Allosteric Logic of the \textit{V. vulnificus} Adenine Riboswitch Resolved by Four-dimensional Chemical Mapping

Siqi Tian\textsuperscript{1}, Wipapat Kladwang\textsuperscript{1}, and Rhiju Das\textsuperscript{1, 2, *}

\textsuperscript{1}Department of Biochemistry, and \textsuperscript{2}Department of Physics, Stanford University, Stanford, California 94305, United States

* Corresponding Author. Phone: (650) 723 5976. E-mail: rhiju@stanford.edu
ABSTRACT
The structural logic that define the functions of gene regulatory RNA molecules may be radically different from classic models of allostery, but the relevant structural correlations have remained elusive in even intensively studied RNA model systems. Here, we present a four-dimensional expansion of chemical mapping called lock-mutate-map-rescue (LM$^2$R), which integrates multiple layers of mutation with nucleotide-resolution chemical mapping. This technique resolves the core mechanism of the adenine-responsive *V. vulnificus* add riboswitch including its gene expression platform, a paradigmatic system for which both Monod-Wyman-Changeux (MWC) conformational selection models and non-MWC alternatives have been proposed. To discriminate amongst these models, we locked each functionally important helix through designed mutations and assessed formation or depletion of other helices via high-throughput compensatory rescue experiments. These LM$^2$R measurements give strong support to the pre-existing correlations predicted by MWC models, disfavor alternative models, and reveal new structural heterogeneities that may be general across ligand-free riboswitches.

**Keywords:** riboswitch, allostery, conformational ensemble, compensatory mutagenesis, SHAPE, secondary structure, multidimensional chemical mapping.
INTRODUCTION

Conformational changes in RNA molecules are ubiquitous features of gene regulation in all living cells and viruses. Recent years have seen an explosion of discoveries of compact *cis*-acting mRNA elements that sense small molecules, proteins, and other RNAs and modulate transcriptional termination, ribosome recruitment, splicing, and other genetic events.\(^1\)\(^-\)\(^3\) Beginning with classic work by Yanofsky and colleagues in the 1970s, these discoveries have been associated with elegant models of allostery in which each molecule interconverts between at least two conformational states.\(^4\)\(^-\)\(^6\) In particular, the allostery is typically presumed to occur through Monod-Wyman-Changeux mechanisms (also called ‘conformational selection’ or ‘population shift’ models)\(^6\)-\(^8\). In these pictures, ligand binding to an aptamer region of the RNA shifts a pre-existing equilibrium between strikingly different secondary structures, resulting in allosteric exposure or sequestration of another sequence-distal region harboring a ribosome binding site or other gene expression platform.\(^9\)-\(^11\) If these multiple structures could be determined in detail, they would each offer a potential new target for biological control and antibiotic development.\(^12\)

Despite decades of work on RNA allostery, experiments directly testing the base pair correlations posited for specific systems have been difficult. A paradigmatic example is the *add* riboswitch, which has become a dominant model system for studying ligand-sensing riboswitches. This compact adenine-sensing element resides in the 5′ untranslated region (UTR) of the *add* adenosine deaminase mRNA of human pathogenic bacterium *Vibrio vulnificus*, and controls translation of the mRNA in response to adenine. Domains of this RNA have been subjected to nearly every *in vitro* biochemical and biophysical technique available, including time-resolved X-ray laser diffraction.\(^13\)-\(^24\) Almost all studies to date have assumed some variation of an MWC model of action.\(^15\) Recently, however, a tour de force study, bringing together multidimensional NMR spectroscopy, NMR relaxation studies, designed model systems, and supporting measurements from stopped-flow and calorimetry techniques, revealed detailed single-nucleotide-resolution base pairing information and intriguing temperature dependences for the complex structural ensemble of the *add* riboswitch.\(^25\) Although similar in some respects to previously proposed MWC models, the new model contradicted a standard MWC assumption\(^7\)-\(^8\),\(^26\): it proposes that, in the absence of adenine, riboswitches that sample the correct aptameric secondary structure do not also concomitantly open their ribosome binding sites and increase gene expression, as predicted by MWC conformational selection. Despite this fundamental distinction, the non-MWC model appears fully consistent with all available data, and it remains unclear what structural correlations, if any, underlie the allosteric logic of the *add* riboswitch. As
a reflection of current uncertainties in the system, the same group that first proposed an non-MWC model for the add RNA has recently revived the MWC framework to interpret newer single molecule measurements on the same system.27

To resolve the allosteric logic of the add riboswitch, we have developed a high-dimensional expansion of single-nucleotidereolution chemical mapping (Figure 1). We previously showed that structural perturbations from mutating each nucleotide of an RNA could be read out at every other nucleotide through chemical mapping to give rich two-dimensional data sets and accurate blind predictions for RNAs with single, well-defined structures (mutate-and-map, M²).28-30 Expanding to three dimensions, testing compensating mutations at pairs of nucleotides through their effects on the reactivities of other nucleotides gave incisive evidence for specific base pairs (mutate-map-rescue, M²R), again for molecules with single dominant structures.31-32 Here, we expand the M²R method to allow dissection of more complex ensembles involving numerous functionally important structures, taking advantage of a Bayesian framework and extensive simulations to connect experimental observables to underlying helix frequencies. By identifying ‘lock’ mutants that stabilize helices posited for each state, we infer how the presence of one helix enhances or suppresses the presence of other helices, through further rounds of M²R on the mutant backgrounds (Figures 1A and B). This four-dimensional chemical mapping workflow (termed lock-mutate-map-rescue, LM²R) resolves fundamental questions raised by the existing models of the add riboswitch and also exposes new ensemble properties that were not expected in prior models.

RESULTS

Inferring helix frequencies through quantitative compensatory rescue
Our key methodological innovation arose from attempts to more quantitatively analyze compensatory mutation experiments for RNA structure read out through chemical mapping (mutate-map-rescue, M²R).31-32 Briefly, if mutation of either side of a base pair (mutant A or mutant B) disrupts the reactivities at other nucleotides in the RNA, but the compensatory double mutation (mutant AB) restores the SHAPE reactivities to the wild type (WT) profile, these observations suggest that the two sites were base paired in the starting WT sequence. Our prior work assumed that such observed compensation would arise with negligible probability if the mutated nucleotides were not paired. To test this assumption and automate the protocol, we carried out simulations of these experiments and developed a quantitative ‘rescue factor’ metric that accurately recovered expert assessments from simulated and experimental data, including the
Tetrahymena ribozyme P4-P6 domain and a blind RNA-puzzle modeling challenge, the GIR1 lariat capping ribozyme (Methods & Supporting Results):

\[
\text{Rescue factor} \equiv 1.0 - \frac{\text{diff}_{(WT,AB)}}{\max(\text{diff}_{(WT,A)}, \text{diff}_{(WT,B)})},
\]

where the Euclidean difference \( \text{diff} \) between two SHAPE profiles \( d \) over probed nucleotides \( i = 1 \) to \( N \) is given by:

\[
\text{diff}_{(A,B)} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (d_{A} - d_{B})^2
\]

These in silico tests confirmed that observation of strong rescue implies high helix frequency (Figure 1C) but unexpectedly also showed that modest rescue ratios could be informative. Values lower than one but greater than zero corresponded to ‘partial’ rescue, where the reactivities at some nucleotides revert to wild type (WT) values but others do not (Figure 1D, center). While individual base pair tests were noisy, averaging rescue factors over tests on three or more base pairs gave values that showed a striking correlation with helix frequencies (Figure 1C).

Furthermore, absence of rescue (rescue factor close to zero; Figure 1D, bottom) gave upper bounds on the helix frequency (Supporting Figure S1). To estimate uncertainties in these estimates, we used a Bayesian framework to derive posterior probabilities of each helix frequency given the observed rescue factor, calibrated based on these simulations. Below, we present cumulative distributions of these posterior probabilities in figures and tables (see also Supporting Figure S1); for ease of reading in the main text, we give simply median values of the helix frequency posterior probability distributions. We also developed a separate category-based classification to check that compensatory rescue of different base pairs within the same helix gave concordant results, and that automated analysis reproduced manual analysis (Supporting Results).

These simulation and experimental studies suggested that compensatory rescue measurements, even ones that resulted in partial or no detectable rescue, could be used to estimate helix frequencies and uncertainties for those values, without relying on complex fits\(^{16}\), though we sought more detailed tests, as described next.

The add riboswitch system

To test the applicability of M^2R helix frequency estimation to understanding complex RNA structural ensembles, we applied it to the paradigmatic \( V. vulnificus \) add adenine-sensing
riboswitch including its expression platform (Figure 2). On one hand, the NMR analysis of this system offered detailed data for validating or falsifying helix frequencies. On the other hand, that study and other analyses have revealed an unresolved but fundamental question in how the RNA’s multiple structures give rise to function. There is currently no controversy about the riboswitch’s adenine-bound state (termed *holo*; Figure 2A, right). Helices P1, P2, and P3 are formed in the 5′-UTR, defining an adenine binding aptamer at nucleotides 15–81, with a well-defined tertiary structure that has been determined through crystallography. Also, a linker including an additional hairpin (P5, nts 89–110) connects the aptamer to the Shine-Dalgarno sequence and AUG start codon (nts 112–122), which are open and available for ribosome recruitment. In the absence of adenine, all available models posit that the riboswitch forms a secondary structure ensemble dominated by a new P4 domain that pairs with the Shine-Dalgarno sequence and the AUG start codon (nts 112–122) with nucleotides that were originally part of P1 (nts 75–88, here called the ‘switch region’\(^\text{15}\), cross-hatched in Figure 2A, left). Concomitantly, the aptamer helices P1 and P2 are rearranged into an alternative structure. In keeping with the previous NMR study, these conformations with a rearranged adenine aptamer secondary structure are grouped into a state *apoB*. The major uncertainty regards what happens when adenine is not present but the RNA nevertheless samples a secondary structure poised to bind adenine, with P1, P2, and P3 formed. This state is termed *apoA*.

At 20–30 °C, several studies have estimated that the fraction of the riboswitch in the adenine-binding-ready *apoA* conformation occurs with a frequency of ~40% both *in vitro* and *in cellulo*.\(^\text{13, 33}\) An MWC conformational selection model would assume that this thermally sampled state should have not only the same adenine aptamer secondary structure as adenine-bound *holo*, but also the same open Shine-Dalgarno conformation as the adenine-bound *holo*, without the P4 domain. That is, *apoA* should be ON, able to recruit ribosomes. A key prediction of the MWC conformational selection model is therefore an exact structural identity of the gene expression platform in the molecule’s active configuration, whether sampled at thermal equilibrium (*apoA*) or stabilized by binding of ligands at the distal pocket (*holo*) (compare Figure 2A, center and right). (We note that the MWC model allows for local rearrangements in the aptamer binding pocket in going from *apoA* to *holo*, but these adenine-induced changes are not allowed to propagate to the gene expression platform, as the fixed stem P1 spatially separates the aptamer and the expression platform in either *apoA* and *holo*.)
At first glance, the NMR-informed portrait is similar to the MWC conformational selection model, with high-population and low-population adenine-free states that display distinct structures incompatible and compatible with adenine binding (apoB and apoA), and a third adenine-bound state (holo) that is similar to the adenine-free apoA. However, the key feature of the MWC conformational selection model is lost (Figure 2B). In the absence of ligand, the NMR measurements could not directly detect opening of the gene expression platform at low frequency in the absence of ligand, i.e., transient sampling of the ON conformation. While this lack of signal could be due to ambiguities in the NMR measurements, both ligand-free states apoA and apoB were interpreted to be OFF with respect to gene expression, in conflict with the MWC conformational selection model. The new model, referred to as the ‘non-MWC’ model henceforth, suggests instead a mixed conformation for apoA, where a shortened P1, P2, and P3 form the adenine aptamer secondary structure but also some parts of the P4 domain (here denoted P4A and P4B) are retained, blocking the Shine-Dalgarno site, i.e., apoA should be mostly OFF. The non-MWC model does not explain how subsequent adenine binding to the aptamer of apoA might communicate through P1 to release the Shine-Dalgarno site and promote gene expression, raising the prospect of a radically different mechanism for allostery in RNA compared to classic MWC systems. Posed as a simple question, then: is apoA ON (MWC) or OFF (non-MWC)?

Before carrying out extensive mutagenesis and compensatory rescue experiments, we tested whether one-dimensional SHAPE profiling on the WT RNA might discriminate between MWC and non-MWC models (Figure 2 and SI Figure S3). In the absence of adenine, the add riboswitch SHAPE profile (Figure 2C) showed protections within the ‘switch region’, which is predicted to have pairings in both apoA and apoB, in both MWC and non-MWC models; other regions show variable levels of protection. Incubation with 5 mM adenine (Figure 2D) gave protections in the 5’-most aptamer region, consistent with formation of P1, P2, and P3, and RNA tertiary structure formation at and around the adenine binding pocket, as expected for the holo state. In addition, adenine-dependent increases in SHAPE reactivity in the expression platform, the Shine-Dalgarno ribosome binding site and the AUG codon (nts 112-122), supported the adenine-coupled opening of the expression platform that is the hallmark of riboswitch function (Figure 2D). Unexpectedly, in between these two regions, nucleotides 105-111 also modestly increased in SHAPE reactivity upon adenine binding, although, in both the MWC and non-MWC models, this segment is predicted to be sequestered into a P5 helix in all states (Figure 2). However, this observation can be accommodated into both the MWC and non-MWC models by assuming partial opening of P5.
in the adenine-bound holo state (see Discussion, below) and therefore does not discriminate between the models.

