Exposing Hidden High-Affinity RNA Conformational States

Nicole I. Orlovsky[†], **Hashim M. Al-Hashimi**[*,†,‡], **Terrence G. Oas**[*,†,‡]

[†]Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, United States

[‡]Department of Chemistry, Duke University, Durham, North Carolina 27708, United States

Abstract

RNA recognition frequently results in conformational changes that optimize intermolecular binding. As a consequence, the overall binding affinity of RNA to its binding partners depends not only on the intermolecular interactions formed in the bound state but also on the energy cost associated with changing the RNA conformational distribution. Measuring these "conformational penalties" is, however, challenging because bound RNA conformations tend to have equilibrium populations in the absence of the binding partner that fall outside detection by conventional biophysical methods. In this study we employ as a model system HIV-1 TAR RNA and its interaction with the ligand argininamide (ARG), a mimic of TAR's cognate protein binding partner, the transactivator Tat. We use NMR chemical shift perturbations and relaxation dispersion in combination with Bayesian inference to develop a detailed thermodynamic model of coupled conformational change and ligand binding. Starting from a comprehensive 12-state model of the equilibrium, we estimate the energies of six distinct detectable thermodynamic states that are not accessible by currently available methods. Our approach identifies a minimum of four RNA intermediates that differ in terms of the TAR conformation and ARG occupancy. The dominant bound TAR conformation features two bound ARG ligands and has an equilibrium population in the absence of ARG that is below detection limit. Consequently, even though ARG binds to TAR with an apparent overall weak affinity $\left(K_{\mathrm{d}}^{\mathrm{app}} \approx 0.2\mathrm{mM}\right)$, it binds the prefolded conformation with a K_{d} in the nM range. Our results show that conformational penalties can be major determinants of RNA-ligand binding affinity as well as a source of binding cooperativity, with important implications for a predictive understanding of how RNA is recognized and for RNA-targeted drug discovery.
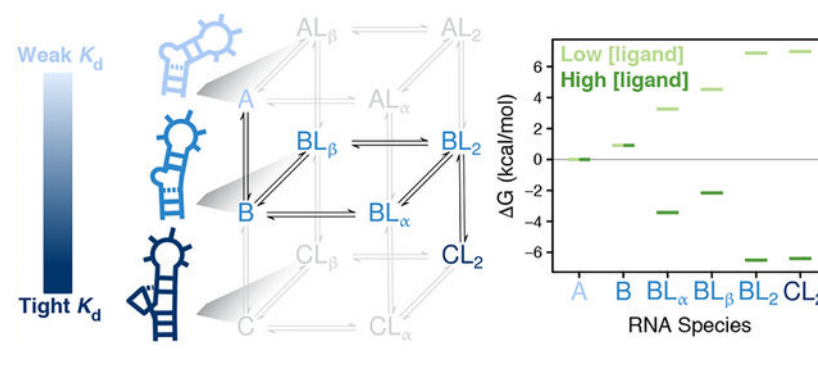
Graphical Abstract

[*]**Corresponding Authors:** hashim.al.hashimi@duke.edu, oas@duke.edu.

INTRODUCTION

The ability of non-coding RNAs to adopt a multitude of conformations underlies their cellular roles in gene expression and regulation.[1–6] These dynamic transitions occur on time scales ranging from picoseconds to tens of seconds and can feature global changes in the orientation of helical elements and local base pair reshuffling that remodel both secondary and tertiary structure.[5,7] Although such alternative conformational states can be short-lived and/or poorly populated in the isolated RNA, they can become long-lived and appreciably populated to play regulatory roles through interactions with binding partners.[8] The high prevalence of conformational changes in RNA on molecular recognition[9–11] implies that binding energies are highly dependent on RNA conformational penalties.[12,13] However, the relative magnitudes of these energetic penalties are difficult to experimentally quantify due to challenges in detecting extremely low population species in the absence of ligand. Detection and characterization of such transient species is further complicated by highly cooperative interactions that can deceivingly appear two-state.[14,15] Significant advances in nuclear magnetic resonance (NMR) spectroscopy[16–20] have allowed characterization of protein and more recently RNA conformational dynamics occurring on the microsecond to millisecond time scale by relaxation dispersion (RD) techniques including $R_{1\rho}$,[21–23] CPMG, [24,25] and CEST,[8,26] but challenges remain with respect to detection limitations and complicated data analysis. Developing a predictive understanding regarding the role of RNA conformational thermodynamics is of critical importance[27] for RNA-targeted drug discovery[28–30] and bioengineering.[31–33]

Prior studies have successfully determined complex mechanisms by globally fitting data collected as a function of multiple independent variables in order to reshape the population distribution of relevant conformational states.[34–38] We apply this strategy in our study of trans-activation response element (TAR) RNA from human immunodeficiency virus type-1 (HIV-1), which is one of the first paradigmatic examples of RNA adaptive recognition. Its conformation-dependent binding interaction with the viral protein Tat is required for full-length transcription elongation[39–41] and has therefore been the subject of numerous studies aimed at understanding the principles behind dynamic ligand recognition. TAR is also a drug target for the development of anti-HIV therapeutics and it has been shown that the ligand-bound RNA conformation is an important determinant of the ligand's inhibitory activity. [42–44] Previous NMR-derived structures show that unliganded TAR adopts a mainly bent (average bend angle ~47°) and flexible inter-helical conformation in which bulge residue

U23 stacks on A22, while spacer bulge residues C24 and U25 are looped out and flexible. [45–48] The ligand argininamide (ARG) has been used extensively in TAR binding studies because it recapitulates the functional structural changes that TAR undergoes upon binding to the HIV-1 Tat arginine-rich motif (ARM). [45,49–53] ARG binding arrests inter-helical motions and stabilizes a coaxial conformation in which bulge residue U23 forms a base triple with helical residues A27 and U38, and junctional residues A22 and U40 become stably hydrogen bonded. [49,53]

Several studies measuring TAR binding to Tat-like peptides or ARG-derived compounds have reported evidence for two binding sites, [43,54–57] while others report single apparent affinities, [44,50,58–61] and it remains unclear what role the other Tat arginine residues located in the arginine-rich RNA binding domain (RBD, residues G48-R57) [62] play *in vivo*. Overall, the results in the literature hint at conformational heterogeneity present in the TAR·ARG bound state that has not been fully explored. Here, we show that bulge-dependent binding occurs at two sites on TAR and that the tightest affinity occurs on the base tripled TAR state. The apical loop of the TAR construct used for all experiments in this study is replaced with a UUCG loop to simplify the analysis and ensure that no dispersion is present in the unliganded state. [63] This substitution has been shown to have minimal effects on ARG binding or ARG-dependent conformational equilibria. [53,64] We globally fit NMR chemical shift perturbation (CSP) data to a six-state model and then test the predictions generated from our model using NMR RD in the presence of ligand.

RESULTS AND DISCUSSION

Qualitative Inspection of Chemical Shift Perturbations′ Ligand and Temperature Dependence Reveals Three Conformations and Two Binding Sites.

We carried out NMR 2D HSQC experiments as a function of both ligand concentration and temperature to quantitatively dissect the coupled conformational change and binding reactions that occur when the small-molecule ligand, ARG (Figure 1B), binds to TAR RNA (Figure 1A). Chemical shifts are exceptionally sensitive to local electronic environment, making them excellent reporters of conformational change and proximity to ligands, although distinguishing the two can be challenging. Close inspection of the raw CSP data (Figure 1C) enables us to build basic elements of a binding model on the basis of the following three criteria.

1. Nonlinear trajectories observed in the 2D HSQC spectra [65] in Figure 1C under the fast-exchange limit are indicative of multi-site binding. [66–68] The data is consistent with fast exchange because no significant line broadening is observed at any of the ligand concentrations and temperatures tested (Figures 1C and S2).

2. Temperature-dependent chemical shifts [69–72] at low ligand concentration (see residue U23-C1′ in Figure 1C) are indicative of conformational equilibria between multiple unliganded RNA species.

3. Temperature-dependent chemical shifts at high ligand concentration (see residue G28-C8 in Figure 1C) are indicative of conformational equilibria between multiple ARG-bound RNA species.

In addition to the perturbations addressed by these observations, ARG-dependent perturbations are also observed across a wide variety of residues including helical regions outside the apparent binding sites, raising the immediate question of whether these distant perturbations are due to long-range conformational effects or whether they reflect weak interaction with three or more ARG molecules. Strong temperature but no ligand dependencies are observed in the chemical shifts of resonances in the UUCG loop even though bulge-dependent equilibria are expected to be independent of the apical loop for the TAR sequence used here.[64] This observation suggests that more than two conformational equilibria may contribute to the observed chemical shifts. A model that accounts for all of these additional RNA species would be under-determined without a substantial amount of additional data, and the equations may not have analytical solutions (for ligand binding to two non-identical sites).[73] To circumvent this issue and focus on ligand-dependent CSPs that probe ARG binding, we used a bulgeless TAR variant to help identify the resonances reporting only on bulge-dependent ARG binding and conformational change.

Bulge Deletion Identifies Chemical Shifts Not Perturbed by Two Primary Binding Sites.

Quantitative interpretation of the observed CSP temperature and ligand dependencies depends on the assumption that the underlying data reports on the explicit equilibria included in the binding model. Ideally, the binding model would include all equilibria but due to experimental and computational limitations, we are forced to build minimal complexity models. Because we have chosen to model the bulge-dependent binding and conformational equilibria presumed to be critical for functionally important TAR-Tat interactions in the cell,[39,40,74,75] it is important for accurate modeling that we exclude CSPs from global fitting that are bulge-independent. To identify these resonances, we carried out NMR CSP mapping experiments on a construct lacking the UCU bulge (bulgeTAR) as a function of temperature and ARG concentration. Specifically, comparison of the NMR spectra between the two constructs both in the absence (Figure S3A) and presence (Figure S3B) of ARG reveals that ARG-induced perturbations in the lower helix do not depend on the bulge. Importantly, when we compare the NMR spectra of TAR and bulgeTAR in the presence of ARG, any peaks in the ARG-bound spectra that do not overlay between TAR and bulgeTAR identify a specific bulge-dependent binding site or a binding-dependent conformational change that can only occur in the construct with the trinucleotide bulge. The aromatic resonances for A20 and G21 are nearly identical for both RNA variants under ARG-saturating conditions (Figure S3B), which indicates that the lower helix does not report on bulge-dependent binding. As expected, resonances near the bulge that are known to play a role in ARG recognition (A22, G26, A27, U38, C39, U40)[49] have large chemical shift differences ($\delta = \delta_{\text{bulge-ARG}} - \delta_{\text{TAR-ARG}}$) in the presence of ARG. More interesting, however, are resonances located farther away from the bulge that do not overlay well in the ARG-bound spectra, namely G28-C8 and C29-C6. This result provides support for a second bulge-dependent binding site in the upper helix above the bulge, (Figure 2B and Figure S4) consistent with a recently published NMR study that showed the TAR·Tat RBD complex was stabilized by two arginine residues: R52 below the bulge and R49 above the bulge.[76] Additionally, these results suggest that using a simplified small-molecule system to study RNA–protein binding can faithfully capture key features of the larger system and is an effective approach for dissecting complex binding equilibria. For the purpose of accurately

modeling the TAR bulge-dependent binding equilibria, we exclude resonances from global fitting whose chemical shifts are sensitive to bulge-independent binding or to melting as discussed further in the Supporting Information.

Global Bayesian Fitting Provides Posterior Probability Distributions of Thermodynamic and Structural Parameters.

Global fitting of the TAR CSPs caused by changes in ligand concentration and temperature to a thermodynamic model can provide estimates of reaction energies and chemical shift fingerprints (i.e., structural information) of the various populated RNA species. Each data point collected at a given temperature and ARG concentration for a given resonance is modeled as the observed chemical shift position δ_{obs}. For systems exchanging between n states on the fast NMR time scale, the δ_{obs} at a given ligand concentration and temperature is an ensemble-averaged value given by

$$\delta_{obs} = \sum_{i=1}^{n} p_i \delta_i \tag{1}$$

for states $i = 1, \ldots, n$. The population of a given species p_i, expressed as

$$p_i = \frac{e^{-\Delta G_i/RT}}{Q} = \frac{e^{-\Delta G_i/RT}}{\sum_{i=1}^{n} e^{-\Delta G_i/RT}} \tag{2}$$

is the probability of the system being in state i, where $\sum_{i=1}^{n} p_i = 1$, R is the ideal gas constant, T is the absolute temperature in Kelvin, and G_i is the Gibbs free energy difference between state i and a reference state. The partition function $Q = \sum_{i=1}^{n} e^{-\Delta G_i/RT}$ represents the sum of the statistical weights of all microstates distinguishable by binding site occupancy and cooperatively formed conformational minima (see Methods for model-dependent Q derivations).

