Synthetic protein-sensing riboswitches

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

Affiliation:

[1]Department of Biotechnology and Food Engineering, Technion - Israel Institute of Technology, Haifa, Israel 32000.

[2]Department of Computer Science, Technion - Israel Institute of Technology, Haifa, Israel 32000.

[3]School of Computer Science, Interdisciplinary Center, Herzeliya, Israel.

[4]Russell Berrie Nanotechnology Institute, Technion - Israel Institute of Technology, Haifa 32000.

*Correspondence to: roeeamit@technion.ac.il

Abstract

We study translational regulation by a 5' UTR sequence encoding the binding site of an RNA-binding protein (RBP), using a reporter assay and SHAPE-seq, in bacteria. We tested constructs containing a single hairpin, based on the binding sites of the coat RBPs of bacteriophages MS2, PP7, GA, and Qβ, positioned in the 5' UTR. With specifically-bound RBP present, either weak repression or up-regulation is observed, depending on the binding site. SHAPE-seq data for a representative construct exhibiting up-regulation indicate a partially-folded hairpin and a hypo-modified upstream flanking region, which we attribute to an intermediate structure that apparently blocks translation. RBP binding stabilizes the fully-folded hairpin state and thus facilitates translation. This indicates 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. Finally, the combination of two binding sites, positioned in the 5' UTR and N-terminus 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 behavior. This implies that RNA-RBP interactions can provide a platform for constructing gene regulatory networks that are based on translational, rather than transcriptional, regulation.

Keywords

Riboswitch, RNA-binding protein, RBP-RNA interactions, phage coat proteins, MS2, PP7, translational repression, hairpin, translation stimulation.

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 (Win and Smolke, 2008; Green et al., 2014; Xie et al., 2011; Wroblewska et al., 2015). RNA-based regulatory components have many advantages. Several RNA components have been shown to be functional in multiple organisms (Harvey et al., 2002; Suess et al., 2003; Desai and Gallivan, 2004; Buxbaum et al., 2015). RNA can respond rapidly to stimuli, enabling a faster regulatory response than that possible at the transcriptional level (Hentze et al., 1987; St Johnston, 2005; Saito et al., 2010; Lewis et al., 2017). The response range of RNA components can be very wide (Green et al., 2014). RNA molecules form a variety of stable secondary and tertiary structures, which support diverse functions. Moreover, a single RNA molecule may contain multiple functional structures, which enables modularity. For example, distinct sequence domains within a molecule (Lewis et al., 2017; Khalil and Collins, 2010) may target different metabolites or nucleic acid molecules (Isaacs et al., 2006; Werstuck and Green, 1998). All of these characteristics make RNA an appealing target for engineered applications (Hutvágner and Zamore, 2002; Rinaudo et al., 2007; Delebecque et al., 2011; Xie et al., 2011; Chen and Silver, 2012; Ausländer et al., 2014; Green et al., 2014; Sachdeva et al., 2014; Pardee et al., 2016).

Perhaps the most well-known class of RNA-based regulatory modules are ribsowitches (Werstuck and Green, 1998; Winkler and Breaker, 2005; Wittmann and Suess, 2012; Serganov and Nudler, 2013). 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 (Buskirk et al., 2004; Bayer and Smolke, 2005; Green et al., 2014), namely riboswitches that sense a nucleic acid molecule, provide a transcription-based feedback mechanism, which can result in slow temporal dynamics (Green et al., 2014). Implementing network modules using RNA-binding proteins (RBPs) could enable multicomponent connectivity without compromising response time or circuit stability.

Regulatory networks require both inhibitory and up-regulatory modules. The vast majority of known RBP regulatory mechanisms are inhibitory (Cerretti et al., 1988; Lim and

Peabody, 2002; Romaniuk et al., 1987; Brown et al., 1997; Sacerdot et al., 1998, Schlax et al., 2001). 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 (Hattman et al., 1991; Wulczyn and Kahmann, 1991). 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 (Gregorio et al., 1999; Boutonnet et al., 2004), adopting a riboswitch-like approach (Ausländer et al., 2014), and utilizing an RNA-binding version of the TetR protein (Goldfless et al., 2012). However, despite these notable efforts, RBP-based translational stimulation is still difficult to design in most organisms.

