*] was the log-based concentration of inhibitor, IC_{50} is the concentration of small molecule at which 50% of TAR-Tat binding was observed, and *m* is the slope of the linear portion of the sigmoidal curve. IC_{50} and *m* were allowed to float during analysis. The inhibition constants (K_i) were then calculated using the Prizm software (GraphPad Software Inc.) employing equation,

$$\log(C_0) = \log(10^{\alpha K_d}) \times (1 + \frac{[FTat]_T}{K_d}),$$

where $[FTat]_T$ is the concentration of F-Tat and K_d is the dissociation constant between F-Tat and E-TAR. The K_is reported are based on the 95% confidence interval. This displacement assay was first benchmarked using known TAR binders neomycin B, paromomycin, kanamycin, ribostamycin, tobramycin, and argininamide (**Figure 5**).

NMR spectroscopy

Uniformly $^{13}C/^{15}N$ isotopically labeled TAR (**Figure 1**) was prepared using *in-vitro* transcription and purification on PAGE as described previously[45]. NMR buffer solution consisted of 15 mM sodium phosphate, 0.1 mM EDTA, 25 mM NaCl, and 10% D_2O with pH ~6.4. The chemical shift mapping experiments were performed by recording 2D HSQC spectra following addition of small molecules from a stock solutions (100 mM) to $^{13}C/^{15}N$ labeled TAR (~0.1-0.2 mM). The chemical shift perturbation magnitudes ($\Delta\delta_{av}$) shown in **Figure 17** were calculated using

$\Delta\delta_{av} = \sqrt{(\Delta\delta_H)^2 + (\frac{\gamma_C}{\gamma_H}(\Delta\delta_C))^2}$, where $\Delta\delta_H$ and $\Delta\delta_C$ were the change in ^{1}H and ^{13}C chemical shift (in ppm), respectively, and γ_C and γ_H are the gyromagnetic ratios for carbon and hydrogen, respectively. Chemical shift perturbations were used to measure the dissociation constant (K_d) for butirosin (**Figure 7**), by fitting chemical shift perturbations at varying butirosin concentrations using the equation fit with Origin software (OriginLab Corporation),

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

where δ_{obs} is the observed weight averaged chemical shift perturbation measured for each ligand concentration, δ_{free} is the chemical shift in the ligand-free state (in ppm), $[But]_T$ is the total butirosin concentration for each point, $[TAR]_T$ is the total TAR concentration based on UV absorbance measurements at 260 nm, $\Delta\delta_T$ is the measured change in chemical shifts between the free and bound states (in ppm). $\Delta\delta_t$ and K_d were allowed to float during the fit using Origin software (OriginLab Corporation).

NMR Validation of Bound TAR Structures

The populations of TAR conformers shown in **Figure 15** were determined using a partition function: $P = \frac{e^{-\Delta G_i / R}}{\sum_i e^{-\Delta G / R}}$. The apparent K_ds were calculated using each docking score and the best-fit line determined in **Figure 2**. The six hit molecules were re-docked against the 20 $TAR^{NMR\text{-}MD}$ conformers 5 times with populations for each conformer calculating an average over the 5 individual runs (shown in **Figure 15**). Sisomicin and amikacin were found to have the largest standard deviation in score between runs (8.7 kcal/mol and 6.0 kcal/mol respectively)

while the other four small molecules had significantly smaller standard deviations <2.6 kcal/mol. Correspondingly, sisomicin and amikacin were found to have the widest distribution of populations across the 20 TAR^{NMR-MD} conformers. **Figure 17** was constructed showing the TAR conformer with highest frequency selection over the five runs for mitoxantrone (conformer 15), sisomicin(conformer 15), netilmicin (conformer 18), butirosin (conformer 18), and DMA (conformer 14) in the lowest scoring binding pose. Amikacin was found to bind two distinct conformers with equal frequency, shown is the lowest scoring conformer of the two (18).

A description of the database of small molecules used in the virtual screen can be found at http://www.lsi.umich.edu/facultyresearch/centers/chemicalgenomics/ccg-resources/libraries. It is of note that the library recently expanded to nearly >3x the original size used in our study. The small molecules in the in-house library were manually drawn from published reports of verified RNA-binding small molecules using the program ChemDraw (CambridgeSoft) and saving files as in sdf format. A total of 49,166 and 2060 small molecules from the library of CCG and in-house, respectively, were screened against the 20 TAR NMR-MD structures. 57 small molecules (excluding spermine) were purchased commercially and originally screened *In Vivo* using the TAR-Tat fluorescence polarization assay. Four small molecules (including spermine) were from the CCG library with three small molecules from the in-house library were confirmed as positive inhibitors of the TAR-Tat interactions and subject to subsequent analysis *in vitro*, *in cell*, and *in vivo*.

Table S1 | Screening small molecules randomly. The small molecules were randomly selected from the CCG and in-house RNA library respectively and purchased when commercially available. All compounds were dissolved in water or DMSO (up to 50mM) and tested using the same three assays used to test the compounds selected using docking simulations. This includes a fluorescence polarization-based Tat displacement assay (FP), a fluorescence binding assay (F), followed by NMR chemical shift mapping experiments upon any observable changes in the polarization assays (N). Compounds were screened using FP and FI with only a small subset tested by NMR. All compounds that not tested using NMR are indicated (X). Compounds not tested in the FP assays due to interactions with the Tat peptide are indicated using "nd". For the FP assay, a positive hit was defined as a compound inducing 20% decrease in polarization upon the addition of 1 mM compound. Stars (*) indicate compounds tested with fluorescein (as opposed to 2-AP) labeled TAR in the FI assay. No compounds yielded detectable perturbations in the NMR chemical shift mapping experiments.

