

Disentangling contact and ensemble epistasis in a riboswitch

Daria R. Wonderlick^{1,4}, Julia R. Widom^{1,2,3}, Michael J. Harms^{1,2*}

1. Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR 97403
2. Institute for Molecular Biology, University of Oregon, Eugene, OR 97403
3. Oregon Center for Optical, Molecular, & Quantum Science, University of Oregon, Eugene, OR 97403
4. Present address: Biophysics Graduate Program, Stanford University, Stanford, CA 94305

*Corresponding author

harmsm@uoregon.edu

Abstract

Mutations introduced into macromolecules often exhibit epistasis, where the effect of one mutation alters the effect of another. Knowing the mechanisms that lead to epistasis is important for understanding how macromolecules work and evolve, as well as for effective macromolecular engineering. Here we investigate the interplay between “contact epistasis” (epistasis arising from physical interactions between mutated residues) and “ensemble epistasis” (epistasis that occurs when a mutation redistributes the conformational ensemble of a macromolecule, thus changing the effect of the second mutation). We argue that the two mechanisms can be distinguished in allosteric macromolecules by measuring epistasis at differing allosteric effector concentrations. Contact epistasis manifests as non-additivity in the microscopic equilibrium constants describing the conformational ensemble. This epistatic effect is independent of allosteric effector concentration. Ensemble epistasis manifests as non-additivity in thermodynamic observables—such as ligand binding—that are determined by the distribution of ensemble conformations. This epistatic effect strongly depends on allosteric effector concentration. Using this framework, we experimentally investigated the origins of epistasis in three pairwise mutant cycles introduced into the adenine riboswitch aptamer domain by measuring ligand binding as a function of allosteric effector concentration. We found evidence for both contact and ensemble epistasis in all cycles. Further, we found that the two mechanisms of epistasis could interact with each other. For example, in one mutant cycle we observed 6 kcal/mol of contact epistasis in a microscopic equilibrium constant. In that same cycle, the maximum epistasis in ligand binding was only 1.5 kcal/mol: shifts in the ensemble masked the contribution of contact epistasis. Finally, our work yields simple heuristics for identifying contact and ensemble epistasis based on measurements of a biochemical observable as a function of allosteric effector concentration.

Statement of significance

Mutations to protein or RNA molecules often have different effects when introduced individually versus together. To understand and engineer biological macromolecules, we must identify the mechanistic origins of this phenomenon. Here, we measured the interplay between direct, physical interactions between mutations (“contact epistasis”) and indirect interactions mediated by conformational ensembles (“ensemble epistasis”). We introduced pairs of mutations into an RNA molecule that transitions between several different conformations. We found epistasis arising from both contacts and the ensemble, and that the two mechanisms could synergize with one another. Our work reveals that one must consider the effects of mutations on multiple conformations to understand epistasis and suggests a few rules-of-thumb for disentangling contact and ensemble epistasis in other macromolecules.

Introduction

Mutations often exhibit epistasis, where the effect of one mutation alters the effect of another (1, 2). Intramolecular epistasis—interactions between mutations within the same macromolecule—is useful for dissecting molecular function (3, 4), can shape molecular evolution (2, 5–7), and must be properly captured for macromolecular engineering (8–10). Knowledge of the mechanisms of epistasis is thus important for both understanding and designing macromolecules.

Broadly, there are two sources of epistasis in macromolecules. The most intuitive is mediated by structural contacts between the sites. Such contacts could be simple one-to-one interactions (e.g. ion pairs, base pairs, hydrogen bonds, etc.) (7, 11, 12) or larger networks of interactions (e.g. pathways of residues connecting distant sites) (4, 13). The second mechanism arises from the ensemble of conformations populated by the macromolecule (14–16). This epistasis arises when a macromolecule populates several conformations that respond differently to the same mutation. A mutation at one site redistributes the relative populations of the conformations, thus altering the effect of a second mutation. The relative contributions of contact and ensemble mechanisms to epistasis across macromolecules are unknown.

We recently found that environment-dependence can reveal ensemble epistasis (14). If one changes the environment such that the populations of the conformational ensemble are altered, one can change the contribution of ensemble epistasis. Many environmental factors—including pH, temperature, ionic strength, etc.—can, in principle, tune ensemble epistasis. For the current study, we used the concentration of allosteric effector as our environmental perturbation. This is because the effector redistributes the ensemble of an allosteric macromolecule—and thus the measurable output of the system—in a predictable fashion. We recently demonstrated ensemble epistasis

experimentally using the lac repressor, which exhibited epistasis in DNA binding as a function of its allosteric effector IPTG (17). Intriguingly, we also found that redistribution of the ensemble could not fully explain the epistasis we observed—suggesting that other mechanisms besides ensemble epistasis were at play in the protein.

Our previous work raised several questions. How can we distinguish between epistasis arising from contacts, the ensemble, or both simultaneously? What are the relative magnitudes of these two forms of epistasis? How do they influence one another? Is ensemble epistasis within the purview of proteins alone, or does it arise in other classes of macromolecules?

To address these questions, we investigated the contributions of contact and ensemble epistasis between pairs of mutations in the well-characterized adenine riboswitch aptamer domain. This RNA molecule has a three-state conformational ensemble that allows it to bind to adenine in a magnesium-dependent fashion (18–25). We measured magnesium-dependent epistasis between mutations we expected to have varying contributions of contact and ensemble epistasis, finding evidence for both mechanisms in all mutant cycles. Our work reveals that contact and ensemble epistasis can have comparable magnitudes, and that the two sources of epistasis can interact in highly nonlinear ways. Further, our results suggest that one can use environment-dependence as a useful heuristic for disentangling ensemble and contact epistasis in macromolecules.

Materials and Methods

RNA sample preparation

We synthesized sequence variants of the *V. vulnificus* adenine riboswitch aptamer domain by *in vitro* transcription of the corresponding DNA oligonucleotides (HiScribe T7 High Yield RNA Synthesis Kit, New England Biolabs; oligonucleotides ordered from Eurofins). The wildtype RNA sequence is shown with mutated positions in boldface: 3'GGGAAGAUAAUCCUAUAUGAUAUGGUUUGGGAGUUUCUACCAAGAGCCUUAACUCUUGAUUAUCUUCCC. The riboswitch product was purified by running the *in vitro* transcription reaction on a 12% denaturing polyacrylamide gel, identifying the 71-nt product by UV shadowing, and extracting the product from the gel by electroelution. The purified RNA product was concentrated by ethanol precipitation, desalted, quantified by Nanodrop absorbance spectroscopy, and resuspended in 50 mM Tris-HCl before being stored at -80 °C until further use.

Measurement of 2-aminopurine binding

We measured binding of 2-aminopurine (2AP), a fluorescent analog of adenine, to the riboswitch aptamer domain using a previously published protocol (19, 25). Riboswitch aliquots were thermally denatured at 90 °C for 1 min before refolding on ice for approximately 10 min. Increasing concentrations of RNA (0.125-8 µM) were combined with 50 nM 2AP, 100 mM KCl, and 0.1-100 mM MgCl₂ in 96-well black bottom plate. The assay plate was shaken in a Molecular Devices SpectraMax i3 fluorescence plate reader at 37 °C for 5 minutes to allow the reaction mixtures to reach equilibrium. 2AP exhibits high fluorescence in water and quenched fluorescence when bound to the riboswitch. We excited each sample at 310 nm and measured emission between 335 and 450 nm. We integrated each emission spectrum and converted the area to fraction of 2AP

bound using a negative control (0 nM RNA, 0 nM 2AP) and a positive control (0 nM RNA, 50 nM 2AP). Because 2AP fluorescence was negligible in samples where the fluorophore was completely bound by RNA, we used a negative control lacking 2AP entirely to universally mimic this fully quenched state. We averaged two technical replicates for each condition per experiment and collected at least three biological replicates for each experimental condition.

RNA ensemble modeling

As described in the results, we used a previously validated thermodynamic model of the riboswitch ensemble to extract information about the populated conformations (23). This model describes the riboswitch with two apo conformations—extended (E) and docked (D)—each of which can form a complex with 2AP (E·A and D·A). We tested three variants of the model: a two-state model with conformations E and D·A, a three-state model with conformations E, D, and D·A, and a four-state model with conformations E, E·A, D, and D·A. We also tried subvariants of these models in which we fixed the values of the different equilibrium constants. As described in the results, we found that the three-state model with K_{dock} set to 1.0 was able to reproduce our binding data without overfitting. A mathematical description of the three-state model follows; the two-state and four-state models are given in the supplement (Fig S1).

The three-state model is shown schematically in Fig 1A. The concentrations of the E, D, and D·A conformations are given by:

$$[E] = \frac{[R]_T - [A]_T + [A]}{1 + K_{dock}(K_{MG}[M]_T)^{n_{MG}}} \quad 1.$$

$$[D] = [E]K_{dock}(K_{MG}[M]_T)^{n_{MG}} \quad 2.$$

$$[D \cdot A] = [D]K_{2AP}[A] \quad 3.$$

where K_{dock} is the equilibrium between E and D in the absence of Mg^{2+} or 2AP, K_{2AP} is the affinity of 2AP for the D conformation, K_{MG} is the relative affinity of the D and E conformations for Mg^{2+} , n_{MG} is the difference in the number of Mg^{2+} ions bound by the D and E conformations, $[R]_T$ is the total concentration of RNA, $[M]_T$ is the total concentration of Mg^{2+} , $[A]_T$ is the total concentration of 2AP, and $[A]$ is the concentration of free 2AP. Our experimental observable, the fraction of 2AP bound to RNA, is given by:

$$\theta = \frac{[D \cdot A]}{[A]_T} = \frac{([R]_T - [A]_T - [A])K_{dock}(K_{MG}[M])^{n_{MG}}K_{2AP}[A]}{[A]_T(1 + K_{dock}(K_{MG}[M])^{n_{MG}})} \quad 4.$$

This model has four explicit parameters: K_{dock} , K_{MG} , n_{MG} , and K_{2AP} . It also has two implicit parameters: $[M]$ (the concentration of free Mg^{2+}) and $[A]$ (the concentration of free 2AP).

We implemented this model as a function that returns the concentrations of all relevant species ($[A]$, $[M]$, $[E]$, $[D]$, and $[D \cdot A]$) given the values of the thermodynamic parameters (K_{dock} , K_{2AP} , K_{MG} , and n_{MG}) and total species concentrations ($[R]_T$, $[M]_T$, and $[A]_T$). This function encodes the thermodynamic relationships above (equations 1-4) and enforces mass-balance relationships (e.g. $[A]_T = [A] + [D \cdot A]$). Internally, this function guesses values for $[A]$ and then iterates to self-consistency between the thermodynamic and mass-balance relationships. We derived the model allowing for any of the binding equilibria to be in a stoichiometric or non-stoichiometric regime. The full derivation is given in Appendix S1.

To estimate the values of the thermodynamic parameters consistent with our binding data, we used a Bayesian Markov-Chain Monte Carlo (MCMC) strategy to sample over parameter combinations. We analyzed all experimental conditions simultaneously for each genotype, globally estimating the thermodynamic parameters. We used the following likelihood function:

$$\mathcal{L} = -\frac{1}{2} \sum_i \left[\frac{(\theta(\vec{p}, \vec{c})_{i, \text{calc}} - \theta_{i, \text{obs}})^2}{\sigma_i^2} + \ln(2\pi\sigma_i^2) \right] \quad 5.$$

where $\theta_{i, \text{obs}}$ and σ_i are the mean and standard deviation of the measured 2AP fractional saturation at condition i ; $\theta_{i, \text{calc}}$ is the calculated 2AP fractional saturation under these conditions given a vector of parameters (\vec{p}) and total concentrations (\vec{c}). If our experimental σ_i was < 0.05 , we set its value to 0.05, thus enforcing a minimum uncertainty on the mean of any observed data point. We performed MCMC sampling on the natural logs of all equilibrium constants. We generated 15,000 MCMC samples from 100 MCMC walkers for each genotype, discarding the first 1,000 as burn in. We checked for convergence by comparing results from multiple independent runs. For the final three-state, three-parameter model we used in our analysis, we generated a million MCMC samples from six different sampling runs. We used the emcee 3.1.0 (26) and the “likelihood” python libraries (<https://github.com/harmslab/likelihood>) for these calculations. We implemented the riboswitch binding models in Python 3.9 extended with numpy 1.21.1 (27), pandas 1.3.1 (28), and scipy 1.6.2 (29).

Data availability

All experimental data, software, and scripts to reproduce analyses reported in this manuscript are available on github (<https://github.com/harmslab/riboswitch-epistasis>).

Results

Selecting the adenine riboswitch as a model system

We set out to measure the relative contributions of contact and ensemble epistasis using mutant cycles introduced into the adenine riboswitch aptamer domain, a 71-nucleotide RNA molecule composed of three helices surrounding an adenine binding pocket (30) (Fig 1A, Fig 2A). We selected this RNA molecule as our model system because it has a relatively simple, well-defined conformational ensemble consisting of two conformations: an extended form (E) and a compact, docked-loop form (D) (Fig 1A). Relative to the E conformation, the D conformation has extensive base pairs, hydrogen bonds, and hydrophobic packing between the two helical arms of the riboswitch. This forms an adenine binding pocket unique to the D conformation; therefore, the D conformation has much higher affinity for adenine than the E conformation. Mg^{2+} ions are key allosteric effectors for the riboswitch ensemble because they nonspecifically neutralize repulsion between backbone phosphate groups to promote formation of the compact D conformation, thus promoting adenine binding (23, 31, 32).