Because the raw data establish a mechanism that involves multiple RNA conformational states and multiple liganded states, we employed a Bayesian approach that has several advantages over least-squares fitting methods, particularly for estimation of correlated parameters in multi-component biophysical models.[77–80] Least-squares methods that obtain point estimates for parameter values rely on the assumption that the error in the data is normally distributed around the mean value, but this assumption need not be true for more complicated systems in which the variance in the data has complex sources (Methods) and parameters are often correlated. A Bayesian approach, by contrast, estimates complete joint probability distributions for each parameter and in so doing estimates the associated uncertainty over all potential parameter space, providing more reliable estimates of parameter uncertainty. Additionally, Bayesian inference has the advantage over standard approaches in its ability to formally incorporate prior knowledge with the observed data, which together are used to maximize the log likelihood of the proposed model, or the conditional probability of observing the data given the model. This approach is represented by Bayes's rule, $p(\theta|y) \propto p(y|\theta) \, p(\theta)$, where $p(\theta|y)$ is the posterior probability distribution,

$p(y|\theta)$ is the likelihood, and $p(\theta)$ is the prior probability distribution for a specified model with parameters θ and data y. Our implementation of Bayesian fitting is described in Methods.

A Six-State Thermodynamic Model Matches the Information Content of the CSP Data.

The model we developed for data fitting assumes explicit conformational changes and binding equilibria to quantitatively interpret the observed CSP temperature and ligand dependencies. As noted earlier, the curvature observed in the ligand-dependent 2D CSP trajectories provides evidence for at least two binding sites, and the temperature-dependent CSPs at both low and high ligand concentrations indicate conformational equilibria between at least three species. Taken together, these observations suggest the existence of a minimum of 12 distinct RNA species: three conformational states (referred to herein as states A, B, and C) and three bound forms of each conformational state, corresponding to ARG bound to site α or β, or both. This scheme is depicted in Figure 1D.

Ideally, there would be sufficient information in the experimental data to determine all of the equilibrium constants. Unfortunately, because some states are insufficiently populated to allow estimation of their contributions to observed CSPs, we are forced to reduce the complexity of the model to six RNA species: only the unliganded form of the A conformation (A), only the fully liganded form of the C conformation (CL_2), and all possible liganded states of the B conformation (B, BL_α, BL_β, and BL_2). Our rationale for including these states and not the other six is described in detail in the Supporting Information. Although our model-fitting approach does not strictly require any prior assumptions regarding the structural identities of the A, B and C conformations, we propose that they correspond to the flexible, bent inter-helical form of TAR[10,46,47,81] (state A); the coaxially stacked state observed at high Mg^{2+} concentration[82–85] where junctional residues A22 and U40 below the bulge remain unpaired (state B); and the coaxially stacked state stabilized with a A22·U40 base pair and U23-A27·U38 base triple[49] (state C), as depicted in Figure 1E. These predictions are based on earlier NMR studies that report hydrogen bond alignments between A22-U40 and U23-A27 in excess ARG but not in Mg^{2+},[53] where both ARG and Mg^{2+} similarly reduce the TAR helix bend angle relative to the unliganded state according to transient electric birefringence, NMR, and circular dichroism studies.[47,59,84,86]

At intermediate ARG concentrations there must be significantly populated singly bound species, which must have either the B or C conformations, or both. As noted above, the minimal model must include one or more singly bound species in the same conformational state as the doubly bound species. The non-sigmoidal binding curves observed for several resonances would not be observed if the two binding sites were on different conformational states. Given the assumption that the unliganded C state is negligible, the B conformation is the best candidate for the singly bound species, i.e., BL_α and/or BL_β as opposed to CL_α and/or CL_β. We fit for $K_a^{B\alpha}$ and $K_a^{B\beta}$ distributions but also consider the possibility that CL_α and/or CL_β become sufficiently populated to contribute to the observed CSPs and therefore also test fits including $K_a^{C\alpha}$ and/or $K_a^{C\beta}$. We conclude that the models including CL_α and/or CL_β do not fit the data significantly better than the simpler and less parametrized model with

only BL_α and BL_β. This conclusion is consistent with our preference for a model whose complexity is matched to the information content of the data. Additional experiments beyond the scope of the present study might provide sufficient information to determine the populations of some of the six species not included in our minimal model. An example of such an experiment would be relaxation kinetics, which have the capacity to detect transiently populated species whose equilibrium populations are undetectable.[37]

The Unliganded C Conformation of TAR Is Poorly Populated But Has High Affinity for ARG.

For the six-state model, the fitting equation is given by

$$\delta_{obs} = \sum_{i=1}^{n=6} p_i \delta_i = \delta_A p_A + \delta_B p_B + \delta_{BL_\alpha} p_{BL_\alpha} + \delta_{BL_\beta} p_{BL_\beta} + \delta_{BL_2} p_{BL_2} + \delta_{CL_2} p_{CL_2} \quad (3)$$

where the population of each species is found as described in Methods. The populations and therefore the observed chemical shifts are a function of the site binding affinities for $BL_\alpha \left(K_a^{B\alpha} = 1/K_d^{B\alpha} \right)$ and $BL_\beta \left(K_a^{B\beta} = 1/K_d^{B\beta} \right)$ and equilibrium constants for $A \rightleftharpoons B$ (K_{AB}) and $BL_2 \rightleftharpoons CL_2$ ($K_{BL_2CL_2}$). Because the slopes of chemical shifts versus ligand concentration do not vary significantly with temperature for all resonances, (representatives shown in inset in Figure 1C), we assume that the site binding constants are insensitive to temperature at the experimental conditions tested and the only detectable temperature dependency arises from conformational equilibria given by

$$K_{eq} = e^{\Delta H/R(1/T_m - 1/T)} \quad (4)$$

where ΔH is the enthalpy change and T_m is the melting temperature in Kelvin for a given conformational transition. Here, the parameter vector θ is given by $\theta = \left(K_a^{B\alpha}, K_a^{B\beta}, \Delta H_{AB}, \Delta H_{BL_2CL_2}, T_m^{AB}, T_m^{BL2CL2}, \delta_{r,i}, \epsilon_j \right)$ where $\delta_{r,i}$ represents the vector of basis chemical shifts for species i (i = A, B, BL_α, BL_β, BL_2, and CL_2) per resonance r, and ϵ_j is the measurement error in δ_{obs} in units of ppm for chemical shift groups j (j = C6/C8, C1′, C2, H2/H6/H8/H1′, H1/H3, and overlapped resonances). The error terms were grouped according to scale of raw chemical shift values, as described in Methods.

As listed in Table 1, the site binding affinities estimated from the marginal posterior distributions are 66 (+7/−6) μM and 550 (+50/−40) μM for the α and β sites on state B, respectively. The enthalpy change of the stacking equilibrium (A $\rightleftharpoons$ B) is estimated to be −7.9 (+0.5/−0.4) kcal mol^{-1}, while the derived entropy change for the same reaction is quite unfavorable at −29.4 (+0.4/−1.2) cal mol^{-1} K^{-1}. This reduction in entropy is expected for a stacking reaction. The $BL_2 \rightleftharpoons CL_2$ equilibrium is slightly less exothermic at −2.9 (+0.4/−0.4) kcal mol^{-1} and also less entropically unfavorable at 9.9 (+1.6/−1.3) cal mol^{-1} K^{-1}. The derived free energy change for both reactions is slightly positive (unfavorable) at 25 °C (Figure 3C, red), but as depicted in Figure 4, both A and B are significantly populated in the absence of ligand and both BL_2 and CL_2 are significantly populated at high ARG concentrations. Figure 4 also shows that the only other significantly populated species is

BL$_a$, whose population reaches ~35% at the midpoint of the apparent binding curve, but is poorly populated at low and high ARG concentrations.

In addition to the six-state model, we tested several models (Figure S5) as potential fits to the data. Although fits to models with species $n \geq 7$ were not statistically superior to the simpler, less parametrized six-state model, they provide additional support for the existence of tight binding to an undetectable C species. For instance, the K_d to at least one site on the C conformation was consistently estimated to be below 0.009 mM for fits to $n \geq 7$-state models. Under such conditions, the population of the singly bound C conformation would be below 5%, so its contribution to δ_{obs} would be negligible and largely indeterminable, which is consistent with the highly uncertain distributions obtained for δ_{CL_a} and/or δ_{CL_β} in those fits. Overall, these results are consistent with a model in which binding to at least one of the sites on the C conformation is much tighter than binding to that site in the B conformation.

Fitted CSPs Give Insight into Structures of Intermediate Species.

To answer the question of whether the A $\rightleftharpoons$ B equilibrium reports on coaxial stacking, we compared the fitted chemical shifts for the B state with measured chemical shifts for TAR in Mg^{2+} and additionally for bulgeTAR in ionic strength conditions comparable to those in the CSP titration (Figure 5A). It is known that Mg^{2+} shifts the equilibrium to >90% stacked, [84,85,87] so we would expect the chemical shifts of the fitted B state to match those seen in Mg^{2+} if temperature perturbs the same stacking equilibrium. However, a complication with this previously used assumption is that Mg^{2+} also makes direct contacts with resonances near the bulge, so it may be difficult to separate the chemical shift contributions due to stacking/unstacking versus those due to localization of Mg^{2+}. The bulgeless variant, however, provides a cleaner fingerprint of the chemical shift for the stacked state since it is presumably never unstacked. The disadvantage of this comparison is that it provides no information on the stacked chemical shift for bulge residues U23, C24, and U25 because they are absent in the variant. As shown in Figure 5A, many of the B-state shifts agree in direction and approximate magnitude with shifts for bulgeTAR, supporting the hypothesis that the A $\rightleftharpoons$ B equilibrium is the unstacked $\rightleftharpoons$ stacked equilibrium. Interestingly but perhaps not surprisingly, the chemical shifts of resonances thought to be in close contact with Mg^{2+} ions (C24-C1′, C24-C6, U25-C1′, U25-C6, G26-C8, A27-C8) are in opposite direction to the chemical shifts for TAR·Mg^{2+}. Relative to the A state, C39-C6 shifts upfield and C39-C1′ shifts downfield in B, consistent with the base becoming more stacked and the sugar pucker becoming more A-form, as would be expected for residues directly above the bulge in the coaxially stacked B state.

Comparing the chemical shifts for CL$_2$ relative to BL$_2$ gives insight into the structural differences between the stacked/base triple state and the stacked state. Relative to the BL$_2$ state, resonances in CL$_2$ that lose stacking interactions on the basis of downfield chemical shifts include U23-C6, G26-C8, G28-C8/H8, C29-H6, while resonances that become more stacked in the base-tripled state include A22-H2, C24-C6, U25-C6, and C29-C6 (Figure 5B). U25-C1′ has a large upfield chemical shift in CL$_2$ indicating a shift from C3′-endo to C2′-endo (deviating away from RNA A-form geometry). C39-C6 remains unchanged between the BL$_2$ and CL$_2$ states, while its sugar (C39-C1′) becomes slightly more helical

(downfield chemical shift), although it is already adopting a more A-form geometry in BL_2 relative to unliganded A and B. The upfield shift in H2 of A22 is consistent with increased stacking and stabilization of the A22-U40 base pair in CL_2.

Testing the Model and Characterizing a Binding Intermediate by NMR RD.

Our model predicts detectable populations of the intermediate species BL_a at intermediate ligand concentrations, so we can exploit this information to gain insight into the kinetics of ligand binding using NMR $R_{1\rho}$ RD. Additionally, $R_{1\rho}$ experiments allow independent determination of chemical shifts for comparison to those obtained from analysis of the CSPs. Measurements were first carried out in the absence of ARG. Consistent with previous results,[63] no dispersion is observed in the free TAR construct at temperatures between 5 and 25 °C. Because our model predicts populations of roughly 30% and 15% for the unliganded B state at 5 and 25 °C, respectively (Figure 6A), and sufficiently large ^{13}C chemical shifts ($\Delta\omega = \omega_B - \omega_A > 1$ ppm) for several resonances (Figure 5A), the absence of dispersion indicates that the A $\rightleftharpoons$ B transition must be fast ($\tau = 1/k_{ex} \ll 1 \mu s$) and outside the detection limits. Our CSP experiments revealed that binding occurs on the fast exchange NMR time scale ($k_{ex} (s^{-1}) > \Delta\omega (rad\ s^{-1})$)[68] at all tested temperatures and therefore we can predict that for average chemical shift differences of 2 ppm in ^{13}C on a 700 MHz spectrometer, the exchange process ($k_{ex} = k_1 + k_{-1}$) will be faster than ~2000 s^{-1}. A simplified description of ligand binding between one ligand and one receptor with a single site yields $k_1 = k_{on}[L]$ and $k_{-1} = k_{off}$. Thus, the total observed exchange ($k_{ex} = k_{on}[L] + k_{off}$) scales with increasing free ligand concentration [L]. Experiments were carried out at low ligand concentration to avoid reaching exchange rates too fast to be detected by RD.

