Allosteric Mechanism of the *V. vulnificus* Adenine Riboswitch Resolved by Four-dimensional Chemical Mapping

Siqi Tian[1], Wipapat Kladwang[1], and Rhiju Das[1, 2, *]

[1]Department of Biochemistry, and [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 interconversions that mediate the gene regulatory functions of RNA molecules may be different from classic models of allostery, but the relevant structural correlations have remained elusive in even intensively studied systems. Here, we present a four-dimensional expansion of chemical mapping called lock-mutate-map-rescue (LM^2R), 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, 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 compensatory rescue evaluated by chemical mapping. These LM^2R measurements give strong support to the pre-existing correlations predicted by MWC models, disfavor alternative models, and suggest additional 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. Recent years have seen an explosion of discoveries of *cis*-acting mRNA elements that sense small molecules, proteins, RNAs, and other environmental conditions and then 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 distinct RNA secondary structures.[4-6] If these multiple states could be characterized in detail and quantitatively modeled, they would offer potential new substrates for biological control and new targets for antibiotic development.[7]

Despite this strong interest in RNA conformational change, directly testing the base pair correlations posited to underlie RNA allostery have been difficult for most systems. Experimental barriers remain even for the compact and self-contained riboswitches, which occur in eukaryotes and appear pervasive throughout bacteria. A riboswitch is a *cis*-regulatory RNA domain composed of two segments: (1) an 'aptamer' segment that forms an intricate tertiary structure to specifically bind small molecule ligands, such as nucleobases or amino acids, and (2) a 'gene expression platform' segment that interacts with molecular machines like the ribosome to modulate gene expression.[8-10] In many cases, the coupling of ligand binding to changes in the gene expression platform has been reconstituted *in vitro* in small model RNAs, enabling detailed biophysical analysis. Nevertheless, in most of these cases, the mechanism underlying the coupling of ligand binding to gene expression remains under question, suggesting the need for new concepts and new biophysical methods to understand RNA allostery.

A particularly well-studied – but still incompletely understood – example of an allosteric RNA is the *add* riboswitch, a ~120-nucleotide element that resides in the 5´ untranslated region (UTR) of the *add* adenosine deaminase mRNA of human pathogenic bacterium *Vibrio vulnificus*. This riboswitch can directly bind adenine and controls translation of the mRNA in response to adenine *in vivo*. Domains of this RNA have been subjected to nearly every biochemical and biophysical technique available, including crystallography, single molecule measurements, and time-resolved X-ray laser diffraction.[11-22] Almost all work on the *add* riboswitch to date have assumed some variation of a Monod-Wyman-Changeux model (MWC, also called 'conformational selection' or 'population shift' models).[6, 13, 23-24] In the absence of adenine, the RNA dominantly forms a stable *apo* secondary structure whose aptamer region is in a fold incompatible with adenine binding and

whose gene expression platform, a ribosome binding site and AUG start codon, is sequestered into helices (blue, Figure 1A). High adenine concentrations stabilize the ligand-binding secondary structure in the aptamer region and rearrange the secondary structure around the gene expression platform so that it is freed to recruit the ribosome and translation machinery (gold, Figure 1A). The signature prediction of the MWC model is that the *holo* RNA secondary structure should also exist in the absence of adenine, albeit as a high energy conformation within the *apo* ensemble (Figure 1A, middle panel; gold). In terms of the riboswitch functional domains, when the RNA occasionally fluctuates to form an aptamer secondary structure compatible with ligand binding, the gene expression platform is predicted to concomitantly open, even without the adenine ligand. This prediction for the ligand-free ensemble is analogous to the classic MWC model of oxygen binding by hemoglobin, involving interconversion between thermodynamically stable hemoglobin structures (tense and relaxed, dominating *apo* and *holo*), with transient sampling of the *holo* oxygen-binding structure even in the absence of oxygen. This prediction for hemoglobin and numerous other protein systems have been supported through decades of studies on protein allostery that include direct detection of transient *holo*-like states in the *apo* ensemble even in the absence of small molecules needed to fully stabilize the *holo* conformation.[24-25]

Recently, a tour de force study highlighted an alternative to the MWC mechanism for the *add* riboswitch (Figure 1B). Multidimensional NMR spectroscopy, NMR relaxation studies, designed model systems, and supporting measurements from stopped-flow kinetics and calorimetry revealed detailed single-nucleotide-resolution base pairing information and intriguing temperature dependences for the complex structural ensemble of the *add* riboswitch.[26] Although similar in some respects to previously proposed MWC models, the new model contradicted a standard MWC assumption[23-25]: it proposed 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 (Figure 1B, middle panel; mixed blue and gold, red rectangle outline). 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 mechanism 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] Understanding the mechanism of allostery is particularly important for developing strategies for controlling riboswitches. For example, potential antibiotics that attempt to constitutively turn on expression of the *add* gene by

selectively stabilizing the aptamer secondary structure may not succeed if the MWC model is incorrect.

Here, we develop an experimental approach that resolves the allosteric mechanism of the *add* riboswitch. To help avoid previous confusion in state definitions, we first cast the question of mechanism into one of determining the statistical correlation or anticorrelation of specific helix elements. We then expand a previously reported method that integrates compensatory rescue with chemical mapping readouts (mutate-map-rescue, M^2R)[28-29] to enable dissection of these helix-helix correlations. The method takes advantage of a Bayesian framework and simulations to connect experimental observables to underlying helix frequencies, and gives experimental results on the wild type *add* sequence that agree with predictions from all available models. Then, by identifying 'lock' mutants that stabilize helices posited for each state and carrying out further rounds of M^2R in these lock mutant backgrounds, we infer how the presence of one helix enhances or suppresses the presence of other helices. The resulting data unambiguously discriminate between MWC and non-MWC models and also suggest additional structural heterogeneity for the *add* riboswitch. We end by discussing how the presented method may be useful for generally dissecting allostery in RNA-based gene regulation.

RESULTS

Defining the question in terms of helix-helix correlations

Before presenting our experimental strategy to dissect RNA allostery, we need to state the problem in a way that can be answered in a binary fashion. As shown in Figures 1A and 1B, a conventional way to phrase questions about allostery is to posit a small number of states and then to design experiments to probe properties of those states. For example, for the *add* riboswitch, previous models partitioned the states in the *apo* ensemble into *apoA* and *apoB*, depending on whether the aptamer secondary structure is correctly formed or not, respectively. We would then design experiments to discriminate whether *apoA* has its gene expression platform open or closed (Figure 1A vs. Figure 1B). However, in practice, this way of posing the question can be confusing. Figures 1C and 1D show cartoons of the ensemble of RNA conformations that illustrate the potential complexities. First, there may be RNA secondary structures present in the actual *apo* ensemble that include some but not all helices of the aptamer secondary structures (Figures 1C and 1D), and it then becomes ambiguous whether they should count as members of the *apoA* or *apoB* state. In addition, there may be secondary structures that include or are entirely composed of helices not shown in any of the presented models (marked with pink arrow, Figures

1C and 1D). Indeed, in principle, there are a vast number of secondary structures that need to be considered. Not only are there several possible helices that can be formed by the RNA, but there are an exponentially large number of combinations of such helices, each giving rise to a distinct secondary structure.

We sought to define the allosteric mechanism in terms of the possible helices that the RNA can form, rather than this vast number of possible secondary structures. In the presence of adenine, there are three helices P1, P2, and P3 that define an adenine-binding aptamer secondary structure at nucleotides 15-81, with a well-defined tertiary structure that has been determined through crystallography. 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). The P4 domain includes potentially three helices, here called P4A, P4B, and P4C. P4A contains nucleotides that were originally part of P1 (the 'switch region'[13], nts 76–81, cross-hatched in Figure 1A, left). When P4A forms, some or all of the aptamer helices may rearrange into an alternative secondary structure or, potentially, a large set of alternative secondary structures. NMR experiments directly detected two alternative helices, P1B and P2B (Figure 1A), but do not rule out other helices being present at significant fractions if their conformational dynamics renders them invisible to the spectroscopy methods applied. Despite these potential complexities, there are a small number of Watson-Crick helices that feature in the current MWC and non-MWC models: P1, P2, P3, P4A, P4B, P4C, and P5. Additional helices may also be relevant to the *add* riboswitch mechanism. For example, numerous helices arose in preliminary modeling of the riboswitch based on conventional 1-dimensional SHAPE and mutate-and-map measurements (Supplementary Materials; Figure 1 Supplement 1, Figure 1 Supplement 2, and Figure 1 Supplement 3). These helices (called here P6, P8 to P17) are further tested below.

Given the list of candidate helix elements, we can define the question that discriminates between MWC and non-MWC models, even if the *apo* ensemble is highly heterogeneous (Figures 1C and 1D). In the absence of adenine, the riboswitch may occasionally sample the helices P1 or P2 (or both simultaneously), which are signatures of the adenine-bound state. The MWC conformational selection model would assume that any such thermally sampled secondary structure should also have an open gene expression platform, just like the adenine-bound *holo* and without the P4A or

P4B domains. That is, there should be an anticorrelation between the presence of, e.g., P1 and P4B, in the ligand-free ensemble (Figure 1C). There should be similar anticorrelations between, e.g., P2 and P4A. (For clarity of presentation, here and below, we will use P1 and P4B (colored gold and blue, respectively, in figures) to illustrate core concepts and data in detail, and then summarize additional independent measurements for other helix-helix pairs, such as P2 and P4A.)

More quantitatively, suppose that P1 and P4B arise with 50% frequency in the ligand-free *add* riboswitch structural ensemble. If there is no allosteric 'communication' between these two elements, the joint frequency of observing both P1 and P4B should be 50% x 50% = 25%. The MWC model predicts that the actual joint frequency of these two elements is significantly lower: the formation of P1 in the aptamer destabilizes the P4B helix that sequesters the gene expression platform, even in the absence of adenine. In terms of a single numerical value, we seek to determine the correlation $g(\mathrm{P1,P4B})$:

$$g(\mathrm{P1,P4B}) = \frac{F(\mathrm{P1,P4B})}{F(\mathrm{P1}) \times F(\mathrm{P4B})} \qquad (1)$$

where F is the helix frequency, a number between 0% and 100%.

The MWC model predicts a strong anticorrelation: $g(\mathrm{P1,P4B}) \ll 1$. In contrast, the non-MWC model proposed based on NMR experiments makes different predictions (Figure 1B and 1D). In the absence of ligand, the NMR measurements could not directly detect opening of the gene expression platform (loss of P4A, P4B, or P4C) at low frequency in the absence of ligand. The non-MWC model 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 P4B and P4C) are retained (middle panel of Figure 1B, and structures marked with red arrow in Figure 1D). The non-MWC model predicts no correlation between elements of the aptamer region and elements of the P4 domain, so that, for example, $g(\mathrm{P1,P4B}) \approx 1$.

Our strategy involves experimentally determining the helix frequencies in the *add* riboswitch structural ensemble $F(\mathrm{P1})$, $F(\mathrm{P4B})$, etc., based on a method described in the next section. To then determine the correlation for two elements P1 and P4B, we design mutations to lock P4B and see if the frequency of the P1 element increases, stays the same, or decreases. Such

measurements allow inference of $g(\mathrm{P1,P4B})$ through the relation of joint probabilities to conditional probabilities, $F(\mathrm{P1,P4B}) = F(\mathrm{P1|P4B})F(\mathrm{P4B})$. Equation (1) becomes

$$g(\mathrm{P1,P4B}) = \frac{F(\mathrm{P1|P4B})}{F(\mathrm{P1})} \quad (2)$$

Here $F(\mathrm{P1|P4B})$ is the frequency of helix P1 conditional on the formation of P4B and would be estimated from experiments that lock P1 and assess the P4B helix frequency. In a 'flipped' experiment, we also lock P1 and then assess if the helix frequency of P4B increases, stays the same, or decreases. The results enable an independent evaluation of the same correlation value through a similar relation:

$$g(\mathrm{P1,P4B}) = \frac{F(\mathrm{P4B|P1})}{F(\mathrm{P4B})} \quad (3)$$

Correlation values for other pairs of helices, such as $g(\mathrm{P2,P4B})$, are defined analogously to equations (1)-(3).

Inferring helix frequencies through quantitative compensatory rescue

To estimate correlations between helix elements, we need a way to estimate frequencies of each helix element in an RNA's structural ensemble, the variables *F*(P1), *F*(P4B), etc. in equations (1)-(3). We discovered an approach to such helix frequency inference when revisiting the use of compensatory mutation experiments for RNA structure read out through SHAPE chemical mapping (selective 2´-hydroxyl acylation with primer extension).[28-29] Mutate-map-rescue (M^2R) experiments (Figure 2) enable incisive confirmation or falsification of candidate base pairs. Strong evidence for a candidate base pair comes from observations that mutation of either side of a base pair (mutant A or mutant B) disrupts the SHAPE profile of the RNA; and that the compensatory double mutation (mutant AB) restores the SHAPE measurements to the wild type (WT) profile. In our prior work on a four-helix junction from the *E. coli* 16S ribosomal RNA (nts 126-235), we were able to successfully use such 'quartets' of SHAPE profiles (WT, mutant A, mutant B, mutant AB) to discriminate between three conflicting structural models derived from prior SHAPE data on the wild type RNA alone, from mutate-and-map data (SHAPE profiles of all single mutants), and from phylogenetic & crystallographic models of the protein-bound ribosome (illustrative data for three base pairs given in Figure 2A).[28] However, we were also initially surprised to observe compensatory rescue for helices that should have been mutually exclusive in the 16S rRNA. These helices involved the same nucleotides partnered to nucleotides

differing by a single-nucleotide register shift (P4a and shift-P4a, not to be confused with P4A in the *add* riboswitch; see blue and green base pairs in Figure 2A). Resolving this paradox, studies on designed mutants and quantitative fits suggested that both registers were indeed present but in different subsets of the RNA's structural ensemble, at helix frequencies of ~20% and ~80%.[28]

In our studies on the 16S rRNA four-way junction, the strongly populated helix shift-P4a was associated with mutate-map-rescue data that showed unambiguous rescue while the less populated helix P4a was associated with 'partial' rescue. In the latter case, disruptions in SHAPE reactivity induced by single mutations were alleviated by their concomitant compensatory mutation, but new features distinct from the wild type SHAPE profile also appeared (yellow arrow, bottom panel of Figure 2A). We speculated that such 'partial rescue' might generally be a hallmark of base pairs that are present at low frequency in an RNA's structural ensemble. We envisioned the structural ensembles shown in Figure 2B, illustrating a case where a candidate base pair is present at 50% frequency in the wild type RNA structural ensemble (structures with gold highlight rectangle in top left, Figure 2B). The compensatory double mutant (structures with gold highlight rectangle in bottom right, Figure 2B) might retain the structures with the candidate base pair at approximately 50% frequency, while the other members of the structural ensemble would be more likely to rearrange (bottom structure of bottom right panel, Figure 2B). The single mutants would rearrange all structures, including the ones that had the target base pair (other structures, bottom left and top right panels, Figure 2B), providing an assessment of the maximum level of disruption that might be restored by the compensatory mutant. In general, however, this picture is not expected to be exact. Some compensatory double mutations might over- or under-stabilize the structures with the target base pair. The RNA may also be excessively sensitive or insensitive to disruption by single mutations. Therefore, we expected that testing all the base pairs of a target helix through separate mutate-map-rescue experiments for each base pair would be critical for assessing helix frequencies, and a careful treatment of uncertainties would be important for establishing statistical confidence.

To test this idea, we first developed a quantitative 'rescue factor' metric that allowed us to avoid human inspection that might bias assessments:

$$\text{Rescue factor} \equiv 1.0 - \frac{RMSD\,(\text{WT,AB})}{\max(RMSD\,(\text{WT,A}), RMSD\,(\text{WT,B}))} \quad , \qquad (4)$$

where the root mean squared difference *RMSD* between two SHAPE profiles *d* over probed nucleotides $i = 1$ to N is given by:

$$RMSD(\mathrm{A,B}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[d_A(i) - d_B(i)]^2} \quad (5)$$

In equation (4), the term $RMSD(\mathrm{WT,AB})$ captures the extent of rescue, as compared to a maximal amount of disruption of the SHAPE profile given by the maximum of $RMSD(\mathrm{WT,A})$ and $RMSD(\mathrm{WT,B})$. We confirmed that this rescue factor recovered expert assessments for presence or absence of rescue in the previous 16S rRNA data as well as in newly collected M^2R experimental data, including the *Tetrahymena* ribozyme P4-P6 domain and a blind RNA-puzzle modeling challenge, the GIR1 lariat capping ribozyme (Methods & Appendix Results). 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 again that automated analysis reproduced manual analysis (Appendix Results).

We then sought to test our hypothesis of a quantitative relationship between rescue factor and helix frequency in an RNA's structural ensemble. To accumulate good statistics and ensure good comparisons to 'ground truth', we carried out simulations of M^2R experiments for structural ensembles of hundreds of naturally derived RNA sequences using the *RNAstructure* simulation package. These *in silico* tests confirmed that observation of strong rescue over all the base pairs of a helix implies high helix frequency (Figure 2C). Importantly, these tests 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 2D, center). As expected, individual base pair tests were noisy, but averaging rescue factors over tests on three or more base pairs gave values that showed a striking correlation with helix frequencies (compare gray and black points, Figure 2C). Furthermore, absence of rescue (rescue factor close to zero; Figure 2D, bottom) gave upper bounds on the helix frequency (Figure 2 Supplement 1). 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. These posterior probability distributions represent our belief, informed by experiment, at all helix frequencies from 0% to 100% and provide a complete representation of our uncertainty; 'fatter' distributions correspond to larger uncertainties in helix frequencies. For ease of reading in the main text and in summary tables, we give median values of the helix frequency posterior probability distributions. Our simulation and experimental studies suggested that compensatory rescue measurements,

including ones that resulted in partial or no detectable rescue, could be used to estimate helix frequencies and uncertainties for those values, though we sought more detailed experimental tests, as described next.

Tests of helix frequency estimation on wild type* add riboswitch *sequence

To test the applicability of M^2R helix frequency estimation for understanding complex RNA structural ensembles, we first applied the method to the *add* riboswitch with and without adenine (Figures 3 and 4). These measurements were necessary to determine baseline values of each helix frequency – the denominators of equations (2) and (3) – before then assessing whether the frequency rises or decreases in response to locking other helices through designed mutations. In addition, prior studies, including detailed NMR analysis, offered detailed helix frequency data on the wild type *add* sequence that might validate or falsify our strategy of measuring rescue factor and using simulations to infer helix frequency.

Overall, compensatory rescue experiments agreed well with prior models. In the absence of adenine, all prior models of the *add* riboswitch predicted that the RNA would show a diverse set of helices. Indeed, aptamer helices P1 (53%) and P2 (31%) were detectable, as they gave partial rescue in mutate-map-rescue experiments. For illustration, Figure 3A shows SHAPE profile quartets and rescue factors for each base pair of P1; the inferred probability distributions over P1 helix frequencies is given as a gold curve underneath, and extends across helix frequencies ranging from 0.2 to 0.75. Helices P3 and P5 were expected to be present, and both were detected by compensatory rescue experiments, with median helix frequencies of 73% and 75%, respectively (brown curves, Figure 4; Figure 3 Supplement 1 gives SHAPE profiles). In addition, compensatory rescue experiments recovered evidence for helices P4A (48%), P4B (55%), P4C (68%), P1B (44%), and, at a lower level, P2B (17%), all proposed as part of *apoB* (Figure 1; left-hand structure in top panel, Figure 3; Figure 3B shows SHAPE profiles and associated partial rescue for each base pair of P4B). The posterior distributions of these helix frequencies, shown as blue curves in Figures 4A and 4B, were centered at intermediate frequencies, as expected. Other helices proposed to occur at low frequencies from M^2-guided modeling (P6, P8 through P17)[14], gave no evidence for their presence (pink curves, Figure 4; Figure 3 Supplement 1); those experiments thus served as negative controls for the M^2R helix frequency estimation method. Overall, these measurements confirmed that the adenine-free state of the *add* riboswitch is quite heterogeneous, with evidence for intermediate frequencies for all helices predicted in prior studies and no further helices.

Turning to the adenine-bound state, all prior models predicted that the riboswitch would shift its helix frequencies to a better defined secondary structure with helices P1, P2, and P3 in the aptamer and P5 (*holo*, Figure 1; right-hand structure in top panel, Figure 3). M^2R experiments testing each of these helices showed visually striking rescue in their SHAPE profiles as mutations disrupted adenine binding, and compensatory mutations restored the adenine-bound profiles (Figure 3C shows SHAPE profiles and rescue factors for each base pair of P1). 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 3C; and gold and brown curves in Figure 4C and D). Measuring rescue factors for other helices, including P1B, P2B, P4A, P4B, and P4C predicted for the adenine-free state and other helices proposed to occur at low frequencies from M^2-guided modeling (P6, P8 through P17)[14], gave no evidence for their presence when adenine was bound (example SHAPE profile quartets for P4B in Figure 3D; all quartets in Figure 3 Supplement 1; and blue and pink curves, Figure 4C and D). In distributions of posterior probabilities, there was a clear separation of these low-frequency helices from the high-frequency helices (Figure 4C and D; and Table 1). Helices P6, P8 through P17 (Figure 3 Supplement 2) again served as negative controls and gave low helix frequencies by compensatory rescue, as expected (pink distributions in Figure 3C and 3D; Table 1). In addition to these measurements based on mutations of each base pair of a helix, 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 (Figure 3 Supplement 3). Overall, M^2R data and helix frequency analysis for the *apo* and *holo* ensembles detected the helices expected from prior studies and provided baseline helix frequency values to infer the correlation or anticorrelation of each pair of these helices.

Four-dimensional Lock-Mutate-Map-Rescue enables measurement of helix-helix correlations

The measurements above showed that locking the aptamer secondary structure through adenine binding reduced the frequencies of helices characteristic of the closing of the gene expression platform (compare upward shift of gold curves – and downward shift of blue curves – between Figures 4A and 4C). Discrimination between MWC and non-MWC models involves determining whether this same allosteric communication between helices occurs even in the absence of adenine ligand. We therefore sought to measure whether locking helices characteristic of the aptamer secondary structure (such as P1) through mutation, rather than adenine binding, also

reduced the frequencies of helices characteristic of the closing of the gene expression platform (such as P4B), and vice versa.

To identify appropriate locking mutants, we noted that a number of double-base-pair compensatory mutants for each of P1, P2, P1A, P4A, and P4B (Figure 3 Supplement 3) exhibited SHAPE profiles that differed from WT RNA but agreed with each other, suggesting that they had independently locked the same state (Figure 5; data for all tested mutants given in Figure 5 Supplement 1). For example, numerous P1- or P2-locking mutants exhibited increased reactivity in the expression distal gene platform region (nts 105-122) relative to the adenine-free WT RNA, at a level comparable to the adenine-bound WT RNA. This observation already suggested that mutation-based locking of P1 or P2 opens the gene expression platform, favoring the MWC model over the non-MWC model, although it depends on an assumption that increase in SHAPE reactivity correlates with reduced base pairing, which does not always hold.[30-31] We also observed similarities in the gene expression platform between these adenine-free P1 and P2-stabilizing mutants and adenine-bound WT RNA in profiles acquired with chemical modifiers besides SHAPE reagents (dimethyl sulfate, glyoxal, ribonuclease V1, terbium(III), and ultraviolet irradiation at 302 nm; Figure 5 Supplement 2). For other helices P1B, P4A, and P4B, we were able to find double-base-pair mutants that gave similar SHAPE profiles to each other, supporting the use of those mutants as helix-locked variants (Figure 5). Interestingly, for each of these three target helices, there was one double-base-pair mutant with a profile nearly indistinguishable from other ones, with spikes in SHAPE reactivity at positions 78 and 85, which also appear in the wild type RNA but not the P1 or P2-lock mutants (red arrows, Figure 5). This observation suggested that these double-base-pair mutants for P1B, P4A, and P4B stabilized the same state; we therefore chose these variants as the lock mutants. Further evidence that these mutants were appropriately locked came from excellent two-state fits that recovered the adenine-free wild type chemical mapping profiles across five modifiers from the profile for any of these P1B- P4A- or P4B- lock mutants with the profiles for any of the P1 or P2-locking mutants (Figure 5 Supplement 2). (P2B and P4C were short two-base-pair helices; since only one candidate double-base-pair lock mutation could be tested and did not allow such cross-checks, we chose to not go forward with P2B and P4C locking experiments.)

Compensatory mutation experiments in locked mutation backgrounds allowed us to confidently determine whether MWC-predicted anticorrelations exist in the adenine-free structural ensemble of the riboswitch. Figure 6A shows M^2R data for the P1 helix in the context of P4B-stabilizing

mutations (lock-P4B; U82A, C83G, G115C, A116U). In contrast to the observation of partial rescue of P1 in the wild type RNA (Figure 3A), these data show no rescue for P1 in the lock-P4B mutational background. The inferred P4B helix frequency drops from 53% to 5%; the shift in posterior probability distribution to lower helix frequency values is clear (faded gold to gold curves; Figure 6B). Stated simply, locking P4B strongly reduces P1. These results support the anticorrelation between P1 and P4B expected from the MWC model, as opposed to the independence of the helices expected from the non-MWC model.

To statistically evaluate the MWC vs. non-MWC model, we computed the correlation value $g(\mathrm{P1},\mathrm{P4B})$ through equation (2), taking into account the full uncertainties in helix frequencies shown in Figure 6B. The resulting posterior distribution for $g(\mathrm{P1},\mathrm{P4B})$ (Figure 7A) is highly skewed away from 1 (the expectation of the non-MWC model) to values close to zero (expectation of the MWC model). Statistical significance in biological studies is often summarized through a *p*-value for a null hypothesis. In our Bayesian framework, an analogous value is the total posterior support for non-MWC models, i.e. the integral of the posterior probability distribution in Figure 7A for $g(\mathrm{P1},\mathrm{P4B}) > 1$. This posterior support is small, with a value of 3.1×10^{-3}, disfavoring the non-MWC model and supporting the MWC model.