In recent work (Katz et al., 2017), we employed a synthetic biology and *in vivo* SHAPE-seq approach (Lucks et al., 2011; Spitale et al., 2013; Flynn et al., 2016) 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 (Saito et al., 2010). Our findings indicate that structure-binding RBPs [coat proteins from the bacteriophages GA (Gott et al., 1991), MS2 (Peabody, 1993), PP7 (Lim and Peabody, 2002), and Qβ (Lim et al., 1996)] can generate a range of translational responses, from previously-observed down-regulation (Saito et al., 2010) to, surprisingly, up-regulation. The mechanism for downregulation is most likely steric hindrance of the initiating ribosome by the RBP-hairpin 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.

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 one of eleven putative binding sites in the 5' UTR [see Supp. Fig. 1 and (Johansson et al., 1998; Lim and Peabody, 2002)]. 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 Table S1). Altogether, we designed 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. The normalized dose response functions for the 5' UTR constructs are plotted as a heatmap in Fig. 1a in order of increasing response, with the highest-repression variants depicted at the bottom. There are striking differences between the dose-response behaviors observed for the 5' UTR (Fig. 1a) and N-terminus configurations (Katz et al., 2017). First, the observed repression is significantly weaker for the 5' UTR hairpins, and at most amounts to about a factor of two reduction from basal levels. This is markedly different from the 20- to 50-fold effect observed for RBPs bound to hairpins in the N-terminus region. Second, some of the variants exhibit a distinct up-regulatory dose response of about 3- to 5-fold, which was not observed for the N-terminus configurations.

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 Supplementary Information). The resultant K_{RBP} values obtained for each RBP—binding-site pair are plotted as a heatmap in Fig. 1b. This heatmap is quite similar to the K_{RBP} heatmap for the N-terminus variants (Katz et al., 2017). 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 distinct dose-response signal in the 5' UTR in the presence of QCP, while no response was detected in the N-terminus configurations. Conversely, QCP and PCP, which displayed a binding affinity to MS2-wt, MS2-U(-5)C, and PP7-USLSBm in the N-terminus, respectively, did not generate a significant dose response to the 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.

To highlight the different types of dose responses 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. For example, in the inset of Fig. 1c we plot the individual dose response curves that were used to compute the red bar in the MS2-U(-5)C line of the plot. In particular, for MS2-U(-5)C, MS2-wt, and PP7-wt, together with MCP, GCP and PCP, respectively, nearly 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. 2a-b).

To better understand the different types of dose-response behavior in the 5' UTR, in Fig. 1d we focus on the dose-responses for the PCP binding site variants PP7-USs (blue) and PP7-wt (red), both at δ=-29. 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 Supp. Fig. 1 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 is encoded into the structure of the RNA molecule itself (see also Supp. Fig. 3c-d).

A structural-based mechanism for translational up-regulation that has been observed in natural bacterial systems is the destabilization of anti-Shine-Dalgarno/Shine-Dalgarno (aSD/SD) secondary structure by the Com protein (Hattman et al., 1991; Wulczyn and Kahmann, 1991). To check for this, 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 an aSD sequence (CU-rich – Fig. 1e-right schema, red) or an AU-rich segment that is highly unlikely to form stable secondary structures (Fig. 1e-right schema, blue

and Supp. Table 1). We then plot the basal expression level for 15 RBP—binding-site pairs, containing the original δ=-29 spacer (green, labeled random), the δ=-29 spacer containing the CU-rich aSD sequence (red), and the δ=-29 AU-rich spacer lacking the aSD sequence (blue). The data (Fig.1e – left heatmap) show that the constructs with a CU-rich flanking region exhibit low basal expression levels as compared with the other constructs, as previously observed (Levy et al., 2017), while the AU-rich and random flanking sequences do not seem to affect basal expression in a consistent fashion (red and blue). However, both the up-regulatory and down-regulatory dose responses persist independently of the flanking region content (Fig. 1e - right heatmap). Therefore, the up-regulatory response observed for four of the binding sites is not likely to be related to structures that sequester the RBS.