To test the predictions from the CSP analysis, we carried out off-resonance $R_{1\rho}$ measurements at 1.5 mM RNA and 0.2 mM ARG which should give rise to chemical exchange corresponding to 9% of the BL_a species and 1% of the BL_β species for resonances with large enough ^{13}C chemical shifts relative to the ground state ($\Delta\omega = \omega_{BLa} - \omega_{GS} > 1$ ppm and $\Delta\omega = \omega_{BL\beta} - \omega_{GS} > 1$ ppm). We note here that because the observed ground state (GS) is in fast exchange between A $\rightleftharpoons$ B as described above, the GS chemical shift is not simply ω_A; rather it is given by $\omega_{GS} = \omega_A(1 - p_B) + \omega_B p_B$, where p_B is calculated from the Bayesian-fitted parameters according to

$$p_B = \frac{K_{AB}}{1 + K_{AB}} = \frac{e^{\Delta H_{AB}/R\left(1/T_m^{AB} - 1/T\right)}}{1 + e^{\Delta H_{AB}/R\left(1/T_m^{AB} - 1/T\right)}} \tag{5}$$

At 5 °C, we observe apparent two-state dispersion profiles at four resonances located near the primary bulge binding site (Figure 6B). No ligand-dependent dispersion was observed at 25 °C, likely because the reaction is too fast for detection by $R_{1\rho}$, as reaction rates scale with increasing temperature. Global fitting of the data at 5 °C revealed the apparent on-rate constant to be 3×10^6 M^{-1} s^{-1} and the apparent off-rate for binding to be about 8000 s^{-1}. The k_{on}^{app} and k_{off}^{app} values were calculated using eq 19 and eq 20 as described in Methods. The population obtained by RD ($5.4 \pm 0.5\%$) is in fairly close agreement with the predicted 9%, although it is important to note that there is no RD evidence for the 1% predicted

exchange with a third species. This could be explained by the chemical shifts for the BL_β state being too small (<1 ppm) for RD detection, as we do not have accurate estimations for the BL_β chemical shifts from the global fitting of the CSP data (Supporting Information). Nevertheless, the chemical shifts obtained by RD are in strikingly good agreement with the chemical shifts for species BL_α determined from the CSP fitting (Figure 6C), further supporting that the RD is reporting on the binding to the α site on the B conformation.

Furthermore, the apparent dissociation constant K_d^{app} can be calculated from the RD fitted parameters using $K_d^{app} = k_{off}^{app}/k_{on}^{app}$ for a two-state system. In this case, because of the fast A $\rightleftharpoons$ B pre-equilibrium, the K_d^{app} is scaled by the population of the B state, so $K_d^{app} = K_{AB}k_{off}^{app}/((1 + K_{AB})k_{on}^{app})$, where K_{AB} (at 5 °C) is obtained from the CSP results. The result for K_d^{app} obtained by RD (0.22 ± 0.05 mM) is in fairly good agreement with the 0.1 mM measured for $K_d^{B\alpha}$ by the global CSP analysis. Measure ments were also carried out at high ligand concentration (10 mM ARG) to probe the kinetics for base triple formation (BL_2 $\rightleftharpoons CL_2$) since our model predicts a sufficiently high population of $CL_2(p_{CL_2} \approx 40 - 60\%)$, but no dispersion was observed at all tested temperatures (Figure S11). It is possible that base triple formation is too fast and outside the detection limits of RD, and more experiments in the future will be needed to explore this hypothesis further.

CONCLUSION

Returning to the complete 12-state model introduced earlier, we can begin to make predictions about this more detailed mechanism based on the analysis of our simpler six-state model. As a consequence of thermodynamic linkage and free energy conservation, not all equilibrium constants in the full 12-state model are independent, i.e.,

$$K_a^{C\alpha}K_a^{C\beta} = \frac{K_a^{B\alpha}K_a^{B\beta}\beta_{BL_2CL_2}}{K_{BC}} \tag{6}$$

Our fitted minimal model provides estimates at 25 °C for $K_a^{B\alpha}$, $K_a^{B\beta}$, and $\beta_{BL_2CL_2}$ (1.4×10^4 M^{-1}, 1.8×10^3 M^{-1} and 0.84, respectively). Our inability to detect any form of C except CL_2 gives an upper limit for K_{BC} of 0.05, and substitution of these estimates into eq 6 gives an estimate for the product of the two C-state binding constants of at least 3×10^{12} M^{-2}. This estimate implies that the upper limit for both binding dissociation constants $K_d^{C\alpha}$ and $K_d^{C\beta}$ must be 0.6 μM, assuming that they are equal. If they are not the same, then one of the C sites must be even tighter (Figure S12).

Because this affinity is much tighter than that of the B conformation ($K_d^{C\alpha}K_d^{C\beta} < K_d^{B\alpha}K_d^{B\beta}$ by 4 orders of magnitude), the probability of finding a doubly bound C species (CL_2) is much higher than finding a singly bound C species (CL_α or CL_β). In other words, it is much more

thermodynamically favorable for ARG to bind to both sites on conformation C than it is to have only one ligand bound. The same is not true for the B conformation, whose ligand binding is sufficiently weak that the most energetically favorable species is the singly bound B state (BL_a) at intermediate ARG concentrations (see population of BL_a in Figure 4).

Considering our results in the context of energetic penalties previously measured for RNA systems, the enthalpy we measure for A $\rightleftharpoons$ B ($H_{AB} \approx -7.9$ kcal mol^{-1}) is in fairly close agreement with the enthalpy reported for formation of a single terminal AU-AU base stack ($H_{stack} = -6.82 \pm 0.72$ kcal mol^{-1}).[88] Since formation of the TAR B state involves coaxial stacking between the upper and lower helix, our data is consistent with the conclusion that the H measured for A $\rightleftharpoons$ B reflects the enthalpy of stacking. This occurs between junctional bases at the bulge that are likely to be as flexible as the terminal bases studied previously.[88] Additionally, consistent with our results of ~ 7 kcal mol^{-1} penalty for forming the C conformation in the absence of ligand (Figure 7), no evidence has ever been observed suggesting the base triple is detectable in the free RNA, and an RDC-informed MD ensemble also never observed the base triple for either HIV-1 or HIV-2 TAR.[89,90]

Our results suggest that preferentially tight affinity to high-energy RNA states may be a ubiquitous mechanism by which ligands confer selectivity to otherwise promiscuous RNA binding partners. The binding energy from binding to a tight-affinity ligand can explain how the RNA overcomes the high conformational energetic penalty for formation of the binding-competent conformation. Lack of selectivity is among the main challenges in RNA drug design and in understanding processes regulated by RNA. Our study provides direct evidence that undetectable RNA species are the key thermodynamic players in a paradigmatic example of RNA adaptive recognition and underscores the importance of developing methods to study low-populated states that have gone undetected. While the presented approach only provides thermodynamic information regarding binding intermediates, its integration with other methods such as NMR RD and stopped-flow techniques may afford a deeper and more complete kinetic characterization of the binding reaction. In addition, extending NMR measurements to other sugar resonances should be straightforward, and this will afford the opportunity to structurally characterize the intermediates to higher resolution.[91,92] We predict that the mechanism uncovered here is not limited to TAR and anticipate that further experiments using the approach outlined here can help to quantitatively characterize multi-state RNA–ligand interactions.

METHODS

Sample Preparation.

HIV-1 TAR RNA capped with a UUCG loop (sequence GGCAGAUCUGAGCUUCGGCUCUCUGCC) and bulgeless mutant (sequence GGCAGAGAGCUUCGGCUCUCUGCC) were prepared by in vitro transcription using DNA templates containing the T7 promoter (Integrated DNA Technologies). DNA templates were annealed in 3 mM MgCl$_2$ by heating to 95 °C for 5 min and cooling on ice for 30 min. The transcription reaction was carried out at 37 °C for 12 h with T7 RNA polymerase (New England BioLabs) using ^{13}C/^{15}N-labeled nucleotide triphosphates (Cambridge Isotope Laboratories, Inc.). RNA was purified using 20% (w/v) denaturing polyacrylamide gel

electrophoresis (PAGE) with 8 M urea and 1× TBE buffer followed by excision from the gel by electroelution in 1× TAE buffer and further purified by ethanol precipitation overnight at −20 °C. Purified RNA was dissolved in water to 50 μM RNA and annealed by heating to 95 °C for 5 min and cooling on ice for 1 h. Prior to NMR experiments, ^{13}C/^{15}N-labeled RNA samples were exchanged into phosphate NMR buffer (15 mM NaH$_2$PO$_4$/Na$_2$HPO$_4$, 25 mM NaCl, 0.1 mM EDTA, 10% (v/v) D$_2$O at pH 6.4) to the desired RNA concentrations (0.2 mM for HSQC titrations and 1.5 mM for RD). Concentrations were measured using a nanodrop UV spectrometer by the absorbance reading of heat-denatured RNA at 260 mm.

NMR CSP Experiments.

NMR experiments were carried out on Bruker Avance III 600 MHz and Bruker Avance III 700 MHz spectrometers equipped with a 5 mm triple resonance cryogenic probe. Assignments for TAR in the unbound and ARG-bound states were previously published.[93] CSPs as a function of temperature and ligand concentration were monitored for aromatic and aliphatic resonances using 2D [^{13}C,^{1}H] heteronuclear single quantum coherence (HSQC) spectra and for imino resonances using 1D [^{1}H] SOFAST spectra. For each titration point, successive additions of ARG (L-argininamide dihydrochloride, Sigma) from a concentrated stock dissolved in phosphate NMR buffer were added directly to an NMR sample tube containing 450 μL of 0.2 mM ^{13}C/^{15}N-labeled RNA, and the tube was inverted at least six times to ensure homogeneous mixing. The increase in total sample volume due to ARG addition remained less than 5% of the starting volume and therefore had a negligible effect on the RNA concentration. A total of seven titration points per temperature were used for the TAR titration and six for the bulgeless mutant titration. Experiments for both constructs were carried out at six different temperatures (1, 2.5, 5, 10, 15, and 25 °C). The temperature of the NMR probe was calibrated using a methanol sample and spectra was internally referenced relative to a trace amount of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). Spectra were processed using NMRPipe,[94] and peak positions were measured using Sparky. [95]

NMR RD Experiments.

^{13}C $R_{1\rho}$ measurements were carried out at 25 and 5 °C on a Bruker Avance III 700 MHz spectrometer using 1.5 mM ^{13}C/^{15}N-labeled RNA. On- and off-resonance measurements were collected at the spin lock powers (Ω) and spin lock offsets (ω_1) listed in Table S1 using a 1D acquisition scheme which uses Hartmann–Hahn cross-polarization transfer to selectively excite a single peak at a time as described previously.[22,63,92] To account for sample heating effects due to high-power spin locks,[96] the pulse sequence contains a heat-compensation element that applies far off-resonance ^{13}C spin locks during the recycle delay after acquisition at the highest field strength for a time $T_{relax}^{max} - T_{relax}$ such that the net amount of radiofrequency (RF) introduced into the sample is constant across all relaxation delay times.[21,64] For preparation of samples, ARG from a concentrated stock dissolved in phosphate NMR buffer was added directly to the NMR sample tube containing 1.5 mM ^{13}C/^{15}N-labeled RNA following the procedure used for titration experiments to achieve the desired ARG concentrations (0.2 and 10 mM). Raw data was processed using NMRPipe to

determine peak intensities at each delay time point, which were then fit to a mono-exponential function to estimate the $R_{1\rho}$ values needed for further analysis via eq 18.

Statistical Thermodynamic Framework.