**Three-dimensional mutate-map-rescue recovers ensemble helix frequencies**

We then tested if quantitative mutate-map-rescue (Figure 1) might confirm or contradict prior NMR results and enable discrimination between MWC and non-MWC models. We carried out M^2R method on the separate adenine-bound and adenine-free ensembles. Prior to rescue experiments, which involve double compensatory mutants, we systematically scanned single mutants of each nucleotide to its complement, followed by SHAPE profiling across the riboswitch, which gave the two-dimensional patterns in Figure 3. Automated analyses of these mutate-and-map (M^2) data to determine the dominant secondary structure and to infer an approximate structural ensemble supported both the MWC and non-MWC models but did not carry the precision to discriminate between them (Supporting Figure S4). Nevertheless, the M^2 data were suggestive. As expected for a multi-state RNA, several mutations produced dramatic perturbations across the SHAPE profile. Compared to the WT sequence, many of these mutants gave protections or enhancements in the 5’-most aptamer region of the riboswitch (nts 25-32) and concomitant exposures or protections (respectively) near the 3’ end of the riboswitch (nts 105-122), which included both the start codon and Shine-Dalgarno ribosome binding sequence (arrows, Figure 3A-B). Such correlations between structuring the aptameric region and opening the gene expression platform, even in the absence of ligand, qualitatively supported predictions of the MWC conformational selection model, but we sought stronger, quantitative tests through deeper mutational analysis.

Compensatory rescue experiments assessed helix frequencies in the presence and absence of adenine and quantitatively agreed with prior models (Figure 3E-F). First, in all prior models, the adenine-bound riboswitch (holo, Figure 2) was expected to display P1, P2, and P3 in the aptamer and P5. Indeed, in the presence of 5 mM adenine, SHAPE measurements for single and compensatory mutations targeting these helices and rescue factor analysis gave helix frequencies of 79%, 64%, 56% and 29%, respectively (Figure 3E,3G). Each of these cases showed visually striking rescue in their SHAPE profiles as mutations disrupted adenine binding, and compensatory mutations restored the adenine-bound profiles (Figure 3C, top ‘quartet’ of profiles); the lower helix frequency of P5 was consistent with its assumed partial opening, as noted with one-dimensional mapping. Measuring rescue factors for other tested helices, including P1B, P2B, P4A, P4B, and P4C predicted for the apoB state and numerous helices (P6, P7,
through P17; SI Figure S4) proposed to occur at low frequencies from M^2-guided modeling, gave no evidence for their presence. We derived upper bounds on these helices’ frequencies of 16% or lower; in cumulative distributions of posterior probabilities, there was a clear separation from the high-frequency helices (Figure 3C, bottom quartet of profiles; Figures 3E-3G; SI Figure S5; and Table 1). In addition, we used a double-base-pair mutation scheme for M^2R, which induced stronger perturbations and rescuing effects, and supported the same pairings as above (SI Figure S6 and Table 2).

In addition to their predictions above for the adenine-bound state, all prior models of the add riboswitch expected the RNA to show a diverse set of helices in the absence of adenine, reflecting an admixture of apoA vs. apoB. Helices P3 and P5 are expected to be present at high frequencies, and both are detected by compensatory rescue experiments, with median helix frequencies of 73% and 75%, respectively (Figure 3F). In all prior models (Figure 2), other helices were predicted to be present at lower but significant frequencies. Compensatory rescue experiments recovered evidence for the apoB-specific helices P4A (48%), P1B (44%), and P2B (17%); the apoA-specific helices P1 (53%) and P2 (31%); and the apoB helices whose presence in apoA is under question, P4B (55%) and P4C (68%) (Figure 3D, 3F, 3H). As with the adenine-bound conditions, the double-base-pair M^2R results were consistent with the single-base-pair M^2R results (SI Figure S6). In addition, helices P6, P7, through P17 (SI Figure S5) again served as negative controls and gave low helix frequencies by compensatory rescue, as expected (pink cumulative distributions in Figure 3H; Table 1). In principle, the results for helices P4B and P4C could discriminate between the models under question (Figure 2A-B), as the non-MWC model predicts their frequencies to be close to 100%, while the MWC conformational selection model predicts their frequencies to be lower, in the same range as P4A (48%; compare yellow and blue cumulative distributions in Figure 3H). While the low helix frequency estimates favor the predictions from the MWC conformational selection model, remaining uncertainties in these values (broadness of probability distributions in Figure 3H; and Tables 1-2) precluded incisive discrimination between the models based purely on M^2R experiments with the WT add sequence, and we turned to the four-dimensional chemical mapping method that involves M^2R in mutational backgrounds that lock each riboswitch state.

*Four-dimensional Lock-Mutate-Map-Rescue enables dissection of ensemble states*
To stringently discriminate between the MWC and non-MWC models, we sought to measure the presence or absence of helices P4B and P4C in the *apoA* state, taking advantage of mutant backgrounds that ‘lock’ this state and the ability of quantitative compensatory rescue to estimate helix frequencies (lock-mutate-map-rescue, LM²R). While our original single-mutant scan (Figure 3B and SI Figure S4) suggested several such mutants, we were concerned that their effects involved destabilizing the P4 domain and might bias our measurements towards favoring the MWC model. Instead, we took advantage of mutations outside P4 that stabilized P1 or P2, components of the aptamer secondary structure that define *apoA* in all available models. A number of double-base-pair compensatory mutants for P1 and P2 exhibited SHAPE profiles that agreed well with each other (Figure 4A), supporting their use as mutant backgrounds for isolating *apoA*. We noted that all of these *apoA*-stabilizing mutants exhibited increased reactivity in the expression platform region (nts 105-122) relative to the adenine-free WT RNA, at a level comparable to the adenine-bound WT RNA (Figure 4A); this observation already suggested that *apoA* is similar to *holo* in this region, favoring the MWC model over the non-MWC model. We also observed similarities in the gene expression platform between these adenine-free *apoA*-stabilizing mutants and adenine-bound WT RNA with other chemical modifiers (dimethyl sulfate, glyoxal, ribonuclease V1, terbium(III), and ultraviolet irradiation at 302 nm; SI Figure S7).

We selected two mutants for lock-mutate-map-rescue probing *apoA*: lock-P1 (A19C, U20G, A76C, U77G) and lock-P2 (U28G, A29C, U41G, G42C). As a consistency check, we confirmed that compensatory rescue gave a strong P1 helix frequency (72%) when P2 was locked. Similarly, compensatory rescue targeting P2 gave a significant helix frequency (42%) when P1 was locked (Figure 4C, SI Figure S8, and Table 2). As a further check, both models under question predicted the absence of P4A in *apoA*; indeed, in either the lock-P1 or lock-P2 background, we did not detect compensatory rescue targeting P4A, and could set an upper bound on its frequency of 4%. Finally, we tested the helix frequency of P4B, which was predicted to be absent (0%) in *apoA* by the MWC model and present (100%) in the non-MWC model (Figure 2A). We saw no evidence for compensatory rescue of P4B in either *apoA*-stabilizing background. Measurements in the lock-P1 background and in the lock-P2 background gave P4B helix frequencies of 4%, in both cases (Figure 4C; and compare blue and yellow curves in Figure 4E). Separate experiments on a variant MutP2 (A29C, A30G, U40C, U41G) used to stabilize *apoA* in prior studies similarly gave no evidence for P4B (SI Figure S9, Table 3). Taken together, these measurements provided strong evidence against the non-MWC model and favor the MWC model.
To further dissect the adenine-free *add* riboswitch ensemble, we selected analogous mutant backgrounds that stabilized *apoB*, based on double-base-pair mutants used to stabilize P1B (U17G, C18G, G43C, G44C) and P4A (U77G, G78C, C119G, A120C) (Figure 4B). For the lock-P1B mutant, compensatory-rescue experiments confirmed the helices expected in *apoB* by both the MWC and non-MWC models (P2B, P4A, and P4B at 53%, 55%, and 80%, respectively; Figure 4F). Experiments with the lock-P4A mutant also gave the expected rescue for P4B (55%) but, unexpectedly, not for P1B or P2B (6% or 9%). The latter results suggest previously undetected conformational heterogeneity in the aptamer region in *apoB* (Figure 4D, Supporting Results). Further supporting the inability to trap *apoB* by locking any single helix, the reactivity profiles of lock-P4A and lock-P1B were distinct (Figure 4B).

As an independent test of the helix content in *apoA* vs. *apoB*, we carried out a ‘flipped’ lock-mutate-map-rescue experiment compared to the experiments above. Rather than locking P1 and P2 and assessing the presence of P4B, we locked P4B and assessed the presence of P1 and P2. In the MWC model (Figure 2A), a P4B-stabilizing mutant background should isolate *apoB*, and give no compensatory rescue for P1 and P2. In the non-MWC model (Figure 2B), where P4B is common to both the *apoA* and *apoB*, the P4B-stabilizing mutant background should give the same results as the WT RNA, and give partial compensatory rescue for P1 and P2. In fact, the P4B-stabilizing background (lock-P4B; U82G, C83G, G115C, A116C) gave a SHAPE profile better matching the *apoB*-stabilizers than the WT RNA sequence, favoring the MWC model over the non-MWC model. Furthermore, compensatory rescue experiments in the lock-P4B background showed no rescue for P1 and P2 (median helix frequencies estimated at 5% and 6%, respectively), again favoring the MWC model over the non-MWC model (SI Figure S10).

**DISCUSSION**

*Four-dimensional chemical mapping for complex RNA ensembles*

Complex secondary structure ensembles underlie many, and perhaps most, RNA regulatory elements. Understanding how a panoply of structures underlies the allosteric logic of these elements requires assessing not just the frequencies of different helices in the RNA structural ensemble but how the presence of one helix enhances or suppresses the frequency of others. Four-dimensional chemical mapping offers an experimental route to this information that is unique in its high throughput and use of standard equipment (capillary electrophoretic sequencers). The strategy (Figure 1) is to ‘lock’ one helix into place through mutations (dimension 1) and to then introduce mutations elsewhere (dimension 2). Perturbation of the chemical mapping profile at
other nucleotides (dimension 3) provides evidence of involvement by the mutated nucleotide in RNA structure, perhaps a second helix. If compensatory mutation of a candidate partner in the second helix (dimension 4) then gives no, partial, or complete rescue of the mapping profile, the frequency of the second helix can be inferred to be low, medium, or high, with uncertainties estimated from simulation. The current throughput of PCR assembly and capillary electrophoretic sequencing enables single-nucleotide-resolution mapping of many hundreds of variants per experiment at the signal-to-noise needed for this four-dimensional lock-mutate-map-rescue (LM^2R) workflow. With this throughput, dissection of the co-occurrence or exclusivity of variable helices underlying a 100 to 200 nucleotide RNA domain becomes feasible.

**Resolving the core mechanism of a paradigmatic riboswitch**

This work demonstrates application of four-dimensional chemical mapping to resolve a fundamental biochemical question. Applying the methodology to the *add* riboswitch from *V. vulnificus* gives the model shown in Figure 5, consistent with a Monod-Wyman-Changeux-like ‘standard model’ of allostery rather than a non-MWC revision proposed after detailed NMR experiments. The two models differed in one core aspect: how ligand sensing in an aptamer region is allosterically communicated to expose a ribosome binding site and turn on *add* mRNA gene expression (Figure 2). Early M^2 experiments showed that, even in the absence of adenine, mutations that gave protections in the 5’ region of the *add* riboswitch also gave exposures in the 3’ region including the ribosome binding site (Figure 3A). Furthermore, different mutants that locked the correct adenine secondary structure without adenine exhibited SHAPE data that were indistinguishable from the adenine-bound riboswitch in the gene expression platform and, more broadly, at all nucleotides downstream of the aptamer region (Figure 4A-B). The most incisive analysis came from LM^2R. In experiments locking P1 and P2, which are specific to the aptameric secondary structure, the rescue data give evidence against the formation of P4 helices that would sequester the ribosome binding site. In other words, the *add* gene expression platform is ON when the sequence-separate aptamer region is folded even in the absence of the adenine ligand (Figure 2; Figure 5, *apoA*). Conversely, LM^2R experiments locking P1B, P4A, or P4B turn the riboswitch OFF and give evidence against the formation of P1 and P2 helices.

These data favor anti-correlation between aptamer secondary structure formation and structures sequestering the ribosome binding site even without adenine binding, as predicted in the MWC conformational selection model, and disfavor a non-MWC revision that omits this coupling (compare Figure 2A and Figure 2B). Stated differently, allosteric communication of the aptameric
region and the gene expression platform is an intrinsic property of the folding landscape of the *add* riboswitch sequence that can be established without adenine. A simple conceptual understanding of this property involves considering a “switch region”\(^{15}\) (nucleotides 75–81, Figure 5) that is able to form base pairs with nucleotides at the 5’ end of the RNA (anti-switch region) but is also able to form pairs (in P4A) with four nucleotides in the Shine-Dalgarno ribosome binding site at the 3’ end of the untranslated region (SD, Figure 5). Depending on whether the former or latter pairing is ‘chosen’ by the switch region, the RNA either closes the P1 helix of the aptamer secondary structure (P1) or closes the P4 domain that sequesters the ribosome binding site. As noted above, despite its similarity to MWC allosteric mechanisms for proteins, this model for the *add* riboswitch shows much more dramatic changes in structure than is typically seen in protein allostery, where conformational shifts typically preserve secondary structure \(^{8,26}\).