Quantifying epistasis between mutations to the riboswitch requires measuring how mutations affect a biological function when introduced individually versus together. As a proxy for the biological function of the riboswitch, we measured its binding to 2-aminopurine (2AP), a fluorescent analog of adenine. The 2AP binding energy is given by:

$$\Delta G_{obs} = -RT \ln \left(\frac{[D \cdot A]}{([D] + [E])[A]} \right) \quad 6.$$

where the conformations are those labeled in Fig 1A. We can calculate the concentrations of each of these species using Equations 1-3, which require knowing the total concentrations of the molecular components (Mg^{2+} , RNA, and 2AP), two apparent equilibrium constants (K_{MG} and

K_{2AP}), and an ion uptake coefficient (n_{MG}). We can rewrite our 2AP binding energy in terms of these equilibrium constants by:

$$\Delta G_{obs} = -RT \ln \left(\frac{(K_{MG}[M]_T)^{n_{MG}} K_{2AP}}{1 + (K_{MG}[M]_T)^{n_{MG}}} \right) \quad 7.$$

where $[M]_T$ is the total Mg^{2+} concentration. The 2AP binding energy can also be calculated directly from experimental measurements of the fraction of 2AP bound to RNA using:

$$\Delta G_{obs} = -RT \ln \left(\frac{\theta[A]_T}{[R]_T - \theta[A]_T} \right) \quad 8.$$

where $[A]_T$ is the total 2AP concentration, $[R]_T$ is the total RNA concentration, and θ is the fraction of 2AP bound to RNA (Equation 4). Describing ΔG_{obs} in terms of the system equilibrium constants (Equation 7) or fraction of 2AP bound to RNA (Equation 8) allows us to calculate a single thermodynamic observable that encapsulates the riboswitch's conformational ensemble without needing to directly measure the population of each conformation spectroscopically.

Contact and ensemble epistasis will give different signals

We next asked how contact and ensemble epistasis would manifest conceptually and mathematically within the riboswitch system. Contact epistasis occurs when two mutations interact within a conformation, thus changing the energy of that conformation. Consider the predicted effects of a pair of mutations ($a \rightarrow A$ and $b \rightarrow B$) at residues that form complementary contacts in the docked conformation (D) of the riboswitch (Fig 1B). When introduced individually, these mutations create non-complementary interactions and shift the equilibrium towards the extended form (E). When introduced together, however, complementarity is restored, thus pushing the equilibrium back towards the docked conformation. Such contact epistasis will manifest in the microscopic equilibrium between these two conformations. This can be quantified using the

biological standard state free energy for this equilibrium, $\Delta G^{\circ'} = -RT\ln(K_{\text{eq}})$, yielding the following expression for contact epistasis:

$$\Delta\Delta G_{\text{contact}} = (\Delta G_{AB}^{\circ'} - \Delta G_{aB}^{\circ'}) - (\Delta G_{Ab}^{\circ'} - \Delta G_{ab}^{\circ'}). \quad 9.$$

Contact epistasis is independent of the concentrations of other species in solution—such as that of Mg^{2+} —because it arises from interactions within a conformation rather than a non-linear redistribution of conformations. Thermodynamically, contact epistasis occurs in standard state free energies to reflect that the interactions are concentration independent.

Ensemble epistasis, by contrast, occurs when a mutation redistributes the relative populations within the ensemble, thus changing the effect of a second mutation on an ensemble-averaged observable. Consider another pair of mutations ($x \rightarrow X$ and $y \rightarrow Y$) to the riboswitch (Fig 1C). If each mutation changes one or more of the equilibrium constants (Equation 7), each of the xy , Xy , xY and XY riboswitch variants will have different populations of the E, D, and D·A conformations at a given Mg^{2+} concentration. The altered equilibrium constants change the apparent 2AP affinity (ΔG_{obs}) for each variant in a nonlinear fashion. One can calculate ensemble epistasis as:

$$\Delta\Delta G_{\text{ensemble}}([\text{Mg}^{2+}]) = (\Delta G_{\text{obs},XY} - \Delta G_{\text{obs},xY}) - (\Delta G_{\text{obs},xy} - \Delta G_{\text{obs},xy}). \quad 10.$$

Ensemble epistasis depends on the Mg^{2+} concentration because the population of the ensemble—and thus ΔG_{obs} —is controlled by $[\text{Mg}^{2+}]$. At low Mg^{2+} concentration, only the E conformation is populated. At high Mg^{2+} concentration, we expect ΔG_{obs} to plateau because 2AP becomes limiting. Additional Mg^{2+} can no longer change the relative population of D·A and thus can no longer alter ΔG_{obs} . At intermediate Mg^{2+} concentrations, multiple conformations are populated, thus maximizing the redistributive effects of each mutation and the magnitude of ensemble epistasis (shown schematically in Fig 1C).

In summary, contact epistasis appears in the standard state free energy of a specific conformation, while ensemble epistasis appears in an ensemble-averaged observable, such as a free energy of 2AP binding. Because contact epistasis could, in principle, strongly perturb the population of the conformational ensemble, we expect a complicated interplay between the two mechanisms. Without experimental knowledge of the relative magnitudes of each mechanism, however, we cannot predict what this synergy might look like for a real macromolecule.

Mutant cycle design

We next designed three mutant cycles for which we predicted different contributions of contact and ensemble epistasis. We built these cycles from four mutations that were previously reported to alter the riboswitch's 2AP and Mg²⁺ affinity (Fig 2A) (19, 20). All cycles involve c60g (Fig 2A; blue), paired with either a35u, g38c, or c50a (Fig 2A; gray, orange, and pink). For consistency, we measured epistasis across all cycles as the difference between the effect of c60g introduced into a mutant riboswitch versus the wildtype riboswitch. In the wildtype background, c60g disrupts the complementarity of a Watson-Crick-Franklin base pair at the docked-loop interface of the D conformation, leading to a decrease in 2AP affinity (19, 20).

We expected the a35u/c60g pair to exhibit minimal epistasis of either type. a35u has been reported to slightly increase ligand binding affinity by adding an extra hydrogen bond to the docked-loop interface (20). a35 is not in direct contact with c60, and the two sites are not in allosterically linked regions of the riboswitch (Fig 2A). We therefore anticipated little epistasis between these positions.

We expected the g38c/c60g pair to give strong contact epistasis (Fig 1B). g38 and c60 form a Watson-Crick-Franklin base pair at the docked-loop interface of the wildtype riboswitch's D

conformation (Fig 2A) (19, 20). When introduced individually, these mutations disrupt this base pair and thus greatly decrease 2AP binding affinity. When both mutations are introduced together, c38/g60 form a reversed Watson-Crick-Franklin base pair, which has been shown to restore 2AP binding activity (19).

We expected the c50a/c60g pair to exhibit significant ensemble epistasis but little contact epistasis (Fig 1C). Positions 50 and 60 are \sim 40 Å apart in the crystal structure (Fig 2A). c60 is at the docked-loop interface, while c50 is directly adjacent to the 2AP binding pocket. The c50a mutation has been shown previously to reduce 2AP binding affinity by destabilizing the ligand binding pocket (20). Because the docked-loop region and 2AP binding pocket are in allosteric communication, we anticipated epistasis between these two mutations; however, we predicted this would be mediated by the ensemble rather than a direct contact given the separation between the positions.

Defining the thermodynamic ensemble

With our system selected and our mutant cycles designed, we next set out to measure how the mutations perturbed the riboswitch ensemble. Previously, Leiply and Draper validated a thermodynamic model with two apo conformations of the adenine riboswitch—extended (E) and docked (D)—each of which can bind 2AP (E·A and D·A) (23). In a separate study (31), they found they could model RNA-Mg²⁺ interactions with an ion uptake coefficient that captured the number of excess Mg²⁺ ions bound by D relative to E. Put together, this results in a five parameter thermodynamic model (Fig S1) with the following terms: K_{dock} (the equilibrium between E and D in the absence of Mg²⁺), K_{2AP} (the affinity of the riboswitch for 2AP), K_{link} (the coupling between Mg²⁺/2AP binding and formation the D conformation), K_{MG} (the Mg²⁺ affinity of the D

conformation relative to the E conformation), and n_{MG} (the difference in the number of Mg^{2+} ions bound by the D and E conformations).

To resolve the parameters in this model, we measured 2AP binding over 4 orders of magnitude in Mg^{2+} concentration: 0.1, 1, 10, and 100 mM. This covers the physiological range of Mg^{2+} concentrations (33). We varied RNA concentrations over ~2 orders of magnitude, from 0.125 to 8 μM , while maintaining a constant, limiting 2AP concentration of 50 nM. The results of these experiments for all eight constructs are shown in Fig 2B. As expected, we observed increasing 2AP binding with increasing amounts of RNA for all variants. Further, we observed a link between 2AP binding and Mg^{2+} concentration, with the midpoint of the 2AP binding curves decreasing with increasing Mg^{2+} concentration.

We used a Bayesian Markov Chain Monte Carlo (MCMC) strategy to sample model parameters for each riboswitch variant that could reproduce our experimental observations. The four-state/five-parameter model reproduced our measured binding data well (Fig S2). We noted, however, that the E·A conformation never accounted for more than 0.02% of the total RNA concentration for all datasets. We therefore asked if we could simplify our model by ignoring the E·A conformation, thus eliminating the model parameter K_{link} . This three-state/four-parameter model reproduced our experimental data as well as the original model (Fig S2). We also noticed strong covariation between K_{MG} and K_{dock} across MCMC samples, so we tried a model in which we set K_{dock} to 1.0. In this formulation, K_{MG} is an apparent constant that accounts for both the energy of Mg^{2+} binding and the relative stability of the D versus E conformations. This three-state/three-parameter model still reproduced our experimental data well (Fig 2B).

Finally, we asked if we could simplify the model further by describing the system with only two conformations, E and D·A. This two-state model makes the riboswitch cooperative,

meaning Mg^{2+} binding necessitates 2AP binding. This two-state model could not reproduce our measured 2AP binding data (Fig S2). We used an Akaike Information Criterion (AIC) test to validate our visual inspection of the binding curves. The AIC test favors models with high likelihoods while penalizing models with excess parameters (34). Our AIC test favored the three-state/three-parameter model (Table S1). All fit parameters for the final three-state/three-parameter model are given in Table S2.

Mutant Cycle #1: a35u/c60g

We next analyzed the epistasis within each mutant cycle, starting with a35u/c60g. These bases do not contact one another and are distant from both the Mg^{2+} and 2AP binding sites (Fig 2A). We therefore expected relatively weak contact and ensemble epistasis for this mutant pair.

We first looked for contact epistasis in the a35u/c60g mutant cycle. We extracted marginal distributions for the values of $\Delta G^{\circ'}_{2AP}$, $\Delta G^{\circ'}_{MG}$ and n_{MG} from the MCMC samples for the relevant variants (Fig 3A-C) and calculated epistasis in each parameter using Equation 9. c60g had no detectable effect on $\Delta G^{\circ'}_{2AP}$ when introduced alone, but increased $\Delta G^{\circ'}_{2AP}$ by 0.6 [0.4, 0.8] kcal/mol in the a35u background. (In this notation, we write the median effect followed by the 95% credibility interval extracted from the marginal distribution in brackets). If the mutations exhibited no epistasis, we would expect the a35u/g38c double mutant to have $\Delta G^{\circ'}_{2AP}$ corresponding to the green point on Fig 3A. The actual change is larger, reflecting measurable epistasis in $\Delta G^{\circ'}_{2AP}$ (0.52 [0.28, 0.73]). We saw little evidence of epistasis in $\Delta G^{\circ'}_{MG}$ or n_{MG} : a35u and c60g both increased $\Delta G^{\circ'}_{MG}$ by ~0.5 kcal/mol (Fig 3B) and neither mutation had a measurable effect on n_{MG} (Fig 3C)

We next looked for evidence of ensemble epistasis in the a35u/c60g mutant cycle. We used Equation 7 to calculate ΔG_{obs} versus Mg^{2+} concentration for all four variants (Fig 3D). Each variant gave decreasing ΔG_{obs} with increasing Mg^{2+} concentration, reflecting the shift in equilibrium from E to D·A (Fig S3). ΔG_{obs} plateaued for all variants when the concentration of 2AP became limiting: additional Mg^{2+} could no longer change the relative concentration of D·A and thus could no longer alter ΔG_{obs} . We then calculated epistasis in ΔG_{obs} using Equation 10 (Fig 3E). At low Mg^{2+} concentration, our uncertainty was high and the signal indistinguishable from zero; for Mg^{2+} above 1 mM, we observed approximately constant epistasis in ΔG_{obs} (0.31 [0.17,0.43] kcal/mol).

We next investigated the interplay between contact and ensemble epistasis in the a35u/c60g mutant cycle by calculating magnesium-dependent epistasis in ΔG_{obs} after removing contact epistasis. We assumed that the parameter values for a35u/c60g double mutant were given by the sum of the median effects of a35u and c60g introduced alone, thereby eliminating epistasis in $\Delta G^{\circ'}_{2\text{AP}}$, $\Delta G^{\circ'}_{\text{MG}}$ and n_{MG} . These non-epistatic parameters are shown as the green points for a35u/c60g in Fig 3A-C. The epistasis in ΔG_{obs} we observe in the absence of contact epistasis is shown as a green line in Fig 3E. The curve behaves as expected for a system exhibiting ensemble epistasis. It peaks when the diversity of the ensemble is maximized and tails off to zero when the system populates a single conformation at high and low Mg^{2+} concentrations (Fig S3). The ensemble contributes maximally to the epistasis in this mutant cycle around 1 mM Mg^{2+} (labeled arrow, Fig 3E). At higher Mg^{2+} concentrations, contact epistasis in 2AP binding ($\Delta G^{\circ'}_{2\text{AP}}$) is the source of epistasis between the mutations (labeled arrow, Fig 3E).

Mutant Cycle #2: g38c/c60g

We next turned to the g38c/c60g mutant cycle. We expected these mutations to yield high-magnitude contact epistasis because the nucleotides form a Watson-Crick-Franklin base pair in the D conformation that inverts complementarity over the mutant cycle (Fig 2A).