As an independent test of the MWC vs. non-MWC models, we carried out the 'flipped' P1-P4B lock-mutate-map-rescue (LM^2R) experiments, in which we locked P1 through mutations (lock-P1; A19U, U20A, A76U, U77A) and assessed the helix frequency of P4B. Again, in contrast to the observation of partial rescue of P4B in the wild type RNA (Figure 3B), the data in the lock-P1 context show no rescue for P4B. Upon locking P1, the P4B helix frequency drops, from 55% to 4%, and there is a striking shift in the posterior probability distribution to lower helix frequencies (faded blue to blue curves; Figure 6D). The posterior support for the non-MWC model $[g(\mathrm{P1},\mathrm{P4B}) > 1]$ from this 'flipped' lock-P1 experiment is somewhat better than the lock-P4B experiment above, but still small (0.039; full posterior distribution in Figure 7B). If the data for the two experiments are combined, the posterior support for the non-MWC model becomes quite small, 4×10^{-5}. Another way to summarize the results is to estimate whether the correlation value $g(\mathrm{P1},\mathrm{P4B})$ for P1 and P4B is not just less than 1 but much less than 1 – a strong anticorrelation. Indeed, the full posterior distribution in Figure 7C is highly skewed towards zero. The median for this correlation value is 0.052 (Table 2), and with 95% posterior support, $g(\mathrm{P1},\mathrm{P4B})$ is less than 0.2. The data agree well with strong anticorrelation predicted by the MWC model but not the non-MWC model.

Further data support the MWC-predicted anticorrelation of aptamer secondary structure helices and the closure of the gene expression platform. LM^2R experiments testing correlation of P1 and P4A were not possible due to overlap of nucleotides between the helices; locking one helix precludes compensatory mutagenesis of base pairs in the other helix. However, we could carry out complete LM^2R experiments probing the correlation between P2 and P4B and the correlation between P2 and P4A. Full data are given in Figure 6-supplement 1, and final posterior distributions for $g(\mathrm{P2,P4A})$ and $g(\mathrm{P2,P4B})$ are shown in Figures 7D and Figure 7E. Both sets of experiments again showed strong skewing of these two correlation values towards zero (anticorrelation; median correlation values of 0.069 and 0.089, respectively; see Table 2). These results again strongly disfavor the non-MWC model (posterior support: $4.1\mathrm{x}10^{-3}$and $1.5\mathrm{x}10^{-3}$, respectively) and supporting the MWC model. We emphasize that these two sets of experiments are independent of each other and of the $g(\mathrm{P1,P4B})$ experiments above.

To gain further insights into the *apo* ensemble, we also carried out LM^2R experiments for pairs of helices that were predicted to co-occur (correlate, rather than anticorrelate) in prior models of the *apo* ensemble (Figure 1). On one hand, LM^2R experiments gave evidence of weak correlation between the two helices characteristic of the aptamer secondary structure, P1 and P2: $g(\mathrm{P1,P2})$ is inferred to be somewhat greater than 1 (Figure 7F). On the other hand, we saw no evidence of correlation for pairs of helices expected to co-occur in the riboswitch when the aptamer secondary structure is not formed (*apoB*, Figure 7G). LM^2R experiments gave posterior distributions for $g(\mathrm{P1B,P4A})$, $g(\mathrm{P1B,P4B})$, and $g(\mathrm{P4A,P4B})$ centered around 1 (Figures 7G-I), indicating no correlation between the occurrence of these helices in the ligand-free ensemble. While the uncertainties in these LM^2R measurements allow correlation values of these helices to be slightly greater than 1, the data are most consistent with a model in which helices P4A, P1B, P4B each appear with significant frequency in the *apo* ensemble, but do not necessarily co-occur, resulting a heterogeneous structural ensemble. Therefore, a single secondary structure such as *apoB* may not be an appropriate description of the structural ensemble. Further evidence for structural heterogeneity came from the incomplete match in SHAPE profiles for mutants that lock different helices of the putative *apoB* state. For example, lock-P4A and lock-P1B showed reproducible differences in SHAPE reactivities at the 5´ end of the RNA (nts 13-20) that were larger than, e.g., differences between lock-P1 and lock-P2 (Figure 5), and we did not discover any alternative locking mutations that brought them into agreement.

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 mechanism 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. This study has introduced a 'four-dimensional' expansion of chemical mapping that offers an experimental route to this information. The strategy 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. Importantly, quantitative inferences of helix-helix correlations from this lock-mutate-map-rescue (LM^2R) workflow do not require interpretation of how SHAPE reactivity corresponds to Watson-Crick pairing probability, which can lead to erroneous conclusions.[28] In that sense, the method is analogous to three-dimensional and four-dimensional pulse sequences that enable structural conclusions to be derived from NMR. Inferences from multidimensional NMR are based on the changes of spectroscopic signals (chemical shifts, magnetization transfer rates) rather than on absolute values of those signals, whose relationships to features of macromolecule structure are not yet understood in quantitative detail.[32]

The LM^2R method currently has limitations regarding the length of RNAs that can be probed. The current throughput of PCR assembly and capillary electrophoretic sequencing enables single-nucleotide-resolution mapping of approximately 192 constructs per experiment at the signal-to-noise needed for statistically sound conclusions. With this throughput, *in vitro* dissection of the co-occurrence or anticorrelation of helices underlying a ~120 nucleotide RNA domain was feasible over a 1-2 month period. However, for RNAs much longer than 200 nucleotides, the expense and time of preparing mutations would likely become prohibitive.

The LM^2R method also currently has limitations regarding which kinds of functional elements can be interrogated for correlation or anticorrelation. Our study has focused on helix elements, which can be selectively probed and selectively locked through mutation of helix base pairs.

However, other functional elements in RNA folding and gene regulation include single-stranded regions and non-Watson-Crick pairings that underlie tertiary structure. For example, in the *add* riboswitch the functional elements in gene expression are the single-stranded AUG and Shine-Dalgarno ribosome binding site. We used the helices in the P4 domain to allow locking and probing (by compensatory rescue) of this region. A more general approach to lock and probe such single-stranded elements might be to sequester them with designed oligonucleotide probes and to assess dissociation constants for such probes, respectively. For tertiary contacts or non-Watson-Crick paired junctions, it is less clear how to generally lock the elements or to carry out compensatory mutagenesis to probe the elements, in which case, LM^2R may need to continue relying on Watson-Crick helices near these elements as proxies.

As a final caveat, the current LM^2R method relies on the RNA system being at equilibrium so that the correlation equations (1)-(3) are valid and also so that extensive computer simulations can be carried out that allow conversion of observed rescue factors to helix frequencies. It will be important to evaluate whether LM^2R, or some time-dependent extension of the method, could be applied to RNAs that change their structures in non-equilibrium scenarios involving co-transcriptional RNA folding, recruitment of energy-dissipating machines like the ribosome or RNA polymerase, or other behaviors that depend on intrinsically kinetic mechanisms.[4, 33]

Resolving the core mechanism of the* add *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* supports a Monod-Wyman-Changeux (MWC, or population shift, or conformational selection) model of allostery rather than a non-MWC revision proposed after detailed NMR experiments.[26, 34] The two prior models differed in one core aspect: whether structuring of the molecule's aptamer region is allosterically communicated to expose a ribosome binding site and turn on *add* mRNA gene expression, even in the absence of adenine (Figure 1C and 1D). In LM^2R experiments locking P1 and P2, which are specific to the aptameric secondary structure, the rescue data showed depletion 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 1A, 1C; Figure 8, *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 helix formation and the 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 1A and Figure 1B; or Figure 1C and Figure 1D). 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. 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.[24-25]

Comparison to prior data and proposals

Figure 8 summarizes our current model of the *add* riboswitch. Our results indicate a strong anticorrelation of helices P4A and P4B with the helices P1 and P2, and so we continue to show the *apo* ensemble as split between *apoB* and *apoA* of the MWC model (Figure 1), but now highlight alternative or partial pairings (dashed lines) to better convey the likely heterogeneity of these states. This model 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 suggested that the riboswitch is rather 'leaky', forming the aptamer secondary structure at ~40% frequency at 20–30 °C, even without adenine.[11-21, 35] Second, X-ray crystallographic studies have been limited to constructs with just the aptameric region[22, 36-43]; these structures all show P1, P2, and P3 helices, in accord with Figure 8. Third, the detailed NMR measurements that suggested the alternative *apoA* (Figure 1B) were also consistent with the *apoA* base pairings in the MWC model (Figure 1A): 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; our LM^2R data strongly support the latter scenario (Figures 6 and 7). 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 (Figure 5 Supplement 1), both the NMR study and our LM^2R measurements detect suppression of P4A in this background (Figure 6 Supplement 2), supporting the anticorrelation between P2 and P4A in Figure 8. Fifth, our SHAPE and multi-probe chemical mapping data suggest partial opening of P5 in *apoA* and

holo (brown curves, Figure 4B), and our LM^{2}R analysis suggests only partial formation of P2B and P1B when P4A is locked (Figure 6 Supplement 1 and Figure 7 Supplement 1D). These observations are again consistent with NMR studies which could unambiguously detect P5 and estimate 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 (dashed lines in *apoA* of Figure 6) and stabilization upon adenine binding to *holo*.[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 ligand present.[27] Seventh, the prior detailed NMR analysis led to a compelling 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 whether *apoA* involves opening the gene expression platform, approximately the same temperature compensation occurs in the MWC conformational selection model of (see Figure 8 Supplement 1).

There is now concordance across numerous measurements for an MWC model that posits strong secondary structure similarities between an *apoA* state that transiently forms the aptamer secondary structure without adenine ligand and the ligand-bound *holo* state. We also propose herein that further heterogeneity is likely to be present in members of the *apo* ensemble that do not present the aptamer secondary structure, grouped into *apoB*. LM^{2}R measurements indicate weak or no correlation of the tested signature helices of *apoB* (P1B, P4A, and P4B; Figure 7). Furthermore, numerous alternative secondary structures are possible for this state (see, e.g., Figure 1 Supplement 3). The presence of these myriad alternative helices, each at low population, would explain their detection difficulty with this and prior bulk equilibrium techniques but their appearance in single-molecule measurements.[17] Supporting this picture, higher-order M^2 analysis in single mutant backgrounds isolate and confirm numerous alternative secondary structures for the *add* riboswitch (Figure 8 Supplement 2 and Figure 8 Supplement 3). We note that there are no functional reasons for the *apoB* ensemble 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 the P4 domain may shift in different solution conditions and flanking sequences while still being compatible 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 *apo* structure while retaining other members of the *apo* ensemble with the P4 domain could evolve quickly and 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.[29, 44] 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 RNA chemical mapping signals in cells and tissues, the lock-mutate-map-rescue approach developed here offers the possibility of dissecting complex structural ensembles for these shape-shifting molecules *in situ*.

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 and R35 GM122579 (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 (AAAGAAACAACAACAACAAC) on the 3´ end. The sequence of interest is flanked at each end by a hairpin with single-stranded buffering.[45] The full sequence probed was:

GGACACGACUCGAGUAGAGUCGAAUGCGCUUCAUAUAAUCCUAAUGAUAUGGUUUGGGAGUUUCUACCAAGAGCCUUAAACUCUUGAUUAUGAAGUCUGUCGCUUUAUCCGAAAUUUUAUAAAGAGAAGACUCAUGAAUUACUUUGACCUGCCGACCGGAGUCGAGUAGACUCCAACAAAAGAAACAACAACAACAAC.

Hairpin sequences are in gray; *add* riboswitch sequence is underlined. The primer assembly scheme and plate orders for all constructs were automatically designed by Primerize-2D.[46]

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

M^2, M^2R and LM^2R chemical mapping were carried out in 96-well format as described previously [28, 47]. 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 $MgCl_2$, 50 mM Na-HEPES pH 8.0 (with or without 5 mM adenine), and returned to ambient temperature (24 °C). 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 ambient 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 (ddH_2O); 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 $TbCl_3$ in ddH_2O; FMN, 2 mM flavin mononucleotide in ddH_2O. 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 photo 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 ddH_2O (and removal of light source) for 'no modification' controls, FMN, and UV.

CE Data Processing

The HiTRACE 2.0 software was used to analyze CE (capillary electrophoresis) data.[48-50] Electrophoretic traces were aligned and baseline subtracted using linear and non-linear alignment routines as previously described.[45] 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.[45] 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 M^2, M^2R, LM^2R and multi-probe mapping, have been deposited at the RNA Mapping Database (http://rmdb.stanford.edu)[30] 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, ADD71_STD_0001, ADD128_STD_0001, ADD140_STD_0001, ADD140_DCP_0001, ADD140_DCP_0002, RNAPZ5_RSQ_0001, TRP4P6_RSQ_0001, 16SFWJ_RSQ_0001.

Simulations to infer helix frequencies and correlation values to rescue factors

All RNA families with length between 100 and 250 nt were screened from the Rfam database (http://rfam.xfam.org/) 11.0 FASTA file.[51] A sequence from each family was randomly picked as representative, and its *in silico* base pair probability (BPP) matrix was simulated. Base pairing positions to be tested by M^2R were selected from this BPP matrix, for BPP values greater than or equal to 1% over any candidate helices that could form 3 Watson-Crick pairs or more. (The 1% BPP cutoff conveys our knowledge that these candidate helices are likely to have finite but potentially low frequency, based on prior literature analysis or experiments; changing the cutoff to lower values such as 0.1% gave similar results and do not change conclusions of the manuscript). For each helix, the rescue factor metric (Equation 4), averaged over simulated M^2R quartets for each base pair in the helix, and the mean BPP across those base pairs were calculated. The MATLAB source code for Rfam simulation and rescue factor estimates are available in a GitHub repository: https://github.com/t47io/m2r_simulations. To generate a posterior distribution over helix frequencies F given an experimental rescue factor for a helix, all RFAM samples were collected with the same number of base pairs as the candidate helix and simulated rescue factor in the same bin as the experimental rescue factor (binwidth of 0.05 was chosen to allow for sufficient sampling). The helix frequencies F(helix) of these samples (as shown in Figures 4 and 6) were visualized as the posterior distribution through a kernel-density estimate (with the *ksdensity* function in MATLAB). The representative helix frequency (presented in Table 1 and in main text) was taken as the median helix frequency of these RFAM samples.

Evaluating correlation between two helices requires determining the posterior distribution over the correlation value $g(\text{helix 1}, \text{helix 2})$, given measurements of $F(\text{helix 1})$ and $F(\text{helix 1} | \text{helix 2})$ (see e.g., equation 2). Two sets of RFAM helix frequency samples corresponding to the experimental rescue factor for helix 1 in the wild type background and for helix 1 in the lock-helix2 mutational background were collected. The posterior distribution of

$g(\text{helix 1},\text{helix 2}) = \frac{F(\text{helix 1}|\text{helix 2})}{F(\text{P1})}$ was then estimated based on taking each helix frequency sample from the latter set (lock-helix2) and dividing by the helix frequency for the former set (wild type). Combination of posterior distributions for a lock-mutate-map-rescue experiments and its 'flipped' variant (in which helix 1 was locked, and helix 2 subjected to mutate-map-rescue) was carried out by multiplying kernel density estimates (KDE) of the posterior distributions $g(\text{helix 1},\text{helix 2})$ for the two separate experiments, assuming a flat prior over correlation values. Changing the bandwidth of the KDE estimates changed posterior support values presented in the main text by less than a factor of 2. For $g(\text{P1},\text{P4B})$, $g(\text{P2},\text{P4A})$, and $g(\text{P2},\text{P4B})$, which discriminated between the MWC and non-MWC models, the combination of posterior distributions was also carried out with exponential fits to the individual posterior distributions; final results for posterior support were the same within a factor of 2.

TABLES

Wild type [a]			In locked contexts [b]		
[adenine]	0 mM	5 mM			0 mM
P1	**53%**	**79%**	Lock P1	*P2*	**42%**
P2	**31%**	**64%**		*P4B*	**4%**
P3	**73%**	**56%**	Lock P2	*P1*	**76%**
P5	**75%**	**29%**		*P4A*	**4%**
P1B	**44%**	**3%**		*P4B*	**4%**
P2B	**17%**	**17%**	Lock P4A	*P1B*	**9%**
P4A	**48%**	**4%**		*P2B*	**6%**
P4B	**55%**	**4%**		*P2*	**6%**
P4C	**68%**	**4%**		*P4B*	**55%**
P6	**6%**	**5%**	Lock P1B	*P2B*	**53%**
P8	**9%**	**4%**		*P4A*	**55%**
P9	**12%**	**3%**		*P4B*	**80%**
P10	**3%**	**3%**		*P3*	**82%**
P11	**12%**	**16%**	Lock P4B	*P1B*	**8%**
P12	**10%**	**6%**		*P2B*	**4%**
P13	**9%**	**4%**		*P1*	**5%**
P14	**4%**	**4%**		*P2*	**6%**
P15	**12%**	**36%**		*P4A*	**47%**
P16	**14%**	**8%**	Mut P2	*P1*	**38%**
P17	**4%**	**4%**		*P4A*	**8%**
				P4B	**13%**

Table 1. Helix frequency estimates for single base-pair M^2R, double base-pair M^2R, LM^2R experiments.

Median helix frequencies from Rfam simulations corresponding to each experimentally observed rescue factor are reported. The experimentally observed rescue factor for each helix was averaged across all tested base pairs. Full posterior distributions are presented in main text Figure 4 and Figure 7 Supplement 1.

a. Median helix frequencies inferred from mutate-map-rescue, compensatory rescue read out by chemical mapping across the transcript.

b. Lock-mutate-map-rescue, mutate-map-rescue carried out in the context of mutations that 'lock' the specified helices.

Helix-helix	Correlation value
P1-P4B	0.052
P2-P4B	0.069
P2-P4A	0.089
P1-P2	1.315
P4A-P4B	0.902
P1B-P4A	0.44
P1B-P4B	0.534

Table 2. Helix-helix correlation estimates from LM^{2}R experiments.

Median values are reported. Full posterior distributions are presented in main text Figure 6 and Figure 7 Supplement 1.

LEGENDS

Figure 1. Models of the *add* riboswitch.

(A-B) Three-state models for conformational changes and ligand binding in the *add* riboswitch. In two prior models, *apoB* state (left) is OFF (Shine-Dalgarno sequence and AUG start codon sequestered) and is also incompetent for ligand binding due to the aptamer's rearrangement (blue background in both regions). The ligand-bound state *holo* (right) is ON with Shine-Dalgarno sequence and AUG codon (gene expression platform, cross-hatched) exposed and the aptamer region folded into a secondary structure compatible with binding adenine (gold background in both regions; adenine is purple). The *apoA* state (center), defined as the transient adenine-free conformation with an aptamer fold ready for ligand binding (gold background), differs in the two models. (*A*) In the Monod-Wyman-Changeux model , *apoA* is assumed to be ON, with a similar expression platform to *holo*. (*B*) In a non-MWC model proposed in a prior study[26], the *apoA* state is 'mixed', with the aptamer fold ready for binding (gold) but the expression platform turned OFF as the Shine-Dalgarno sequence and AUG codon are sequestered into helices P4B and P4C (blue). *(C-D)* Cartoons of possible ligand-free structural ensembles allowing for myriad alternative secondary structures with only some helices shared with those in *(A-B)* or completely different helices. This description does not assume a three-state decomposition of (*A-B*), which does not take into account the possibility of other possible helices (pink arrows). In an MWC scenario *(C)*, helices characteristic of the aptamer secondary structure (such as P1, highlighted gold) and helices characteristic of closed gene expression platform (such as P4B, highlighted blue) appear in different members of the structural ensemble, but never together. The helix frequencies are anticorrelated (equation above cartoon). In non-MWC scenario *(D)*, both kinds of helices appear together in 'mixed' secondary structures (marked with red arrows) at joint frequencies similar to the product of their individual occurrence frequencies (equation above cartoon).

Figure 1 Supplement 1. One-dimensional SHAPE profiles of *add* riboswitch constructs.

Three different constructs were probed by SHAPE: *(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) including the first 15 nt of the coding region. Reactivities are normalized based on GAGUA penta-loop reference hairpins that flank the RNA (not shown). Ligand-dependent reactivity changes in the aptamer domain (nts 15-83) occur in all three constructs. The 113-nt construct did not show significant

switching for the Shine-Dalgarno sequence (nts 112-114) or AUG start codon (nts 120-122), highlighted in gray, with cross-hatched boxes between panels. The 128-nt construct was therefore used for all measurements in this manuscript. See also reference.[52] Measurements were acquired in 10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, in the absence or presence of ligand (5 mM adenine).

Figure 1 Supplement 2. Mutate-and-map (M^2) experiments on the *add* riboswitch. Measurements were acquired in 10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, in (*A*) the absence of adenine ligand and (*B*) in the presence of ligand (5 mM adenine).

Figure 1 Supplement 3. Candidate alternative helices from M^2 analyses.
(A) Helices modeled to be in the adenine riboswitch based on M^2 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.[47] While not recovering the actual helix frequencies in the solution ensemble, this initial analysis revealed a large number of possible helices that needed to be considered in modeling the ensemble, particularly for the adenine-free state.
(B) Helices P6, P8-17 identified in the M^2-guided structure analysis of *(A)* as well as a different model of the adenine-free ensemble based on the REEFFIT method[14] and the M^2 data. These helices were individually tested by single base-pair M^2R (see main text and Figure 4). The presented structures are for illustration of the helices only, and do not imply co-existence in the same structure. Helix name P7 was skipped to avoid a naming conflict with a previous REEFFIT investigation under different solution conditions.[14]

Figure 2. Mutate-map-rescue (M^2R) enables inference of helix frequencies.
(A) Interrogation of base pairs by compensatory mutagenesis and chemical mapping, illustrated on a four-way junction (nts 126-235) of the *E. coli* 16S ribosomal RNA. Experiments involve 'quartets' of SHAPE profiles for the wild type RNA, variants with single mutations on each side of a candidate base pair, and a variant with double mutations that could rescue Watson-Crick pairing of the candidate pair. Experiments involving possible pairings of G201 result in no obvious rescue (G201C-C207G), clear rescue (G201C-C217G), or an intermediate result (G201C-U216G), termed 'partial rescue'. In partial rescue, some disruptions observed in single

mutants are rescued, but other features are not rescued or arise in the double compensatory mutant (marked with yellow arrow). Data are reproduced from ref.[28].

(B) Cartoon of structural ensembles that explain the 'partial rescue' scenario. A fraction of structures in the wild type ensemble have target base pair (cyan/blue, within helix outlined in gold); this picture assumes that those structures are disrupted in the single mutant but restored with similar fraction in the double mutant ensemble.

(C) Correlation of observed rescue factor against helix frequency in simulated mutate-map-rescue experiments. Data are for helices with length longer than 2-bp for RNAs across 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 shaded in grays based on 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 M^2R quartets illustrating full compensatory rescue (high rescue factor), partial rescue (medium rescue factor) and no rescue (low rescue factor).

Figure 2 Supplement 1. Correlation between rescue factor and helix frequency from Rfam simulations of mutate-map-rescue (M^2R).

(A) Boxplots showing helix frequency vs. rescue factor, over all simulated helices. Purple lines mark median, open boxes cover 25th to 75th percentile, cyan whiskers show 5th and 95th percentile, and gray plus signs label outliers. Rescue factor ranges are colored in red, yellow, or green for low or no rescue (0 to 0.3), partial rescue (0.3 to 0.7), and strong rescue (0.7 and higher).

(B) Boxplots showing how uncertainties in helix frequency (e.g. the 25th to 75th percentile range spanned by open boxes) decrease with the number of base pairs tested by mutate-map-rescue: 3, 4, 5, 6, and ≥ 7 base-pairs.

Figure 3. Experimental data and frequency estimates for helices P1 and P4B in the *add* riboswitch structural ensembles. Top: two example secondary structures showing sequences of the *add* riboswitch and helices probed. The presented structures are for illustration of the helices only, and do not imply co-existence in the same structure. In (*A*-*D*), each panel shows (top) experimentally measured mutate-map-rescue (M^2R) quartets for each base pair of a probed helix and (below) posterior distributions over helix frequency, estimated from the mean rescue factor in the M^2R experimental data and simulations (see Figure 2 Supplement 1). Experiments are shown

probing (*A*) P1 helix without adenine, (*B*) P4B helix without adenine, (*C*) P1 helix with 5 mM adenine, and (*D*) P4B helix with 5 mM adenine. Measurements were acquired in 10 mM $MgCl_2$, 50 mM Na-HEPES pH 8.0.

Figure 3 Supplement 1. All 188 single base-pair M^2R 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 2 Supplement 1.

Figure 3 Supplement 2. All 82 double base-pair M^2R 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 a previous study[26] is also included. Rescue factor for each quartet is given in the title and colored as in Figure 2 Supplement 1. Mutants chosen to 'lock' P1, P2, P4A, P1B, and P4B are labeled by green lock symbol.

Figure 4. Summary of helix frequencies inferred for the *add* riboswitch.
Posterior distributions over helix frequencies *(A)* without adenine ligand and *(B)* with 5 mM adenine ligand. Curves have been smoothed through kernel density estimation. Sub-panels separate helices expected to have high frequencies with adenine bound (gold); 'positive control' helices expected to be present irrespective of adenine (brown); helices expected to be present only in the absence of adenine (blue); additional helices that are not expected to have high frequencies (pink). (*C*-*D*) same distributions as in (*A*-*B*) but represented as cumulative distributions and overlaid together without smoothing.