As described in the Results and Discussion section, qualitative analysis of the CSP data led us to conclude that the minimum complexity system necessary to explain the temperature and ligand dependencies must include three conformational states (A, B, C) each potentially capable of binding two ligands. To further minimize complexity, the two binding sites (α and β) are assumed to be independent; that is, occupancy at one site does not detectably influence the intrinsic affinity at the other site. The partition function Q for such a system can be derived by rewriting the following microscopic association and conformational equilibrium constants:

$$K_a^{A\alpha} = \frac{[AL_\alpha]}{[A][L]} = \frac{[AL_2]}{[AL_\beta][L]} \quad K_a^{A\beta} = \frac{[AL_\beta]}{[A][L]} = \frac{[AL_2]}{[AL_\alpha][L]}$$

$$K_a^{B\alpha} = \frac{[BL_\alpha]}{[B][L]} = \frac{[BL_2]}{[BL_\beta][L]} \quad K_a^{B\beta} = \frac{[BL_\beta]}{[B][L]} = \frac{[BL_2]}{[BL_\alpha][L]}$$

$$K_a^{C\alpha} = \frac{[CL_\alpha]}{[C][L]} = \frac{[CL_2]}{[CL_\beta][L]} \quad K_a^{C\beta} = \frac{[CL_\beta]}{[C][L]} = \frac{[CL_2]}{[CL_\alpha][L]}$$

$$K_{AB} = \frac{[B]}{[A]} \qquad\qquad K_{BC} = \frac{[C]}{[B]}$$

(7)

in terms of RNA concentrations, summing together all 12 RNA forms,

$$[R_T] = [A] + [AL_\alpha] + [AL_\beta] + [AL_2] + [C] + [CL_\alpha] + [CL_\beta] + [CL_2] + [B] + [BL_\alpha] + [BL_\beta] + [BL_2]$$

(8)

and dividing by a chosen reference state (most simply set as the free A state) to obtain

$$Q = 1 + K_a^{A\alpha}[L] + K_a^{A\beta}[L] + K_a^{A\alpha}K_a^{A\beta}[L]^2 + K_{AB} + K_{AB}K_a^{B\alpha}[L] + K_{AB}K_a^{B\beta}[L]$$
$$+ K_{AB}K_a^{B\alpha}K_a^{B\beta}[L]^2 + K_{AB}K_{BC} + K_{AB}K_{BC}K_a^{C\alpha}[L] + K_{AB}K_{BC}K_a^{C\beta}[L]$$
$$+ K_{AB}K_{BC}K_a^{C\alpha}K_a^{C\beta}[L]^2$$

(9)

Although this 12-state model (Figure 1D) was used for simulations, extensive testing revealed that the information content of the data is insufficient to fit more parameters than those necessary to quantify the interconversion of the six states highlighted in Figure 1D. The summation of all RNA forms in the six-state model is

$$[R_T] = [A] + [B] + [BL_\alpha] + [BL_\beta] + [BL_2] + [CL_2]$$

(10)

which gives the partition function

$$Q = 1 + K_{AB} + K_{AB}K_a^{B\alpha}[L] + K_{AB}K_a^{B\beta}[L] + K_{AB}K_a^{B\alpha}K_a^{B\beta}[L]^2$$
$$+ K_{AB}K_{BL_2CL_2}K_a^{B\alpha}K_a^{B\beta}[L]^2$$

(11)

The population term for each species is found by taking its statistical weight and dividing by the partition function. Note that the equilibrium constants K_{BC} and $K_{BL_2CL_2}$ are related through

$$K_{BC} = \frac{K_a^{B\alpha} K_a^{B\beta} K_{BL_2CL_2}}{K_a^{C\alpha} K_a^{C\beta}} \tag{12}$$

The free ligand concentration term [L] is a solution to a cubic equation of the form

$$a[L]^3 + b[L]^2 + c[L] + d = 0 \tag{13}$$

where the coefficients a, b, c, and d are defined as

$$
\begin{aligned}
a = & -K_a^{A\alpha} K_a^{A\beta} - K_{AB}\left(K_a^{B\alpha} K_a^{B\beta} + K_a^{C\alpha} K_a^{C\beta} K_{BC}\right)\\
b = & K_a^{A\alpha}\left(K_a^{A\beta}([L_T] - 2[R_T]) - 1\right) - K_a^{A\beta}\\
& + K_{AB}\left(K_a^{B\alpha}\left(K_a^{B\beta}([L_T] - 2[R_T]) - 1\right) - K_a^{B\beta} + K_a^{C\alpha} K_{BC}\left(K_a^{C\beta}([L_T] - 2[R_T]) - 1\right)\right.\\
& \left. - K_a^{C\beta} K_{BC}\right)\\
c = & K_a^{A\alpha}([L_T] - [R_T]) + K_a^{A\beta}[L_T] - K_a^{A\beta}[R_T]\\
& + K_{AB}\left(K_a^{B\alpha}[L_T] - K_a^{B\alpha}[R_T] + K_a^{B\beta}[L_T] - K_a^{B\beta}[R_T]\right.\\
& \left. + K_{BC}\left(K_a^{C\alpha}[L_T] - K_a^{C\alpha}[R_T] + K_a^{C\beta}[L_T] - K_a^{C\beta}[R_T] - 1\right)\right) - 1\\
d = & [L_T](K_{AB} + K_{AB} K_{BC} + 1)
\end{aligned}
\tag{14}
$$

for the 12-state model and

$$
\begin{aligned}
a = & -K_a^{B\alpha} K_a^{B\beta} K_{AB} - \left(1 + K_{BL_2CL_2}\right)\\
b = & K_{AB}\left(-K_a^{B\beta} + K_a^{B\alpha}\left(-1 + K_a^{B\beta}\left(1 + K_{BL_2CL_2}\right)([L_T] - 2[R_T])\right)\right)\\
c = & -1 + K_{AB}\left(-1 + \left(K_a^{B\alpha} + K_a^{B\beta}\right)([L_T] - [R_T])\right)\\
d = & [L_T](1 + K_{AB})
\end{aligned}
\tag{15}
$$

for the six-state model. The general, model-independent algebraic solution[97,98] for free ligand concentration [L] is given by

$$[L] = \begin{cases} c_0 + s_c \sinh\left(\frac{1}{3}\sinh^{-1}(\chi)\right) & \text{if } p > 0 \\[2mm] c_0 + s_c \cosh\left(\frac{1}{3}\cosh^{-1}(\chi)\right) & \text{if } p < 0 \wedge \chi \geq 1 \\[2mm] c_0 - s_c \cosh\left(\frac{1}{3}\cosh^{-1}(-\chi)\right) & \text{if } p < 0 \wedge \chi \leq 1 \\[2mm] \max\left(\begin{aligned} & c_0 + s_c \cos\left(\frac{1}{3}\cos^{-1}(\chi)\right) \\ & c_0 - s_c \sin\left(\frac{\pi}{6} - \frac{1}{3}\cos^{-1}(\chi)\right) \\ & c_0 - s_c \sin\left(\frac{\pi}{6} + \frac{1}{3}\cos^{-1}(\chi)\right) \end{aligned}\right) & \text{if } p < 0 \wedge |\chi| < 1 \end{cases}$$

(16)

where

$$\begin{aligned} c_0 &= -\frac{b}{3a} \\[2mm] p &= \frac{c}{a} - \frac{b^2}{3a^2} \\[2mm] s_c &= \frac{2\sqrt{|p|}}{\sqrt{3}} \\[2mm] \chi &= \frac{3\sqrt{3}\left(-\frac{2b^3}{27a^3} + \frac{bc}{3a^2} - \frac{d}{a}\right)}{2|p|^{3/2}} \end{aligned}$$

(17)

and a, b, c, and d are model-dependent terms as shown in eqs 14 and 15.

Bayesian Analysis of CSP Data.

Bayesian inference was used to fit the CSP data. This fitting method applies Bayes's rule, $p(\theta|y) = p(y|\theta)\, p(\theta)/p(y)$, which states that for a specified mathematical model of the data y based on parameters θ, the probability distributions of parameter values given the data, $p(\theta|y)$ (called the posterior distributions), are the product of the likelihood of the data given the parameters, $p(y|\theta)$ (a measure of goodness-of-fit), and user-specified prior probability distributions, $p(\theta)$, normalized by the probability of the data, $p(y)$. In simpler terms, the posterior $\propto$ prior $\times$ likelihood. The calculation of the $p(y)$ term needed for calculating the likelihood involves a multivariate integral that does not have an analytical solution except in the simplest of cases such as single-parameter models,[80,99] so Bayesian techniques have been developed that use Markov-chain Monte Carlo (MCMC) rejection sampling to approximate the likelihood function and respective posterior distributions. Here, we use the software package Stan, which employs a version of MCMC sampling known as the no-U-turn (NUTS) Hamiltonian Monte Carlo (HMC) sampler which is efficient and robust for models with complicated posteriors.[100,101] The analysis was carried out using in-house scripts written in the Stan statistical programming language run under the R programming interface (RStan).

The Stan interface allows the user to define a likelihood function given by the model equations and conditioned on the data. For each unknown parameter, a prior distribution is

specified with a corresponding mean μ and standard deviation σ. Briefly, at each sampling iteration, the algorithm takes a number of discretized leapfrog parameter steps to simulate the trajectory of a particle along a potential energy field. The parameter values at a point in this field correspond to the log of the posterior density we wish to maximize. Essentially, the leapfrog steps are a discrete way to approximate the continuous function represented by the model's equations. Highly curved posteriors (i.e., complicated nonlinear functions) are especially difficult to sample from, partly because an infinite number of steps would be required to accurately traverse the trajectory. In the work here, the polynomial terms in the free ligand expression (eq 16) give rise to a highly nonlinear dependence of the model on free ligand concentration. This variable is model-dependent and based on the combination of binding and equilibrium constants (as shown in eqs 7–17), which makes the selection of informative priors very important.

The Bayesian fitting algorithm aims to maximize the log-likelihood for master eq 1 and in doing so treats each model parameter as a random variable with its own probability distribution for parameter vector $\boldsymbol{\theta}$. The general parameter vector for all models tested can be expressed as $\theta = \left(K_d^{X_s}, \Delta H_{Y_1 \rightleftharpoons Y_2}, T_m^{Y_1 \rightleftharpoons Y_2}, \delta_{r,i}, \epsilon_j \right)$, where superscript X refers to a conformational state (A, B, or C), subscript s refers to a binding site (α or β), $Y_1 \rightleftharpoons Y_2$ refers to a given conformational equilibrium (such as A $\rightleftharpoons$ B), $\delta_{r,i}$ represents the vector of basis chemical shifts for species i per resonance r, and ϵ_j is the measurement error for chemical shift categories discussed in the following paragraph.

For all thermodynamic parameters, Gaussian prior distributions were used with standard deviations up to 50% of their means. Gaussian priors were also used for chemical shift parameters with standard deviations of 1.0 ppm for carbon and 0.25 ppm for proton. For a given resonance, the prior distributions used for all species' chemical shifts were identical to avoid biasing the posteriors. In addition to the thermodynamic and chemical shift parameters, the Bayesian fitting algorithm treats the measurement error as a random unknown variable and therefore fits for the an error term referred to as ϵ_j. As mentioned in the Results and Discussion, Bayesian techniques are suitable for handling data whose variance may have several sources. Because our fitting approach uses the raw measured chemical shift values of all included resonance types and avoids scaling or normalizing the data, the scale of the measurement error varies depending on the nucleus type (^{13}C vs ^1H) and even atom position (C6/C8 vs C1$'$). For this reason, we fit for six error terms corresponding to the following categories of chemical shifts: C6/C8, C1$'$, C2, non-exchangeable protons, exchangeable protons, and resonances with significant overlap at any point in the titration. An Inverse-Gamma distribution with shape and scale parameters (α, β) was chosen for the error priors, where the shape and scale terms are related to the mean μ and standard deviation σ by $\alpha = (\mu^2 + 2\sigma^2)/\sigma^2$ and $\beta = \mu(\mu^2 + \sigma^2)/\sigma^2$.

MCMC sampling was carried out using four independent chains and 4000 iterations per chain, the first half of which comprise the initial warm-up phase where the algorithm optimizes for sampling efficiency and are therefore discarded before posterior analysis. The post-warm-up samples used for approximating the parameter posterior distributions consist of 2000 draws per chain, i.e., 8000 total draws. Convergence was assessed (Figure S6) via

the potential scale reduction factor $\hat{R}$ [102] for each parameter where $\hat{R}$ values approaching 1 but 1.05 are ideal and $\hat{R} \geq 1.2$ indicate lack of convergence.[103] Prior to data fitting, the model was extensively tested for robustness by subjecting simulated data to the fitting procedure and evaluating the effects of various prior distributions on posterior distributions. As a benchmark, data with noise were simulated for 20 resonances under experimental conditions identical to that of the collected data (seven ligand concentrations and six temperatures), and the simulated data was subjected to global fitting using the equations described above. The fitting was able to converge within 2000 iterations only when the correct parameter distributions were found. Setting priors very far from simulated values prevented convergence, indicating that the algorithm did not falsely converge to incorrect parameter values.

We note that using a Bayesian approach was not only advantageous to a standard nonlinear least-squares approach but it was integral for quantitative analysis of a model that matched the level of complexity present in the raw data described here. For data sets simulated to the six-state model both with and without noise, least-squares global fitting failed to recover the simulated parameters, whereas Bayesian global fitting correctly recovered the parameter distributions for converged fits.