SHAPE-seq data is consistent with expression levels

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 with the AU-rich spacer to SHAPE-seq using 2-methylnicotinic acid imidazole (NAI) suspended in anhydrous dimethyl sulfoxide (DMSO), with DMSO as a non-modifiying control (see Supplementary Information). We constructed *in vitro* and *in vivo* SHAPE-seq libraries for both the fully-induced and non-induced RBP states to generate a total of eight different libraries for these constructs. We computed the SHAPE-seq read ratio, which is defined as the ratio of modified to unmodified normalized number of reads per position (see Supplementary Information). In Fig. 2, we plot the *in vitro* and *in vivo* SHAPE-seq read ratios for PP7-wt-induced (Fig. 2a), PP7-wt-non-induced (Fig. 2b), PP7-USs-induced (Fig. 2c), and PP7-USs-non-induced (Fig. 2d) cases. We chose to modify a segment that includes the entire 5' UTR (binding site, linker, and RBS regions marked in green, yellow, and purple respectively), and in addition another ~140 nt of the mCherry reporter gene (see inset schematics in Fig. 2a-d). An initial examination of the read ratio data reveals a noisy signal with little similarity between the *in vivo* and *in vitro* read ratios. In particular, an especially large deviation between the *in vitro* and *in vivo* read ratio signals is observed for all configurations in the region upstream of and including the binding sites (green box). For three of the four cases (Fig. 2a,c,d) there seems to be little similarity between the *in vivo* and *in vitro* signals in the linker, RBS, and reporter gene regions. Only for the non-induced PP7-wt, which unlike PP7-USs is translationally inactive (Fig. 2b),

does there seem to be good agreement between the *in vitro* and *in vivo* read ratio signals starting from the RBS and into the reporter gene sequence. This is in agreement with the previous observation, that less-structured mRNA transcripts are typically more highly expressed (Rouskin et al., 2014). The SHAPE-seq data indicate that molecular processes *in vivo* may be altering the RNA structure in the 5' UTR region, and especially near the binding site, independent of the induction state.

To better assess the structural footprint embedded within the read ratio signals, we employed a position-based correlation analysis. If the read ratios of two segments are correlated in a statistically significant manner (i.e. Pearson p-value <0.01 – see Supplementary Information), then the segments' propensities to be modified by NAI must be similar. This, in turn, indicates that their structures must also be similar. In addition, by applying correlation analysis to subsets of the full library length, we can obtain correlation levels for the different functional elements within the construct (binding site, linker, RBS). We tested the analysis on the *in vitro* SHAPE-seq data. Here, we expect the induced and non-induced read ratios to be identical for each of the constructs because there is no RBP present *in vitro*. When correlating the induced and non-induced *in vitro* samples for a single construct using either a 41 or 21 nucleotide running window (Supp. Figs. 3 and 4, respectively), statistically-significant correlation is observed for the entire read length, for both the PP7-wt and PP7-USs samples. Moreover, the confidence intervals for the correlation coefficient values are tightly constrained for both samples.

Next, we correlated the *in vitro* read ratios of the PP7-wt and PP7-USs constructs. Here, we expected statistically significant correlation for regions inside the 5' UTR. However, due to the two-nucleotide deletion in the PP7-USs binding site as compared with the PP7-wt, no further statistically significant correlation is expected for downstream segments. Instead, statistically significant correlation should emerge in the downstream segment when correlating segments that are shifted by two nucleotides from one another. In Fig. 3a-b, we plot the Pearson coefficient and p-value results for the 21-nucleotide running window cross-correlation of the read ratios of PP7-wt and PP7-USs that is carried out with 5 different offsets: -2 nucleotides (blue), -1 nucleotide (red), no offset (orange), +1 nucleotide (purple), and +2 nucleotides (green). The data show that statistically significant correlation is measured for the no-offset case only in the 5' UTR region (-66 to -10), while for segments further downstream (position > 16), statistically significant

correlation is observed only for the -2-nucleotide shift, as expected. Interestingly, in the intervening region that spans positions -11 to +14 nt, no statistically significant correlation was observed, which indicates that the two constructs may have different structures in this region.