Compound Name	Vendor (ID)	Library	Assays		
			FP	FI	N
(6-hydroxy-4-methyl-2-thioxo-1,2-dihydro-5-pyrimidinyl)acetic acid	Chembridge (5728034)	CCG	–	–*	X
1-(1,1-dioxidotetrahydro-3-thienyl)-3,5-dimethyl-1H-pyrazole	Chembridge (5772356)	CCG	–	–	X
1-(3-chloro-2-buten-1-yl)pyrrolidine hydrochloride	Chembridge (5920818)	CCG	–	–*	X
1-[2-amino-2-(hydroxyimino)ethyl]pyridinium chloride	Chembridge (5511239)	CCG	–	–*	X
1-methyl-4-(3-oxo-1-buten-1-yl)pyridinium iodide	Chembridge (5567408)	CCG	nd	+*	-
1,1'-[[2,5-bis(methylthio)-3,4-thienediyl]bis(methylene)]dipyridinium diiodide	Chembridge (5119164)	CCG	nd	–*	X

1,1'-methylenebis(3-methylpyridinium) dibromide	Chembridge (5480934)	CCG	–	–	X
1,3,5-triazine-2,4,6-triamine - 1H-tetrazol-5-amine (1:1)	Chembridge (5510986)	CCG	–	–	X
2-(4-chlorophenoxy)-N-[3-(dimethylamino)propyl]acetamide	Chembridge (5533542)	CCG	–	–	X
2-(ethylthio)ethyl imidothiocarbamate hydrochloride	Chembridge (5104084)	CCG	–	–	X
2-[2-(2,5-dioxo-1-pyrrolidinyl)-1-methylethylidene]hydrazinecarboximidamide nitrate	Chembridge (5210259)	CCG	–	+	X
2-amino-1-(2-amino-2-oxoethyl)-5-bromopyridinium iodide	Chembridge (5122052)	CCG	–	–*	x
2-amino-1,3-thiazole-5-sulfonic acid	Chembridge (5140941)	CCG	–	+	x
2-amino-2-deoxyhexopyranose hydrochloride	Chembridge (5152335)	CCG	–	–	x
2,2'-(methylenedisulfonyl)diacetic acid	Chembridge (5144188)	CCG	–	–	x
2,3-dihydro-1,4-benzodioxin-2-ylmethyl imidothiocarbamate hydrobromide	Chembridge (5705537)	CCG	–	–	x
3-{[(2-hydroxyethyl)amino]carbonyl}-1-methylpyridinium iodide	Chembridge (5365449)	CCG	nd	–	x
3-hydroxy-1-methylquinolinium methyl sulfate	Chembridge (5175189)	CCG	–	–*	x
4-({[3-(methoxycarbonyl)phenyl]amino}carbonyl)-1-methylpyridinium iodide	Chembridge (5566387)	CCG	nd	–*	x
4-[5-(4-fluorophenyl)-1,3-oxazol-2-yl]-1-methylpyridinium benzenesulfonate	Chembridge (5605903)	CCG	–	–*	x
4-amino-N,N,N-trimethylbenzenaminium iodide	Chembridge (5132278)	CCG	nd	–*	x
4-bromo-N,N,N-trimethylbenzenaminium methyl sulfate	Chembridge (5139915)	CCG	–	–	x
5-(2-anilinovinyl)-1-methyl-3,4-dihydro-2H-pyrrolium methyl sulfate	Chembridge (5233738)	CCG	nd	–*	x
5-(3,3-dimethyl-2-oxobutyl)-1-methyl-2-(3,4,5-trimethoxyphenyl)-1H-imidazo[4,5-c]pyridin-5-ium bromide	Chembridge (5540344)	CCG	–	–*	x

5-(ethoxycarbonyl)-1,3-bis(2-hydroxyethyl)-2-methyl-1H-3,1-benzimidazol-3-ium iodide	Chembridge (5119924)	CCG	nd	–	x
5-[2-(4-methoxyphenyl)-2-oxoethyl]-1-methyl-2-(2,4,5-trimethoxyphenyl)-1H-imidazo[4,5-c]pyridin-5-ium bromide	Chembridge (5554490)	CCG	–	+*	x
6-amino-3-pyridinesulfonic acid	Chembridge (5140823)	CCG	–	–*	x
6-hydroxy-2-(4-morpholinylmethyl)-3(2H)-pyridazinone	Chembridge (5404082)	CCG	–	–*	x
dimethyl 4-amino-5-hydroxy-2-(2-methoxy-2-oxoethyl)-2H-pyrrole-2,3-dicarboxylate	Chembridge (5807958)	CCG	–	–*	x
disodium 3,3'-thiodi(1-propanesulfonate)	Chembridge (5135337)	CCG	nd	–	x
ethyl 4-[2-amino-4-(5-methyl-2-oxo-2,3-dihydro-1H-imidazol-4-yl)-1,3-thiazol-5-yl]butanoate hydrobromide	Chembridge (5819199)	CCG	–	–*	x
N-(2-nitrobenzylidene)-4H-1,2,4-triazol-4-amine	Chembridge (5570335)	CCG	–	–*	x
N-[3-(3,5-dimethylphenoxy)propyl]-1-butanamine	Chembridge (5571803)	CCG	nd	–*	x
N-[3-(dimethylamino)propyl]-5-nitro-2-furamide	Chembridge (5528274)	CCG	–	+*	–
N-[6-(dimethylamino)hexyl]-2-(hydroxyimino)acetamide hydrochloride	Chembridge (5556986)	CCG	–	–	x
N~5~-[(acetylamino)(amino)methylene]ornithine	Chembridge (5626440)	CCG	–	–	x
potassium 4-carboxy-1,2,5-thiadiazole-3-carboxylate hydrate	Chembridge (5131016)	CCG	–	–*	x
Adenosine	Sigma (A9251)	RNA	–	–*	x
Amprolium	Sigma (31592)	RNA	–	–*	x
Anisomycin	Sigma (A9789)	RNA	–	–*	x
Dopamine	Sigma (H8502)	RNA	–	–	x
Doxycycline	Sigma (D9891)	RNA	nd	–*	x
Glycine	Sigma (G7126)	RNA	–	–	x
Isoniazid	Sigma (I3377)	RNA	–	–*	x
L-leucine	Sigma (L8000)	RNA	–	–	x
L-threonine	Sigma (T8625)	RNA	–	–	x
L-Valine	Sigma (V0500)	RNA	–	–	x
Lincomycin	Sigma (L2774)	RNA	–	–*	x