Analysis of RD Data.

Data for two-state exchange were fit to

$$
\begin{aligned}
R_{1\rho} = {}& R_1\cos^2\theta + R_2\sin^2\theta \\
&+ \frac{\sin^2\theta\, p_{GS} p_{ES} \Delta\omega_{ES}^2 k_{ex}}{\left\{ \frac{\omega_{GS}^2 \omega_{ES}^2}{\omega_{eff}^2} + k_{ex}^2 - \sin^2\theta\, p_{GS} p_{ES} \Delta\omega_{ES}^2 \left(1 + \frac{2k_{ex}^2\left(p_{GS}\omega_{GS}^2 + p_{ES}\omega_{ES}^2\right)}{\omega_{eff}^2 k_{ex}^2}\right) \right\}}
\end{aligned}
\tag{18}
$$

where R_1 and R_2 are the longitudinal and transverse relaxation rates, ω is the change in chemical shift between the low population excited state (ES) and highly populated ground state (GS), expressed as $\omega = \omega_{ES} - \omega_{GS}$. The strength of the carrier spinlock power is ω_{SL} and the tilt angle in the rotating frame is defined as $\theta = \tan^{-1}(\omega_{SL}/\,\Omega)$ given by $\Omega = \Omega - \omega_{rf}$, the difference between average spin lock offset (Ω_{ave}) and reference frequency (ω_{rf}). Average spin lock offset is calculated by $\Omega_{ave} = p_{GS}\omega_{GS} + p_{ES}\omega_{ES}$ where p_{GS} and p_{ES} are populations of the GS and ES, respectively. The spinlock strengths at the GS and ES are given by $\omega_{GS}^2 = (\omega_{GS} + \omega_{rf})^2 + \omega_{SL}^2$ and $\omega_{ES}^2 = (\omega_{ES} + \omega_{rf})^2 + \omega_{SL}^2$, respectively, and their resonance offsets from the spinlock carrier correspond to ω_{GS} and ω_{ES}. $\omega_{ES} = \omega_{ES} - \Omega_{GS}$ defines the difference in offset for transitions from ES to GS. The effective spinlock field strength is calculated as $\omega_{eff}^2 = \Delta\Omega^2 + \omega_{SL}^2$. Finally, the exchange rate constant is expressed as $k_{ex} = k_1 + k_{-1} = p_{ES}k_{ex} + p_{GS}k_{ex}$ and can be converted to the form $k_{ex} = k_{on}^{app}[L] + k_{off}^{app}$ to determine the on and off rate constants. Free ligand concentration [L] is calculated from the relation $[L] = [L_T] - p_{bound}[R_T]$, where the total ligand and RNA concentrations are known a priori, and the population of bound RNA (p_{bound}) is obtained from the RD fitted value for p_{ES} in the case of low ligand concentration where the excited species is the bound state, or

p_{GS} in the case of excess ligand where the excited state is the free RNA. This allows us to rewrite k_{on}^{app} and k_{off}^{app} as

$$k_{on}^{app} = \frac{p_{bound}k_{ex}}{[L_T] - p_{bound}[R_T]} \tag{19}$$

and

$$k_{off}^{app} = k_{ex}(1 - p_{bound}) \tag{20}$$

Note that the rate constants are considered apparent rather than intrinsic due to the presence of RD-invisible $A \underset{[k_{BA}]}{\overset{[k_{AB}]}{\rightleftharpoons}} B$ pre-equilibrium. The apparent association and dissociation rate constants $k_{on}^{app} = k_{on}k_{AB}/(k_{off} + k_{AB})$ and $k_{of}^{app} = k_{off}k_{BA}/(k_{off} + k_{AB})^{104}$ are measured, whereas the intrinsic k_{on} and k_{off} cannot be determined here because k_{AB} and k_{BA} are unknown.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

(1). Mattick JS RNA regulation: a new genetics? Nat. Rev. Genet 2004, 5, 316–23. [PubMed: 15131654]

(2). Sharp PA The centrality of RNA. Cell 2009, 136, 577–80. [PubMed: 19239877]

(3). Cruz JA; Westhof E The dynamic landscapes of RNA architecture. Cell 2009, 136, 604–9. [PubMed: 19239882]

(4). Rinnenthal J; Buck J; Ferner J; Wacker A; Fürtig B; Schwalbe H Mapping the landscape of RNA dynamics with NMR spectroscopy. Acc. Chem. Res 2011, 44, 1292–1301. [PubMed: 21894962]

(5). Dethoff EA; Chugh J; Mustoe AM; Al-Hashimi HM Functional complexity and regulation through RNA dynamics. Nature 2012, 482, 322. [PubMed: 22337051]

(6). Ganser LR; Kelly ML; Herschlag D; Al-Hashimi HM The roles of structural dynamics in the cellular functions of RNAs. Nat. Rev. Mol. Cell Biol 2019, 20, 474–489. [PubMed: 31182864]

(7). Bothe JR; Nikolova EN; Eichhorn CD; Chugh J; Hansen AL; Al-Hashimi HM Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. Nat. Methods 2011, 8, 919–931. [PubMed: 22036746]

(8). Zhao B; Guffy SL; Williams B; Zhang Q An excited state underlies gene regulation of a transcriptional riboswitch. Nat. Chem. Biol 2017, 13, 968–974. [PubMed: 28719589]

(9). De Guzman RN; Turner RB; Summers MF Protein–RNA recognition. Biopolymers 1998, 48, 181–195. [PubMed: 10333745]

(10). Patel DJ Adaptive recognition in RNA complexes with peptides and protein modules. Curr. Opin. Struct. Biol 1999, 9, 74–87. [PubMed: 10047585]

(11). Hermann T; Patel DJ Adaptive recognition by nucleic acid aptamers. Science 2000, 287, 820–5. [PubMed: 10657289]

(12). Varani G RNA–protein intermolecular recognition. Acc. Chem. Res 1997, 30, 189–195.

(13). Becker WR; Jarmoskaite I; Vaidyanathan PP; Greenleaf WJ; Herschlag D Demonstration of protein cooperativity mediated by RNA structure using the human protein PUM2. RNA 2019, 25, 702–712. [PubMed: 30914482]

(14). Solomatin SV; Greenfeld M; Herschlag D Implications of molecular heterogeneity for the cooperativity of biological macromolecules. Nat. Struct. Mol. Biol 2011, 18, 732–4. [PubMed: 21572445]

(15). Motlagh HN; Wrabl JO; Li J; Hilser VJ The ensemble nature of allostery. Nature 2014, 508, 331–9. [PubMed: 24740064]

(16). Palmer AG; Kroenke CD; Loria JP In Nuclear Magnetic Resonance of Biological Macromolecules - Part B; James TL, Dötsch V, Schmitz U, Eds.; Methods in Enzymology 339; Academic Press, 2001; pp 204–238. [PubMed: 11462813]

(17). Korzhnev DM; Skrynnikov NR; Millet O; Torchia DA; Kay LE An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. J. Am. Chem. Soc 2002, 124, 10743–10753. [PubMed: 12207529]

(18). Abergel D; Palmer AG Approximate solutions of the Bloch-McConnell equations chemical equations for two-site chemical exchange. ChemPhysChem 2004, 5, 787–793. [PubMed: 15253305]

(19). Trott O; Palmer AG Theoretical study of $R_1\rho$ rotating-frame and R_1 free-precession relaxation in the presence of n-site chemical exchange. J. Magn. Reson 2004, 170, 104–112. [PubMed: 15324763]

(20). Palmer AG; Massi F Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. Chem. Rev 2006, 106, 1700–1719. [PubMed: 16683750]

(21). Hansen AL; Nikolova EN; Casiano-Negroni A; Al-Hashimi HM Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon $R_1\rho$ NMR spectroscopy. J. Am. Chem. Soc 2009, 131, 3818–9. [PubMed: 19243182]

(22). Xue Y; Kellogg D; Kimsey IJ; Sathyamoorthy B; Stein ZW; McBrairty M; Al-Hashimi HM Characterizing RNA excited states using NMR relaxation dispersion. Methods Enzymol 2015, 558, 39–73. [PubMed: 26068737]

(23). Rangadurai A; Szymaski ES; Kimsey IJ; Shi H; Al-Hashimi HM Characterizing micro-to-millisecond chemical exchange in nucleic acids using off-resonance $R_1\rho$ relaxation dispersion. Prog. Nucl. Magn. Reson. Spectrosc 2019, 112–113, 55–102.

(24). Vallurupalli P; Bouvignies G; Kay LE Increasing the exchange time-scale that can be probed by CPMG relaxation dispersion NMR. J. Phys. Chem. B 2011, 115, 14891–14900. [PubMed: 22077866]

(25). Strebitzer E; Nußbaumer F; Kremser J; Tollinger M; Kreutz C Studying sparsely populated conformational states in RNA combining chemical synthesis and solution NMR spectroscopy. Methods 2018, 148, 39–47. [PubMed: 29753787]

(26). Zhao B; Hansen AL; Zhang Q Characterizing slow chemical exchange in nucleic acids by carbon CEST and low spin-lock field $R_1\rho$ NMR spectroscopy. J. Am. Chem. Soc 2014, 136, 20–23. [PubMed: 24299272]

(27). Herschlag D; Allred BE; Gowrishankar S From static to dynamic: the need for structural ensembles and a predictive model of RNA folding and function. Curr. Opin. Struct. Biol 2015, 30, 125–133. [PubMed: 25744941]

(28). Thomas JR; Hergenrother PJ Targeting RNA with small molecules. Chem. Rev 2008, 108, 1171–1224. [PubMed: 18361529]

(29). Connelly CM; Moon MH; Schneekloth JS The emerging role of RNA as a therapeutic target for small molecules. Cell Chemical Biology 2016, 23, 1077–1090. [PubMed: 27593111]

(30). Ganser LR; Lee J; Rangadurai A; Merriman DK; Kelly ML; Kansal AD; Sathyamoorthy B; Al-Hashimi HM High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble. Nat. Struct. Mol. Biol 2018, 25, 425–434. [PubMed: 29728655]

(31). Liang JC; Bloom RJ; Smolke CD Engineering biological systems with synthetic RNA molecules. Mol. Cell 2011, 43, 915–926. [PubMed: 21925380]

(32). Chang AL; Wolf JJ; Smolke CD Synthetic RNA switches as a tool for temporal and spatial control over gene expression. Curr. Opin. Biotechnol 2012, 23, 679–688. [PubMed: 22305712]

(33). Berens C; Groher F; Suess B RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. Biotechnol. J 2015, 10, 246–257. [PubMed: 25676052]

(34). Henkels CH; Kurz JC; Fierke CA; Oas TG Linked folding and anion binding of the Bacillus subtilis ribonuclease P protein. Biochemistry 2001, 40, 2777–2789. [PubMed: 11258888]

(35). Reining A; Nozinovic S; Schlepckow K; Buhr F; Fürtig B; Schwalbe H Three-state mechanism couples ligand and temperature sensing in riboswitches. Nature 2013, 499, 355–359. [PubMed: 23842498]

(36). Daniels KG; Tonthat NK; McClure DR; Chang Y-C; Liu X; Schumacher MA; Fierke CA; Schmidler SC; Oas TG Ligand concentration regulates the pathways of coupled protein folding and binding. J. Am. Chem. Soc 2014, 136, 822–825. [PubMed: 24364358]

(37). Daniels KG; Suo Y; Oas TG Conformational kinetics reveals affinities of protein conformational states. Proc. Natl. Acad. Sci. U. S. A 2015, 112, 9352–9357. [PubMed: 26162682]

(38). Frankel EA; Strulson CA; Keating CD; Bevilacqua PC Cooperative interactions in the hammerhead ribozyme drive pKa shifting of G12 and its stacked base C17. Biochemistry 2017, 56, 2537–2548. [PubMed: 28485924]

(39). Feng S; Holland EC HIV-1 Tat trans-activation requires the loop sequence within TAR. Nature 1988, 334, 165–7. [PubMed: 3386755]

(40). Garcia JA; Harrich D; Soultanakis E; Wu F; Mitsuyasu R; Gaynor RB Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. EMBO J 1989, 8, 765–78. [PubMed: 2721501]

(41). Marciniak RA; Calnan BJ; Frankel AD; Sharp PA HIV-1 Tat protein trans-activates transcription in vitro. Cell 1990, 63, 791–802. [PubMed: 2225077]

(42). Murchie AIH; Davis B; Isel C; Afshar M; Drysdale MJ; Bower J; Potter AJ; Starkey ID; Swarbrick TM; Mirza S; Prescott CD; Vaglio P; Aboul-ela F; Karn J Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. J. Mol. Biol 2004, 336, 625–638. [PubMed: 15095977]