As with the previous mutant cycle, we looked for evidence of contact epistasis in the g38c/c60g mutant cycle by considering the standard state free energies for the equilibria defining the system. We calculated the marginal distributions of $\Delta G^{\circ'}_{2AP}$, $\Delta G^{\circ'}_{MG}$ and n_{MG} for all four variants (Fig 4A-C). The parameters were well constrained for the wildtype, c60g, and g38c/c60g variants, but not g38c introduced alone. The g38c variant exhibits low 2AP binding, even at high Mg^{2+} and RNA concentrations (Fig 2B). Based on our MCMC samples, we are confident that g38c differs from the other variants; however, this could arise because the mutation increases $\Delta G^{\circ'}_{MG}$, increases $\Delta G^{\circ'}_{2AP}$, or some combination of the two (Fig 4D). This leads to a strong inverse correlation in estimates of $\Delta G^{\circ'}_{2AP}$ and $\Delta G^{\circ'}_{MG}$ (Fig 4D).

We calculated contact epistasis in $\Delta G^{\circ'}_{2AP}$, $\Delta G^{\circ'}_{MG}$ and n_{MG} for the g38c/c60g mutant cycle by drawing from each parameter's MCMC distributions for each variant and applying Equation 9. We observed strong covariation in epistasis between $\Delta G^{\circ'}_{2AP}$ and $\Delta G^{\circ'}_{MG}$ (Fig 4E). The centroid of the distribution yields estimates of the epistasis in each parameter as 1 and -6 kcal/mol, respectively (red point, Fig 4E). Only 0.4% of samples have epistasis < 1 kcal/mol in both $\Delta G^{\circ'}_{2AP}$ and $\Delta G^{\circ'}_{MG}$, strongly suggesting contact epistasis is present in one or both parameters.

To look for ensemble epistasis in the g38c/c60g mutant cycle, we first calculated ΔG_{obs} vs. Mg^{2+} concentration for all four variants (Fig 4F). Because of the strong covariation in our estimates of $\Delta G^{\circ'}_{2AP}$ and $\Delta G^{\circ'}_{MG}$, our confidence in ΔG_{obs} for g38c is high despite low confidence in the absolute magnitudes of these parameters. We found that ΔG_{obs} plateaued for all variants

except g38c. This reflects the fact that 2AP binding approaches saturation for all variants except g38c (Fig S3).

We observed striking magnesium-dependent epistasis in ΔG_{obs} for the g38c/c60g mutant cycle (Fig 4G, Equation 10). The value is indistinguishable from zero at low and high Mg^{2+} concentrations. We see epistasis of -1.5 [-1.9, -1.3] kcal/mol, peaking around the point where the riboswitch binds Mg^{2+} and 2AP (Fig S3). The negative sign indicates the double mutant variant is 1.5 kcal/mol *better* at binding 2AP than we would expect based on the effects of the mutations alone. This is driven largely by the deleterious effect of g38c when introduced alone (blue curve; Fig 4F) and aligns with previous measurements of restored 2AP binding upon re-establishing a Watson-Crick-Franklin base pair between these sites in the D conformation (19,20). Previous work suggests that mutation g38c is more deleterious than c60g because g38 also coordinates an a33·a66 non-canonical base pair, forming a base quartet platform that further stabilizes the D conformation (20).

To decompose the interplay between ensemble and contact epistasis for the g38c/c60g mutant cycle, we eliminated contact epistasis from ΔG_{obs} by removing epistasis in $\Delta G^{\circ'}_{2\text{AP}}$, $\Delta G^{\circ'}_{\text{MG}}$, and n_{MG} as before. For the wildtype, c60g, and g38c/c60g variants, we used the median value for each parameter from its MCMC distribution (Fig 4A-C). To obtain $\Delta G^{\circ'}_{2\text{AP}}$ and $\Delta G^{\circ'}_{\text{MG}}$ for g38c, we subtracted the effect of c60g on each parameter in the wildtype background from the values for the g38c/c60g double mutant (green points, Fig 4A-B). This removal of contact epistasis led to the solid green curve in Fig 4G, which shows that ensemble redistribution alone contributes up to -0.5 kcal/mol of epistasis at low Mg^{2+} concentrations (labeled arrow, Fig 4G). Contact epistasis further amplifies this effect to yield the maximum epistasis of -1.5 kcal/mol (labeled arrow, Fig 4G). This contact epistasis is driven primarily by epistasis in Mg^{2+} affinity because

eliminating epistasis in $\Delta G^{\circ'}_{\text{MG}}$ gives a value entirely outside the MCMC samples for g38c (green point; Fig 4B).

Mutant Cycle #3: c50a/c60g

We next turned to analyzing epistasis in the c50a/c60g mutant cycle. We expected this mutant cycle to yield minimal contact epistasis but significant ensemble epistasis between allosteric sites. These two positions are distant from one another but connect two allosterically linked regions of the protein: the docked helices (c60g) and the adenine binding site (c50a).

When introduced alone, we found that c50a had a deleterious effect of 1.1 [0.9,1.3] kcal/mol on 2AP binding (Fig 5A), but no detectable effect on Mg^{2+} binding (Fig 5B,C). This matches previous observations and makes sense given its proximity to the adenine binding pocket (20). When c50a is combined with c60g, the riboswitch becomes almost unresponsive to Mg^{2+} as measured by 2AP binding. Even at the highest RNA and Mg^{2+} concentrations we studied, we saw only small shifts in 2AP binding (Fig 2B). As a result, the values of $\Delta G^{\circ'}_{\text{2AP}}$ and $\Delta G^{\circ'}_{\text{MG}}$ for the double mutant are poorly defined.

To investigate contact epistasis for the c50a/c60g mutant cycle we looked at the joint distribution of $\Delta G^{\circ'}_{\text{2AP}}$ and $\Delta G^{\circ'}_{\text{MG}}$ for the double mutant (Fig 5D). Unlike the previous mutant cycle, we saw no covariation in our estimates of $\Delta G^{\circ'}_{\text{2AP}}$ and $\Delta G^{\circ'}_{\text{MG}}$: $\Delta G^{\circ'}_{\text{MG}}$ is essentially unconstrained by the data. There is some evidence for contact epistasis, even for these poorly defined parameters, as only 1.1% of MCMC samples had $\Delta G^{\circ'}_{\text{2AP}}$ and $\Delta G^{\circ'}_{\text{MG}}$ that both had epistasis less than 1 kcal/mol. This result is tentative, however, given the poor constraints on parameters given by the experimental data for c50a/g60c.

To look for ensemble epistasis, we next calculated ΔG_{obs} vs. Mg^{2+} concentration for the wildtype, c50a, and c60g variants (Fig 5F). Because the value of $\Delta G^{\circ'}_{\text{MG}}$ was unconstrained for the c50a/c60g variant, we could not use Equation 7 to estimate ΔG_{obs} . We therefore turned to Equation 8, which allows us to estimate ΔG_{obs} at each Mg^{2+} concentration directly from the 2AP binding curves (Fig 2B). ΔG_{obs} for c50a/c60g estimated using Equation 8 is shown as points in Fig 5F. We observed no apparent change in ΔG_{obs} for c50a/c60g as a function of Mg^{2+} concentration, reflecting its unresponsiveness to Mg^{2+} ions. To calculate epistasis in ΔG_{obs} using Equation 10, we used the average value of ΔG_{obs} for c50a/c60g across all measured Mg^{2+} concentrations (3.8 ± 0.5 kcal/mol; solid purple line, Fig 5F).

We observed magnesium-dependent epistasis in ΔG_{obs} for the c50a/c60g mutant cycle (Fig 5G): the value starts negative, then saturates just above 1 mM Mg^{2+} . At low concentrations, the epistasis arises from the difference in the responsiveness of the variants to Mg^{2+} . At high Mg^{2+} concentrations, the flat epistasis curve arises because the four variants reach different final populations of D·A (Fig S3). The epistasis is positive because the double mutant is worse at binding 2AP than would be expected based on the individual effects of the mutations.

We next investigated the interplay between contact and ensemble epistasis in the c50a/c60g mutant cycle. As before, we calculated epistasis in ΔG_{obs} after removing epistasis in $\Delta G^{\circ'}_{2\text{AP}}$, $\Delta G^{\circ'}_{\text{MG}}$ and n_{MG} for all variants. For this analysis, we used the median values of our estimates for wildtype, c50a, and c60g variants and the green points for c50a/c60g (Fig 5A-C). The resulting ensemble epistasis in ΔG_{obs} is shown as a green line in Fig 5G. For this mutant cycle, the ensemble contributes a maximum of 0.7 kcal/mol of epistasis at Mg^{2+} concentrations below 1 mM (labeled arrow, Fig 5G). At higher Mg^{2+} concentrations, contact epistasis dominates (labeled arrow, Fig

5G). Given our uncertainty in the thermodynamic parameters, we cannot resolve what contact(s) and corresponding magnitudes lead to the observed epistasis.

General patterns of concentration-dependence

We next asked if there were general rules that might allow one to make hypotheses about the source of epistasis given an experimental observable, even without detailed information about the thermodynamic ensemble under study. The riboswitch ensemble provides a useful framework for posing this question. We can rewrite our energetic observable as follows, pulling out the 2AP binding constant:

$$\Delta G_{obs} = -RT \ln \left(\frac{(K_{MG}[M]_T)^{n_{MG}}}{1 + (K_{MG}[M]_T)^{n_{MG}}} \right) - RT \ln(K_{2AP}). \quad 11.$$

When we do so, we see that ΔG_{obs} reduces to a Hill binding model (left term) with a Mg^{2+} -independent offset controlled by 2AP affinity (right term). Such a model is generic and applicable to many systems. We can therefore think about the interplay of contact and ensemble epistasis in the adenine riboswitch as a model for many macromolecules of interest. What patterns arise if we introduce contact epistasis in K_{2AP} (target affinity), K_{MG} (allosteric effector affinity) and n_{MG} (allosteric effector cooperativity)?

In Figure 6, we calculated epistasis in ΔG_{obs} for otherwise non-epistatic mutant cycles in which we injected varying amounts of epistasis into these equilibrium constants. We observe that each equilibrium constant gives a distinct effector-dependent epistatic signal. First, epistasis in 2AP affinity has no concentration dependence: it acts as a global offset to ΔG_{obs} (Fig 6A). We observed this in our experimental data: epistasis in ΔG°_{2AP} led to a global offset (Fig 3E). The affinity for 2AP is independent of the population of the 2AP binding-competent conformation and thus does not depend on Mg^{2+} concentration.

Contact epistasis in Mg^{2+} affinity gives more complex patterns. In the absence of contact epistasis in K_{MG} , we observe a peak in epistasis (black curve, Fig 6B; Fig 1C). The peak arises from the redistribution of conformations within the ensemble: there is no contact epistasis anywhere in the system. As we add contact epistasis, it modulates the signal for ensemble epistasis in a nonlinear fashion: the contact epistasis itself redistributes the ensemble, amplifying the ensemble epistasis. The largest effects of epistasis in K_{MG} occur at low Mg^{2+} concentrations where a change in Mg^{2+} binding can alter the relative population of the D·A conformation, and thus ΔG_{obs} (Fig S3). This epistasis is lost at high Mg^{2+} because the molecule is saturated with Mg^{2+} and 2AP binding now controls ΔG_{obs} (Fig S3). We observed this pattern for the g38c/c60g mutant cycle, where strong contact epistasis in $\Delta G^{\circ'}_{MG}$ amplified ensemble epistasis at low and moderate Mg^{2+} concentrations, then dropped to zero at high Mg^{2+} concentrations (Fig 4G).

Finally, epistasis in ion uptake leads to a different pattern of magnesium-dependent epistasis. As with epistasis in K_{MG} , contact epistasis in n_{MG} plays a role at low Mg^{2+} concentration, then goes to zero when other factors outside Mg^{2+} binding are the most important determinants of the energetics of the system. Because the ion uptake coefficient controls the sensitivity to Mg^{2+} —mathematically, it is equivalent to a Hill coefficient (31)—contact epistasis in n_{MG} manifests as a linear dependence between the log of Mg^{2+} concentration and epistasis. We saw no evidence for this particular form of epistasis in our riboswitch mutant cycles.

This rogues' gallery of epistatic effects provides a useful lens to generate hypotheses about the origins of epistasis in a macromolecule, even without a detailed model of the thermodynamic ensemble for that system. If one were to observe a fixed epistatic signal across multiple concentrations of an allosteric effector, it would suggest contact epistasis on some feature of the macromolecule that does not depend on the effector concentration (Fig 6A). If one observed a peak

in epistasis that dropped to zero at both high and low effector concentrations—concentration regimes in which few conformations are populated—the pattern would suggest ensemble epistasis and minimal contact epistasis (Fig 6B). If one observed high epistasis at low effector concentration that dropped to zero at high effector concentration, it would suggest contact epistasis in the parameters controlling effector binding (Fig 6B). And, finally, if one saw a monotonic change in epistasis with the logarithm of effector concentration, it would suggest contact epistasis in binding cooperativity (Fig 6C).

These rules of thumb are not definitive and require system-specific validation (to say nothing of the complexity that could arise if multiple classes of contact epistasis contribute to the observed signal). However, they do provide a useful starting point. One might imagine a high-throughput experiment conducted at a few allosteric effector concentrations. A simple analysis of the effector-dependent epistasis in the system would provide initial hints about the energetic couplings within the system. This could set up deeper biophysical investigation, possibly leading to fitting a thermodynamic model to the system (as demonstrated in this work) to extract the relevant thermodynamic parameters and epistatic mechanisms (16, 17).