Figure 5. Double-base-pair mutations to lock each helix of the adenine riboswitch. One-dimensional SHAPE reactivity of candidate constructs for locking each *add* riboswitch helix. In each candidate, two base pairs of the helix were switched to alternative Watson-Crick pairs (Figure 3 Supplement 2). For each helix targeted, at least two sets of mutations gave SHAPE reactivities distinct from wild type but similar to each other. Red arrows mark residues in switch region that are reactive in wild type and strongly reactive in lock mutants chosen for P1B, P4A, and P4B.

Figure 5 Supplement 1. All double-base-pair mutations tested as possible locking mutations. One-dimensional SHAPE reactivity of candidate constructs for locking each *add* riboswitch helix. The MutP2 variant is derived from reference [26]; it shows less SHAPE increase in the gene expression platform (nts 105-122) than other tested lock P2 mutants.

Figure 5 Supplement 2. 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. χ^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 data normalization for a quantitative fitting.

(B) Comparison of fitted SHAPE reactivity profile (pink) to observed WT (black) in absence of adenine. 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, although discrepancies (nts 79-80) suggest that there are more than the two states represented by lock-P1 and lock-P4A in the WT adenine-free ensemble.

(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 (nts 82-140) with adenine-bound WT (purple) even in the absence of adenine (solid gold and green).

Figure 6. Anticorrelation between P1 and P4B helices in ligand-free *add* riboswitch structural ensemble, inferred through lock-mutate-map-rescue (M^2R). *(A-B)* M^2R quartets probing P1 in the context of lock-P4B mutations show no rescue. (*C-D*) M^2R quartets probing P4B in the context of lock-P1 mutations show no rescue. Panels *(A,C)* give M^2R quartets for each base pair of the probed helix in the context of the mutations locking the other helix. Panels (*B,D*) give posterior probability distributions over helix frequency, estimated from the experimental M^2R rescue factors in locked background and wild type background (solid and dotted curves, respectively).

Figure 6 Supplement 1. All 82 single base-pair LM^2R quartets for *add* riboswitch. Helices P2 and P4B are tested in context of lock-P1 stabilizing mutations; helices P1, P4A, and P4B are tested in lock-P2 stabilizing mutations; helices P2, P1B, P2B, and P4B are tested in context of lock-P4A stabilizing mutations; helices P2B, P4A, P4B, and P3 are tested in context of lock-

P1B stabilizing mutations; helices P1, P2, P1B, P2B, and P4A are tested in context of lock-P4B stabilizing mutations; all under no adenine condition. Rescue factor for each quartet is given in the title and colored red, yellow, or green as in Figure 2 Supplement 1.

Figure 6 Supplement 2. All 15 single base-pair LM^{2}R 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 2 Supplement 1.

Figure 7. Correlation analysis for functionally important helix-helix pairs in the ligand-free *add* riboswitch structural ensemble. Posterior probability distributions, smoothed through kernel density estimation, for correlation of (*A*-*C*) P1 and P4B, based on LM^{2}R experiments (*A*) locking P4B and probing P1, (*B*) locking P1 and probing P4B, and (*C*) combination of experiments of (A) and (B); and (*D*-*I*) other pairs of helices, based on combination of experiments locking each helix and probing the other and vice versa. Helix pairs are (*D*) P2 and P4A, (*E*) P2 and P4B, (*F*) P1 and P2; (*G*) P1B and P4A; (*H*) P1B and P4B; and (*I*) P4A and P4B.

Figure 7 Supplement 1. Posterior probability distributions over helix frequencies and helix-helix correlations for individual LM^{2}R experiments. In each panel, top two sub-panels show posterior probability distributions, smoothed through kernel density estimation, over helix frequency, estimated from the experimental M^{2}R rescue factors in locked background and wild type background (solid and dotted curves, respectively). In each panel, bottom two sub-panels show posterior probability distributions, smoothed through kernel density estimation, for correlation of the helix-helix pairs inferred from the corresponding top subpanels. Helix pairs are (*A*) P2 and P4A, (*B*) P2 and P4B, (*C*) P1 and P2; (*D*) P1B and P4A; (*E*) P1B and P4B; and (*F*) P4A and P4B. These curves are the distributions that are combined to give final posterior distributions for helix-helix correlations shown in Figure 7 *(D)-(I)*, respectively.

Figure 8. Current model for the *add* riboswitch structural ensemble.
The proposed model favors a Monod-Wyman-Changeux (population shift, conformational selection) model of allostery. The ligand-free *apoB* state (left) sequesters the Shine-Dalgarno sequence and AUG codon in the P4 domain, precluding gene expression. Our measurements show that sampling of the aptamer secondary structure is strongly anticorrelated with formation of the P4 domain; it is therefore valid to represent it as a separate state (*apoA*, middle). This *apoA* state is structurally similar to the ligand-bound *holo*, including not just the aptamer region but

also the gene expression platform, as predicted by an MWC framework (right). Our measurements and other studies are consistent with sampling of additional alternative structures in the *apoB* state (dashed blue lines), a partially formed P5 in *apoA*, and additional stabilization of the aptamer through coupled tertiary contacts upon adenine binding (dashed black lines in *apoA*, solid yellow lines in *holo*).

Figure 8 Supplement 1. Simulation of temperature dependence on switching efficiency. Thermodynamic parameters used are: ΔH_{Kpre} = 52 kJ·mol^{-1}, ΔS_{Kpre} = 167 J·mol^{-1}·K^{-1}, ΔH_{Kd} = 238 kJ·mol^{-1}, ΔS_{Kd} = 110 J·mol^{-1}·K^{-1}. Ligand concentrations: [L_{low}] = 0.01 μM; [L_{high}] = 1.0 μM for *(top)* panels and [L_{low}] = 1.5 μM for *(bottom)* heatmaps. The *(top)* population linear plots used [*RNA*] = 1.5 nM; while *(bottom)* temperature vs. [RNA] heatmap simulated [*RNA*] = 0 – 100 nM for T = 0 – 40 °C. Parameters were derived from prior study reviewing the temperature compensation model.[53]

Figure 8 Supplement 2. Models of additional structures that the *add* riboswitch can form, based on mutate-and-map 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. Models are for dominant structures of the RNA based on higher-order mutate-and-map data presented in Figure 8 Supplement 3. 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 ease at which these alternative structures can be stabilized for the *add* riboswitch suggest that they occur for the wild type RNA at low frequency, and might explain deviations of SHAPE profiles and LM^{2}R data across mutants that lock different helices of the *apoB* secondary structure (Figures 5 and 6).

Figure 8 Supplement 3. Higher-order mutate-and-map on selected single mutant backgrounds. *(A)* A109U, *(B)* G78C, *(C)* A116U, *(D)* G44C, *(E)* G37C, and *(F)* G81C. A comparison of 1D SHAPE profile to WT under both no adenine and 5 mM adenine conditions are shown along with complete M^2 datasets; inferred structures are given in Figure 8 Supplement 2.

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. Jones, C. P.; Ferre-D'Amare, A. R., Long-Range Interactions in Riboswitch Control of Gene Expression. *Annu Rev Biophys* **2017,** *46*, 455-481.
8. Fernández-Luna, M. T.; Miranda-Ríos, J., Riboswitch folding: One at a time and step by step. *RNA Biol.* **2008,** *5* (1), 20-23.
9. Gilbert, S. D.; Batey, R. T., Riboswitches: Fold and function. *Chem. Biol.* **2006,** *13* (8), 805-807.
10. Breaker, R. R., Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* **2012,** *4* (2), a003566.
11. 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.
12. 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.
13. Batey, R. T., Structure and mechanism of purine binding riboswitches. *Q. Rev. Biophys.* **2012,** *45* (3), 345-381.
14. 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.
15. Stoddard, C. D.; Widmann, J.; Trausch, 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. *J. Mol. Biol.* **2013,** *425* (10), 1596-1611.

16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. Lemay, J.-F.; Lafontaine, D. A., Core requirements of the adenine riboswitch aptamer for ligand binding. *RNA* **2007,** *13* (3), 339-350.
22. 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.; Zatsepin, 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.
23. Gunasekaran, K.; Ma, B.; Nussinov, R., Is allostery an intrinsic property of all dynamic proteins? *Proteins* **2004,** *57* (3), 433-43.
24. Changeux, J. P.; Edelstein, S. J., Allosteric mechanisms of signal transduction. *Science* **2005,** *308* (5727), 1424-8.
25. 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.
26. 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.
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. 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.
29. Tian, S.; Das, R., RNA structure through multidimensional chemical mapping. *Q. Rev. Biophys.* **2016,** *49* (e7), 1-30.

30. Cordero, P.; Lucks, J. B.; Das, R., An RNA Mapping DataBase for curating RNA structure mapping experiments. *Bioinformatics* **2012,** *28* (22), 3006-3008.
31. Deng, F.; Ledda, M.; Vaziri, S.; Aviran, S., Data-directed RNA secondary structure prediction using probabilistic modeling. *RNA* **2016,** *22* (8), 1109-19.
32. Wuthrich, K., NMR studies of structure and function of biological macromolecules (Nobel Lecture). *J Biomol NMR* **2003,** *27* (1), 13-39.
33. Watters, K. E.; Strobel, E. J.; Yu, A. M.; Lis, J. T.; Lucks, J. B., Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat Struct Mol Biol* **2016,** *23* (12), 1124-1131.
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. 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.
36. Stoddard, C. D.; Widmann, J.; Trausch, 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.
37. Zhang, J.; Ferré-D'Amaré, Adrian R., Dramatic Improvement of Crystals of Large RNAs by Cation Replacement and Dehydration. *Structure* **2014,** *22* (9), 1363-1371.
38. 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.
39. Edwards, A. L.; Batey, R. T., A Structural Basis for the Recognition of 2f 2Binding by the Purine Riboswitch in the *Journal of Molecular Biology* **2009,** *385* (3), 938-948.
40. 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.
41. 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.
42. 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.
43. 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.
44. Leamy, K. A.; Assmann, S. M.; Mathews, D. H.; Bevilacqua, P. C., Bridging the gap between in vitro and in vivo RNA folding. *Q Rev Biophys* **2016,** *49*, e10.
45. 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.

46. Tian, S.; Das, R., Primerize-2D: automated primer design for RNA multidimensional chemical mapping. *Bioinformatics* **2017**.
47. 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.
48. 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.
49. 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.
50. 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.
51. 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.
52. Ali, M. Conformational transitions in RNA probed in riboswitch systems. Stanford University, 2011.
53. 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.

APPENDIX

APPENDIX RESULTS

Conventional SHAPE and Mutate-and-Map Measurements

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 (main text Figure 1). In the absence of adenine, the *add* riboswitch SHAPE profile (main text Figure 1 Supplement 1) showed protections and exposures throughout the aptamer region (nts 14-80), and incubation with 5 mM adenine gave reduced reactivity throughout this 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, weak 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 (main text Figure 1 Supplement 1C). 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 (main text Figure 1A and 1B). 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) and therefore does not discriminate between the models.

Overall, while consistent with available models, these SHAPE data do not discriminate between them due to difficulties in interpreting these data: changes in SHAPE reactivity do not necessarily reflect changes in base pairing frequencies. Nucleotides that do not form Watson-Crick pairs can appear protected or reactive depending on changes in their local environment. Conversely. nucleotides that form Watson-Crick pairs can appear partially reactive, especially if the pairs are at the edges of helices. For these reasons, we turned to mutate-and-map measurements (described next) and higher-dimensional techniques (compensatory rescue; described in the main text) that do not require detailed interpretations of the absolute reactivity values at particular nucleotides, but rather changes in reactivities integrated across all nucleotides.

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 main text Figure 1 Supplement

2. 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 (main text Figure 1 Supplement 3).

Mutate-Map-Rescue (M^2R) – Application Generality and Automated Classification

Our prior work on mutate-map-rescue (M^2R) 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 M^2R approach by applying it to other RNA elements. First, we deployed M^2R 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 M^2-based model with a 2-register shift (Appendix Figure 2). For each base pair tested, we compared the SHAPE profiles of WT (wild type), two 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, M^2R 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 M^2R on the P4-P6 domain of the *Tetrahymena* ribozyme[3] to test for presence of crystallographic P5c against an alternative suggested by mutate-and-map analysis (Appendix Figure 2). Finally, M^2R-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 M^2R on a larger scale, we turned to *in silico* simulation to generate a much greater number of RNA test cases. Specifically, we asked: do M^2R-validated base pairs have high base-pairing probability (BPP) *in silico*? To answer this question, we sought to automate the calling of M^2R result with an automated classifier that would evaluate M^2R 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:

$$\text{Rescue factor} \equiv 1.0 - \frac{RMSD(\text{WT,AB})}{\max(RMSD(\text{WT,A}),\ RMSD(\text{WT,B}))}$$

where

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

and d_A and d_B are vectors of normalized SHAPE profiles, with the same length of N (see Appendix 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 (Appendix Figure 2).

Next, we simulated M^2R *in silico* on 325 RNA families from the Rfam database[6] 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%. Appendix Figure 3 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 (main text Figure 2C). 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 (Appendix Figure 3), 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.) (Appendix Figure 1). Moreover, the BPP distribution of the simulated *in silico* counterparts for *add* riboswitch follows the same trend as Rfam tests, in agreement with its generality. The performance of such a simple classifier was acceptable and supported our heuristic for the rescue factor (Appendix Figure 3). 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 6); 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 (Appendix Figure 1).

Adenine Riboswitch Construct Design

We performed the M^2R pipeline on a 128-nt *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.[7] We chose this longer *add* version after reproducing prior work in our laboratory[8-9] finding that the sequence context has an effect on the folding landscape of this riboswitch (main text Figure 1 Supplement 1). Although the 112-nt construct showed weak ligand-responsive chemical reactivity changes in the SD region under the salt conditions used in NMR [9], it does not show noticeable switching under our in vitro solution conditions. The extended construct gives chemical reactivity changes under all conditions tested.

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 *add* riboswitch samples two states: *apoA* and *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 *apoA* and *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) [10], glyoxal [11], RNase V1 [12], terbium(III) [13-14], and ultraviolet irradiation at 302 nm [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 *apoA* vs. *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 *apoA* state and 52% of the *apoB* state recovered the data measured for the adenine-free wild type riboswitch within experimental errors, except at nucleotides 79-81 (main text Figure 5 Supplement 1A-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 [16]. These values for the *apoA* vs. *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.

APPENDIX 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^2R 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:

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

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:

$$\max(RMSD(\mathrm{WT,A}), RMSD(\mathrm{WT,B})) \qquad (1)$$

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):

$$\frac{RMSD(\mathrm{WT,AB})}{\max(RMSD(\mathrm{WT,A}),\ RMSD(\mathrm{WT,B}))} \qquad (2)$$

[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., in revisions), 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.

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 χ^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 σ_{WT}, σ_{apoA}, and σ_{apoB} are errors (standard deviation) of d_{WT}, d_{apoA}, and d_{apoB}. The parameter α is the fraction of the *apoA* state, ranging from 0 to 1. χ^2 is summed over all or a subset of nucleotide positions *i* as specified below. χ^2 scores are plotted against α, and the α value with minimum χ^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.

APPENDIX FIGURE LEGENDS

Appendix Figure 1. Base-pair-wise classification of M^2R and LM^2R 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^2R, *(B)* double base-pair M^2R, and *(C)* LM^2R and MutP2 are shown. Each symbol represents one quartet testing a single base-pair (or 2 adjacent base-pairs in *(B)*).

Appendix Figure 2. M^2R quartets and helices tested for P4-P6 and GIR1.
(A) M^2R 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) M^2R quartets of GIR1 ribozyme testing P5, alt-P5, P4, and pk2-5.

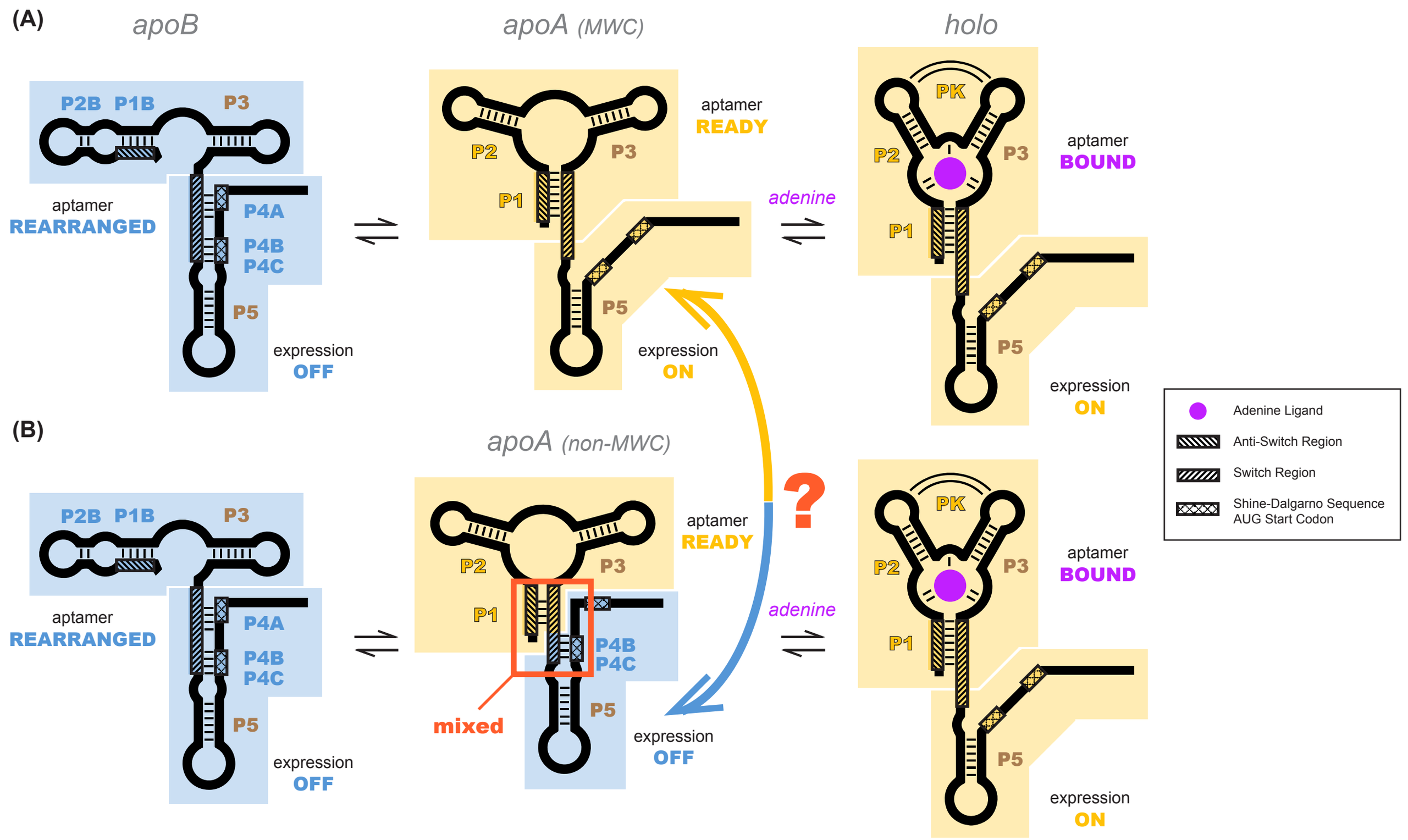
Appendix Figure 3. Performance of 4-bin auto-score classifier. Histogram of *in silico (A*) training data (16S-FWJ, P4P6, GIR1, Hox, PB2; total of 162 quartets), *(B)* test data (*add* single and double base-pair M^2R; total of 242 quartets), and (*C)* Rfam family (total of 62484 quartets) simulations are shown by their classification, and grouped by the *in silico* predicted base-pairing probability (BPP).

APPENDIX 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.

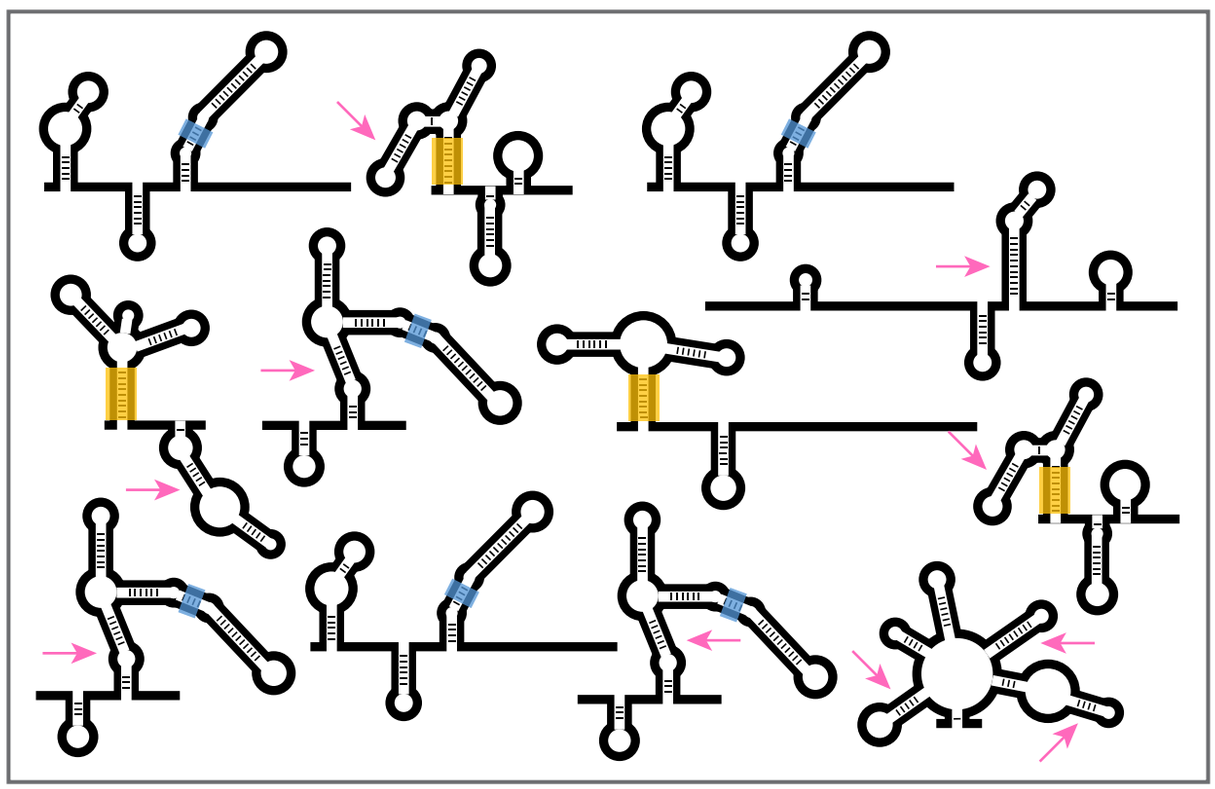
3. Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A., Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* **1996,** *273* (5282), 1678-1685.
4. Xue, S.; Tian, S.; Fujii, K.; Kladwang, W.; Das, R.; Barna, M., RNA regulons in Hox 5´ UTRs confer ribosome specificity to gene regulation. *Nature* **2014,** *517* (7532), 33-38.
5. Hagey, R.; Tian, S.; Pham, E.; Elazar, M.; Kladwang, W.; Taubenberger, J.; Das, R.; Glenn, J., A novel RNA secondary structural element modulates Influenza A virus packaging. **2017**, submitted.
6. 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 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.

20. Tian, S.; Das, R., Primerize-2D: Automated primer design for RNA multidimensional chemical mapping. *Bioinformatics, in press* **2016**.
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.



