Deep mutational scanning of an *S. pneumoniae* FMN riboswitch reveals robustness during mouse infection but diverging adaptive landscapes in response to targeting antibiotics

Rebecca Korn and Michelle M. Meyer*

Boston College Biology Department, Chestnut Hill MA 02467

*corresponding author: m.meyer@bc.edu

ABSTRACT

RNA-based control of gene expression is common across all domains of life, yet shows extraordinary variety in terms of environmental cues recognized, mechanisms of action, and magnitude of gene expression changes. The fitness benefit conferred by RNA-mediated regulation is often subtle and challenging to measure and may be highly influenced by the environment, yet selective forces are sufficient to generate and maintain complex RNA structures that allow rapid and robust response to stimuli. In this work we perform deep mutational scanning to measure the relative fitness conferred by all single mutants of an FMN riboswitch to the opportunistic pathogen *Streptococcus pneumoniae* under seven different conditions. We find that in standard culture conditions the RNA sequence is not under selective pressure, however the dynamic conditions encountered during mouse lung infection reveal an element that is under weak purifying selection, but robust to strongly deleterious mutations. Furthermore, assessment of fitness in the presence of two different targeting antibiotics exposes diverging adaptive landscapes that reveal suites of mutations conferring either constitutive activation or repression of gene expression. Analysis of these landscapes highlights a complex relationship between antibiotic concentration and the fitness benefit conferred by individual mutations as well as provides insights into the mechanism of action for this RNA with lessons broadly applicable to interpretation of bacterial transcriptomics.

INTRODUCTION

The interplay between genotype and phenotype is central to understanding how organisms adapt to their environments, evolve new traits, and respond to external challenges. In bacteria, regulation that occurs after transcription initiation plays an important role in allowing rapid adjustment to environmental changes. Regulatory RNAs such as trans-acting small RNAs (sRNAs) and cis-acting elements, including riboswitches and RNA thermometers, modulate gene expression in response to environmental cues, tuning metabolic processes, inducing stress responses, and activating resistance mechanisms[1]. Riboswitches, RNAs that respond to specific small molecule metabolites to alter gene expression, often control essential metabolic pathways and transporters. These RNAs act as molecular sensors that directly integrate metabolic cues with regulatory features to allow conditional intrinsic transcription termination or translation initiation. Riboswitches have emerged as potential targets for antimicrobials[1–8], and due to their dynamic nature and relatively small size, have also proven to be valuable biophysical model systems for studying the relationships between RNA sequence, structure, and function *in vitro*[9,10]. However, our understanding of how sequence changes in such RNAs drive functional changes in regulation within the cell, and how these regulatory shifts ultimately affect an organism's phenotype and fitness in diverse or dynamic environments, remains limited.

Fitness landscapes measured for other RNAs have yielded invaluable insights into the RNA folding landscape[11], the effects of epistasis[12], the degree to which fitness peaks are isolated from one another[13–15], and the impacts of environmental perturbations on RNA function[16,17]. However, such studies are generally conducted *in vitro* to enable assessment of large numbers of variants[11,13–17] or utilize engineered genetic selection schemes to partition functional and non-functional sequences[12]. Thus, how changes in ncRNA sequence influence regulatory function and ultimately organism phenotype remain unstudied in the context of dynamic regulatory RNAs that are dependent on cellular components for biological function.

In this work, we generated fitness landscapes for a 372 nucleotide *Streptococcus pneumoniae* flavin mononucleotide (FMN) riboswitch[18] that assess all 1,116 single point mutations under seven different conditions including *in vivo* infection and antibiotic perturbed environments. *S. pneumoniae* is a significant global health concern as the leading cause of lower respiratory infection worldwide, contributing to over 1 million deaths globally in 2016[19]. In the United States, antibiotic-resistant pneumococcal diseases cause around 1.2 million illnesses and 7,000 deaths annually[20]. There are two FMN riboswitches in *S. pneumoniae* that function as cis-regulatory "off" switches; one preceding the biosynthesis operon (*ribB/A,D,E,H*) and the other preceding the gene encoding the riboflavin transport protein RibU. Riboflavin (vitamin B_2) is an essential nutrient, and its derivatives, FMN and flavin adenine dinucleotide (FAD), serve as critical cofactors in cellular metabolism, catalyzing redox reactions[1]. The FMN riboswitch is one of the most frequent and widely distributed riboswitches, occurring across a broad swath of diverse bacterial clades, with especially high prevalence across potential pathogenic species[21,22]. Thus, this riboswitch is specifically targeted by several different antimicrobial compounds that ultimately inhibit both FMN synthesis and transport[23–29].

We show that the FMN riboswitch preceding RibU is under minimal selection in standard culture conditions, but within *in vivo* mouse lung infections the RNA is under weak purifying selection, and robust to mutations that are strongly deleterious to fitness in this environment. Furthermore, antibiotic perturbation of the *in vitro* environment with antibiotics identified suites of mutations with constitutively active and repressed behavior, which in turn enable insights into the RNA mechanism of action and mechanisms of drug tolerance. Our study uniquely assesses a regulatory RNA function within its native context to highlight the complex interplay between environment, selective pressure on the organism from the environment, and potential phenotypes accessible via single mutations.

RESULTS

Single mutations to the FMN riboswitch are neutral in chemically defined medium without riboflavin

The *S. pneumoniae* FMN riboswitch preceding *ribU* spans 372 nucleotides, from the transcription start site to a transcription termination site just upstream of the translational start[30] **(Supplementary Figure S1).** The riboswitch includes a canonical six-stem FMN-binding aptamer with an extended P3 stem, followed by an extended stem after the aptamer that appears to enable intrinsic transcription termination, although the base-pairing changes required to connect ligand binding to termination is unclear **(Fig. 1F)**. To assess the impact of each of the 1,116 possible single-point mutants on *S. pneumoniae* fitness, we created a library of *S. pneumoniae* strains that integrates these mutations into the native locus accompanied by an antibiotic marker at an upstream neutral location[31]. Our library maximizes the proportion of single point-mutations via combination of six individually constructed libraries, each of which contains a single segment comprised of a synthetic oligo doped with 1% of each alternative base at each position (See Supplemental Fig S2A for construction methodology).

With this library we assessed the relative fitness of each strain compared to the wild-type sequence in a chemically defined medium in the absence of riboflavin (CDM – riboflavin) by culturing our strain library to mid-log phase (~3 doublings), harvesting a pair of T1 samples, and subsequently inoculating cultures in a complete defined medium (CDM) lacking riboflavin (technical triplicate) and allowing an additional ~10 doublings before harvesting cells (T2). Genomic DNA was extracted from both T1 and T2 samples, and the FMN riboswitch preceding *ribU* amplified via PCR with unique molecular identifiers (UMIs) to enable accurate quantification of each variant. The resulting amplicons were sequenced to obtain 0.6 - 7.3 million reads for each sample **(Fig. S2B, Fig. S3, Table S1)**. We observe that the distribution of wildtype versus mutants with 1-6 mutations is similar in both the T1 and T2 samples, with about 60% of reads corresponding to the wildtype sequence **(Fig. 1A)**. We also find that all single point mutations are represented in the datasets at both the T1 and T2 conditions indicating that we assessed the fitness of all mutants **(Fig. S2C, Table S2)**. We did not examine fitness of double mutants because the number of double mutants is large (1,245,456), and our construction methodology does not create all pairs of double mutants.

To quantify the fitness of each single mutant, we used the DiMSum pipeline[32]. This software package extracts the frequencies of each single mutant variant before (T1) and after (T2) selection, and calculates fitness relative to the wild-type sequence based on these frequencies (see Methods), where a fitness of 0 is equivalent to the wild-type sequence, a fitness >0 is increased frequency in T2, and a fitness <0 is a decreased frequency in T2. DiMSum analyzed 1.0 – 12.5 million deconvoluted reads for each condition (**Fig. S3A-C**). We find that in CDM without riboflavin, no individual single-point mutation exhibited a fitness greater than 1 (beneficial) or less than -1 (detrimental) (**Fig. 1B-E, Table S2**). Thus, single-point mutations do not confer a substantial fitness advantage or disadvantage to *S. pneumoniae* in culture, and the FMN riboswitch is not under strong selective pressure. This is consistent with the lack of change in the distribution in our mutation frequency profile **(Fig. 1A)** under these conditions.

To determine whether any underlying patterns might be detectable in the low-magnitude changes observed, we assessed our populations of mutants, focusing on specific regions of the riboswitch (ligand-binding aptamer, extended P3 stem (ext. P3 stem), extended terminal stem (ext. term. stem), and putative terminator) **(Fig. 1F)**. We find that in the aptamer and the ext. P3 stem, transversions have a statistically significant positive impact on fitness compared to transitions **(Fig. 2A, Table S3)**. However, the difference in fitness median (Δmedian) from transversions to transitions is 0.043 and 0.053 for the aptamer and the ext. P3 stem respectively and there are no significant differences between compensatory and disruptive mutations (**Fig. S5A**), suggesting only very small impacts on fitness. Thus, the intact riboflavin biosynthetic

operon (SP_0178 - SP_0175) appears to provide sufficient riboflavin for growth in culture lacking riboflavin and constitutive expression of the transporter at levels accessible by single mutations is not detrimental under these conditions. Given the lack of selective pressure apparent in this condition, going forward we will consider this condition our negative control.

Single mutations to the FMN riboswitch are neutral in chemically defined medium containing riboflavin

To assess whether addition of riboflavin to the medium would alter the fitness benefit conferred by expression changes to RibU, the riboflavin transporter, we grew the library for ~10 generations in the presence of riboflavin, and harvested cells for amplicon sequencing as described above. Similar to the T1 and -riboflavin condition, about 60% of the reads corresponded to the wildtype sequence (**Fig. 1A**), and no mutants had a fitness >1 or <-1 in the presence of riboflavin. These findings indicate that there is not a strong fitness advantage or disadvantage to *S. pneumoniae* for any individual mutations under this condition **(Fig. 1B-E, Table S2)**.

When comparing all regions of the riboswitch to the condition lacking riboflavin, there is no significant difference in fitness compared to the -riboflavin condition for any of the regions **(Fig. 1B-E, Table S3)**. Similarly to the -riboflavin condition, transversion mutations are slightly higher fitness than transitions with minimal ∆medians **(Fig. 2B, Table S3).** In the aptamer there are statistically significant differences between compatible mutations (those do not break Watson-Crick pairing due to the G•U wobble) at paired bases compared with both mutations disruptive of base-pairing or those at unpaired positions, but the ∆medians are substantially less than one **(Fig. S5B, Table S3)**. Similarly, in the terminator region there are statistically significant differences in fitness between mutations disruptive to base-pairing and both mutations to unpaired bases and mutations compatible with the base-pairing, but the ∆medians are very small **(Fig. S5B, Table S3)**. Thus, the addition of riboflavin to the culture medium does not substantially change the lack of strong selective pressure on the FMN riboswitch preceding RibU in *S. pneumoniae*.