(43). Davis B; Afshar M; Varani G; Murchie AIH; Karn J; Lentzen G; Drysdale M; Bower J; Potter AJ; Starkey ID; Swarbrick T; Aboul-ela F Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots. J. Mol. Biol 2004, 336, 343–356. [PubMed: 14757049]

(44). Davidson A; Patora-Komisarska K; Robinson JA; Varani G Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of Tat protein. Nucleic Acids Res 2011, 39, 248–256. [PubMed: 20724442]

(45). Aboul-ela F; Karn J; Varani G The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. J. Mol. Biol 1995, 253, 313–332. [PubMed: 7563092]

(46). 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 1996, 24, 3974–3981. [PubMed: 8918800]

(47). Long KS; Crothers DM Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. Biochemistry 1999, 38, 10059–10069. [PubMed: 10433713]

(48). Al-Hashimi HM; Gosser Y; Gorin A; Hu W; Majumdar A; Patel DJ Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. J. Mol. Biol 2002, 315, 95–102. [PubMed: 11779230]

(49). Puglisi JD; Chen L; Frankel AD; Williamson JR Role of RNA structure in arginine recognition of TAR RNA. Proc. Natl. Acad. Sci. U. S. A 1993, 90, 3680–3684. [PubMed: 7682716]

(50). Long KS; Crothers DM Interaction of human immunodeficiency virus type 1 Tat-derived peptides with TAR RNA. Biochemistry 1995, 34, 8885–95. [PubMed: 7612630]

(51). Tao J; Chen L; Frankel AD Dissection of the proposed base triple in human immunodeficiency virus TAR RNA indicates the importance of the Hoogsteen interaction. Biochemistry 1997, 36, 3491–3495. [PubMed: 9131998]

(52). Brodsky A NMR evidence for a base triple in the HIV-2 TAR C-G.C+ mutant- argininamide complex. Nucleic Acids Res 1998, 26, 1991–1995. [PubMed: 9518494]

(53). Pitt SW; Majumdar A; Serganov A; Patel DJ; Al-Hashimi HM Argininamide binding arrests global motions in HIV-1 TAR RNA: comparison with Mg2+-induced conformational stabilization. J. Mol. Biol 2004, 338, 7–16. [PubMed: 15050819]

(54). Ferner J; Suhartono M; Breitung S; Jonker HRA; Hennig M; Wöhnert J; Göbel M; Schwalbe H Structures of HIV TAR RNA-ligand complexes reveal higher binding stoichiometries. ChemBioChem 2009, 10, 1490–1494. [PubMed: 19444830]

(55). Schneeberger EM; Breuker K Native top-down mass spectrometry of TAR RNA in complexes with a wild-type Tat peptide for binding site mapping. Angew. Chem., Int. Ed 2017, 56, 1254–1258.

(56). Kumar S; Maiti S Effect of different arginine methylations on the thermodynamics of Tat peptide binding to HIV-1 TAR RNA. Biochimie 2013, 95, 1422–1431. [PubMed: 23541506]

(57). Suryawanshi H; Sabharwal H; Maiti S Thermodynamics of peptide-RNA recognition: the binding of a tat peptide to TAR RNA. J. Phys. Chem. B 2010, 114, 11155–11163. [PubMed: 20687526]

(58). Boudier C; Humbert N; Chaminade F; Chen Y; de Rocquigny H; Godet J; Mauffret O; Fosse P; Mely Y Dynamic interactions of the HIV-1 Tat with nucleic acids are critical for Tat activity in reverse transcription. Nucleic Acids Res 2014, 42, 1065–1078. [PubMed: 24153111]

(59). Tan R; Frankel AD Circular dichroism studies suggest that TAR RNA changes conformation upon specific binding of arginine or guanidine. Biochemistry 1992, 31, 10288–10294. [PubMed: 1384695]

(60). Churcher MJ; Lamont C; Hamy F; Dingwall C; Green SM; Lowe AD; Butler PJG; Gait MJ; Karn J High affinity binding of TAR RNA by the human immunodeficiency virus type-1 tat protein requires base-pairs in the RNA stem and amino acid residues flanking the basic region. J. Mol. Biol 1993, 230, 90–110. [PubMed: 8450553]

(61). Stelzer AC; Kratz JD; Zhang Q; Al-Hashimi HM RNA dynamics by design: biasing ensembles towards the ligand-bound state. Angew. Chem., Int. Ed 2010, 49, 5731–3.

(62). Muniz L; Egloff S; Ughy B; Jady BE; Kiss T Controlling cellular P-TEFb activity by the HIV-1 transcriptional transactivator Tat. PLoS Pathog 2010, 6, No. e1001152. [PubMed: 20976203]

(63). Lee J; Dethoff EA; Al-Hashimi HM Invisible RNA state dynamically couples distant motifs. Proc. Natl. Acad. Sci. U. S. A 2014, 111, 9485–9490. [PubMed: 24979799]

(64). Dethoff EA; Hansen AL; Musselman C; Watt ED; Andricioaei I; Al-Hashimi HM Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. Biophys. J 2008, 95, 3906–3915. [PubMed: 18621815]

(65). Note that the absence of 2D curved trajectories does not prove single-site binding; however, the existence of curvature under fast exchange requires multi-site binding.

(66). Arai M; Ferreon JC; Wright PE Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. J. Am. Chem. Soc 2012, 134, 3792–3803. [PubMed: 22280219]

(67). Quinternet M; Starck JP; Delsuc MA; Kieffer B Unraveling complex small-molecule binding mechanisms by using simple NMR spectroscopy. Chem. - Eur. J 2012, 18, 3969–74. [PubMed: 22336999]

(68). Furukawa A; Konuma T; Yanaka S; Sugase K Quantitative analysis of protein-ligand interactions by NMR. Prog. Nucl. Magn. Reson. Spectrosc 2016, 96, 47–57. [PubMed: 27573180]

(69). Freier SM; Albergo DD; Turner DH Solvent effects on the dynamics of (dG-dC)3. Biopolymers 1983, 22, 1107–31. [PubMed: 6850059]

(70). Pieters JML; Mellema J-R; Van Den Elst H; Van Der Marel GA; Van Boom JH; Altona C Thermodynamics of the various forms of the dodecamer d(ATTACCGGTAAT) and of its constituent hexamers from proton NMR chemical shifts and UV melting curves: three-state and four-state thermodynamic models. Biopolymers 1989, 28, 717–740. [PubMed: 2706311]

(71). Turner DL; Williams RJP 1H and 13C-NMR investigation of redox-state-dependent and temperature-depentent conformation changes in horse cytochrome c. Eur. J. Biochem 1993, 211, 555–562. [PubMed: 8382154]

(72). Davies DB; Pahomov VI; Veselkov AN NMR determination of the conformational and drug binding properties of the DNA heptamer d(GpCpGpApApGpC) in aqueous solution. Nucleic Acids Res 1997, 25, 4523–31. [PubMed: 9358161]

(73). Onufriev A; Ullmann GM Decomposing complex cooperative ligand binding into simple components: connections between microscopic and macroscopic models. J. Phys. Chem. B 2004, 108, 11157–11169.

(74). Berkhout B; Jeang K-T Trans activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the trans-acting-responsive hairpin: a quantitative analysis. J. Virol 1989, 63, 5501–5504. [PubMed: 2479775]

(75). Roy S; Delling U; Chen CH; Rosen CA; Sonenberg N A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. Genes Dev 1990, 4, 1365–1373. [PubMed: 2227414]

(76). Pham VV; Salguero C; Khan SN; Meagher JL; Brown WC; Humbert N; de Rocquigny H; Smith JL; D'Souza VM HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. Nat. Commun 2018, 9, 4266. [PubMed: 30323330]

(77). Wang S-Q; Li H-X Bayesian inference based modelling for gene transcriptional dynamics by integrating multiple source of knowledge. BMC Syst. Biol 2012, 6, S3.

(78). Thijssen B; Dijkstra TM; Heskes T; Wessels LF BCM: toolkit for Bayesian analysis of computational models using samplers. BMC Syst. Biol 2016, 10, 100. [PubMed: 27769238]

(79). Epstein M; Calderhead B; Girolami MA; Sivilotti LG Bayesian statistical inference in ion-channel models with exact missed event correction. Biophys. J 2016, 111, 333–348. [PubMed: 27463136]

(80). Saa PA; Nielsen LK Construction of feasible and accurate kinetic models of metabolism: a Bayesian approach. Sci. Rep 2016, 6, 29635. [PubMed: 27417285]

(81). Musselman C; Al-Hashimi HM; Andricioaei I iRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. Biophys. J 2007, 93, 411–422. [PubMed: 17449677]

(82). Ippolito JA; Steitz TAA 1.3-Å resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation. Proc. Natl. Acad. Sci. U. S. A 1998, 95, 9819–9824. [PubMed: 9707559]

(83). Olejniczak M; Gdaniec Z; Fischer A; Grabarkiewicz T; Bielecki L; Adamiak RW; Bielecki Ł; Adamiak RW The bulge region of HIV-1 TAR RNA binds metal ions in solution. Nucleic Acids Res 2002, 30, 4241–4249. [PubMed: 12364603]

(84). Casiano-Negroni A; Sun X; Al-Hashimi HM Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. Biochemistry 2007, 46, 6525–35. [PubMed: 17488097]

(85). Merriman DK; Yuan J; Shi H; Majumdar A; Herschlag D; Al-Hashimi HM Increasing the length of poly-pyrimidine bulges broadens RNA conformational ensembles with minimal impact on stacking energetics. RNA 2018, 24, 1363–1376. [PubMed: 30012568]

(86). Zacharias M; Hagerman PJ The bend in RNA created by the trans-activation response element bulge of human immunodeficiency virus is straightened by arginine and by Tat-derived peptide. Proc. Natl. Acad. Sci. U. S. A 1995, 92, 6052–6056. [PubMed: 7597079]

(87). Merriman DK; Xue Y; Yang S; Kimsey IJ; Shakya A; Clay M; Al-Hashimi HM Shortening the HIV-1 TAR RNA bulge by a single nucleotide preserves motional modes over a broad range of time scales. Biochemistry 2016, 55, 4445–56. [PubMed: 27232530]

(88). Wang Y; Gong S; Wang Z; Zhang W The thermodynamics and kinetics of a nucleotide base pair. J. Chem. Phys 2016, 144, 115101. [PubMed: 27004898]

(89). Frank AT; Stelzer AC; Al-Hashimi HM; 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 2009, 37, 3670–3679. [PubMed: 19369218]

(90). Salmon L; Bascom G; Andricioaei I; Al-Hashimi HM A general method for constructing atomic-resolution RNA ensembles using NMR residual dipolar couplings: the basis for interhelical motions revealed. J. Am. Chem. Soc 2013, 135, 5457–5466. [PubMed: 23473378]

(91). Cherepanov AV; Glaubitz C; Schwalbe H High-resolution studies of uniformly 13C,15N-labeled RNA by solid-state NMR spectroscopy. Angew. Chem., Int. Ed 2010, 49, 4747–4750.

(92). Clay MC; Ganser LR; Merriman DK; Al-Hashimi HM Resolving sugar puckers in RNA excited states exposes slow modes of repuckering dynamics. Nucleic Acids Res 2017, 45, No. e134. [PubMed: 28609788]

(93). Zhang Q; Sun X; Watt ED; Al-Hashimi HM Resolving the motional modes that code for RNA adaptation. Science 2006, 311, 653–6. [PubMed: 16456078]

(94). Delaglio F; Grzesiek S; Vuister GW; Zhu G; Pfeifer J; Bax A NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 1995, 6, 277–293. [PubMed: 8520220]

(95). Lee W; Tonelli M; Markley JL NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 2015, 31, 1325–7. [PubMed: 25505092]

(96). Wang AC; Bax A Minimizing the effects of radio-frequency heating in multidimensional NMR experiments. J. Biomol. NMR 1993, 3, 715–720. [PubMed: 8111234]

(97). Nickalls RWD A new approach to solving the cubic: Cardan's solution revealed. Mathematical Gazette 1993, 77, 354.

(98). Wang ZX; Jiang RF A novel two-site binding equation presented in terms of the total ligand concentration. FEBS Lett 1996, 392, 245–249. [PubMed: 8774854]

(99). Annis J; Miller BJ; Palmeri TJ Bayesian inference with Stan: a tutorial on adding custom distributions. Behav. Res. Methods 2017, 49, 863–886. [PubMed: 27287444]

(100). Carpenter B; Hoffman MD; Brubaker M; Lee D; Li P; Betancourt M The Stan math library: reverse-mode automatic differentiation in C++. arXiv:1509.07164 [cs.MS], arXiv Preprint, 2015 https://arxiv.org/abs/1509.07164.

