bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RBP-RNA interactions in the 5' UTR lead to structural changes that alter translation

Noa Katz[1], Roni Cohen[1], Oz Solomon[1,3], Beate Kaufmann[1], Orna Atar[1], Zohar Yakhini[2,3], Sarah Goldberg[1], and Roee Amit[1,4]*

[1] Department of Biotechnology and Food Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel.
[2] Department of Computer Science, Technion - Israel Institute of Technology, Haifa 32000, Israel.
[3] School of Computer Science, Interdisciplinary Center, Herzeliya 46150, Israel.
[4] Russell Berrie Nanotechnology Institute, Technion - Israel Institute of Technology, Haifa 32000, Israel.

* To whom correspondence should be addressed. Tel: 972-77-8871894; Fax: 972-4-8293399; Email: roeeamit@technion.ac.il

ABSTRACT

We study translational regulation by a 5' UTR sequence encoding the binding site of an RNA-binding protein (RBP) in bacteria, using a reporter assay and Selective 2′-hydroxyl acylation analysed by primer extension sequencing (SHAPE-Seq). We tested constructs containing a single hairpin, based on the binding sites of the coat RBPs of bacteriophages GA, MS2, PP7, and Qβ, positioned in the 5' UTR of a reporter gene. With specifically-bound RBP present, either weak repression or up-regulation is observed, depending on the binding site and its flanking sequence. SHAPE-Seq data for a representative construct exhibiting up-regulation, indicates a partially-folded hairpin and non-reactive upstream and downstream flanking region, which we attribute to intermediate structures that apparently blocks translation. RBP binding stabilizes the fully-folded hairpin state and thus facilitates translation, suggesting that the up-regulating constructs are RBP-sensing riboswitches. This finding is further supported by lengthening the binding-site stem, which in turn destabilizes the translationally-inactive state, and abolishes the up-regulating behavior. Finally, we found that the combination of two binding sites, positioned in the 5' UTR and gene-header of the same transcript, can yield a cooperative regulatory response. Together, we show that the interaction of an RBP with its RNA target facilitates structural changes in the RNA, which is reflected by a controllable range of binding affinities and dose response behaviors. Thus, demonstrating that RNA-RBP interactions can provide a platform for constructing gene regulatory networks that are based on translational, rather than transcriptional, regulation.

INTRODUCTION

One of the main goals of synthetic biology is the construction of complex gene regulatory networks. The majority of engineered regulatory networks have been based on transcriptional regulation, with only a few examples based on post-transcriptional regulation (1–4), although RNA-based regulatory components have many advantages. Several RNA components have been shown to be functional in multiple organisms (5–9). RNA can respond rapidly to stimuli, enabling a faster regulatory response

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

as compared with transcriptional regulation (10–13). From a structural perspective, RNA molecules can form a variety of biologically functional secondary and tertiary structures (2), which enables modularity. For example, distinct sequence domains within a molecule (13, 14) may target different metabolites or nucleic acid molecules (15, 16). All of these characteristics make RNA an appealing target for engineered-based applications (17–19, 3, 20, 21, 2, 22, 23).

Perhaps the most well-known class of RNA-based regulatory modules are riboswitches (16, 24–26). Riboswitches are noncoding mRNA segments that regulate the expression of adjacent genes via structural change, effected by a ligand or metabolite. However, response to metabolites cannot be easily used as the basis of a regulatory network, as there is no convenient feedback or feed-forward mechanism for connection of additional network modules. Riboregulators (27, 28, 2), namely riboswitches that sense a nucleic acid molecule, provide a transcription-based feedback mechanism, which can result in slow temporal dynamics (2). Implementing network modules using RNA-binding proteins (RBPs) could enable an alternative multicomponent connectivity for gene-regulatory networks that is not based solely on transcription factors. Such a network architecture can offer unique advantages from potentially faster response time to generating increased complexity and selectivity when implemented together with standard gene regulatory modules.

Regulatory networks require both inhibitory and up-regulatory modules. The vast majority of known RBP regulatory mechanisms are inhibitory (29–34). A notable exception is the phage RBP Com, whose binding was demonstrated to destabilize a sequestered ribosome binding site (RBS) of the Mu phage *mom* gene, thereby facilitating translation (35, 36). Several studies have attempted to engineer activation modules utilizing RNA-RBP interactions, based on different mechanisms: recruiting the eIF4G1 eukaryotic translation initiation factor to specific RNA targets via fusion of the initiation factor to an RBP (37, 38), adopting a riboswitch-like approach (21), and utilizing an RNA-binding version of the TetR protein (39). However, despite these notable efforts, RBP-based translational stimulation is still difficult to design in most organisms.

In a recent study (40), we employed a synthetic biology and *in vivo* SHAPE-Seq approach (41–43) to study repression controlled by an RBP bound to a hairpin within the N-terminus of a reporter gene, in bacteria. Here, we focus on regulation by an RBP bound within the 5' UTR of bacterial mRNA, following a design introduced by (12). Our findings indicate that structure-binding RBPs [coat proteins from the bacteriophages GA (44), MS2 (45), PP7 (46), and Qβ (47)] can generate a range of translational responses, from previously-observed down-regulation (12) to, surprisingly, up-regulation. The mechanism for downregulation is most likely steric hindrance of the initiating ribosome by the RBP-mRNA complex. For the 5' UTR sequences that exhibit up-regulation, RBP binding seems to facilitate a transition from an RNA structure with a low translation rate, into another RNA structure with a higher translation rate. These two experimental features indicate that the up-regulatory elements constitute protein-sensing riboswitches. Our findings imply that RNA-RBP interactions can provide a platform for constructing gene regulatory networks that are based on translational, rather than transcriptional, regulation.

MATERIAL AND METHODS

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Design and construction of binding-site plasmids

Binding-site cassettes (see Supp. Table 1) were ordered as double-stranded DNA minigenes from either Gen9 or Twist Bioscience. Each minigene was ~500 bp long and contained the following parts: EagI restriction site, ~40 bases of the 5' end of the Kanamycin (Kan) resistance gene, pLac-Ara constitutive promoter, ribosome binding site (RBS), and a KpnI restriction site. In addition, each cassette contained one or two wild-type or mutated RBP binding sites, either upstream or downstream to the RBS (see Supp. Table 1), at varying distances. All binding sites were derived from the wild-type binding sites of the coat proteins of one of the four bacteriophages GA, MS2, PP7, and Qβ. For insertion into the binding-site plasmid backbone, minigene cassettes were double-digested with EagI-HF and either KpnI or ApaLI (New England Biolabs [NEB]). The digested minigenes were then cloned into the binding-site backbone containing the rest of the mCherry gene, terminator, and a Kanamycin resistance gene, by ligation and transformation into *E. coli* TOP10 cells (ThermoFisher Scientific). All the plasmids were sequence-verified by Sanger sequencing. Purified plasmids were stored in 96-well format, for transformation into *E. coli* TOP10 cells containing one of the four fusion-RBP plasmids (see below).

Design and construction of fusion-RBP plasmids

RBP sequences lacking a stop codon were amplified via PCR off either Addgene or custom-ordered templates (Genescript or IDT, see Supp. Table 2). All RBPs presented (GCP, MCP, PCP, and QCP) were cloned into the RBP plasmid between restriction sites KpnI and AgeI, immediately upstream of an mCerulean gene lacking a start codon, under the so-called RhlR promoter (containing the *rhlAB* las box (48)) and induced by N-butyryl-L-homoserine lactone (C_4-HSL). The backbone contained an Ampicillin (Amp) resistance gene. The resulting fusion-RBP plasmids were transformed into *E. coli* TOP10 cells. After Sanger sequencing, positive transformants were made chemically-competent and stored at -80°C in 96-well format.

Transformation of binding-site plasmids

Binding-site plasmids stored in 96-well format were simultaneously transformed into chemically-competent bacterial cells containing one of the fusion plasmids, also prepared in 96-well format. After transformation, cells were plated using an 8-channel pipettor on 8-lane plates (Axygen) containing LB-agar with relevant antibiotics (Kan and Amp). Double transformants were selected, grown overnight, and stored as glycerol stocks at -80°C in 96-well plates (Axygen).

SHAPE-Seq experimental setup

LB medium supplemented with appropriate concentrations of Amp and Kan was inoculated with glycerol stocks of bacterial strains harboring both the binding-site plasmid and the RBP-fusion plasmid and grown at 37°C for 16 hours while shaking at 250 rpm. Overnight cultures were diluted 1:100 into semi-poor medium. Each bacterial sample was divided into a non-induced sample and an

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

induced sample in which RBP protein expression was induced with 250 nM N-butanoyl-L-homoserine lactone (C_4-HSL), as described above.

Bacterial cells were grown until OD_{600}=0.3, 2 ml of cells were centrifuged and gently resuspended in 0.5 ml semi-poor medium. For *in vivo* SHAPE modification, cells were supplemented with a final concentration of 30 mM 2-methylnicotinic acid imidazole (NAI) suspended in anhydrous dimethyl sulfoxide (DMSO, Sigma Aldrich) (42), or 5% (v/v) DMSO. Cells were incubated for 5 min at 37°C while shaking and subsequently centrifuged at 6000 g for 5 min. RNA isolation of 5S rRNA was performed using TRIzol-based standard protocols. Briefly, cells were lysed using Max Bacterial Enhancement Reagent followed by TRIzol treatment (both from Life Technologies). Phase separation was performed using chloroform. RNA was precipitated from the aqueous phase using isopropanol and ethanol washes, and then resuspended in RNase-free water. For the strains harboring PP7-wt δ=-29 and PP7-USs δ=-29, column-based RNA isolation (RNeasy mini kit, QIAGEN) was performed. Samples were divided into the following sub-samples (except for 5S rRNA, where no induction was used):

1. induced/modified (+C_4-HSL/+NAI)
2. non-induced/modified (-C_4-HSL/+NAI)
3. induced/non-modified (+C_4-HSL/+DMSO)
4. non-induced/non-modified (-C_4-HSL/+DMSO).

In vitro modification was carried out on DMSO-treated samples (3 and 4) and has been described elsewhere (43). Briefly, 1500 ng of RNA isolated from cells treated with DMSO were denatured at 95°C for 5 min, transferred to ice for 1 min and incubated in SHAPE-Seq reaction buffer (100 mM HEPES [pH 7.5], 20 mM $MgCl_2$, 6.6 mM NaCl) supplemented with 40 U of RiboLock RNAse inhibitor (Thermo Fisher Scientific) for 5 min at 37°C. Subsequently, final concentrations of 100 mM NAI or 5% (v/v) DMSO were added to the RNA-SHAPE buffer reaction mix and incubated for an additional 5 min at 37°C while shaking. Samples were then transferred to ice to stop the SHAPE-reaction and precipitated by addition of 3 volumes of ice-cold 100% ethanol, followed by incubation at -80°C for 15 min and centrifugation at 4°C, 17000 g for 15 min. Samples were air-dried for 5 min at room temperature and resuspended in 10 µl of RNAse-free water.

Subsequent steps of the SHAPE-Seq protocol, that were applied to all samples, have been described elsewhere (49), including reverse transcription (steps 40-51), adapter ligation and purification (steps 52-57) as well as dsDNA sequencing library preparation (steps 68-76). In brief, 1000 ng of RNA were converted to cDNA using the reverse transcription primers (for details of primer and adapter sequences used in this work see Supp. Table 3) for mCherry (#1) or 5S rRNA (#2) that are specific for either the mCherry transcripts (PP7-USs δ=-29, PP7-wt δ=-29). The RNA was mixed with 0.5 µM primer (#1) or (#2) and incubated at 95°C for 2 min followed by an incubation at 65°C for 5 min. The Superscript III reaction mix (Thermo Fisher Scientific; 1x SSIII First Strand Buffer, 5 mM DTT, 0.5 mM dNTPs, 200 U Superscript III reverse transcriptase) was added to the cDNA/primer mix, cooled down to 45°C and subsequently incubated at 52°C for 25 min. Following inactivation of the reverse

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transcriptase for 5 min at 65°C, the RNA was hydrolyzed (0.5 M NaOH, 95°C, 5 min) and neutralized (0.2 M HCl). cDNA was precipitated with 3 volumes of ice-cold 100% ethanol, incubated at -80°C for 15 minutes, centrifuged at 4°C for 15 min at 17000 g and resuspended in 22.5 µl ultra-pure water. Next, 1.7 µM of 5' phosphorylated ssDNA adapter (#3) (see Supp. Table 3) was ligated to the cDNA using a CircLigase (Epicentre) reaction mix (1xCircLigase reaction buffer, 2.5 mM $MnCl_2$, 50 µM ATP, 100 U CircLigase). Samples were incubated at 60°C for 120 min, followed by an inactivation step at 80°C for 10 min. cDNA was ethanol precipitated (3 volumes ice-cold 100% ethanol, 75 mM sodium acetate [pH 5.5], 0.05 mg/mL glycogen [Invitrogen]). After an overnight incubation at -80°C, the cDNA was centrifuged (4°C, 30 min at 17000 g) and resuspended in 20 µl ultra-pure water. To remove non-ligated adapter (#3), resuspended cDNA was further purified using the Agencourt AMPure XP beads (Beackman Coulter) by mixing 1.8x of AMPure bead slurry with the cDNA and incubation at room temperature for 5 min. The subsequent steps were carried out with a DynaMag-96 Side Magnet (Thermo Fisher Scientific) according to the manufacturer's protocol. Following the washing steps with 70% ethanol, cDNA was resuspended in 20 µl ultra-pure water and were subjected to PCR amplification to construct dsDNA library as detailed below.