**Comparison to prior data and proposals**

The model in Figure 5 is consistent with all data collected in this study as well as all measurements on the *V. vulnificus add* riboswitch published to date. First, a number of prior studies applied techniques that do not directly read out base pairing but are sensitive to ligand binding and can quantify energetics and kinetics. Functional measurements *in vitro* and *in vivo*, use of fluorescent reporters, and single-molecule force experiments all have given *apoA* frequencies of ~40% at 20–30°C. \(^{13,23,33}\) Second, X-ray crystallographic studies have been limited to constructs with just the aptameric region \(^{24,35-42}\); these structures all show P1, P2, and P3 helices, in accord with Figure 5. Third, the detailed NMR measurements that suggested the alternative *apoA* (Figure 2B) were also consistent with the *apoA* base pairings in the MWC model (Figure 2A): nuclear Overhauser effect (NOE) spectroscopy unambiguously established that pairings in both P1 and P4A occur in the absence of adenine, but could not infer whether these helices might co-occur or be mutually exclusive (Figure 2A-B). Fourth, analogous to our locking approach, the NMR study used a variant MutP2 (A29C, A30G, U40C, U41G) to stabilize the aptamer secondary structure. While our chemical mapping measurements suggest that this mutant only partially stabilizes P2 (SI Figure S5), both the NMR study and our LM\(^2\)R measurements detect P1 and no P4A in this background (SI Figure S9), supporting the MWC model in Figure 5. Fifth, our SHAPE and multi-probe chemical mapping data suggest partial opening of P5 in *apoA* and *holo* (Figure 2G-H), and our LM\(^2\)R analysis suggests only partial formation of P2 in *apoA* (Figure 3E) and partial formation of P2B in *apoB* (Figure 3F). These observations are again consistent with NMR studies which could unambiguously detect P5 and precisely measure
relative frequencies of, e.g., P2 vs. P2B, but could not establish these helices’ absolute frequencies. Sixth, newer single molecule FRET measurements detect transient formation of the L2/L3 tertiary contact prior to adenine binding in apoA and stabilization upon adenine binding to holo\textsuperscript{27}. These data are consistent with, and indeed were interpreted within, an MWC framework in which formation of the aptameric secondary structure occurs in concert with opening of the gene expression platform, even without adenine to shift the equilibrium in favor of the ‘ON’ state. Seventh, the prior detailed NMR analysis led to a compelling non-MWC model of riboswitch temperature robustness that involves compensating the improving affinity of the aptamer for adenine at colder temperatures by a pre-equilibrium that favored apoB over apoA at those temperatures. Although the precise predictions of the switching efficiencies of this model are affected by the interpretation of whether apoA is ON or OFF, approximately the same temperature compensation occurs in the MWC conformational selection model of Figure 5 (SI Figure S11).

There is now concordance across numerous measurements for an MWC model that posits strong secondary structure similarities between apoA and holo. We also propose herein that further heterogeneity is likely to be present in apoB, the ‘OFF’ state, as LM\textsuperscript{2}R measurements with different helix stabilizers gave different results for this state, suggesting incomplete correlation of the tested ‘signature’ helices (P1B and P4A). Numerous alternative secondary structures are possible for this state (see, e.g., SI Figure S4). The presence of these myriad helices, each at low population, would explain their detection difficulty with this and prior bulk equilibrium techniques but their appearance in single-molecule measurements.\textsuperscript{19} Supporting this picture, higher-order M\textsuperscript{2} analysis in single mutant backgrounds (SI Figure S12) isolate and confirm numerous alternative apoB structures. We note that there are no functional reasons for apoB to maintain a single secondary structure – it simply has to disallow adenine binding while keeping the gene expression platform closed in P4. Without selection for a pure single structure, we therefore suggest that apoB has remained structurally heterogeneous, and the exact populations of its helix pairings outside P4 may shift in different solution conditions and flanking sequences while still being consistent with riboswitch function. We further speculate that ‘non-functional’ states of other riboswitches and RNA gene regulatory modules may have highly heterogeneous structures and, indeed, this feature might explain why those states have been refractory to conventional structural biology approaches developed primarily to dissect protein structure/function. This picture also implies that antibiotics targeting any specific apo structure of the add and other riboswitches are unlikely to succeed, as single mutations that disrupt the
targeted structure while still maintaining the apo ensemble should be easy to find and would offer resistance.

**General applicability of 4D RNA chemical mapping**

The next challenge for understanding riboswitches and other cis-regulatory RNA elements is to test how the structural ensembles defined through *in vitro* studies are retained or altered by co-transcriptional effects, protein binding, helicases, crowding, noise due to low numbers of molecules, and other complexities of these molecules’ native biological environments. Amongst available biochemical and biophysical techniques, chemical mapping methods read out by sequencing have unusual promise in delivering single-nucleotide-resolution structural information in such environments. As techniques improve to edit genomes across organisms and to amplify and measure chemical mapping signals *in cellulo*, the lock-mutate-map-rescue approach developed here offers the possibility of dissecting complex structural ensembles for these shape-shifting molecules.

**ACKNOWLEDGMENTS**

We thank M. Ali, S. Doniach, and C. VanLang for early discussions and experiments on the add riboswitch. We thank Drs. H. Schwalbe and B. Fürtig for personal communication regarding temperature compensation and NMR analysis and sharing a manuscript before publication. We acknowledge financial support from NIH R01 GM102519 (to R.D.).

**METHODS**

**RNA Synthesis and Construct Design**

Double-stranded DNA templates were prepared by PCR assembly of DNA oligomers with maximum length of 60 nt ordered from IDT (Integrated DNA Technologies). DNA templates contain a 20-nt T7 RNA polymerase promoter sequence (TTCTAATACGACTCACTATA) on the 5’ end and a 20-nt Tail2 sequence (AAAGAACAACAAACAACAAAC) on the 3’ end. The sequence of interest is flanked by one hairpin with single-stranded buffering region on each end. The primer assembly scheme and plate orders for all constructs were automatically designed by Primerize-2D.

PCR reactions, including 100 pmol of terminal primers and 1 pmol of internal primers were carried out as previously described. PCR products were purified using Ampure XP magnetic beads (Agencourt) on a 96-well microplate format following manufacturer’s instructions. DNA
concentrations were measured on a Nanodrop 1000 spectrophotometer (Thermo Scientific). *In vitro* transcription reactions were described previously, followed by similar purification (using Ampure XP beads with externally added 10% PEG-8000) and quantification steps.

**Chemical Modification**

M², M²R and LM²R chemical mapping were carried out in 96-well format as described previously. 29, 31 Prior to chemical modification, 1.2 pmol of RNA was heated to 90 °C for 2 min and cooled on ice for 2 min to remove secondary structure heterogeneity, then folded for 20 min at 37 °C in 15 μL of 10 mM MgCl₂, 50 mM Na-HEPES pH 8.0 (with or without 5 mM adenine). RNA was modified by adding 5 μL of freshly made SHAPE reagent, 5 mg/mL 1M7 (1-methyl-7-nitroisatoic anhydride) dissolved in anhydrous DMSO. Modification reactions were incubated at room temperature for 12 min and then quenched by 5 μl of 0.5 M Na-MES pH 6.0. Quenches also included 1 μL of poly(dT) magnetic beads (Ambion) and 0.065 pmol of FAM-labeled Tail2-A20 primer for reverse transcription. Samples were separated using magnetic stands, washed thoroughly with 70% ethanol, and air-dried. Beads were resuspended in 5.0 μL reverse transcription mix with SuperScript III (Thermo Fisher), then incubated at 48 °C for 30 min. RNAs were degraded by adding 5 μL 0.4 M NaOH and incubating at 90 °C for 3 min. Solutions were cooled down on ice then neutralized with 3 μL acid quench (1.4 M NaCl, 0.6 M HCl, and 1.3 M Na-acetate). Fluorescent labeled cDNA was recovered by magnetic bead separation, rinsed 70% ethanol, and air-dried. The beads were resuspended in 10 μL Hi-Di formamide (Applied Biosystems) with 0.0625 μL ROX-350 ladder (Applied Biosystems) and eluted for 20 min. The eluants were loaded onto capillary electrophoresis sequencers (ABI3100 or ABI3730).

Multi-probe chemical mapping was performed with procedures similar to SHAPE chemical mapping, with variations in the modification and quench steps. Specific preparations were as follows: 1% dimethyl sulfate (DMS), mixing 1 μL 10.5 M DMS into 9 μL ethanol, and then 90 μL doubly deionized water (ddH₂O); 0.4% glyoxal, dilution to 1/100x of 8.8 M glyoxal; RNase V1, serial dilution to 1/1000x in storage buffer (50% glycerol, 50 mM Tris-HCl pH 7.4, 100 mM NaCl, 0.1 mM EDTA); Terbium(III), 4 mM TbCl₃ in ddH₂O; FMN, 2 mM flavin mononucleotide in ddH₂O. Volumes of 5 μL of these modifier stocks were added to folded RNA solution for 12 min, except for UV treatment, in which samples were exposed directly under a hand-held 302-nm UV lamp for 3 min. FMN phot oxidation reactions were placed on a visible-light box during the entire 12 min reaction. Modifications were quenched by 5 μL of 2-mercaptoethanol for 1M7,
DMS, glyoxal and RNase V1; 72 mM EDTA for Terbium(III); or ddH2O (and removal of light source) for ‘no mod’ controls, FMN, and UV.

**CE Data Processing**

The Hit RACE 2.0 software was used to analyze CE (capillary electrophoresis) data. Electrophoretic traces were aligned and baseline subtracted using linear and non-linear alignment routines as previously described. Sequence assignment was accomplished semi-automatically with human supervision. Band intensities were obtained by fitting profiles to Gaussian peaks and integrating. Normalization, correction for signal attenuation, and background subtraction were enabled by inclusion of referencing hairpin loop residues (GAGUA) at both 5’ and 3’ ends, 10x dilution replicates, and no-modification controls. Briefly, true values for saturated peaks were obtained from 10x dilutions. Signal attenuation was corrected from 5’ to 3’ ends based on the relative reactivity between 5’ and 3’ referencing hairpin loop intensities. Reactivities of SHAPE profiles were normalized against GAGUA, while other modifiers in multi-probe mapping were normalized to subsets of GAGUA which are reactive to that particular modifier.

**Data Deposition**

All chemical mapping datasets, including M², M²R, LM²R and multi-probe mapping, have been deposited at the RNA Mapping Database (http://rmdb.stanford.edu) under the following accession codes: ADD140_1M7_0001, ADD140_1M7_0002, ADD140_1M7_0003, ADD140_1M7_0004, ADD140_1M7_0005, ADD140_1M7_0006, ADD140_1M7_0007, ADD140_1M7_0008, ADD140_1M7_0009, ADD140_1M7_0010, ADD140_1M7_0011, ADD140_1M7_0012, ADD140_1M7_0013, ADD140_1M7_0014, ADD140_RSQ_0001, ADD140_RSQ_0002, ADD140_RSQ_0003, ADD140_RSQ_0004, ADD140_LCK_0001, ADD140_LCK_0002, ADD140_LCK_0003, ADD140_LCK_0004, ADD140_LCK_0005, ADD140_LCK_0006, ADD128_STD_0001, ADD128_STD_0001, ADD140_STD_0001, ADD140_DCP_0001, ADD140_DCP_0002, RNAPZ5_RSQ_0001, TRP4P6_RSQ_0001, 16SFWJ_RSQ_0001.

**LEGENDS**

Table 1. Helix frequency estimate results for single base-pair M²R
Median helix frequencies from Rfam simulations are reported along with 90% confidence range from the same bin. The rescue factor for each helix is averaged across all tested base pairs.

**Table 2. Helix frequency estimate results for double base-pair M²R.**

Median helix frequencies from Rfam simulations are reported along with 90% confidence range from the same bin. The rescue factor for each helix is averaged across all tested base pairs.

**Table 3. Helix frequency estimate results for LM²R constructs and MutP2 from previous study.**

Median helix frequencies from Rfam simulations are reported along with 90% confidence range from the same bin. The rescue factor for each helix is averaged across all tested base pairs.

**Figure 1. Schematic of four-dimensional chemical mapping for RNA structural ensembles, the lock-mutate-map-rescue (LM²R) method.**

(A) Secondary structure ensemble is dissected by locking one helix in place by double base-pair mutations (green circles). The co-existence of a helix in question (cyan line) is then tested by single base-pair mutations (red circles) M²R in the locked ensemble.

(B) Base-pairing probability plot and example quartet of LM²R result. Positions of locked mutations and M²R pair are marked by horizontal and vertical lines. The helix frequency estimate is colored in gray scale. An example SHAPE profile quartet shows the perturbations and rescue effects.

(C) Correlation of observed rescue factor against helix frequency for helices predicted with length longer than 2-bp from 325 Rfam families mutated and folded in silico. Each data point represents a helix whose rescue factor and simulated helix frequency has been averaged across all its base-pairs. Helices are colored by their length. The rescue factor estimates the extent of similarity restored in a double mutant’s reactivity compared to WT (wild type), scaled by perturbations observed in single mutants.

(B) In silico M²R quartets illustrating full compensatory rescue (high rescue factor), partial rescue (mid rescue factor) and no rescue (low rescue factor).

**Figure 2. MWC and non-MWC model of add riboswitch.**

(A-B) In both models, apoB state (left) is OFF (Shine-Dalgarno sequence and start codon sequestered) and also incompetent for ligand binding due to the aptamer’s rearrangement. The ligand-bound state holo (right) is ON with Shine-Dalgarno sequence and AUG codon (gene
expression platform) exposed. The \textit{apoA} state (center), defined as the conformations with an aptamer fold ready for ligand binding, differ in the two models. In the Monod-Wyman-Changeux model (A), \textit{apoA} is assumed to be ON, with a similar expression platform to \textit{holo}, while in the non-MWC model proposed in an NMR study (B), the \textit{apoA} state is proposed to be OFF, with the expression platform sequestered into helices P4B and P4C. (C-D) One-dimensional SHAPE reactivity of \textit{add} construct (nts 13-140) in the absence or presence of ligand 5 mM adenine. (A-D) Locations of helices in the MWC and non-MWC models are marked as follows: For the aptamer domain, OFF-state helices are P1B and P2B (blue); ON-state helices are P1 and P2 (gold). For the expression platform, OFF-state helices are P4A-C (blue). Helices P3 and P5 (brown) are shared across all models. The switch region, anti-switch region, Shine-Dalgarno sequence and AUG codon are labeled by the same patterns in (A-D).