(101). Carpenter B; Gelman A; Hoffman MD; Lee D; Goodrich B; Betancourt M; Brubaker M; Guo J; Li P; Riddell A Stan: a probabilistic programming language. J. Statistical Software 2017, 76, 1–32.

(102). Gelman A; Rubin DB Inference from iterative simulation using multiple sequences. Statistical Science 1992, 7, 457–472.

(103). Brooks SP; Gelman A General methods for monitoring convergence of iterative simulations. J. Comput. Graphical Statistics 1998, 7, 434–455.

(104). Sugase K; Dyson HJ; Wright PE Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 2007, 447, 1021–1025. [PubMed: 17522630]
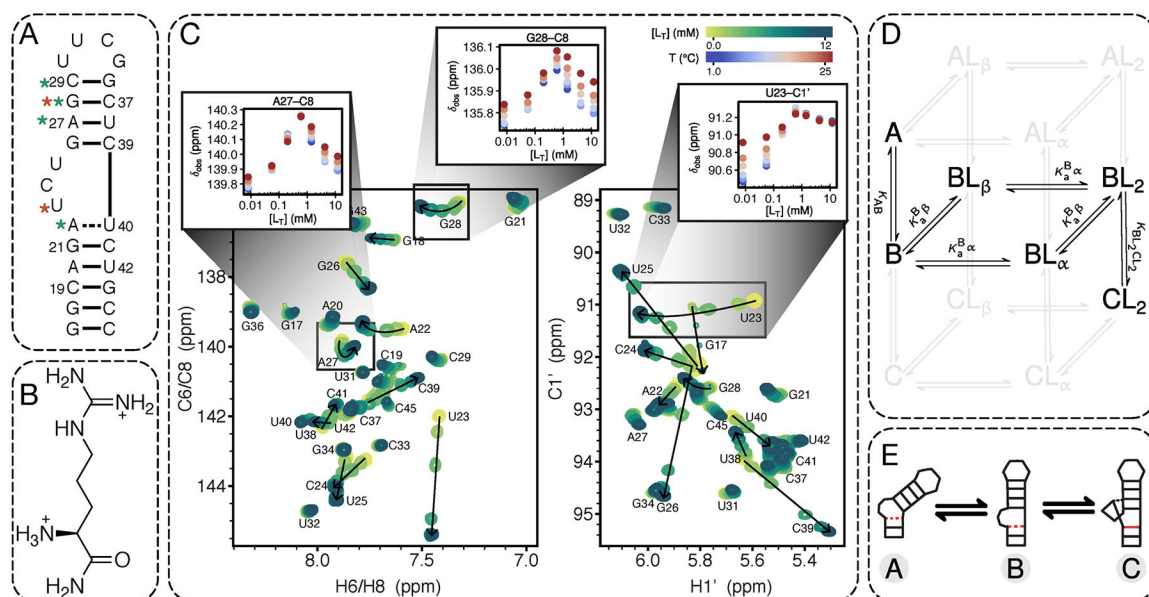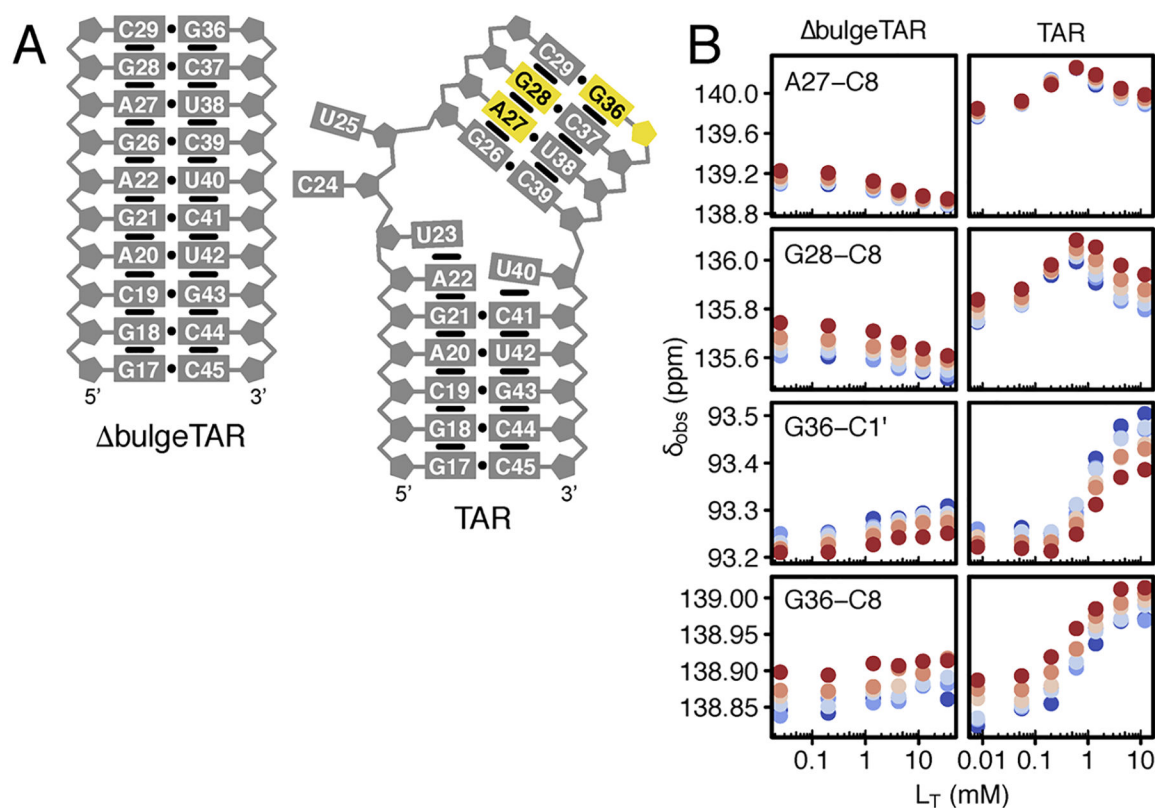
Figure 1.
NMR evidence for multi-site binding coupled with conformational equilibria. (A) Secondary structure of TAR construct used in this work. Red asterisk highlights residues that are particularly sensitive to temperature-dependent conformational equilibria and green asterisk highlights residues that show curved CSP trajectories indicating multiple binding sites. (B) Chemical structure of argininamide (ARG). (C) Chemical shifts as a function of ARG concentration at varying temperatures (blue, 1 °C to red, 25 °C) for representative resonances. Examples are shown for residues sensitive to multiple binding sites (A27-C8 and G28-C8) evidenced by curved CSP trajectories, and residues sensitive to conformational equilibria (U23-C1′ and G28-C8) evidenced by CSPs that vary significantly with temperature. Note that G28-C8 is sensitive to the $BL_2 \rightleftharpoons CL_2$ equilibria in addition to multi-site binding, while U23-C1′ reports on $A \rightleftharpoons B$ but is almost entirely insensitive to $BL_2 \rightleftharpoons CL_2$. Shown are the aromatic C6/C8-H6/H8 (left) and sugar C1′-H1′ (right) spectra at 25 °C as a function of increasing ARG concentration showing that the peaks shift in accordance with fast exchange. Spectra of all raw data can be found in Figure S2. (D) Minimal six-state model consistent with the observed CSP trends. In lighter gray are the equilibria that complete the statistical thermodynamic description for a system with three conformations and two independent binding sites per conformation, giving a total of 12 possible RNA species. (E) Cartoon depictions of the three TAR conformational states discussed in the text.