Discussion

We set out to investigate the relative contributions of contact and ensemble epistasis between mutations in the adenine riboswitch. We found evidence for both forms of epistasis in all three mutant cycles. Our work reveals that ensemble epistasis—previously measured in proteins (17)—also occurs in RNA and is likely a generic feature of allosteric macromolecules. Further, we found that contact and ensemble epistasis interact in a nonlinear fashion. Even when the magnitude of ensemble epistasis was low (~ 0.5 kcal/mol) in one mutant cycle, the ensemble could strongly

attenuate the magnitude of contact epistasis (from ~ 6 kcal/mol to ~ 1.5 kcal/mol, Fig 4E,G). Thus, understanding intramolecular epistasis in this molecule requires accounting for both contact and ensemble mediated effects.

Interplay between ensemble and contact epistasis

In our previous theoretical and experimental work, we observed that mutations led to high-magnitude epistasis in thermodynamic observables defined by the ensemble (14, 17), even without any epistasis between mutations at the level of equilibrium constants. This was *not* the case for adenine riboswitch. If we removed epistasis from equilibrium constants ($\Delta G^{\circ'}_{2AP}$, $\Delta G^{\circ'}_{MG}$, and n_{MG}), ensemble epistasis in ΔG_{obs} hovered around kT (0.6 kcal/mol). This can be seen in the green curves shown in Fig 3E, 4G, and 5G, which have peak magnitudes of 0.3 (a35u/c60g), -0.5 (g38c/c60g), and 0.7 (c50a/c60g) kcal/mol, respectively.

When we included contact epistasis, however, we observed much larger concentration-dependent epistasis in ΔG_{obs} for two of the three cycles: g38c/c60g (-1.5 kcal/mol), and c50a/c60g (1.5 kcal/mol). Concentration-dependent epistasis folds together contact and ensemble mechanisms in a nonlinear fashion. This is seen most clearly for g38c/c60g. This cycle had strong contact epistasis because it inverted the polarity of a Watson-Crick-Franklin G-C base pair. Contact epistasis in either $\Delta G^{\circ'}_{2AP}$ or $\Delta G^{\circ'}_{MG}$ is necessary to explain the observed binding curves (Fig 4B, Fig 4C). Although we could not unambiguously determine which parameter exhibited epistasis due to strong covariance, the median estimate of contact epistasis was -6 kcal/mol in $\Delta G^{\circ'}_{MG}$ and 1 kcal/mol in $\Delta G^{\circ'}_{2AP}$ (red point, Fig 4E). This result is plausible, as the above value of $\Delta G^{\circ'}_{MG}$ is consistent with a previous measurement of the g38/c60 interaction energy (19), as well as the expected energy for a G-C base pair (37). Further, we would anticipate that epistasis

from an inverted G-C base pair in the docked-loop conformations would manifest largely in $\Delta G^{\circ'}_{\text{MG}}$ rather than indirectly in $\Delta G^{\circ'}_{\text{2AP}}$ (Fig 1B).

Despite this strong contact epistasis (-6 kcal/mol) in the g38c/c60g mutant cycle, we observed epistasis of only -1.5 kcal/mol in our ensemble observable (ΔG_{obs}). The riboswitch's conformational ensemble attenuates the strong contact epistasis, making epistasis in the observable much lower than the epistasis in the underlying equilibrium constant. We note that even MCMC samples with contact epistasis up to -45 kcal/mol in $\Delta G^{\circ'}_{\text{MG}}$ and 10 kcal/mol in $\Delta G^{\circ'}_{\text{2AP}}$ (distribution, Fig 4E) were accommodated with only ~1.5 kcal/mol epistasis in ΔG_{obs} (gray area, Fig 4G). Although these massive energetic effects on $\Delta G^{\circ'}_{\text{MG}}$ and $\Delta G^{\circ'}_{\text{2AP}}$ are chemically implausible, they illustrate an important point. Epistasis in ΔG_{obs} arises due to shifts in the relative population of ensemble conformations—E, D, and D·A—rather than directly from the energetic effects of mutations on each conformation. Massive effects on the energy of each conformation due to physical contacts can be subsumed by the ensemble if these effects do not meaningfully redistribute the probabilities of the conformations.

The importance of multistate design

Our work points to the fundamental importance of conformational landscapes in modulating the effects of mutations and epistasis between them. Efforts to engineer or design macromolecules by focusing on individual conformations may face unexpected results if the macromolecule populates multiple conformations in solution.

As a hypothetical example, consider an attempt to optimize the adenine riboswitch's ligand-bound conformation for stronger ligand binding affinity. Perhaps a pair of mutations are introduced to form another Watson-Crick-Franklin base pair to bridge the docked-loop interface.

Structure-based calculations on the ligand-bound conformation might predict a certain favorable energetic interaction between the two mutations. They might incorrectly calculate the energetic effect on ligand binding, however, by failing to examine how the new base pair reshapes the conformational landscape. These considerations point to the value of multi-state design to better account for the interplay between contact and ensemble sources of epistasis (38, 39), especially since many biological macromolecules must toggle between multiple conformational states under physiological conditions for function.

Implications for allostery

This work also provides an intriguing lens for attempting to understand the molecular basis of allostery. Epistasis and allostery between sites have similar energetic origins: introduction of a new stabilizing or destabilizing interaction at one site alters the activity at another. Despite this basic equivalence, allostery and epistasis differ in their concentration-dependence. Mutations are a fixed perturbation, independent of environment. In contrast, allosteric interactions depend on the concentrations of allosteric effectors. By combining measurements of epistasis (fixed perturbation to the probability of a conformation) with an allosteric perturbation (addition of allosteric effector) we gain insight into the thermodynamic basis for communication between sites in a macromolecule. In particular, we can separate epistatic effects into “contact” and “ensemble” categories, which may allow us to disentangle communication occurring *through* a molecule (13) from communication occurring through a redistribution of conformational states (15).

We see a hint of long-distance communication through a conformation rather than the ensemble for c50a/c60g. For this mutant cycle, the vast majority of MCMC samples exhibited epistasis in $\Delta G^{\circ'}_{\text{MG}}$ (Fig 5E), even though these two mutations are ~ 34 Å apart in the structure.

One way to view this result is that communication is occurring via one of the conformations in the ensemble, rather than through redistributing the ensemble. The thermodynamic parameters for this cycle were poorly determined, so this observation remains tentative. We bring it up here, however, as an example of a promising way to disentangle through-conformation from ensemble-based mechanisms of long-distance communication in macromolecules: look for evidence of epistasis between distant mutations in the microscopic equilibria that define the ensemble (e.g., $\Delta G^{\circ'}_{\text{MG}}$) versus epistasis in the overall thermodynamic observable for the system (e.g., ΔG_{obs}).

Identifying sources of epistasis

Finally, our work suggests a few “rules of thumb” that researchers could use to make hypotheses about the roles of contact and ensemble mechanisms in a macromolecule of interest, even given limited measurements. This is particularly apropos for high-throughput studies, where one might be able to measure binding, or some other thermodynamic observable, as a function of allosteric effector. One might imagine using these rules to classify sites for further study, or to help develop a more rigorous mechanistic model of an ensemble for further characterization.

We observed a few basic patterns in our experimental (Figs 3-5) and model studies (Fig 6). First, contact epistasis in the observed conformation (D·A, for the riboswitch) gives rise to epistasis independent of effector concentration (Fig 6A). Second, ensemble epistasis with no contact epistasis yields a peak in epistasis centered at the allosteric effector concentration that maximizes ensemble diversity (black line, Fig 6B) (14). Third, a convolution of contact and ensemble epistasis appears as a peak in epistasis, but with non-zero epistasis at high or low effector concentration (colored lines, Fig 6B). Fourth, contact epistasis in cooperativity (i.e., the ion uptake coefficient or a Hill coefficient) appears as epistasis that changes linearly as with the log of the

allosteric effector concentration (Fig 6C). Intriguingly, we also observed very similar patterns of epistasis versus allosteric effector in our previous study of ensemble epistasis in the lac repressor (17).

Further work will be needed to generalize these results, with a particular focus on understanding how these patterns would change if several classes of contact epistasis were in play. It will also be important to explore different sorts of ensembles, with different numbers of conformations and linkages between allosteric effector and observable. This said, given the generic form of the mathematical model of the riboswitch ensemble (Equation 11), we expect these findings to generalize well. Overall, our work highlights the importance of considering effects of mutations on multiple conformations to understand epistasis and suggests a few rules-of-thumb for disentangling contact and ensemble epistasis in macromolecules.

Author contributions:

DRW, JRW, and MJH designed the research; DRW performed the research; DRW and MJH developed software and analyzed the data; DRW wrote the manuscript; DRW, JRW, and MJH edited the manuscript.

Declaration of interests:

The authors declare no competing interests.

Acknowledgments:

We would like to thank members of the Harms lab, past and present, for their invaluable input. This work was funded by a National Science Foundation CAREER Award DEB-1844963 (M.J.H.).

Bibliography

1. Fisher, R. 1918. The correlation between relatives on the supposition of mendelian inheritance. *Trans. Roy. Soc.* .
2. Phillips, P.C. 2008. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet.* 9:855–867.
3. Hopf, T.A., A.G. Green, B. Schubert, S. Mersmann, C.P.I. Schärfe, J.B. Ingraham, A. Toth-Petroczy, K. Brock, A.J. Riesselman, P. Palmedo, C. Kang, R. Sheridan, E.J. Draizen, C. Dallago, C. Sander, and D.S. Marks. 2019. The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics.* 35:1582–1584.
4. Ortlund, E.A., J.T. Bridgeman, M.R. Redinbo, and J.W. Thornton. 2007. Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis. *Science.* 317:1544–1548.
5. Bloom, J.D., L.I. Gong, and D. Baltimore. 2010. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. *Science.* 328:1272–1275.
6. Field, S.F., and M.V. Matz. 2010. Retracing Evolution of Red Fluorescence in GFP-Like Proteins from Faviina Corals. *Molecular Biology and Evolution.* 27:225–233.
7. Harman, J.L., P.N. Reardon, S.M. Costello, G.D. Warren, S.R. Phillips, P.J. Connor, S. Marqusee, and M.J. Harms. 2022. Evolution avoids a pathological stabilizing interaction in the immune protein S100A9. *Proceedings of the National Academy of Sciences.* 119:e2208029119.
8. Miton, C.M., and N. Tokuriki. 2016. How Mutational Epistasis Impairs Predictability in Protein Evolution and Design. *Protein Science.* 25:1260–1272.
9. Luo, Y., G. Jiang, T. Yu, Y. Liu, L. Vo, H. Ding, Y. Su, W.W. Qian, H. Zhao, and J. Peng. 2021. ECNet is an evolutionary context-integrated deep learning framework for protein engineering. *Nat Commun.* 12:5743.
10. Siedhoff, N.E., A.-M. Illig, U. Schwaneberg, and M.D. Davari. 2021. PyPEF—An Integrated Framework for Data-Driven Protein Engineering. *J. Chem. Inf. Model.* 61:3463–3476.
11. Storz, J.F. 2018. Compensatory mutations and epistasis for protein function. *Current Opinion in Structural Biology.* 50:18–25.
12. Hopf, T.A., C.P.I. Schärfe, J.P.G.L.M. Rodrigues, A.G. Green, O. Kohlbacher, C. Sander, A.M.J.J. Bonvin, and D.S. Marks. 2014. Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife.* 3:e03430.
13. Ribeiro, A.A.S.T., and V. Ortiz. 2016. A Chemical Perspective on Allostery. *Chem. Rev.* 116:6488–6502.

14. Morrison, A.J., D.R. Wonderlick, and M.J. Harms. 2021. Ensemble epistasis: thermodynamic origins of nonadditivity between mutations. *Genetics*.
15. Motlagh, H.N., J.O. Wrabl, J. Li, and V.J. Hilser. 2014. The ensemble nature of allostery. *Nature*. 508:331–339.
16. Otwinowski, J. 2018. Biophysical Inference of Epistasis and the Effects of Mutations on Protein Stability and Function. *Molecular Biology and Evolution*. 35:2345–2354.
17. Morrison, A.J., and M.J. Harms. 2022. An experimental demonstration of ensemble epistasis in the lac repressor. *bioRxiv*. 2022.10.14.512271.
18. Mandal, M., and R.R. Breaker. 2004. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol*. 5:451–463.
19. Lemay, J.-F., J.C. Penedo, R. Tremblay, D.M.J. Lilley, and D.A. Lafontaine. 2006. Folding of the Adenine Riboswitch. *Chemistry & Biology*. 13:857–868.
20. Gilbert, S.D., C.E. Love, A.L. Edwards, and R.T. Batey. 2007. Mutational Analysis of the Purine Riboswitch Aptamer Domain [†]. *Biochemistry*. 46:13297–13309.
21. Stoddard, C.D., S.D. Gilbert, and R.T. Batey. 2008. Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA*. 14:675–684.
22. Neupane, K., H. Yu, D.A.N. Foster, F. Wang, and M.T. Woodside. 2011. Single-molecule force spectroscopy of the ad adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Research*. 39:7677–7687.
23. Leiply, D., and D.E. Draper. 2011. Effects of Mg ²⁺ on the Free Energy Landscape for Folding a Purine Riboswitch RNA. *Biochemistry*. 50:2790–2799.
24. Reining, A., S. Nozinovic, K. Schlepckow, F. Buhr, B. Fürtig, and H. Schwalbe. 2013. Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature*. 499:355–359.
25. Stoddard, C.D., J. Widmann, J.J. Trausch, J.G. Marcano-Velázquez, R. Knight, and R.T. Batey. 2013. Nucleotides Adjacent to the Ligand-Binding Pocket are Linked to Activity Tuning in the Purine Riboswitch. *Journal of Molecular Biology*. 425:1596–1611.
26. Foreman-Mackey, D., D.W. Hogg, D. Lang, and J. Goodman. 2013. emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*. 125:306.
27. Harris, C.R., K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T.E. Oliphant. 2020. Array programming with NumPy. *Nature*. 585:357–362.