Oxythiamine	Sigma (O4000)	RNA	–	–	x
Pefloxacin	Sigma (P0106)	RNA	–	–*	x
Puromycin	Sigma (P7255)	RNA	nd	–*	x
S-adenosyl-methionine	Sigma (A7007)	RNA	–	–	x
Tetracycline	Sigma (87128)	RNA	–	–*	x
Theophylline	Sigma (T1633)	RNA	–	–*	x
Thiamin Pyrophosphate	Sigma (C8754)	RNA	–	–	x
Tyramine	Sigma (T90344)	RNA	–	–*	x
Vancomycin	Sigma (V2002)	RNA	nd	–*	x

References

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

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

3 Zhang, Q., Stelzer, A. C., Fisher, C. K. & Al-Hashimi, H. M. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**, 1263-1267 (2007).

4 Latham, M. P., Zimmermann, G. R. & Pardi, A. NMR chemical exchange as a probe for ligand-binding kinetics in a theophylline-binding RNA aptamer. *J. Am. Chem. Soc.* **131**, 5052-5053 (2009).

5 Vaiana, A. C. & Sanbonmatsu, K. Y. Stochastic gating and drug-ribosome interactions. *J. Mol. Biol.* **386**, 648-661 (2009).

6 Frank, A. T., Stelzer, A. C., Al-Hashimi, H. M. & Andricioaei, I. 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 (2009).

7 Chen, Y., Campbell, S. L. & Dokholyan, N. V. Deciphering protein dynamics from NMR data using explicit structure sampling and selection. *Biophys J* **93**, 2300-2306 (2007).

8 Clore, G. M. & Schwieters, C. D. 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 (2004).

9 Abagyan, R., Totrov, M. & Kuznetsov, D. ICM - A new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488-506 (2004).

10 Lang, P. T. *et al.* DOCK 6: Combining techniques to model RNA-small molecule complexes. *RNA* **15,** 1219-1230 (2009).

11 Cheng, T., Li, X., Li, Y., Liu, Z. & Wang, R. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **49**, 1079-1093 (2009).

12 Guilbert, C. & James, T. L. Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *J. Chem. Inf. Model.* **48**, 1257-1268 (2008).

13 Ippolito, J. A. & Steitz, T. A. 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 (1998).

14 Aboul-ela, F., Karn, J. & Varani, G. 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, (1996).

15 Yang, M. Discoveries of Tat-TAR interaction inhibitors for HIV-1. *Curr. Drug Targets Infect. Disord.* **5**, 433-444 (2005).

16 Bradrick, T. D. & Marino, J. P. Ligand-induced changes in 2-aminopurine fluorescence as a probe for small molecule binding to HIV-1 TAR RNA. *RNA* **10**, 1459-1468 (2004).

17 Matsumoto, C., Hamasaki, K., Mihara, H. & Ueno, A. A high-throughput screening utilizing intramolecular fluorescence resonance energy transfer for the discovery of the molecules that bind HIV-1 TAR RNA specifically. *Bioorg. Med. Chem. Lett.* **10**, 1857-1861 (2000).

18 Nandi, C. K., Parui, P. P., Brutschy, B., Scheffer, U. & Gobel, M. Fluorescence correlation spectroscopy at single molecule level on the Tat-TAR complex and its inhibitors. *Biopolymers* **89**, 17-25 (2008).

19 Blount, K. F., Tor, Y., Hamasaki, K. & Ueno, A. Using pyrene-labeled HIV-1 TAR to measure RNA-small molecule binding Aminoglycoside antibiotics, neamine and its derivatives as potent inhibitors for the RNA-protein interactions derived from HIV-1 activators. *Nucleic Acids Res.* **31**, 5490-5500 (2003).

20 Tao, J. & Frankel, A. D. Specific binding of arginine to TAR RNA. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2723-2726 (1992).

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

22 Davidson, A. *et al.* Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11931-11936 (2009).

23 White, R. J. & Durr, F. E. Development of mitoxantrone. *Invest. New Drugs* **3**, 85-93 (1985).

24 Parolin, C. *et al.* New anti-human immunodeficiency virus type 1 6-aminoquinolones: mechanism of action. *Antimicrob. Agents Chemother.* **47**, 889-896 (2003).

25 DeJong, E. S., Chang, C. E., Gilson, M. K. & Marino, J. P. Proflavine acts as a Rev inhibitor by targeting the high-affinity Rev binding site of the Rev responsive element of HIV-1. *Biochemistry* **42**, 8035-8046 (2003).

26 Kaul, M., Barbieri, C. M. & Pilch, D. S. Fluorescence-based approach for detecting and characterizing anti biotic-induced conformational changes in ribosomal RNA: Comparing aminoglycoside binding to prokaryotic and eukaryotic ribosomal RNA sequences. *J. Am. Chem. Soc.* **126**, 3447-3453 (2004).

27 Stelzer, A. C., Kratz, J. D., Zhang, Q. & Al-Hashimi, H. M. RNA dynamics by design: biasing ensembles towards the ligand-bound state. *Angew. Chem. Int. Edit.* **49**, 5731-5733 (2010).