SHAPE-Seq library preparation and sequencing

To produce the dsDNA for sequencing 10ul of purified cDNA from the SHAPE procedure (see above) were PCR amplified using 3 primers: 4nM mCherry selection (#4) or 5S rRNA selection primer (#5), 0.5µM TruSeq Universal Adapter (#6) and 0.5µM TrueSeq Illumina indexes (one of #7-26) (Supp. Table 3) with PCR reaction mix (1x Q5 HotStart reaction buffer, 0.1 mM dNTPs, 1 U Q5 HotStart Polymerase [NEB]). A 15-cycle PCR program was used: initial denaturation at 98°C for 30 s followed by a denaturation step at 98°C for 15 s, primer annealing at 65°C for 30 s and extension at 72°C for 30 s, followed by a final extension 72°C for 5 min. Samples were chilled at 4°C for 5 min. After cool-down, 5 U of Exonuclease I (ExoI, NEB) were added, incubated at 37°C for 30 min followed by mixing 1.8x volume of Agencourt AMPure XP beads to the PCR/ExoI mix and purified according to manufacturer's protocol. Samples were eluted in 20 µl ultra-pure water. After library preparation, samples were analyzed using the TapeStation 2200 DNA ScreenTape assay (Agilent) and the molarity of each library was determined by the average size of the peak maxima and the concentrations obtained from the Qubit fluorimeter (Thermo Fisher Scientific). Libraries were multiplexed by mixing the same molar concentration (2-5 nM) of each sample library, and library and sequenced using the Illumina HiSeq 2500 sequencing system using either 2X51 paired end reads for the 5S-rRNA control and *in vitro* experiments or 2x101 bp paired-end reads for all other samples. See Supp. Table 4 for read counts for all experiments presented in the manuscript.

Analysis of fluorescence expression assay

See Supplementary Information.

SHAPE-Seq reactivity analysis

See Supplementary Information.

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tandem cooperativity fit and analysis

See Supplementary Information.

RESULTS

RBP-binding can effect either up- or down-regulation

We studied the regulatory effect generated by RNA-binding phage coat proteins for GA (GCP), MS2 (MCP), PP7 (PCP), and Qβ (QCP) when co-expressed with a reporter construct containing 1 of 11 putative binding sites in the 5' UTR [see Supp. Fig. 1 and previously described in (40, 46, 50)]. We positioned each of the 11 binding sites at three or four different locations upstream of the RBS, that varied from δ=-21 to δ=-31 nt measured relative to the ATG of the mCherry reporter gene (see Supp. Table 1). Altogether, we constructed 44 reporter constructs (including non-hairpin controls), and co-transformed with all four RBPs, resulting in a total of 176 regulatory strains. RBP levels were induced by addition of N-butyryl-L-homoserine lactone (C_4HSL), at 24 different concentrations (see Supp. Fig. 2 for experiment schematic and sample data-set). The normalized dose-response curves for the 5' UTR constructs with matching RBPs (e.g., MS2 binding site with MCP) are plotted as a heat-map in Fig. 1A. In all cases, the data for both the mCherry rate of production and mean mCerulean levels are normalized by the respective maximal value. The dose response functions are arranged in order of increasing response, with the highest-repression variants depicted at the bottom. The results show that the observed repression is generally weak, and at most amounts to about a factor of two reduction from basal levels (turquoise). This is markedly different from the strong repression effect observed for when the same hairpins were positioned within the ribosomal initiation region downstream of the AUG (i.e. "gene-header" - δ<13) (40). Second, some of the variants exhibit a distinct up-regulatory dose-response (variants positioned at the top of the map at variant #>140) of up to ~3-fold, which was not previously observed for these RBPs.

From the dose-response curves, we computed the effective dissociation constant, which is defined as the fitted K_d, normalized by the maximal mCerulean expression level (see Supp. Information). The resultant K_{RBP} values obtained for each RBP—binding-site pair are plotted as a heat-map in Fig. 1B. This heatmap is quite similar to the K_{RBP} heatmap obtained for the gene-header region, despite the difference in fold regulatory effect (40). We found similar effective dissociation constant values (up to an estimated fit error of 10%) for all binding-site positions, for each of the native binding sites (MS2-wt, PP7-wt, and Qβ-wt), and for the mutated sites with a single mutation in the loop region [MS2-U(-5)C and MS2-U(-5)G]. However, deviations in effective dissociation constant were observed for several of the mutated sites. In particular, both Qβ-USLSLm and Qβ-LSs generated a down-regulatory dose-response signal in the 5' UTR in the presence of QCP, while no response was detected in the gene-header configurations. Conversely, QCP and PCP, which displayed a binding affinity to MS2-wt, MS2-U(-5)C, and PP7-USLSBm in the gene-header configuration, respectively (40), did not generate a significant dose response to either of these sites in the 5' UTR. Thus, it seems that binding affinity depends on the molecular structure that forms *in vivo*, which may be different from the *in silico* expectation, and likely also depends on flanking sequences.

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To highlight the different types of dose-responses (up- or down-regulation) in the 5' UTR region, we plot the regulatory response for each binding site that was tested in Fig. 1C. Each bar represents the fold change for a single RBP—binding-site combination, averaged over several binding-site positions. Fold change is defined as the ratio between the rates of mCherry production for the RBP-induced and non-induced cases. In particular, for MS2-U(-5)C, MS2-wt, and PP7-wt, together with MCP, GCP and PCP, respectively, a modest two to three-fold up-regulation was observed. For all binding sites, the precise positioning within the 5' UTR did not seem to play a strong regulatory role, as the regulatory response remained fairly constant across binding site locations (Supp. Fig. 3A-B). To provide a more detailed description of the up-regulatory response, we plot in the inset of Fig. 1C the individual dose response curves that were used to compute the red bar in the MS2-U(-5)C line of the plot. The panel shows a sigmoidal response for three of the four positions. However, the basal production-rate levels are not fully repressed providing an explanation for the modest fold-response observed for these constructs.

To better understand the different types of dose-response behavior in the 5' UTR, we focus on the PCP binding site variants PP7-USs (red) and PP7-wt (blue), both at δ=-29 (Fig. 1D). There is a single U-A base-pair deletion in the upper stem for PP7-USs as compared with PP7-wt, while the rest of the 5' UTR is identical (see Fig. 1E and Supp. Table 1). When we examine the expression-level dose responses measured for these variants, we see that the PP7-wt response function exhibits a low production rate in the absence of induction (~150 a.u./hr), while at full induction it rises to an intermediate production rate (~450 a.u./hr). For PP7-USs, the basal rate of production level at zero induction is ~ 1100 a.u./hr, and declines upon induction to levels similar to that observed for PP7-wt at full induction. The comparable levels at full induction for both strains indicate that the effect of the bound protein on expression is similar for both constructs. These results imply that the difference in expression levels in the absence of protein may be due to differences in underlying structures of the RNA molecules in the 5' UTR, even though a putative structural depiction of these variants based on RNAfold (51) and their sequence alone, does not predict significant structural differences in this region (Fig. 1E).

In vitro SHAPE-Seq data is consistent with predicted structures

In order to unravel the connection between the structure of the 5' UTR and resultant dose-response functions, we subjugated the PP7-wt and PP7-USs constructs at δ=-29 to SHAPE-Seq both *in vitro* and *in vivo* using 2-methylnicotinic acid imidazole (NAI) suspended in anhydrous dimethyl sulfoxide (DMSO), with DMSO-treated cells as a non-modified control (see Methods and Supp. Information). We chose to modify a segment that includes the entire 5' UTR, and in addition another ~140 nt of the mCherry reporter gene. We hypothesized that SHAPE-seq data can provide a foot-print or echo for the mRNA structure in the 5'UTR and ribosomal initiation region with and without a bound RBP (see Supp. Fig. 4 for SHAPE-Seq analysis of 5S-rRNA as positive control).

In Fig. 2A and Supp. Fig. 5, we plot the reactivity signals as a function of nucleotide obtained for both the PP7-wt (blue line) and PP7-USs (red line) constructs at δ=-29 using *in vitro* SHAPE-Seq.

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The reactivity of each base corresponds to the propensity of that base to be modified by NAI (for the definition of reactivity see SI). For each data-point in the plots, error-bars are computed from two biological replicates for each variant, and additional boot-strapping analysis (see Supp. Information for detailed description of the analysis approach). A direct comparison of the reactivity scores (Supp. Fig. 5) show that both constructs generate a fluctuating reactivity signature that varies from no reactivity to values of ~1. Furthermore, a close examination of both *in vitro* signals reveals a global 2 nt offset between the PP7-wt and PP7-USs reactivity scores downstream to the binding sites (Supp. Fig. 5A). The observed offset coincides with a wide-region of statistically distinguishable reactivities computed using Z-factor analysis (see SI for definition and Supp. Fig. 5B). Since the two constructs differ by a deletion of two nucleotides, we reasoned that in order to facilitate a proper alignment between the PP7-USs and PP7-wt reactivity scores downstream to the binding sites, the reactivities at positions -45 and -38 of PP7-wt should be omitted from the plot (Fig. 2A). When doing so both *in vitro* reactivity signals look nearly identical for the entire modified segment of the RNA. This is further confirmed by z-factor analysis (lower panel), which only yields significant distinguishability for a narrow segment within the coding region (~+30 nt).

Next, we used the *in vitro* reactivity data to compute the structure of the variants and guiding the computational prediction using our experimental data. It was shown previously (52–55) that using experimental constraints for RNA 2D structure computation can increase the accuracy of the predicted structure, increasing its similarity to the solved structure. In this computation, the algorithm utilizes both the sequence and the reactivity to generate the structure. In brief, RNAfold together with RNApvmin (see SI for more detail) implement the algorithm developed by (52), which uses the reactivity scores to constrain the free-energy computations of the different structures. This is done by "biasing" the way each base is responsive to the general free energy minimization computation taking into account *a priory* the experimental reactivity data. Thus, the free-energy minimizing structure that results is different from the one that would be obtained from computations that are based on the sequence alone. In Fig. 2B we plot the structures for both variants, as computed using constraints from the *in vitro* SHAPE-Seq data, which reveals that the extracted structures (Fig. 2B) are similar to the initial non-constrained RNAfold computation (Fig.1E). A closer examination of all the computed structures so far reveals two 5' UTR features that consistently appear. The first corresponds to the binding site (-56 to -30) as expected, while the second corresponds to a downstream satellite structure (-23 to -10). The secondary hairpin encodes a putative short anti-Shine-Dalgarno (aSD) motif (CUCUU) (56), which may partially sequester the RBS. While RBS-sequestration by an aSD motif can explain the up-regulation effect observed for PP7-wt, it cannot at the same time explain the down-regulatory phenomenon observed for PP7-USs, nor its high basal production rate levels.

In vivo SHAPE-Seq reveals different structures for PP7-wt and PP7-USs

To resolve this discrepancy, we proceeded to carry out the SHAPE-Seq protocol *in vivo* on induced and non-induced samples for both the PP7-wt and PP7-USs δ=-29 variants. We used biological duplicates for every variant/induction level pair. In Fig. 3A, we plot the non-induced (RBP-)

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reactivity obtained for PP7-wt (blue) and PP7-USs (red). The data shows that PP7-USs is more reactive across nearly the entire segment, including all of the 5' UTR and >50 nt into the coding region. Z-factor analysis reveals that this difference is statistically significant for a large portion of the 5' UTR and the coding region, suggesting that the PP7-USs is overall more reactive and thus less structured than the PP7-wt fragment. Alternatively, in Fig. 3B we show that in the induced state (RBP+) both constructs exhibit a weaker reactivity signal that is statistically indistinguishable in the 5' UTR (i.e. Z-factor ~0). Moreover, the region associated with the binding site is unreactive (marked in grey), while both the adjacent upstream and downstream regions exhibit a moderate reactivity signal. To further explore the reactivity signal of the 5'UTR in the induced cases, we plot the induced versus non-induced reactivities for each construct (Supp. Fig. 6). The plots reveal that for the PP7-wt construct (Supp. Fig. 6A), the binding site location coincides with a statistically distinguishable protected region that becomes non-reactive upon induction. For PP7-USs (Supp. Fig. 6B), no such identification can be made due to the radically different reactivity signals observed for the two states. Taken together, PCP-mCerulean induction seems to trigger structural changes in the mRNA molecules. For PP7-USs, RBP binding likely leads to a moderate re-structuration of the 5'UTR, which in turn triggers reduced translation. Whereas, for the PP7-wt construct a signature for RBP binding can be discerned and taking into account the nearly identical reactivity with PP7-USs in the induced case a likely structural shift ensues as well.

To generate a structural insight, we implemented the constrained structure computation that was used for the *in vitro* samples. This was done in order to derive structures of the RNA molecules that are consistent with the reactivity data obtained for the different induction states. The structures with nucleotides overlaid by base-pairing probabilities are plotted in Fig. 3C. In the top schema, we plot the derived PP7-USs non-induced variant, which is non-structured in the 5' UTR exhibiting a predominantly yellow and white coloring of the individual nucleotide base-pairing probabilities. By contrast, in the PP7-wt non-induced structure (bottom) there are three predicted closely spaced smaller hairpins that span from -60 to -10 that are predominantly colored by yellow and red except in the predicted loop regions. Both top and bottom structures are markedly different from the *in vitro* structures (Fig. 2B). Neither displays the PP7-wt or PP7-USs binding site, and the secondary aSD hairpin only appears in the PP7-wt non-induced strain. In the induced state, a structure reminiscent of the *in vitro* structure is recovered for both variants with three distinct structural features visible in the 5' UTR: the upstream flanking hairpin, the binding site, and downstream CUCUU anti-Shine Dalgarno hairpin. These variety of predicted structures for each state *in vivo* suggests that the level of translation may be mostly dependent on a particular arrangement of sub-structures in the 5' UTR, and to a lesser extent on the presence of the aSD motif. Consequently, the reactivity data and structural analysis indicates that the deletion of the two nucleotides which encode the PP7-USs binding site together with the translational machinery, are sufficient to trigger large-scale structural changes across the 5' UTR, which in turn lead to the divergent expression levels at the non-induced level. Conversely, the binding of PCP-mCerulean is sufficient for the stabilization of the binding site, which in turn stabilizes the satellite structures in the flanking regions leading to an indistinguishable expression level in the induced states.