The FMN riboswitch is robust to strongly deleterious single point mutations during mouse lung infection, but is under weak purifying selection

FMN biosynthesis is dispensable for *S. pneumoniae* during *in vivo* infection due to the availability of FMN in the environment. However, the fitness of strains with transposon insertions into *ribU* is severely attenuated with *in vivo* mouse models[33]. In addition, in our past assessment of fitness conferred by the FMN riboswitch preceding RibU we observed a severe *in vivo* fitness defect for a truncation mutant that resulted in constitutively repressed expression[31]. To assess whether any single mutant could recapitulate this phenotype, we inoculated 24 mice with our strain library, allowed the infection to progress for 20-24 hours, and subsequently harvested and homogenized the lung and plated on blood plates to collect the living cells. Genomic DNA was then extracted from colonies collected off these plates and amplicon sequencing performed similarly to our experiment above. A pre-inoculation sample was collected as separate T1 sample for each set of mice infected during this experiment ($T1_{mouse}$). Close to 70% of the reads coming from both the $T1_{mouse}$ and the mice are the wildtype sequence, similar to the + and -riboflavin conditions **(Fig. 1A)**, suggesting that few mutants are strongly beneficial or detrimental in the *in vivo* infection environment.

When compared to the -riboflavin condition, mutations to all regions of the riboswitch displayed significantly negative impacts on fitness, with ∆medians of 0.158 to 0.248 that are substantially higher than those of the + or -riboflavin populations (**Fig. 1B-E, Table S3**). In the infection environment, transversions are significantly more detrimental to fitness than transitions in all regions with ∆medians ranging from 0.246 to 0.620 **(Fig. 2C, Table S3)**. P3 stem mutations disruptive of predicted base-pairing are significantly more deleterious than compatible mutations (∆medians = 0.280 and 0.323), and mutations disruptive of base-pairing in the terminator are

significantly more deleterious compared to mutations at unpaired bases (Δmedian = 0.461) **(Fig. 2D, Table S3)**.

Together, these data indicate that no mutations have strong negative or positive impacts on the fitness of *S. pneumoniae* during mouse infection. However, many mutations, especially transversions, are slightly deleterious. This observation is consistent with purifying selection on the riboswitch sequence where transversions that disrupt structural features are selected against more compared to less disruptive transitions. When our observations of single mutants are combined with previous findings that transposon interruption[33], or constitutive repression[31], of *ribU* results in large fitness defects within the *in vivo* infection environment, our data strongly imply that this sequence is mutationally robust. No single mutations seem to be capable of changing the biological activity in such a way that results in a very deleterious phenotype within the infection environment.

Mutations to the FMN riboswitch aptamer domain are beneficial in the presence of ribocil-C

The antibiotic ribocil-C has been shown to interact directly with FMN riboswitches[25,26]. To determine what mutations to the FMN riboswitch preceding RibU confer fitness changes in the presence of this antibiotic, the mutant strain library was grown in CDM (lacking riboflavin) in the presence of 2.5μM and 10μM ribocil-C. These cultures only reached OD_{600} of ~0.8 and <0.4 respectively, and were harvested at 12.5 hours after only ~8 or 9 doublings. At 2.5 μM ribocil C, the distribution of wild-type and mutated variants is similar to our T1 and non-selective conditions. However, at 10μM ribocil-C the distribution of wild-type and mutated variants changed substantially with wild-type making up only 30% of the total reads and most reads containing 0-2 mutations. These findings indicate that some mutants are more beneficial than wildtype in the presence of 10μM ribocil-C **(Fig. 1A)**, but that at lower (but still inhibitory) antibiotic concentrations of ribocil-C there is less selective pressure.

We find that the positively selected mutants at 10uM ribocil are enriched in the aptamer compared to the -riboflavin condition **(Fig. 1B, Table S3)**. This difference in fitness is both substantial in magnitude (Δmedian = 1.442) and statistically significant, indicating that many mutations to the aptamer confer a large fitness advantage in the presence of 10uM ribocil-C. In the presence of 2.5μM ribocil-C, there is no such substantial enrichment, suggesting that mutations to the aptamer provide a larger benefit at higher ribocil-C concentrations. In other regions of the riboswitch, the distribution of fitness values across both antibiotic concentrations suggests similar selective pressures at both antibiotic concentrations (**Fig. 1C-E**). We directly compared the fitness of each mutant at the two different concentrations and we observe that fitness values are highly correlated (r = 0.7505). The negatively selected mutants display similar fitness values under both conditions, but the magnitude of positive selection is much larger at 10μM compared to 2.5 μM ribocil-C **(Fig S6A)**.

Visualizing the locations of beneficial and detrimental mutations at 10μM ribocil-C on the riboswitch secondary structure shows that mutations to ~83% of positions in the aptamer region increase fitness **(Fig. 3A)**, with many beneficial single-point mutations exhibiting fitness values greater than three (**Fig. 3B, Table S2**). The majority of mutations that confer positive fitness in the aptamer are transversions (fitness Δmedian between transitions and transversions is 2.469) (**Fig. 3C, Table S3**). In assessing all aptamer positions with respect to their base-pairing status, we see that mutations disruptive to Watson-Crick pairing within the stems are much more beneficial than those that are compensatory (Δmedian = 2.407) or those that are not Watson-Crick paired (Δmedian = 1.900) (**Fig. 3D, Table S3**). Furthermore, regions that are not paired in the secondary structure, but may be involved in tertiary interactions or ligand binding, show a wider distribution of effects, and are more positively selected compared to mutations compatible with base pairing (Δmedian = 0.507), but not significantly so. This pattern suggests that mutations impairing the ability of the aptamer to bind ligand are strongly positively selected at 10μM ribocil-

C. Despite the more modest selection pressure present at 2.5μM ribocil-C, similar trends are observed **(Fig. S6B,C)**.

Across other regions of the riboswitch, we generally observe that the relative fitness of mutants has a wider spread in both ribocil-C perturbed conditions compared to the non-selective conditions, and many regions show evidence of statistically significant changes in fitness **(Fig. 1C-E)**, which are supported by differences in the effects of transitions and transversions (**Fig. 3C, Fig. S6B**). However, the median fitness change of mutations in these regions is small compared to that observed for the aptamer (Δmedian = 0.076 to 0.8035) (**Fig. 1C-E, Table S3**). The largest magnitude changes are to positions classified within the putative terminator where we observe that mutations are negatively selected at 2.5μM ribocil-C. However, at 10μM a subset of these mutations overlapping the aptamer structure are positively selected, thus narrowing the Δmedian from 0.8035 to 0.5158, and rendering the median difference from the non-selective conditions insignificant **(Fig. 1E**). Thus, mutations in the putative terminator region are slightly deleterious to fitness in both concentrations of ribocil-C, but this affect is relatively small compared to the effect of positive selection for mutations that disrupt the aptamer structure at 10μM ribocil-C.

Mutations to the putative terminator region are strongly beneficial in the presence of high concentrations of roseoflavin

Although both ribocil-C and roseoflavin target the FMN riboswitch, they are not similar chemically, and structural studies show different modes of interaction[25,26,28,34]. Furthermore, roseoflavin is thought to be imported via the riboflavin transporter (RibU), under control of this riboswitch, and roseoflavin resistance mutants have been raised in FMN riboswitches preceding the transporter in *S. aureus*[28,35–37]. Thus, we hypothesized that an alternative set of mutations might confer a fitness defect or benefit on *S. pneumoniae* in the presence of roseoflavin. To identify these mutants, the mutant strain library was grown with 256nM roseoflavin for ~10 doublings. Most of the reads at 256nM roseoflavin (RoF) contained 1-3 mutations, with wildtype only making up ~10% of the total reads **(Fig. 1A).** This value is even more pronounced than that observed for 10μM ribocil-C, suggesting that the wildtype sequence is even less fit compared to the positively selected mutants in the presence of 256 nM roseoflavin compared to 10μM ribocil-C.

The mutations that are strongly positively selected at 256nM roseoflavin are clustered to the putative terminator region, and mutations in this region confer fitness values significantly more positive than in the -riboflavin condition, with a high Δmedian of 2.067 (**Fig. 1E, Fig. 4A,B Table S3**) and many individual single-point mutations displaying fitness values greater than four **(Fig. 5A,B, Table S2**). In the terminator region transversions are significantly more beneficial than transitions, with a Δmedian of 3.102 **(Fig. 5AB, Table S3)**, and mutations disruptive to predicted base-pairing have a significant positive impact on fitness compared to unpaired nucleotides (Δmedian = 3.128) or to mutants that are compatible with base-pairing (Δmedian = 3.125) **(Fig. 5B, Table S3)**. This indicates that disrupting the putative terminator stem is strongly beneficial in the presence of high roseoflavin concentrations.

In addition, we observe that across the riboswitch, the spread of relative fitness is larger in the presence of the antibiotic and there are small magnitude, but statically significant differences from fitness values in our control -riboflavin population **(Fig. 1B-E)**. Furthermore, mutations compatible with predicted base-pairing display fitness that is more similar to that of the wild-type sequence (narrower distributions around zero) compared to mutations that disrupt base-pairing or are predicted to be unpaired (**Fig. 5A,B**). The most notable changes outside the putative terminator region are to the extended P3 stem, where mutations, especially transversions that are largely disruptive to predicted secondary structure, are negatively selected while transitions/compatible mutations do not have a change in fitness compared to the wild-type **(Fig. 1C, Fig. 5A,B, Table S3)**. Thus, maintaining the structure and wildtype sequence of the extended P3 stem is important for *S. pneumoniae* fitness in the presence of roseoflavin.

Low roseoflavin concentrations reveal a complex adaptive landscape

We performed similar experiments at a lower, but still inhibitory, concentration of roseoflavin (64nM, corresponding to 2x the IC_{50} of the parental strain) to find that the percentage of wildtype reads is slightly higher compared to the 256nM condition suggesting weaker selective pressure at 64nM roseoflavin (**Fig. 1A**). In assessing which regions of the RNA are under positive selection in this condition, we see that compared to the -riboflavin condition, mutations to both the aptamer and putative terminator confer fitness values that are significantly more positive than -riboflavin and have large ∆medians, 1.206 and 2.123 respectively (**Fig. 1B,E, Table S3**). This diverges somewhat from our findings with 256nM roseoflavin where we observed this magnitude of positive fitness change predominantly within the putative terminator. Comparing the fitness changes between the two different antibiotic concentrations, we observe that the fitness changes at the two roseoflavin concentrations are correlated (along marked diagonal) (r = 0.6637) (**Fig. 5E, Table S4**). However, the degree of correlation between two concentrations is lower than that of the ribocil-C selected populations. In particular there are a large number of mutations to the aptamer region that are positively selected at 64nM but not at 256nM roseoflavin, while the positively selected mutations to the putative terminator are strongly positively selected at both concentrations.

The positive selection on mutations in the aptamer at 64nM roseoflavin is primarily driven by transversion mutations (∆median=1.972) (**Fig. 5C, Table S3**). Similarly, the fitness of mutations that disrupt base-pairing is significantly more positive than that of compensatory mutations or mutations at unpaired nucleotides (∆medians=1.749 and 1.424 respectively)(**Fig. 5D, Table S3**). This pattern suggests that disruptive mutations to the aptamer structure confer a benefit in the presence of this lower concentration of roseoflavin, mimicking the pattern we see in 10μM ribocil-C. Directly comparing the fitness of mutants at the low roseoflavin concentration (64nM) with their fitness in ribocil-C (10μM) we observe that there is significant correlation (r = 0.5778), most notably among mutations to the aptamer (**Fig. 5F, Table S4**). This correlation with the 10μM ribocil-C condition is unique to the 64nM roseoflavin concentration, as comparison of the 256nM roseoflavin concentration with the 10μM ribocil-C reveals a much weaker correlation (r = 0.1251), with few positions positively selected in both conditions (**Fig. 5G, Table S4**).