We next computed the correlation coefficient function between the *in-vitro* and *in-vivo* read ratios, for each of the induced and non-induced samples of PP7-wt and PP7-USs, using a 21 nucleotide running window (Fig 3c-d). For sequence elements that are upstream of the RBS (green and yellow shaded regions, position < -16), no statistically-significant correlation is observed between the *in vitro* and *in vivo* samples for any of the constructs. For elements that are downstream to the RBS, no statistically-significant correlation is observed for the PP7-USs constructs. However, statistically-significant correlation is observed in this region for the PP7-wt constructs, with especially low p-values measured for the PP7-wt non-induced construct (Fig. 3c – red line). These results are consistent with an interpretation that the degree of correlation between the *in vitro* and *in vivo* SHAPE-seq signals depends on the density of ribosome occupancy of the RNA, and the frequency of translation initiation events. Thus, the construct whose RNA structures *in vivo* and *in vitro* are most similar is also the lowest-expressing construct. Finally, the statistically-significant correlation observed for the 5' UTR for both constructs *in vitro* (Fig. 3a-b), together with the lack of statistical significance when correlating the *in vitro* and *in vivo* samples for the 5' UTR (Fig. 3c-d), indicate that the presence of cellular factors, which may bind the 5' UTR, are likely associated with structural changes to the molecule *in vitro*.

RBP-induced up-regulation is related to structural changes in the 5' UTR

Given the lack of statistically-significant insight obtained for the 5' UTR region that encompasses the binding site using the *in vitro* and in *vivo* correlation analysis, we wondered whether additional correlation computations between the various *in vivo* read ratios could shed light on the structure in this region. We computed the positioned-based correlation coefficient function for the 5' UTR segments (-61 to -31 nt) between the non-induced (Fig. 4a-blue) and induced (Fig. 4a-red) PP7-wt and PP7-USs ratios, using a higher-resolution eleven-nucleotide running window. For the induced read ratios, we observe statistically-significant correlation between the induced states of the PP7-wt and PP7-USs constructs (p-value < 0.01 – dashed line) in a sequence segment which encodes both the putative A-bulge and the 6-nucleotide loop (green

and yellow shaded regions, respectively). Conversely, a much narrower region of statistically-significant correlation, which encompasses only part of the 6-nucleotide loop, is observed for the non-induced states. In particular, the regions encoding the lower stem and the flanking region upstream to the binding site do not exhibit statistically-significant correlation, despite being identical in sequence. Next, we computed the correlation between the induced and non-induced states for each construct, again using an eleven-nucleotide running window (Fig. 4b). For PP7-USs, a statistically-significant result appears in the region encompassing the lower stem, A-bulge, upper stem, and loop, while for the PP7-wt constructs, statistically-significant correlation appears in the loop region only.

When examining the different cross-correlation results for the non-induced PP7-wt construct, we note that for this construct only, no statistically-significant cross-correlation was observed for the upstream flanking region and lower stem, in any of the correlation data. To try to understand the features that make this segment unique, we plot the read ratio data for both PP7-wt (Fig. 4c) and PP7-USs (Fig. 4d) over a narrow -66 to -26 position range. We notice that the flanking region at positions -59 to -51 nt of PP7-wt is less modified in the non-induced state (red) than in the induced state (blue). For the PP7-USs construct, a relatively consistent level of modification over this range is observed for both the induced and non-induced states. Together, the SHAPE-seq data and associated correlation analysis suggest that in the non-induced state of the PP7-wt construct, the structure of the 5' UTR may include a partially-folded binding site (upper stem and loop) adjacent to a modification-protected upstream flanking region, lower stem, and A-bulge. Therefore, it is possible that the differing regulatory phenomena observed for (up-regulating) PP7-wt and (down-regulating) PP7-USs may originate in structural differences between the constructs' non-induced states, that are localized to the flanking region upstream and immediately adjacent to the binding sites.