**Figure 3. Mutate-map-rescue (M²R) evaluates ensemble helix frequency in 5 mM adenine (A.C,E,G) and no adenine (B.D,F,H).**

(A-B) Mutate-and-map (M²) datasets of \textit{add} riboswitch. Mutants that show anti-correlation of SHAPE reactivity pattern in the aptamer domain and expression platform are marked by the orange arrows. (C-D) Representative \textit{in vitro} M²R quartets of helices P1 and P4A. In each quartet, four SHAPE profiles are shown from left (5’) to right (3’): WT, mutant A, mutant B, and compensatory double mutant AB. Flanking sequences used for SHAPE reactivity normalization are not shown. (E-F) Base-pairing probability plot of single base-pair M²R on \textit{add} riboswitch. Expected helix frequencies (100\% \textit{holo} for 5 mM adenine; 60\% \textit{apoB} + 40\% \textit{apoA} for no adenine) are shown in the bottom-left half, and helix-wise estimates by M²R are shown in the top-right half. (G,H) Cumulative distributions of posterior probabilities of each helix frequency, given measured rescue factors.

**Figure 4. Lock-mutate-map-rescue (LM²R) dissects helix frequency in locked ensemble.**

(A-B) One-dimensional SHAPE reactivity of lock-P1, lock-P2, lock-P4A and lock-P1B constructs compared to WT in the absence of ligand adenine. Targeting the \textit{apoA} state, lock-P1 and lock-P2 exhibited similar SHAPE reactivity profiles (except for partial opening in P2 region of lock-P1), and show higher reactivity in the expression platform compared to WT. Targeting the \textit{apoB} state, lock-P4A and lock-P1B also exhibited similar reactivity patterns (except for partial opening in P1B region and more stringent protection of the expression platform in lock-P4A) and gave lower reactivities in the expression platform compared to WT.
(C) Base-pairing probability plot of LM²R tests in lock-P1 and lock-P2 constructs. Compensatory rescue in either lock background recover high frequencies for helices P1 and P2 but not for P4A and P4B.

(D) Base-pairing probability plot of LM²R tests in lock-P4A and lock-P1B constructs. Compensatory rescue experiments for P4A and P4B give high helix frequency in either lock background but not for P1 and P2. In lock-P4A, helices P1B and P2B have low helix frequency as well, suggesting heterogeneity in aptamer domain folding in the absence of ligand. Helices locked in place are highlighted by green boxes; helices not tested herein are in dashed boxes.

(E,F) Cumulative distributions of posterior probabilities of each helix frequency, given measured rescue factors, in each of the four locked-helix backgrounds.

Figure 5. Current model for the add structural ensemble.
The proposed model favors the MWC model of allostery. Sampling of the aptamer secondary structure in the absence of ligand (apoA) correlates with opening of the P4 domain and gene expression, analogous to ligand-bound holo. The ligand-free apoB state sequesters the Shine-Dalgarno sequence and AUG codon in the P4 domain, precluding gene expression. Our measurements and other studies are consistent with additional alternative structures in the apoB state (dashed blue lines) and stabilization of the aptamer through coupled tertiary contacts upon adenine binding in the holo state (solid yellow lines).

REFERENCES
1. Henkin, T. M., Riboswitch RNAs: Using RNA to sense cellular metabolism. Genes Dev. 2008, 22 (24), 3383-3390.
2. Nudler, E.; Mironov, A. S., The riboswitch control of bacterial metabolism. Trends Biochem. Sci. 2004, 29 (1), 11-17.
3. Gesteland, R. F.; Cech, T.; Atkins, J. F., The RNA world: The nature of modern RNA suggests a prebiotic RNA world. 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2006; p xxiii, 768 p.
4. Yanofsky, C., Transcription Attenuation: Once Viewed as a Novel Regulatory Strategy. Journal of Bacteriology 2000, 182 (1), 1-8.
5. Gutiérrez-Preciado, A.; Henkin, T. M.; Grundy, F. J.; Yanofsky, C.; Merino, E., Biochemical Features and Functional Implications of the RNA-Based T-Box Regulatory Mechanism. Microbiology and Molecular Biology Reviews 2009, 73 (1), 36-61.
6. Monod, J.; Wyman, J.; Changeux, J.-P., On the nature of allosteric transitions: A plausible model. Journal of Molecular Biology 1965, 12 (1), 88-118.
7. Gunasekaran, K.; Ma, B.; Nussinov, R., Is allostery an intrinsic property of all dynamic proteins? Proteins 2004, 57 (3), 433-43.
8. Changeux, J. P.; Edelstein, S. J., Allosteric mechanisms of signal transduction. Science 2005, 308 (5727), 1424-8.
9. Fernández-Luna, M. T.; Miranda-Ríos, J., Riboswitch folding: One at a time and step by step. RNA Biol. 2008, 5 (1), 20-23.
10. Gilbert, S. D.; Battey, R. T., Riboswitches: Fold and function. Chem. Biol. 2006, 13 (8), 805-807.
11. Breaker, R. R., Riboswitches and the RNA world. Cold Spring Harb. Perspect. Biol. 2012, 4 (2), a003566.
12. Jones, C. P.; Ferre-D'Amare, A. R., Long-Range Interactions in Riboswitch Control of Gene Expression. Annu Rev Biophys 2017, 46, 455-481.
13. Lemay, J.-F.; Desnoyers, G.; Blouin, S.; Heppell, B.; Bastet, L.; St-Pierre, P.; Massé, E.; Lafontaine, D. A., Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. PLoS Genet. 2011, 7 (1), e1001278.
14. Mandal, M.; Breaker, R. R., Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat. Struct. Mol. Biol. 2004, 11 (1), 29-35.
15. Battey, R. T., Structure and mechanism of purine binding riboswitches. Q. Rev. Biophys. 2012, 45 (3), 345-381.
16. Cordero, P.; Das, R., Rich structure landscapes in both natural and artificial RNAs revealed by mutate-and-map analysis. PLoS Biol. 2015, 11 (11), e1004473.
17. Stoddard, C. D.; Widmann, J.; Trausch, J. J.; Marciano-Velázquez, J. G.; Knight, R.; Battey, R. T., Nucleotides adjacent to the ligand-binding pocket are linked to activity tuning in the purine riboswitch. J. Mol. Biol. 2013, 425 (10), 1596-1611.
18. Rieder, R.; Lang, K.; Graber, D.; Micura, R., Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. ChemBioChem 2007, 8 (8), 896-902.
19. Neupane, K.; Yu, H.; Foster, D. A. N.; Wang, F.; Woodside, M. T., Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. Nucleic Acids Res. 2011, 39 (17), 7677-7687.
20. Lemay, J.-F.; Penedo, J. C.; Tremblay, R.; Lilley, D. M. J.; Lafontaine, Daniel A., Folding of the adenine riboswitch. Chem. Biol. 2006, 13 (8), 857-868.
21. Greenleaf, W. J.; Frieda, K. L.; Foster, D. A. N.; Woodside, M. T.; Block, S. M., Direct observation of hierarchical folding in single riboswitch aptamers. Science 2008, 319 (5863), 630-633.
22. Wickiser, J. K.; Cheah, M. T.; Breaker, R. R.; Crothers, D. M., The kinetics of ligand binding by an adenine-sensing riboswitch. Biochemistry 2005, 44 (40), 13404-13414.
23. Lemay, J.-F.; Lafontaine, D. A., Core requirements of the adenine riboswitch aptamer for ligand binding. RNA 2007, 13 (3), 339-350.
24. Stagno, J. R.; Liu, Y.; Bhandari, Y. R.; Conrad, C. E.; Panja, S.; Swain, M.; Fan, L.; Nelson, G.; Li, C.; Wendel, D. R.; White, T. A.; Coe, J. D.; Wiedorn, M. O.; Knoska, J.; Oberthuer, D.; Tuckey, R. A.; Yu, P.; Dyba, M.; Tarasov, S. G.; Weierstall, U.; Grant, T. D.; Schwieters, C. D.; Zhang, J.; Ferre-D’Amare, A. R.; Fromme, P.; Draper, D. E.; Liang, M.; Hunter, M. S.; Boutet, S.; Tan, K.; Zuo, X.; Ji, X.; Barty, A.; Zatspin, N. A.; Chapman, H. N.; Spence, J. C.; Woodson, S. A.; Wang, Y. X., Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. Nature 2017, 541 (7636), 242-246.
25. Reining, A.; Nozimov, S.; Schlepckow, K.; Buhr, F.; Furtig, B.; Schwalbe, H., Three-state mechanism couples ligand and temperature sensing in riboswitches. Nature 2013, 499 (7458), 355-359.
26. Eaton, W. A.; Henry, E. R.; Hofrichter, J.; Mozzarelli, A., Is cooperative oxygen binding by hemoglobin really understood? Nat Struct Mol Biol 1999, 6 (4), 351-358.
27. Warhaut, S.; Mertinkus, K. R.; Hollthaler, P.; Furtig, B.; Heilemann, M.; Hengesbach, M.; Schwalbe, H., Ligand-modulated folding of the full-length adenine riboswitch probed by NMR and single-molecule FRET spectroscopy. Nucleic Acids Res. 2017.
28. Kladwang, W.; Cordero, P.; Das, R., A mutate-and-map strategy accurately infers the base pairs of a 35-nucleotide model RNA. RNA 2011, 17 (3), 522-534.
29. Kladwang, W.; VanLang, C. C.; Cordero, P.; Das, R., A two-dimensional mutate-and-map strategy for non-coding RNA structure. Nat. Chem. 2011, 3 (12), 954-962.
30. Miao, Z.; Adamiak, R. W.; Blanchet, M.-F.; Boneicki, M.; Bujnicki, J. M.; Chen, S.-J.; Cheng, C.; Chojnowski, G.; Chou, F.-C.; Cordero, P.; Cruz, J. A.; Ferré-D’Amare, A. R.; Das, R.; Ding, F.; Dokholyan, N. V.; Dunin-Horkawicz, S.; Kladwang, W.; Krokhotin, A.; Lach, G.; Magnus, M.; Major, F.; Mann, T. H.; Masquida, B.; Matelska, D.; Meyer, M.; Peselis, A.; Popenda, M.; Purzycka, K. J.; Serganov, A.; Stasiewicz, J.; Szachniuk, M.; Tandon, A.; Tian, S.; Wang, J.; Xiao, Y.; Xu, X.; Zhang, J.; Zhao, P.; Zok,
T.; Westhof, E., RNA-Puzzles round II: Assessment of RNA structure prediction programs applied to three large RNA structures. *RNA* **2015**, *21* (6), 1066-1084.
31. Tian, S.; Cordero, P.; Kladwang, W.; Das, R., High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA* **2014**, *20* (11), 1815-26.
32. Tian, S.; Das, R., RNA structure through multidimensional chemical mapping. *Q. Rev. Biophys.* **2016**, *49* (e7), 1-30.
33. Reining, A.; Nozinovic, S.; Schlepcow, K.; Buhr, F.; Furtig, B.; Schwalbe, H., Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature* **2013**, *499* (7458), 355-9.
34. Furtig, B.; Nozinovic, S.; Reining, A.; Schwalbe, H., Multiple conformational states of riboswitches fine-tune gene regulation. *Current opinion in structural biology* **2015**, *30*, 112-24.
35. Stoddard, C. D.; Widmann, J.; Traus, J. J.; Marcano-Velázquez, J. G.; Knight, R.; Batey, R. T., Nucleotides Adjacent to the Ligand-Binding Pocket are Linked to Activity Tuning in the Purine Riboswitch. *Journal of Molecular Biology* **2013**, *425* (10), 1596-1611.
36. Zhang, J.; Ferré-D’Amare, Adrian R., Dramatic Improvement of Crystals of Large RNAs by Cation Replacement and Dehydration. *Structure* **2014**, *22* (9), 1363-1371.
37. Gilbert, S. D.; Reyes, F. E.; Edwards, A. L.; Batey, R. T., Adaptive Ligand Binding by the Purine Riboswitch in the Recognition of Guanine and Adenine Analogs. *Structure* **2009**, *17* (6), 857-868.
38. Edwards, A. L.; Batey, R. T., A Structural Basis for the Recognition of 2′-Deoxyguanosine by the Purine Riboswitch. *Journal of Molecular Biology* **2009**, *385* (3), 938-948.
39. Serganov, A.; Yuan, Y.-R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J., Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. *Chemistry & Biology* **2004**, *11* (12), 1729-1741.
40. Daldrop, P.; Reyes, Francis E.; Robinson, David A.; Hammond, Colin M.; Lilley, David M.; Batey, Robert T.; Brenk, R., Novel Ligands for a Purine Riboswitch Discovered by RNA-Ligand Docking. *Chemistry & Biology* **2011**, *18* (3), 324-335.
41. Delfosse, V.; Bouchard, P.; Bonneau, E.; Dagenais, P.; Lemay, J.-F.; Lafontaine, D. A.; Legault, P., Riboswitch structure: an internal residue mimicking the purine ligand. *Nucleic Acids Research* **2010**, *38* (6), 2057-2068.
42. Gilbert, S. D.; Love, C. E.; Edwards, A. L.; Batey, R. T., Mutational Analysis of the Purine Riboswitch Aptamer Domain. *Biochemistry* **2007**, *46* (46), 13297-13309.
43. Leamy, K. A.; Assmann, S. M.; Mathews, D. H.; Bevilaqua, P. C., Bridging the gap between in vitro and in vivo RNA folding. *Q Rev Biophys.* **2016**, *49*, e10.
44. Kladwang, W.; Mann, T. H.; Becka, A.; Tian, S.; Kim, H.; Yoon, S.; Das, R., Standardization of RNA chemical mapping experiments. *Biochemistry* **2014**, *53* (19), 3063-3065.
45. Tian, S.; Das, R., Primerize-2D: automated primer design for RNA multidimensional chemical mapping. *Bioinformatics* **2017**.
46. Kim, H.; Cordero, P.; Das, R.; Yoon, S., HiTRACE-Web: An online tool for robust analysis of high-throughput capillary electrophoresis. *Nucleic Acids Res.* **2013**, *41* (W1), W492-W498.
47. Yoon, S.; Kim, J.; Hum, J.; Kim, H.; Park, S.; Kladwang, W.; Das, R., HiTRACE: High-throughput robust analysis for capillary electrophoresis. *Bioinformatics* **2011**, *27* (13), 1798-1805.
48. Lee, S.; Kim, H.; Tian, S.; Lee, T.; Yoon, S.; Das, R., Automated band annotation for RNA structure probing experiments with numerous capillary electrophoresis profiles. *Bioinformatics* **2015**, *31* (17), 2808-2815.
49. Cordero, P.; Lucks, J. B.; Das, R., An RNA Mapping DataBase for curating RNA structure mapping experiments. *Bioinformatics* **2012**, *28* (22), 3006-3008.
| Lock | p1 53% (35% - 62%) | 79% (72% - 91%) |
|------|------------------|------------------|
| p2   | 31% (4% - 58%)   | 64% (58% - 72%)  |
| p3   | 73% (66% - 91%)  | 56% (40% - 68%)  |
| p5   | 75% (58% - 87%)  | 29% (14% - 55%)  |
| p1b  | 44% (14% - 55%)  | 3% (2% - 9%)     |
| p2b  | 17% (2% - 55%)   | 17% (2% - 55%)   |
| p4a  | 48% (11% - 66%)  | 4% (2% - 26%)    |
| p4b  | 55% (28% - 67%)  | 4% (2% - 26%)    |
| p4c  | 68% (58% - 80%)  | 4% (2% - 23%)    |
| p6   | 6% (3% - 19%)    | 5% (2% - 16%)    |
| p8   | 9% (3% - 32%)    | 4% (2% - 26%)    |
| p9   | 12% (2% - 31%)   | 3% (2% - 15%)    |
| p10  | 3% (1% - 10%)    | 3% (1% - 15%)    |
| p11  | 12% (3% - 31%)   | 16% (3% - 29%)   |
| p12  | 10% (2% - 37%)   | 6% (2% - 27%)    |
| p13  | 9% (3% - 32%)    | 4% (2% - 26%)    |
| p14  | 4% (1% - 23%)    | 4% (2% - 15%)    |
| p15  | 12% (3% - 31%)   | 36% (12% - 51%)  |
| p16  | 14% (2% - 38%)   | 8% (3% - 31%)    |
| p17  | 4% (2% - 26%)    | 4% (2% - 16%)    |