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Longer stems can flip the regulatory response of PP7-wt and MS2-wt binding sites

Based on the collective findings of the various SHAPE-Seq libraries, we hypothesized that we can change the nature of the regulation of PP7-wt from up-regulating to down-regulating by altering the sequence of the flanking regions. We reasoned that different flanking regions together with the translational machinery can trigger a non-structured 5' UTR that will mimic the effect of the U-A base-pair deletion in PP7-USs. Thus, upon the binding of the RBP and subsequent stabilization of the binding site structure in the 5'UTR, the up-regulation effect can be changed to repressed translation. We chose to test this hypothesis via two approaches: first, by mutating the CUCUU motif into an A-track, thus preventing the formation of the secondary hairpin, and second, by increasing the length of the lower stem thereby increasing the stability of the binding site at the expanse of alternative structures such as the one shown in Fig. 3C (PP7-wt, induced - bottom right structure). In Fig. 4A, we present a sample of mutated structures predicted with RNAfold (using only sequence information), where the mutated flanking sequence for each variant is overlaid by color: red for the CU-rich variant (top-left), blue for the A-rich variant (top-right), yellow for the extended stem (+3 nt) PP7-wt variant (bottom-right) and green for the original PP7-wt variant (bottom-left).

To check the effect of varying the aSD motif, we produced ten additional constructs at δ=-29 with a PP7-nB, PP7-USs, PP7-wt, MS2-wt, or MS2-U(-5)C binding site, in which the sequence between the binding site and the RBS encoded either a strong aSD sequence (CU-rich), or an A-rich segment that is highly unlikely to form stable secondary structures with the RBS (see Supp. Table 1). We first modeled these new variants and found radically different structures in the 5' UTR than the ones deduced for the original variants (see Fig. 4A-top left, marked with red bold label in Fig. 4B). For the A-rich flanking region (Fig. 4A-top right, A-track marked with blue bold label in Fig. 4B), the predicted structure displays an extended hairpin feature encompassing most of the 5' UTR, but the region upstream to the RBS and AUG is predicted to be single-stranded as expected. Conversely, the predicted structure for the CU-rich downstream flanking region (Fig. 4A-top left - highlighted in red, CU-rich segment marked with red bold label in Fig. 4B) displays a longer RBS-sequestering hairpin as compared with the original sequence (Fig. 4A-bottom left). We plot the basal expression level for 15 RBP—binding-site pairs, containing the original δ=-29 spacer (green), the δ=-29 spacer containing the CU-rich aSD sequence (red), and the δ=-29 A-rich spacer lacking the aSD sequence (blue). The data (Fig.4B – left heatmap) shows that the constructs with a CU-rich flanking region exhibit low basal expression levels as compared with the other constructs, as previously observed (56), while the different A-rich variants and original flanking sequences of each type do not seem to affect basal expression in a consistent fashion. However, both the up-regulatory and down-regulatory dose responses persist independently of the flanking region content (Fig. 4B- right heatmap), as compared with the response recorded for the original flanking sequences (Fig. 1C).

To check the effect of increasing stem length and binding site stability, we designed six new variants for each of the MS2-wt and PP7-wt binding sites by extending the length of the lower stem by 3, 6, and 9 base-pairs, with both GU and GC repeats (see Supp. Table 1). Thus, in this case we

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

effectively mutated both the upstream and downstream flanking regions of the original PP7-wt and MS2-wt constructs by insertions of 3,6, or 9 nucleotides each. Structural modelling of these variants reveal that the stem extension alters the secondary structural features that flank the binding site (e.g. short aSD-SD motif) and, as expected, increases the stability of the binding site stem by making it longer (see the sample PP7-wt+3bp structure in Fig. 4A-bottom – right – insertions highlighted in yellow). When examining the dose-response functions for all the configurations (Fig. 4C-D) the up-regulating dose-response functions observed for both PP7-wt with PCP and MS2-wt with MCP were converted to down-regulating responses. In all cases, the basal expression level for the non-induced state was increased by ~10-fold (Fig. 4D – left heatmap), and the down-regulatory effect that was observed upon RBP induction did not seem to reach saturation, with an apparent increase in the effective dissociation constant by 2 to 3-fold (Supp. Fig. 7). Consequently, basal expression level is highly dependent on the particular choice of 5' UTR sequence that flanks the binding site and resultant 5' UTR structure. This, in turn, can lead to either up- or down-regulation depending on the final structure of the complex in the RBP-bound state.

A tandem of binding sites can exhibit cooperativity

Finally, we further explored the regulatory potential of constructs containing one hairpin placed in the 5' UTR of a gene ($\delta<0$) with another hairpin placed in the gene-header region of the reporter ($\delta>0$) (Fig. 5A), in the presence of RBP. To do so, we constructed an additional 28 variants with combinations of MS2, PP7, and Qβ binding sites (see Supp. Table 1). In Fig. 5B-D, we plot the dose responses of the tandem variants in the presence of MCP, PCP, and QCP as heatmaps arranged in order of increasing basal mCherry rate of production. Overall, the basal mCherry production rate for all the tandem variants is lower as compared with the single-binding-site variants located in the 5' UTR. In addition, approximately half of the variants generated a significant regulatory response in the presence of the RBP, while the other half seem to be repressed at the basal level, with no RBP-related effect detected.

For MCP (Fig. 5B), we observed strong repression for four of the ten variants tested, with the MS2-U(-5)G binding site positioned in the N-terminus for all four repressed variants. With different N-terminus binding sites [MS2-wt, Qβ-wt, or MS2-U(-5)C], basal mCherry rate of production was reduced to nearly zero. For PCP (Fig. 5C), a similar picture emerges, with several variants exhibiting a strong dose-response repression signature, while no regulatory effect observed for others. Interestingly, the variants in the top six in terms of basal mCherry production rate all encode the PP7-nB binding site in the 5' UTR. Moreover, all eight variants with a PP7-nB positioned in the 5' UTR exhibit a down-regulatory response. These observations are consistent with the data shown in Fig. 4C-D, where the binding sites with longer stems resulted in higher basal mCherry rate of production, presumably due to increased hairpin stability. For other PP7 binding-site combinations a lower basal level, and hence lower fold-repression effect, is observed.

In Fig. 5D, we present the dose-response heatmaps obtained for QCP. Here, we used the same tandem variants as for MCP, due to the binding cross-talk between both proteins shown

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

previously (40). Interestingly, the dose responses for these tandems in the presence of QCP vary substantially as compared with that observed for MCP. While the site MS2-U(-5G) is still associated with higher basal expression when positioned in the N-terminus, only three variants (as compared with five for MCP) do not seem to respond to QCP. In particular, two variants, each containing MS2-U(-5)G in the N-terminus and MS2-U(-5)C in the 5' UTR, exhibit a 2-fold up-regulation dose response, as compared with a strong down-regulatory effect for MCP.

Finally, we measured the effective cooperativity factor *w* (see Supp. Information for fitting model) for repressive tandem constructs in the presence of their corresponding cognate RBPs. In Fig. 5E, we plot a sample fit for a MS2-U(-5)C/MS2-U(-5)G tandem in the presence of MCP. The data shows that when taking into account the known K_{RBP} values that were extracted for the single-binding-site N-terminus and 5' UTR positions, a fit with no cooperativity (w=1) does not explain the data well (red line). However, when the cooperativity parameter is not fixed, a good description for the data is obtained for $50<w<80$ (best fit at w=73). In Fig. 5F we plot the extracted cooperativity parameter for each of the 16 tandems displaying a regulatory response with known K_{RBP} values for both sites (see Supp. Fig. 8 and Supp. Table 4 for fits and parameter values, respectively). Altogether, at least 6 of the 16 tandems exhibited cooperative behavior. For MCP and QCP five of the six relevant tandems displayed strong cooperativity ($w>25$). For PCP, only two of the ten tandems displayed weak cooperativity ($1<w<25$); these tandems had less than 30 nt between the two PCP binding sites. The cooperative behavior, which reflects overall increase in affinity of the RBP to the molecule when there is more than one binding site present, may also indicate increased stability of the hairpin structures. An increased stability can explain two additional features of the tandems, that were not observed for the single binding site constructs: the QCP up-regulatory response observed for the MS2-U(-5)C/MS2-U(-5)G tandem, and the decreased basal mCherry rate of production levels. Overall, the effective dissociation constant of the tandem- and single-binding-site constructs together with the RBPs can be varied over a range of specificities that spans approximately an order of magnitude, depending on the chosen 5' UTR and gene-header sequences.

DISCUSSION

Synthetic biology approaches have been increasingly used in recent years to map potential regulatory mechanisms of transcriptional and translational regulation, in both eukaryotic and bacterial cells (56–61). Here, we built on the design introduced by (12) to quantitatively study RBP-based regulation in bacterial 5' UTRs, using the combined synthetic biology and SHAPE-Seq approach that we recently applied to the study of hairpins with bacterial gene-header regions (40). Using a library of RNA variants, we were able to identify three RNA-based regulatory mechanisms: weak repression of translational initiation by a hairpin-RBP complex (as in Fig. 1), a previously un-observed translational stimulation phenomenon due to RBP binding and change in mRNA structure in the 5' UTR (described mainly in Fig. 2-4), and a cooperative increase in overall binding affinity of the RBP to the molecule when two similar binding sites are encoded (one in the 5' UTR and the other in the gene-header, as in Fig. 5).

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our expression level data on the single-binding-site constructs, SHAPE-Seq data, and follow-up structural analysis suggest that a “densely” structured 5’ UTR is associated with an inhibited-translation-initiation state. Inhibited translation is alleviated by RBP binding, which seems to stabilize the binding-site hairpin while simultaneously weakening flanking structures in the 5’ UTR, leading to translation stimulation. Our experiments revealed no particular structural feature that was associated with this regulatory switch, such as the release of a sequestered RBS, which has been reported before as a natural mechanism for translational stimulation (35, 36). Consequently, the up-regulation phenomenon that we observed is a transition from a strongly-repressing densely structured 5’ UTR to a weakly-repressing loosely structured 5’ UTR that occurs upon RBP binding.

Structural changes that are triggered in RNA by the binding of a ligand frequently occur in nature. The most common class of such structures are called “riboswitches”, where the structure change typically either sequesters or releases an RBS leading to inhibition or up-regulation of translation, see for instance (7, 26, 62–66). In this work, we demonstrated that RNA binding proteins can function as such ligands. In the two cases studied in detail, we demonstrated that upon induction the RBP triggered structural changes in the RNA molecule. This result is not surprising for several reasons: first, the size of the RBPs are comparable to typical structural feature on RNAs, and thus they are likely to affect the stability of nearby structural elements. Second, it is believed that RNA structures fluctuate between closely related ensemble of structures (67–70), and thus binding of an RBP can easily shift the energetic equilibrium of this ensemble leading to a different cumulative translation rate. Third, interaction with the translational machinery can substantially alter the underlying structure. We and others (40, 71–73) have shown that mRNAs that are being strongly translated are predominantly non-structured. However, bound RBPs in the 5’ UTR near the RBS are likely to hinder translation initiation. This slowdown can, in turn, trigger re-structuring of the RNA molecule leading to a further slowdown of translation, and to a radically different reactivity signature for the RBP bound and unbound states as was observed here and in our previous work (40). Consequently, the definition of a “riboswitch” should be widened to include all forms of ligand-triggered structural changes (i.e. not only to a small metabolite or an ion), which in turn lead to variations in translation. Given this wider definition, the constructs reported here can be considered as synthetic protein-sensitive riboswitches.

We emphasize that our results were obtained in *E. coli*, and thus may be specific to bacterial systems. However, the up-regulatory phenomenon was surprisingly robust: it was observed for 4 of 11 of the binding sites positioned in the 5’ UTR (depicted in Supp. Fig. 1), for two different binding site structures (MS2-wt and PP7-wt), and for all four RBPs. Moreover, it was not sensitive to changes in the downstream flanking region. Finally, given the propensity of RBPs to alter the RNA structure via direct interaction or inhibition of translation initiation, it is tempting to speculate that such an interaction may be a generic 5’ UTR mechanism that could be extended to other RBPs in any organism as was shown for the 3’UTR in human cell lines (74, 75).