Conversely, the fitness of mutations at 64nM roseoflavin in other regions of the riboswitch are similar to that found in at 256nM roseoflavin. Specifically, mutations to the putative terminator are positively selected (**Fig. 5E, Table S4**), with transversions, particularly those that disruptive base-pairing are strongly positively selected (**Fig. 5C, D, Table S3**). Mutations to the P3 stem are negatively selected, with transversions, especially those disruptive of predicted base-pairing displaying more negative fitness values. Thus, the two different concentrations of roseoflavin reveal a pattern where some mutations are beneficial (mutations disruptive to the putative terminator) or detrimental (mutations disruptive to the extended P3 stem) compared to the wild-type sequence at both concentrations, while mutations disruptive to the aptamer are largely beneficial at low concentration, but are much more likely to be neutral at high roseoflavin concentration.

A biological replicate reinforces findings from technical replicates

The reported data is from one FMN riboswitch doped oligo library (Lib1). However, a second library was made in the same manner (separate *S. pneumoniae* transformations) as a biological replicate (Lib2). The same amplicon sequencing experiment was repeated for Lib2, and two technical replicates from each *in vitro* condition were sequenced on the NextSeq2000 (**Table S1**). To determine if the biological replicates Lib1 and Lib2 report similar results from the amplicon sequencing experiment, we compared the conditions tested between each library. For the six *in vitro* conditions tested, the r values were 0.03717 for -riboflavin, -0.004272 for +riboflavin and

ranged from 0.4723 to 0.7012 across the four antibiotic perturbed conditions **(Fig. 6A, Table S4, FigS7)**. Considering that no mutation in the + or -riboflavin condition has a fitness > |1| for Lib1, and very few do for Lib2 (**Table S5**), the low r values are not surprising, as the fitness changes likely largely correspond to noise under such conditions where there is minimal selection.

The main difference between the sequenced libraries is that the error on the fitness of Lib2 is greater than the error on the fitness of Lib1 in each condition **(Fig. S4, Table S2, Table S5)**. This is likely due to the decrease in technical replicates sequenced from Lib2 compared to Lib1 (**Table S1**). The Lib2 data predominantly serves to reaffirm the Lib1 results **(Fig. S7-10, Table S4, and Table S6)**. In Lib1 in the presence of ribocil-C, four mutants confer a fitness defect <-4 in the ext. term. stem. However, in Lib2 the same mutants in ribocil-C do not have a fitness > |1| (**Fig. 6A, Table S2, Table S5**). Since these fitness values were not replicated, we did not assess them further.

In both the ribocil-C and roseoflavin conditions, the more positive the fitness, the smaller the error, and the more negative the fitness, the larger the error (**Fig. 6A, Fig. S10, Table S2, Table S5**). This is due to the nature of the measurement, where positive fitness values correspond to a higher read count and negative fitness values a lower read count in the raw data. Thus, in our experiment we are more confident in mutants with positive fitness, than those with negative fitness. Going forward, we focus our attention on mutants that are beneficial, rather than detrimental under our selective conditions as these values are higher confidence.

Roseoflavin and ribocil-C selected mutants display contrasting gene expression levels

To validate the findings from our high-throughput experiment, we selected a panel of mutations to sample the different regions of the riboswitch found to be beneficial in the presence of high antibiotic concentrations for further analysis. Three mutants beneficial in the presence of 256nM roseoflavin (RoF1, RoF2, RoF3), four mutants beneficial in the presence of 10μM ribocil-C (RibC1, RibC2, RibC3, RibC4), and two negative control mutations in non-selected sites (NC1, NC2) were chosen **(Fig. 6B)** and *S. pneumoniae* strains carrying each individual mutation were constructed and validated (see Methods). To enable a more comprehensive comparison, we assessed five additional strains alongside our panel: the *S. pneumoniae* TIGR4 parental strain (316), and previously published mutants: a wildtype containing a chloramphenicol resistance cassette (WT^c), a ligand-insensitive mutant (mLI), an "always on" mutant (mON), and an "always off" mutant (mOFF)[31]. None of the previously constructed mutants are single point mutations, but rather represent more severe changes to the sequence that range from a double mutant (mLI) to large scale deletion of the aptamer (mOFF), or putative terminator stem (mON).

The fourteen strains were first assessed against a panel of roseoflavin concentrations ranging from 8nM to 16μM. The IC_{50}s for the RibC strains, the NC strains, 316, WT^c, mLI, and mON are ~32nM roseoflavin. The IC_{50}s for RoF1, RoF2, RoF3, and mOFF are nearly 30x higher, around 1μM roseoflavin **(Fig. 6C, Table S7)**. Thus, the roseoflavin resistant mutants and the mOFF mutant have IC_{50}s an order of magnitude higher than the RibC and control strains. This demonstrates that the mutants positively selected in the presence of roseoflavin identified via the high-throughput experiment are resistant to roseoflavin

To determine if the RibC mutants displayed a similar resistance to ribocil-C, the same panel of fourteen *S. pneumoniae* strains was assessed for growth across a concentration range of 250nM to 4μM ribocil-C. The IC_{50} for all fourteen strains is 1μM – 2μM ribocil-C **(Fig. 6D, Table S7)**. Although RibC4 and WT^c have a reproducibly higher IC_{50} than unmodified *S. pneumoniae* TIGR4 (316), the magnitude of this change is small and the concentration is still much lower than that in our selection (10μM). Upon careful inspection of the maximum OD_{600}s at moderate concentrations of ribocil-C, the roseoflavin-resistant mutants RoF1, RoF2, RoF3, and the mOFF mutant, showed significantly decreased maximum OD_{600}s at 4μM compared to wildtype 316 (**Fig. 6E, Table S7**). This is consistent with our observation that mutations to the putative terminator display slightly negative fitness values in ribocil-C, and suggests that disrupting the terminator

region potentially makes strains simultaneously more sensitive to ribocil-C at high concentrations, and resistant to roseoflavin.

Based on the results above we hypothesized that mutants positively selected in the presence of 256nM roseoflavin are likely constitutively repressed mutants similar to the mOFF mutant. To confirm this, we examined the gene expression of our panel of mutants in the presence and absence of riboflavin via beta-galactosidase reporter assays similar to previous studies[31]. We find that the RoF mutants, while still displaying some riboflavin dependency, express the RibU-LacZ translational fusion at lower levels than the riboflavin repressed WT^c, regardless of whether riboflavin is present in the medium **(Fig. 6F, Table S8)**. In contrast we see that the RibC mutants all constitutively express the RibU-LacZ fusion at higher levels than the riboflavin repressed WT^c, and often higher than WT^c in the absence of riboflavin **(Fig. 6F, Table S8)**. Thus, in the presence of 256nM roseoflavin, constitutive repression of the transporter is beneficial, while in the presence of ribocil-C, constitutive activation is beneficial.

Roseoflavin resistant mutants reveal a translational repression mechanism of action

Previous 3'-end sequencing data and sequence analysis identified a transcription termination site associated with this riboswitch just upstream of the translational start (Fig. S1)[30]. However, how this terminator enabled FMN-dependent termination was unclear from predicted base-pairing. The most positively selected mutants in the presence of roseoflavin cluster to this region of the transcript, with mutations to positions in the base of the aptamer and the region connecting the aptamer to the ext. term. stem displaying the strongest positive impacts on fitness in roseoflavin **(Fig. 4A, Table S2)**. Based on our validation experiments, these mutations likely all decrease gene expression. However, they also uniformly decrease the stability of the predicted base-pairings, and specifically destabilize the formation of the potential intrinsic terminator stem close to the site of termination.

This finding is surprising, as destabilization of the terminator should result in activation of gene expression rather than repression. This unexpected finding motivated us to look further at the entire transcript sequence to determine if there were any potential interactions with sequence following the putative terminator. Investigation of the downstream region revealed the presence of extensive potential base-pairing between the Shine-Dalgarno sequence and the region of the RNA constituting one side of the intrinsic terminator and the polyU sequence itself (anti-Shine-Dalgarno sequence) **(Fig. 7B).** The mutations positively selected in the presence of roseoflavin also serve to destabilize an alternative pairing that allows the ribosome to access the Shine-Dalgarno sequence **(Fig. 7A)**, ultimately resulting in constitutive repression of gene expression as demonstrated by our panel of mutants **(Fig. 6F).** Consistent with this, mutations to the terminator in and immediately following the aptamer (bases 224-235) confer a higher fitness than those at the base of the terminator closest to the termination site (bases 355-367) as mutation of bases closer to the aptamer only serves to destabilize the structure that exposes the Shine-Dalgarno sequence, while mutation of bases closer to the site of termination may serve to destabilize both potential structures (**Table S2**). This finding strongly suggests that the mechanism of action for this riboswitch is via regulation of translation initiation rather than transcription attenuation, despite the close association of a putative intrinsic terminator proximal to the aptamer.

To determine whether the gene expression changes were transcriptional or translational, we performed qPCR to determine *ribU* transcript levels for our WT^C control strain grown in the absence of riboflavin and in the presence of 256nM roseoflavin, 10μM ribocil-C, or riboflavin. As an internal control, we also assessed levels of *ribD*, which is regulated by the other FMN riboswitch and believed to be transcriptionally regulated. Our results show that levels of *ribU* transcript are not substantially altered in the presence vs. absence of riboflavin, and may very modestly increase in the presence of ribocil-C or roseoflavin **(Fig. 7C, Table S9)**. In contrast, our beta-galactosidase assays show approximately a 6-fold decrease in expression in the presence

vs absence of riboflavin **(Fig. 6F, Table S8)**. Conversely, levels of *ribD* decrease approximately 4-fold in riboflavin, ribocil-C, and roseoflavin compared the absence of riboflavin, indicating that the altered conditions do impact FMN riboswitch functionality within *S. pneumoniae* and are altering transcript levels of the biosynthesis operon **(Fig. 7D, Table S9)**. To further support this mechanism of action, we generated a mutant that disrupts the Shine-Dalgarno/anti-Shine Dalgaro sequence interaction (termed SD1) in our beta-galactosidase reporter strain (**Fig. 7A, B**). We find that SD1 is constitutively active at higher levels than WT^C in both the absence and presence of riboflavin (**Fig. 7E**). These results support the translational regulation of *ribU*, and our proposed mechanism of action based on the beneficial mutations selected by roseoflavin.

DISCUSSION

Our experiments have revealed a host of information that informs how RNA structure responds single point mutations as well as uncovered important insights regarding *S. pneumoniae* response to antibiotics and the mechanism of action for this RNA. By monitoring the growth of a strain library of FMN riboswitch variants in seven different environments, we illuminated entire suites of mutations displaying decreased or increased expression, as well as stumbled upon a strategy increasing the fitness of *S. pneumoniae* in the presence of FMN riboswitch targeting antibiotics.

The simultaneous assessment of many mutations allowed us revise our mechanism of action for this riboswitch. While riboswitch aptamers are typically well-conserved and recognizable via sequence analysis alone, the features that determine how ligand is mediated into genetic regulation or "expression platforms" are not well conserved across species and can be challenging or impossible to deduce from sequence analysis alone[38]. The presence of an intrinsic terminator identified via 3'-end sequencing in close proximity to the aptamer biased our initial hypotheses toward assuming a transcriptional mechanism of action. However, the suite of strongly positively selected mutants in this putative terminator the presence of roseoflavin with repressed gene expression lead us determine that this FMN riboswitch regulates via a translational repression mechanism. This finding underscores the challenges associated with prediction of riboswitch mechanism of action in the absence of detailed experiments, even when coupled to high-throughput 3'-end sequencing data, and points toward using increased caution with interpreting the results of 3'-sequencing data.