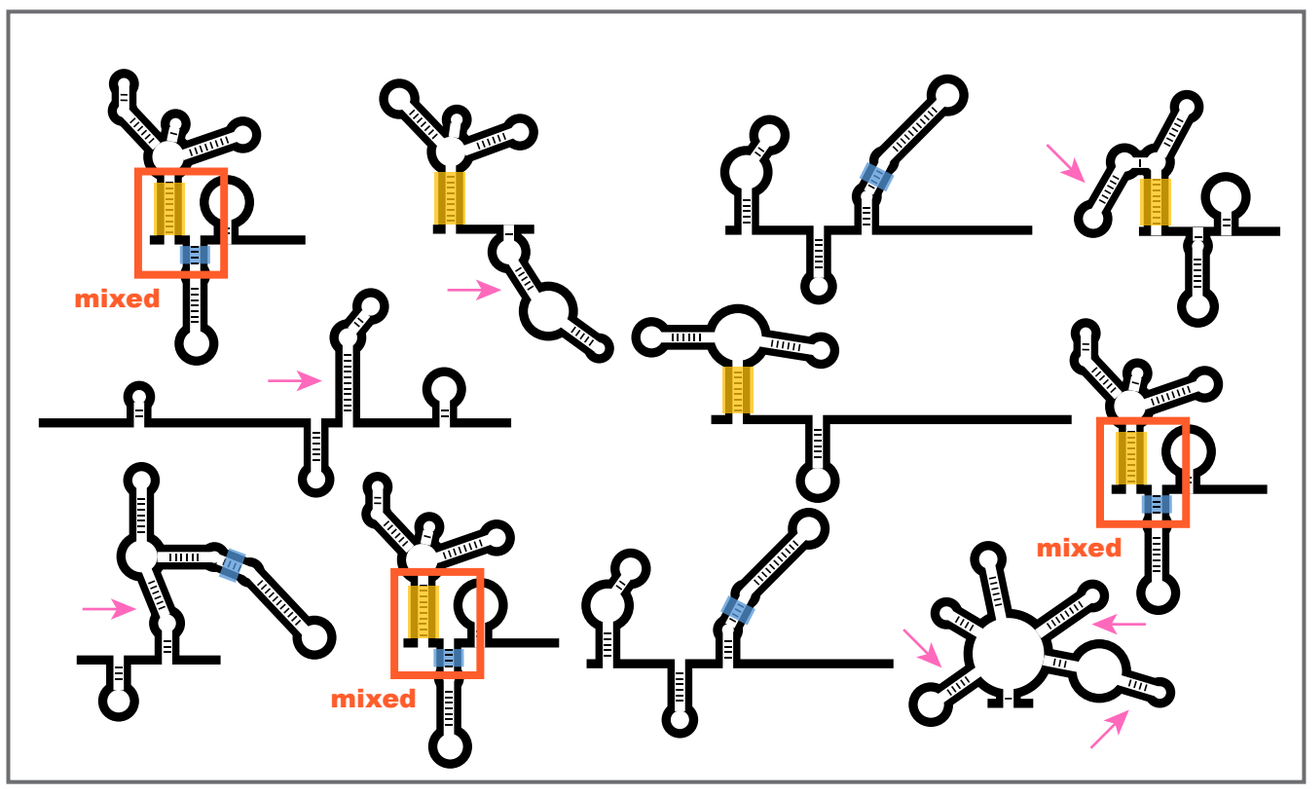
(C) **Anti-Correlation**

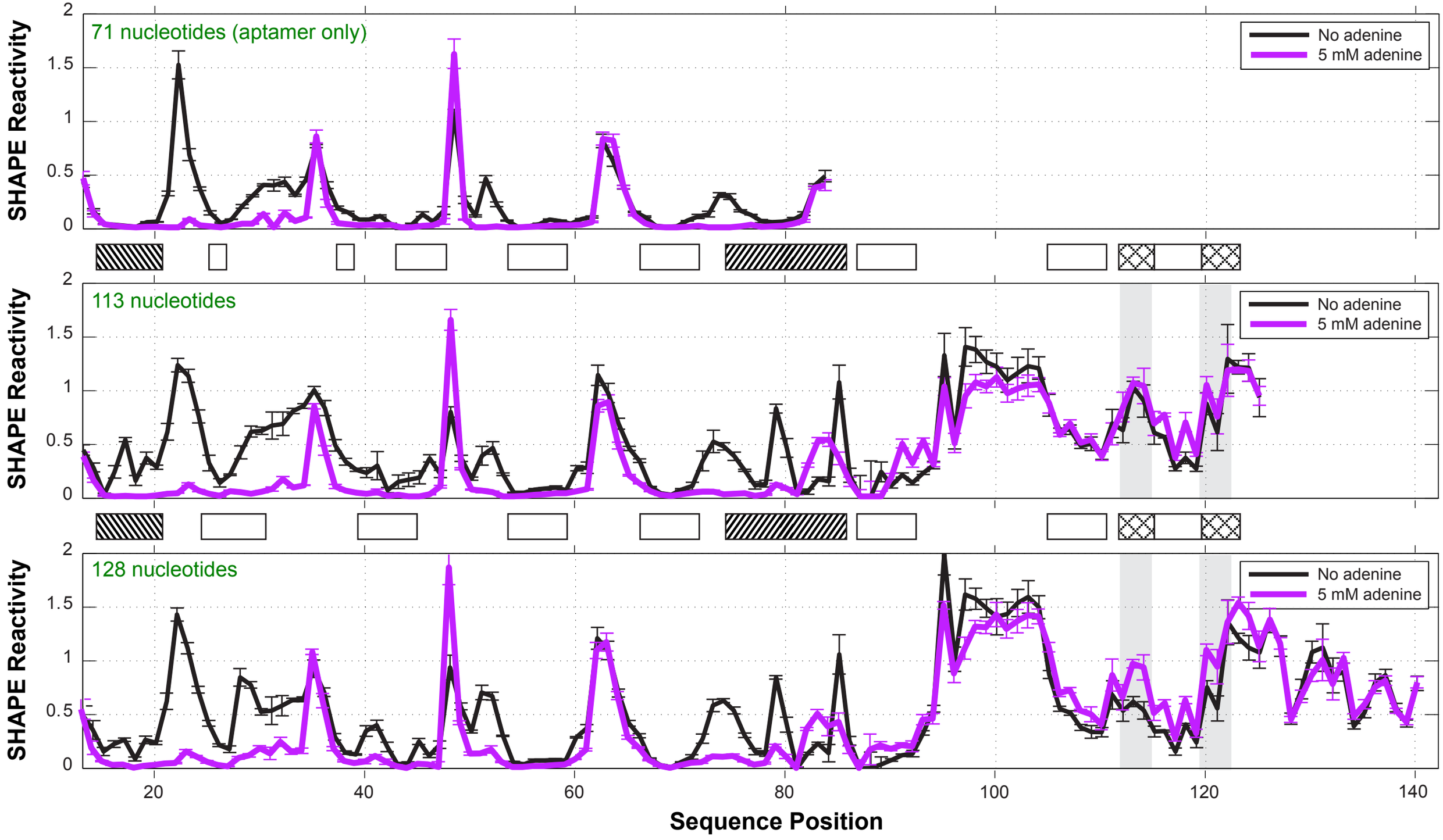
$$\frac{F(\mathbf{P1}, \mathbf{P4B})}{F(\mathbf{P1})\,F(\mathbf{P4B})} \ll 1$$

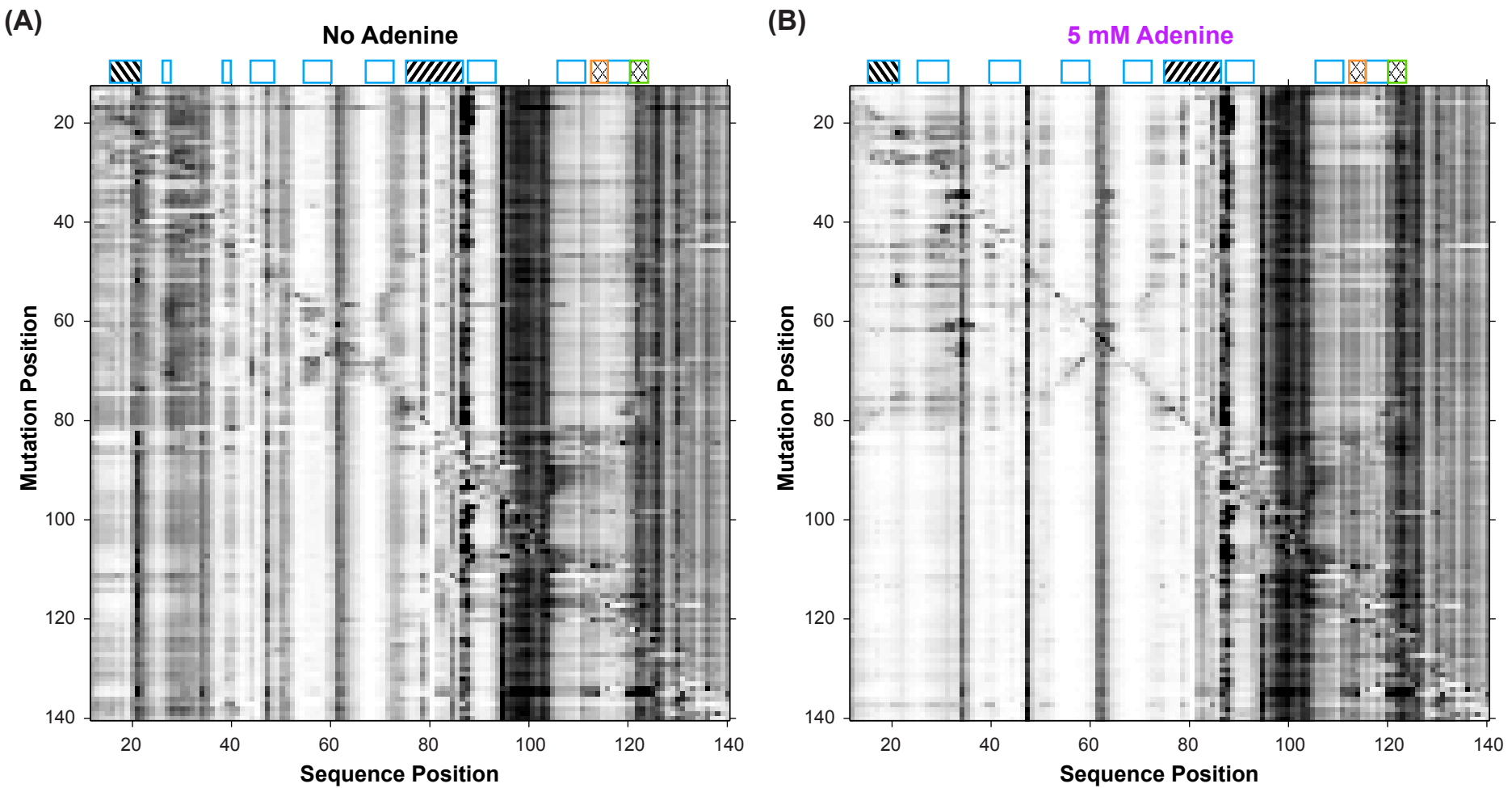


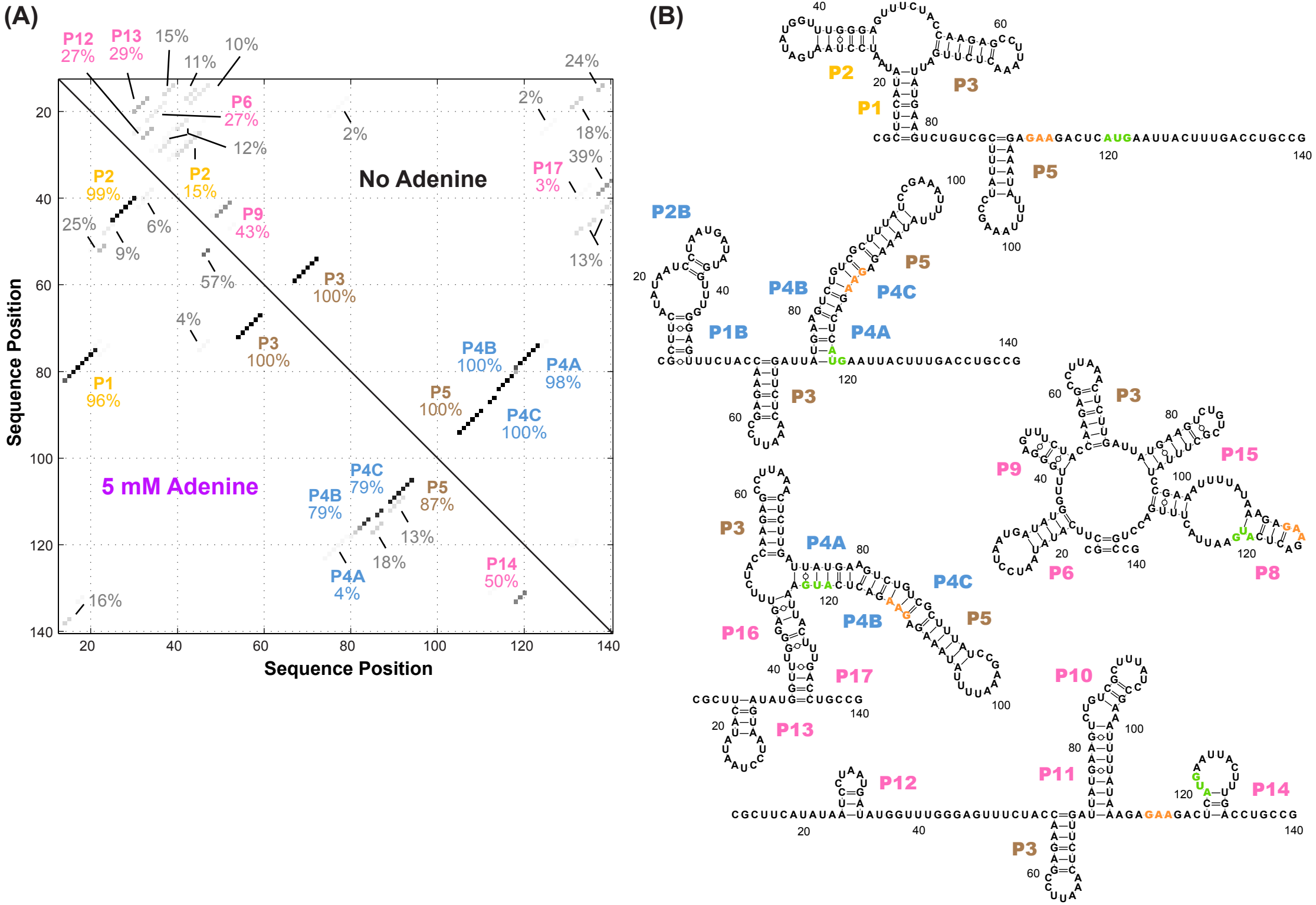
(D) **No Correlation**

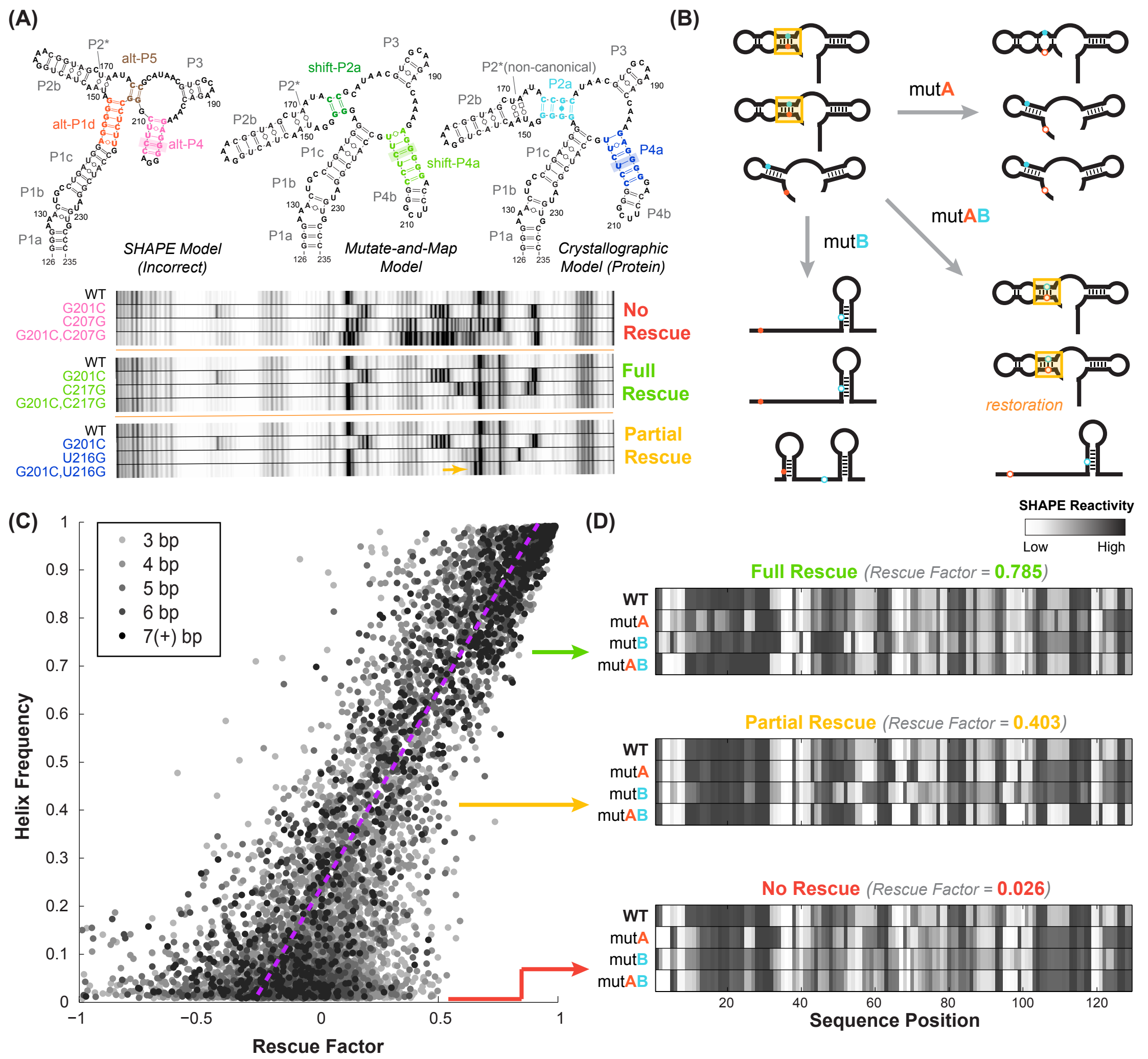
$$\frac{F(\mathbf{P1}, \mathbf{P4B})}{F(\mathbf{P1})\,F(\mathbf{P4B})} \sim 1$$

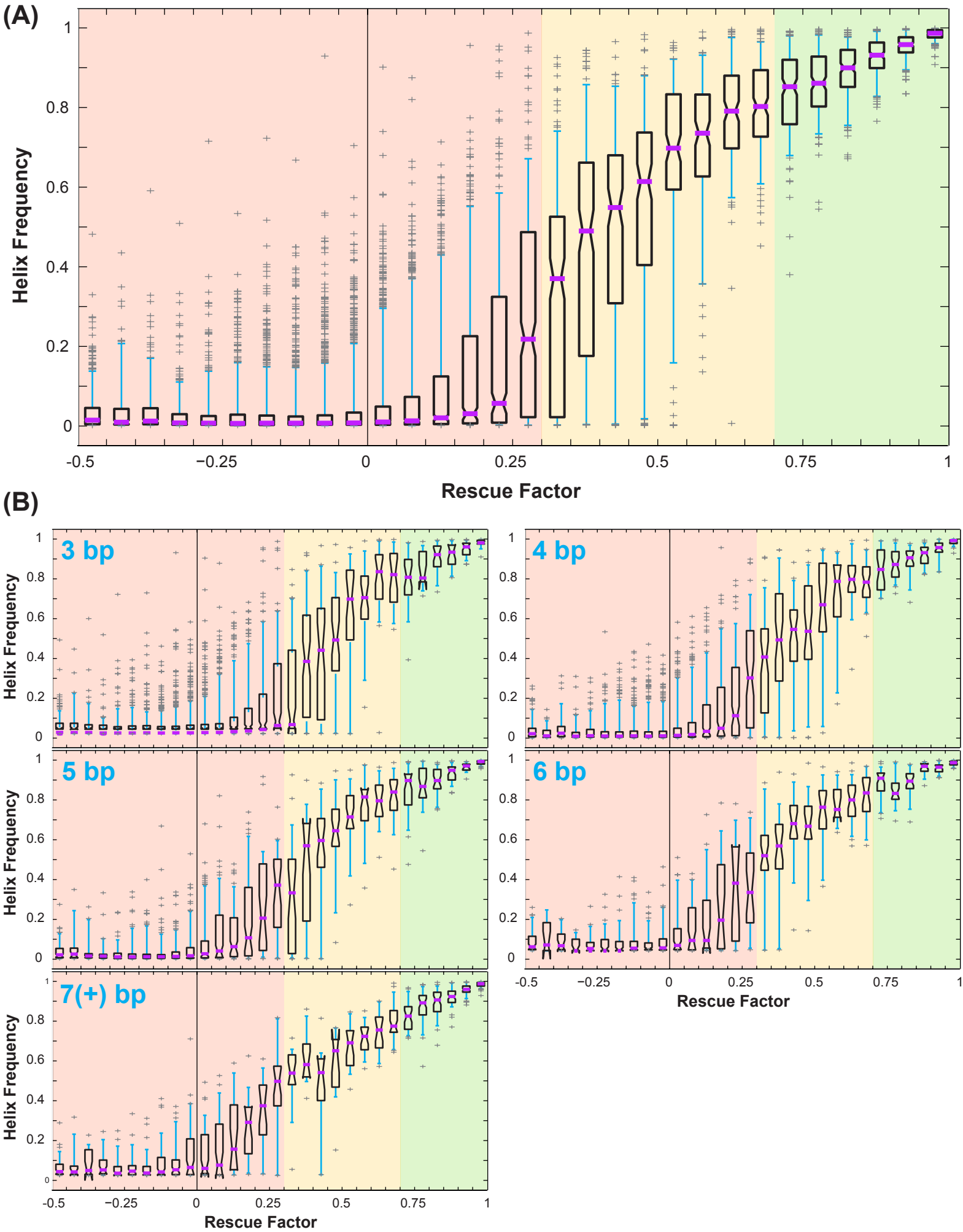


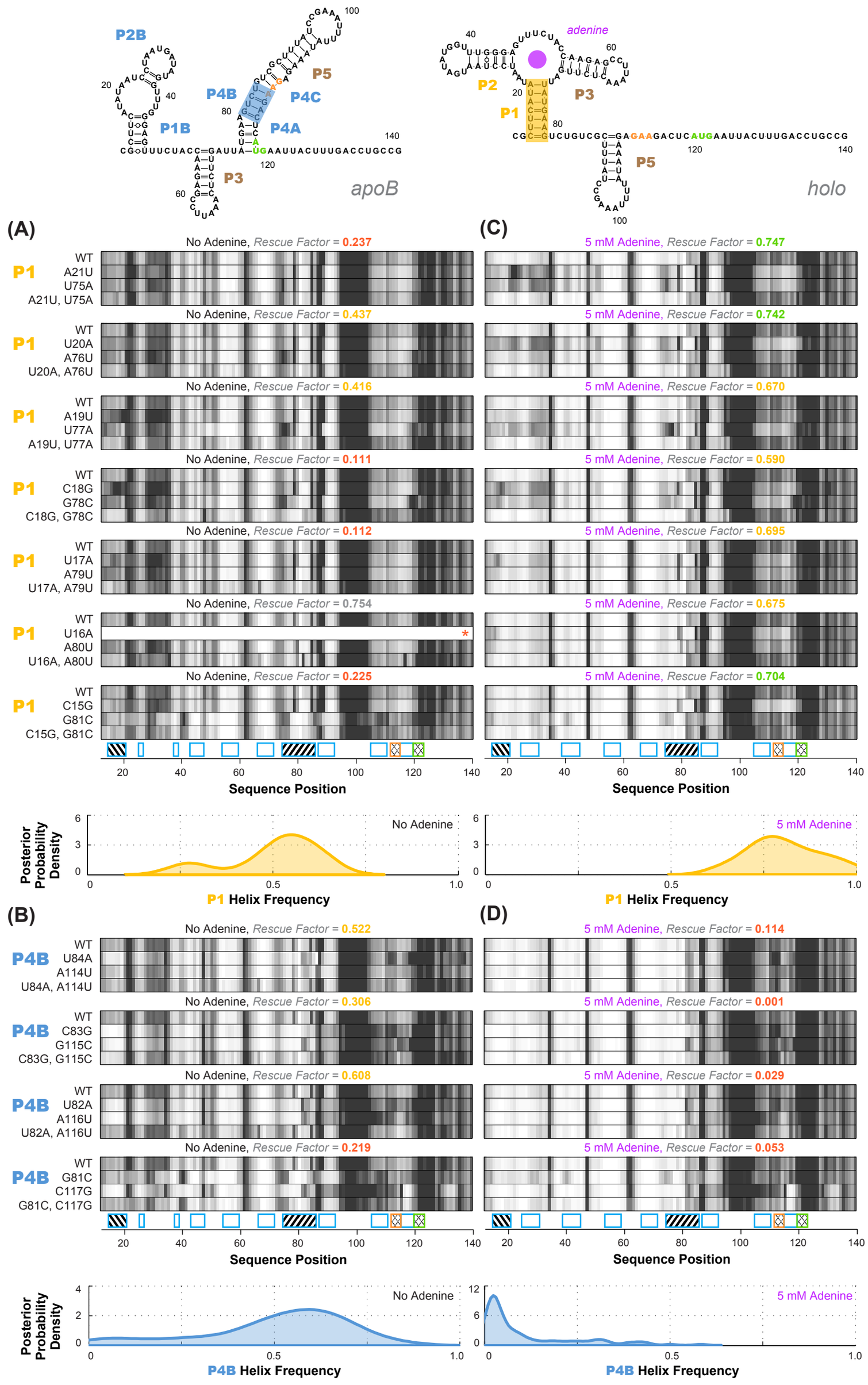

71 nucleotides (aptamer only)
113 nucleotides
128 nucleotides
No adenine
5 mM adenine
SHAPE Reactivity
Sequence Position
0
0.5
1
1.5
2
20
40
60
80
100
120
140