28 Lapidot, A., Berchanski, A. & Borkow, G. Insight into the mechanisms of aminoglycoside derivatives interaction with HIV-1 entry steps and viral gene transcription. *FEBS J.* **275**, 5236-5257 (2008).

29 Lapidot, A., Vijayabaskar, V., Litovchick, A., Yu, J. G. & James, T. L. Structure-activity relationships of aminoglycoside-arginine conjugates that bind HIV-1 RNAs as determined by fluorescence and NMR spectroscopy. *FEBS Lett.* **577**, 415-421 (2004).

30 Faber, C., Sticht, H., Schweimer, K. & Rosch, P. Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.* **275**, 20660-20666 (2000).

31 Blount, K. F., Zhao, F., Hermann, T. & Tor, Y. Conformational constraint as a means for understanding RNA-aminoglycoside specificity. *J. Amer. Chem. Soc.* **127**, 9818-9829 (2005).

32 Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789-796 (2009).

33 Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR* **6**, 277-293 (1995).

34 Goddard, T. D. & Kneller, D. G. *SPARKY 3*. (University of California, San Franscisco).

35 Ding, L., Zhang, X. X., Chang, W. B., Lin, W. & Yang, M. Studies of binding constants and interaction of drugs to trans-activation response RNA by capillary electrophoresis. *Anal. Chim. Acta* **543**, 249-253 (2005).

36 Hamasaki, K. & Ueno, A. Aminoglycoside antibiotics, neamine and its derivatives as potent inhibitors for the RNA-protein interactions derived from HIV-1 activators. *Bioorg. Med. Chem. Lett.* **11**, 591-594 (2001).

37 He, M. Z. *et al.* Synthesis and assay of isoquinoline derivatives as HIV-1 Tat-TAR interaction inhibitors. *Bioorg. Med. Chem. Lett.* **15**, 3978-3981 (2005).

38 Ironmonger, A. *et al.* Scanning conformational space with a library of stereo- and regiochemically diverse aminoglycoside derivatives: the discovery of new ligands for RNA hairpin sequences. *Org. Biomol. Chem.* **5**, 1081-1086 (2007).

39 Mayer, M. *et al.* Synthesis and testing of a focused phenothiazine library for binding to HIV-1 TAR RNA. *Chem. Biol.* **13**, 993-1000 (2006).

40 Mei, H.-Y. *et al.* Inhibition of an HIV-I Tat-derived peptide binding to TAR RNA by aminoglycoside antibiotics. *Bioorg. Med.Chem. Lett.* **5**, 2755-2760 (1995).

41 Xu, Y. L. *et al.* Synthesis and evaluation of novel neamine derivatives effectively targeting to RNA. *Bioorg. Med. Chem. Lett.* **19**, 2103-2106 (2009).

42 Zhao, P., Jin, H. W., Yang, Z. J., Zhang, L. R. & Zhang, L. H. Solid-phase synthesis and evaluation of TAR RNA targeted beta-carboline-nucleoside conjugates. *Org. Biomol. Chem.* **6**, 3741-3750 (2008).

43 Zhao, F. *et al.* Molecular recognition of RNA by neomycin and a restricted neomycin derivative. *Angew. Chem. Int. Edit.* **44**, 5329-5334 (2005).

44 Lundblad, J. R., Laurance, M. & Goodman, R. H. Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *J. Mol. Endocrin.* **10**, 607-612 (1996).

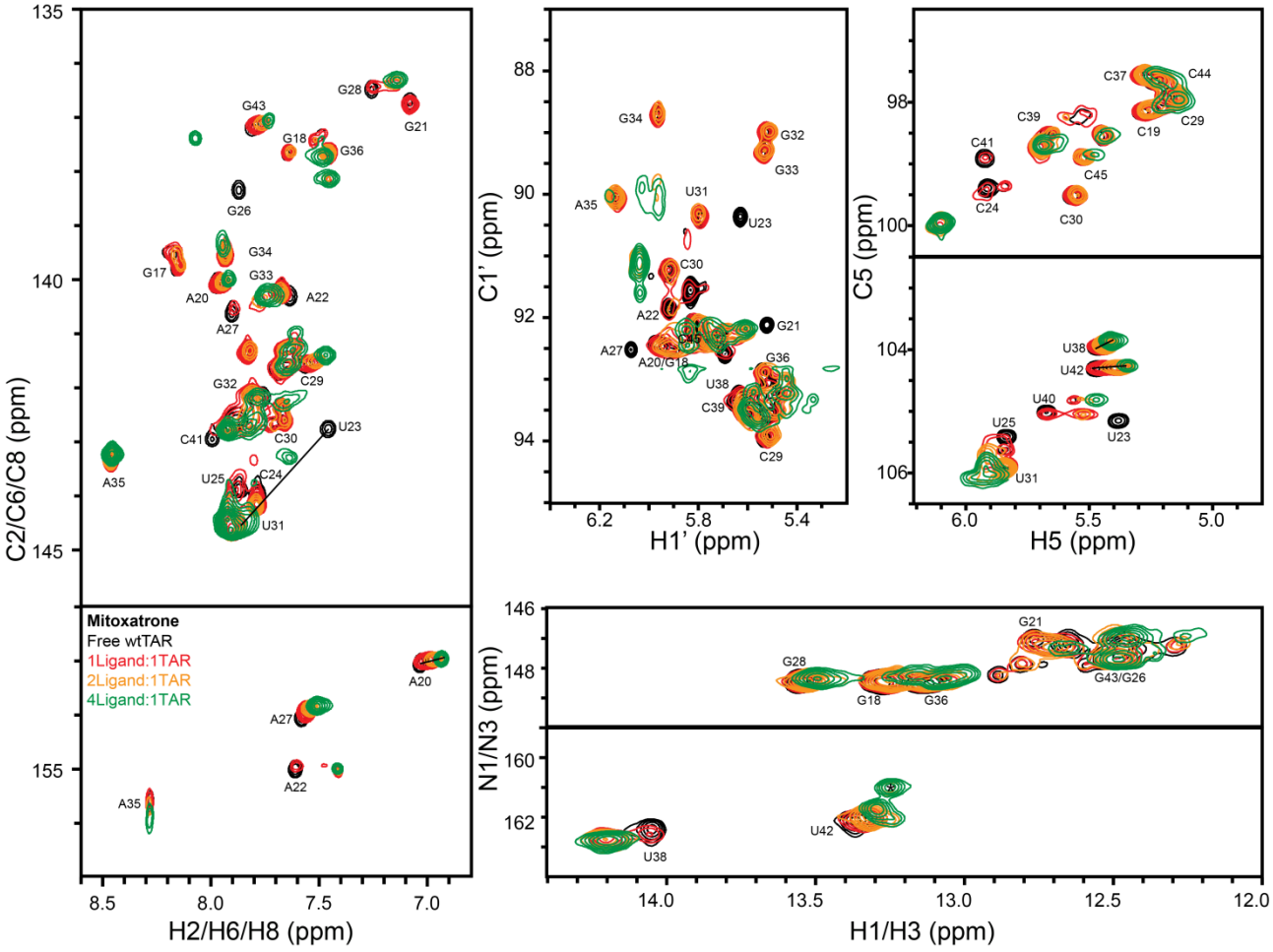
45 Zhang, Q., Sun, X., Watt, E. D. & Al-Hashimi, H. M. Resolving the motional modes that code for RNA adaptation. *Science* **311**, 653-656 (2006).