In addition to the mechanistic insights gained from assessment of our mutations, the fitness landscapes can also shed light on biological forces acting on non-canonical features of the RNA. The extended P3 stem of this FMN riboswitch extended compared to the canonical FMN aptamer[21]. Although this sequence is common amongst *S. pneumoniae* strains, with very few mutations identified in analysis of this sequence across 300 clinical isolates[30], this feature is not shared in amongst other Streptococcus species suggesting relatively recent introduction[39]. This sequence consists of ~120 additional nucleotides, many of them predicted to pair with each other. Similar such extensions are common in many structured RNAs, and their origins, the evolutionary pressures upon them, and what if any biological function they may have are not well understood[40]. In the presence of roseoflavin, mutations to the P3 stem, particularly those that are disruptive of base-pairing structure, result in significantly decreased fitness **(Fig. 5A, B)**. We see a similar pattern, although with a smaller magnitude change in fitness, in data collected from *in vivo* lung infection **(Fig. 2C)**. Our findings indicate that this deviation from the canonical FMN aptamer is under selection to maintain structure, and during *in vivo* infection this region, which is likely a more recent addition to the sequence, is the one where purifying selection is most apparent.

Beyond specific insights into RNA structure and function, our study also reveals the complex relationship between drugs targeting metabolic control, the native metabolite ligands, and the processes inhibited by such compounds. The mechanism of roseoflavin resistance that results in positive selection of constitutively repressed mutants is relatively straightforward **(Fig. 4, Fig. 6F)**. As a result of the constitutively repressed *ribU* expression, these mutants reduce their

transport of roseoflavin into the cell, and thus are protected from its effects. In contrast, in the presence of ribocil-C, advantageous mutations occur predominantly within the aptamer. These mutations are likely disruptive to aptamer structure, with transversions showing much greater effect on fitness than transitions **(Fig. 3A, C)**, likely impede ligand binding, whether that ligand is ribocil-C, roseoflavin, or FMN, and ultimately result in increased gene-expression **(Fig 6F)** due to destabilization of the ligand-bound aptamer.

The connection between the constitutive gene expression observed and the fitness benefit conferred in the presence of ribocil-C is less clear. This mechanism is further obscured by the lack of explicit ribocil-C resistance for mutants in monoculture **(Fig. 3 and Fig. 6D**). However, the ribocil-C sensitivity of the constitutively repressed roseoflavin resistant mutants **(Fig. 6E, F)** leads us to a potential mechanism for the selective pressure observed. Based on our data, we speculate that constitutively active expression of the transporter enables such mutants to better scavenge the limited FMN or riboflavin available in the environment from surrounding cells that maybe undergoing autolysis. This would give such mutants an advantage in the presence of ribocil-C, which enters the cell via other means, inhibits both FMN biosynthesis and transport, and is a competitive inhibitor with cellular FMN for riboswitch binding[25,26,28]. Thus, any additional FMN present inside the cell negatively impacts drug efficacy due to this competition. Consistent with this, other studies have found that most frequent mechanism of resistance to roseoflavin is constitutive expression of the biosynthetic operon[23]. In our model the increased sensitivity of our constitutively repressed mutants to ribocil-C reflects their reduced ability to perform such scavenging. This model also explains the advantage of such mutants at lower roseoflavin concentrations. At 64nM roseoflavin, we observe a balance where either constitutive expression of *ribU,* to allow scavenging in the presence of the drug, or constitutive repression of *ribU* to prevent drug entry into the cell are positively selected. This situation changes at higher concentration of roseoflavin, where preventing entry of the drug becomes much more positively selected.

Our study also shows the extent of natural selection on the riboswitch under conditions the organism is likely to encounter. We find that single mutations are overwhelmingly neutral in the absence of riboflavin in culture. When riboflavin is added to medium, a similar portrait emerges due the presence of functional riboflavin biosynthesis machinery. However, when the library is assessed within an *in vivo* lung infection model, weak purifying selection is apparent, with many mutations displaying relatively small changes in fitness **(Fig. 2)**. However, in aggregate it is clear that many mutants are mildly deleterious across all regions of the sequence (**Fig. 1B-E**) with transversions always appearing more deleterious than transitions. Due to the substantial defect exhibited by our mOFF truncation mutant compared to the wildtype *S. pneumoniae* in the lung infection model[31], we expected to observe much greater impacts on *in vivo* fitness. However, the positively selected roseoflavin resistant mutants we validated show 5 to 10-fold greater gene-expression than our truncation mOFF mutant, which is only incrementally above background **(Fig 6F)**. The combination of these results suggests that there are no single mutants capable of conferring such a deleterious phenotype to *S. pneumoniae*, and thus the sequence is inherently robust to deleterious mutations, as has been demonstrated repeatedly for natural protein sequences[41–44], but is not necessarily a general trait of all functional sequences[13]**.**

In conclusion, in this study we start to shed light on the selective pressures necessary to create and maintain complex RNA regulators within biological systems. The complexity of the FMN aptamer that we observe conserved across broad swaths of the bacterial kingdom is highly tuned component of an important regulatory pathway, yet *S. pneumoniae* fitness is robust to single mutation changes to its sequence. However, under perturbation with two different antibiotics the adaptive landscapes reveal the extent to which the RNA functionality is sensitive to single mutations that result in either constitutive repression or activation of the RNA, as well as illuminate the details of gene-regulation and mechanisms governing antibiotic sensitivity.

MATERIALS AND METHODS

Doped oligo library construction

Six doped oligos, with 5' and 3' constant regions, were ordered from IDT to create a SP_0488 FMN riboswitch mutant library that would be mutated at each nucleotide. The second strand of each doped oligo was synthesized to make the DNA double stranded by combining tris-HCl (pH 8.0), $MgCl_2$, primer, and doped oligo library for each of the six doped oligo libraries. Each doped oligo was PCR amplified, and purified using the Zymo Clean & Concentrator kit. 3' and 5' homology arms were also amplified via PCR and purified using the Zymo Clean & Concentrator kit. The three regions for each doped oligo DNA library were then assembled in two Gibson Assemblies: the first with the doped region and the 3' homology arm, and the second with the 5' homology arm. The six doped oligo libraries were then mixed together in an equimolar pool and transformed into *S. pneumoniae*[45]. Primers used in this study are listed in **Table S10**.

Amplicon sequencing culture conditions and sample collection

A doped oligo library starter culture was washed three times and used to inoculate 10mL of semi-defined minimal media (SDMM), and grown until mid-log phase. This culture was then washed three times, and used to inoculate three replicates of each experimental condition at an OD_{600} of 0.001. The experimental conditions were in the presence of riboflavin, in the absence of riboflavin, 256nM roseoflavin, and 10μM ribocil-C, grown in chemically defined media (CDM) without riboflavin. After eight hours of growth, each culture was spun down, the supernatant pipetted off, and the media replaced with fresh media and ligand. Each culture was grown until it reached an OD_{600} of ~1, or until it had been 12.5 hours. At this point, the cultures were pelleted, the supernatants pipetted off, and the pellets stored at -80°C to use for amplicon sequencing preparation.

In vivo FMN riboswitch mutant strain library challenge

Fitness experiments were performed by challenging mice with the FMN riboswitch mutant strain library. The library was initially grown to mid-log phase, at which point it was washed and resuspended in 1x PBS. Mice were inoculated with $2.5*10^7$ cells via intranasal injection, and sacrificed 24 hours later, where the lungs were removed and homogenized in 1x PBS as described previously[31]. 10μl of homogenized lungs were used for serial dilutions to estimate the bacterial burden. The remainder of the homogenized lungs were plated on chloramphenicol blood agar plates, grown overnight, and the resulting colonies collected for amplicon sequencing.

Amplicon sequencing preparation

The cell pellets were taken out of the -80°C freezer, and their gDNA was isolated. The quantity was determined using a nanodrop, and each sample was diluted to 50ng/μl. 800ng of gDNA was then used in an initial 400μl 5 cycle PCR, spread over 8 PCR tubes, used to add the unique molecular identifiers (UMIs). The products were column purified using the Zymo Clean & Concentrator kit, and the quantity was determined using the nanodrop. 400ng was then used in a 400μl 30-cycle PCR, spread out over 8 PCR tubes, for amplification of the libraries. These products were then purified using the Zymo Gel Purification kit, and quantified using the Qubit. Samples were mixed together to create an equimolar pool. The pool was quantified using the Qubit, and run on either the NextSeq 2000, using the P1 2x300bp kit, or the MiSeq using the v3 reagent kit. Primers used in this study are listed in **Table S10**.

Identification of antibiotic resistant mutants

After receiving the Fastq files from the NextSeq 2000 or MiSeq, the paired end reads were initially merged using Vsearch[46]. PCR bias was then corrected for by using AmpUMI twice, the first time using the UMIs attached on the 5' end, the second time using the UMIs attached on the 3' end[47].

The deep mutational scanning software DiMSum was then used to identify mutants that are either beneficial, detrimental, or neutral in each condition[32]. DiMSum calculates fitness for each variant compared to the wildtype sequence in each replicate as the natural logarithm of the ratio between sequencing counts in a replicate's output and input samples relative the wildtype sequence according to equation 1 below and subsequently produces a weighted score across the replicates for each condition.

$$fitness_i = \ln\left(\frac{N_i^{output}}{N_i^{input}} \times \frac{N_{wt}^{input}}{N_{wt}^{output}}\right)$$

FMN RNA mutant strain construction

Oligos ordered from IDT containing the targeted mutations with additional overhangs for Gibson Assembly were used in an initial 30-cycle PCR to add the mutant and overhang to both halves of the DNA containing the 5' and 3' homology arms, the chloramphenicol resistant cassette, and the FMN riboswitch. The PCR products were purified using the Zymo Clean & Concentrator kit, and were used for Gibson Assembly. The Gibson Assembly products were then diluted 1:4, and used in a 30-cycle PCR for amplification. The PCR products were purified using the Zymo Clean & Concentrator kit and quantified using the nanodrop. 50ng DNA was used to transform into wildtype, 316 (TIGR4)[45]. Transformants were plated on chloramphenicol blood plates, and 12 colonies from each mutant were picked, grown for six hours, and analyzed via colony PCR. Glycerol stocks were made for each mutant. PCR products amplifying the genome region of interest from promising mutants were sent to Eton Biosciences for Sanger Sequencing, and one from each mutant was selected to make starters with and use for experiments. Primers used in this study are listed in **Table S10**.

FMN RNA mutant growth assays

One starter per strain was washed three times in 1xPBS, and then used to inoculate 7.5ml of chemically defined media (CDM) without riboflavin. Cultures were grown for ~2 hours. The OD_{600} was determined, and cultures were diluted to an OD_{600} of 0.015 in 7ml CDM-riboflavin. Culture and ligand were mixed to obtain the desired ligand concentration, and 200µl of each condition was plated into three wells of a 96-well plate. The plates were incubated in a BioTek BioSpa 8, where their OD_{600} was read every 30 minutes. IC_{50}s were determined by identifying the antibiotic concentration at which growth is inhibited >50% compared to growth in the -riboflavin condition at 8.5 hours.

LacZ reporter strain construction and beta-galactosidase assays

Genomic DNA from strains carrying the RibC, RoF, and NC mutant strains was PCR amplified with primers SP0647+cat_F and lacZ_SP0488_R to generate a product containing the chloramphenicol resistance cassette and the riboswitch through the first three codons of *ribU* under the control of its native promoter. For each mutant these fragments were combined via Gibson Assembly with flanking sequences (amplified by primers SP0647_F/CAT+SP0647_R, and SP0488+lacZ_F/lacZ_R respectively) that correspond to a soluble *lacZ* sequence, and SP_0647, to result in a translational fusion between *ribU* and *lacZ*. These full-length products were PCR amplified and transformed into a previously reported strain wherein the endogenous beta-galactosidase in *S. pneumoniae* TIGR4 (SP_0648) was replaced with a portion of a soluble *lacZ* and a kanamycin resistance marker[31]. Transformants were screened for correct integration of the Chl cassette via PCR and presence of the mutation confirmed via Sanger sequencing of the PCR product. Primers used in this study are listed in **Table S10**.