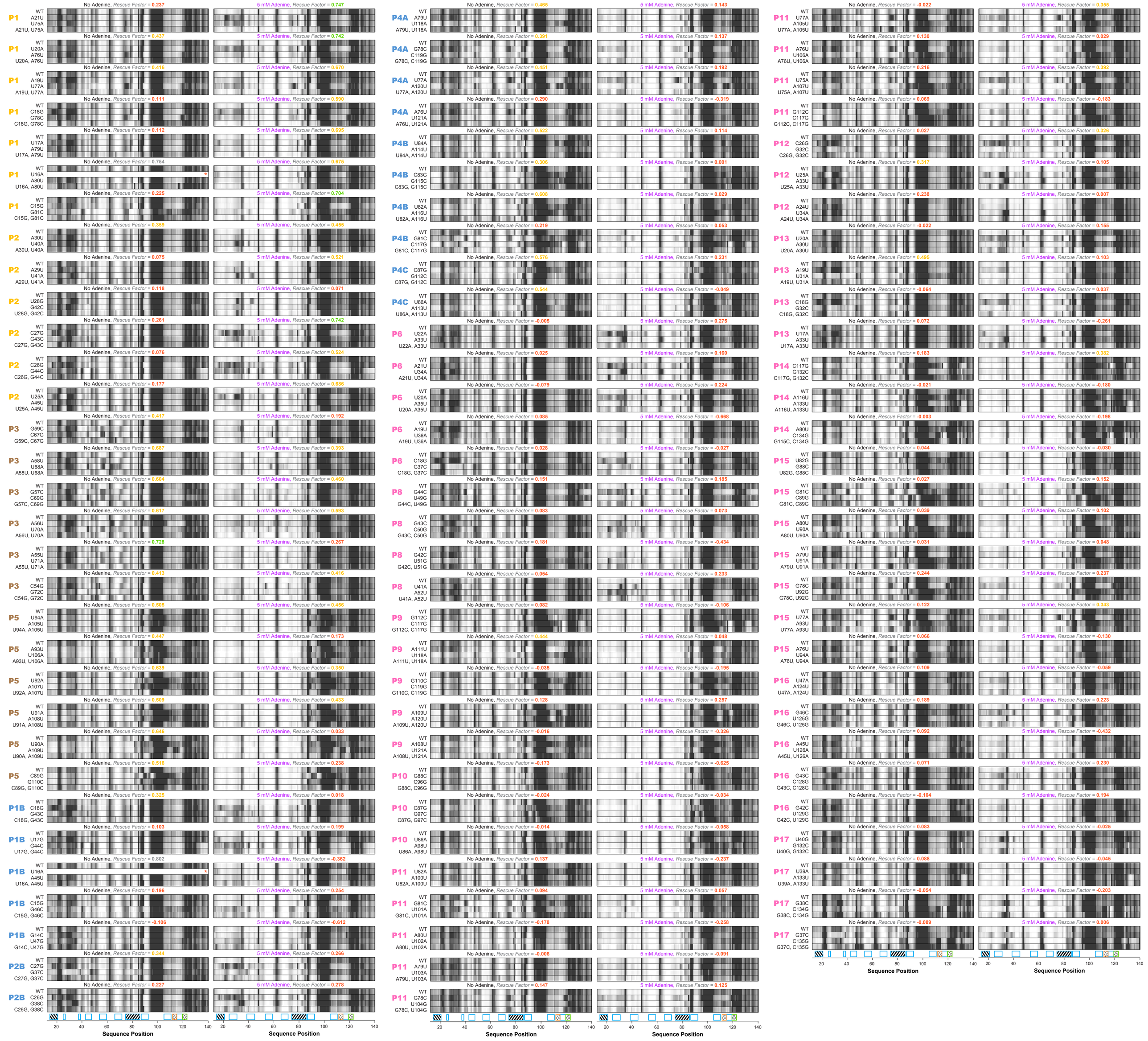

(A)
No Adenine
Mutation Position
Sequence Position
20
40
60
80
100
120
140
(B)
5 mM Adenine
Mutation Position
Sequence Position
20
40
60
80
100
120
140


(A)
No Adenine
5 mM Adenine
Sequence Position
Sequence Position
P12 27%
P13 29%
15%
11%
10%
P6 27%
12%
2%
24%
2%
18%
39%
P17 3%
13%
P2 99%
P2 15%
P9 43%
25%
6%
9%
57%
P3 100%
4%
P3 100%
P1 96%
P4B 100%
P4A 98%
P5 100%
P4C 100%
P4C 79%
P4B 79%
P5 87%
13%
18%
P4A 4%
P14 50%
16%
(B)
P2
P3
P1
P5
P2B
P4B
P4C
P4A
P1B
P3
P9
P15
P6
P8
P16
P17
P13
P10
P11
P12
P14

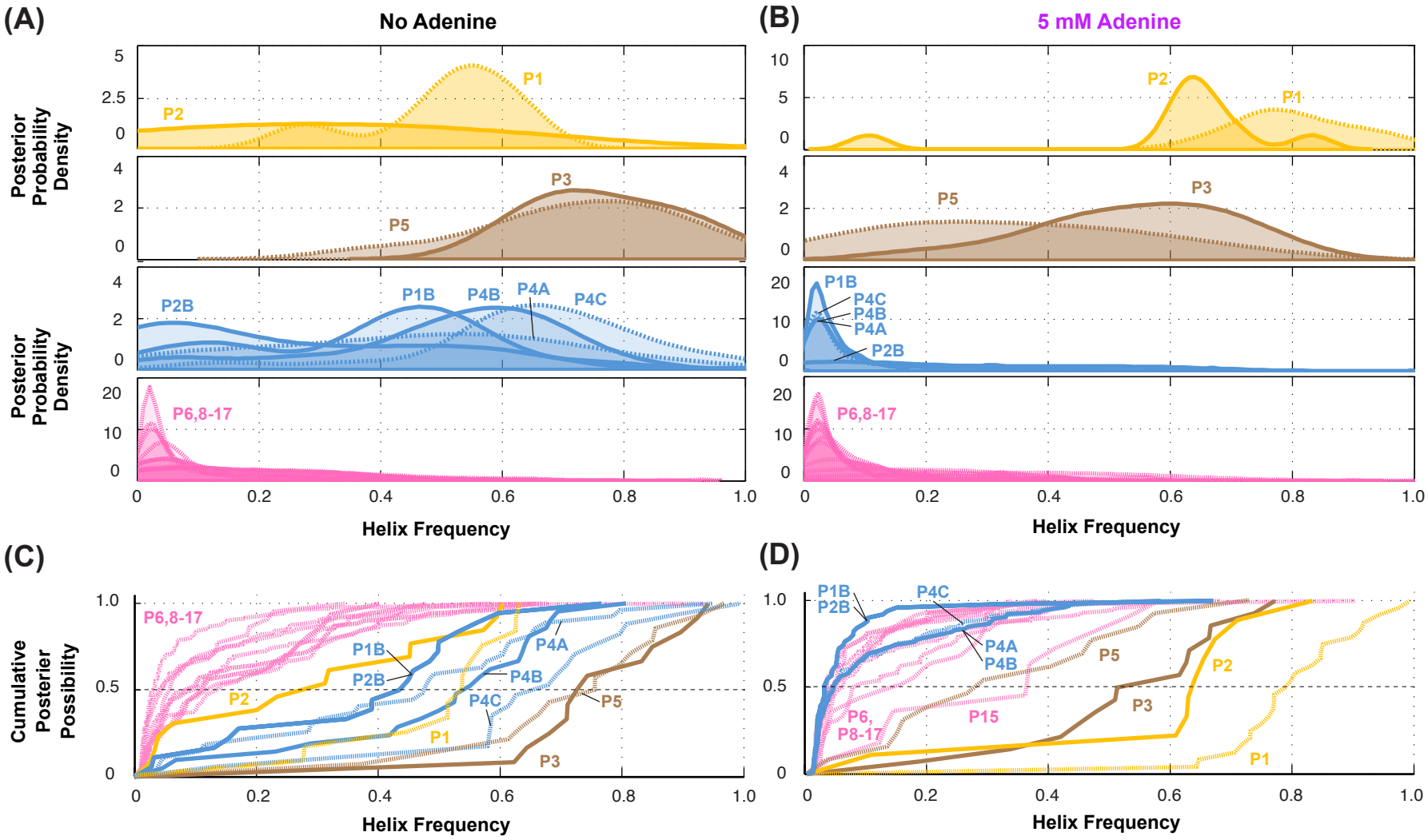

(A)
P2*
alt-P5
P3
P2b
alt-P1d
alt-P4
P1c
P1b
P1a
SHAPE Model (Incorrect)
shift-P2a
shift-P4a
P4b
Mutate-and-Map Model
P2*(non-canonical)
P2a
P4a
Crystallographic Model (Protein)
WT
G201C
C207G
G201C,C207G
No Rescue
WT
G201C
C217G
G201C,C217G
Full Rescue
WT
G201C
U216G
G201C,U216G
Partial Rescue
(B)
mutA
mutB
mutAB
restoration
(C)
3 bp
4 bp
5 bp
6 bp
7(+) bp
Helix Frequency
Rescue Factor
(D)
SHAPE Reactivity
Low
High
Full Rescue (Rescue Factor = 0.785)
Partial Rescue (Rescue Factor = 0.403)
No Rescue (Rescue Factor = 0.026)
WT
mutA
mutB
mutAB
Sequence Position

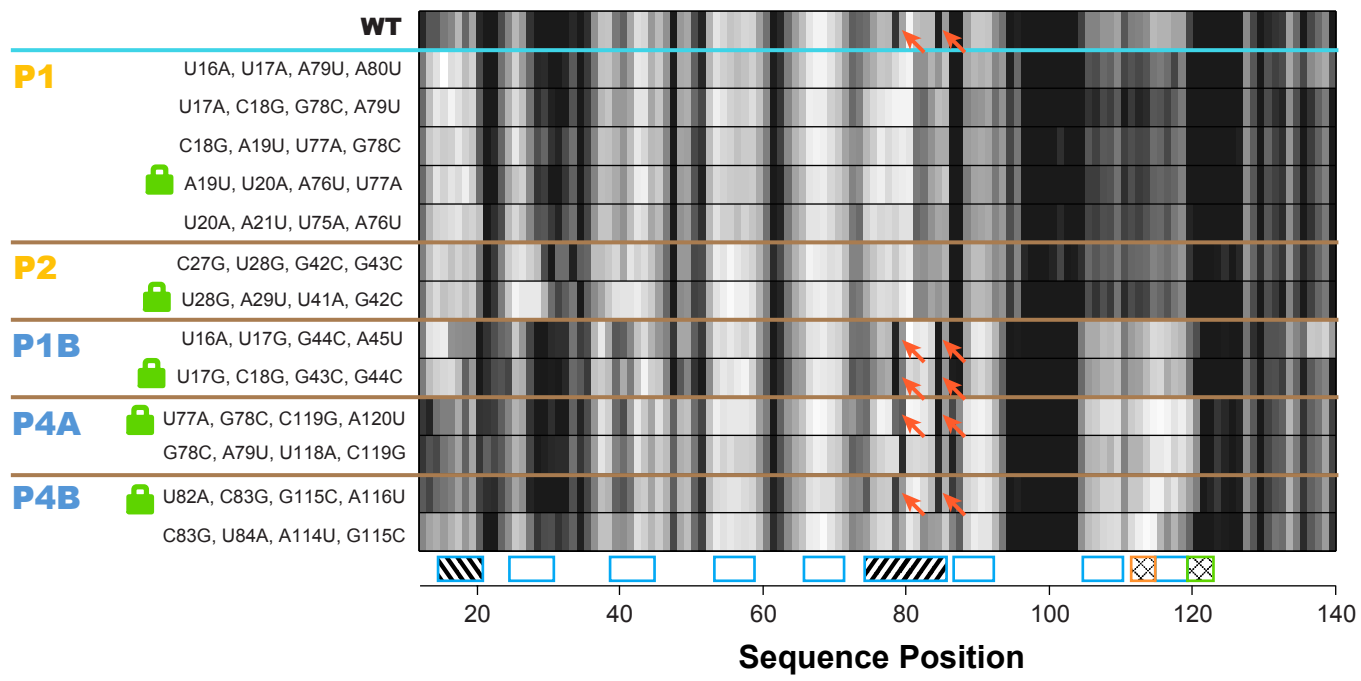

(A)
Helix Frequency
Rescue Factor
(B)
3 bp
4 bp
5 bp
6 bp
7(+) bp
Helix Frequency
Rescue Factor

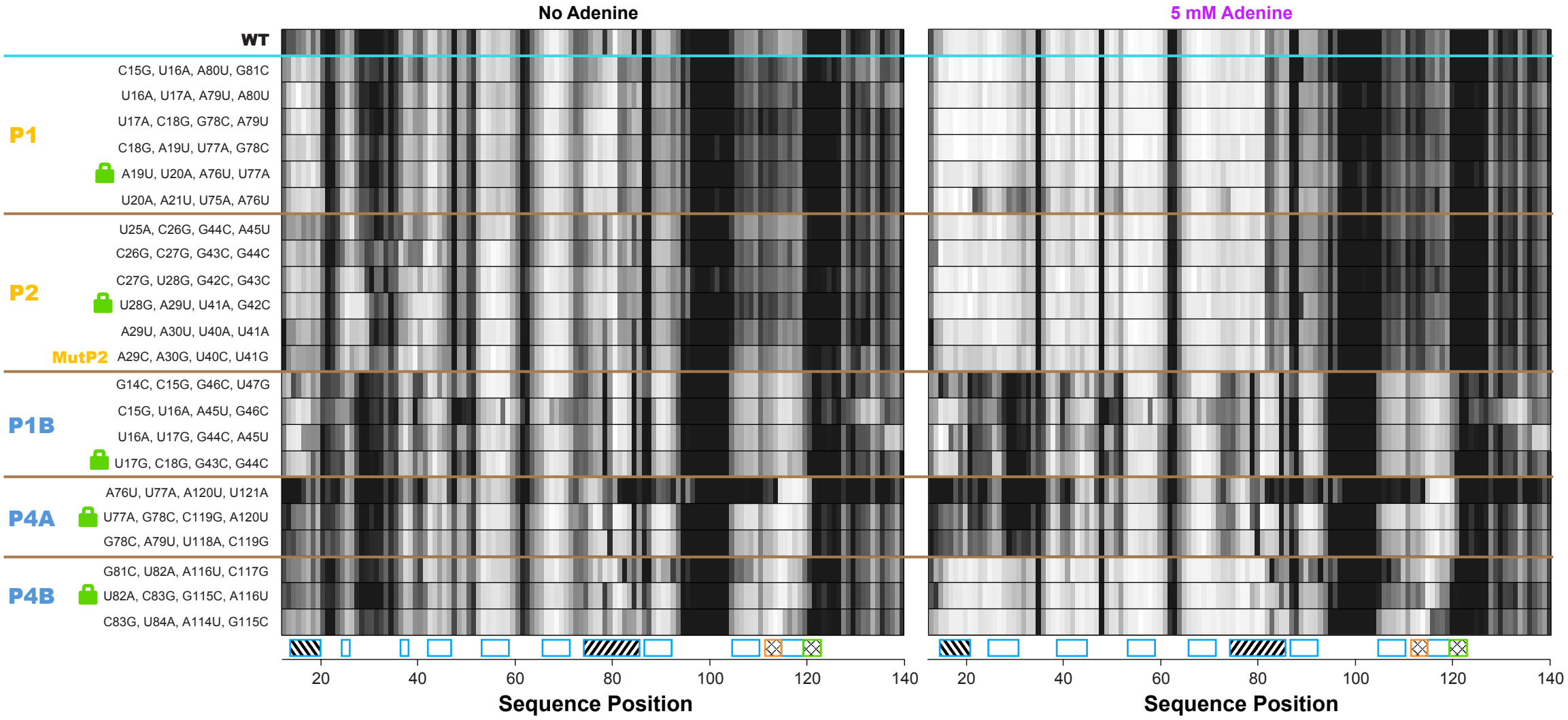

P2B
P1B
P4B
P4C
P4A
P5
P3
apoB
adenine
P2
P1
P3
P5
holo
(A)
No Adenine, Rescue Factor = 0.237
WT
P1 A21U
U75A
A21U, U75A
No Adenine, Rescue Factor = 0.437
WT
P1 U20A
A76U
U20A, A76U
No Adenine, Rescue Factor = 0.416
WT
P1 A19U
U77A
A19U, U77A
No Adenine, Rescue Factor = 0.111
WT
P1 C18G
G78C
C18G, G78C
No Adenine, Rescue Factor = 0.112
WT
P1 U17A
A79U
U17A, A79U
No Adenine, Rescue Factor = 0.754
WT
P1 U16A
A80U
U16A, A80U
No Adenine, Rescue Factor = 0.225
WT
P1 C15G
G81C
C15G, G81C
Sequence Position
Posterior Probability Density
No Adenine
P1 Helix Frequency
(C)
5 mM Adenine, Rescue Factor = 0.747
5 mM Adenine, Rescue Factor = 0.742
5 mM Adenine, Rescue Factor = 0.670
5 mM Adenine, Rescue Factor = 0.590
5 mM Adenine, Rescue Factor = 0.695
5 mM Adenine, Rescue Factor = 0.675
5 mM Adenine, Rescue Factor = 0.704
Sequence Position
5 mM Adenine
P1 Helix Frequency
(B)
No Adenine, Rescue Factor = 0.522
WT
P4B U84A
A114U
U84A, A114U
No Adenine, Rescue Factor = 0.306
WT
P4B C83G
G115C
C83G, G115C
No Adenine, Rescue Factor = 0.608
WT
P4B U82A
A116U
U82A, A116U
No Adenine, Rescue Factor = 0.219
WT
P4B G81C
C117G
G81C, C117G
Sequence Position
Posterior Probability Density
No Adenine
P4B Helix Frequency
(D)
5 mM Adenine, Rescue Factor = 0.114
5 mM Adenine, Rescue Factor = 0.001
5 mM Adenine, Rescue Factor = 0.029
5 mM Adenine, Rescue Factor = 0.053
Sequence Position
5 mM Adenine
P4B Helix Frequency