This work was published:

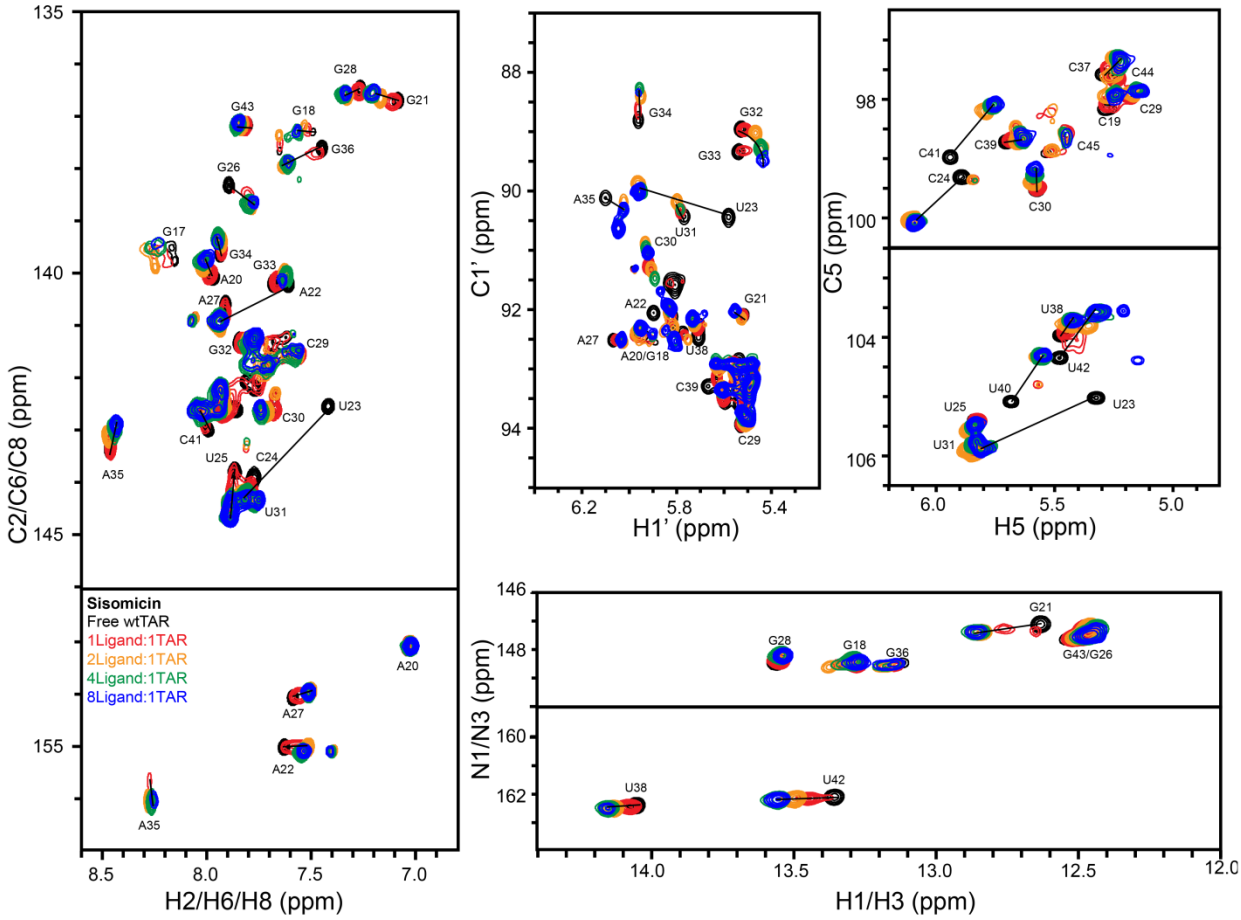
Stelzer, A. C., Frank, A. T., Kratz, J. D., Swanson, M., Gonzalez-Hernandez, M. J., Lee, J., Andricioaei, I., Markovitz, D. M., & Al-Hashimi, H. M. Discovery of Selective Bioactive Small Molecules by Targeting an RNA Dynamic. *Nat. Chem. Biol.* (Accepted) (2011).