The resulting reporter strains were grown in CDM with and without riboflavin to an OD_{600} of ~0.5 and harvested. Cells were resuspended in Z buffer (50 mM Na2HPO4, 40 mM NaH2PO4, 10 mM KCl, 1 mM MgSO4, 50 mM 2-mercaptoethanol) and beta-galactosidase activity assay was performed as described previously[48].

qPCR analysis

RNA was isolated from cultures, with a starting OD_{600} of 0.1, grown for 3 hours under -riboflavin, +riboflavin, 256nM roseoflavin, and 10μM ribocil-C conditions, using the Qiagen RNeasy kit (Qiagen). The resulting RNA was treated with DNase (TURBO DNA Free Kit (Ambion)). The DNase treated RNA was used to generate cDNA with iScript reverse transcriptase supermix for RT-qPCR (BioRad). Quantitative PCR was performed using a Bio-Rad MyiQ. Each sample was normalized against the 50S ribosomal gene, SP_2204 and were measured in technical triplicates. No-reverse transcriptase controls were included for all samples. Primers used in this study are listed in **Table S10**.

References

1. Panchal, V. & Brenk, R. Riboswitches as Drug Targets for Antibiotics. *Antibiotics (Basel)* **10**, (2021).
2. Ellinger, E. *et al.* Riboswitches as therapeutic targets: promise of a new era of antibiotics. *Expert Opin Ther Targets* **27**, 433–445 (2023).
3. Giarimoglou, N. *et al.* A Riboswitch-Driven Era of New Antibacterials. *Antibiotics (Basel)* **11**, 1243 (2022).
4. Hong, W., Zeng, J. & Xie, J. Antibiotic drugs targeting bacterial RNAs. *Acta Pharm Sin B* **4**, 258–265 (2014).
5. Rekand, I. H. & Brenk, R. Ligand design for riboswitches, an emerging target class for novel antibiotics. *Future Med Chem* **9**, 1649–1663 (2017).
6. Mehdizadeh Aghdam, E., Hejazi, M. S. & Barzegar, A. Riboswitches: From living biosensors to novel targets of antibiotics. *Gene* **592**, 244–259 (2016).
7. Machtel, P., Bąkowska-Żywicka, K. & Żywicki, M. Emerging applications of riboswitches - from antibacterial targets to molecular tools. *J Appl Genet* **57**, 531–541 (2016).
8. Deigan, K. E. & Ferré-D'Amaré, A. R. Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs. *Acc Chem Res* **44**, 1329–1338 (2011).
9. Chauvier, A. & Walter, N. G. Regulation of bacterial gene expression by non-coding RNA: It is all about time! *Cell Chem Biol* **31**, 71–85 (2024).
10. Chauvier, A. & Walter, N. G. Beyond ligand binding: Single molecule observation reveals how riboswitches integrate multiple signals to balance bacterial gene regulation. *Curr Opin Struct Biol* **88**, 102893 (2024).
11. Andreasson, J. O. L., Savinov, A., Block, S. M. & Greenleaf, W. J. Comprehensive sequence-to-function mapping of cofactor-dependent RNA catalysis in the glmS ribozyme. *Nat Commun* **11**, 1663 (2020).
12. Li, C., Qian, W., Maclean, C. J. & Zhang, J. The fitness landscape of a tRNA gene. *Science* **352**, 837–840 (2016).
13. Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. & Chen, I. A. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc Natl Acad Sci U S A* **110**, 14984–14989 (2013).
14. Pressman, A. D. *et al.* Mapping a Systematic Ribozyme Fitness Landscape Reveals a Frustrated Evolutionary Network for Self-Aminoacylating RNA. *J Am Chem Soc* **141**, 6213–6223 (2019).

15. Bendixsen, D. P., Collet, J., Østman, B. & Hayden, E. J. Genotype network intersections promote evolutionary innovation. *PLoS Biol* **17**, e3000300 (2019).
16. Peri, G., Gibard, C., Shults, N. H., Crossin, K. & Hayden, E. J. Dynamic RNA Fitness Landscapes of a Group I Ribozyme during Changes to the Experimental Environment. *Mol Biol Evol* **39**, msab373 (2022).
17. Hayden, E. J. & Wagner, A. Environmental change exposes beneficial epistatic interactions in a catalytic RNA. *Proc Biol Sci* **279**, 3418–3425 (2012).
18. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A* **99**, 15908–15913 (2002).
19. GBD 2016 Lower Respiratory Infections Collaborators. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis* **18**, 1191–1210 (2018).
20. CDC. Antibiotic Resistance Threats in the United States, 2013. 1–114 (2013).
21. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23**, 995–1011 (2017).
22. Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nat Biotechnol* **24**, 1558–1564 (2006).
23. Lee, E. R., Blount, K. F. & Breaker, R. R. Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol* **6**, 187–194 (2009).
24. Blount, K. F. *et al.* Novel riboswitch-binding flavin analog that protects mice against Clostridium difficile infection without inhibiting cecal flora. *Antimicrob Agents Chemother* **59**, 5736–5746 (2015).
25. Howe, J. A. *et al.* Selective small-molecule inhibition of an RNA structural element. *Nature* **526**, 672–677 (2015).
26. Howe, J. A. *et al.* Atomic resolution mechanistic studies of ribocil: A highly selective unnatural ligand mimic of the E. coli FMN riboswitch. *RNA Biol* **13**, 946–954 (2016).
27. Ott, E., Stolz, J., Lehmann, M. & Mack, M. The RFN riboswitch of Bacillus subtilis is a target for the antibiotic roseoflavin produced by Streptomyces davawensis. *RNA Biol* **6**, 276–280 (2009).
28. Wang, H. *et al.* Dual-Targeting Small-Molecule Inhibitors of the Staphylococcus aureus FMN Riboswitch Disrupt Riboflavin Homeostasis in an Infectious Setting. *Cell Chem Biol* **24**, 576-588.e6 (2017).
29. Pedrolli, D. B. *et al.* A highly specialized flavin mononucleotide riboswitch responds differently to similar ligands and confers roseoflavin resistance to Streptomyces davawensis. *Nucleic Acids Research* **40**, 8662–8673 (2012).
30. Warrier, I. *et al.* The Transcriptional landscape of Streptococcus pneumoniae TIGR4 reveals a complex operon architecture and abundant riboregulation critical for growth and virulence. *PLoS Pathog* **14**, e1007461 (2018).
31. Warrier, I. *et al.* RNA cis-regulators are important for Streptococcus pneumoniae in vivo success. *PLoS Genet* **20**, e1011188 (2024).
32. Faure, A. J., Schmiedel, J. M., Baeza-Centurion, P. & Lehner, B. DiMSum: an error model and pipeline for analyzing deep mutational scanning data and diagnosing common experimental pathologies. *Genome Biology* **21**, 207 (2020).
33. van Opijnen, T. & Camilli, A. A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res* **22**, 2541–2551 (2012).
34. Serganov, A., Huang, L. & Patel, D. J. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* **458**, 233–237 (2009).

35. Mansjö, M. & Johansson, J. The riboflavin analog roseoflavin targets an FMN-riboswitch and blocks Listeria monocytogenes growth, but also stimulates virulence gene-expression and infection. *RNA Biol* **8**, 674–680 (2011).
36. Gutiérrez-Preciado, A. *et al.* Extensive Identification of Bacterial Riboflavin Transporters and Their Distribution across Bacterial Species. *PLOS ONE* **10**, e0126124 (2015).
37. Pedrolli, D. *et al.* The ribB FMN riboswitch from Escherichia coli operates at the transcriptional and translational level and regulates riboflavin biosynthesis. *FEBS J* **282**, 3230–3242 (2015).
38. Olenginski, L. T., Spradlin, S. F. & Batey, R. T. Flipping the script: Understanding riboswitches from an alternative perspective. *J Biol Chem* **300**, 105730 (2024).
39. Ontiveros-Palacios, N. *et al.* Rfam 15: RNA families database in 2025. *Nucleic Acids Res* **53**, D258–D267 (2025).
40. Srivastava, A., Cai, L., Mrázek, J. & Malmberg, R. L. Mutational patterns in RNA secondary structure evolution examined in three RNA families. *PLoS One* **6**, e20484 (2011).
41. Zheng, J., Guo, N. & Wagner, A. Selection enhances protein evolvability by increasing mutational robustness and foldability. *Science* **370**, eabb5962 (2020).
42. Guo, H. H., Choe, J. & Loeb, L. A. Protein tolerance to random amino acid change. *Proc Natl Acad Sci U S A* **101**, 9205–9210 (2004).
43. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* **19**, 596–604 (2009).
44. Wagner, A. Mutational robustness accelerates the origin of novel RNA phenotypes through phenotypic plasticity. *Biophys J* **106**, 955–965 (2014).
45. Bricker, A. L. & Camilli, A. Transformation of a type 4 encapsulated strain of Streptococcus pneumoniae. *FEMS Microbiology Letters* **172**, 131–135 (1999).
46. Mathon, L. *et al.* Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. *Molecular Ecology Resources* **21**, 2565–2579 (2021).
47. Clement, K., Farouni, R., Bauer, D. E. & Pinello, L. AmpUMI: design and analysis of unique molecular identifiers for deep amplicon sequencing. *Bioinformatics* **34**, i202–i210 (2018).
48. Miller, J. H. *A Short Course in Bacterial Genetics : A Laboratory Manual and Handbook for Escherichia Coli and Related Bacteria*. (Plainview, N.Y. : Cold Spring Harbor Laboratory Press, 1992).
49. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res* **36**, W70-74 (2008).