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

Based on both the SHAPE-seq and expression level measurements, we propose a kinetic state model to explain the observed up- and down-regulatory dose responses (Fig. 5a). A kinetic-state-model (Ritort, 2007; Alemany et al., 2012) combines free energy computations for metastable kinetic states with transition processes between those states to assess the dynamical behavior of the non-equilibrium system. In the non-induced state of the PP7-wt strain (Fig. 5a,

top schemas), the upstream flanking region, together with the first several nucleotides of the binding site, seem to form a partially-folded structure (Fig. 5a, top-left). This intermediate state is characterized by two SHAPE-seq signatures: the first is hypo-modification of the upstream flanking region (position -59 to -51 nt - upstream of the binding site), which supports a modification-protected structure (e.g. inactive and bound 30S, as shown in Fig. 5a-top-left - left structure), and the second is a possibly-stabilizing short stem and loop structure in the -44 to -38 nt region. When the mRNA is in this intermediate state, it is apparently translationally inactive, based on the expression level measurements. To explain why the molecule is predominantly in this translationally-inactive intermediate state, we propose that there are large energy barriers, characterized by slow transition rates, separating the inactive state from the adjacent states (k_{-2} and k_3 - Fig. 5a, top-left arrows). In this scenario, it is plausible that one of the adjacent states is the fully-folded-hairpin, translationally-active state (Fig. 5a, top-left - right structure), but that the slow transition rate (k_3 in this case) effectively prevents the system from accessing this state.

For the non-induced case of PP7-USs, we hypothesize that the energy barrier between the intermediate and the folded states is smaller than for the PP7-wt strain. This is a reasonable assumption as PP7-USs has one less base-pair in the upper stem (Fig. 5a - top-center - left structure), thus reducing the stability of this structure. In kinetic terms, this implies that for this molecule, the rates k_{-2} and k_3 could be significantly faster, thereby resulting in a more rapid transition out of the inactive intermediate state. Thus, such a molecule's structure is likely to fluctuate between translationally-active, partially-unfolded, and fully-folded intermediate states. Finally, in the induced state (Fig. 5a, bottom), based on the SHAPE-seq results (Fig. 4a,c,d), RBP binding is assumed to stabilize the fully-folded binding-site state as compared with the other states, for both constructs. Since both bound structures are expected to be preferred to the other metastable states, dose responses for the PP7-wt and PP7-USs are expected to exhibit approximately the same reporter level at maximal induction, which is what was observed (Fig. 1d).

Based on this model, we hypothesized that we might change the direction of the PP7-wt dose response function from up-regulating to down-regulating by lowering the free-energy of the fully-folded state, which could also lower the energy barrier separating the fully-folded and intermediate states, thus increasing transition rate k_3 (Fig. 5a - top-right). We reasoned that such

an effect could be achieved by increasing the length of the bottom stem of the binding site, leading to a more stable structure (characterized by a lower free energy). To test this hypothesis, 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 (Supp. Fig. 5). In all twelve cases studied, the free energy associated with the binding sites as computed by NUPACK (Zadeh et al., 2011) was reduced as compared with the wild-type binding sites (Supp. Fig. 5). When examining the dose-response functions for all the configurations (Fig. 5b-c), 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, 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 (Fig. 5b - inset).

A tandem of binding sites can exhibit cooperativity

Finally, we 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 N-terminus region of the reporter ($\delta_{+}>0$) (Fig. 6a), in the presence of RBP. To do so, we constructed an additional 28 variants with combinations of MS2, PP7, and Qβ binding sites. In Fig. 6b-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. 6b), 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. 6c), a similar picture emerges, with several variants exhibiting a strong dose-response repression signature, while no regulatory effect is 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. 5b-c, 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. 6d, 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 previously (Katz et al., 2017). 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 Supplementary Information for fitting model) for repressive tandem constructs in the presence of their corresponding cognate RBPs. In Fig, 6e, 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 $50<w<80$. In Fig. 6f 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. 6 and Table S4 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.

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. Here, we built on the design introduced by (Saito et al., 2010) 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 bacterial gene N-terminus regions (Katz et al., 2017). 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, a previously-unobserved translational stimulation phenomenon due to RBP binding and change in mRNA structure, and an 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 N-terminus).