Figure 2.
Evidence for bulge-dependent secondary binding site in the TAR upper helix. (A) Comparison of secondary structures of the TAR and the bulgeless mutant in the absence of ligand. Resonances colored in yellow correspond to the putative secondary binding site. (B) CSPs as a function of ARG concentration and temperature for residues that are sensitive to the bulge-dependent secondary binding site on TAR that is abolished in the bulgeless mutant. Color coding of temperatures follows that of Figure 1C. The data supports that the secondary binding site is localized to the upper helix region between residues A27 and G36. The curved trajectories seen for A27-C8 and G28-C8 in TAR indicate they are sensitive to both the α and β sites. This is consistent with previous results that the α site ARG is stacked above A22 and contacts G26 just below A27. Note that the chemical shifts for G36-C1′ and G36-C8 in TAR are only sensitive to the β site ligand.
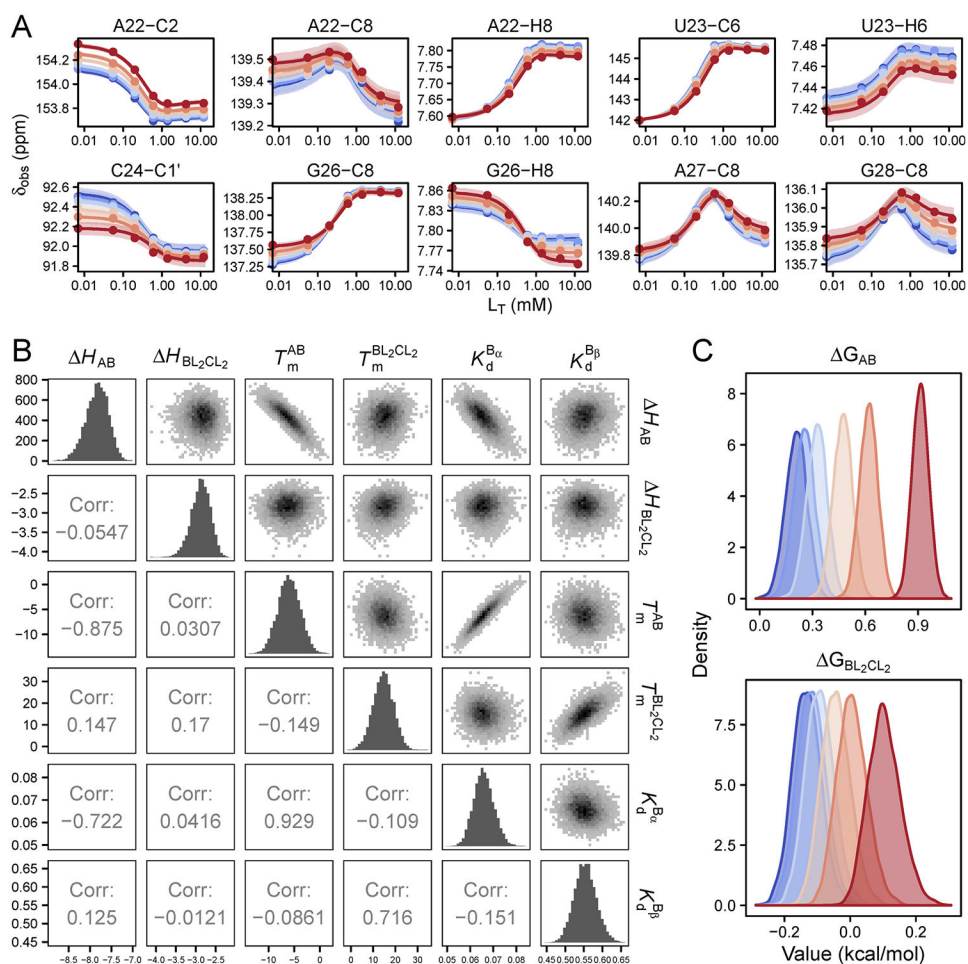
Figure 3.
Bayesian global fitting of CSPs. (A) Representative fits to 10 resonances out of a global fitting set containing 54 resonances. Shown are the 95% confidence interval regions at each temperature. Fits for all 54 resonances included in the global fit are shown in Figure S8. The points and curves are colored from 1 °C (blue) to 25 °C (red). (B) Joint posterior distributions of thermodynamic parameters with marginal posterior distributions down the diagonal. (C) Marginal posterior distributions of ΔG values for the A $\rightleftharpoons$ B and $BL_2 \rightleftharpoons CL_2$ reactions at each temperature. They are colored using the same color scheme used in panel A.
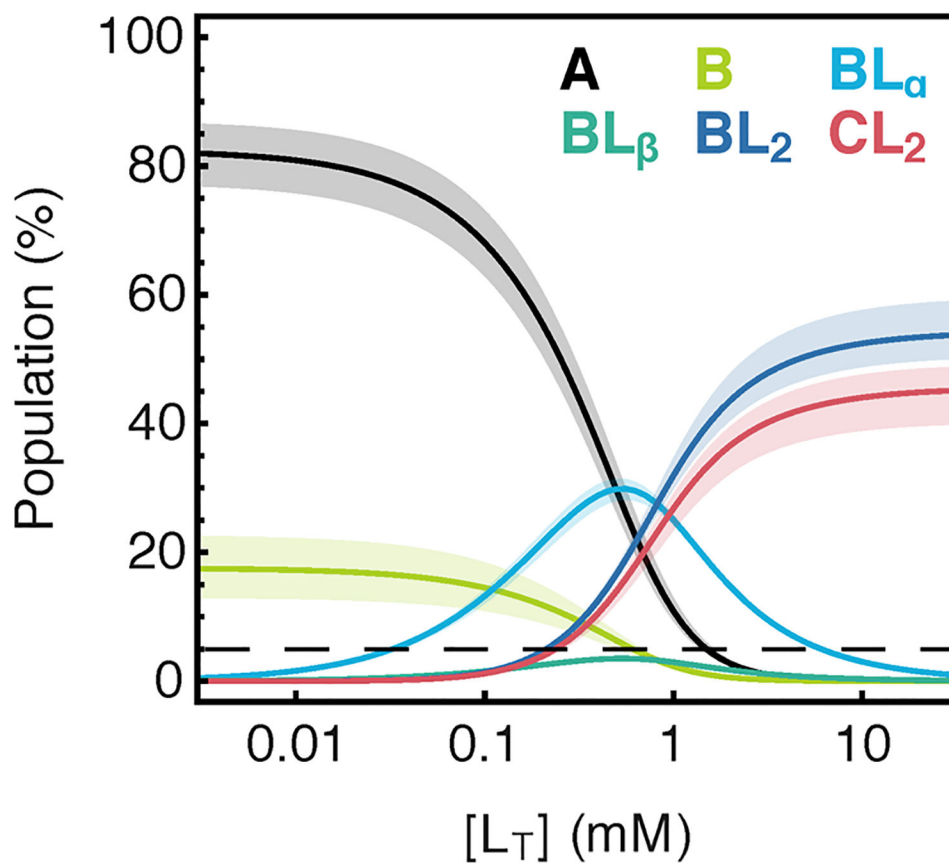
Figure 4.
Populations vs [ARG] for all species in the six-state model calculated using fitted parameter confidence intervals at 25 °C. The dotted line refers to the ~5% population cutoff for detection of species by the CSP titration method.
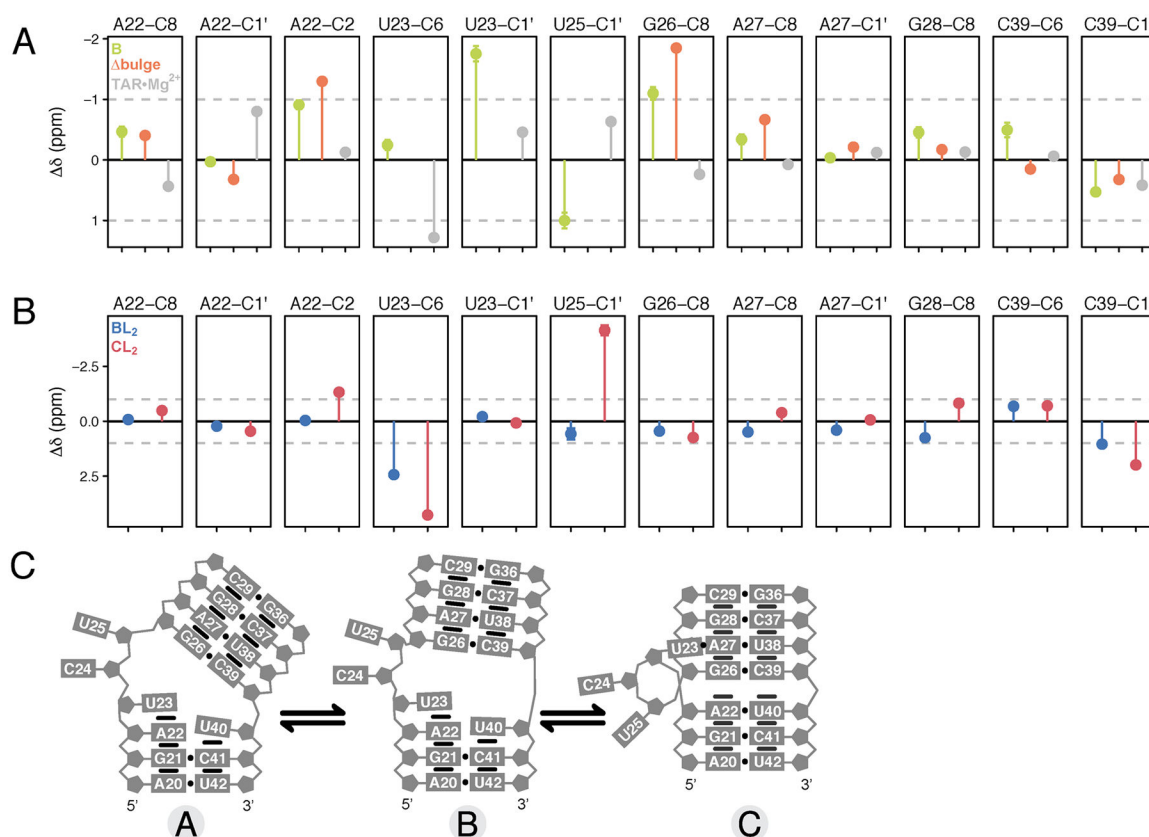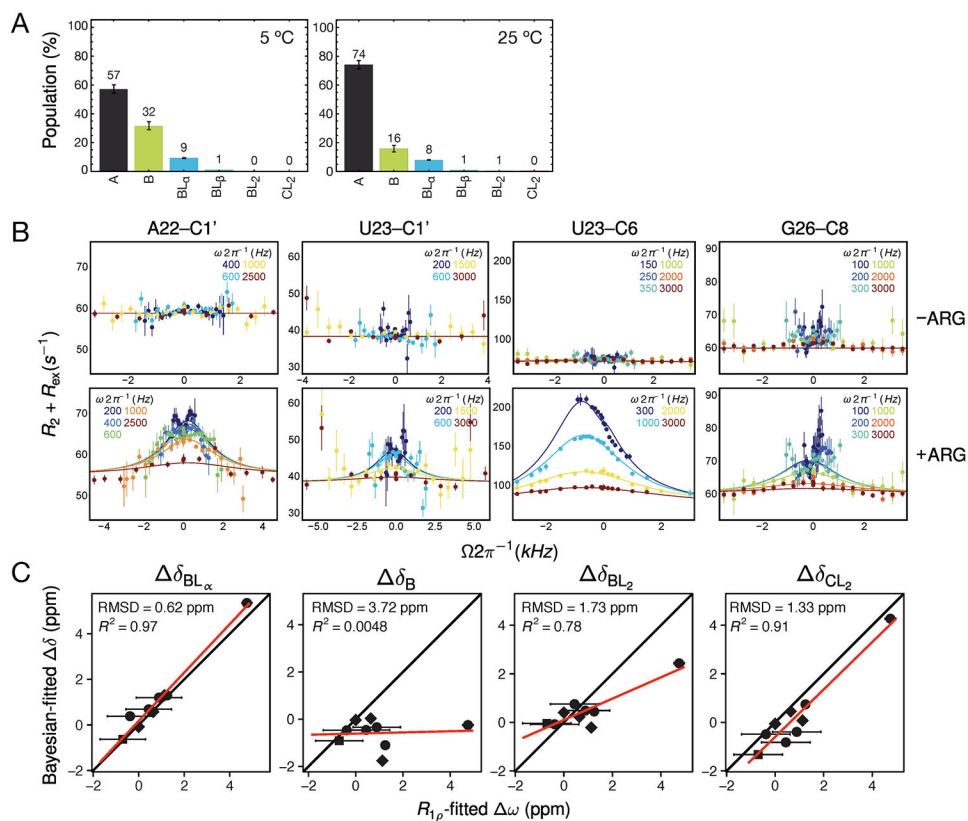
Figure 5.
Bayesian-fitted chemical shifts of intermediate species. (A) Comparison of chemical shift changes $\Delta\delta$ obtained for species B by Bayesian global fitting (green) with chemical shifts measured for ΔbulgeTAR in 25 mM NaCl (orange) and TAR in 25 mM NaCl plus 3 mM free $MgCl_2$ (buffer exchanged) (gray) for representative resonances. (B) Comparison of $\Delta\delta$ between species BL_2 (blue) and CL_2 (red) by Bayesian global fitting. The reference state is the unliganded A state and all $\Delta\delta$ are shown relative to the Bayesian-fitted δ_A values. Error bars represent the standard deviation determined by Bayesian fitting. (C) Schematic of the structural changes between the three different conformational states.

Figure 6.
Testing the model by NMR RD. (A) The predicted distribution of RNA states at the conditions under which the RD measurements were run (0.2 mM ARG, 1.5 mM RNA) at 5 and 25 °C. (B) Off-resonance RD data for the resonances with detectable chemical exchange at 5 °C. Lines represent global fit to BM equations. (C) Comparison of fitted chemical shifts obtained by RD with the chemical shifts obtained by Bayesian fitting to CSP data. Closest agreement is with chemical shifts for state BL_a, consistent with RD reporting on the binding to the a site in the B conformation.
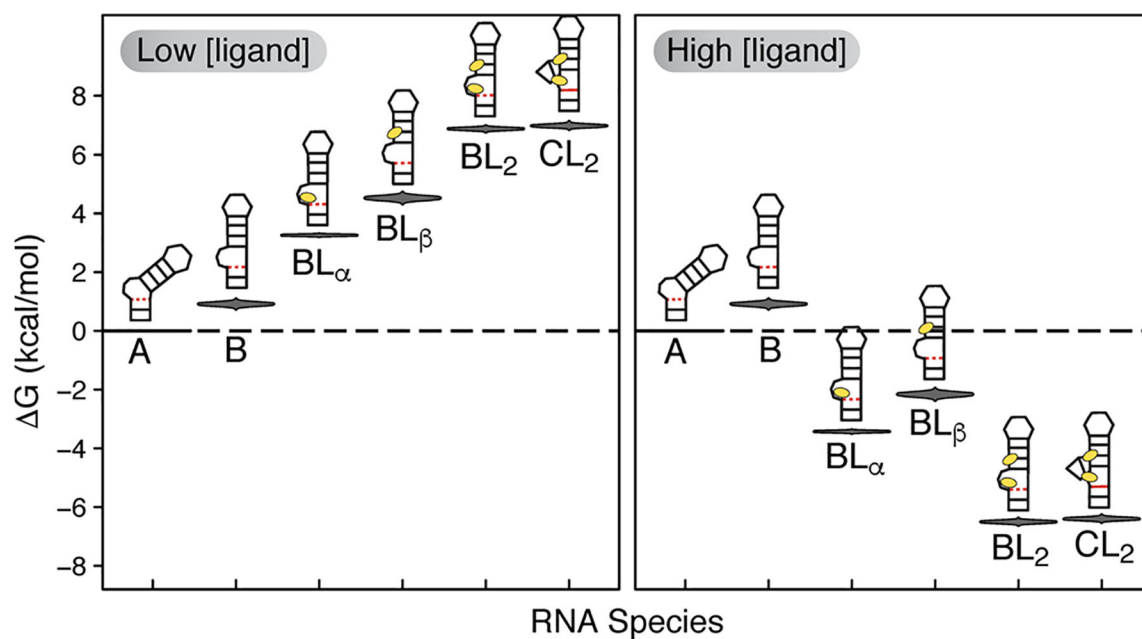
Figure 7.
Energetic distributions of the six detectable RNA species showing the effect of ligand concentration on the RNA energy landscape. The violin plots show the ΔG posterior distributions calculated at 25 °C relative to the A state.

Table 1.

Thermodynamic Parameter Estimates and Their 5% and 95% Confidence Intervals

Parameter	estimate	5% C.I.	95% C.I.
H_{AB} (kcal mol^{-1})	−7.86	−8.33	−7.45
$\Delta H_{BL_2CL_2}\left(\text{kcal mol}^{-1}\right)$	−2.85	−3.27	−2.47
$T_m^{AB}(_{\circ}C)$	−6.08	−9.63	−2.46
$T_m^{B_2CL2}(_{\circ}C)$	14.55	6.92	22.05
S_{AB} (cal mol^{-1}K^{-1})^a	−29.44	−30.83	−28.23
$\Delta S_{BL_2CL_2}\left(\text{cal mol}^{-1}\text{K}^{-1}\right)^a$	−9.90	−11.45	−8.55
G_{AB} (kcal mol^{-1})^{a,b}	0.91	0.84	0.99
$\Delta G_{BL_2CL_2}\left(\text{kcal mol}^{-1}\right)^{a,b}$	0.11	0.03	0.19
$K_d^{B\alpha}(\text{mM})$	0.066	0.060	0.073
$K_d^{B\beta}(\text{mM})$	0.55	0.51	0.60

^{a}Derived parameter.

^{b}At 25 °C.