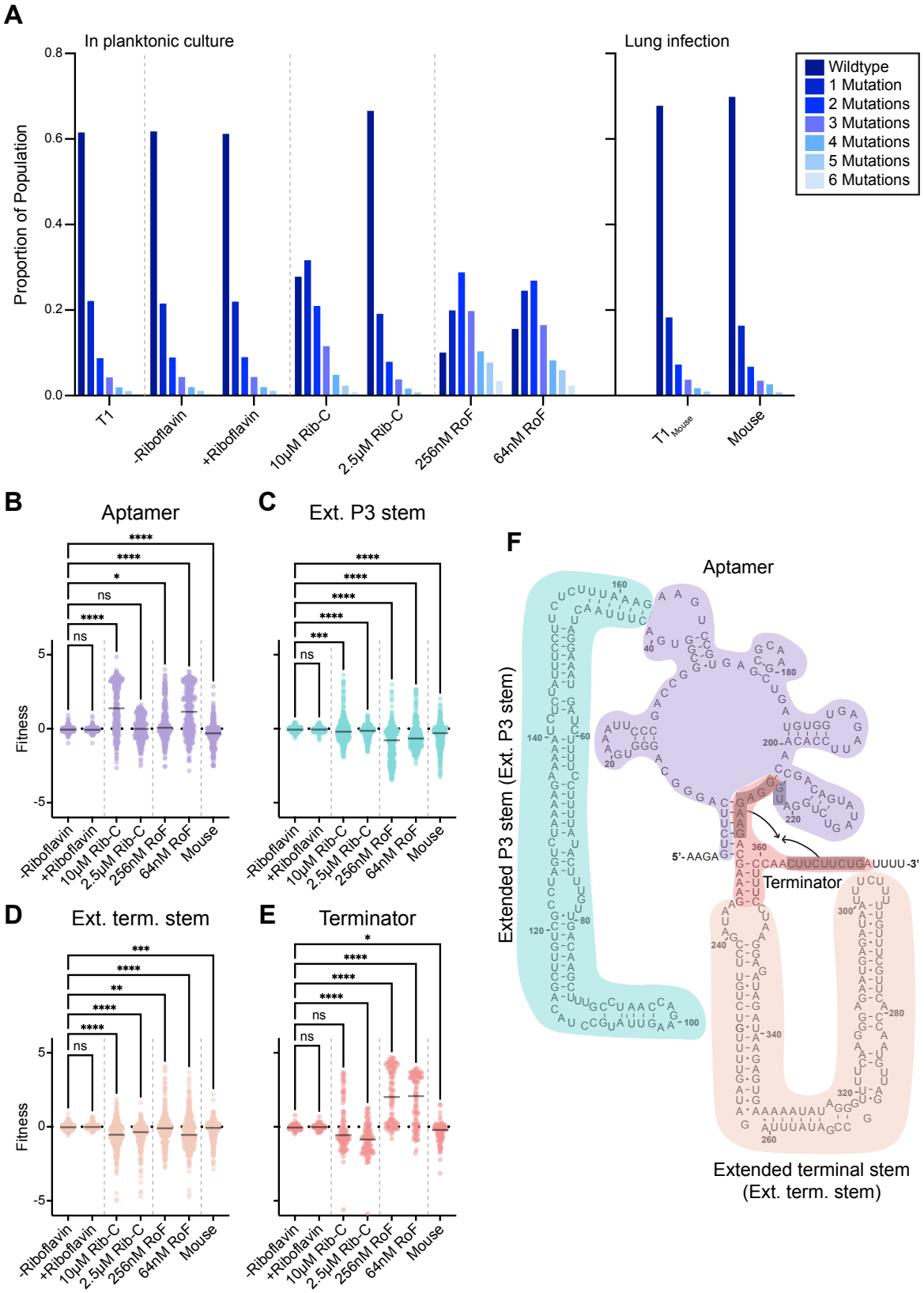

Figure 1: The mutant strain library is under selective pressure when challenged with antibiotics or within lung infection. A) The proportion of processed reads that have 0 (wildtype

sequence) to 6 mutations for each condition and their respective T1. **B-E)** Fitness of mutations to different regions of the riboswitch color coded with diagram **F** relative to the wildtype sequence in seven different environments: chemically defined media (CDM) with no riboflavin (-Riboflavin), CDM+ 140 ng/mL riboflavin (+Riboflavin), CDM+ 10μM ribocil-C (10μM Rib-C), CDM+ 2.5μM ribocil-C (2.5μM Rib-C), CDM+ 256nM roseoflavin (256nM RoF), CDM+ 64nM roseoflavin (64nM RoF), and within a mouse lung infection (Mouse). For each region, the fitness of the population of single mutants is compared to the fitness in the -Riboflavin condition using Kruskal-Wallis test with Dunn's test for multiple comparisons (adj. $p>0.05$ (ns), adj. $p<0.05$ (*), adj. $p<0.01$ (**), adj. $p<0.001$ (***), $p<0.0001$ (****)). Population medians and exact p-values for each comparison on **Table S3. F)** The SP_0488 FMN riboswitch secondary structure colored according to region. Secondary structure derived from alignment to FMN aptamer consensus structure[39], and folding extended stems with the RNAfold webserver[49].

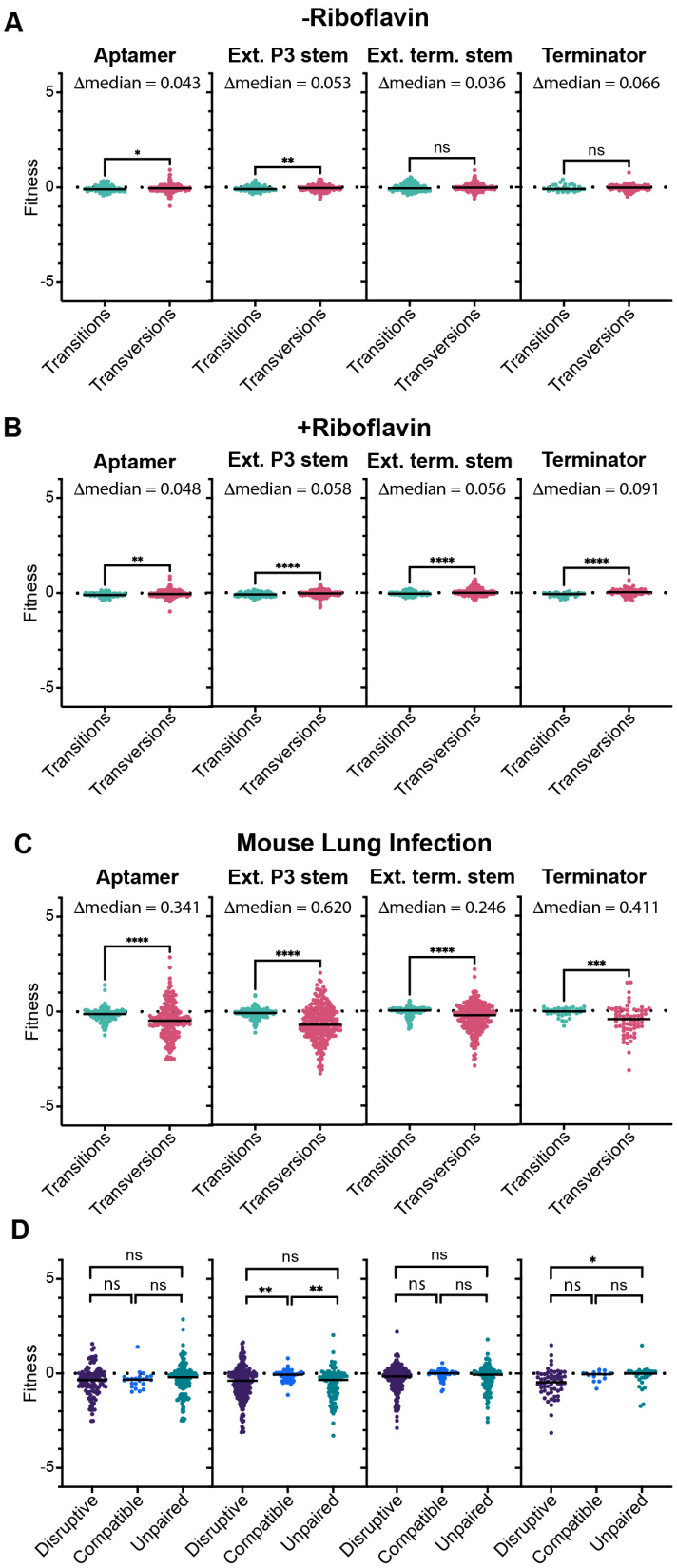

Figure 2: The FMN riboswitch is not under selection in culture but is under weak purifying selection during mouse lung infection. A) Fitness of transition versus transversion mutations

by region of the riboswitch in the absence of riboflavin show only small magnitude differences between population medians. Populations compared using Mann-Whitney test ($p>0.05$ (ns), $p< 0.05$ (*), $p<0.01$ (**), $p<0.001$ (***), $p<0.0001$ (****)). Exact p-values for each comparison on Table S3**. B)** Fitness of transition versus transversion mutations by region in the presence of riboflavin show only small magnitude differences between population medians. Statistical significance determined as in A. **C)** Fitness of transition versus transversion mutations by region of the riboswitch selected via mouse lung infection shows that transversions are negatively selected compared to transitions in all regions of the RNA. Statistics as in A. **D**) Fitness of mutations in mouse lung infection in the ext. P3 stem that are disruptive to predicted secondary structure or unpaired is significantly lower than that of compatible mutations. Fitness of mutations that disrupt predicted base-pairing in the terminator is significantly lower than that for positions predicted to be unpaired. Populations compared using a Kruskal-Wallis test with Dunn's test for multiple comparisons ($p>0.05$ (ns), $p< 0.05$ (*), $p<0.01$ (**), $p<0.001$ (***), $p<0.0001$ (****)). Population medians and exact p-values for each comparison on Table S3.

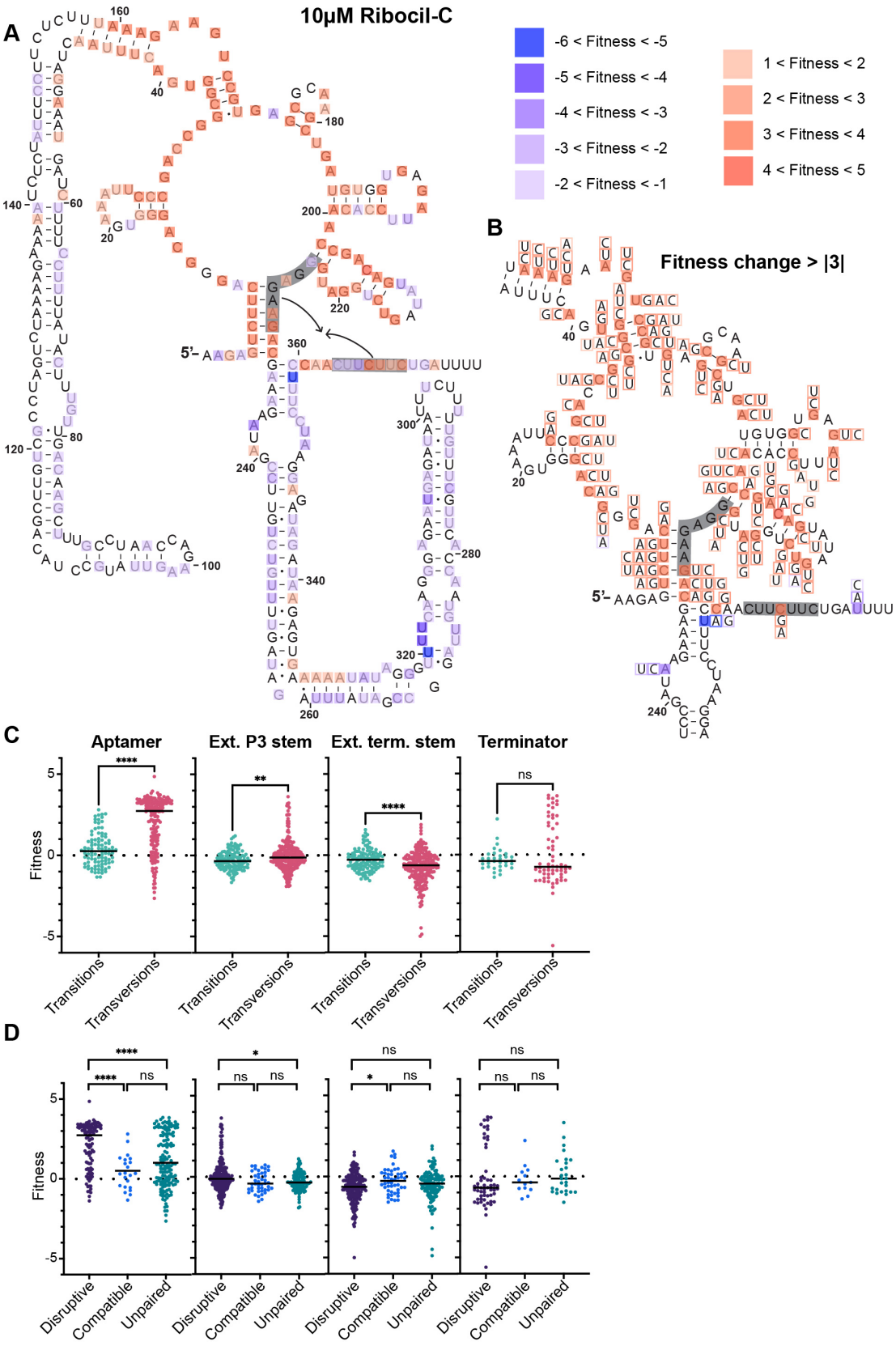

Figure 3: Mutations to the aptamer are beneficial in the presence of ribocil-C. A) Heatmap of mutation fitness in ribocil-C relative to wildtype mapped to riboswitch secondary structure.