| Lock | p1 5% (2% - 14%) | 82% (73% - 91%) |
|------|------------------|------------------|
| p2   | 6% (2% - 14%)   | 84% (79% - 97%)  |
| p3   | 64% (58% - 72%) | 84% (79% - 97%)  |
| p5   | 71% (54% - 87%) | 6% (2% - 26%)    |
| p1b  | 3% (2% - 9%)    | 3% (2% - 15%)    |
| p2b  | 4% (1% - 23%)   | 10% (2% - 37%)   |
| p4a  | 9% (3% - 32%)   | 4% (2% - 13%)    |
| p4b  | 42% (9% - 61%)  | 5% (2% - 19%)    |
| p4c  | 10% (2% - 37%)  | 3% (1% - 14%)    |
| p6   | 8% (2% - 24%)   | 6% (3% - 19%)    |
| p13  | 4% (2% - 12%)   | 8% (2% - 30%)    |

Table 1

| Lock | p2 42% (21% - 69%) |
|------|-------------------|
| p4b  | 4% (2% - 12%)     |

| Lock | p1 76% (66% - 88%) |
|------|-------------------|
| p4a  | 4% (2% - 16%)     |
| p4b  | 4% (2% - 12%)     |

| Lock | p1b 9% (2% - 19%) |
|------|------------------|
| p2b  | 6% (2% - 27%)    |
| p2   | 6% (2% - 14%)    |
| p4b  | 55% (28% - 67%)  |

| Lock | p4a 55% (28% - 67%) |
|------|-------------------|
| p4b  | 80% (30% - 88%)   |
| p3   | 82% (69% - 93%)   |

| Lock | p1b 8% (3% - 31%) |
|------|------------------|
| p2b  | 4% (1% - 23%)    |
| p1   | 5% (2% - 14%)    |
| p2   | 6% (2% - 14%)    |
| p4a  | 47% (7% - 75%)   |

| Mut | p1 30% (1% - 62%) |
|-----|------------------|
| p4a | 8% (1% - 41%)    |
| p4b | 13% (2% - 56%)   |

Table 2

| LM2R |
|------|
| p1   |
| p2   |
| p3   |

Table 3
Figure 2

(A) 

(B) 

(C) 

(D)
Supplemental Information for “Allosteric Logic of the V. vulnificus Adenine Riboswitch Resolved by Four-dimensional Chemical Mapping”

Siqi Tian¹, Wipapat Kladwang¹, and Rhiju Das¹,²,*

¹Department of Biochemistry, and ²Department of Physics, Stanford University, Stanford, California 94305, United States

* Corresponding Author

This Supplemental Information document contains the following sections:

Supporting Results ........................................................................................................................................... 2
  Mutate-Map-Rescue (M²R) – Application Generality and Automated Classification ...................................... 2
  Current Limitations of M²R and Classifier ....................................................................................................... 3
  Adenine riboswitch construct design .............................................................................................................. 4
  Multi-probe chemical mapping tests partitioning predicted by the MWC model .............................................. 4
Supporting Methods ............................................................................................................................................ 5
  In silico Partition, Autoscore Classifier and Rfam Sampling ......................................................................... 5
  Structural Equilibrium Fitting ......................................................................................................................... 6
SI Figures ............................................................................................................................................................ 8
Supporting References ....................................................................................................................................... 11
Supporting Results
Mutate-Map-Rescue (M²R) – Application Generality and Automated Classification

Before interrogating add riboswitch ensembles, we sought a method that would enable rapid and unbiased readout of helix frequencies. Our prior work on mutate-map-rescue (M²R) tested RNA structures by disrupting a base pair by single mutation, and testing whether compensatory double mutations restore such perturbations, but relied on manual inspection to assess rescue, and focused on a single model system, a four-way junction domain of *E. coli* 16S rRNA.¹ We first sought to assess the generality of M²R approach by applying it to other RNA elements. First, we deployed M²R on the lariat-capping GIR1 ribozyme,² to test helix P5 from crystallographic model against its alternative version (alt-P5), which was predicted by our M²-based model with a 2-register shift (Figure 2B). For each base pair tested, we compared the SHAPE profiles of WT (wild type), 2 single mutants, and the double mutant (rescue) in a ‘quartet’. With the local perturbations caused by single mutants in the P5/alt-P5 region, double mutants based on the P5 successfully restored the SHAPE profile to WT, while double mutants based on the alt-P5 failed to rescue. Thus, M²R provided evidence for the P5 helix but not the register-shifted alt-P5, consistent with the crystallographic structure of the ribozyme. In addition, we applied M²R on the P4-P6 domain of the *Tetrahymena* ribozyme³ to test for presence of crystallographic P5c against an alternative suggested by mutate-and-map analysis (SI Figure S1). Finally, M²R-validated secondary structure models were confirmed experimentally on two RNA domains by functional assays in vitro or in cellulo: a human *HoxA9* mRNA IRES (internal ribosomal entry site) domain ⁴ and a stem loop domain from Influenza A virus ⁵.

In order to evaluate the accuracy of M²R on a larger scale, we turned to in silico simulation to generate a much greater number of RNA test cases. Specifically, we asked: do M²R-validated base pairs have high base-pairing probability (BPP) in silico? To answer this question, we sought to automate the calling of M²R result with an automated classifier that would evaluate M²R data with results matching human inspection, and this study eventually led to the definition of the ‘rescue factor’ applied to the adenine riboswitch. Our first scoring function took into account the following two factors: 1) the amount of perturbation seen in each of the two single mutants, compared to wild type profiles; and 2) the similarity (rescuing effect) between double mutant and WT. We captured the latter effect through a rescue factor metric:

\[
\frac{\text{diff}(\text{WT},\text{AB})}{\max(\text{diff}(\text{WT},A),\text{diff}(\text{WT},B))}
\]

where
and $d_A$ and $d_B$ are vectors of normalized SHAPE profiles, with the same length of $N$ (see Supporting Methods for further details). The classifier returned a result from three categories: Validated, Falsified, or Uncertain; the last one was assigned for cases that either 1) single mutants failed to introduce discernible perturbations, or 2) the rescuing effect was ‘half-way’ and hard to assign. After training with the *in vitro* $M^2R$ data on the GIR1, P4-P6, and IRES measurements as well as *in silico* simulated counterparts, we obtained a classifier that recovered human expert calls (SI Figure S2).

Next, we simulated $M^2R$ *in silico* on 325 RNA families from the Rfam database whose lengths were in range between 100 and 250 nt. More than 37,000 base pairs were tested from helices that were longer than 2 bp and with predicted base pairing probability (BPP) greater than 1%. SI Figure S13 shows the BPP distribution of the 4-category classification. These results confirmed that the Validated and Falsified categories corresponded to base pairs with high and low BPP. Unexpectedly, we also observed a quantitative correlation between the *in silico* predicted BPP and the ‘rescue factor’ metric that underlies our classification, even for base pairs that appear with BPP frequencies between 0.3 and 0.7 (Figure 2A). Furthermore, for helices of longer length, this relationship became more well-defined. We reasoned that this correlation would enable estimation of BPP based on helix length and the ‘rescue factor’ metrics measured and averaged over all the base pairs for a given helix (SI Figure S13), and that this quantitative metric would be more informative than a qualitative 4-category classification. Further support for using the ‘rescue factor’ to estimate a posterior probability for tested helix frequencies arose from our studies on the adenine riboswitch with and without adenine, which could be compared to ‘gold standard’ helix frequency estimates from NMR analysis, as described in the main text.

**Current Limitations of $M^2R$ and Classifier**

We trained an automated classifier based on previous $M^2R$ *in vitro* data combined with *in silico* simulations. The classifier was later independently tested on *add* single-base-pair $M^2R$, double-base-pair $M^2R$, and LM$^2R$ datasets, showing agreement with calling by an expert human referee (S.T.) (Figure 3A). Moreover, the BPP distribution of the simulated *in silico* counterparts for *add* riboswitch follows the same trend as Rfam tests (SI Figure S14), in agreement with its generality. The performance of such a simple classifier was acceptable and supported our heuristic for the rescue factor (SI Figure S13). Our preliminary attempt on 4-category classification gave conservative conclusions for $M^2R$ quartets, with ‘Partially Validated’ as weaker cases for incomplete compensatory rescue. For final analyses, we chose to
present final rescue factors and corresponding helix frequencies (see, e.g., main text Figures 3 and 4); the classifier based on the rescue factor gives a more qualitative picture of the results, but allows simple visual checks that for each helix, different base pairs tested for compensatory rescue give concordant ‘calls’ for the frequency of the helix (Figure S14).

Adenine riboswitch construct design
We performed the M\textsuperscript{2}R pipeline on a 128-nt \textit{add} riboswitch construct, which has 15 extra nucleotides on the 3’ end into the coding sequence compared to a 112-nt one used in a recent study.\textsuperscript{7} We chose this longer \textit{add} version after reproducing prior work in our laboratory\textsuperscript{8-9} finding that the sequence context has an effect on the folding landscape of this riboswitch (SI Figure S5A). Although the 112-nt construct showed weak ligand-responsive chemical reactivity changes in the SD region under the salt conditions used in NMR \textsuperscript{9}, it does not show noticeable switching under our in vitro solution conditions. The extended construct gives chemical reactivity changes under all conditions tested. Besides M\textsuperscript{2} (SI Figure S5B-C), we performed single-base-pair M2R on the \textit{add} riboswitch (Figure 3, SI Figure S6).

Multi-probe chemical mapping tests partitioning predicted by the MWC model
As discussed in the main text introduction, the MWC model requires that, even in the absence of adenine ligand, the \textit{add} riboswitch samples two states: \textit{apoA} and \textit{apoB}, that have secondary structures compatible and incompatible with aptamer adenine binding, respectively, and that these states are more likely or less likely, respectively, to have free ribosome binding sites. We tested this model with the aid of the lock-P1 and lock-P4A mutants, which isolated the \textit{apoA} and \textit{apoB} states, respectively, and chemical mapping to monitor the riboswitch structural ensemble at nucleotide resolution. To stringently test the prediction, we expanded our chemical mapping protocol to include five additional modifiers beyond SHAPE: DMS (dimethyl sulfate) \textsuperscript{10}, glyoxal \textsuperscript{11}, RNase V1 \textsuperscript{12}, terbium(III) \textsuperscript{13-14}, and ultraviolet irradiation at 302 nm \textsuperscript{15}, giving a total number of 1920 measurements measured in three sequences that would test a two-state partitioning. As expected from the model of \textit{apoA} vs. \textit{apoB}, these multiple modifier profiles varied significantly between the lock-P1 and lock-P4A mutants, especially in the expression platform regions (see, e.g., nts 98-112). Nevertheless, as predicted from the MWC model, a simple linear combination of the profiles, assuming 48\% of the \textit{apoA} state and 52\% of the \textit{apoB} state recovered the data measured for the adenine-free wild type riboswitch within experimental errors, except at nucleotides 79-81 (SI Figure 7A-B). We note that there was only one parameter optimized in this fit; scaling of each data profile was carried out based on well-defined flanking hairpins as normalization standards.\textsuperscript{16} These values for the \textit{apoA} vs. \textit{apoB} state frequencies agree within errors with helix frequencies measured above in compensatory rescue experiments as well as with the prior NMR experiment at similar temperature (40\%}
apoA at 30 °C)\(^{17}\). However, the discrepancies at nucleotides 79-81 were reproducible, suggesting that there are additional secondary structures, potentially many at low frequencies, beyond those detected here in the ligand-free ensembles.