28. team, T. pandas development. 2022. *pandas-dev/pandas: Pandas*. .

29. Virtanen, P., R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, and P. van Mulbregt. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 17:261–272.

30. Rieder, R., K. Lang, D. Graber, and R. Micura. 2007. Ligand-Induced Folding of the Adenosine Deaminase A-Riboswitch and Implications on Riboswitch Translational Control. *ChemBioChem*. 8:896–902.

31. Leiply, D., and D.E. Draper. 2010. The dependence of RNA tertiary structure stability on Mg²⁺ concentration: interpretation of the Hill equation and coefficient. *Biochemistry*. 49:1843–1853.

32. Noeske, J., H. Schwalbe, and J. Wohner. 2007. Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain. *Nucleic Acids Research*. 35:5262–5273.

33. Tyrrell, J., J.L. McGinnis, K.M. Weeks, and G.J. Pielak. 2013. The Cellular Environment Stabilizes Adenine Riboswitch RNA Structure. *Biochemistry*. 52:8777–8785.

34. Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 19:716–723.

35. Stagno, J.R., Y. Liu, Y.R. Bhandari, C.E. Conrad, S. Panja, M. Swain, L. Fan, G. Nelson, C. Li, D.R. Wendel, T.A. White, J.D. Coe, M.O. Wiedorn, J. Knoska, D. Oberthuer, R.A. Tuckey, P. Yu, M. Dyba, S.G. Tarasov, U. Weierstall, T.D. Grant, C.D. Schwieters, J. Zhang, A.R. Ferré-D’Amaré, P. Fromme, D.E. Draper, M. Liang, M.S. Hunter, S. Boutet, K. Tan, X. Zuo, X. Ji, A. Barty, N.A. Zatsepin, H.N. Chapman, J.C.H. Spence, S.A. Woodson, and Y.-X. Wang. 2017. Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature*. 541:242–246.

36. Serganov, A., Y.-R. Yuan, O. Pikoškaya, A. Polonskaya, L. Malinina, A.T. Phan, C. Hobartner, R. Micura, R.R. Breaker, and D.J. Patel. 2004. Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. *Chemistry & Biology*. 11:1729–1741.

37. Vendeix, F.A.P., A.M. Munoz, and P.F. Agris. 2009. Free energy calculation of modified base-pair formation in explicit solvent: A predictive model. *RNA*. 15:2278–2287.

38. Davey, J.A., and R.A. Chica. 2012. Multistate approaches in computational protein design. *Protein Science*. 21:1241–1252.

39. Löffler, P., S. Schmitz, E. Hupfeld, R. Sterner, and R. Merkl. 2017. Rosetta:MSF: a modular framework for multi-state computational protein design. *PLOS Computational Biology*. 13:e1005600.

FIGURES

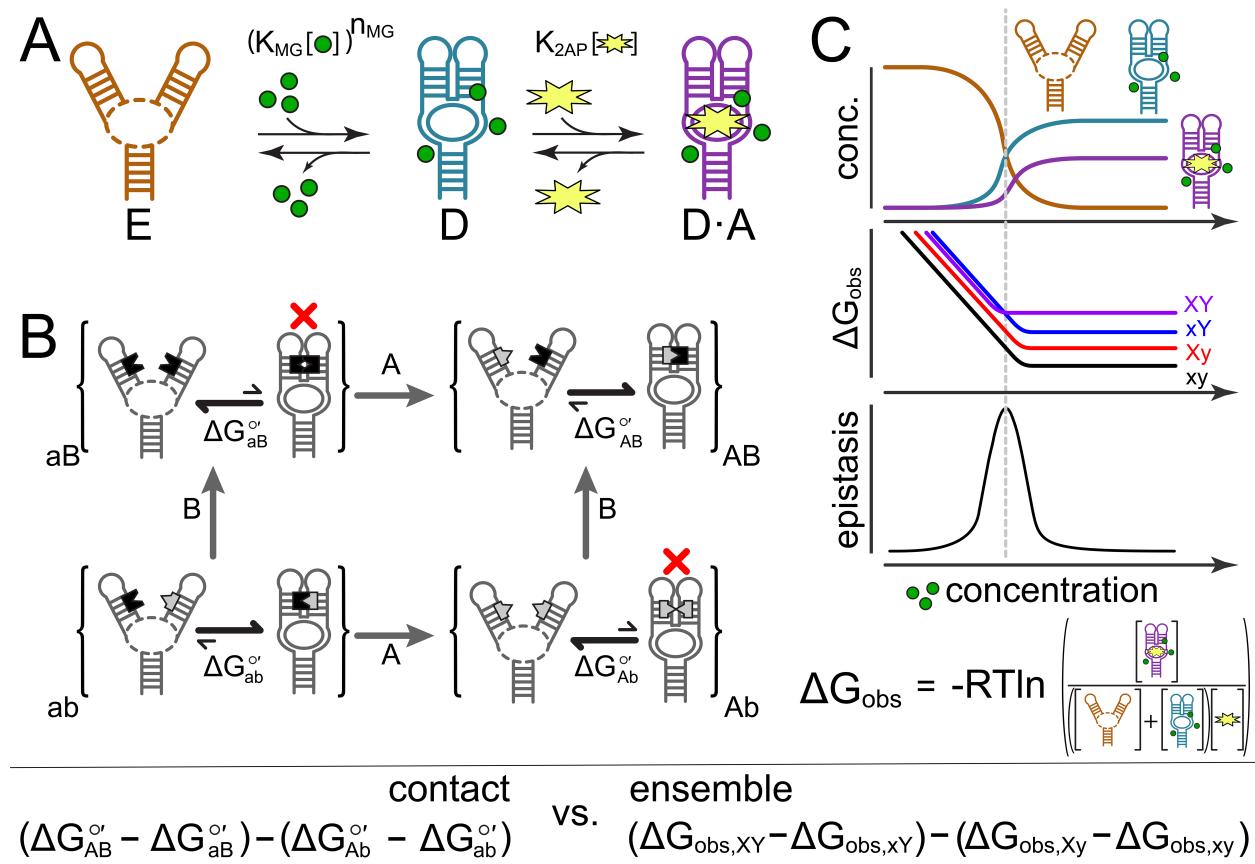


Fig 1. Differentiating contact and ensemble epistasis in the adenine riboswitch. A) A thermodynamic model of the riboswitch ensemble. The molecule can be extended (E), docked (D), or docked and bound to 2AP (D·A). The equilibria are defined by three parameters: K_{MG} , n_{MG} , and K_{2AP} . The equilibria are tuned by the concentration of Mg^{2+} ions (green circles) and 2AP (yellow star). B) Epistasis arising from a contact between two sites. RNA molecules within braces represent an ensemble for a specific genotype (ab, Ab, aB, or AB). The equilibrium between the E and D forms is shown by arrows within the braces and measured by $\Delta G^{\circ'}$. The A and B mutations are indicated by bold gray arrows between ensembles. Bases at these sites can either be indented (black) or protruding (gray). Mutation A flips the left base from indented to protruding; mutation B flips the right base from protruding to indented

B flips the right base from protruding to indented. Mismatches are indicated with a red “ \times ”. C) Epistasis arising from the ensemble. The cartoon graphs show concentrations of E, D, and D·A for wildtype (top); ΔG_{obs} for four genotypes (xy, Xy, xY, and XY) taken from a hypothetical mutant cycle (middle); and epistasis in ΔG_{obs} calculated from the middle plot (bottom). These series are plotted as a function of Mg^{2+} concentration. The mathematical expressions for contact and ensemble epistasis are displayed at the bottom of the figure.

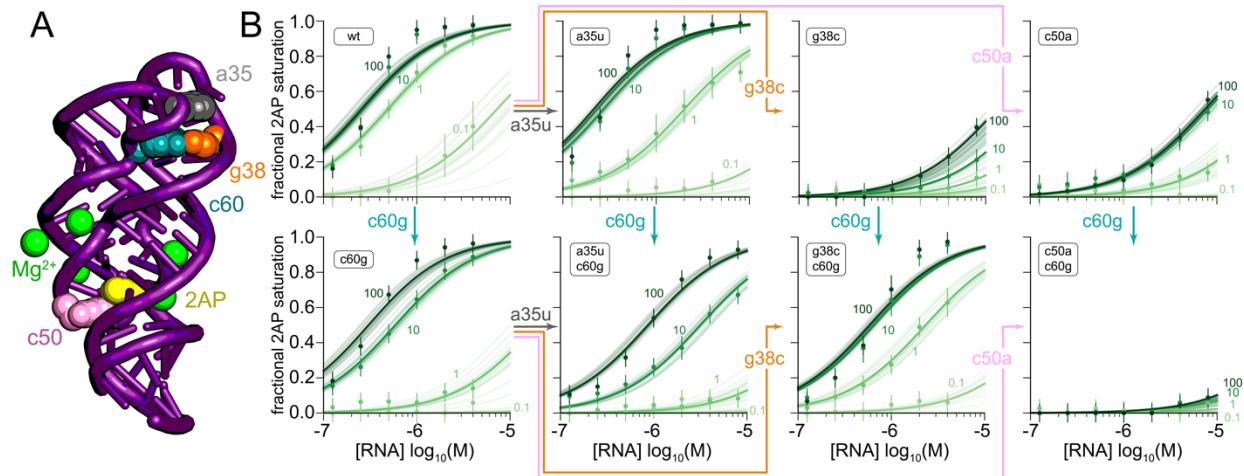


Fig 2. 2AP binding behavior can be captured by a thermodynamic model of the riboswitch ensemble. A) Crystal structure of the adenine riboswitch aptamer domain (PDB ID: 5SWE) (35). The overall architecture is shown as a purple cartoon. Bases mutated in this study are indicated with colored spheres (labeled on structure). The bound adenine base is shown as yellow spheres; probable Mg^{2+} ion locations are shown as green spheres, taken from PDB: 1Y26 (36). B) Each panel shows 2AP binding for the genotype indicated. Columns are (left to right) wildtype, a35u, g38c, and c50a. Rows are without (top) and with (bottom) the c60g mutation. The arrows between panels define three mutant cycles: a35u/c60g, g38c/c60g, and c50a/c60g. Each panel shows the fraction of 2AP bound, measured by relative 2AP fluorescence compared to free and fully quenched controls, as a function of riboswitch concentration. The green shade indicates the total Mg^{2+} concentration, labeled on each plot in millimolar. Each point is the average of at least three experimental replicates, with standard deviations indicated as error bars. Lines were calculated using parameters taken from 50 randomly selected MCMC samples of the three-state/three-parameter model shown in Fig 1A.

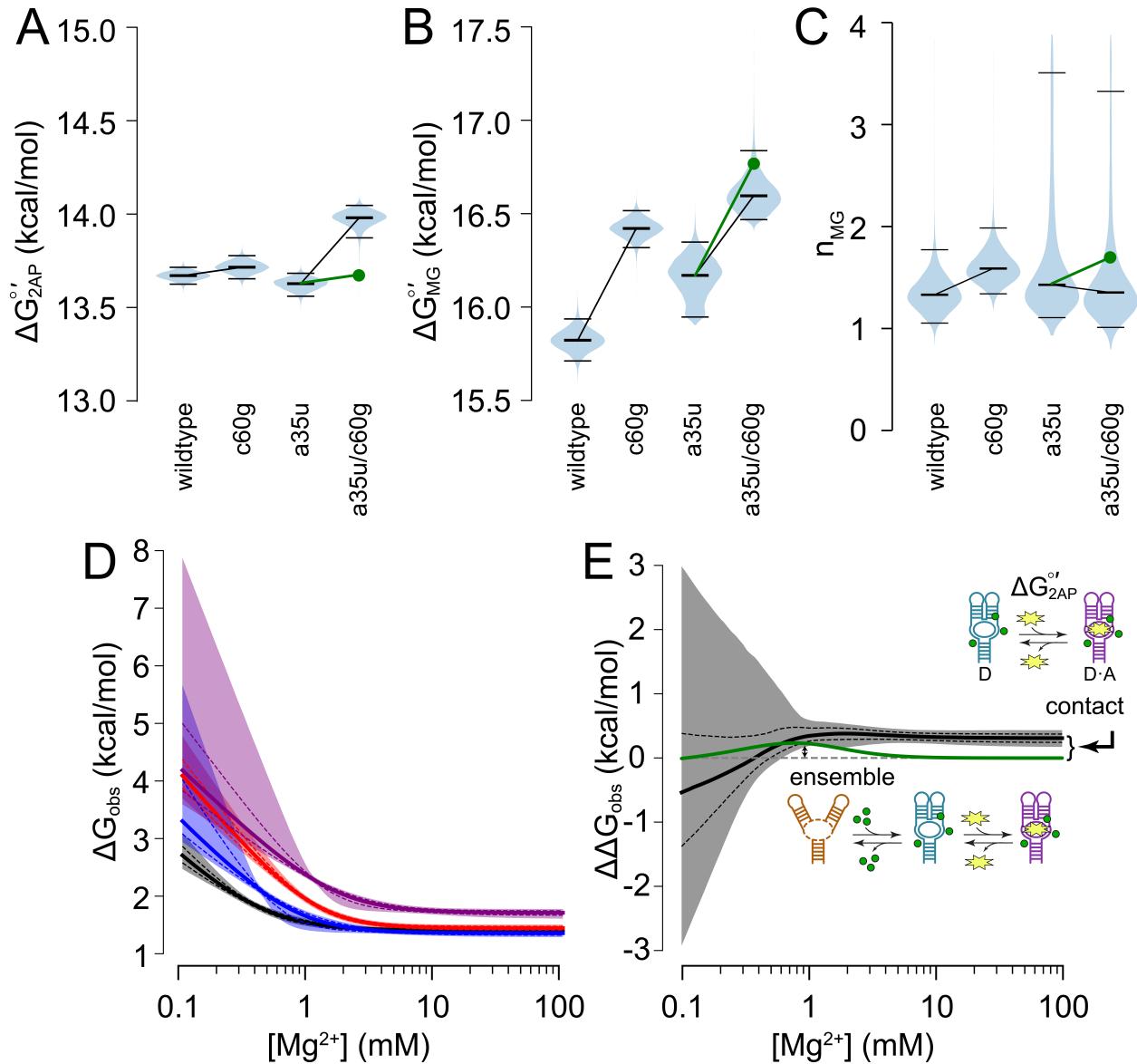


Figure 3. Contact and ensemble epistasis in the a35u/c60g cycle. Panels A-C show marginal distributions of $\Delta G^{\circ'}_{2AP}$, $\Delta G^{\circ'}_{MG}$, and n_{MG} taken from across all MCMC samples. The area of the violin plot encompasses all values seen. The central bar is the median; the top and bottom bars mark the 95% credibility interval. The black lines between series indicate the median effect of the c60g mutation in the wildtype (left) or a35u (right) backgrounds. Epistasis appears as a background-dependent difference in the slope of the line; the green lines and points indicate the expected value if the mutations do not exhibit epistasis. D) ΔG_{obs} versus Mg^{2+} concentration

calculated using Equation 7 for wildtype (black), c60g (red), a35u (blue) and a35u/c60g (purple). Solid lines indicate the median of the MCMC samples; dashed lines indicate the standard deviation; areas indicate the 95% credibility region. E) Epistasis in ΔG_{obs} calculated from the curves in panel D. The green line indicates ΔG_{obs} calculated in the absence of contact epistasis (green points in panels A-C). The small arrows with cartoons indicate the Mg^{2+} concentrations where epistasis arises from the ensemble (1 mM) versus contacts (>10 mM).