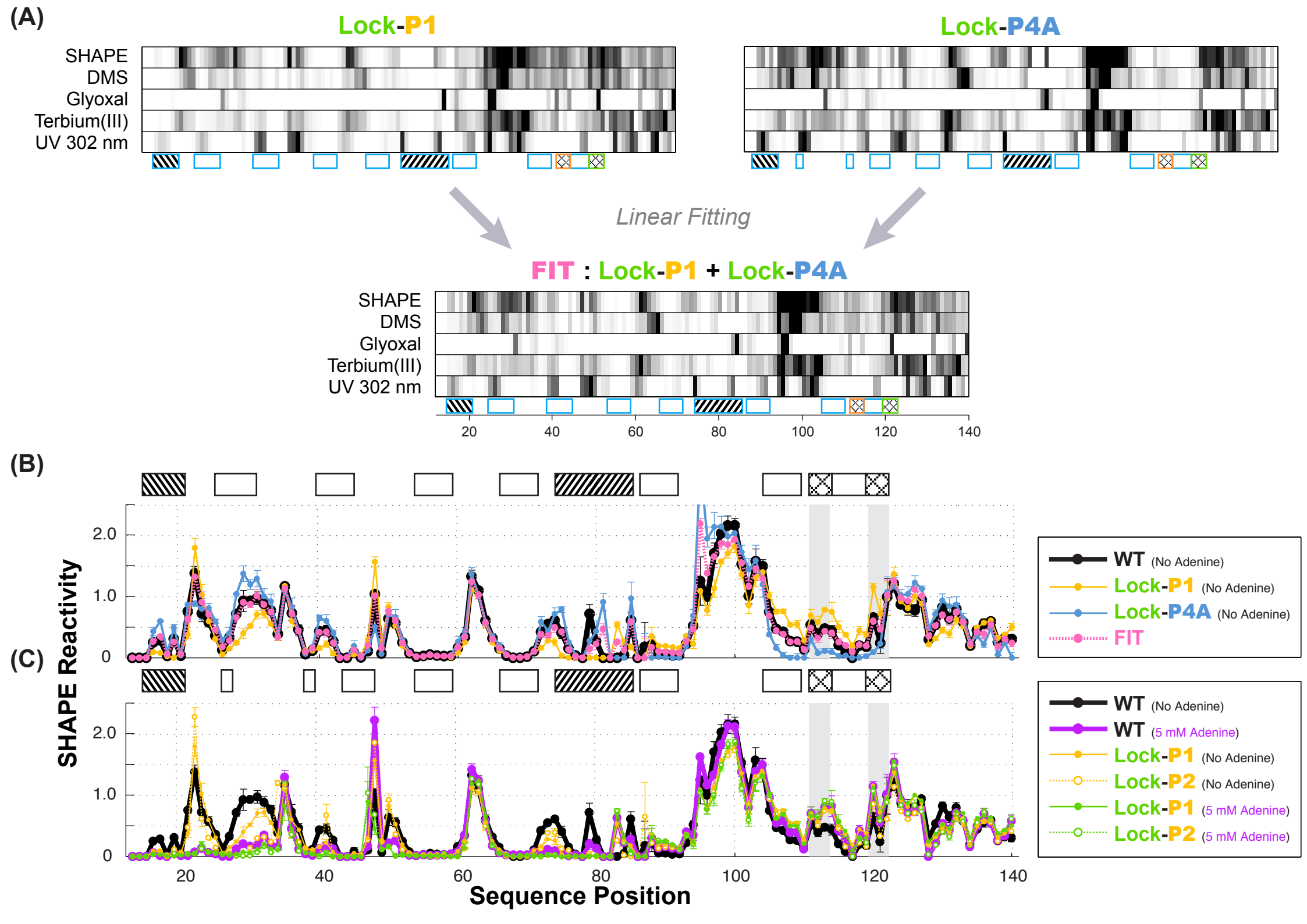

P1 WT A21U U75A A21U, U75A
No Adenine, Rescue Factor = 0.237
5 mM Adenine, Rescue Factor = 0.747
P1 WT U20A A76U U20A, A76U
No Adenine, Rescue Factor = 0.437
5 mM Adenine, Rescue Factor = 0.742
P1 WT A19U U77A A19U, U77A
No Adenine, Rescue Factor = 0.416
5 mM Adenine, Rescue Factor = 0.670
P1 WT C18G G78C C18G, G78C
No Adenine, Rescue Factor = 0.111
5 mM Adenine, Rescue Factor = 0.590
P1 WT U17A A79U U17A, A79U
No Adenine, Rescue Factor = 0.112
5 mM Adenine, Rescue Factor = 0.695
P1 WT U16A A80U U16A, A80U
No Adenine, Rescue Factor = 0.754
5 mM Adenine, Rescue Factor = 0.675
P1 WT C15G G81C C15G, G81C
No Adenine, Rescue Factor = 0.225
5 mM Adenine, Rescue Factor = 0.704
P2 WT A30U U40A A30U, U40A
No Adenine, Rescue Factor = 0.359
5 mM Adenine, Rescue Factor = 0.455
P2 WT A29U U41A A29U, U41A
No Adenine, Rescue Factor = 0.075
5 mM Adenine, Rescue Factor = 0.521
P2 WT U28G G42C U28G, G42C
No Adenine, Rescue Factor = 0.118
5 mM Adenine, Rescue Factor = 0.071
P2 WT C27G G43C C27G, G43C
No Adenine, Rescue Factor = 0.261
5 mM Adenine, Rescue Factor = 0.742
P2 WT C26G G44C C26G, G44C
No Adenine, Rescue Factor = 0.076
5 mM Adenine, Rescue Factor = 0.524
P2 WT U25A A45U U25A, A45U
No Adenine, Rescue Factor = 0.177
5 mM Adenine, Rescue Factor = 0.686
P3 WT G59C C67G G59C, C67G
No Adenine, Rescue Factor = 0.417
5 mM Adenine, Rescue Factor = 0.192
P3 WT A58U U68A A58U, U68A
No Adenine, Rescue Factor = 0.687
5 mM Adenine, Rescue Factor = 0.393
P3 WT G57C C69G G57C, C69G
No Adenine, Rescue Factor = 0.604
5 mM Adenine, Rescue Factor = 0.460
P3 WT A56U U70A A56U, U70A
No Adenine, Rescue Factor = 0.617
5 mM Adenine, Rescue Factor = 0.593
P3 WT A55U U71A A55U, U71A
No Adenine, Rescue Factor = 0.728
5 mM Adenine, Rescue Factor = 0.267
P3 WT C54G G72C C54G, G72C
No Adenine, Rescue Factor = 0.413
5 mM Adenine, Rescue Factor = 0.416
P5 WT U94A A105U U94A, A105U
No Adenine, Rescue Factor = 0.505
5 mM Adenine, Rescue Factor = 0.456
P5 WT A93U U106A A93U, U106A
No Adenine, Rescue Factor = 0.447
5 mM Adenine, Rescue Factor = 0.173
P5 WT U92A A107U U92A, A107U
No Adenine, Rescue Factor = 0.639
5 mM Adenine, Rescue Factor = 0.350
P5 WT U91A A108U U91A, A108U
No Adenine, Rescue Factor = 0.509
5 mM Adenine, Rescue Factor = 0.433
P5 WT U90A A109U U90A, A109U
No Adenine, Rescue Factor = 0.646
5 mM Adenine, Rescue Factor = 0.033
P5 WT C89G G110C C89G, G110C
No Adenine, Rescue Factor = 0.516
5 mM Adenine, Rescue Factor = 0.238
P1B WT C18G G43C C18G, G43C
No Adenine, Rescue Factor = 0.325
5 mM Adenine, Rescue Factor = 0.018
P1B WT U17G G44C U17G, G44C
No Adenine, Rescue Factor = 0.103
5 mM Adenine, Rescue Factor = 0.199
P1B WT U16A A45U U16A, A45U
No Adenine, Rescue Factor = 0.802
5 mM Adenine, Rescue Factor = -0.362
P1B WT C15G G46C C15G, G46C
No Adenine, Rescue Factor = 0.196
5 mM Adenine, Rescue Factor = 0.254
P1B WT G14C U47G G14C, U47G
No Adenine, Rescue Factor = -0.106
5 mM Adenine, Rescue Factor = -0.612
P2B WT C27G G37C C27G, G37C
No Adenine, Rescue Factor = 0.344
5 mM Adenine, Rescue Factor = 0.266
P2B WT C26G G38C C26G, G38C
No Adenine, Rescue Factor = 0.227
5 mM Adenine, Rescue Factor = 0.278
P4A WT A79U U118A A79U, U118A
No Adenine, Rescue Factor = 0.465
5 mM Adenine, Rescue Factor = 0.143
P4A WT G78C C119G G78C, C119G
No Adenine, Rescue Factor = 0.391
5 mM Adenine, Rescue Factor = 0.137
P4A WT U77A A120U U77A, A120U
No Adenine, Rescue Factor = 0.451
5 mM Adenine, Rescue Factor = 0.192
P4A WT A76U U121A A76U, U121A
No Adenine, Rescue Factor = 0.290
5 mM Adenine, Rescue Factor = -0.319
P4B WT U84A A114U U84A, A114U
No Adenine, Rescue Factor = 0.522
5 mM Adenine, Rescue Factor = 0.114
P4B WT C83G G115C C83G, G115C
No Adenine, Rescue Factor = 0.306
5 mM Adenine, Rescue Factor = 0.001
P4B WT U82A A116U U82A, A116U
No Adenine, Rescue Factor = 0.608
5 mM Adenine, Rescue Factor = 0.029
P4B WT G81C C117G G81C, C117G
No Adenine, Rescue Factor = 0.219
5 mM Adenine, Rescue Factor = 0.053
P4C WT C87G G112C C87G, G112C
No Adenine, Rescue Factor = 0.576
5 mM Adenine, Rescue Factor = 0.231
P4C WT U86A A113U U86A, A113U
No Adenine, Rescue Factor = 0.544
5 mM Adenine, Rescue Factor = -0.049
P6 WT U22A A33U U22A, A33U
No Adenine, Rescue Factor = -0.005
5 mM Adenine, Rescue Factor = 0.275
P6 WT A21U U34A A21U, U34A
No Adenine, Rescue Factor = 0.025
5 mM Adenine, Rescue Factor = 0.160
P6 WT U20A A35U U20A, A35U
No Adenine, Rescue Factor = -0.079
5 mM Adenine, Rescue Factor = 0.224
P6 WT A19U U36A A19U, U36A
No Adenine, Rescue Factor = 0.085
5 mM Adenine, Rescue Factor = -0.668
P6 WT C18G G37C C18G, G37C
No Adenine, Rescue Factor = 0.028
5 mM Adenine, Rescue Factor = -0.027
P8 WT G44C U49G G44C, U49G
No Adenine, Rescue Factor = 0.151
5 mM Adenine, Rescue Factor = 0.185
P8 WT G43C C50G G43C, C50G
No Adenine, Rescue Factor = 0.083
5 mM Adenine, Rescue Factor = 0.073
P8 WT G42C U51G G42C, U51G
No Adenine, Rescue Factor = 0.181
5 mM Adenine, Rescue Factor = -0.434
P8 WT U41A A52U U41A, A52U
No Adenine, Rescue Factor = 0.054
5 mM Adenine, Rescue Factor = 0.233
P9 WT G112C C117G G112C, C117G
No Adenine, Rescue Factor = 0.082
5 mM Adenine, Rescue Factor = -0.106
P9 WT A111U U118A A111U, U118A
No Adenine, Rescue Factor = 0.444
5 mM Adenine, Rescue Factor = 0.048
P9 WT G110C C119G G110C, C119G
No Adenine, Rescue Factor = -0.035
5 mM Adenine, Rescue Factor = -0.195
P9 WT A109U A120U A109U, A120U
No Adenine, Rescue Factor = 0.128
5 mM Adenine, Rescue Factor = 0.257
P9 WT A108U U121A A108U, U121A
No Adenine, Rescue Factor = -0.016
5 mM Adenine, Rescue Factor = -0.326
P10 WT G88C C96G G88C, C96G
No Adenine, Rescue Factor = -0.173
5 mM Adenine, Rescue Factor = -0.625
P10 WT C87G G97C C87G, G97C
No Adenine, Rescue Factor = -0.024
5 mM Adenine, Rescue Factor = -0.034
P10 WT U86A A98U U86A, A98U
No Adenine, Rescue Factor = -0.014
5 mM Adenine, Rescue Factor = -0.058
P11 WT U82A A100U U82A, A100U
No Adenine, Rescue Factor = 0.137
5 mM Adenine, Rescue Factor = -0.237
P11 WT G81C U101A G81C, U101A
No Adenine, Rescue Factor = 0.094
5 mM Adenine, Rescue Factor = 0.057
P11 WT A80U U102A A80U, U102A
No Adenine, Rescue Factor = -0.178
5 mM Adenine, Rescue Factor = -0.258
P11 WT A79U U103A A79U, U103A
No Adenine, Rescue Factor = -0.006
5 mM Adenine, Rescue Factor = -0.091
P11 WT G78C U104G G78C, U104G
No Adenine, Rescue Factor = 0.147
5 mM Adenine, Rescue Factor = 0.125
P11 WT U77A A105U U77A, A105U
No Adenine, Rescue Factor = -0.022
5 mM Adenine, Rescue Factor = 0.355
P11 WT A76U U106A A76U, U106A
No Adenine, Rescue Factor = 0.130
5 mM Adenine, Rescue Factor = 0.029
P11 WT U75A A107U U75A, A107U
No Adenine, Rescue Factor = 0.216
5 mM Adenine, Rescue Factor = 0.392
P11 WT G112C C117G G112C, C117G
No Adenine, Rescue Factor = 0.069
5 mM Adenine, Rescue Factor = -0.183
P12 WT C26G G32C C26G, G32C
No Adenine, Rescue Factor = 0.027
5 mM Adenine, Rescue Factor = 0.326
P12 WT U25A A33U U25A, A33U
No Adenine, Rescue Factor = 0.317
5 mM Adenine, Rescue Factor = 0.105
P12 WT A24U U34A A24U, U34A
No Adenine, Rescue Factor = 0.238
5 mM Adenine, Rescue Factor = 0.007
P13 WT U20A A30U U20A, A30U
No Adenine, Rescue Factor = -0.022
5 mM Adenine, Rescue Factor = 0.155
P13 WT A19U U31A A19U, U31A
No Adenine, Rescue Factor = 0.495
5 mM Adenine, Rescue Factor = 0.103
P13 WT C18G G32C C18G, G32C
No Adenine, Rescue Factor = -0.064
5 mM Adenine, Rescue Factor = 0.037
P13 WT U17A A33U U17A, A33U
No Adenine, Rescue Factor = 0.072
5 mM Adenine, Rescue Factor = -0.261
P14 WT C117G G132C C117G, G132C
No Adenine, Rescue Factor = 0.183
5 mM Adenine, Rescue Factor = 0.382
P14 WT A116U A133U A116U, A133U
No Adenine, Rescue Factor = -0.021
5 mM Adenine, Rescue Factor = -0.180
P14 WT A80U C134G G115C, C134G
No Adenine, Rescue Factor = -0.003
5 mM Adenine, Rescue Factor = -0.198
P15 WT U82G G88C U82G, G88C
No Adenine, Rescue Factor = 0.044
5 mM Adenine, Rescue Factor = -0.030
P15 WT G81C C89G G81C, C89G
No Adenine, Rescue Factor = 0.027
5 mM Adenine, Rescue Factor = 0.152
P15 WT A80U U90A A80U, U90A
No Adenine, Rescue Factor = 0.039
5 mM Adenine, Rescue Factor = 0.102
P15 WT A79U U91A A79U, U91A
No Adenine, Rescue Factor = 0.031
5 mM Adenine, Rescue Factor = 0.048
P15 WT G78C U92G G78C, U92G
No Adenine, Rescue Factor = 0.244
5 mM Adenine, Rescue Factor = 0.237
P15 WT U77A A93U U77A, A93U
No Adenine, Rescue Factor = 0.122
5 mM Adenine, Rescue Factor = 0.343
P15 WT A76U U94A A76U, U94A
No Adenine, Rescue Factor = 0.066
5 mM Adenine, Rescue Factor = -0.130
P16 WT U47A A124U U47A, A124U
No Adenine, Rescue Factor = 0.109
5 mM Adenine, Rescue Factor = -0.059
P16 WT G46C U125G G46C, U125G
No Adenine, Rescue Factor = 0.189
5 mM Adenine, Rescue Factor = 0.223
P16 WT A45U U126A A45U, U126A
No Adenine, Rescue Factor = 0.092
5 mM Adenine, Rescue Factor = -0.432
P16 WT G43C C128G G43C, C128G
No Adenine, Rescue Factor = 0.071
5 mM Adenine, Rescue Factor = 0.230
P16 WT G42C U129G G42C, U129G
No Adenine, Rescue Factor = -0.104
5 mM Adenine, Rescue Factor = 0.194
P17 WT U40G G132C U40G, G132C
No Adenine, Rescue Factor = 0.083
5 mM Adenine, Rescue Factor = -0.025
P17 WT U39A A133U U39A, A133U
No Adenine, Rescue Factor = 0.088
5 mM Adenine, Rescue Factor = -0.045
P17 WT G38C C134G G38C, C134G
No Adenine, Rescue Factor = -0.054
5 mM Adenine, Rescue Factor = -0.203
P17 WT G37C C135G G37C, C135G
No Adenine, Rescue Factor = -0.089
5 mM Adenine, Rescue Factor = 0.006
20 40 60 80 100 120 140
Sequence Position

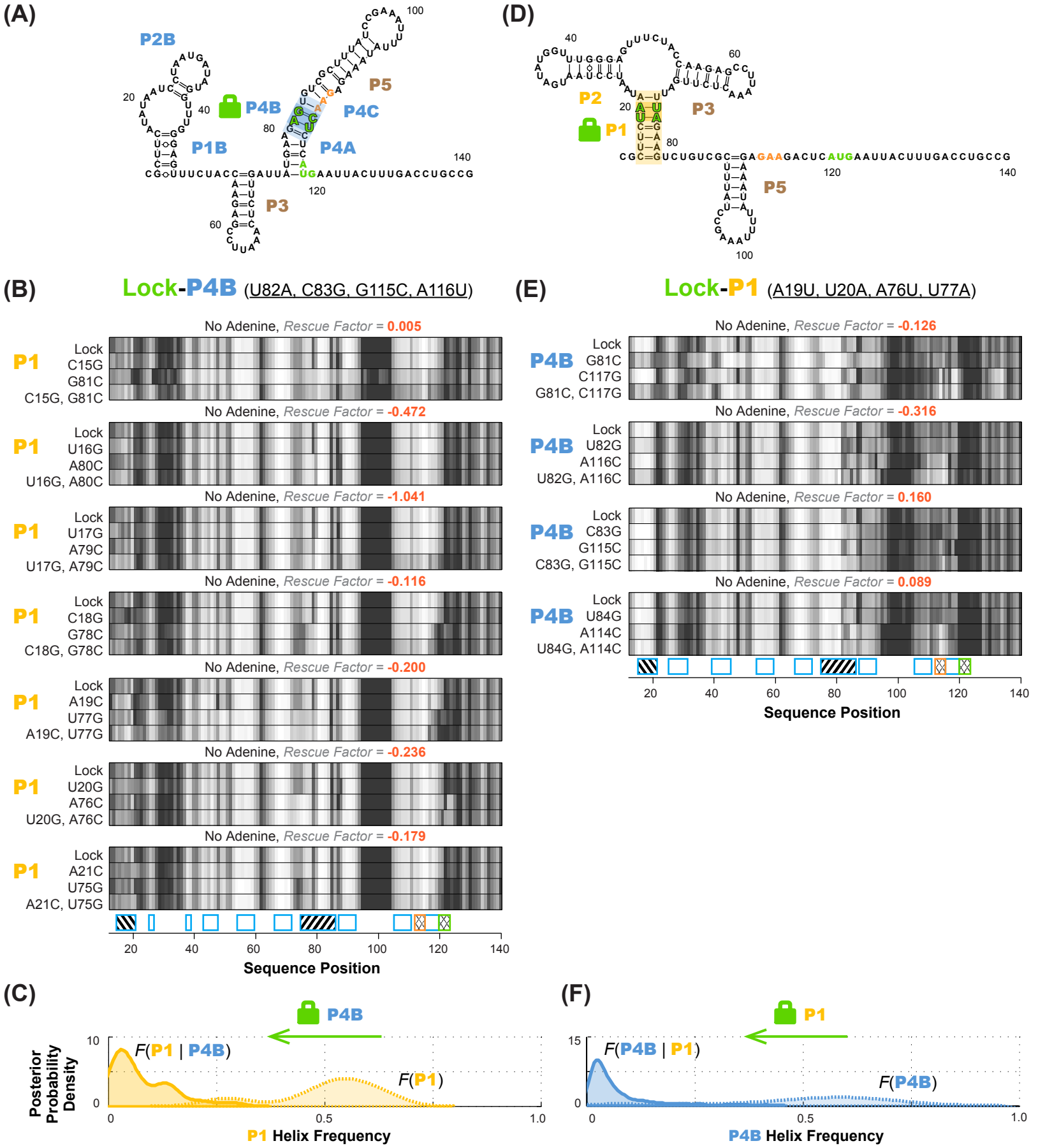
P1
WT
C15G, U16A
A80U, G81C
C15G, U16A, A80U, G81C
No Adenine, Rescue Factor = -0.086
5 mM Adenine, Rescue Factor = 0.712
P1
WT
U16A, U17A
A79U, A80U
U16A, U17A, A79U, A80U
No Adenine, Rescue Factor = -0.027
5 mM Adenine, Rescue Factor = 0.740
P1
WT
U17A, C18G
G78C, A79U
U17A, C18G, G78C, A79U
No Adenine, Rescue Factor = -0.476
5 mM Adenine, Rescue Factor = 0.749
P1
WT
C18G, A19U
U77A, G78C
C18G, A19U, U77A, G78C
No Adenine, Rescue Factor = -0.289
5 mM Adenine, Rescue Factor = 0.783
P1
WT
A19U, U20A
A76U, U77A
A19U, U20A, A76U, U77A
No Adenine, Rescue Factor = -0.508
5 mM Adenine, Rescue Factor = 0.710
P1
WT
U20A, A21U
U75A, A76U
U20A, A21U, U75A, A76U
No Adenine, Rescue Factor = -0.540
5 mM Adenine, Rescue Factor = 0.599
P2
WT
U25A, C26G
G44C, A45U
U25A, C26G, G44C, A45U
No Adenine, Rescue Factor = 0.095
5 mM Adenine, Rescue Factor = 0.827
P2
WT
C26G, C27G
G43C, G44C
C26G, C27G, G43C, G44C
No Adenine, Rescue Factor = -0.266
5 mM Adenine, Rescue Factor = 0.772
P2
WT
C27G, U28G
G42C, G43C
C27G, U28G, G42C, G43C
No Adenine, Rescue Factor = -0.750
5 mM Adenine, Rescue Factor = 0.737
P2
WT
U28G, A29U
U41A, G42C
U28G, A29U, U41A, G42C
No Adenine, Rescue Factor = -0.631
5 mM Adenine, Rescue Factor = 0.714
P2
WT
A29U, A30U
U40A, U41A
A29U, A30U, U40A, U41A
No Adenine, Rescue Factor = 0.062
5 mM Adenine, Rescue Factor = 0.869
P3
WT
C54G, A55U
U71A, G72C
C54G, A55U, U71A, G72C
No Adenine, Rescue Factor = 0.498
5 mM Adenine, Rescue Factor = 0.761
P3
WT
A55U, A56U
U70A, U71A
A55U, A56U, U70A, U71A
No Adenine, Rescue Factor = 0.118
5 mM Adenine, Rescue Factor = 0.739
P3
WT
A56U, G57C
C69G, U70A
A56U, G57C, C69G, U70A
No Adenine, Rescue Factor = 0.694
5 mM Adenine, Rescue Factor = 0.799
P3
WT
G57C, A58U
U68A, C69G
G57C, A58U, U68A, C69G
No Adenine, Rescue Factor = 0.569
5 mM Adenine, Rescue Factor = 0.840
20 40 60 80 100 120 140
Sequence Position
P3
WT
A58U, G59C
C67G, U68A
A58U, G59C, C67G, U68A
No Adenine, Rescue Factor = 0.544
5 mM Adenine, Rescue Factor = 0.765
P5
WT
C89G, U90A
A109U, G110C
C89G, U90A, A109U, G110C
No Adenine, Rescue Factor = 0.595
5 mM Adenine, Rescue Factor = 0.258
P5
WT
U90A, U91A
A108U, A109U
U90A, U91A, A108U, A109U
No Adenine, Rescue Factor = 0.481
5 mM Adenine, Rescue Factor = 0.152
P5
WT
U91A, U92A
A107U, A108U
U91A, U92A, A107U, A108U
No Adenine, Rescue Factor = 0.379
5 mM Adenine, Rescue Factor = -0.138
P5
WT
U92A, A93U
U106A, A107U
U92A, A93U, U106A, A107U
No Adenine, Rescue Factor = 0.444
5 mM Adenine, Rescue Factor = -0.109
P5
WT
A93U, U94A
A105U, U106A
A93U, U94A, A105U, U106A
No Adenine, Rescue Factor = 0.212
5 mM Adenine, Rescue Factor = -0.031
P1B
WT
G14C, C15G
G46C, U47G
G14C, C15G, G46C, U47G
No Adenine, Rescue Factor = -0.094
5 mM Adenine, Rescue Factor = -0.090
P1B
WT
C15G, U16A
A45U, G46C
C15G, U16A, A45U, G46C
No Adenine, Rescue Factor = -0.472
5 mM Adenine, Rescue Factor = -0.094
P1B
WT
U16A, U17G
G44C, A45U
U16A, U17G, G44C, A45U
No Adenine, Rescue Factor = 0.026
5 mM Adenine, Rescue Factor = -0.050
P1B
WT
U17G, C18G
G43C, G44C
U17G, C18G, G43C, G44C
No Adenine, Rescue Factor = -0.001
5 mM Adenine, Rescue Factor = -0.011
P2B
WT
C26G, C27G
G37C, G38C
C26G, C27G, G37C, G38C
No Adenine, Rescue Factor = 0.052
5 mM Adenine, Rescue Factor = 0.154
P4A
WT
A76U, U77A
A120U, U121A
A76U, U77A, A120U, U121A
No Adenine, Rescue Factor = 0.041
5 mM Adenine, Rescue Factor = -0.308
P4A
WT
U77A, G78C
C119G, A120U
U77A, G78C, C119G, A120U
No Adenine, Rescue Factor = -0.159
5 mM Adenine, Rescue Factor = -0.097
P4A
WT
G78C, A79U
U118A, C119G
G78C, A79U, U118A, C119G
No Adenine, Rescue Factor = 0.441
5 mM Adenine, Rescue Factor = -0.102
P4B
WT
G81C, U82A
A116U, C117G
G81C, U82A, A116U, C117G
No Adenine, Rescue Factor = 0.278
5 mM Adenine, Rescue Factor = 0.036
20 40 60 80 100 120 140
Sequence Position
P4B
WT
U82A, C83G
G115C, A116U
U82A, C83G, G115C, A116U
No Adenine, Rescue Factor = 0.440
5 mM Adenine, Rescue Factor = -0.019
P4B
WT
C83G, U84A
A114U, G115C
C83G, U84A, A114U, G115C
No Adenine, Rescue Factor = 0.204
5 mM Adenine, Rescue Factor = -0.693
P4C
WT
U86A, C87G
G112C, A113U
U86A, C87G, G112C, A113U
No Adenine, Rescue Factor = 0.179
5 mM Adenine, Rescue Factor = -0.200
P6
WT
C18G, A19U
U36A, G37C
C18G, A19U, U36A, G37C
No Adenine, Rescue Factor = 0.120
5 mM Adenine, Rescue Factor = -0.085
P6
WT
A19U, U20A
A35U, U36A
A19U, U20A, A35U, U36A
No Adenine, Rescue Factor = -0.084
5 mM Adenine, Rescue Factor = 0.045
P6
WT
U20A, A21U
U34A, A35U
U20A, A21U, U34A, A35U
No Adenine, Rescue Factor = -0.152
5 mM Adenine, Rescue Factor = -0.012
P6
WT
A21U, U22A
A33U, U34A
A21U, U22A, A33U, U34A
No Adenine, Rescue Factor = -0.143
5 mM Adenine, Rescue Factor = 0.049
P13
WT
U17A, C18G
G32C, A33U
U17A, C18G, G32C, A33U
No Adenine, Rescue Factor = 0.001
5 mM Adenine, Rescue Factor = 0.034
P13
WT
C18G, A19U
U31A, G32C
C18G, A19U, U31A, G32C
No Adenine, Rescue Factor = -0.053
5 mM Adenine, Rescue Factor = 0.048
P13
WT
A19U, U20A
A30U, U31A
A19U, U20A, A30U, U31A
No Adenine, Rescue Factor = -0.013
5 mM Adenine, Rescue Factor = 0.117
20 40 60 80 100 120 140
Sequence Position
MutP2
WT
A29C, A30G
U40C, U41G
A29C, A30G, U40C, U41G
No Adenine, Rescue Factor = 0.180
5 mM Adenine, Rescue Factor = 0.787
20 40 60 80 100 120 140
Sequence Position


(A)
No Adenine
(B)
5 mM Adenine
Posterior Probability Density
P1
P2
P3
P5
P1B
P2B
P4A
P4B
P4C
P6,8-17
Helix Frequency
(C)
(D)
Cumulative Posterior Possibility
P15
P6, P8-17


WT
P1
U16A, U17A, A79U, A80U
U17A, C18G, G78C, A79U
C18G, A19U, U77A, G78C
A19U, U20A, A76U, U77A
U20A, A21U, U75A, A76U
P2
C27G, U28G, G42C, G43C
U28G, A29U, U41A, G42C
P1B
U16A, U17G, G44C, A45U
U17G, C18G, G43C, G44C
P4A
U77A, G78C, C119G, A120U
G78C, A79U, U118A, C119G
P4B
U82A, C83G, G115C, A116U
C83G, U84A, A114U, G115C
20
40
60
80
100
120
140
Sequence Position

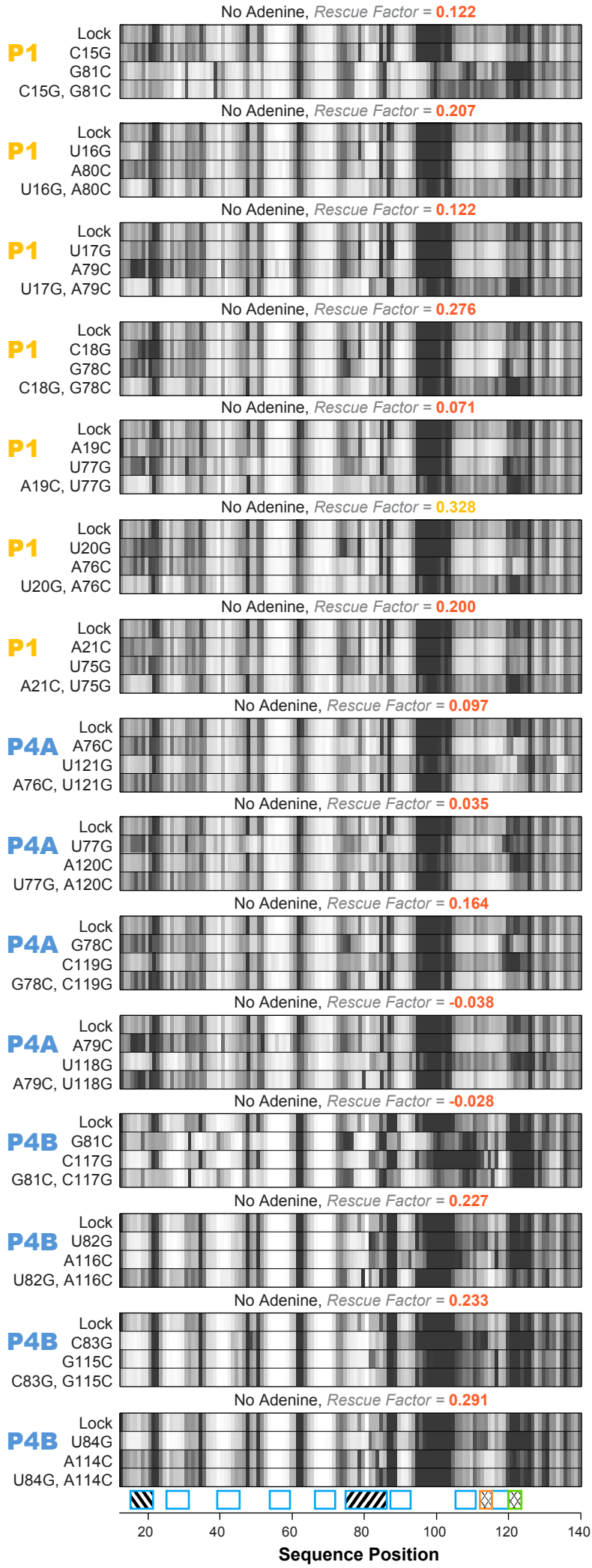

No Adenine
5 mM Adenine
WT
P1
C15G, U16A, A80U, G81C
U16A, U17A, A79U, A80U
U17A, C18G, G78C, A79U
C18G, A19U, U77A, G78C
A19U, U20A, A76U, U77A
U20A, A21U, U75A, A76U
P2
U25A, C26G, G44C, A45U
C26G, C27G, G43C, G44C
C27G, U28G, G42C, G43C
U28G, A29U, U41A, G42C
A29U, A30U, U40A, U41A
MutP2
A29C, A30G, U40C, U41G
P1B
G14C, C15G, G46C, U47G
C15G, U16A, A45U, G46C
U16A, U17G, G44C, A45U
U17G, C18G, G43C, G44C
P4A
A76U, U77A, A120U, U121A
U77A, G78C, C119G, A120U
G78C, A79U, U118A, C119G
P4B
G81C, U82A, A116U, C117G
U82A, C83G, G115C, A116U
C83G, U84A, A114U, G115C
20
40
60
80
100
120
140
Sequence Position
Sequence Position