Our expression level data on the single-binding-site constructs, SHAPE-seq data, and follow-up experiments suggest that structure upstream of the binding site correlates with a blocked-translation-initiation state. The blocked translation can be alleviated by RBP binding, which may simultaneously stabilize the binding-site hairpin and destabilize the alternative ``blocking" mRNA structure, thus leading to up-regulation. What type of structure can generate such an effect? Our combined expression level and SHAPE-seq results seem to rule out the release of a sequestered RBS, which has been reported before as a natural mechanism for translational stimulation (Hattman et al., 1991; Wulczyn and Kahmann, 1991), and instead support a model of 30S blocking, or trapping, that is alleviated in the RBP-bound state. Such a state could be a consequence of tertiary RNA interactions such as hairpin-hairpin (Holbrook, 2005; Lehnert et al., 1996), A-minor (Geary et al., 2011), kink-turn (Huang and Lilley, 2016), or pseudoknot structures (Schlax et al., 2001; Ehresmann et al., 2004). Either way, what we seem to be observing is two types of RNA structures in the 5' UTR. The first is a fully-folded binding site which seems to stabilize an unbound RBS, thereby leading to a high basal expression level.

The second structure contains a partially-folded binding site and a modification-protected upstream flanking region, which are associated with complete suppression of translation. A bound RBP stabilizes both the binding-site structure and the unbound RBS, but also seems to interfere with translation. Consequently, the up-regulation phenomenon that we observed is a transition from a strongly-repressing structure to a weakly-repressing one, that occurs upon RBP binding. Thus, the constructs that exhibit up-regulation upon RBP induction are, to the best of our knowledge, the first reported examples of synthetic protein-sensitive riboswitches.

How difficult is it to design this up-regulatory phenomenon *de novo*? We emphasize that our results were obtained in *E. coli*, and that the proposed mechanism for the non-induced RBP translation rate involves possible trapping of the 30S ribosomal subunit, 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. This indicates that it may be a generic 5' UTR mechanism, that could be extended to other structure-binding RBPs. It is possible that the free-energy score of the binding sites can be used as a predictor for up-regulation. The binding sites which exhibited up-regulation were all scored in the 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 (see Supp. Fig. 5) exhibited high basal level expression and subsequent down-regulation. Thus, while intermediate stability is not necessarily a guarantee for ascertaining 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 (it is a good indicator for 4 of 7 sites in our case).

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. 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.

Supplemental Information

6 Supplementary Figures

4 Supplementary Tables

1 Supplementary Data set

Acknowledgments

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

Author contribution

NK designed and carried out the expression level experiments and analysis for all constructs. BK and RC designed and carried out the SHAPE-seq experiments. OS and ZY analyzed the SHAPE-seq data. SG and OA assisted and guided the experiments and analysis. RA supervised the study. RA, NA, BK, and SG wrote the manuscript.

Figure Captions

Figure 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 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 red 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 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 (red) and PP7-USs (blue) binding sites at δ=-29 nt from the AUG. (e) Adding CU-rich (red), AU-rich (blue), or random (green) flanking sequences upstream of the RBS. While basal levels are clearly affected

by the presence of a strong CU-rich flanking sequence, the nature of the regulatory effect is apparently not determined by the strength of the aSD.

Figure 2: SHAPE-seq analysis for PP7-wt and PP7-USs strains

SHAPE-seq analysis (see Supplementary Information) for two constructs (PP7-wt and PP7-USs) at δ=-29 with AU-rich flanking sequence in the fully induced and non-induced states (see schematics inside panels). In all panels we present the *in vivo* read ratio in color: blue for induced (a,c), red for non-induced (b,d), and the associated *in vitro* read ratio in gray. (a) PP7-wt induced (blue). (b) PP7-wt non-induced (red). (c) PP7-USs induced (blue). (d) PP7-USs non-induced (red).