How difficult is it to design this up-regulatory phenomenon *de novo*? Our findings suggest that the free-energy score of the binding sites can be used as a predictor for up-regulation independent of flanking sequence content. The binding sites which exhibited up-regulation were all scored in the

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

intermediate range of free-energy values (Supp. Fig. 1); however, all of the more stable binding sites (free energy < -10 kcal/mol) such as PP7-nB and the longer stem variants exhibited high basal level expression and subsequent down-regulation. Thus, while intermediate stability is not necessarily a guarantee for achieving an up-regulatory response (e.g. PP7-USs, Qβ-wt, Qβ-LSs), it is a reasonable indicator which may be utilized during the design process (4 of 7 sites in our case). It is important to emphasize, however, that while such empirical findings may be used as a guide for future designs of synthetic protein-sensing riboswitches, we caution against over-reliance on such parameters. Overall, we found that structural models of RNA do not perform well in the *in vivo* environment, and especially poorly for translationally-active transcripts. This implies that the best approach to designing such elements at the present time is to first characterize experimentally a small library of a variety of designs, and subsequently selecting and optimizing the most suitable parts.

Our work presents an important step in understanding and engineering post-transcriptional regulatory networks. Our findings suggest that generating translational regulation using RBPs, and translational stimulation in particular, may not be as difficult as previously thought, especially given the more inclusive definition riboswitches that we propose. Furthermore, the cooperativity observed for the tandems indicates that binding affinity is also a variable parameter, which can be used in various designs. Consequently, the described constructs add to the growing toolkit of translational regulatory parts and provide a working design for further exploration of both natural and synthetic post-transcriptional gene regulatory networks.

SUPPLEMENTARY DATA

8 Supplementary Figures.

5 Supplementary Tables.

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

The authors would like to acknowledge the Technion's LS&E staff (Tal Katz-Ezov and Anastasia Diviatis) for help with sequencing.

FUNDING

This project received funding the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation (Grant No. 152/11); and Marie Curie Reintegration Grant No. PCIG11-GA-2012-321675.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. Win,M.N. and Smolke,C.D. (2008) Higher-Order Cellular Information Processing with Synthetic RNA Devices. *Science*, **322**, 456–460.

2. Green,A.A., Silver,P.A., Collins,J.J. and Yin,P. (2014) Toehold Switches: De-Novo-Designed Regulators of Gene Expression. *Cell*, **159**, 925–939.

3. Xie,Z., Wroblewska,L., Prochazka,L., Weiss,R. and Benenson,Y. (2011) Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science*, **333**, 1307–1311.

4. Wroblewska,L., Kitada,T., Endo,K., Siciliano,V., Stillo,B., Saito,H. and Weiss,R. (2015) Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery. *Nat Biotech*, **33**, 839–841.

5. Harvey,I., Garneau,P. and Pelletier,J. (2002) Inhibition of translation by RNA-small molecule interactions. *RNA*, **8**, 452–463.

6. Suess,B., Hanson,S., Berens,C., Fink,B., Schroeder,R. and Hillen,W. (2003) Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Res.*, **31**, 1853–1858.

7. Desai,S.K. and Gallivan,J.P. (2004) Genetic Screens and Selections for Small Molecules Based on a Synthetic Riboswitch That Activates Protein Translation. *J. Am. Chem. Soc.*, **126**, 13247–13254.

8. Buxbaum,A.R., Haimovich,G. and Singer,R.H. (2015) In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.*, **16**, 95–109.

9. Green,A.A., Kim,J., Ma,D., Silver,P.A., Collins,J.J. and Yin,P. (2017) Complex cellular logic computation using ribocomputing devices. *Nature*, **548**, 117–121.

10. Hentze,M.W., Caughman,S.W., Rouault,T.A., Barriocanal,J.G., Dancis,A., Harford,J.B. and Klausner,R.D. (1987) Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science*, **238**, 1570–1573.

11. St Johnston,D. (2005) Moving messages: the intracellular localization of mRNAs. *Nat. Rev. Mol. Cell Biol.*, **6**, 363–375.

12. Saito,H., Kobayashi,T., Hara,T., Fujita,Y., Hayashi,K., Furushima,R. and Inoue,T. (2010) Synthetic translational regulation by an L7Ae–kink-turn RNP switch. *Nat. Chem. Biol.*, **6**, 71–78.

13. Lewis,C.J.T., Pan,T. and Kalsotra,A. (2017) RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.*, **18**, 202–210.

14. Khalil,A.S. and Collins,J.J. (2010) Synthetic biology: applications come of age. *Nat. Rev. Genet.*, **11**, 367–379.

15. Isaacs,F.J., Dwyer,D.J. and Collins,J.J. (2006) RNA synthetic biology. *Nat. Biotechnol.*, **24**, 545–554.

16. Werstuck,G. and Green,M.R. (1998) Controlling gene expression in living cells through small molecule-RNA interactions. *Science*, **282**, 296–298.

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. Hutvágner,G. and Zamore,P.D. (2002) A microRNA in a Multiple-Turnover RNAi Enzyme Complex. *Science*, **297**, 2056–2060.

18. Rinaudo,K., Bleris,L., Maddamsetti,R., Subramanian,S., Weiss,R. and Benenson,Y. (2007) A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat Biotech*, **25**, 795–801.

19. Delebecque,C.J., Lindner,A.B., Silver,P.A. and Aldaye,F.A. (2011) Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science*, **333**, 470–474.

20. Chen,A.H. and Silver,P.A. (2012) Designing biological compartmentalization. *Trends Cell Biol.*, **22**, 662–670.

21. Ausländer,S., Stücheli,P., Rehm,C., Ausländer,D., Hartig,J.S. and Fussenegger,M. (2014) A general design strategy for protein-responsive riboswitches in mammalian cells. *Nat. Methods*, **11**, 1154–1160.

22. Sachdeva,G., Garg,A., Godding,D., Way,J.C. and Silver,P.A. (2014) In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. *Nucleic Acids Res.*, **42**, 9493–9503.

23. Pardee,K., Green,A.A., Takahashi,M.K., Braff,D., Lambert,G., Lee,J.W., Ferrante,T., Ma,D., Donghia,N., Fan,M., *et al.* (2016) Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell*, **165**, 1255–1266.

24. Winkler,W.C. and Breaker,R.R. (2005) Regulation of Bacterial Gene Expression by Riboswitches. *Annu. Rev. Microbiol.*, **59**, 487–517.

25. Wittmann,A. and Suess,B. (2012) Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett.*, **586**, 2076–2083.

26. Serganov,A. and Nudler,E. (2013) A Decade of Riboswitches. *Cell*, **152**, 17–24.

27. Buskirk,A.R., Landrigan,A. and Liu,D.R. (2004) Engineering a Ligand-Dependent RNA Transcriptional Activator. *Chem. Biol.*, **11**, 1157–1163.

28. Bayer,T.S. and Smolke,C.D. (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol.*, **23**, 337–343.

29. Romaniuk,P.J., Lowary,P., Wu,H.N., Stormo,G. and Uhlenbeck,O.C. (1987) RNA binding site of R17 coat protein. *Biochemistry (Mosc.)*, **26**, 1563–1568.

30. Cerretti,D.P., Wignall,J., Anderson,D., Tushinski,R.J., Gallis,B.M., Stya,M., Gillis,S., Urdal,D.L. and Cosman,D. (1988) Human macrophage-colony stimulating factor: Alternative RNA and protein processing from a single gene. *Mol. Immunol.*, **25**, 761–770.

31. Lim,F. and Peabody,D.S. (2002) RNA recognition site of PP7 coat protein. *Nucleic Acids Res.*, **30**, 4138–4144.

32. Brown,D., Brown,J., Kang,C., Gold,L. and Allen,P. (1997) Single-stranded RNA recognition by the bacteriophage T4 translational repressor, regA. *J. Biol. Chem.*, **272**, 14969–14974.

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

33. Sacerdot Christine, Caillet Joel, Graffe Monique, Eyermann Flore, Ehresmann Bernard, Ehresmann Chantal, Springer Mathias and Romby Pascale (2002) The Escherichia coli threonyl-tRNA synthetase gene contains a split ribosomal binding site interrupted by a hairpin structure that is essential for autoregulation. *Mol. Microbiol.*, **29**, 1077–1090.

34. Schlax,P.J., Xavier,K.A., Gluick,T.C. and Draper,D.E. (2001) Translational repression of the Escherichia coli alpha operon mRNA: importance of an mRNA conformational switch and a ternary entrapment complex. *J. Biol. Chem.*, **276**, 38494–38501.

35. Hattman,S., Newman,L., Murthy,H.M. and Nagaraja,V. (1991) Com, the phage Mu mom translational activator, is a zinc-binding protein that binds specifically to its cognate mRNA. *Proc. Natl. Acad. Sci.*, **88**, 10027–10031.

36. Wulczyn,F.G. and Kahmann,R. (1991) Translational stimulation: RNA sequence and structure requirements for binding of Com protein. *Cell*, **65**, 259–269.

37. Gregorio,E.D., Preiss,T. and Hentze,M.W. (1999) Translation driven by an eIF4G core domain in vivo. *EMBO J.*, **18**, 4865–4874.

38. Boutonnet,C., Boijoux,O., Bernat,S., Kharrat,A., Favre,G., Faye,J.-C. and Vagner,S. (2004) Pharmacological-based translational induction of transgene expression in mammalian cells. *EMBO Rep.*, **5**, 721–727.

39. Goldfless,S.J., Belmont,B.J., de Paz,A.M., Liu,J.F. and Niles,J.C. (2012) Direct and specific chemical control of eukaryotic translation with a synthetic RNA-protein interaction. *Nucleic Acids Res.*, **40**, e64.

40. Katz,N., Kaufmann,B., Cohen,R., Solomon,O., Atar,O., Yakhini,Z., Goldberg,S. and Amit,R. (2017) An in vivo binding assay for RNA-binding proteins based on repression of a reporter gene. *bioRxiv*, 10.1101/168625.

41. Lucks,J.B., Mortimer,S.A., Trapnell,C., Luo,S., Aviran,S., Schroth,G.P., Pachter,L., Doudna,J.A. and Arkin,A.P. (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci.*, **108**, 11063–11068.

42. Spitale,R.C., Crisalli,P., Flynn,R.A., Torre,E.A., Kool,E.T. and Chang,H.Y. (2013) RNA SHAPE analysis in living cells. *Nat. Chem. Biol.*, **9**, 18–20.

43. Flynn,R.A., Zhang,Q.C., Spitale,R.C., Lee,B., Mumbach,M.R. and Chang,H.Y. (2016) Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat. Protoc.*, **11**, 273–290.

44. Gott,J.M., Wilhelm,L.J. and Uhlenbeck,O.C. (1991) RNA binding properties of the coat protein from bacteriophage GA. *Nucleic Acids Res.*, **19**, 6499–6503.

45. Peabody,D.S. (1993) The RNA binding site of bacteriophage MS2 coat protein. *EMBO J.*, **12**, 595–600.

46. Lim,F. and Peabody,D.S. (2002) RNA recognition site of PP7 coat protein. *Nucleic Acids Res.*, **30**, 4138–4144.

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

47. Lim,F., Spingola,M. and Peabody,D.S. (1996) The RNA-Binding Site of Bacteriophage Qβ Coat Protein. *J. Biol. Chem.*, **271**, 31839–31845.

48. Medina,G., Juárez,K., Valderrama,B. and Soberón-Chávez,G. (2003) Mechanism of Pseudomonas aeruginosa RhlR Transcriptional Regulation of the rhlAB Promoter. *J. Bacteriol.*, **185**, 5976–5983.

49. Watters,K.E., Abbott,T.R. and Lucks,J.B. (2016) Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic Acids Res.*, **44**, e12–e12.

50. Johansson,H.E., Dertinger,D., LeCuyer,K.A., Behlen,L.S., Greef,C.H. and Uhlenbeck,O.C. (1998) A Thermodynamic Analysis of the Sequence-Specific Binding of RNA by Bacteriophage MS2 Coat Protein. *Proc. Natl. Acad. Sci.*, **95**, 9244–9249.

51. Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhoeffer,L.S., Tacker,M. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structures. *Monatshefte Für Chem. Chem. Mon.*, **125**, 167–188.

52. Washietl,S., Hofacker,I.L., Stadler,P.F. and Kellis,M. (2012) RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction. *Nucleic Acids Res.*, **40**, 4261–4272.

53. Zarringhalam,K., Meyer,M.M., Dotu,I., Chuang,J.H. and Clote,P. (2012) Integrating chemical footprinting data into RNA secondary structure prediction. *PloS One*, **7**, e45160.

54. Deigan,K.E., Li,T.W., Mathews,D.H. and Weeks,K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 97–102.

55. Ouyang,Z., Snyder,M.P. and Chang,H.Y. (2013) SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res.*, **23**, 377–387.

56. Levy,L., Anavy,L., Solomon,O., Cohen,R., Brunwasser-Meirom,M., Ohayon,S., Atar,O., Goldberg,S., Yakhini,Z. and Amit,R. (2017) A Synthetic Oligo Library and Sequencing Approach Reveals an Insulation Mechanism Encoded within Bacterial σ54Promoters. *Cell Rep.*, **21**, 845–858.

57. Sharon,E., Kalma,Y., Sharp,A., Raveh-Sadka,T., Levo,M., Zeevi,D., Keren,L., Yakhini,Z., Weinberger,A. and Segal,E. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.*, **30**, 521–530.

58. Dvir,S., Velten,L., Sharon,E., Zeevi,D., Carey,L.B., Weinberger,A. and Segal,E. (2013) Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, E2792-2801.

59. Weingarten-Gabbay,S., Elias-Kirma,S., Nir,R., Gritsenko,A.A., Stern-Ginossar,N., Yakhini,Z., Weinberger,A. and Segal,E. (2016) Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science*, **351**, aad4939.

60. Peterman,N. and Levine,E. (2016) Sort-seq under the hood: implications of design choices on large-scale characterization of sequence-function relations. *BMC Genomics*, **17**, 206.