Supporting Methods

**In silico** Partition, Autoscore Classifier and Rfam Sampling

**In silico** RNA SHAPE profiles were simulated using the *partition* executable from the RNAstructure package version 5.6\(^{18-19}\). The resulting pair-wise probability matrices was projected into a one-dimensional vector to get per-residue base pairing probabilities (BPP). For M\(^2\)R quartet simulations, the targeted base pair was mutated to G-C or C-G pairs (i.e., the ‘Stable’ library from Primerize-2D\(^{20}\)). Flanking sequences (5’ or 3’) were not included for simulation.

The autoscore classifier utilizes a ‘ratio’ metric for classification. Normalized SHAPE profiles are used as input. First, the difference between two SHAPE profiles are calculated by:

\[
diff_{(A,B)} = \frac{1}{N} \sum_{i=1}^{N} (d_A - d_B)^2
\]

Where \(d_A\) and \(d_B\) are vectors of normalized SHAPE reactivity values, with the same length of \(N\).

For a given quartet of SHAPE profiles (WT, A, B, AB), we first determined whether there was pronounced perturbation introduced by either single mutants (A and B) by calculating:

\[
\max\left(\text{diff}_{(WT,A)}, \text{diff}_{(WT,B)}\right)
\]

(1) If the value of (1) is less than a *CUTOFF*, the Uncertain class is assigned. Otherwise, continue to determine the amount of rescuing effect by the double mutant (AB):

\[
\frac{\text{diff}_{(WT,AB)}}{\max(\text{diff}_{(WT,A)}, \text{diff}_{(WT,B)})}
\]

(2)
[2] defines the ‘ratio’ metric, i.e., the ratio between the distance of AB to WT and the maximum distance of A or B to WT. If the value of [2] is less than LOW, the Validated class is assigned due to the relatively small residual of difference of AB to WT compared to single mutants. If the value of (2) is greater than HIGH, the Falsified class is assigned. For values of (2) between LOW and HIGH, the Partial class is assigned, which captures cases that show incomplete rescue or a mixed result.

The classifier was trained on in silico and in vitro data of GIR1, P4P6, 16S-FWJ, HoxA9 4 and PSL2 (Hagey et al., manuscript in preparation), a total of 324 quartets. Data were cropped to the region of interest (i.e., excluding the flanking sequences); in silico simulated reactivity profiles were used directly, while in vitro data were attenuation-corrected and normalized by internal standards 16. For each quarter, a manual score in 1 of the 4 categories was available from expert inspection (by S.T.). The set of parameters CUTOFF (0.1), LOW (0.4), and HIGH (0.7) were then determined by grid search, optimizing for recovery of the expert assessments. The classifier was then tested on in silico and in vitro data for the add riboswitch using the chosen parameters. Both training and test data compose of Falsified vs. Validated cases in a ~2.8:1 ratio.

RNA families with length between 100 and 250 nt were screened from the Rfam database (http://rfam.xfam.org/) 11.0 FASTA file 21. A sequence from each family was randomly picked as representative, and its in silico BPP matrix was simulated. Base pairing positions to be tested by M²R were selected from this BPP matrix, for BPP values greater than or equal to 1%, and in helices that are 3 bp or longer. For each helix, the averaged ‘ratio’ metric and BPP across all simulated M²R quartets in the helix were calculated. Collectively, all simulated Rfam helices were visualized by first grouping the ‘ratio’ values into bins with width of 0.05, and then on boxplots of their BPP distribution within each bin. Intervals of 95% confidence are shown as black whiskers representing the 2.5 – 97.5 percentile of each bin, and boxes representing quartiles. Helix-wise average ‘ratio’ for helices with in vitro M2R data were similarly calculated. The in vitro BPP for those helices are assigned as: for GIR1 and P4P6, 1.0 or 0 based on their existence in crystallographic structures; for 16S-FWJ, 1.0, 0.8, 0.2 or 0 based on the prior M²R results 22. In vitro BPP of add helices in vitro were estimated by the median Rfam BPP of the ‘ratio’ bin the query helix falls into.

Structural Equilibrium Fitting
Equilibrium fractions of each state were determined by assuming that lock-P1 and lock-P4A stabilizers completely stabilize the corresponding structure – their reactivity profiles therefore represent the
reactivity profile for each state. The reactivities of lock-P1 and lock-P4A were taken to fit the WT reactivity by $\chi^2$ score, which is calculated as follows:

$$
\chi^2 = \sum_i \frac{(d_{WT} - d_{PRED})^2}{\sigma_{WT}^2 + \sigma_{PRED}^2}
$$

$$
d_{PRED} = \alpha \cdot d_{apoA} + (1 - \alpha) \cdot d_{apoB}
$$

$$
\sigma_{PRED}^2 = \alpha^2 \cdot \sigma_{apoA}^2 + (1 - \alpha)^2 \cdot \sigma_{apoB}^2
$$

Where $d_{WT}$, $d_{apoA}$, and $d_{apoB}$ are mean SHAPE reactivity profiles of WT, lock-P1, and lock-P4A, and $\sigma_{WT}$, $\sigma_{apoA}$, and $\sigma_{apoB}$ are errors (standard deviation) of $d_{WT}$, $d_{apoA}$, and $d_{apoB}$. The parameter $\alpha$ is the fraction of the $apoA$ state, ranging from 0 to 1. $\chi^2$ is summed over all or a subset of nucleotide positions $i$ as specified below. $\chi^2$ scores are plotted against $\alpha$, and the $\alpha$ value with minimum $\chi^2$ is taken as the best fit. To prevent a single nucleotide position from dominating the fit by their extreme values, nucleotides 95 (SHAPE), 66 (DMS), and 85 (glyoxal) were excluded.
SI Figures

**Figure S1. Correlation between rescue factor and helix frequency from Rfam simulation.**

*(A)* Boxplot of all helices. The purple lines mark median, open boxes cover 25 and 75 percentile, cyan whiskers show 5 and 95 percentile, and gray plus signs label outliers. Rescue factor numbers are colored in red, yellow, or green, partitioned by 0.3 and 0.7.

*(B)* Boxplot of helices by their length: 3, 4, 5, 6, and ≥ 7 base-pairs.

**Figure S2. M²R quartets and helices tested for P4-P6 and GIR1.**

*(A)* M²R quartets of P4-P6 domain testing P5C, alt-P5C, and P6. Rescue factor for each quartet is given in the title and colored as in Figure S1.

*(B-C)* Secondary structures of P4-P6 and GIR1 highlighting tested helices.

*(D)* M²R quartets of GIR1 ribozyme testing P5, alt-P5, P4, and pk2-5.

**Figure S3. One-dimensional SHAPE profiles of add riboswitch constructs with different lengths.**

Three different constructs were probed by SHAPE (1M7): *(top)* a 71-nt aptamer only construct (nts 13-83); *(middle)* a 113-nt construct (nts 13-125) including the expression platform, matching the previous NMR study; *(bottom)* a 128-nt construct (nts 13-140) further including the first 15 nt of the coding region. Reactivities are normalized based on GAGUA penta-loop reference hairpins (not shown). The 113-nt construct did not show significant switching for the Shine-Dalgarno sequence (nts 112-114) or AUG start codon (nts 120-122). See also reference. The ligand-dependent reactivity changes in the aptamer domain are shared across all three constructs.

**Figure S4. Candidate alternative helices from REEFFIT analyses.** *(A)* Helices modeled to be in the adenine riboswitch based on M² data with and without adenine, using prior analysis assuming a dominant secondary structure. To estimate uncertainties, simulated data replicates were prepared by bootstrapping, and percentages give fraction of the simulations in which each helix was inferred to be in the dominant secondary structure. While not recovering the actual helix frequencies in the solution ensemble, this initial analysis revealed the sheer number of possible helices that needed to be considered in modeling the ensemble, particularly for the adenine-free state. *(B)* Helices P6, P8-13 were identified and tested by single base-pair M²R. The presented structures are illustrating the helices only, and do not imply co-existence in the same structure. Helix name P7 was skipped to avoid naming conflict with a previous REEFFIT investigation.
**Figure S5.** All 188 single base-pair M²R quartets for *add* riboswitch. Helices P1 (7 bp), P2 (6 bp), P3 (6 bp), P5 (6 bp), P1B (5 bp), P2B (2 bp), P4A (4 bp), P4B (4 bp), P4C (2 bp), P6 (5 bp), P8 (4 bp), P9 (5 bp), P10 (3 bp), P11 (9 bp), P12 (3 bp), P13 (4 bp), P14 (3 bp), P15 (7 bp), P16 (5 bp), and P17 (4 bp) were tested for both no adenine and 5 mM adenine conditions. Rescue factor for each quartet is given in the title and colored as in Figure S1.

**Figure S6.** All 82 double base-pair M²R quartets for *add* riboswitch. Helices P1, P2, P3, P5, P1B, P2B, P4A, P4B, P4C, P6, and P13 were tested for both no adenine and 5 mM adenine conditions. MutP2 from previous study is also included. Rescue factor for each quartet is given in the title and colored as in Figure S1. Picked stabilizers for P1, P2, P4A, P1B, and P4B are labeled by green lock symbol.

**Figure S7. Multi-probe chemical mapping of stabilizers and linear fitting of P1 and P4A**

* (A) Five-modifier (SHAPE, 1M7, glyoxal, terbium (III), UV 302 nm) reactivity profile for lock-P1 and lock-P4A constructs. $\chi^2$ score-based linear fitting yields *apoA* state as 48% (SHAPE), 58% (DMS), 46% (glyoxal), 52% (terbium (III)) and 68% (UV 302 nm) in population. Some uncertainty remains in the UV shadowing data normalization for a quantitative fitting.

* (B) Comparison of fitted SHAPE reactivity profile to observed WT. Reactivity profiles of lock-P1 (gold) and lock-P4A (blue) are plotted. The fitted profile resolves reactivity differences in the expression platform and resembles WT.

* (C) Comparison of lock-P1 and lock-P2 profiles in the absence or presence of ligand to observed WT profile in the presence of 5 mM adenine. Lock-P1 and lock-P2 mutants show agreement in the expression platform reactivity with WT, even under no ligand condition.

**Figure S8.** All 82 single base-pair LM²R quartets for *add* riboswitch. Helices P2 and P4B are tested by lock-P1 stabilizer; helices P1, P4A, and P4B are tested by lock-P2 stabilizer; helices P2, P1B, P2B, and P4B are tested by lock-P4A stabilizer; helices P2B, P4A, P4B, and P3 are tested by lock-P1B stabilizer; helices P1, P2, P1B, P2B, and P4A are tested by lock-P4B stabilizer; all under no adenine condition. Rescue factor for each quartet is given in the title and colored as in Figure S1.

**Figure S9.** All 15 single base-pair LM²R quartets for *add* riboswitch with the MutP2 construct. Helices P1, P4A, and P4B were tested under no adenine condition. Rescue factor for each quartet is given in the title and colored as in Figure S1.
Figure S10. LM$^2$R stabilizer lock-P4B and helix frequency for apoA state.

(A) 1D SHAPE profiles for lock-P1, lock-P2, lock-P4A, lock-P1B, and lock-P4B stabilizers under both no adenine and 5 mM adenine conditions, with boxes showing successful stabilization resulting in protections from SHAPE, respectively.

(B) Base-pairing probability plot of LM$^2$R test on lock-P4B construct and expected apoA state from the non-MWC model. Helix P4A gives high helix frequency, while P1 and P2 exhibit no evidence. Helices P1B and P2B have low helix frequency as well, suggesting heterogeneity in aptamer domain folding in the absence of ligand. Helix locked in place is highlighted by green box; helices not tested by LM$^2$R are in dashed boxes.

Figure S11. Simulation of temperature dependence on switching efficiency. Thermodynamic parameters used are: $\Delta H_{K_{pre}} = 52 \text{kJ mol}^{-1}$, $\Delta S_{K_{pre}} = 167 \text{J mol}^{-1} \text{K}^{-1}$, $\Delta H_{K_d} = 238 \text{kJ mol}^{-1}$, $\Delta S_{K_d} = 110 \text{J mol}^{-1} \text{K}^{-1}$. Ligand concentrations: $[L_{\text{low}}] = 0.01 \mu\text{M}$; $[L_{\text{high}}] = 1.0 \mu\text{M}$ for (top) panels and $[L_{\text{low}}] = 1.5 \mu\text{M}$ for (bottom) heatmaps. The (top) population linear plots used $[RNA] = 1.5 \text{nM}$; while (bottom) temperature vs. [RNA] heatmap simulated $[RNA] = 0 – 100 \text{nM}$ for $T = 0 – 40 ^\circ \text{C}$. Parameters were derived from prior study reviewing the temperature compensation model.\textsuperscript{23}

Figure S12. Higher-order M$^2$ on selected single mutant backgrounds. Single mutants (A) A109U, (B) G78C, (C) A116U, (D) G44C, (E) G37C, and (F) G81C were selected from a clustering analysis of WT M$^2$ data. A comparison of their 1D SHAPE profile to WT under both no adenine and 5 mM adenine conditions are shown along with complete M$^2$ datasets. (G) Secondary structures inferred for these single mutants are quite diverse, particularly for the adenine-free state (percentage values give bootstrapping confidence that those helices are present in the dominant secondary structure). The presence of these alternative structures would explain deviations of main text LM$^2$R results for apoB-locking backgrounds (Figures 4 and 5) from expectations derived from a single apoB secondary structure (Figure 5).