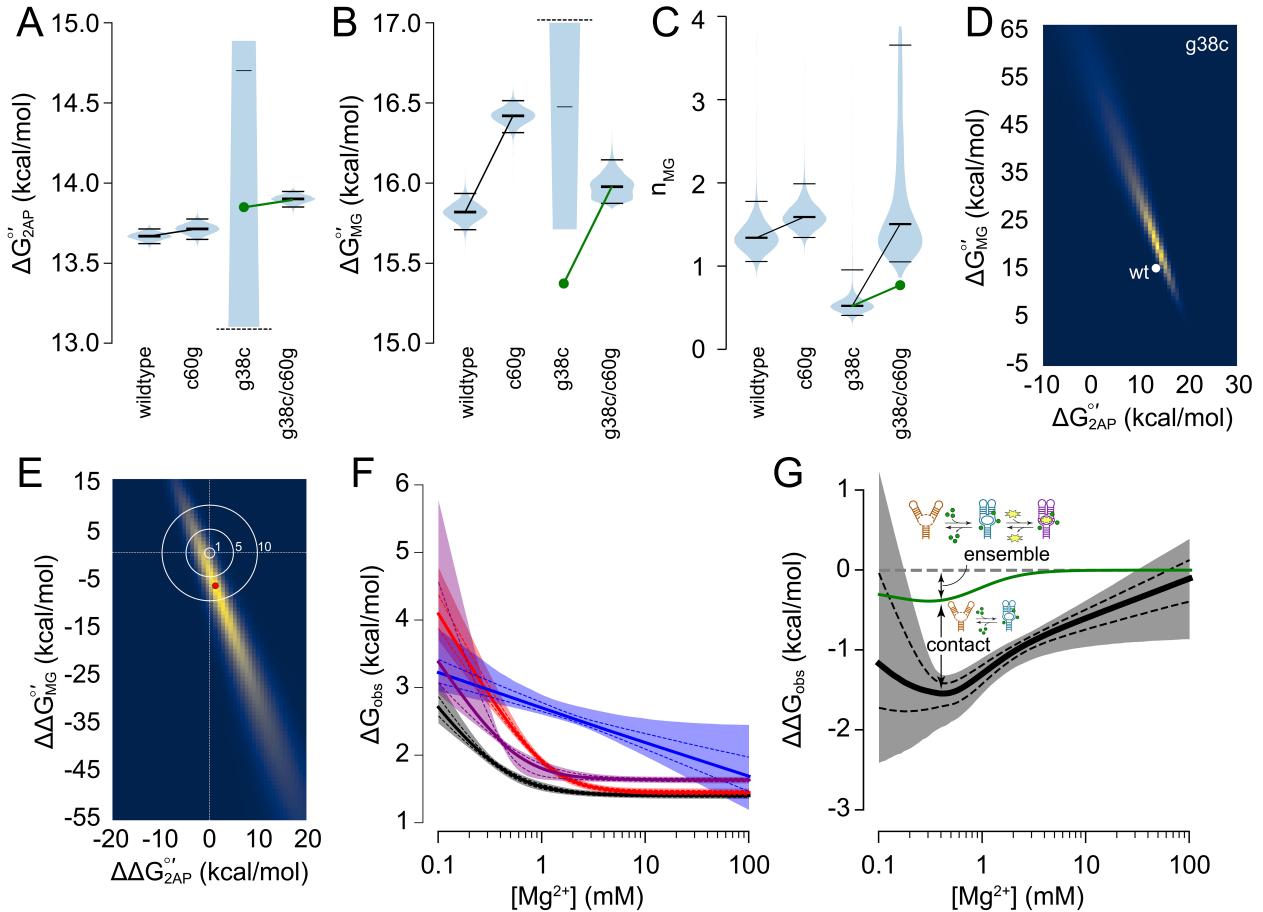


Figure 4. Contact and ensemble epistasis in the g38c/c60g mutant cycle. Panels A-C show marginal distributions of $\Delta G^{\circ'}_{2AP}$, $\Delta G^{\circ'}_{MG}$ and n_{MG} taken across MCMC samples. The area of the violin plot encompasses all values. The central bar is the median; the top and bottom bars mark the 95% credibility interval. The distributions of $\Delta G^{\circ'}_{2AP}$ and $\Delta G^{\circ'}_{MG}$ for the g38c variant are truncated at the dashed lines. The green lines and points indicate the expected value if the mutations exhibit no epistasis. D) Estimates for $\Delta G^{\circ'}_{2AP}$ and $\Delta G^{\circ'}_{MG}$ across MCMC samples for the g38c variant (dark to light encodes low to high frequency). The median wildtype value is shown as a white circle. E) Estimates of epistasis in $\Delta G^{\circ'}_{MG}$ and $\Delta G^{\circ'}_{MG}$ across MCMC samples. The red point indicates the centroid of the distribution. The concentric circles indicate where the total epistasis summed over $\Delta G^{\circ'}_{2AP}$ and $\Delta G^{\circ'}_{MG}$ is 1, 5, and 10 kcal/mol. F) ΔG_{obs} versus Mg^{2+} concentration calculated using Equation 7 for wildtype (black), c60g (red), g38c (blue) and

g38c/c60g (purple). Solid lines indicate the median of the MCMC samples; dashed lines indicate the standard deviation; areas indicate the 95% credibility region. G) Epistasis in ΔG_{obs} calculated from the curves in panel F. The green line indicates ΔG_{obs} in the absence of contact epistasis (green points in panels A-C). The small arrows indicate the maximal contributions of ensemble and contact epistasis near 1 mM Mg²⁺.

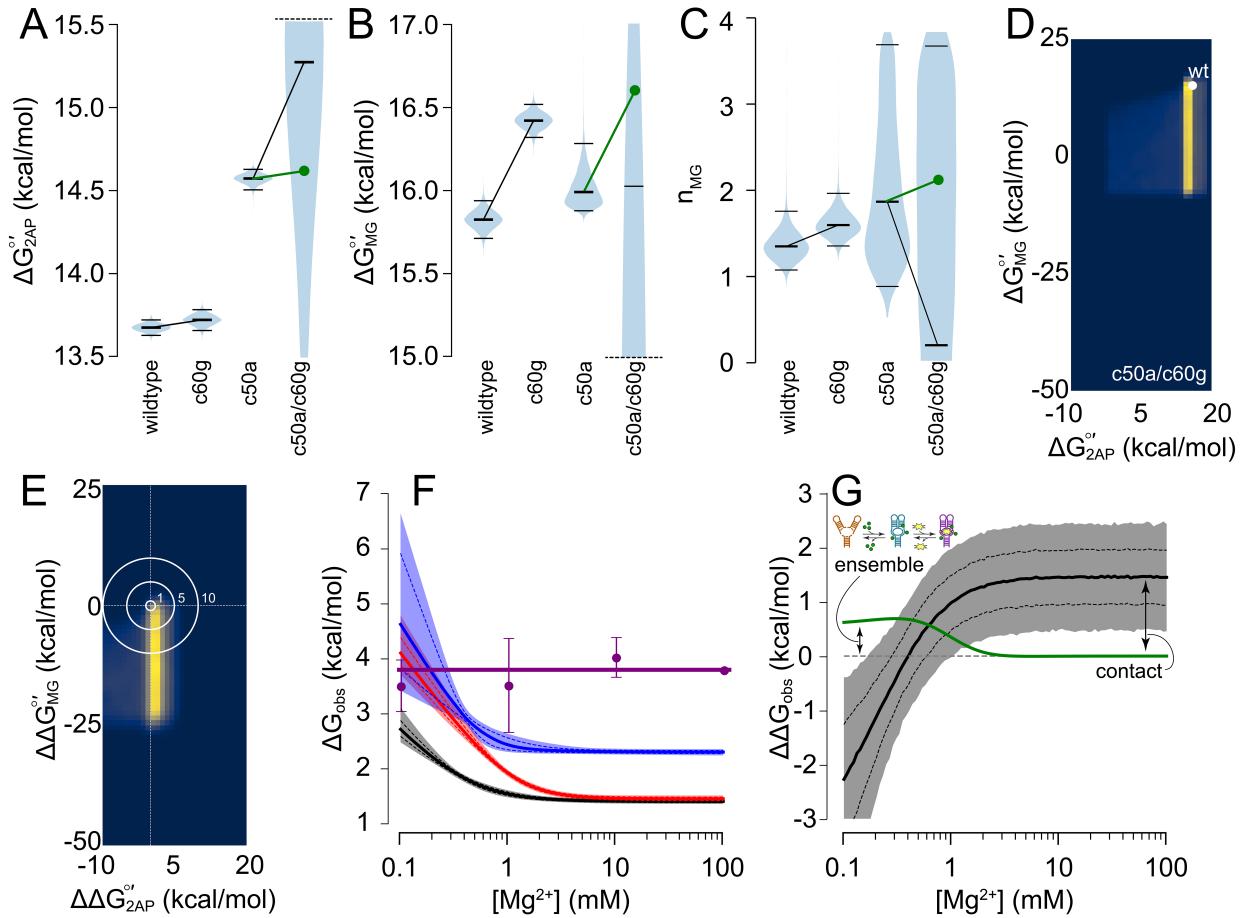


Fig 5. Contact and ensemble epistasis in the c50a/c60g mutant cycle. Panels A-C show marginal distributions of $\Delta G_{2AP}^{\circ'}$, $\Delta G_{MG}^{\circ'}$ and n_{MG} taken from across MCMC samples. The area of the violin plot encompasses all values. The central bar is the median; the top and bottom bars mark the 95% credibility interval. The distributions of $\Delta G_{2AP}^{\circ'}$ and $\Delta G_{MG}^{\circ'}$ for the c50a/c60g variant are truncated at the dashed lines. The green lines and points indicate the expected value if the mutations exhibit no epistasis. D) Estimates of $\Delta G_{2AP}^{\circ'}$ and $\Delta G_{MG}^{\circ'}$ across MCMC samples for the c50a/c60g variant (dark to light encodes low to high frequency). The median wildtype value is shown as a white circle. E) Estimates of epistasis in $\Delta G_{2AP}^{\circ'}$ and $\Delta G_{MG}^{\circ'}$ across MCMC samples. The concentric circles indicate combined epistasis between the two parameters of 1, 5, and 10 kcal/mol. F) ΔG_{obs} versus Mg^{2+} concentration calculated using Equation 7 for wildtype (black), c60g (red), and c50a (blue). Solid lines indicate the median of the MCMC samples; dashed lines represent the 95% credibility interval. G) $\Delta \Delta G_{\text{obs}}$ versus Mg^{2+} concentration. It includes an ensemble model (colored dots) and a contact model (green line with shaded error). A schematic of the ensemble model is shown in the top left.

indicate the standard deviation; areas indicate the 95% credibility region. The purple points indicate the mean and standard deviation of ΔG_{obs} for c50a/c60g calculated from the binding curves in Fig 2 using Equation 8. G) Epistasis in ΔG_{obs} calculated from the curves in panel F. The green line indicates ΔG_{obs} calculated assuming no contact epistasis (green points in panels A-C). The small arrows indicate the Mg^{2+} concentrations where the epistasis may arise from the ensemble versus contacts.