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

61. Kinney,J.B., Murugan,A., Callan,C.G. and Cox,E.C. (2010) Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 9158–9163.

62. Breaker,R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.*, **4**.

63. Trachman,R.J. and Draper,D.E. (2017) Divalent ion competition reveals reorganization of an RNA ion atmosphere upon folding. *Nucleic Acids Res.*, **45**, 4733–4742.

64. Schwalbe,H., Buck,J., Fürtig,B., Noeske,J. and Wöhnert,J. (2007) Structures of RNA switches: insight into molecular recognition and tertiary structure. *Angew. Chem. Int. Ed Engl.*, **46**, 1212–1219.

65. Nguyen,G.T.D.T., Scaife,M.A., Helliwell,K.E. and Smith,A.G. (2016) Role of riboswitches in gene regulation and their potential for algal biotechnology. *J. Phycol.*, 10.1111/jpy.12416.

66. Schluenzen,F., Tocilj,A., Zarivach,R., Harms,J., Gluehmann,M., Janell,D., Bashan,A., Bartels,H., Agmon,I., Franceschi,F., *et al.* (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, **102**, 615–623.

67. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.

68. Ding,Y., Chan,C.Y. and Lawrence,C.E. (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.*, **32**, W135-141.

69. Kutchko,K.M., Sanders,W., Ziehr,B., Phillips,G., Solem,A., Halvorsen,M., Weeks,K.M., Moorman,N. and Laederach,A. (2015) Multiple conformations are a conserved and regulatory feature of the RB1 5' UTR. *RNA N. Y. N*, **21**, 1274–1285.

70. Halvorsen,M., Martin,J.S., Broadaway,S. and Laederach,A. (2010) Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet.*, **6**, e1001074.

71. Kertesz,M., Wan,Y., Mazor,E., Rinn,J.L., Nutter,R.C., Chang,H.Y. and Segal,E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.

72. Ding,Y., Tang,Y., Kwok,C.K., Zhang,Y., C Bevilacqua,P. and M Assmann,S. (2013) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, **505**.

73. Kudla,G., Murray,A.W., Tollervey,D. and Plotkin,J.B. (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science*, **324**, 255–258.

74. Kedde,M., van Kouwenhove,M., Zwart,W., Oude Vrielink,J.A.F., Elkon,R. and Agami,R. (2010) A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nat. Cell Biol.*, **12**, 1014–1020.

75. Ray,P.S., Jia,J., Yao,P., Majumder,M., Hatzoglou,M. and Fox,P.L. (2009) A stress-responsive RNA switch regulates VEGFA expression. *Nature*, **457**, 915–919.

FIGURES LEGENDS

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 1. Translational stimulation upon RBP binding in the 5' UTR. (a) Heatmap of the dose responses of the 5' UTR variants. Each response is divided by its maximal mCherry/mCerulean level, for easier comparison. Variants are arranged in order of increasing fold up-regulation. (b) Normalized K_{RBP}. Blue corresponds to low $K_{RBP,}$ while yellow indicates no binding. If there was no measureable interaction between the RBP and binding site, K_{RBP} was set to 1. (c) Bar graph showing expression fold change of each RBP—binding-site pair for all 11 binding sites, as follows: QCP-mCer (purple), PCP-mCer (green), MCP-mCer (red), and GCP-mCer (blue). Values larger and smaller than one correspond to up- and down-regulation, respectively. (Inset) Dose response function for MS2-U(-5)C with MCP at positions δ=-23,-26,-29, and -31 nt from the AUG. (d) Dose response functions for two strains containing the PP7-wt (blue) and PP7-USs (red) binding sites at δ=-29 nt from the AUG. Each data point is an average over multiple mCerulean and mCherry measurements taken at a given inducer concentration.(e) Structures schemes predicted by RNAfold for the 5' UTR and the first 134 nts of the PP7-wt and PP7-USs constructs (using sequence information only).

Fig. 2. *in vitro* SHAPE-Seq analysis for PP7-wt and PP7-USs strains. (a) *In vitro* reactivity analysis for SHAPE-Seq data (see Supp. Information) obtained for two constructs PP7-wt (blue) and PP7-USs (red) at δ=-29. Error-bars are computed by using boot-strapping re-sampling of the original modified and non-modified libraries for each strain (see sup. information) and also averaged from two biological replicates (b) Inferred *in vitro* structures for both constructs and constrained by the reactivity scores from (a). Each base is colored by its base pairing probability (red-high, yellow-intermediate, and white-low) calculated based on the structural ensemble (using RNAsubopt).

Fig. 3. *in vivo* SHAPE-Seq analysis for PP7-wt and PP7-USs strains. (a-b) Comparison of reactivity analysis computed using *in vivo* SHAPE-Seq data for the non-induced (a) and induced (b) states of PP7-wt (blue) and PP7-USs (red) at δ=-29. Error-bars are computed by using boot-strapping re-sampling of the original modified and non-modified libraries for each strain, and also averaged from two biological replicates (see sup. information). (c) Inferred *in vivo* structures for all 4 constructs and constrained by the reactivity scores show in (a-b). Each base is colored by its base pairing probability (red-high, yellow-intermediate, and white-low) calculated based on the structural ensemble (using RNAsubopt). For both the PP7-wt and PP7-USs the inferred structures show a distinct structural change in the 5' UTR as a result of induction of the RBP.

Fig. 4. Up-regulation is dependent on binding-site free energy. (a) Schematics for four sample structures computed with RNAfold (using sequence information only), where a short segment of the flanking region to the hairpin was mutated in each strain. Three structures contain the PP7-wt hairpin at δ=-29. (Top-left) CU-rich flanking colored in red. (Top-right) A-rich flanking colored in blue. (Bottom-left) original construct with "random" flanking sequence colored in green. (Bottom-right) PP7-wt hairpin encoded with a longer stem colored in yellow. (b) Variants containing 5 distinct hairpins with either a CU-rich (red), A-rich (blue), or original (green) flanking sequences upstream of the RBS. While basal levels are clearly affected by the presence of a strong CU-rich flanking sequence, the

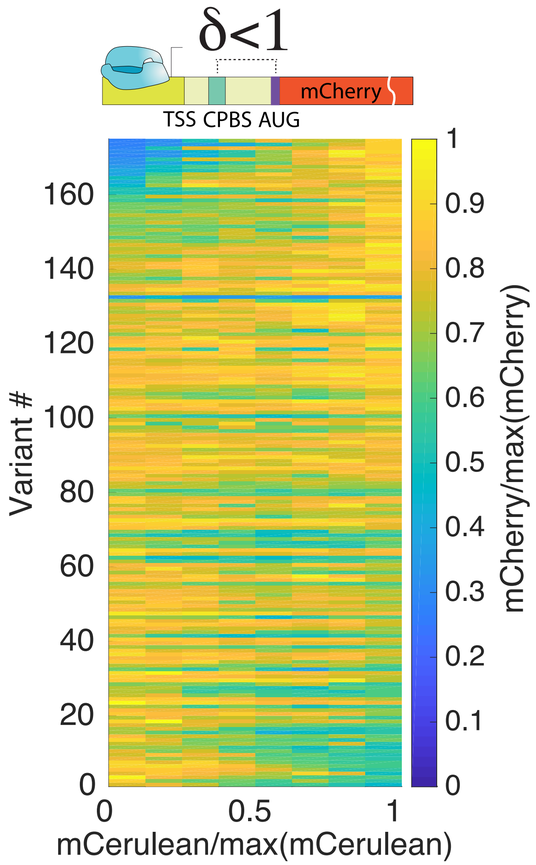
bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nature of the regulatory effect is apparently not determined by the sequence content of the flanking region. (c) Dose response functions for PP7-wt binding sites with an extra 3 (blue), 6 (magenta), and 9 (green) stem base-pairs are shown relative to the dose response for PP7-wt (red). (d) Basal levels and logarithm (base 2) of fold change for dose responses of all extended stem constructs with their corresponding RBPs (MCP or PCP).

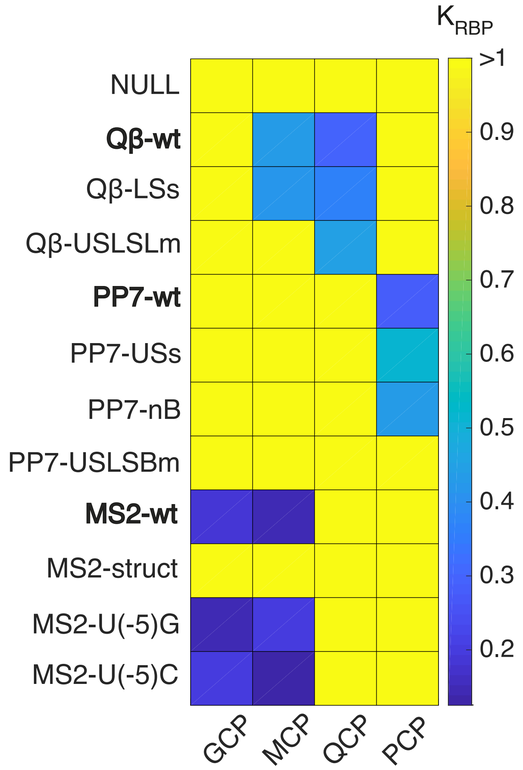
Fig. 5. mRNAs with a tandem of hairpins exhibit higher affinity to RBPs. (a) Schematic of the design of mRNA molecules with a single binding site at the 5' UTR and N-terminus. (b-d) Heatmap corresponding the dose-response function observed for MCP (b), PCP (c) and QCP (d). In all heatmaps, the dose-response is arranged in order of increasing mCherry rate of production, with the lowest-expressing variant at the bottom. The binding-site abbreviations are as follows: For MCP (b) and QCP (d), WT is MS2-wt, U(-5)G is MS2-U(-5)G, U(-5)C is MS2-U(-5)C, and Qβ is Qβ-wt. For PCP (c), WT is PP7-wt, nB is PP7-nB, Bm is PP7-LSLSBm, and USs is PP7-USs. (e) A sample fit using the cooperativity model (see Supp. Information). (f) Bar plot depicting the extracted cooperativity factors *w* for all the tandems that displayed either an up- or down-regulatory effect.