Appendix I | TAR chemical shift perturbations a, Mitoxantrone**.** Shown are 2D HSQC spectra of free 100 μM TAR (black) and upon addition of 100 μM (red), 200 μM (orange), and 400 μM (green) mitoxantrone. **b,** Sisomicin**.** Shown are 2D HSQC spectra of free 200 μM TAR (black) and upon addition of 200 μM (red), 400 μM (orange), 800 μM (green), and 1600 μM (blue) sisomicin. **c,** Netilmicin**.** Shown are 2D HSQC spectra of free 200 μM TAR (black) and upon addition of 200 μM (red) and 400 μM (orange) netilmicin. **d,** Amikacin**.** Shown are 2D HSQC spectra of free 200 μM TAR (black) and upon addition of 200 μM (red), 400 μM (orange), 800 μM (green), and 1600 μM (blue) amikacin. **e,** Butirosin**.** Shown are 2D HSQC spectra of free 100 μM TAR (black) and upon addition of 100 μM (red), 200 μM (orange), 400 μM (green), and 650 μM (blue) butirosin. **f,** DMA. Shown are 2D HSQC spectra of free 200 μM TAR (black) and upon addition of 200 μM (red), 400μM (orange), 800 μM (green), and 1600 μM (blue) DMA.

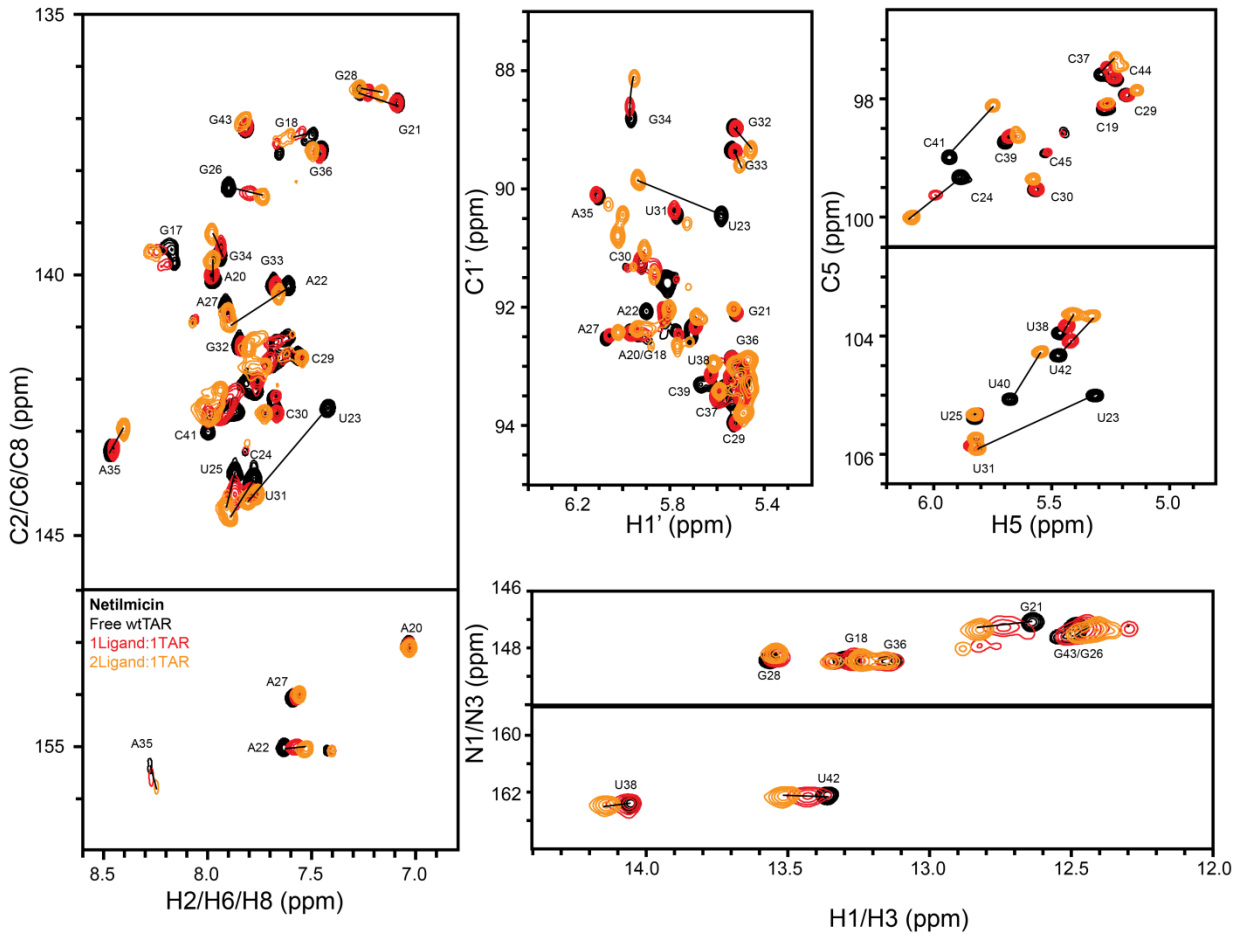
a