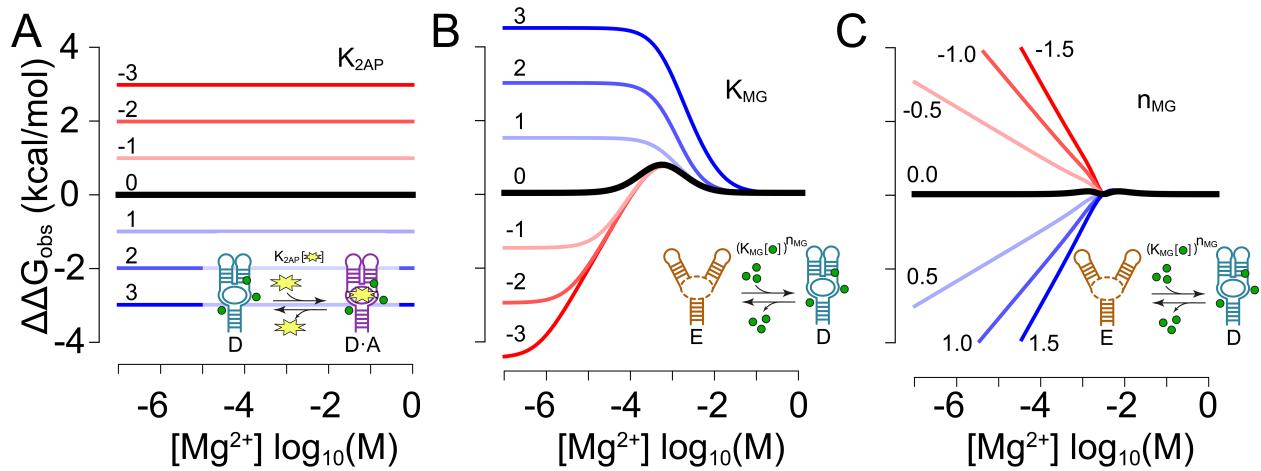


Fig 6. Contact epistasis alters epistasis in ΔG_{obs} . Panels A-C show calculated effects of contact epistasis in ΔG_{obs} vs. Mg^{2+} concentration. The black curves show epistasis in ΔG_{obs} for two mutations that have no contact epistasis. Mutations were modeled as having an effect of +1 kcal/mol (A and B) or +0.5 ions (C). The colored curves are epistasis in ΔG_{obs} given the positive (blue) or negative (red) contact epistasis indicated next to each curve. Cartoon insets indicate which equilibrium is being perturbed.

Table S1: AIC test comparing models of differing complexity.

Model (# states, # parameters)	AIC value	AIC probability
3.3	-1033.26	0.616
3.4	-1031.26	0.226
4.4	-1030.54	0.158
4.5	-996.13	8.654×10^{-9}
2.3	-630.57	3.615×10^{-88}

Table S2: Characteristics of marginal MCMC distributions for 3-state, 3-parameter model.

Values shown are the median, 2.5% (bottom) and 97.5% (top) values from each distribution.

genotype	parameter	median	bottom	top
wt	$\ln(K_{2AP})$	-2.4	-2.4	-2.3
wt	$\ln(K_{MG})$	-6.0	-6.2	-5.8
wt	n_{mg}	1.5	1.1	2.1
a35u	$\ln(K_{2AP})$	-2.3	-2.4	-2.2
a35u	$\ln(K_{MG})$	-6.6	-6.9	-6.2
a35u	n_{mg}	1.6	1.2	4.5
g38c	$\ln(K_{2AP})$	14.5	-4.1	110.3
g38c	$\ln(K_{MG})$	-57.7	-289.5	-7.0
g38c	n_{mg}	0.4	0.2	1.0
c50a	$\ln(K_{2AP})$	-3.9	-4.0	-3.8
c50a	$\ln(K_{MG})$	-6.3	-6.8	-6.1
c50a	n_{mg}	2.3	0.9	4.8
c60g	$\ln(K_{2AP})$	-2.4	-2.5	-2.3
c60g	$\ln(K_{MG})$	-7.0	-7.2	-6.8
c60g	n_{mg}	1.9	1.5	2.4
a35u/c60g	$\ln(K_{2AP})$	-2.9	-3.0	-2.7
a35u/c60g	$\ln(K_{MG})$	-7.3	-7.7	-7.1
a35u/c60g	n_{mg}	1.5	1.1	4.2
g38c/c60g	$\ln(K_{2AP})$	-2.8	-2.8	-2.7
g38c/c60g	$\ln(K_{MG})$	-6.3	-6.6	-6.1
g38c/c60g	n_{mg}	1.7	1.1	4.7
c50a/c60g	$\ln(K_{2AP})$	-5.1	-11.4	21.9
c50a/c60g	$\ln(K_{MG})$	14.8	-6.3	33.0
c50a/c60g	n_{mg}	-0.1	-4.8	4.8

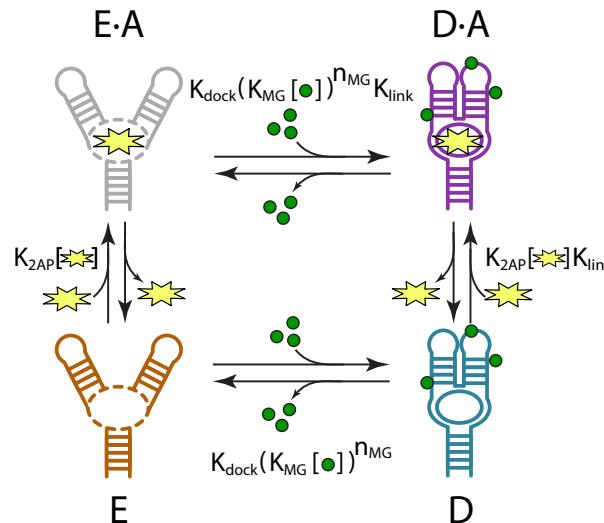
4-state model

4-state, 5 parameter

K_{2AP} K_{MG} n_{MG} K_{dock} K_{link}

4-state, 4 parameter

K_{2AP} K_{MG} n_{MG} K_{link}



$$S \equiv 1 + K_{2AP}[A]$$

$$T \equiv 1 + K_{2AP}K_{link}[A]$$

$$[M] = [M]_T + \left(\frac{n_{mg}T}{T-S} \right) ([R]_T(S-1) - S[A]_T + S[A])$$

$$[E] = \frac{[R]_T}{1 + K_{2AP}[A] + K_{dock}((K_{MG}[M])^{n_{MG}}(1 + K_{2AP}K_{link}[A]))}$$

$$[D] = K_{dock}((K_{MG}[M])^{n_{MG}})[E]$$

$$[E \cdot A] = K_{2AP}[A][E]$$

$$[D \cdot A] = [D]K_{2AP}K_{link}[A]$$

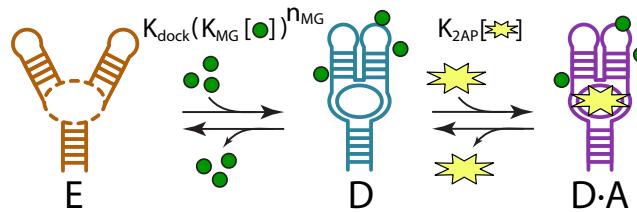
3-state model

3-state, 4 parameter

K_{2AP} K_{MG} n_{MG} K_{dock}

3-state, 3 parameter

K_{2AP} K_{MG} n_{MG}



$$[M] = [M]_T - n_{MG}([A]_T - [A]) - \frac{n_{MG}([A]_T - [A])}{K_{2AP}[A]}$$

$$[E] = \frac{[R]_T - [A]_T + [A]}{1 + K_{dock}(K_{MG}[M])^{n_{MG}}}$$

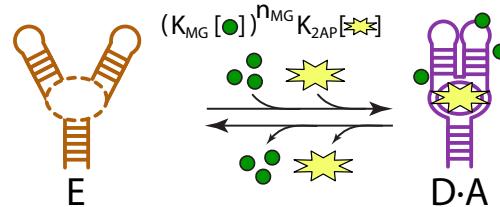
$$[D] = K_{dock}(K_{MG}[M])^{n_{MG}}[E]$$

$$[D \cdot A] = K_{2AP}[A][D]$$

2-state model

2-state, 3 parameter

K_{2AP} K_{MG} n_{MG}



$$[M] = [M]_T - n_{MG}([A]_T - [A])$$

$$[E] = \frac{[R]_T}{1 + K_{2AP}(K_{MG}[M])^{n_{MG}}[A]}$$

$$[D \cdot A] = K_{2AP}(K_{MG}[M])^{n_{MG}}[A][E]$$

Fig S1: Variants of the adenine riboswitch thermodynamic model. Parameters required are shown on the left; model schema are shown in the middle; mathematical expressions for the the different species are shown on the right. The colors and iconography in the schema match those shown in Figure 1A. In the mathematical expressions, [M] and [A] indicate the free Mg²⁺ and 2AP concentrations; [M]_T and [A]_T indicate their total concentrations. [R]_T is the total RNA concentration. To estimate species concentrations, We numerically found values for [A] that satisfied the conservation of mass in RNA, Mg²⁺, and 2AP.

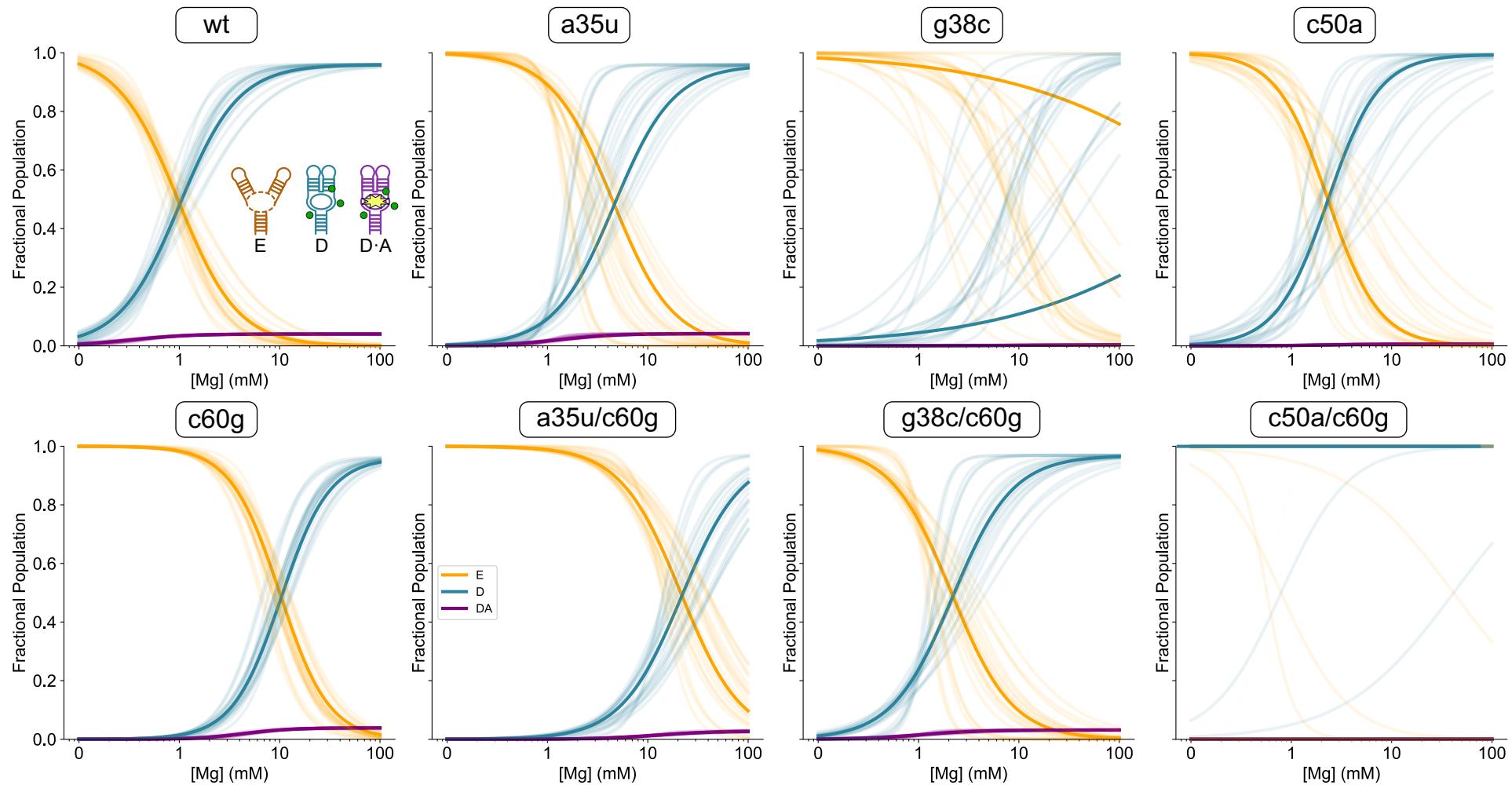


Fig S3. Magnesium-dependent changes in the populations of conformations in the ensemble. Each panel shows the population of E (orange), D (teal), and D·A (purple) as a function of total $[Mg^{2+}]$ for the indicated RNA variant. Faded lines were calculated using parameters taken from 50 randomly selected MCMC samples of the three-state, three-parameter model shown in Fig 1A. The solid lines are the median populations across MCMC replicates.

Appendix S1: Linkage model derivations

1 Four-state model

1.1 Model description

See Fig S1 for the relevant cycle.

1.1.1 Molecular species:

- E is extended RNA.
- D is docked RNA.
- $E \cdot A$ is extended RNA bound to 2AP.
- $D \cdot A$ is docked RNA bound to 2AP.
- D and $D \cdot A$ bind n_{MG} Mg^{2+} ions
- A is the free 2AP concentration.
- M is the free magnesium concentration.