Colors displayed correspond to the fitness of the mutation with largest magnitude change from wild-type. Blue: negative fitness, red: positive fitness. **B)** Inset of aptamer and proximal areas highlighting only positions with fitness changes > |3|, all mutations with a fitness change > |1| for each position are boxed in the color corresponding to the fitness change. **C)** Fitness of transition versus transversion mutations by region of the riboswitch in the presence of ribocil-C shows that transversions are positively selected compared to transitions in the aptamer and ext. P3 stem, and negatively selected in the ext. term. stem. Statistics as on Fig. 2A. **D)** Fitness of mutations in the aptamer disruptive of predicted base-pairing (see Fig. 1F for predicted secondary structure) are significantly more positive compared to those of compatible mutations or mutations to unpaired bases. Statistics and analysis as in Fig. 2D.

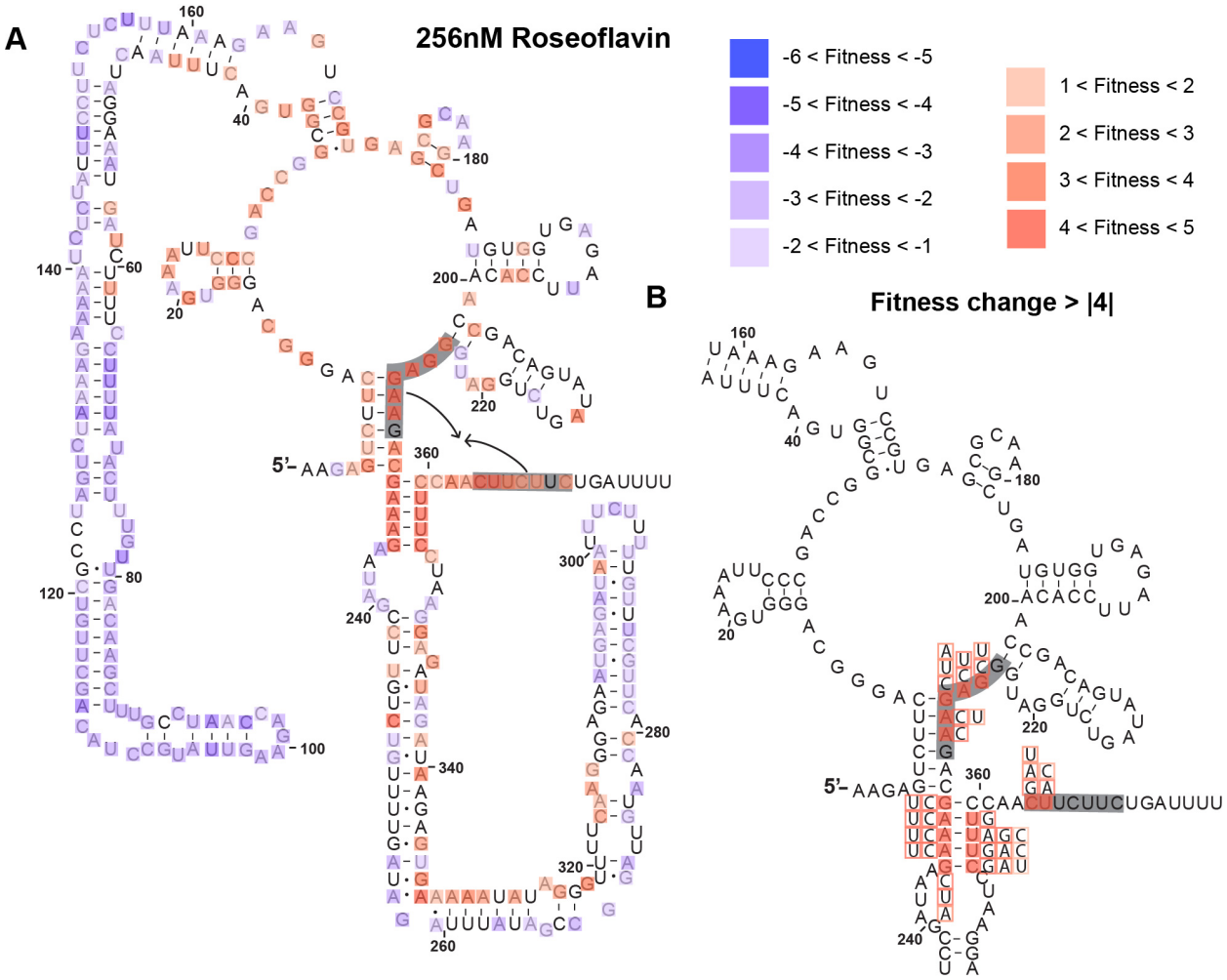

Figure 4: Mutations to the terminator region display the strongest benefit in the presence of roseoflavin. A) Heatmap of mutation fitness in roseoflavin relative to wildtype mapped to riboswitch secondary structure. Colors displayed correspond to the fitness of the mutation with largest magnitude change from wild-type. Blue: negative fitness, red: positive fitness. **B)** Inset of aptamer and proximal areas highlighting only positions with fitness changes > |4|, all mutations with a fitness change > |1| for each position are boxed in the color corresponding to the fitness change.

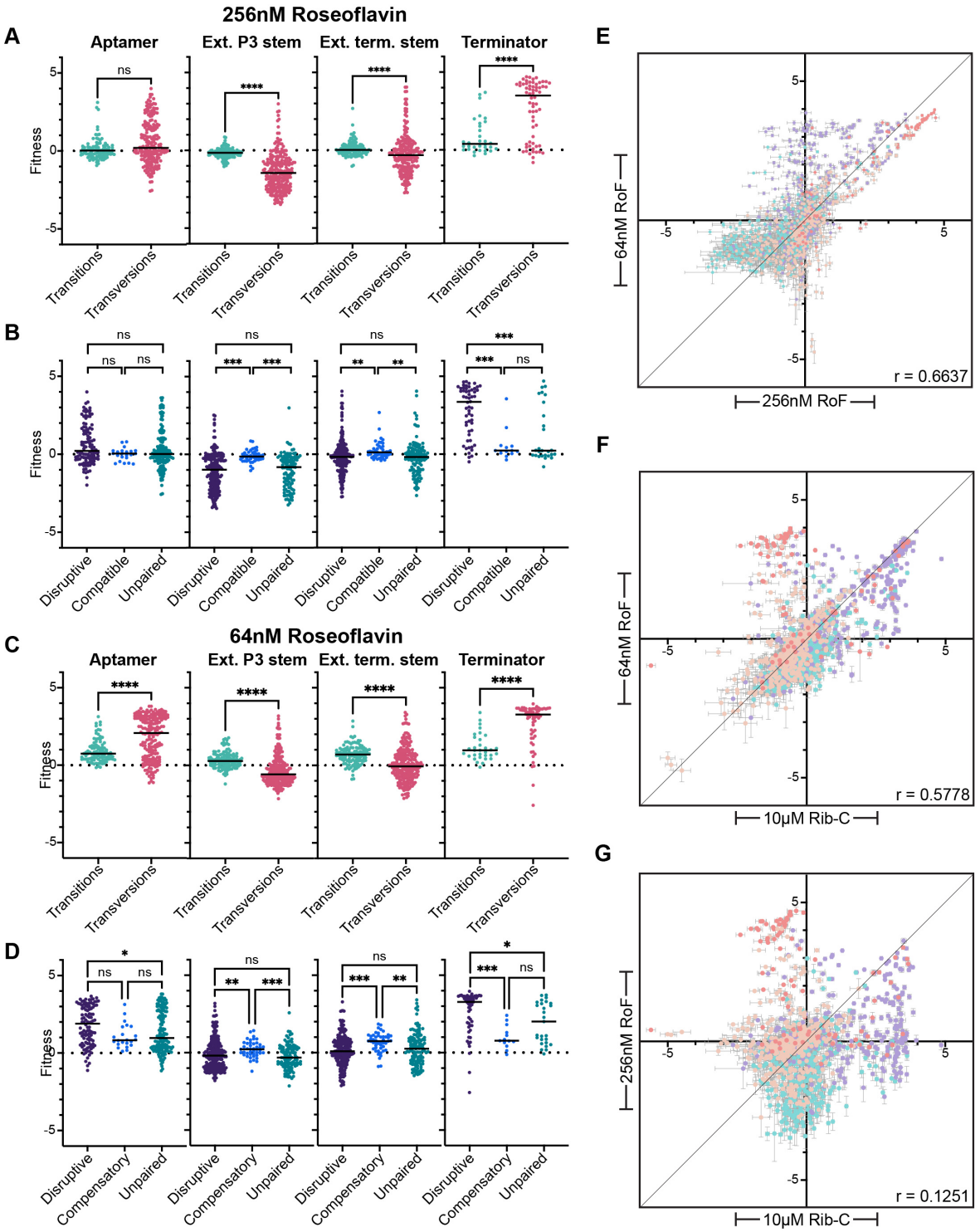

Figure 5: Different concentrations of roseoflavin result in distinct fitness landscapes. A) Fitness of transition versus transversion mutations by region of the riboswitch in the presence of 256nM roseoflavin shows that transversions are negatively selected compared to transitions in both extended stems, and positively selected in the terminator region. Statistics as on Fig. 2A, D) Fitness of mutations in the terminator region disruptive of predicted base-pairing (see Fig. 1F for predicted secondary structure) are significantly more positive compared to those of compatible mutations or mutations to unpaired bases. Mutations disruptive to base-pairing or to unpaired

bases display lower fitness compared to compatible mutations in both extended stems. Statistics and analysis as in Fig. 2D. **C)** Fitness of transition versus transversion mutations by region of the riboswitch in the presence of 64nM roseoflavin shows that transversions are negatively selected compared to transitions in both extended stems, and positively selected in the aptamer terminator regions. Statistics as on Fig. 2A. **D)** Fitness of mutations in the terminator region disruptive of predicted base-pairing (see Fig. 1F for predicted secondary structure) are significantly more positive compared to those of compatible mutations or mutations to unpaired bases. Fitness of mutations in the aptamer region disruptive of base-pairing are significantly more positive compared to mutations at unpaired bases. Compensatory mutations display higher fitness compared to disruptive mutations and mutations at unpaired regions in both extended stems. Statistics and analysis as in Fig. 2D. **E)** Fitness values obtained from two distinct roseoflavin conditions, plotted on x-y axis with calculated error. There is a strong correlation between values obtained (Spearman's correlation co-efficient r shown on graph, $p < 0.0001$, Table S4), with lower error on the positively selected values. Colors correspond to regions of the riboswitch depicted in Fig. 1F. **F)** Fitness values obtained from 64nM roseoflavin and 10μM ribocil-C conditions and **G)** Fitness values obtained from 256nM roseoflavin and 10μM ribocil-C conditions plotted on x-y axis with reported error. There is a significant correlation between both values obtained. Statistics and analysis as in E.