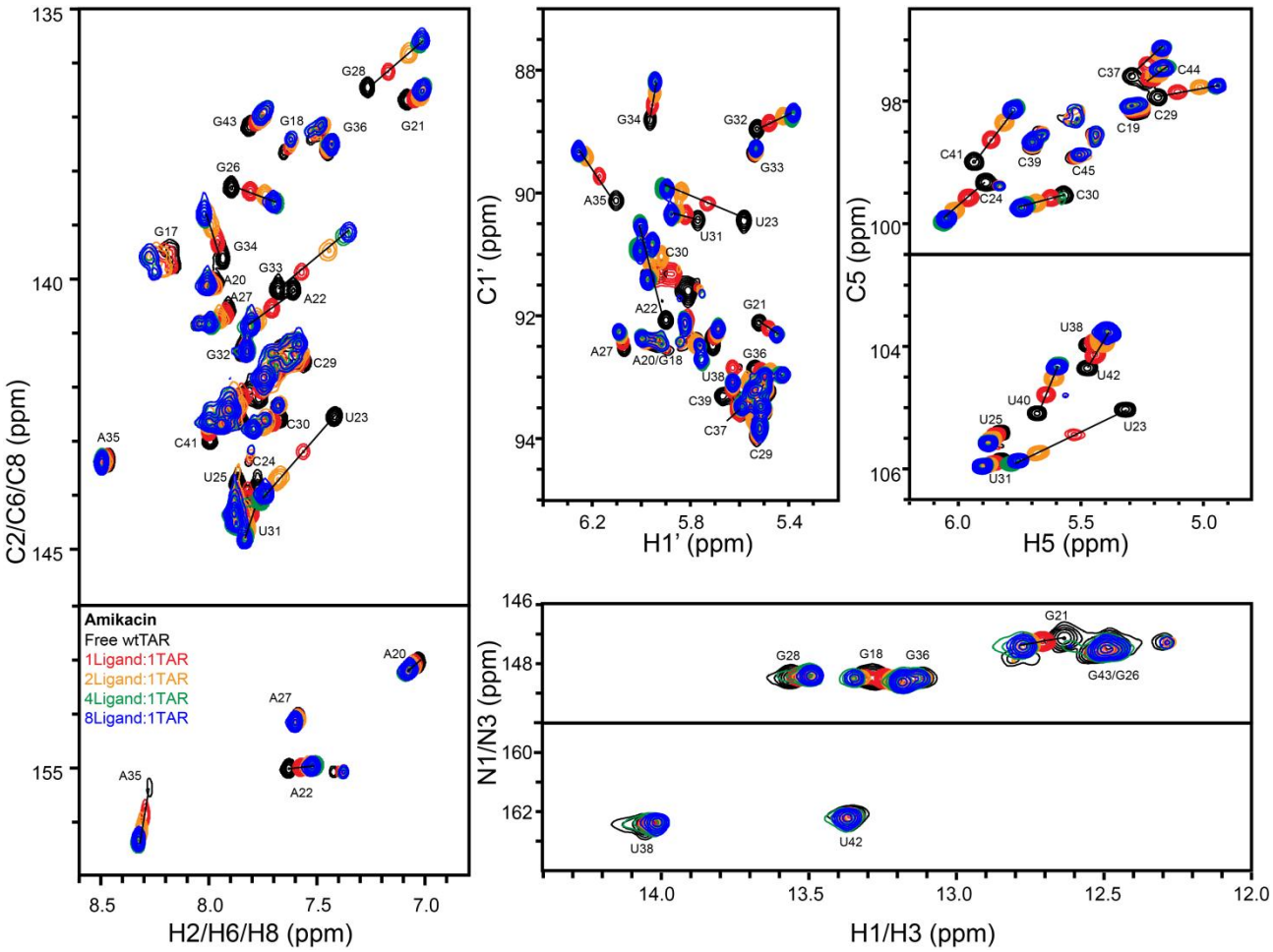
b


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

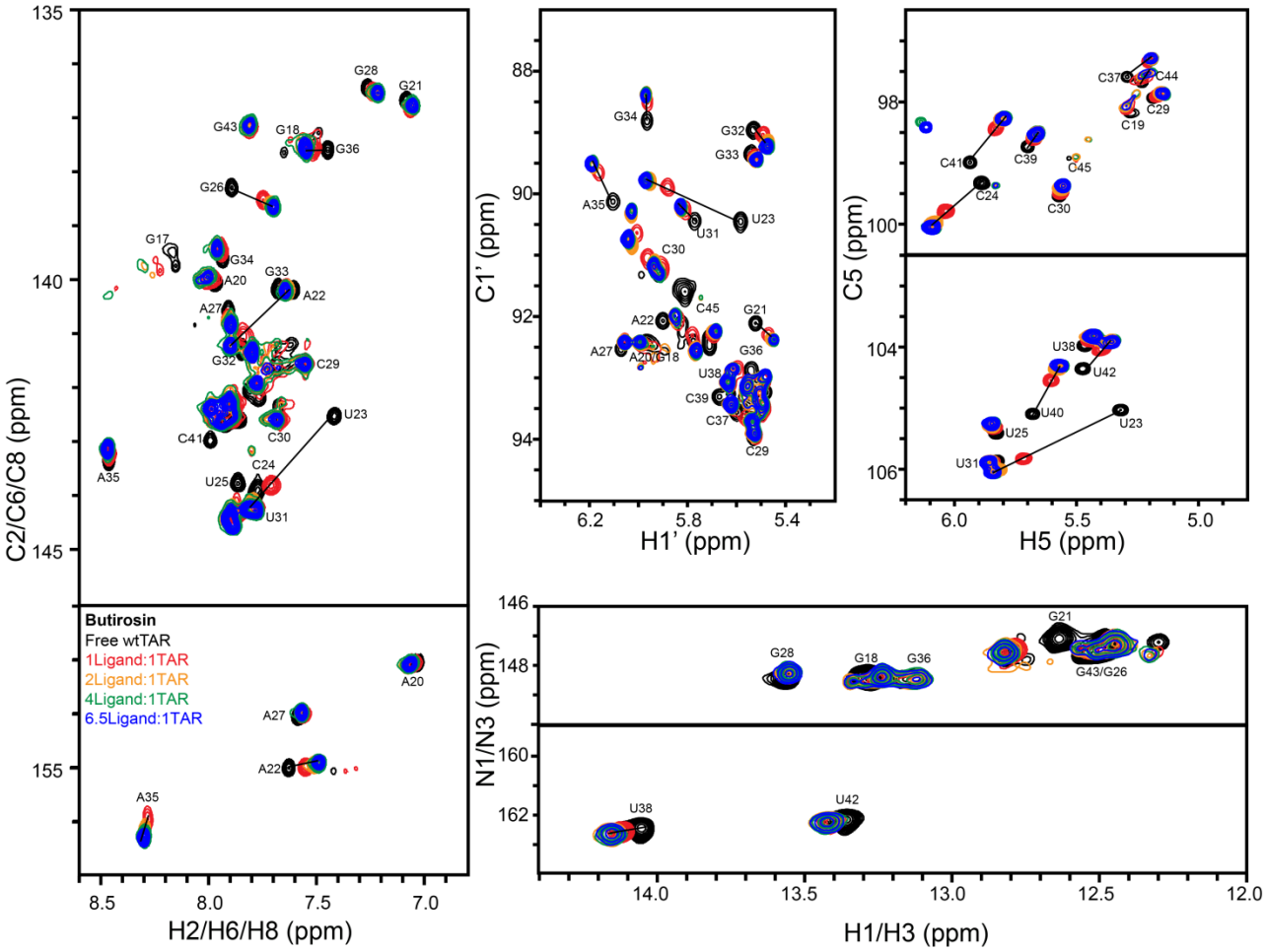
c


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

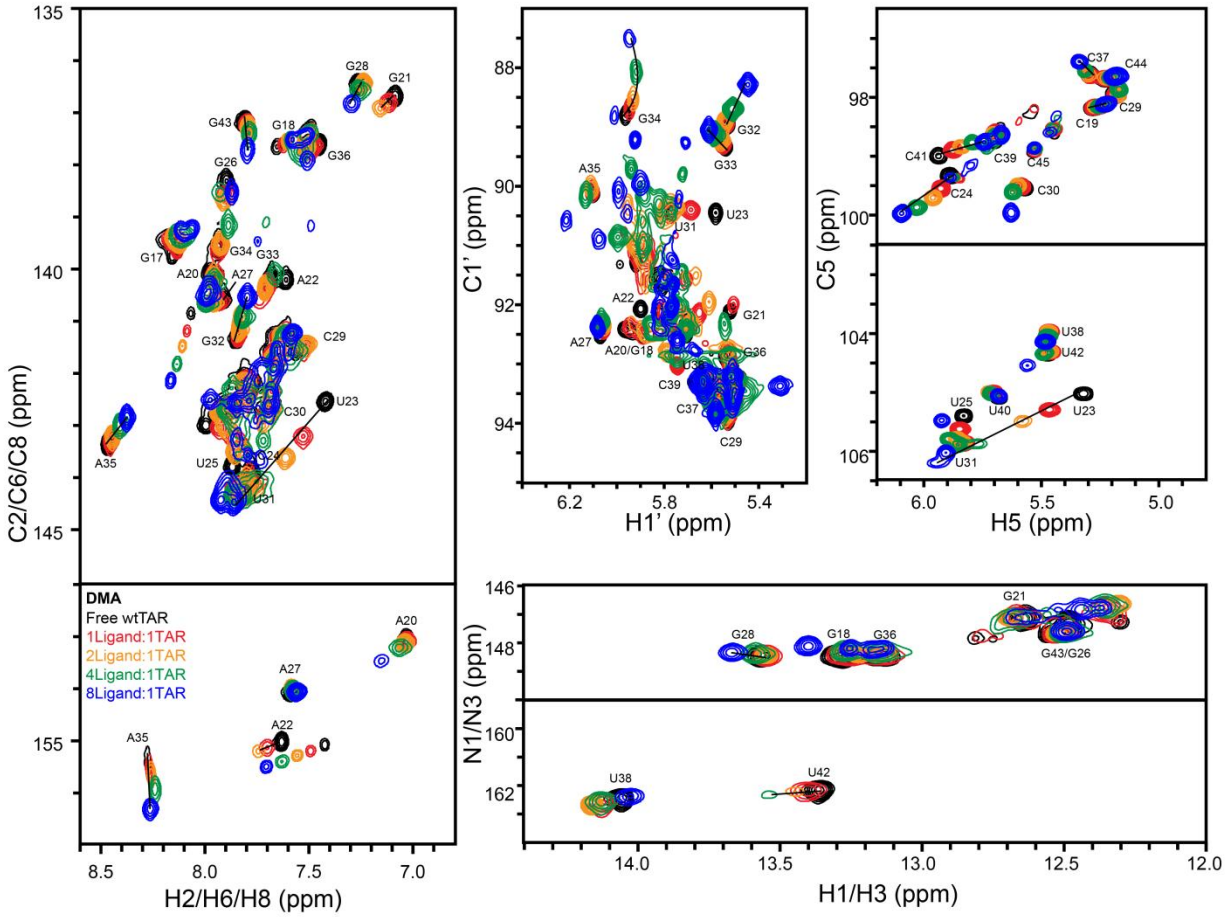
d


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

e


Butirosin
Free wtTAR
1Ligand:1TAR
2Ligand:1TAR
4Ligand:1TAR
6.5Ligand:1TAR
C2/C6/C8 (ppm)
H2/H6/H8 (ppm)
C1' (ppm)
H1' (ppm)
C5 (ppm)
H5 (ppm)
N1/N3 (ppm)
H1/H3 (ppm)
G43/G26

f


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