Figure S13. Performance of 4-bin auto-score classifier. Histogram of in silico (A) training data (16S-FWJ, P4P6, GIR1, Hox, PB2; total of 162 quartets), (B) test data (add single and double base-pair M$^2$R; total of 242 quartets), and (C) Rfam family (total of 62484 quartets) simulations are shown by their classification, and grouped by the in silico predicted base-pairing probability (BPP).

Figure S14. Base-pair-wise classification of M$^2$R and LM$^2$R quartets by the 4-bin auto-score. Table summary of 4-bin classification by human expert and automatic algorithm on in vitro (A) single base-pair
M\textsuperscript{2}R, (B) double base-pair M\textsuperscript{2}R, and (C) LM\textsuperscript{2}R and MutP2 are shown. Each symbol represents one quartet testing a single base-pair (or 2 adjacent base-pairs in (B)).

Supporting References
1. Tian, S.; Das, R., RNA structure through multidimensional chemical mapping. Q. Rev. Biophys. 2016, 49 (e7), 1-30.
2. Meyer, M.; Nielsen, H.; Oliéric, V.; Roblin, P.; Johansen, S. D.; Westhof, E.; Masquida, B., Speciation of a group I intron into a lariat capping ribozyme. Proceedings of the National Academy of Sciences of the United States of America 2014, 111 (21), 7659-7664.
3. Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A., Crystal structure of a group I ribozyme domain: Principles of RNA packing. Science 1996, 273 (5282), 1678-1685.
4. Xue, S.; Tian, S.; Fujii, K.; Kladwang, W.; Das, R.; Barna, M., RNA regulons in Hox 5’ UTRs confer ribosome specificity to gene regulation. Nucleic Acids Research 2013, 41 (D1), D226-D232.
7. Reining, A.; Nozinovic, S.; Schlepckow, K.; Buhr, F.; Furtig, B.; Schwalbe, H., Three-state mechanism couples ligand and temperature sensing in riboswitches. Nature 2013, 499 (7458), 355-359.
8. Ali, M. Conformational transitions in RNA probed in riboswitch systems. Stanford University, 2011.
9. Cordero, P.; Das, R., Rich structure landscapes in both natural and artificial RNAs revealed by mutate-and-map analysis. PLoS Comput. Biol. 2015, 11 (11), e1004473.
10. Cordero, P.; Kladwang, W.; VanLang, C. C.; Das, R., Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. Biochemistry 2012, 51 (36), 7037-7039.
11. Banerjee, A. R.; Jaeger, J. A.; Turner, D. H., Thermal unfolding of a group I ribozyme: The low-temperature transition is primarily disruption of tertiary structure. Biochemistry 1993, 32 (1), 153-163.
12. Zou, C.; Ouyang, Z., Joint modeling of RNase footprint sequencing profiles for genome-wide inference of RNA structure. Nucleic Acids Res. 2015, 43 (19), 9187-9197.
13. Hargittai, M. R.; Musier-Forsyth, K., Use of terbium as a probe of tRNA tertiary structure and folding. RNA 2000, 6 (11), 1672-1680.
14. Harris, D. A.; Walter, N. G., Probing RNA structure and metal-binding sites using terbium(III) footprinting. Curr. Protoc. Nucleic Acid Chem. 2003, 13 (6.8).
15. Kladwang, W.; Hum, J.; Das, R., Ultraviolet shadowing of RNA can cause significant chemical damage in seconds. Sci. Rep. 2012, 2, 517.
16. Kladwang, W.; Mann, T. H.; Becka, A.; Tian, S.; Kim, H.; Yoon, S.; Das, R., Standardization of RNA chemical mapping experiments. Biochemistry 2014, 53 (19), 3063-3065.
17. Reining, A.; Nozinovic, S.; Schlepckow, K.; Buhr, F.; Furtig, B.; Schwalbe, H., Three-state mechanism couples ligand and temperature sensing in riboswitches. Nature 2013, 499 (7458), 355-9.
18. Mathews, D. H.; Disney, M. D.; Childs, J. L.; Schroeder, S. J.; Zuker, M.; Turner, D. H., Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. USA 2004, 101 (19), 7287-7292.
19. Reuter, J.; Mathews, D., RNAstructure: Software for RNA secondary structure prediction and analysis. BMC Bioinform. 2010, 11 (1), 129.
21. Burge, S. W.; Daub, J.; Eberhardt, R.; Tate, J.; Barquist, L.; Nawrocki, E. P.; Eddy, S. R.; Gardner, P. P.; Bateman, A., Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **2013**, *41* (D1), D226-D232.

22. Tian, S.; Cordero, P.; Kladwang, W.; Das, R., High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA* **2014**, *20* (11), 1815-26.

23. Fürtig, B.; Nozinovic, S.; Reining, A.; Schwalbe, H., Multiple conformational states of riboswitches fine-tune gene regulation. *Curr. Opin. Struct. Biol.* **2015**, *30*, 112-124.
SI Figure S1
Rescue Factor = 0.428
WT
U167A
A173U
U167A, A173U
Rescue Factor = -0.053
P5C
Rescue Factor = 0.445
WT
C166G
G174C
C166G, G174C
Rescue Factor = 0.617
P5C
Rescue Factor = 0.791
WT
C165G
G175C
C165G, G175C
Rescue Factor = 0.652
P5C
Rescue Factor = -0.273
WT
U168A
A173U
U168A, A173U
Rescue Factor = -1.714
alt-P5C
Rescue Factor = -0.357
WT
U167G
G174C
U167G, G174C
Rescue Factor = -0.004
alt-P5C
Rescue Factor = 0.247
WT
C166G
G175C
C166G, G175C
Rescue Factor = -0.672
alt-P5C
Rescue Factor = 0.235
WT
C165G
G176C
C165G, G176C
Rescue Factor = 0.274
alt-P5C
Rescue Factor = -0.275
WT
G164C
U177G
G164C, U177G
Rescue Factor = -0.569
alt-P5C
Rescue Factor = 0.404
WT
C216G
G257C
C216G, G257C
Rescue Factor = -0.297
P6
Rescue Factor = -0.100
WT
G215U
U258G
G215U, U258G
Rescue Factor = 0.313
P6
Rescue Factor = -0.752
WT
G141C
U146G
G141C, U146G
Rescue Factor = -0.319
P5
Rescue Factor = -0.100
WT
G141U
U146G
G141U, U146G
Rescue Factor = 0.313
P6

SI Figure S2
No Adenine, Rescue Factor = 0.687
WT A58U
U68A
A58U, U68A
5 mM Adenine, Rescue Factor = 0.393
P3
No Adenine, Rescue Factor = 0.604
WT A56U
U70A
A56U, U70A
5 mM Adenine, Rescue Factor = 0.267
P3
No Adenine, Rescue Factor = 0.617
WT A55U
U71A
A55U, U71A
5 mM Adenine, Rescue Factor = 0.416
P3
No Adenine, Rescue Factor = 0.728
WT A54U
U72A
A54U, U72A
5 mM Adenine, Rescue Factor = 0.593
P3
No Adenine, Rescue Factor = 0.617
WT C54G
G72C
C54G, G72C
5 mM Adenine, Rescue Factor = 0.460
P5
No Adenine, Rescue Factor = 0.617
WT U94A
A105U
U94A, A105U
5 mM Adenine, Rescue Factor = 0.173
P5
No Adenine, Rescue Factor = 0.728
WT U92A
A107U
U92A, A107U
5 mM Adenine, Rescue Factor = 0.350
P5
No Adenine, Rescue Factor = 0.639
WT U90A
A109U
U90A, A109U
5 mM Adenine, Rescue Factor = 0.267
P5
No Adenine, Rescue Factor = 0.516
WT C89G
G110C
C89G, G110C
5 mM Adenine, Rescue Factor = 0.199
P5
No Adenine, Rescue Factor = 0.505
WT U17G
A45U
U17G, A45U
5 mM Adenine, Rescue Factor = 0.033
P1B
No Adenine, Rescue Factor = 0.196
WT C15G
G46C
C15G, G46C
5 mM Adenine, Rescue Factor = 0.018
P1B
No Adenine, Rescue Factor = 0.103
WT C18G
G43C
C18G, G43C
5 mM Adenine, Rescue Factor = 0.199
P1B
No Adenine, Rescue Factor = -0.362
WT U16A
A45U
U16A, A45U
5 mM Adenine, Rescue Factor = -0.362
P1B
No Adenine, Rescue Factor = -0.106
WT G14C
G46C
G14C, G46C
5 mM Adenine, Rescue Factor = -0.612
P1B
No Adenine, Rescue Factor = 0.802
WT U17G
G44C
U17G, G44C
5 mM Adenine, Rescue Factor = -0.362
P1B
No Adenine, Rescue Factor = -0.106
WT G14C
U47G
G14C, U47G
5 mM Adenine, Rescue Factor = -0.612
P1B
No Adenine, Rescue Factor = 0.687
WT A58U
U68A
A58U, U68A
5 mM Adenine, Rescue Factor = 0.393
P3
No Adenine, Rescue Factor = 0.604
WT A56U
U70A
A56U, U70A
5 mM Adenine, Rescue Factor = 0.267
P3
No Adenine, Rescue Factor = 0.617
WT A55U
U71A
A55U, U71A
5 mM Adenine, Rescue Factor = 0.416
P3
No Adenine, Rescue Factor = 0.728
WT A54U
U72A
A54U, U72A
5 mM Adenine, Rescue Factor = 0.593
P3
No Adenine, Rescue Factor = 0.617
WT C54G
G72C
C54G, G72C
5 mM Adenine, Rescue Factor = 0.460
P5
No Adenine, Rescue Factor = 0.617
WT U94A
A105U
U94A, A105U
5 mM Adenine, Rescue Factor = 0.173
P5
No Adenine, Rescue Factor = 0.728
WT U92A
A107U
U92A, A107U
5 mM Adenine, Rescue Factor = 0.350
P5
No Adenine, Rescue Factor = 0.639
WT U90A
A109U
U90A, A109U
5 mM Adenine, Rescue Factor = 0.267
P5
No Adenine, Rescue Factor = 0.516
WT C89G
G110C
C89G, G110C
5 mM Adenine, Rescue Factor = 0.199
P5
No Adenine, Rescue Factor = 0.103
WT C18G
G43C
C18G, G43C
5 mM Adenine, Rescue Factor = 0.199
P1B
No Adenine, Rescue Factor = -0.362
WT U16A
A45U
U16A, A45U
5 mM Adenine, Rescue Factor = -0.362
P1B
No Adenine, Rescue Factor = -0.106
WT G14C
U47G
G14C, U47G
5 mM Adenine, Rescue Factor = -0.612
P1B
No Adenine, Rescue Factor = 0.802
WT U17G
G44C
U17G, G44C
5 mM Adenine, Rescue Factor = -0.362
P1B
No Adenine, Rescue Factor = -0.106
WT G14C
U47G
G14C, U47G
5 mM Adenine, Rescue Factor = -0.612
P1B
|          | No Adenine, Rescue Factor | 5 mM Adenine, Rescue Factor |
|----------|--------------------------|-----------------------------|
| **P2B**  |                          |                             |
| WT       |                          |                             |
| C27G     | 0.344                    | 0.266                       |
| C37C     |                          |                             |
| C27G, G37C |                      |                             |
| **P2B**  |                          |                             |
| WT       |                          |                             |
| C26G     | 0.227                    | 0.278                       |
| G38C     |                          |                             |
| C26G, G38C |                     |                             |
| **P4A**  |                          |                             |
| WT       |                          |                             |
| A79U     | 0.465                    | 0.278                       |
| U115A    |                          |                             |
| A79U, U115A |                   |                             |
| **P4A**  |                          |                             |
| WT       |                          |                             |
| G78C     | 0.290                    | -0.319                      |
| C119G    |                          |                             |
| G78C, C119G |                   |                             |
| **P4A**  |                          |                             |
| WT       |                          |                             |
| U77A     | 0.451                    | -0.522                      |
| A120U    |                          |                             |
| U77A, A120U |                   |                             |
| **P4B**  |                          |                             |
| WT       |                          |                             |
| A76U     | 0.522                    | 0.114                       |
| U121A    |                          |                             |
| A76U, U121A |                  |                             |
| **P4B**  |                          |                             |
| WT       |                          |                             |
| U84A     | 0.306                    | 0.001                       |
| A114U    |                          |                             |
| U84A, A114U |                   |                             |
| **P4B**  |                          |                             |
| WT       |                          |                             |
| C83G     | 0.608                    | 0.029                       |
| G115C    |                          |                             |
| C83G, G115C |                    |                             |
| **P4B**  |                          |                             |
| WT       |                          |                             |
| U82A     | 0.219                    | 0.053                       |
| A116U    |                          |                             |
| U82A, A116U |                   |                             |
| **P4B**  |                          |                             |
| WT       |                          |                             |
| G81C     | 0.576                    | 0.231                       |
| C117G    |                          |                             |
| G81C, C117G |                    |                             |
| **P4C**  |                          |                             |
| WT       |                          |                             |
| C87G     | 0.544                    | -0.049                      |
| G112C    |                          |                             |
| C87G, G112C |                   |                             |
| **P4C**  |                          |                             |
| WT       |                          |                             |
| U86A     | 0.005                    | 0.027                       |
| A113U    |                          |                             |
| U86A, A113U |                  |                             |
| **P6**   |                          |                             |
| WT       |                          |                             |
| U22A     | 0.025                    | -0.079                      |
| A33U     |                          |                             |
| U22A, A33U |                   |                             |
| **P6**   |                          |                             |
| WT       |                          |                             |
| A21U     | 0.025                    | -0.079                      |
| U34A     |                          |                             |
| A21U, U34A |                   |                             |
| **P6**   |                          |                             |
| WT       |                          |                             |
| U20A     | 0.085                    | -0.668                      |
| A35U     |                          |                             |
| U20A, A35U |                   |                             |
| **P6**   |                          |                             |
| WT       |                          |                             |
| A19U     | 0.085                    | -0.668                      |
| U36A     |                          |                             |
| A19U, U36A |                   |                             |
| **P6**   |                          |                             |
| WT       |                          |                             |
| A19U     | 0.085                    | -0.668                      |
| U36A     |                          |                             |
| A19U, U36A |                   |                             |
| Sequence Position | Rescue Factor | Adenine Condition |
|-------------------|---------------|-------------------|
| No Adenine, Rescue Factor = 0.028 | 5 mM Adenine, Rescue Factor = -0.027 |
| WT C18G G37C C18G, G37C | | |
| No Adenine, Rescue Factor = 0.151 | 5 mM Adenine, Rescue Factor = 0.185 |
| WT G44C U49G G44C, U49G | | |
| No Adenine, Rescue Factor = 0.083 | 5 mM Adenine, Rescue Factor = 0.073 |
| WT G43C C50G G43C, C50G | | |
| No Adenine, Rescue Factor = 0.181 | 5 mM Adenine, Rescue Factor = -0.434 |
| WT G42C U51G G42C, U51G | | |
| No Adenine, Rescue Factor = 0.054 | 5 mM Adenine, Rescue Factor = 0.233 |
| WT U41A A52U U41A, A52U | | |
| No Adenine, Rescue Factor = 0.082 | 5 mM Adenine, Rescue Factor = -0.106 |
| WT G112C C117G G112C, C117G | | |
| No Adenine, Rescue Factor = 0.444 | 5 mM Adenine, Rescue Factor = 0.048 |
| WT A111U U118A A111U, U118A | | |
| No Adenine, Rescue Factor = -0.035 | 5 mM Adenine, Rescue Factor = -0.195 |
| WT G88C C96G G88C, C96G | | |
| No Adenine, Rescue Factor = -0.625 | 5 mM Adenine, Rescue Factor = -0.237 |
| WT G81C U101A G81C, U101A | | |
| No Adenine, Rescue Factor = -0.173 | 5 mM Adenine, Rescue Factor = -0.625 |
| WT A80U U102A A80U, U102A | | |
| No Adenine, Rescue Factor = -0.014 | 5 mM Adenine, Rescue Factor = -0.058 |
| WT U86A A98U U86A, A98U | | |
| No Adenine, Rescue Factor = 0.137 | 5 mM Adenine, Rescue Factor = -0.034 |
| WT U82A A100U U82A, A100U | | |
| No Adenine, Rescue Factor = 0.094 | 5 mM Adenine, Rescue Factor = 0.057 |
| WT G81C U101A G81C, U101A | | |
| No Adenine, Rescue Factor = -0.178 | 5 mM Adenine, Rescue Factor = -0.258 |
| WT A80U U102A A80U, U102A | | |
| No Adenine, Rescue Factor = -0.016 | 5 mM Adenine, Rescue Factor = -0.258 |
|    | WT                  | No Adenine, Rescue Factor | 5 mM Adenine, Rescue Factor |
|----|---------------------|---------------------------|-----------------------------|
| P1 | C15G, U16A, A80U, G81C | 0.694                     | 0.799                       |
|    | C15G, U16A, A80U, G81C | 0.118                     | 0.761                       |
| P2 | A56U, G57C, C69G, U70A | -0.476                    | 0.694                       |
|    | A19U, U20A, A76U, U77A | -0.631                    | 0.869                       |
|    | A56U, G57C, C69G, U70A | 0.095                     | 0.599                       |
| P3 | C54G, A55U, U71A, G72C | 0.118                     | 0.799                       |
|    | A55U, A56U, U70A, U71A | -0.508                    | 0.799                       |
|    | A55U, A56U, U70A, U71A | -0.289                    | 0.799                       |
| Position | WT | G57C, A58U | U68A, C69G | G57C, A58U, U68A, C69G | 5 mM Adenine | Rescue Factor |
|----------|----|-------------|-------------|--------------------------|---------------|---------------|
| P3       | No Adenine, Rescue Factor = 0.569 | 5 mM Adenine, Rescue Factor = 0.840 |
| P3       | No Adenine, Rescue Factor = 0.544 | 5 mM Adenine, Rescue Factor = 0.765 |
| P5       | No Adenine, Rescue Factor = 0.595 | 5 mM Adenine, Rescue Factor = 0.258 |
| P5       | No Adenine, Rescue Factor = 0.481 | 5 mM Adenine, Rescue Factor = 0.152 |
| P5       | No Adenine, Rescue Factor = 0.379 | 5 mM Adenine, Rescue Factor = -0.138 |
| P5       | No Adenine, Rescue Factor = 0.212 | 5 mM Adenine, Rescue Factor = -0.050 |
| P2B      | No Adenine, Rescue Factor = 0.044 | 5 mM Adenine, Rescue Factor = -0.308 |
| P4A      | No Adenine, Rescue Factor = 0.159 | 5 mM Adenine, Rescue Factor = -0.097 |
| P4A      | No Adenine, Rescue Factor = 0.441 | 5 mM Adenine, Rescue Factor = -0.102 |
| P4B      | No Adenine, Rescue Factor = 0.278 | 5 mM Adenine, Rescue Factor = 0.036 |