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

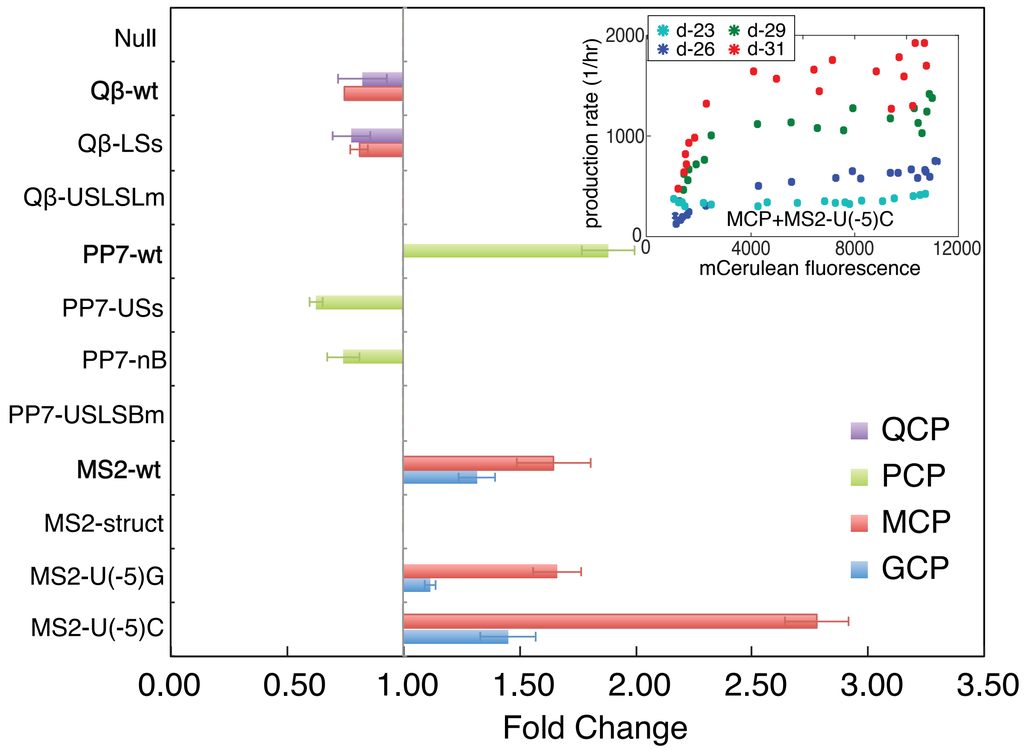
A



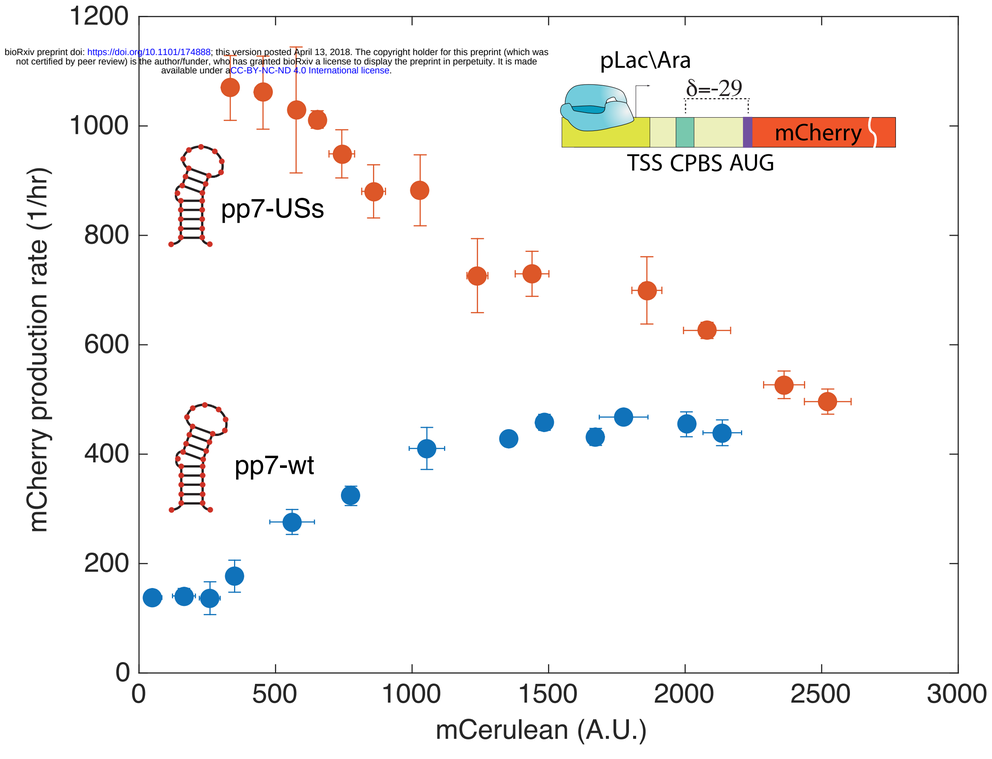
B



C



D



E

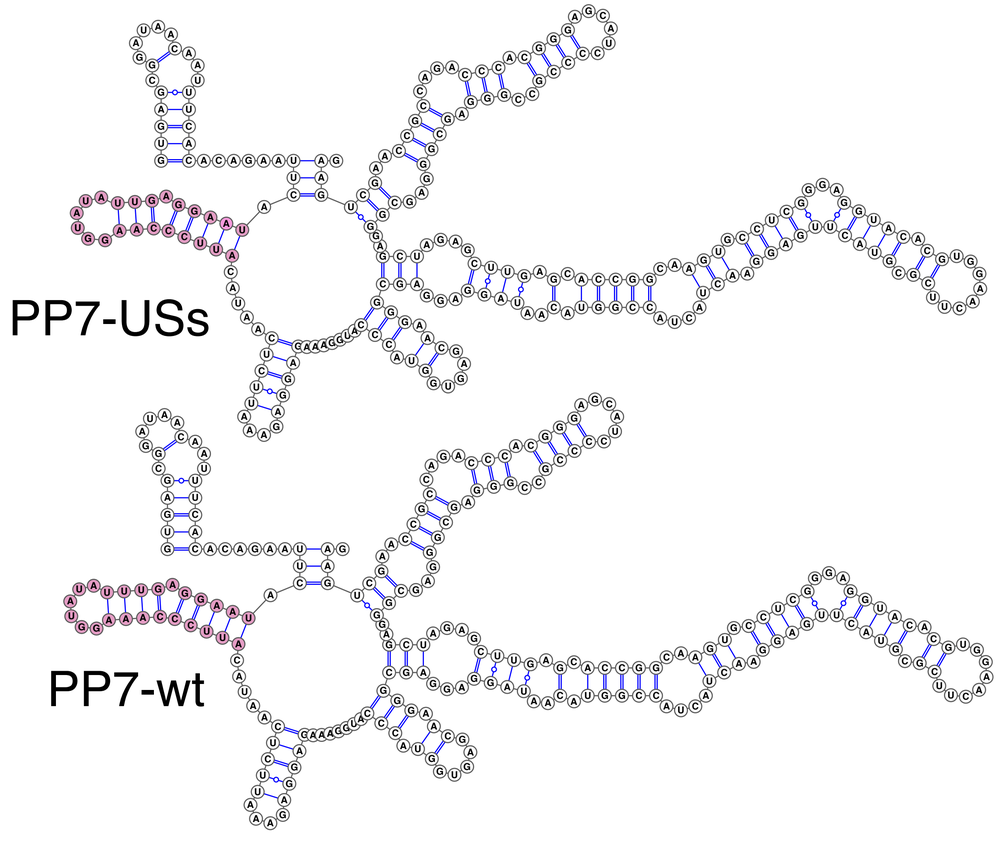
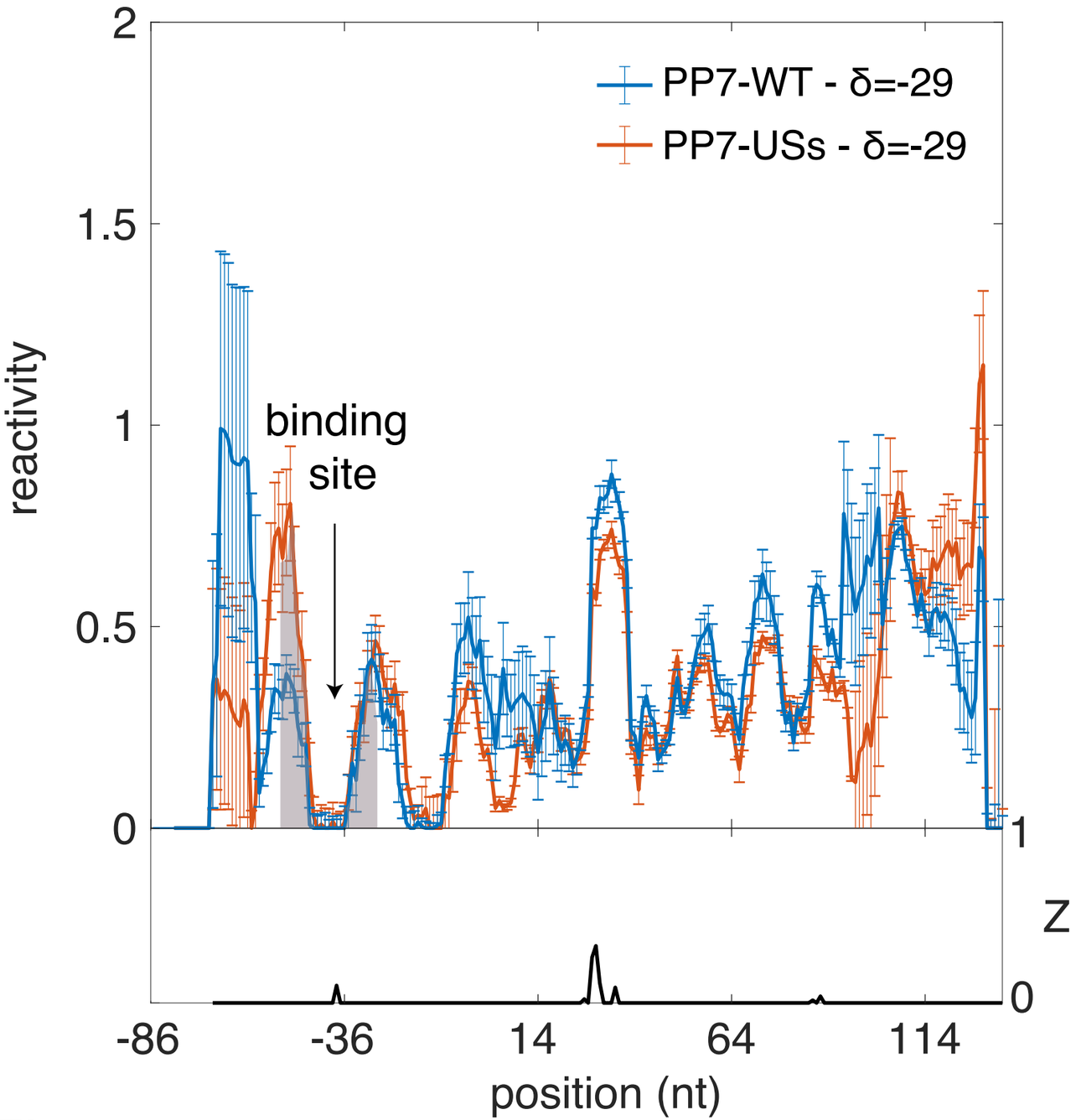


Figure 1

A


PP7-WT - δ=-29
PP7-USs - δ=-29
reactivity
binding site
2
1.5
1
0.5
0
Z
1
0
-86
-36
14
64
114
position (nt)

B

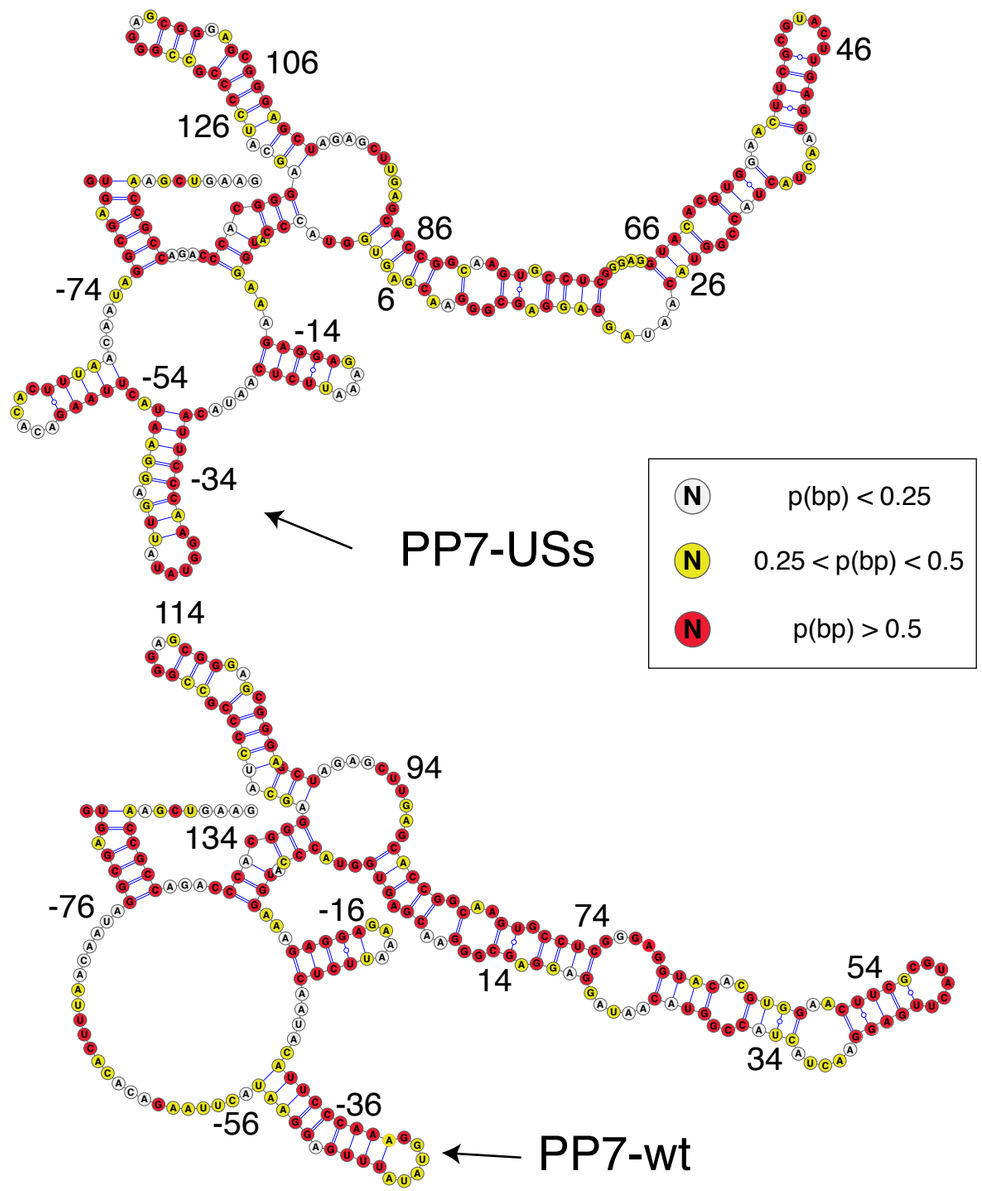

106
46
126
86
66
-74
6
26
-14
-54
-34
PP7-USs
114
N
p(bp) < 0.25
N
0.25 < p(bp) < 0.5
N
p(bp) > 0.5
94
134
-76
-16
74
14
54
34
-36
-56
PP7-wt

Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

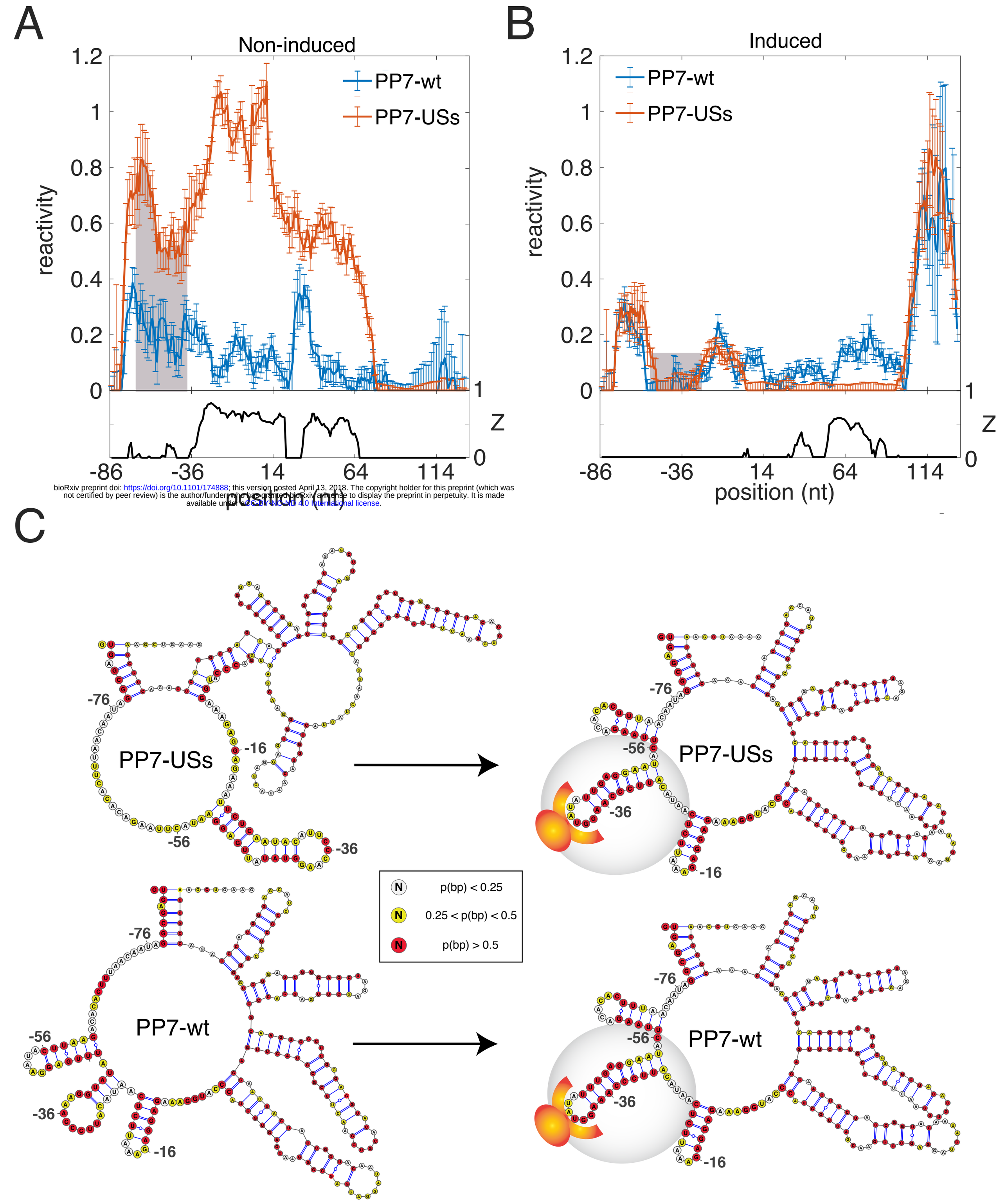
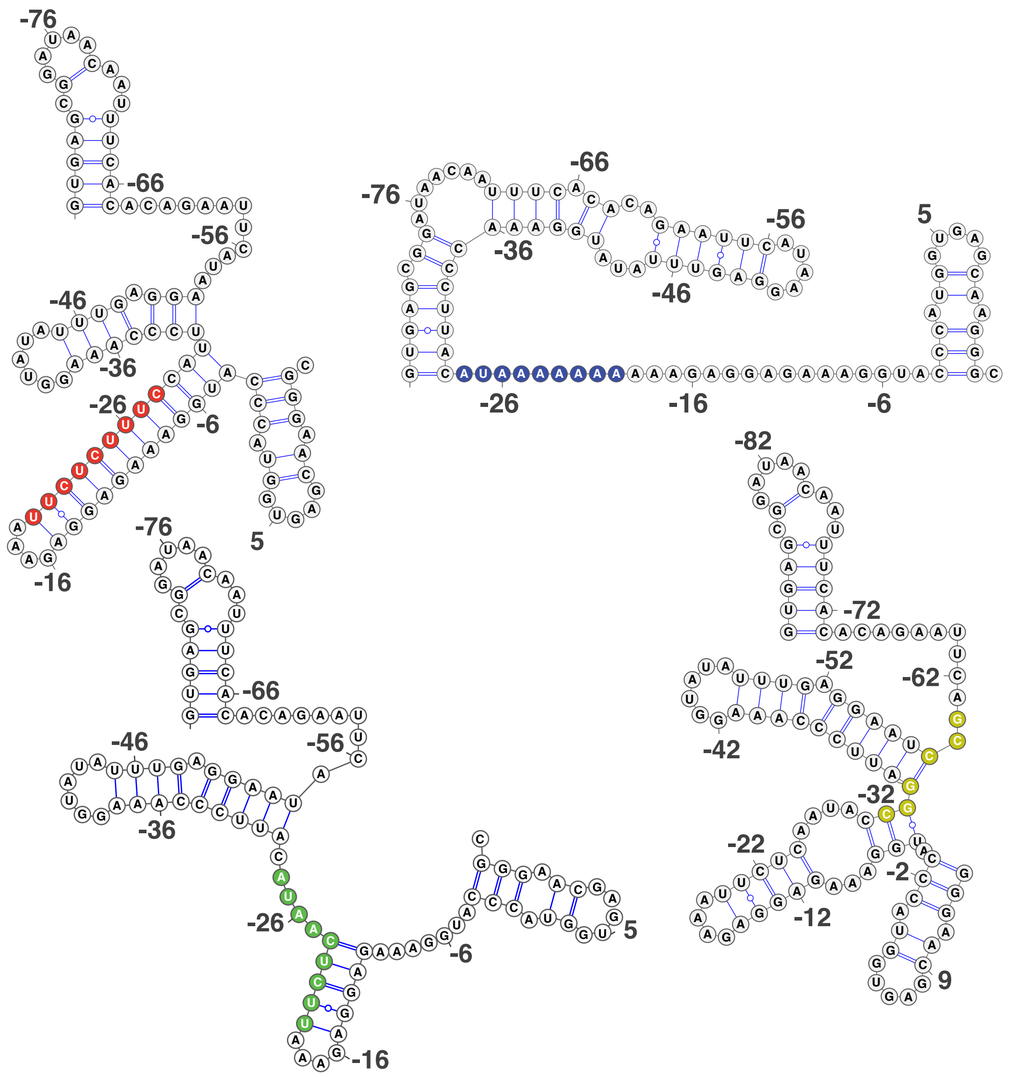
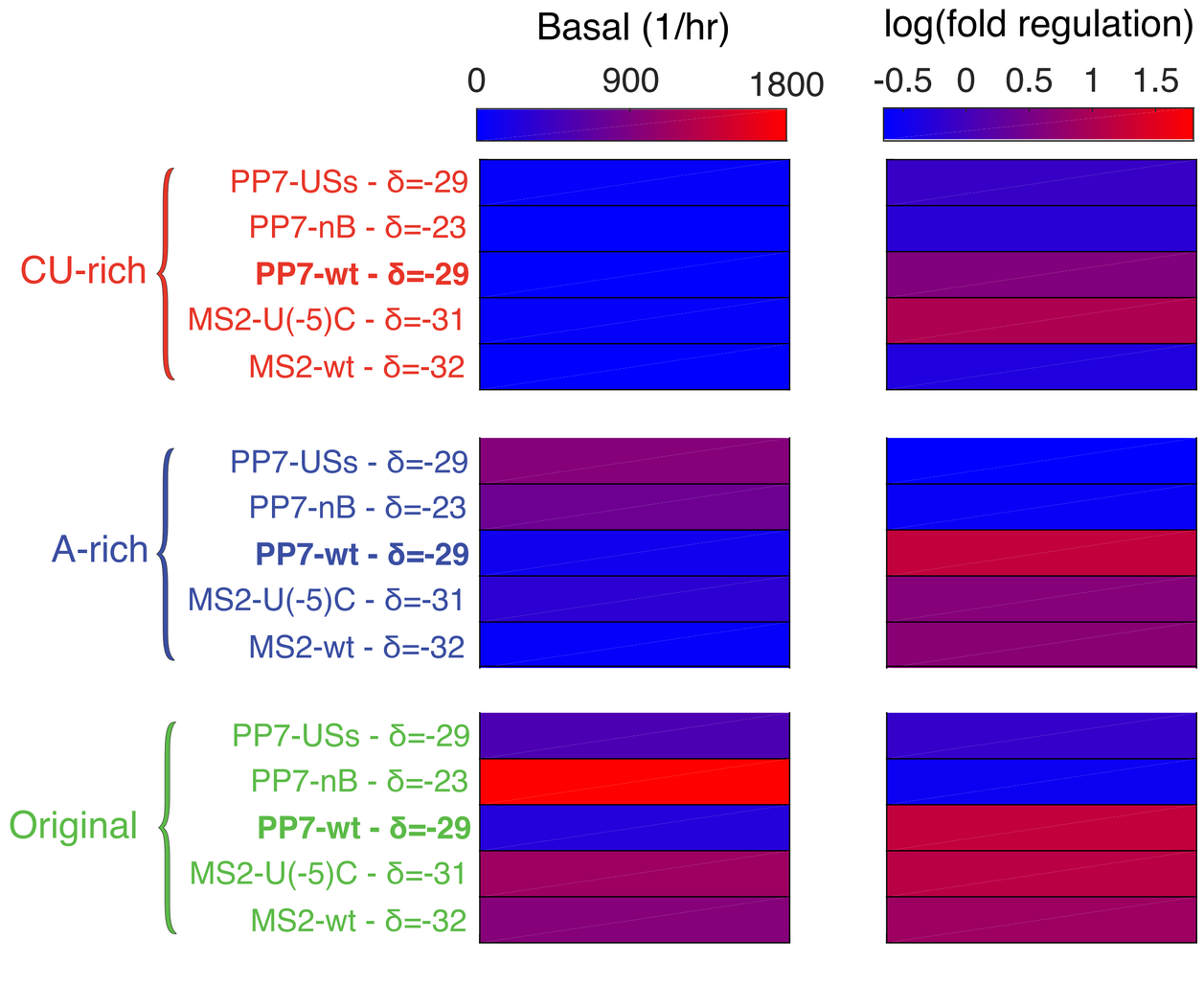