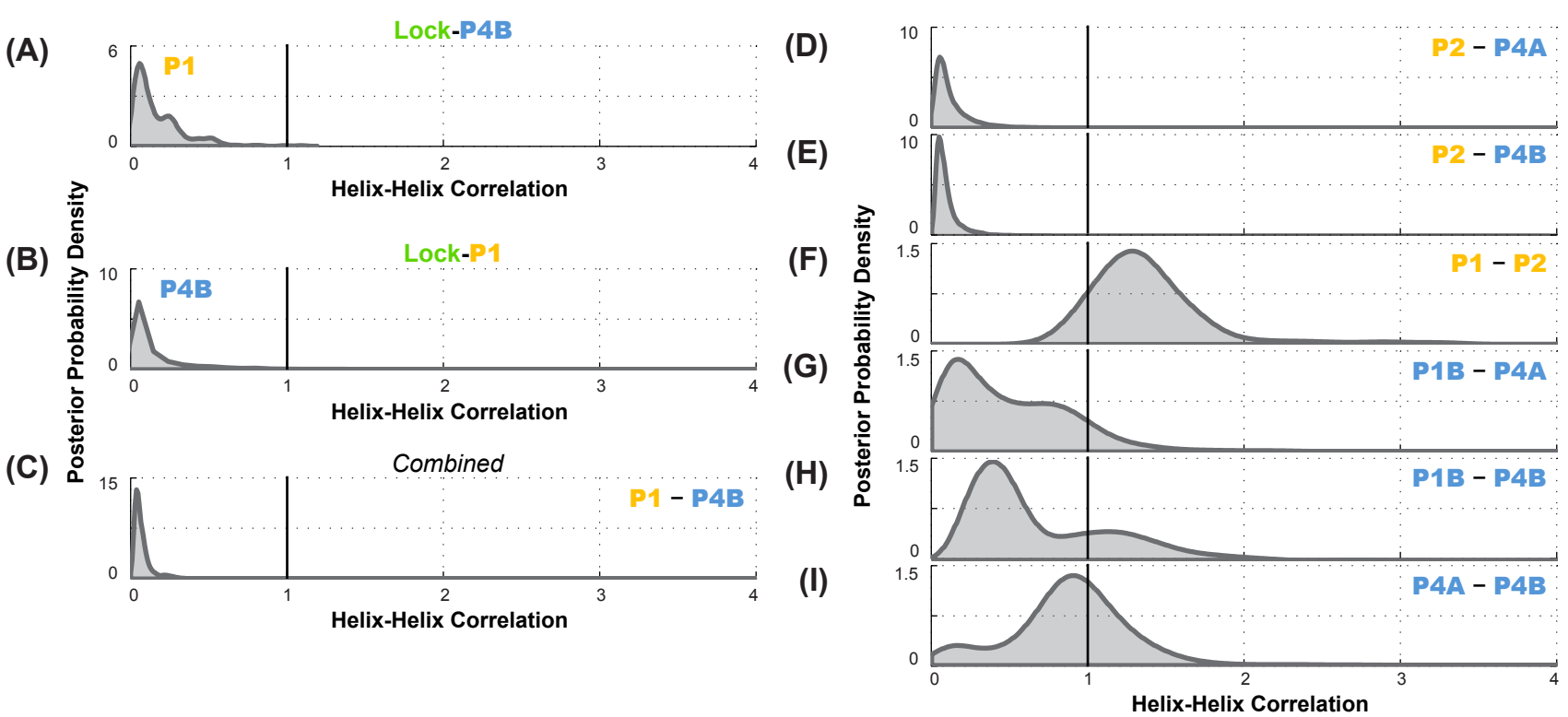

(A)
Lock-P1
Lock-P4A
SHAPE
DMS
Glyoxal
Terbium(III)
UV 302 nm
Linear Fitting
FIT : Lock-P1 + Lock-P4A
20
40
60
80
100
120
140
(B)
(C)
SHAPE Reactivity
2.0
1.0
0
WT (No Adenine)
Lock-P1 (No Adenine)
Lock-P4A (No Adenine)
FIT
WT (No Adenine)
WT (5 mM Adenine)
Lock-P1 (No Adenine)
Lock-P2 (No Adenine)
Lock-P1 (5 mM Adenine)
Lock-P2 (5 mM Adenine)
Sequence Position

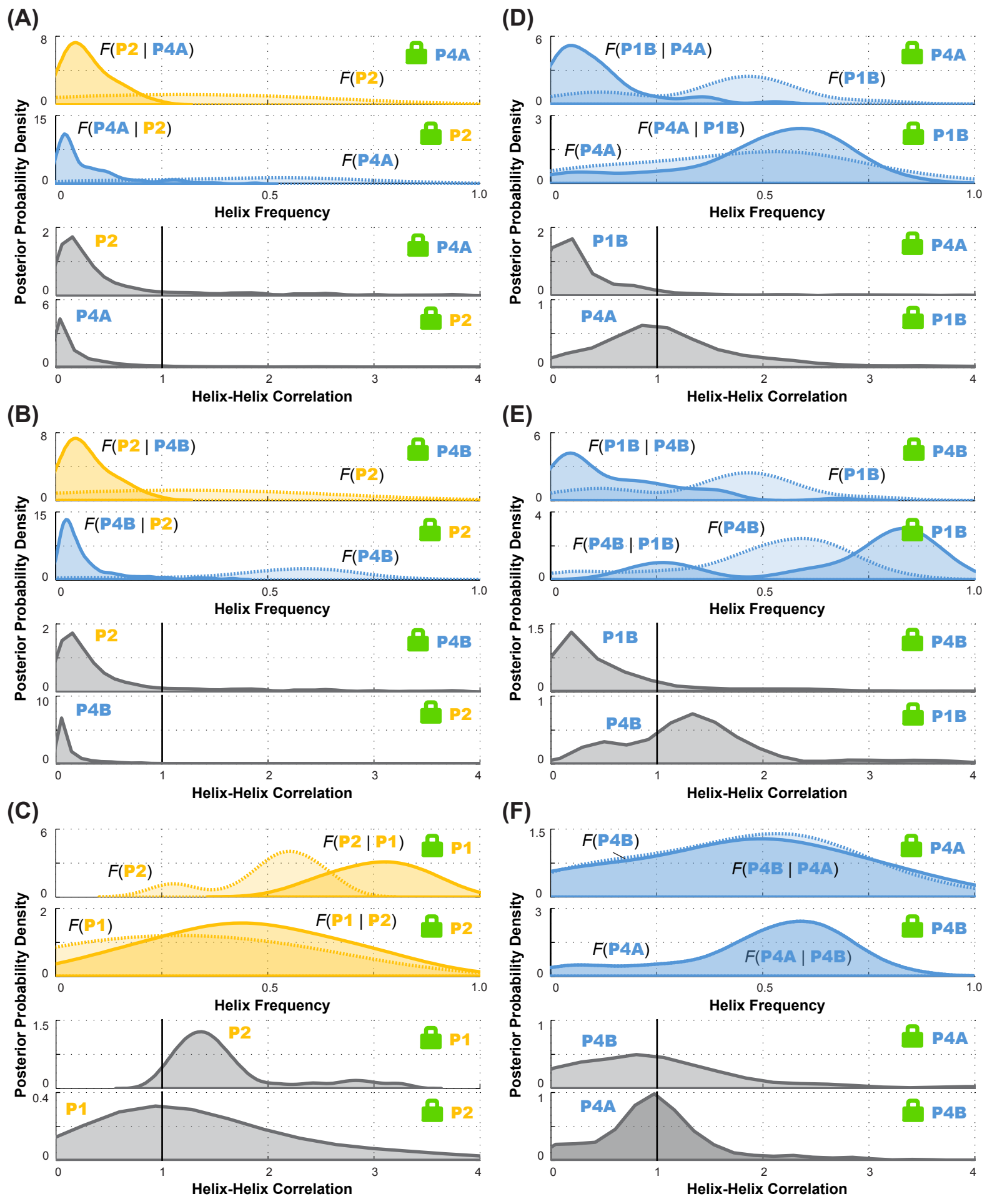

(A)
P2B
P1B
P4B
P4C
P4A
P5
P3
(D)
P2
P1
P3
P5
(B)
Lock-P4B (U82A, C83G, G115C, A116U)
No Adenine, Rescue Factor = 0.005
P1 Lock C15G G81C C15G, G81C
No Adenine, Rescue Factor = -0.472
P1 Lock U16G A80C U16G, A80C
No Adenine, Rescue Factor = -1.041
P1 Lock U17G A79C U17G, A79C
No Adenine, Rescue Factor = -0.116
P1 Lock C18G G78C C18G, G78C
No Adenine, Rescue Factor = -0.200
P1 Lock A19C U77G A19C, U77G
No Adenine, Rescue Factor = -0.236
P1 Lock U20G A76C U20G, A76C
No Adenine, Rescue Factor = -0.179
P1 Lock A21C U75G A21C, U75G
Sequence Position
(E)
Lock-P1 (A19U, U20A, A76U, U77A)
No Adenine, Rescue Factor = -0.126
P4B Lock G81C C117G G81C, C117G
No Adenine, Rescue Factor = -0.316
P4B Lock U82G A116C U82G, A116C
No Adenine, Rescue Factor = 0.160
P4B Lock C83G G115C C83G, G115C
No Adenine, Rescue Factor = 0.089
P4B Lock U84G A114C U84G, A114C
Sequence Position
(C)
P4B
Posterior Probability Density
F(P1 | P4B)
F(P1)
P1 Helix Frequency
(F)
P1
F(P4B | P1)
F(P4B)
P4B Helix Frequency

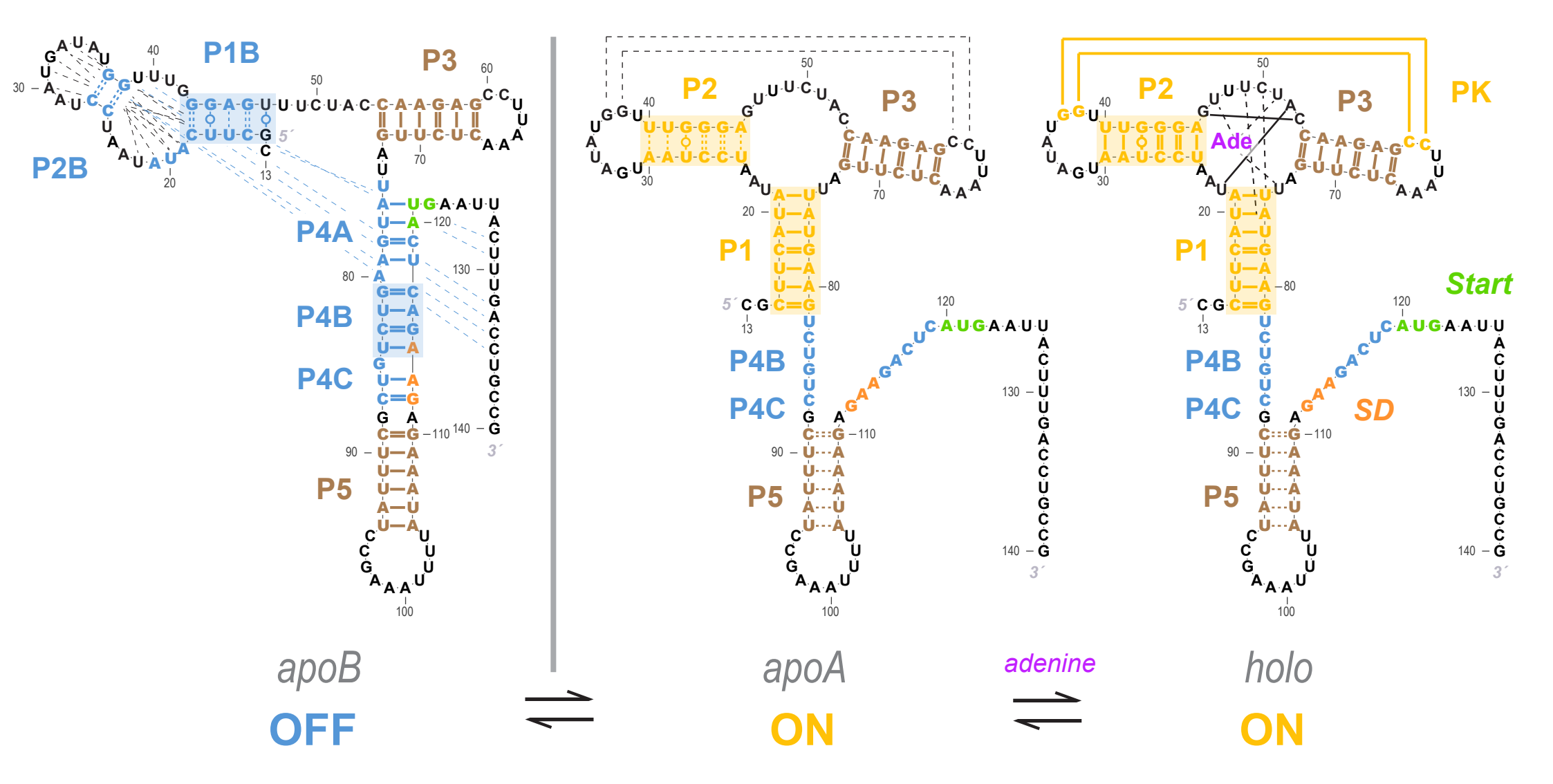
(A) Lock-P1 (A19U, U20A, A76U, U77A)
No Adenine, Rescue Factor = 0.280
P2 Lock U25G A45C U25G, A45C
No Adenine, Rescue Factor = 0.073
P2 Lock C26G G44C C26G, G44C
No Adenine, Rescue Factor = 0.472
P2 Lock C27G G43C C27G, G43C
No Adenine, Rescue Factor = 0.216
P2 Lock U28G G42C U28G, G42C
No Adenine, Rescue Factor = 0.005
P2 Lock A29C U41G A29C, U41G
No Adenine, Rescue Factor = 0.179
P2 Lock A30C U40G A30C, U40G
No Adenine, Rescue Factor = -0.126
P4B Lock G81C C117G G81C, C117G
No Adenine, Rescue Factor = -0.316
P4B Lock U82G A116C U82G, A116C
No Adenine, Rescue Factor = 0.160
P4B Lock C83G G115C C83G, G115C
No Adenine, Rescue Factor = 0.089
P4B Lock U84G A114C U84G, A114C
20 40 60 80 100 120 140
Sequence Position
(B) Lock-P2 (U28G, A29U, U41A, G42C)
No Adenine, Rescue Factor = 0.255
P1 Lock C15G G81C C15G, G81C
No Adenine, Rescue Factor = 0.678
P1 Lock U16G A80C U16G, A80C
No Adenine, Rescue Factor = 0.724
P1 Lock U17G A79C U17G, A79C
No Adenine, Rescue Factor = 0.740
P1 Lock C18G G78C C18G, G78C
No Adenine, Rescue Factor = 0.581
P1 Lock A19C U77G A19C, U77G
No Adenine, Rescue Factor = 0.730
P1 Lock U20G A76C U20G, A76C
No Adenine, Rescue Factor = 0.560
P1 Lock A21C U75G A21C, U75G
No Adenine, Rescue Factor = 0.008
P4A Lock A76C U121G A76C, U121G
No Adenine, Rescue Factor = -0.148
P4A Lock U77G A120C U77G, A120C
No Adenine, Rescue Factor = -0.066
P4A Lock G78C C119G G78C, C119G
No Adenine, Rescue Factor = -0.013
P4A Lock A79C U118G A79C, U118G
No Adenine, Rescue Factor = 0.045
P4B Lock G81C C117G G81C, C117G
No Adenine, Rescue Factor = -0.280
P4B Lock U82G A116C U82G, A116C
No Adenine, Rescue Factor = 0.181
P4B Lock C83G G115C C83G, G115C
No Adenine, Rescue Factor = -0.011
P4B Lock U84G A114C U84G, A114C
20 40 60 80 100 120 140
Sequence Position
(C) Lock-P4A (U77A, G78C, C119G, A120U)
No Adenine, Rescue Factor = -0.366
P2 Lock U25G A45C U25G, A45C
No Adenine, Rescue Factor = -0.104
P2 Lock C26G G44C C26G, G44C
No Adenine, Rescue Factor = -0.020
P2 Lock C27G G43C C27G, G43C
No Adenine, Rescue Factor = -1.035
P2 Lock U28G G42C U28G, G42C
No Adenine, Rescue Factor = -0.445
P2 Lock A29C U41G A29C, U41G
No Adenine, Rescue Factor = -0.024
P2 Lock A30C U40G A30C, U40G
No Adenine, Rescue Factor = -0.017
P1B Lock G14C U47G G14C, U47G
No Adenine, Rescue Factor = 0.037
P1B Lock C15G G46C C15G, G46C
No Adenine, Rescue Factor = -0.011
P1B Lock U16G A45C U16G, A45C
No Adenine, Rescue Factor = -0.085
P1B Lock U17G G44C U17G, G44C
No Adenine, Rescue Factor = 0.040
P1B Lock C18G G43C C18G, G43C
No Adenine, Rescue Factor = -0.180
P2B Lock C26G G38C C26G, G38C
No Adenine, Rescue Factor = 0.392
P2B Lock C27G G37C C27G, G37C
No Adenine, Rescue Factor = 0.508
P4B Lock G81C C117G G81C, C117G
No Adenine, Rescue Factor = 0.317
P4B Lock U82G A116C U82G, A116C
No Adenine, Rescue Factor = 0.509
P4B Lock C83G G115C C83G, G115C
No Adenine, Rescue Factor = 0.366
P4B Lock U84G A114C U84G, A114C
20 40 60 80 100 120 140
Sequence Position
(D) Lock-P4B (U82A, C83G, G115C, A116U)
No Adenine, Rescue Factor = 0.005
P1 Lock C15G G81C C15G, G81C
No Adenine, Rescue Factor = -0.472
P1 Lock U16G A80C U16G, A80C
No Adenine, Rescue Factor = -1.041
P1 Lock U17G A79C U17G, A79C
No Adenine, Rescue Factor = -0.116
P1 Lock C18G G78C C18G, G78C
No Adenine, Rescue Factor = -0.200
P1 Lock A19C U77G A19C, U77G
No Adenine, Rescue Factor = -0.236
P1 Lock U20G A76C U20G, A76C
No Adenine, Rescue Factor = -0.179
P1 Lock A21C U75G A21C, U75G
No Adenine, Rescue Factor = 0.113
P2 Lock U25G A45C U25G, A45C
No Adenine, Rescue Factor = -0.188
P2 Lock C26G G44C C26G, G44C
No Adenine, Rescue Factor = -0.087
P2 Lock C27G G43C C27G, G43C
No Adenine, Rescue Factor = -0.750
P2 Lock U28G G42C U28G, G42C
No Adenine, Rescue Factor = -0.496
P2 Lock A29C U41G A29C, U41G
No Adenine, Rescue Factor = -0.223
P2 Lock A30C U40G A30C, U40G
No Adenine, Rescue Factor = 0.456
P1B Lock G14C U47G G14C, U47G
No Adenine, Rescue Factor = -0.447
P1B Lock C15G G46C C15G, G46C
No Adenine, Rescue Factor = 0.108
P1B Lock U16G A45C U16G, A45C
No Adenine, Rescue Factor = -0.062
P1B Lock U17G G44C U17G, G44C
No Adenine, Rescue Factor = 0.097
P1B Lock C18G G43C C18G, G43C
No Adenine, Rescue Factor = -0.188
P2B Lock C26G G38C C26G, G38C
No Adenine, Rescue Factor = 0.348
P2B Lock C27G G37C C27G, G37C
No Adenine, Rescue Factor = 0.584
P4A Lock A76C U121G A76C, U121G
No Adenine, Rescue Factor = 0.487
P4A Lock U77G A120C U77G, A120C
No Adenine, Rescue Factor = 0.540
P4A Lock G78C C119G G78C, C119G
No Adenine, Rescue Factor = 0.229
P4A Lock A79C U118G A79C, U118G
20 40 60 80 100 120 140
Sequence Position
(E) Lock-P1B (U17G, C18G, G43C, G44C)
No Adenine, Rescue Factor = 0.561
P2B Lock C26G G38C C26G, G38C
No Adenine, Rescue Factor = 0.272
P2B Lock C27G G37C C27G, G37C
No Adenine, Rescue Factor = 0.442
P4A Lock A76C U121G A76C, U121G
No Adenine, Rescue Factor = 0.538
P4A Lock U77G A120C U77G, A120C
No Adenine, Rescue Factor = 0.645
P4A Lock G78C C119G G78C, C119G
No Adenine, Rescue Factor = 0.121
P4A Lock A79C U118G A79C, U118G
No Adenine, Rescue Factor = 0.651
P4B Lock G81C C117G G81C, C117G
No Adenine, Rescue Factor = 0.525
P4B Lock U82G A116C U82G, A116C
No Adenine, Rescue Factor = 0.759
P4B Lock C83G G115C C83G, G115C
No Adenine, Rescue Factor = 0.345
P4B Lock U84G A114C U84G, A114C
No Adenine, Rescue Factor = 0.307
P3 Lock C54G G72C C54G, G72C
No Adenine, Rescue Factor = 0.620
P3 Lock A55C U71G A55C, U71G
No Adenine, Rescue Factor = 0.759
P3 Lock A56C U70G A56C, U70G
No Adenine, Rescue Factor = 0.755
P3 Lock G57C C69G G57C, C69G
No Adenine, Rescue Factor = 0.704
P3 Lock A58C U68G A58C, U68G
No Adenine, Rescue Factor = 0.765
P3 Lock G59C C67G G59C, C67G
20 40 60 80 100 120 140
Sequence Position

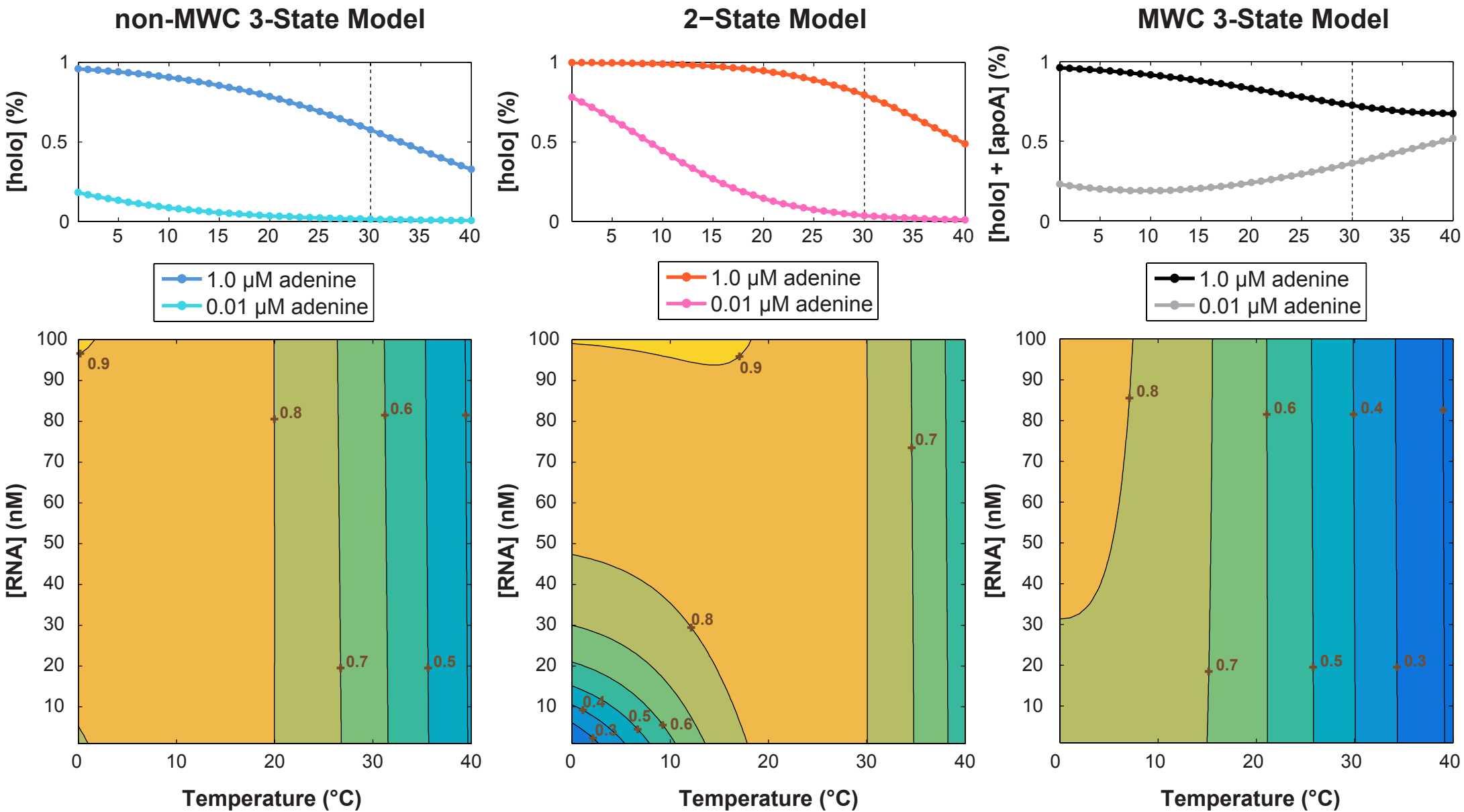
MutP2 (A29C, A30G, U40C, U41G)