20 40 60 80 100 120 140
Sequence Position

20 40 60 80 100 120 140
Sequence Position
No Adenine, Rescue Factor = 0.440
WT
U82A, C83G
G115C, A116U
U82A, C83G, G115C, A116U
5 mM Adenine, Rescue Factor = 0.019

No Adenine, Rescue Factor = 0.204
P4B
WT
C83G, U84A
A114U, G115C
C83G, U84A, A114U, G115C
5 mM Adenine, Rescue Factor = 0.693

No Adenine, Rescue Factor = 0.179
P4C
WT
U86A, C87G
G112C, A113U
U86A, C87G, G112C, A113U
5 mM Adenine, Rescue Factor = 0.200

No Adenine, Rescue Factor = 0.120
P6
WT
C18G, A19U
U36A, G37C
C18G, A19U, U36A, G37C
5 mM Adenine, Rescue Factor = 0.085

No Adenine, Rescue Factor = 0.084
P6
WT
A19U, U20A
A35U, U36A
A19U, U20A, A35U, U36A
5 mM Adenine, Rescue Factor = 0.045

No Adenine, Rescue Factor = 0.152
P6
WT
U20A, A21U
U34A, A35U
U20A, A21U, U34A, A35U
5 mM Adenine, Rescue Factor = 0.012

No Adenine, Rescue Factor = 0.013
P6
WT
A21U, U22A
A33U, U34A
A21U, U22A, A33U, U34A
5 mM Adenine, Rescue Factor = 0.049

No Adenine, Rescue Factor = 0.180
P13
WT
A19U, U20A
A35U, U36A
A19U, U20A, A35U, U36A
5 mM Adenine, Rescue Factor = 0.787

No Adenine, Rescue Factor = 0.034
P13
WT
U17A, C18G
G32C, A33U
U17A, C18G, G32C, A33U
5 mM Adenine, Rescue Factor = 0.048

No Adenine, Rescue Factor = 0.013
P13
WT
C18G, A19U
U31A, G32C
C18G, A19U, U31A, G32C
5 mM Adenine, Rescue Factor = 0.117

No Adenine, Rescue Factor = 0.001
P13
WT
A19U, U20A
A35U, U36A
A19U, U20A, A35U, U36A
5 mM Adenine, Rescue Factor = 0.034

MutP2
WT
A29C, A30G
U40C, U41G
A29C, A30G, U40C, U41G
5 mM Adenine, Rescue Factor = 0.787

No Adenine, Rescue Factor = 0.180
P13
WT
A19U, U20A
A35U, U36A
A19U, U20A, A35U, U36A
5 mM Adenine, Rescue Factor = 0.787
### SI Figure S7

**SHAPE Reactivity**

- **Sequence Position**
- **WT (No Adenine)**
- **Lock-P1 (No Adenine)**
- **Lock-P4A (No Adenine)**
- **FIT**

---

**SHAPE Reactivity**

- **WT (5 mM Adenine)**
- **Lock-P1 (5 mM Adenine)**
- **Lock-P2 (5 mM Adenine)**

---

**Linear Fitting**

- **FIT**: Lock-P1 + Lock-P4A
(A) **Lock-P1** (A19U, U20A, A76U, U77A)

- **P2**
  - Lock U25G A45C U25G, A45C
  - No Adenine, Rescue Factor = 0.280
- **P2**
  - Lock C26G G44C C26G, G44C
  - No Adenine, Rescue Factor = 0.073
- **P2**
  - Lock U28G G42C U28G, G42C
  - No Adenine, Rescue Factor = 0.216
- **P2**
  - Lock A29C U41G A29C, U41G
  - No Adenine, Rescue Factor = -0.126

(B) **Lock-P2** (U28G, A29U, U41A, G42C)

- **P1**
  - Lock C15G G81C C15G, G81C
  - No Adenine, Rescue Factor = 0.255
- **P1**
  - Lock U16G A80C U16G, A80C
  - No Adenine, Rescue Factor = 0.724

(C) **Lock-P4A** (U77A, G78C, C119G, A120U)

- **P2**
  - Lock U25G A45C U25G, A45C
  - No Adenine, Rescue Factor = -0.366
- **P2**
  - Lock C26G G44C C26G, G44C
  - No Adenine, Rescue Factor = -0.104
- **P2**
  - Lock U28G G42C U28G, G42C
  - No Adenine, Rescue Factor = -0.020
- **P2**
  - Lock A29C U41G A29C, U41G
  - No Adenine, Rescue Factor = 0.179

**Sequence Position**

**SI Figure S8**
| Lock          | Rescue Factor |
|--------------|---------------|
| P1 C15G G81C | 0.122         |
| P1 U16G A80C | 0.122         |
| P1 U17G A79C | 0.276         |
| P1 C18G G78C | 0.071         |
| P1 A19C U77G | 0.328         |
| P1 U20G A76C | 0.200         |
| P1 A21C U75G | 0.200         |
| P4A A76C U121G | 0.097   |
| P4A A120C U77G | 0.035   |
| P4A G78C C119G | 0.164   |
| P4A A118G U79C | -0.038  |
| P4B G81C C117G | 0.227   |
| P4B A116C U82G | 0.233   |
| P4B C83G G115C | 0.291   |
(A) 5 mM Adenine

WT
A109U

(B) 5 mM Adenine

WT
G78C

SI Figure S12
**E**

5 mM Adenine

No Adenine

| Sequence Position | Mutation Position |
|-------------------|-------------------|
| 20                |                   |
| 40                |                   |
| 60                |                   |
| 80                |                   |
| 100               |                   |
| 120               |                   |
| 140               |                   |

**F**

5 mM Adenine

No Adenine

| Sequence Position | Mutation Position |
|-------------------|-------------------|
| 20                |                   |
| 40                |                   |
| 60                |                   |
| 80                |                   |
| 100               |                   |
| 120               |                   |
| 140               |                   |

SI Figure S12
(A) Validated
(Falsified)
(Uncertain)
(Partial)

(bpp(i,j))

(C) Validated
(Falsified)
(Uncertain)
(Partial)

(bpp(i,j))
### (A) M²R Single Base-Pair

|        | Manual Score | Auto Score |
|--------|--------------|------------|
| P1     | ??+??????    | +??????+   |
| P2     | ??+??????    | ???????+   |
| P3     | +??????+     | +??????+   |
| P5     | ??+??????    | +??????+   |
| P1B    | +??????+     | +??????+   |
| P2B    | +            | ?          |
| P4A    | +??????+     | +??????+   |
| P4B    | +??????+     | +??????+   |
| P4C    | +??????+     | +??????+   |
| P6     | ??+??????    | ??+??????  |
| P8     | +??????+     | +??????+   |
| P10    | +??????+     | +??????+   |
| P11    | ??+??????    | ??+??????  |
| P12    | ??+??????    | ??+??????  |
| P13    | ??+??????    | ??+??????  |
| P14    | +??????+     | +??????+   |
| P15    | +??????+     | +??????+   |
| P16    | +??????+     | +??????+   |
| P17    | ??+??????    | ??+??????  |

### (B) M²R Double Base-Pair

|        | Manual Score | Auto Score |
|--------|--------------|------------|
| P1     | +??????+     | +??????+   |
| P2     | +??????+     | +??????+   |
| P3     | +??????+     | +??????+   |
| P5     | +??????+     | +??????+   |
| P1B    | ??+??????    | +??????+   |
| P2B    | +            | ?          |
| P4A    | +??????+     | +??????+   |
| P4B    | +??????+     | +??????+   |
| P3     | +??????+     | +??????+   |

### (C) LM²R

|        | Manual Score | Auto Score |
|--------|--------------|------------|
| Lock-P1| +??????+     | +??????+   |
| P2     | +??????+     | +??????+   |
| P4A    | ??+??????    | ??+??????  |
| P4B    | +??????+     | +??????+   |
| Lock-P2| +??????+     | +??????+   |
| P1     | +??????+     | +??????+   |
| P2     | +??????+     | +??????+   |
| P4A    | ??+??????    | ??+??????  |
| P4B    | +??????+     | +??????+   |
| Lock-P4A| +??????+    | +??????+   |
| P1B    | ??+??????    | ??+??????  |
| P2B    | +            | ?          |
| P4A    | +??????+     | +??????+   |
| P4B    | +??????+     | +??????+   |
| Lock-P4B| +??????+    | +??????+   |
| P1B    | +??????+     | +??????+   |
| P2B    | +??????+     | +??????+   |
| P4A    | +??????+     | +??????+   |
| P4B    | +??????+     | +??????+   |

Legend: ✅ Validated ⚫ Falsified ⚫ Uncertain + Partial

SI Figure S14