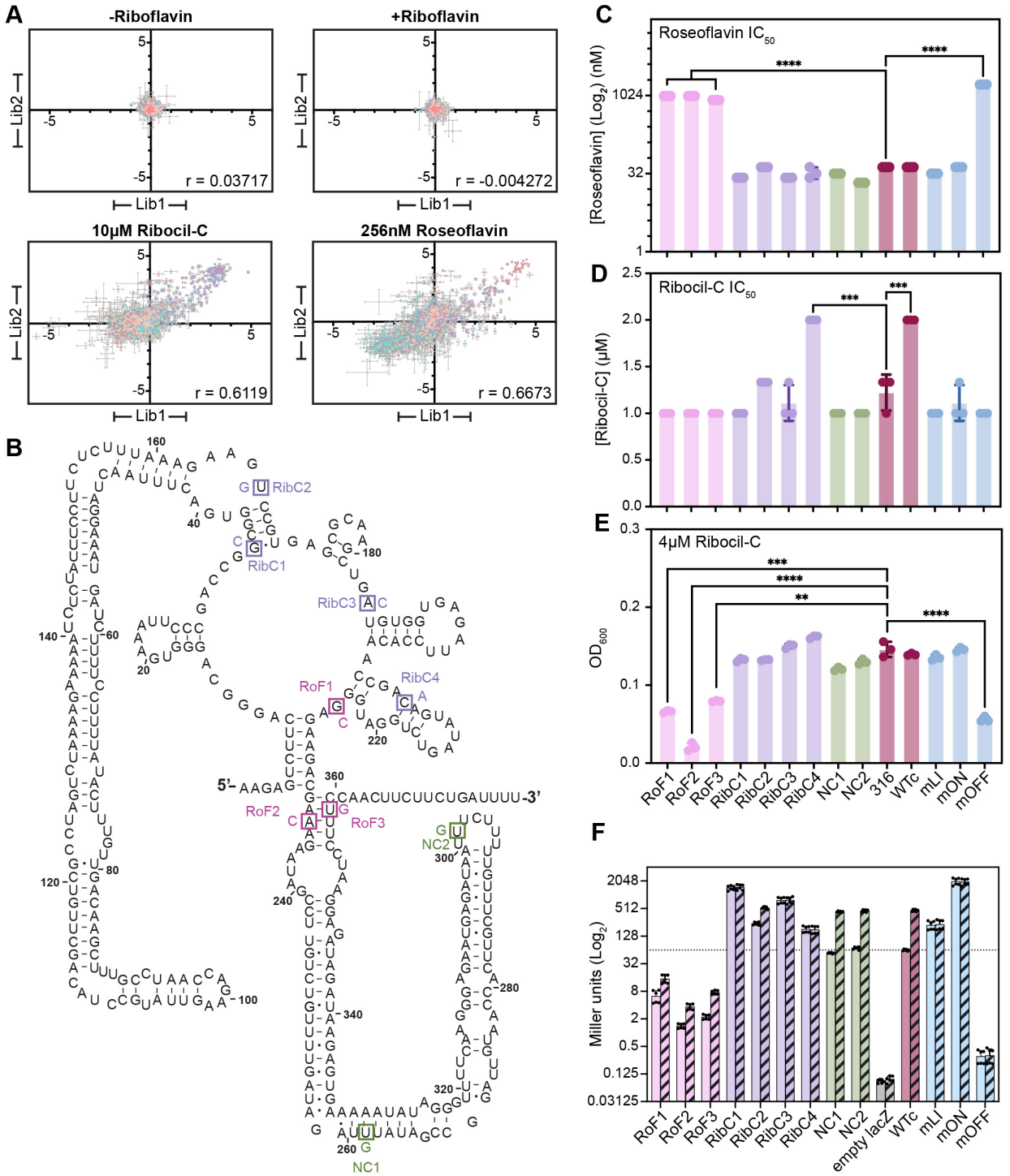

Figure 6: A biological replicate and assessment of selected mutants validates fitness values. A) Fitness values obtained from two distinct libraries, Lib1 and Lib2, plotted on x-y axis with reported error. In + and -riboflavin conditions, most mutations display fitness <|1|, and there is no significant correlation between values in the replicates. For 10μM ribocil-C and 256nM roseoflavin there is strong correlation between values obtained for Lib1 and Lib2. Statistics and analysis as in Fig. 5E. **B)** Roseoflavin selected (RoF), Ribocil-C selected (RibC), and negative

control (NC) mutations shown on the secondary structure of the SP_0488 FMN riboswitch. **C)** IC_{50} of roseoflavin for RoF, RibC, NC, and additional control strains shows that RoF mutants are resistant to roseoflavin. Pink: RoF selected mutants, purple: RibC selected mutants, green: negative controls, red: parental TIGR4 or WT^C control containing only Chl cassette, blue: previously published SP_0488 FMN riboswitch mutants[31]. Each point represents mean of technical triplicates. Data analyzed via a nested one-way ANOVA, compared to wildtype 316 (unmodified strain). All strains without significance indicated are not significant (ns) when compared to wildtype 316, ($p>0.05$ (ns), $p< 0.05$ (*), $p<0.01$ (**), $p<0.001$ (***), $p<0.0001$ (****)) (See Table S7). **D)** IC_{50} of ribocil-C shows that positively selected strains are ultimately not resistant. Replicates and data analysis as in part C above. **E)** Maximum OD_{600} for strains when grown with 4μM ribocil-C for 8.5 hours shows that RoF mutants, and mOFF are more susceptible than the unmodified strain to ribocil-C. Replicates and data analysis as in part C above. **F)** ß-galactosidase assays measure gene expression in the presence (solid bars) and absence (stripped bars) of riboflavin. Error bars represent the standard error of the mean (n=6). Empty lacZ (gray) indicates background levels of ß-galactosidase expression. In the presence of riboflavin, RoF mutations display lower gene expression compared to WT (dashed line at WT^C + riboflavin) and RibC mutations display higher gene expression. NC mutants do not display behavior statistically distinct from WT^C, but all other mutants display gene expression in both + and -riboflavin that is statistically different from WT^C via a Brown-Forsythe ANOVA followed by Dunnett's test of multiple comparisons (Table S8).

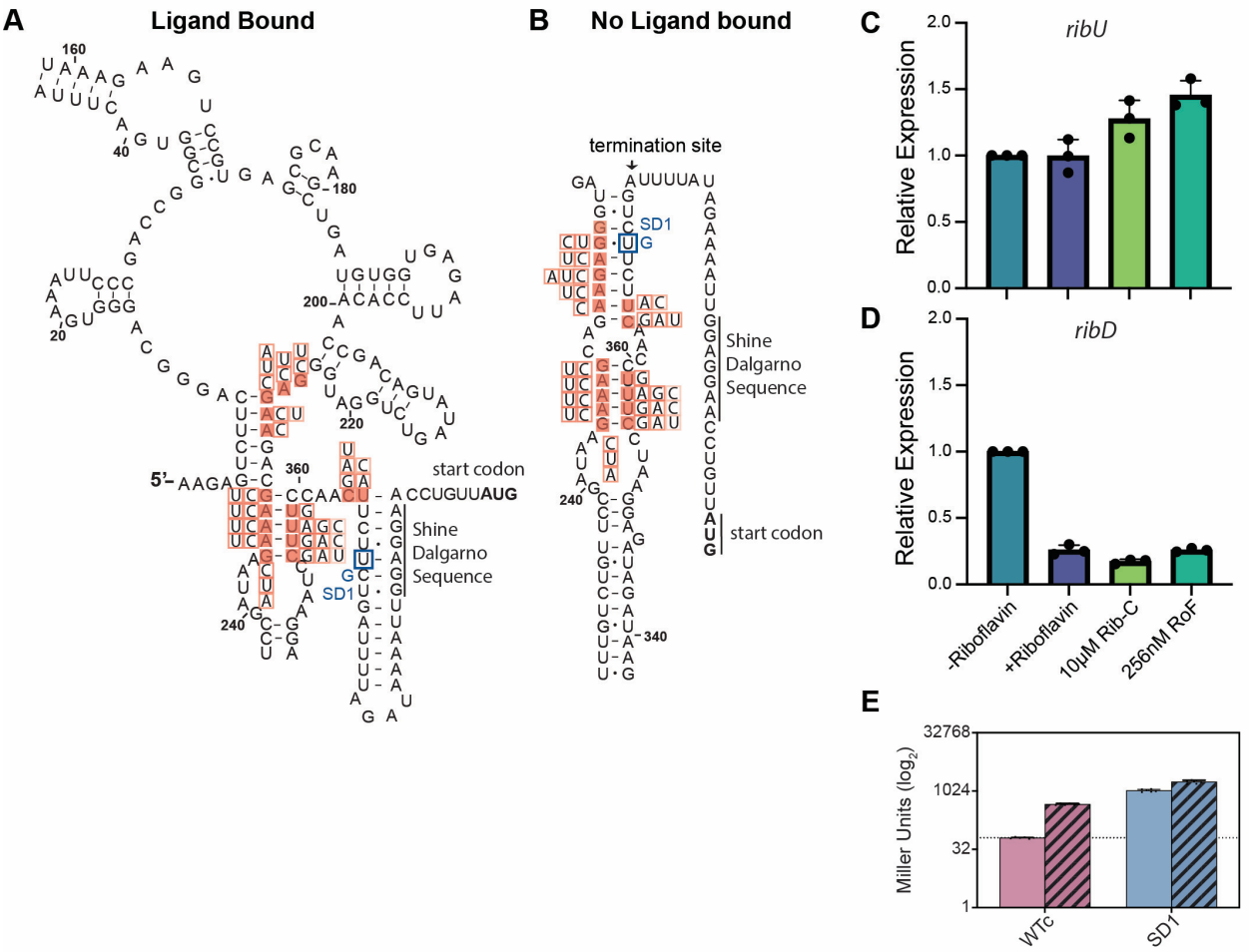

Figure 7: Constitutively repressed roseoflavin resistant mutants suggest a translation inhibition mechanism. A) Proposed structure for the ligand-bound form of the aptamer and downstream sequence including the Shine-Dalgarno and start codon indicated. Positions where a mutation results in fitness >4 in roseoflavin that likely display constitutively repressed expression are indicated. Shine-Dalgarno mutant (SD1) shown in blue. **B)** Proposed structure of the unbound form of the RNA where the prospective terminator forms, but the Shine-Dalgarno sequence is released. Positions where a mutation results in fitness >4 in roseoflavin that likely display constitutively repressed expression are indicated. All such mutations destabilize putative terminator stem, resulting in increased formation of the competing structure. Shine-Dalgarno mutant (SD1) shown in blue. **C)** qPCR analysis of *ribU* transcript in CDM -riboflavin, +riboflavin, +ribocil-C, and +roseoflavin shows minimal change in response to riboflavin and a slight increase in the presence of roseoflavin and ribocil-C. **D)** qPCR analysis of the *ribD* transcript, also regulated via an FMN riboswitch shows the anticipated repression of transcript in the presence of ligands for transcriptional regulation. **E)** ß-galactosidase assays measure gene expression in the presence (solid bars) and absence (stripped bars) of riboflavin. Error bars represent the standard error of the mean (n=6). WT^C the same as in Fig. 6F. In the presence of riboflavin, the SD1 mutant displays a higher gene expression compared to WT^C (dashed line at WT^C + riboflavin).