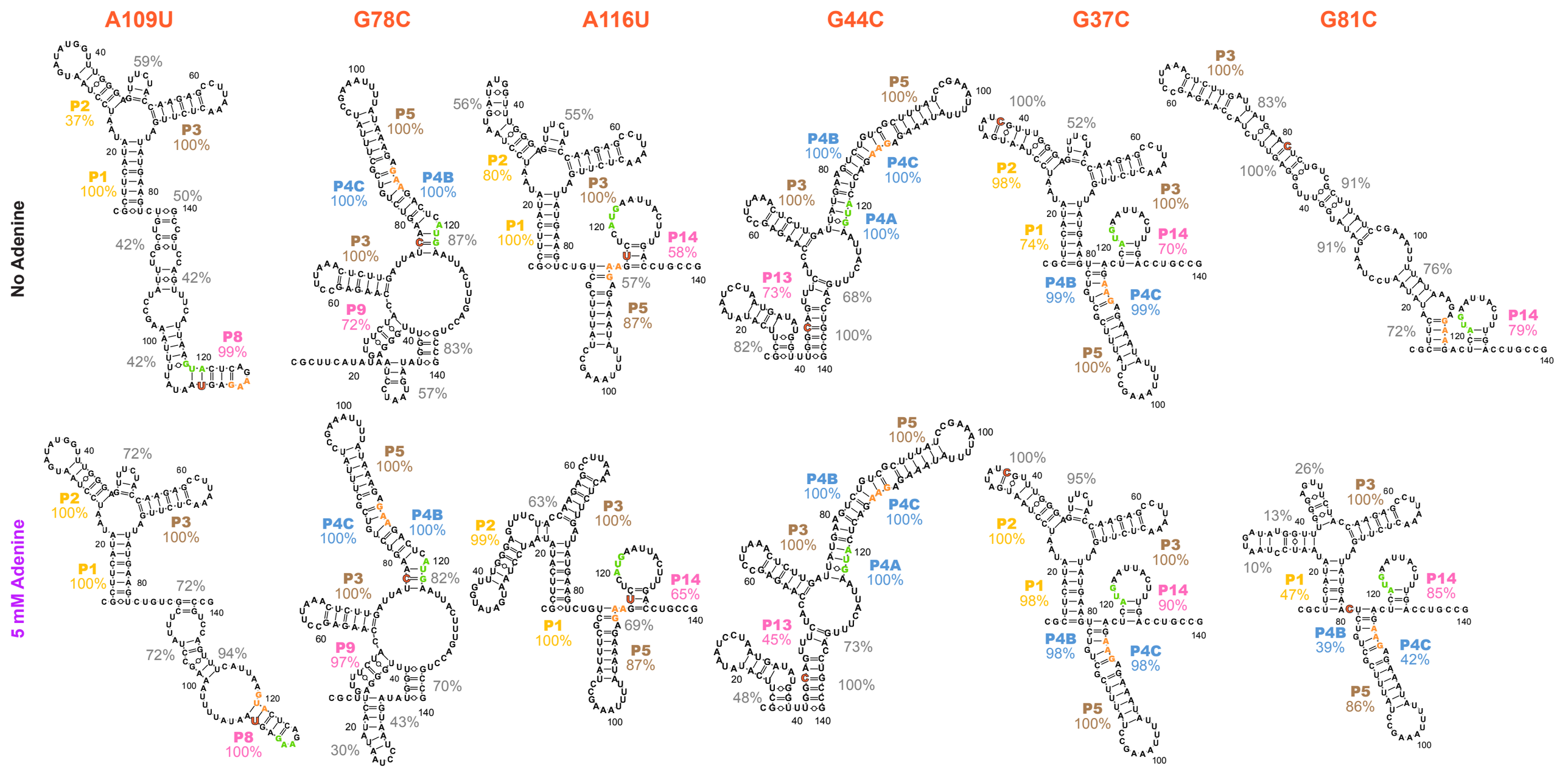


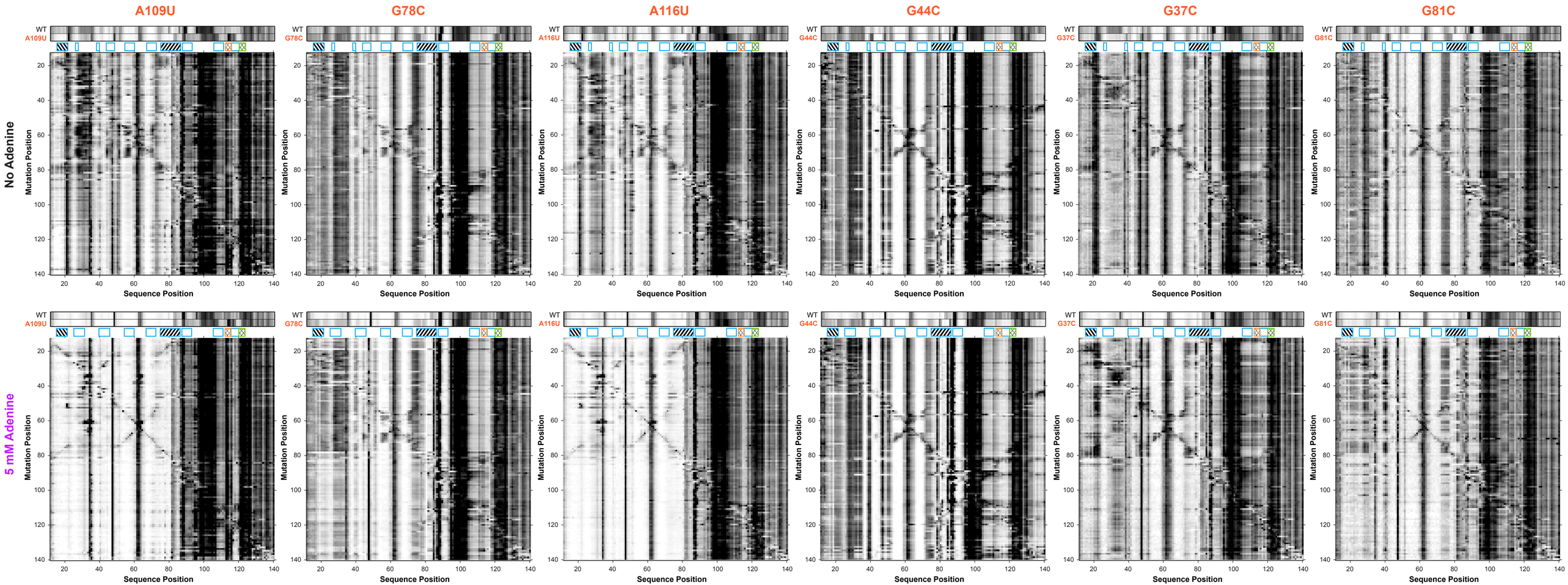

(A)
Lock-P4B
P1
6
0
0
1
2
3
4
Helix-Helix Correlation
(B)
Lock-P1
P4B
10
0
0
1
2
3
4
Helix-Helix Correlation
(C)
Combined
P1 – P4B
15
0
0
1
2
3
4
Helix-Helix Correlation
Posterior Probability Density
(D)
P2 – P4A
10
0
(E)
P2 – P4B
10
0
(F)
P1 – P2
1.5
0
(G)
P1B – P4A
1.5
0
(H)
P1B – P4B
1.5
0
(I)
P4A – P4B
1.5
0
0
1
2
3
4
Helix-Helix Correlation
Posterior Probability Density


(A)
F(P2 | P4A)
F(P2)
P4A
F(P4A | P2)
F(P4A)
P2
Posterior Probability Density
Helix Frequency
P2
P4A
P4A
P2
Helix-Helix Correlation
(B)
F(P2 | P4B)
F(P2)
P4B
F(P4B | P2)
F(P4B)
P2
Posterior Probability Density
Helix Frequency
P2
P4B
P4B
P2
Helix-Helix Correlation
(C)
F(P2 | P1)
F(P2)
P1
F(P1)
F(P1 | P2)
P2
Posterior Probability Density
Helix Frequency
P2
P1
P1
P2
Helix-Helix Correlation
(D)
F(P1B | P4A)
F(P1B)
P4A
F(P4A | P1B)
F(P4A)
P1B
Posterior Probability Density
Helix Frequency
P1B
P4A
P4A
P1B
Helix-Helix Correlation
(E)
F(P1B | P4B)
F(P1B)
P4B
F(P4B)
F(P4B | P1B)
P1B
Posterior Probability Density
Helix Frequency
P1B
P4B
P4B
P1B
Helix-Helix Correlation
(F)
F(P4B)
F(P4B | P4A)
P4A
F(P4A)
F(P4A | P4B)
P4B
Posterior Probability Density
Helix Frequency
P4B
P4A
P4A
P4B
Helix-Helix Correlation


P1B
P2B
P3
P4A
P4B
P4C
P5
P2
P1
PK
Ade
Start
SD
apoB
OFF
apoA
ON
adenine
holo
ON


non-MWC 3-State Model
2−State Model
MWC 3-State Model
[holo] (%)
[holo] + [apoA] (%)
1.0 μM adenine
0.01 μM adenine
[RNA] (nM)
Temperature (°C)
0.9
0.8
0.7
0.6
0.5
0.4
0.3


A109U
G78C
A116U
G44C
G37C
G81C
No Adenine
5 mM Adenine
P1
P2
P3
P4
P4A
P4B
P4C
P5
P8
P9
P13
P14


A109U
G78C
A116U
G44C
G37C
G81C
No Adenine
5 mM Adenine
WT
Mutation Position
Sequence Position
20
40
60
80
100
120
140

(A) M²R Single Base-Pair

	Manual Score	Auto Score
P1	? ? + + ✗ ✗ ✗	✗ + + ✗ ✗ ✓ ✗
P2	? ? ? + ✗ ✗	+ ✗ ✗ ✗ ✗ ✗
P3	+ ✓ ✓ ✓ ✓ +	+ ✓ ✓ ✓ ✓ +
P5	? + + + + +	+ + + + ✓ +
P1B	+ ✗ ✗ ✗ ✗	+ ✗ ✓ ✗ ✗
P2B	+ +	+ ✗
P4A	✓ ✓ ✓ ?	+ + + +
P4B	+ + + ✗	+ + ✓ ✗
P4C	✓ ✓	+ +
P6	? ? ? ✗ ✗ ✗	✗ ✗ ? ✗ ✗ ✗
P8	✗ ✗ ✗ ✗	✗ ✗ ✗ ✗
P9	✗ ✓ ✗ ✗ ✗	✗ + ✗ ✗ ✗
P10	✗ ✗ ?	✗ ✗ ✗
P11	✗ ✗ ? ? ? ✗ ✗ ✓ ✗	✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗
P12	✗ ✗ ?	✗ + ✗
P13	? + ✗ ✗	? + ✗ ✗
P14	✗ ✗ ✗	✗ ✗ ✗
P15	✗ ✗ ✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ✗ ✗ ✗ ✗
P16	? ✗ + + +	✗ ✗ ✗ ✗ ✗
P17	+ ✗ ✗ ✗	✗ ✗ ✗ ✗

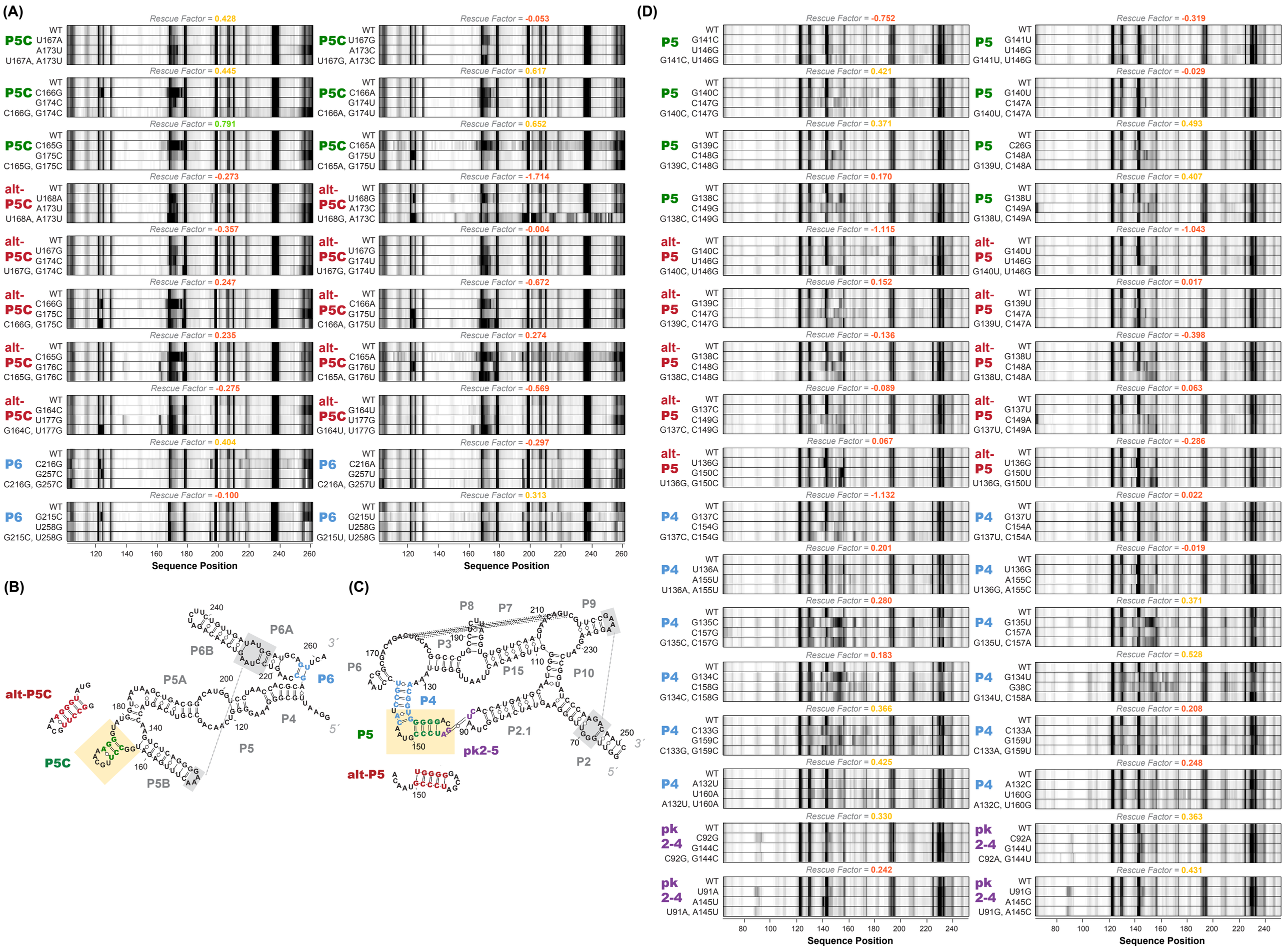
(B) M²R Double Base-Pair

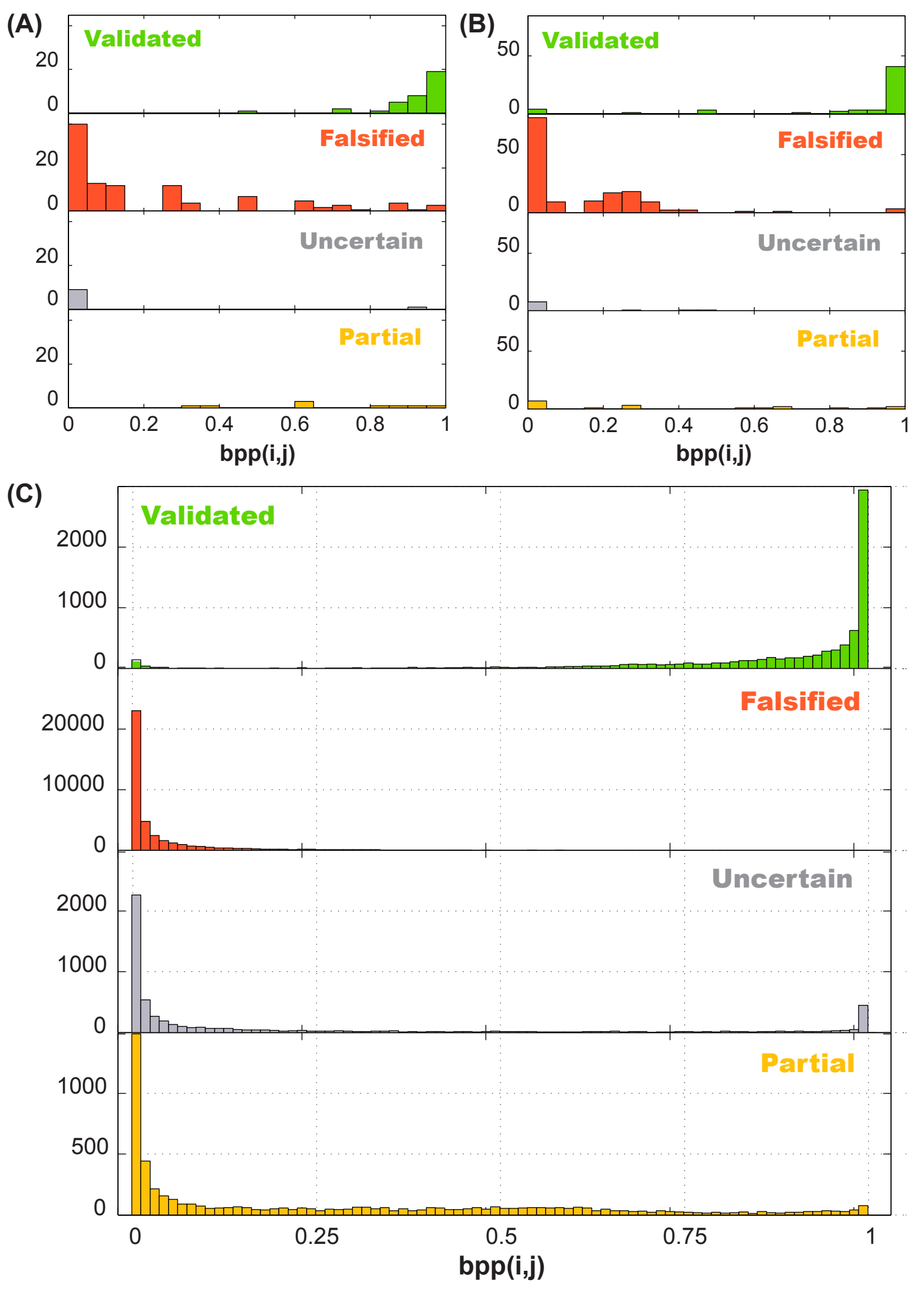
	Manual Score	Auto Score
P1	✗ ✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ✗ ✗ ✗
P2	? ✗ ✗ ✗ ✗	✗ ✗ ✗ ✗ ✗
P3	✓ + ✓ ✓ ✓	+ ✗ ✓ + +
P5	+ + + + +	+ + ✗ + ✗
P1B	? ✗ ✗ +	✗ ✗ ✗ ✗
P2B	+	✗
P4A	✗ + +	✗ ✗ +
P4B	+ + +	✗ + ✗
P4C	+	✗
P6	✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ✗ ✗
P13	✗ ✗ ✗	✗ ✗ ✗

(C) LM²R

	Manual Score	Auto Score
Lock-P1		
P2	+ ✗ ✓ + + +	✗ ✗ + ✗ ✗ ✗
P4B	✗ ✗ ✗ ?	✗ ✗ ✗ ✗
Lock-P2		
P1	✗ ✓ ✓ ✓ ✓ ✓ ✓	✗ ✓ ✓ ✓ + ✓ +
P4A	✗ ✗ ✗ ✗	✗ ✗ ✗ ✗
P4B	✗ ✗ ✗ ?	✗ ✗ ✗ ✗
Lock-P4A		
P1B	✗ ? ✗ ✗ +	? ✗ ✗ ✗ ✗
P2B	? +	✗ +
P2	✗ ✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ✗ ? ?
P4B	+ + ✓ ?	+ + + +
P6	✗ ✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ? ? ?
P13	✗ ✗ ✗ ✗	? ✗ ✗ ?
Lock-P1B		
P2B	✓ ✓	+ ✗
P4A	✓ ✓ ✓ ✓	+ + ✓ ✗
P4B	✓ ✓ ✓ ?	✓ + ✓ ?
P3	✓ ✓ ✓ ✓ ✓ ✓	✗ ✓ ✓ ✓ ✓ ✓
Lock-P4B		
P1B	✓ ✓ ✗ ✗ ✗	+ ? ✗ ✗ ✗
P2B	✗ +	✗ +
P1	? ? ? ✗ ✗ ✗ ✗	✗ ? ? ✗ ✗ ✗ ?
P2	✗ ✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ? ? ?
P4A	? + ✓ ✓	+ + + ✗
Mut-P2		
P1	✗ + + + + + +	✗ ✗ ✗ ✗ ✗ + ✗
P4A	✗ ✗ ✗ ✗	✗ ✗ ✗ ✗
P4B	✗ + ✗ +	✗ ✗ ✗ ✗

Legend: ✓ Validated ✗ Falsified ? Uncertain + Partial


(A)
Rescue Factor = 0.428
P5C WT U167A A173U U167A, A173U
Rescue Factor = -0.053
P5C WT U167G A173C U167G, A173C
Rescue Factor = 0.445
P5C WT C166G G174C C166G, G174C
Rescue Factor = 0.617
P5C WT C166A G174U C166A, G174U
Rescue Factor = 0.791
P5C WT C165G G175C C165G, G175C
Rescue Factor = 0.652
P5C WT C165A G175U C165A, G175U
Rescue Factor = -0.273
alt-P5C WT U168A A173U U168A, A173U
Rescue Factor = -1.714
alt-P5C WT U168G A173C U168G, A173C
Rescue Factor = -0.357
alt-P5C WT U167G G174C U167G, G174C
Rescue Factor = -0.004
alt-P5C WT U167G G174U U167G, G174U
Rescue Factor = 0.247
alt-P5C WT C166G G175C C166G, G175C
Rescue Factor = -0.672
alt-P5C WT C166A G175U C166A, G175U
Rescue Factor = 0.235
alt-P5C WT C165G G176C C165G, G176C
Rescue Factor = 0.274
alt-P5C WT C165A G176U C165A, G176U
Rescue Factor = -0.275
alt-P5C WT G164C U177G G164C, U177G
Rescue Factor = -0.569
alt-P5C WT G164U U177G G164U, U177G
Rescue Factor = 0.404
P6 WT C216G G257C C216G, G257C
Rescue Factor = -0.297
P6 WT C216A G257U C216A, G257U
Rescue Factor = -0.100
P6 WT G215C U258G G215C, U258G
Rescue Factor = 0.313
P6 WT G215U U258G G215U, U258G
120 140 160 180 200 220 240 260
Sequence Position
(B)
alt-P5C
P5C
P6A
P6B
P5A
P6
P4
P5
P5B
3′
5′
(C)
P8
P7
P9
P3
P6
P15
P10
P4
P5
P2.1
pk2-5
P2
alt-P5
3′
5′
(D)
Rescue Factor = -0.752
P5 WT G141C U146G G141C, U146G
Rescue Factor = -0.319
P5 WT G141U U146G G141U, U146G
Rescue Factor = 0.421
P5 WT G140C C147G G140C, C147G
Rescue Factor = -0.029
P5 WT G140U C147A G140U, C147A
Rescue Factor = 0.371
P5 WT G139C C148G G139C, C148G
Rescue Factor = 0.493
P5 WT C26G C148A G139U, C148A
Rescue Factor = 0.170
P5 WT G138C C149G G138C, C149G
Rescue Factor = 0.407
P5 WT G138U C149A G138U, C149A
Rescue Factor = -1.115
alt-P5 WT G140C U146G G140C, U146G
Rescue Factor = -1.043
alt-P5 WT G140U U146G G140U, U146G
Rescue Factor = 0.152
alt-P5 WT G139C C147G G139C, C147G
Rescue Factor = 0.017
alt-P5 WT G139U C147A G139U, C147A
Rescue Factor = -0.136
alt-P5 WT G138C C148G G138C, C148G
Rescue Factor = -0.398
alt-P5 WT G138U C148A G138U, C148A
Rescue Factor = -0.089
alt-P5 WT G137C C149G G137C, C149G
Rescue Factor = 0.063
alt-P5 WT G137U C149A G137U, C149A
Rescue Factor = 0.067
alt-P5 WT U136G G150C U136G, G150C
Rescue Factor = -0.286
alt-P5 WT U136G G150U U136G, G150U
Rescue Factor = -1.132
P4 WT G137C C154G G137C, C154G
Rescue Factor = 0.022
P4 WT G137U C154A G137U, C154A
Rescue Factor = 0.201
P4 WT U136A A155U U136A, A155U
Rescue Factor = -0.019
P4 WT U136G A155C U136G, A155C
Rescue Factor = 0.280
P4 WT G135C C157G G135C, C157G
Rescue Factor = 0.371
P4 WT G135U C157A G135U, C157A
Rescue Factor = 0.183
P4 WT G134C C158G G134C, C158G
Rescue Factor = 0.528
P4 WT G134U G38C G134U, C158A
Rescue Factor = 0.366
P4 WT C133G G159C C133G, G159C
Rescue Factor = 0.208
P4 WT C133A G159U C133A, G159U
Rescue Factor = 0.425
P4 WT A132U U160A A132U, U160A
Rescue Factor = 0.248
P4 WT A132C U160G A132C, U160G
Rescue Factor = 0.330
pk 2-4 WT C92G G144C C92G, G144C
Rescue Factor = 0.363
pk 2-4 WT C92A G144U C92A, G144U
Rescue Factor = 0.242
pk 2-4 WT U91A A145U U91A, A145U
Rescue Factor = 0.431
pk 2-4 WT U91G A145C U91G, A145C
80 100 120 140 160 180 200 220 240
Sequence Position


(A)
Validated
Falsified
Uncertain
Partial
20
0
0
0.2
0.4
0.6
0.8
1
bpp(i,j)
(B)
Validated
Falsified
Uncertain
Partial
50
0
0
0.2
0.4
0.6
0.8
1
bpp(i,j)
(C)
Validated
Falsified
Uncertain
Partial
2000
1000
0
20000
10000
0
2000
1000
0
1000
500
0
0
0.25
0.5
0.75
1
bpp(i,j)