Figure 3

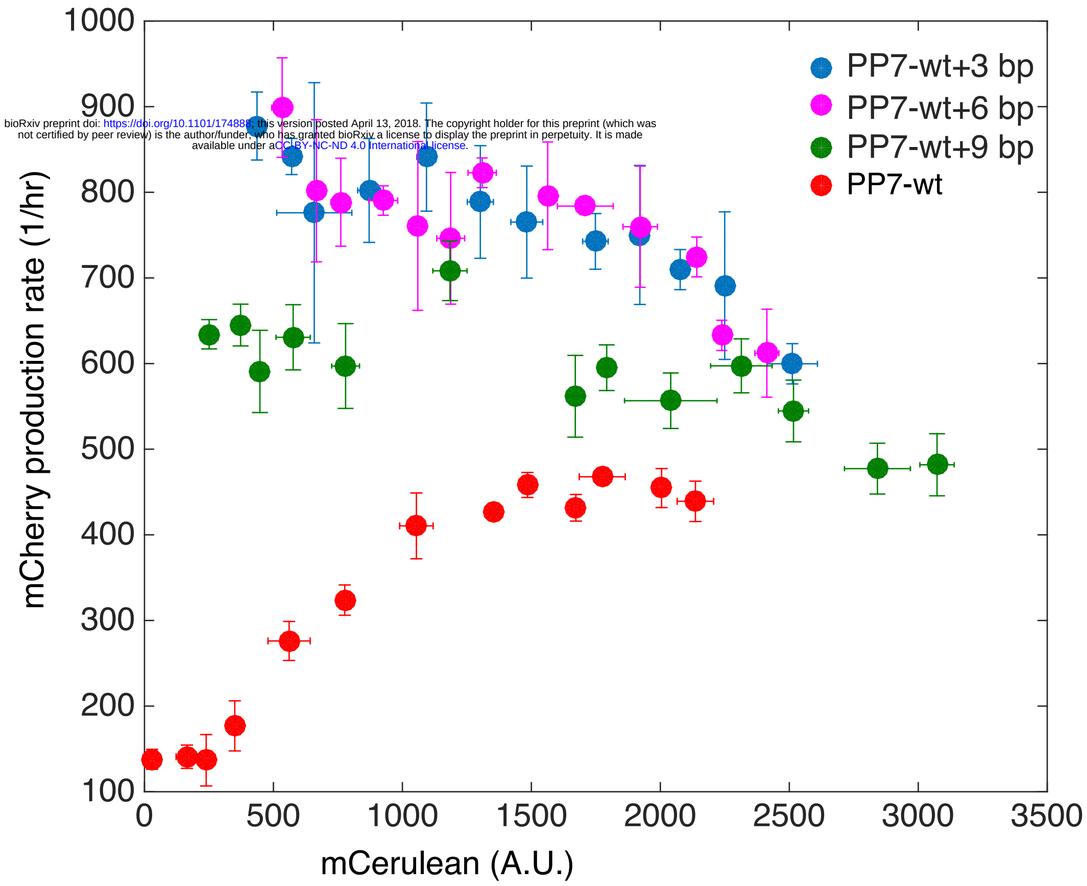
A



B



C



bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D

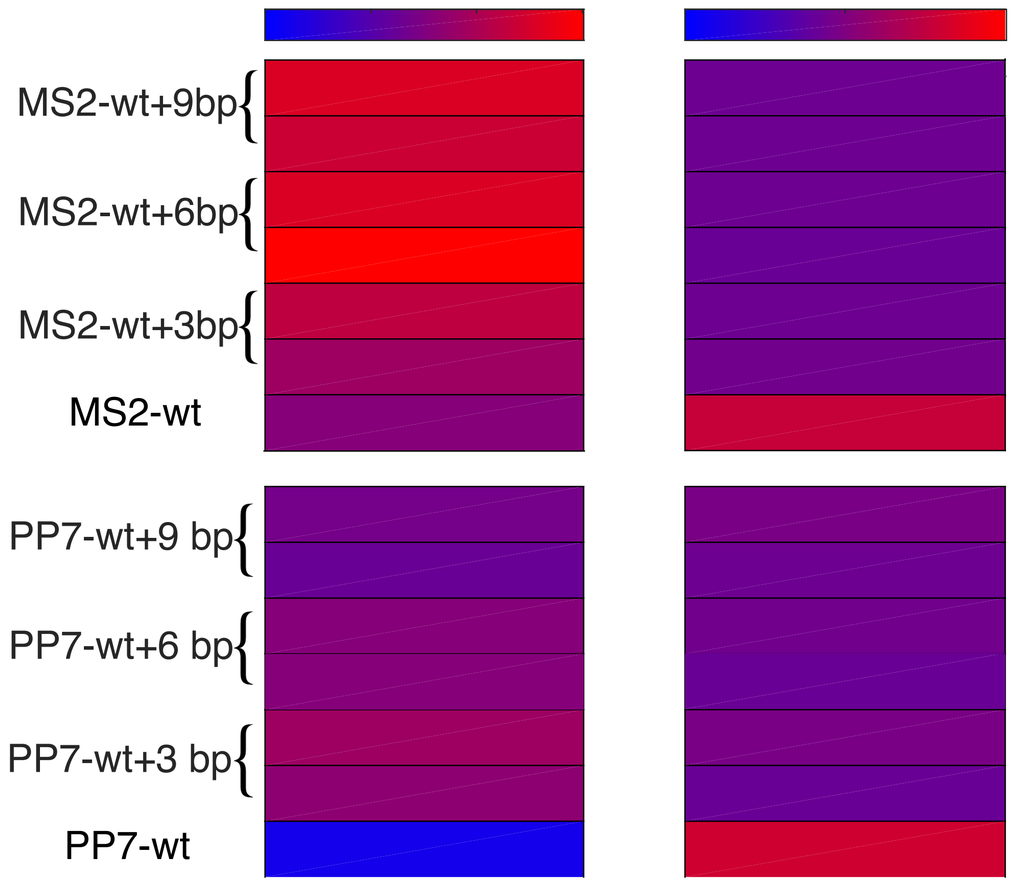
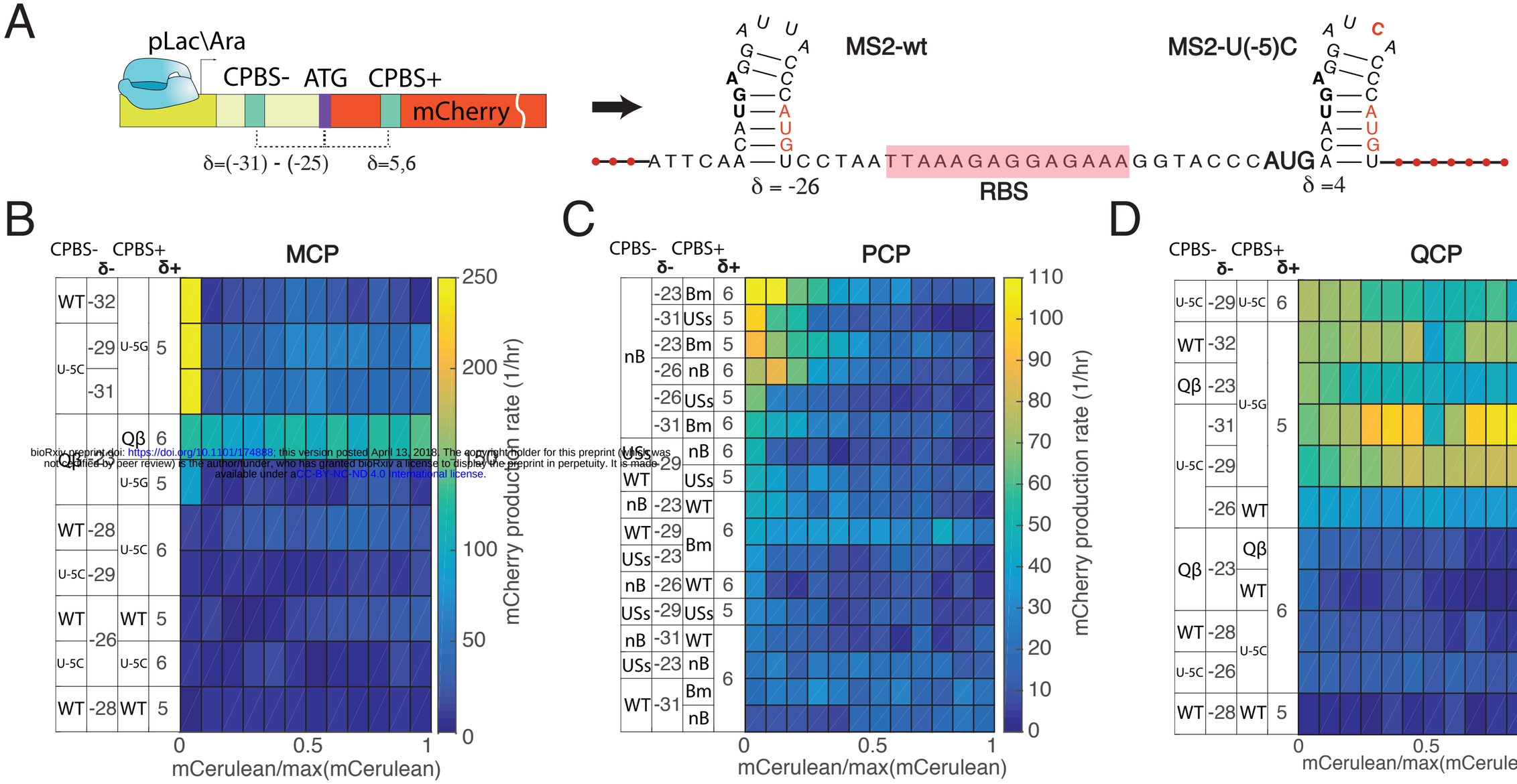
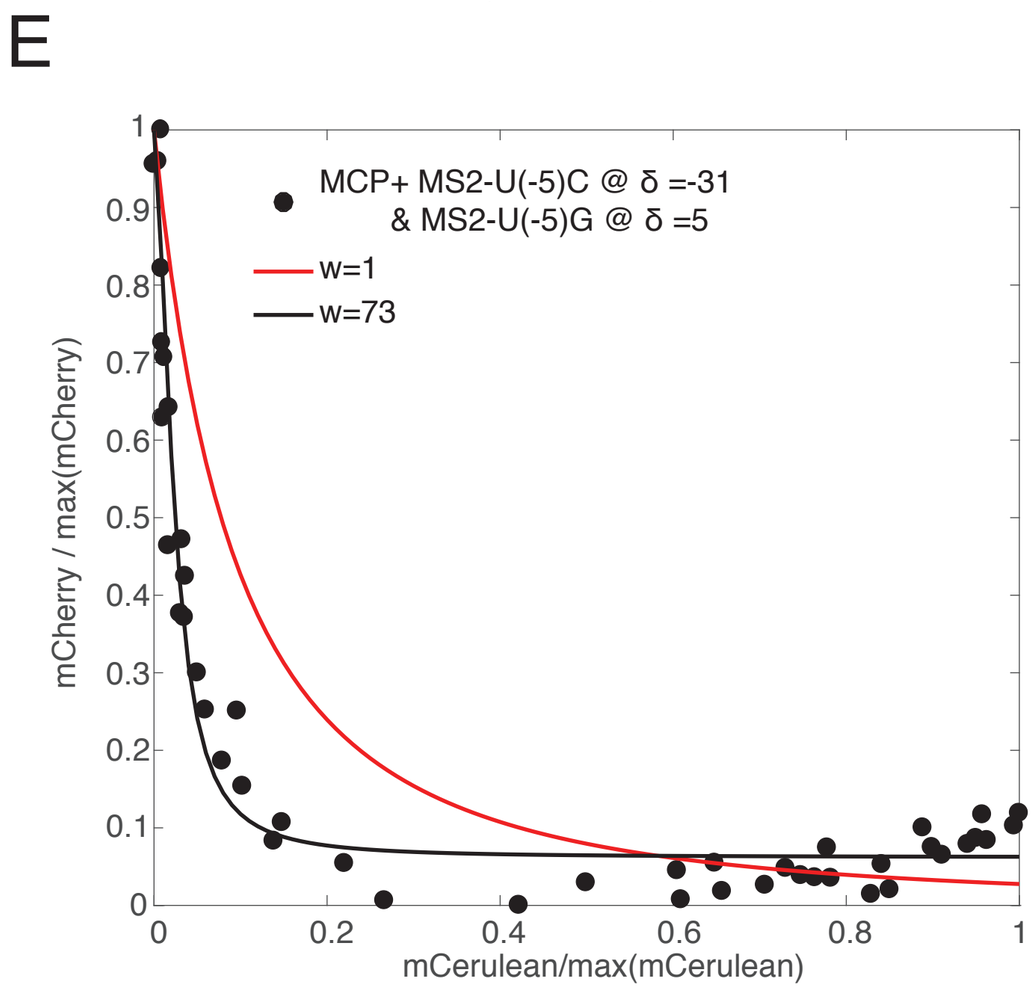


Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/174888; this version posted April 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.





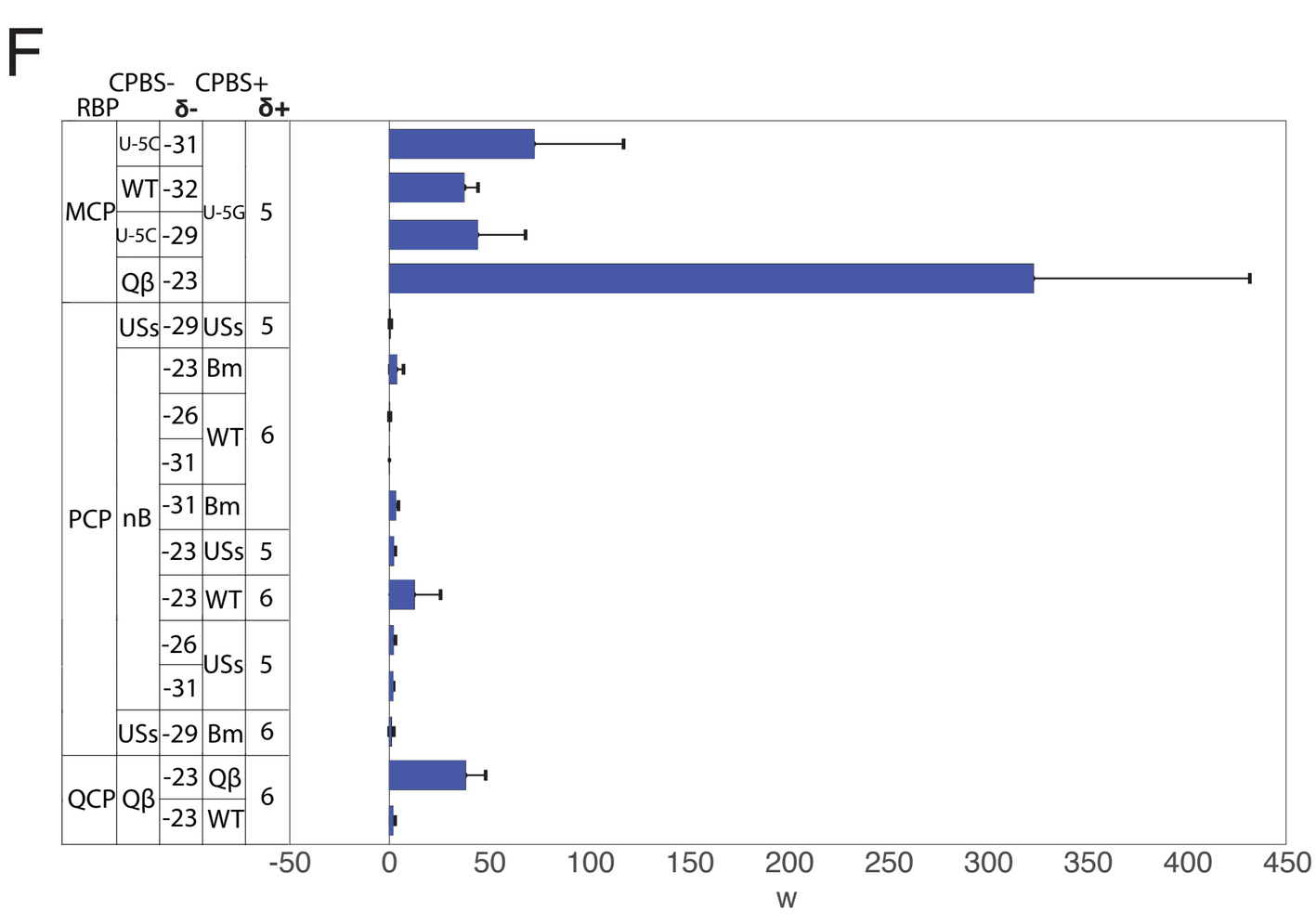


Figure 5