1.1.2 Equilibria

- $E + M \rightleftharpoons D$; $[D] = K_{dock}(K_{MG}[M])^{n_{MG}}[E]$
- $E + A \rightleftharpoons E \cdot A$; $[E \cdot A] = K_{2AP}[E][A]$
- $D + A \rightleftharpoons D \cdot A$; $[D \cdot A] = K_{2AP}K_{link}[D][A]$
- $E \cdot A + M \rightleftharpoons D \cdot A$; $[D \cdot A] = K_{dock}(K_{MG}[M])^{n_{MG}}K_{link}[E \cdot A]$

1.1.3 Final equations:

We can describe the concentrations of all species in terms of constants, total concentrations, and $[A]$. To calculate the concentrations of all species, we guess values of $[A]$, use those values to calculate $[M]$, and iterate to self-consistency between the thermodynamic and mass balance relationships.

$$\begin{aligned}
 S &\equiv 1 + K_{2AP}[A] \\
 T &\equiv 1 + K_{2AP}K_{link}[A] \\
 [M] &= [M]_T + \left(\frac{n_{mg}T}{T - S} \right) ([R]_T(S - 1) - S[A]_T + S[A]) \\
 [E] &= \frac{[R]_T}{1 + K_{2AP}[A] + K_{dock}((K_{MG}[M])^{n_{MG}}(1 + K_{2AP}K_{link}[A]))} \\
 [D] &= K_{dock}((K_{MG}[M])^{n_{MG}})[E] \\
 [E \cdot A] &= K_{2AP}[A][E] \\
 [D \cdot A] &= [D]K_{2AP}K_{link}[A]
 \end{aligned}$$

1.1.4 2AP binding observable:

$$\Theta = \frac{[E \cdot A] + [D \cdot A]}{[A]_T}$$

1.2 Derive model

1.2.1 Find an expression for $[E]$ that depends on constants, total concentrations, $[A]$, and $[M]$.

1. Define all RNA species in terms of thermodynamic parameters and $[E]$, $[M]$, and $[A]$.

$$[E] = [E] \quad (1)$$

$$[D] = K_{dock}((K_{MG}[M])^{n_{MG}})[E] \quad (2)$$

$$[E \cdot A] = K_{2AP}[A][E] \quad (3)$$

$$[D \cdot A] = K_{2AP}K_{link}[D][A] = K_{2AP}K_{link}K_{dock}((K_{MG}[M])^{n_{MG}})[E][A] \quad (4)$$

2. Define the mass-balance relationships for RNA, magnesium ions, and 2AP:

$$[R]_T = [E] + [E \cdot A] + [D] + [D \cdot A] \quad (5)$$

$$[M]_T = [M] + n_{MG}[D] + n_{MG}[D \cdot A] \quad (6)$$

$$[A]_T = [A] + [E \cdot A] + [D \cdot A] \quad (7)$$

3. Rewrite mass-balance relationships using thermodynamic definitions of RNA species:

$$[R]_T = [E] + K_{2AP}[A][E] + K_{dock}((K_{MG}[M])^{n_{MG}})[E] + K_{dock}((K_{MG}[M])^{n_{MG}})K_{2AP}K_{link}[A][E] \quad (8)$$

$$[M]_T = [M] + n_{MG}K_{dock}((K_{MG}[M])^{n_{MG}})[E] + nK_{dock}((K_{MG}[M])^{n_{MG}})K_{2AP}K_{link}[A][E] \quad (9)$$

$$[A]_T = [A] + K_{2AP}[A][E] + K_{dock}((K_{MG}[M])^{n_{MG}})K_{2AP}K_{link}[A][E] \quad (10)$$

4. Rearrange Eq. 8:

$$[E] = \frac{[R]_T}{1 + K_{2AP}[A] + K_{dock}((K_{MG}[M])^{n_{MG}}(1 + K_{2AP}K_{link}[A]))} \quad (11)$$

1.2.2 Find an expression for $[M]$ that depends on constants, total concentrations, and $[A]$.

1. Isolate $(K_{MG}[M])^{n_{MG}}$ by subtracting Eq. 10 from Eq. 8:

$$[R]_T - [A]_T = [E] + K_{dock}((K_{MG}[M])^{n_{MG}})[E] - [A] \quad (12)$$

$$(K_{MG}[M])^{n_{MG}} = \frac{[R]_T - [A]_T - [E] + [A]}{K_{dock}[E]} \quad (13)$$

2. Substitute $(K_{MG}[M])^{n_{MG}}$ from Eq. 13 into Eq. 9 and simplify:

$$[M]_T = [M] + n_{MG}(1 + K_{2AP}K_{link}[A])([R]_T - [A]_T - [E] + [A]) \quad (14)$$

3. Eq. 14 now contains $[E]$. To remove it, we need to substitute an expression for $[E]$ that does not include the term $(K_{MG}[M])^{n_{MG}}$. Subtract Eq. 9 from Eq. 8 multiplied by n_{MG} , then solve for $[E]$.

$$n_{MG}[R]_T - [M]_T = n_{MG}[E] + n_{MG}K_{2AP}[A][E] - [M] \quad (15)$$

$$[E] = \frac{n_{MG}[R]_T - [M]_T + [M]}{n_{MG}(1 + K_{2AP}[A])} \quad (16)$$

4. Substitute $[E]$ from Eq. 16 into Eq. 14:

$$[M]_T = [M] + n_{MG}(1 + K_{2AP}K_{link}[A])([R]_T - [A]_T + [A]) - (1 + K_{2AP}K_{link}[A]) \left(\frac{n_{MG}[R]_T - [M]_T + [M]}{n_{MG}(1 + K_{2AP}[A])} \right) \quad (17)$$

5. Define some convenient terms that depend only on constants and $[A]$.

$$S \equiv 1 + K_{2AP}[A]$$

$$T \equiv 1 + K_{2AP}K_{link}[A]$$

6. Simplify Eq. 17 by substituting S and T and multiplying the entire equation by S :

$$S[M]_T = S[M] + n_{MG}ST([R]_T - [A]_T + [A]) - T(n_{MG}[R]_T - [M]_T + [M]) \quad (18)$$

7. Solve for $[M]$:

$$[M] = [M]_T + \left(\frac{n_{MG}T}{T - S} \right) ([R]_T(S - 1) - S[A]_T + S[A]) \quad (19)$$

2 3-state model

2.1 Model description

For the 3-state model, we eliminate $E \cdot A$, which implicitly eliminates K_{link} (the difference in 2AP binding between E and D). For the relevant cycle, see Fig S1.

2.1.1 Molecular species:

- E is extended RNA.
- D is docked RNA.
- $D \cdot A$ is docked RNA bound to 2AP.
- D and $D \cdot A$ bind n_{MG} Mg^{2+} ions
- A is the free 2AP concentration.
- M is the free magnesium concentration.

2.1.2 Equilibria

- $E + M \rightleftharpoons D$; $[D] = K_{dock}(K_{MG}[M])^{n_{MG}}[E]$
- $D + A \rightleftharpoons D \cdot A$; $[D \cdot A] = K_{2AP}[D][A]$

2.1.3 Final equations:

We can describe the concentrations of all species in terms of constants, total concentrations, and $[A]$. To calculate the concentrations of all species, we guess values of $[A]$ and iterate to self-consistency between the thermodynamic and mass balance relationships.

$$\begin{aligned} [M] &= [M]_T - n_{MG}([A]_T - [A]) - \frac{n_{MG}([A]_T - [A])}{K_{2AP}[A]} \\ [E] &= \frac{[R]_T - [A]_T + [A]}{1 + K_{dock}(K_{MG}[M])^{n_{MG}}} \\ [D] &= K_{dock}(K_{MG}[M])^{n_{MG}}[E] \\ [D \cdot A] &= K_{2AP}[A][D] \end{aligned}$$

2.1.4 2AP binding observable:

$$\Theta = \frac{[D \cdot A]}{[A]_T}$$

2.2 Derive model

2.2.1 Find an expression for $[E]$ that depends on constants, total concentrations, $[A]$, and $[M]$.

1. Define all RNA species in terms of thermodynamic parameters and $[E]$, $[M]$, and $[A]$.

$$[E] = [E] \quad (20)$$

$$[D] = K_{dock}((K_{MG}[M])^{n_{MG}})[E] \quad (21)$$

$$[D \cdot A] = K_{2AP}K_{dock}((K_{MG}[M])^{n_{MG}})[E][A] \quad (22)$$

2. Define the mass-balance relationships for RNA, magnesium ions, and 2AP:

$$[R]_T = [E] + [D] + [D \cdot A] \quad (23)$$

$$[M]_T = [M] + n_{MG}[D] + n_{MG}[D \cdot A] \quad (24)$$

$$[A]_T = [A] + [D \cdot A] \quad (25)$$

3. Rewrite mass-balance relationships using thermodynamic definitions of RNA species:

$$[R]_T = [E] + K_{dock}((K_{MG}[M])^{n_{MG}})[E] + K_{dock}((K_{MG}[M])^{n_{MG}})K_{2AP}[A][E] \quad (26)$$

$$[M]_T = [M] + n_{MG}K_{dock}((K_{MG}[M])^{n_{MG}})[E] + n_{MG}K_{dock}((K_{MG}[M])^{n_{MG}})K_{2AP}[A][E] \quad (27)$$

$$[A]_T = [A] + K_{dock}((K_{MG}[M])^{n_{MG}})K_{2AP}[A][E] \quad (28)$$

4. Subtract Eq. 28 from Eq. 26 to isolate $[E]$:

$$[R]_T - [A]_T = [E] + K_{dock}((K_{MG}[M])^{n_{MG}})[E] - [A] \quad (29)$$

$$[E] = \frac{[R]_T - [A]_T + [A]}{1 + K_{dock}(K_{MG}[M])^{n_{MG}}} \quad (30)$$

2.2.2 Find an expression for $[M]$ that depends on constants, total concentrations, and $[A]$.

1. Isolate $(K_{MG}[M])^{n_{MG}}$ by rearranging Eq. 30:

$$(K_{MG}[M])^{n_{MG}} = \frac{[R]_T - [A]_T - [E] + [A]}{K_{dock}[E]} \quad (31)$$

2. Substitute $(K_{MG}[M])^{n_{MG}}$ from Eq. 31 into Eq. 27 and simplify:

$$[M]_T = [M] + n_{MG}(1 + K_{2AP}[A])([R]_T - [A]_T - [E] + [A]) \quad (32)$$

3. Eq. 32 now contains $[E]$. To remove it, we need to substitute an expression for $[E]$ that does not include the term $(K_{MG}[M])^{n_{MG}}$. Subtract Eq. 27 from Eq. 26 multiplied by n_{MG} , then solve for $[E]$.

$$n_{MG}[R]_T - [M]_T = n_{MG}[E] - [M] \quad (33)$$

$$[E] = \frac{n_{MG}[R]_T - [M]_T + [M]}{n_{MG}} \quad (34)$$

4. Substitute $[E]$ from Eq. 34 into Eq. 32, expand, and solve for $[M]$:

$$[M]_T = [M] + n_{MG}(1 + K_{2AP}[A]) \left([R]_T - [A]_T - \left(\frac{n_{MG}[R]_T - [M]_T + [M]}{n_{MG}} \right) + [A] \right) \quad (35)$$

$$[M] = [M]_T - n_{MG}([A]_T - [A]) - \frac{n_{MG}([A]_T - [A])}{K_{2AP}[A]} \quad (36)$$

3 2-state model

3.1 Model description

For the 2-state model, we eliminate D , which implicitly eliminates K_{dock} (it gets folded into K_{2AP}). For the relevant cycle, see Fig S1.

3.1.1 Molecular species:

- E is extended RNA.
- $D \cdot A$ is docked RNA bound to 2AP.
- $D \cdot A$ binds n_{MG} Mg^{2+} ions
- A is the free 2AP concentration.
- M is the free magnesium concentration.

3.1.2 Equilibria

- $E + M \rightleftharpoons D \cdot A; [D \cdot A] = (K_{MG}[M])^{n_{MG}} K_{2AP}[A][E]$

3.1.3 Final equations:

We can describe the concentrations of all species in terms of constants, total concentrations, and $[A]$. To calculate the concentrations of all species, we guess values of $[A]$ and iterate to self-consistency between the thermodynamic and mass balance relationships.

$$\begin{aligned} [M] &= [M]_T - n_{MG}([A]_T - [A]) \\ [E] &= \frac{[R]_T}{1 + K_{2AP}(K_{MG}[M])^{n_{MG}}[A]} \\ [D \cdot A] &= K_{2AP}(K_{MG}[M])^{n_{MG}}[A][E] \end{aligned}$$

3.1.4 2AP binding observable:

$$\Theta = \frac{[D \cdot A]}{[A]_T}$$

3.2 Derive model

3.2.1 Find an expression for $[E]$ that depends on constants, total concentrations, $[A]$, and $[M]$.

1. Define all RNA species in terms of thermodynamic parameters and $[E]$, $[M]$, and $[A]$.

$$[E] = [E] \quad (37)$$

$$[D \cdot A] = K_{2AP}(K_{MG}[M])^{n_{MG}}[A][E]$$

2. Define the mass-balance relationships for RNA, magnesium ions, and 2AP:

$$[R]_T = [E] + [D \cdot A] \quad (38)$$

$$[M]_T = [M] + n_{MG}[D \cdot A] \quad (39)$$

$$[A]_T = [A] + [D \cdot A] \quad (40)$$

3. Rewrite mass-balance relationships using thermodynamic definitions of RNA species:

$$[R]_T = [E] + K_{2AP}(K_{MG}[M])^{n_{MG}}[A][E] \quad (41)$$

$$[M]_T = [M] + n_{MG}K_{2AP}(K_{MG}[M])^{n_{MG}}[A][E] \quad (42)$$

$$[A]_T = [A] + K_{2AP}(K_{MG}[M])^{n_{MG}}[A][E] \quad (43)$$

4. Rearrange Eq. 41:

$$[E] = \frac{[R]_T}{1 + K_{2AP}(K_{MG}[M])^{n_{MG}}[A]} \quad (44)$$

3.2.2 Find an expression for $[M]$ that depends on constants, total concentrations, and $[A]$.

1. Isolate $(K_{MG}[M])^{n_{MG}}$ by rearranging Eq. 43:

$$(K_{MG}[M])^{n_{MG}} = \frac{[A]_T - [A]}{K_{2AP}[A][E]} \quad (45)$$

2. Substitute $(K_{MG}[M])^{n_{MG}}$ from Eq. 45 into Eq. 42 and simplify:

$$[M]_T = [M] + n_{MG}K_{2AP} \left(\frac{[A]_T - [A]}{K_{2AP}[A][E]} \right) [A][E] \quad (46)$$

$$[M] = [M]_T - n_{MG}([A]_T - [A]) \quad (47)$$