Figure 3: Correlation reveals structural information in SHAPE-seq data

(a) Cross correlation analysis for the *in vitro* SHAPE-seq read ratios of PP7-wt and PP7-USs. The correlation coefficient at each point is an average of 4 separate computations carried out on the four *in vitro* pairs (PP7-wt-induced/PP7-US-induced, PP7-wt-non-induced/PP7-US-induced, PP7-wt-induced/PP7-US-non-induced, PP7-wt-non-induced/PP7-US-non-induced). Correlation was computed at 5 different offsets *n* from PP7-wt, defined as a shift in *n* nucleotides of the PP7-USs read ratio relative to the PP7-wt read ratio, as follows: (blue) -2 nt offset, (red) -1 nt offset, (yellow) no offset, (magenta) +1 offset, (green) +2 offset. (b) p-value plot for the correlation values shown in (a) showing a significance of correlation coefficient values > 0.75. (c-d) Cross-correlation (top) and associated p-values (bottom) for a 21 nt running window for the *in vivo* and associated *in vitro* read ratios, as follows: (c) PP7-wt non-induced (red) and PP7-wt induced (blue). (d) PP7-USs non-induced (red) and PP7-USs induced (blue). Dashed lines correspond to p-value=0.01. We consider results with p-value<0.01 and correlation>0.75 to be statistically-significantly correlated.

Figure 4: Comparison of SHAPE-seq data for PP7-wt and PP7-USs constructs with δ=-29

(a) Position-based correlation function for the SHAPE-seq read ratios of PP7-wt and PP7-USs using a 10-nt running window, for the non-induced (red) and induced (blue) cases. (b) Position-based correlation function for the induced and non-induced ratio signals of PP7-wt (red) and PP7-USs (blue), using a 10-nt running window. Correlation coefficient values above dashed

(dashed-dot) line correspond to P-value < 0.05 (P-value < 0.1). (c-d) *In vivo* (solid line) SHAPE-seq ratios for PPT-wt (c) and PP7-USs (d) induced (blue) and non-induced modification (red) states. Putative single-stranded regions are highlighted in yellow and green on both structure (insets) and ratio signals. *In vitro* ratios are shown by gray dashed lines.

Figure 5: Up-regulation is dependent on binding-site free energy

(a) Schematic model for translation stimulation. Sizes of illustrated states reflect their probabilities. (Top-left) Free-energy landscape of the PP7-wt structure showing a stable intermediate state with a trapped 30S subunit, and a large energy barrier between the intermediate state and the folded-binding-site state. Arrows correspond to the transition rates between the different kinetic states. Thicker arrows correspond to faster rates. In this case, k_2 is a fast process, which traps the molecule in a metastable state, with slow rates k_{-2} and k_3 to transition out of it. (Top-middle) The stable intermediate state is destabilized for the PP7-USs binding site, resulting in a semi-stable partially-folded-binding-site state. This is depicted by the arrows with fast transition rates k_{-2} and k_3, that generate an unstable intermediate kinetic state. (Top-right) Increasing the bottom stem length (PP7-wt+3) reduces the energy barrier between the intermediate and folded-binding-site states, thus destabilizing the intermediate state. This is depicted by the fast k_2, k_3, and slow k_{-2}, k_{-3} rates. (Bottom) Energy landscape with RBP bound. Bound RBP reduces the energy barrier between intermediate and folded-binding-site states. (b) Fold change for binding sites with an extra 3 (blue), 6 (magenta), and 9 (green) stem base-pairs are shown relative to the fold change for PP7-wt (red). (Inset) K_{RBP} calculated for all constructs shown in Supp. Fig. 5, with corresponding RBPs (MCP or PCP). (c) Basal levels and logarithm of fold change for dose responses of all constructs with their corresponding RBPs (MCP or PCP).

Figure 6: 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 Supplementary Information). (f) Bar plot depicting the extracted cooperativity factors *w* for all the tandems that displayed either an up- or down- regulatory effect.

Bibliography

Alemany, A., Mossa, A., Junier, I., and Ritort, F. (2012). Experimental free-energy measurements of kinetic molecular states using fluctuation theorems. Nat. Phys. *8*, 688–694.

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.

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

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.

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

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.

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

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.

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.

Ehresmann, C., Ehresmann, B., Ennifar, E., Dumas, P., Garber, M., Mathy, N., Nikulin, A., Portier, C., Patel, D., and Serganov, A. (2004). Molecular Mimicry in Translational Regulation: The Case of Ribosomal Protein S15. RNA Biol. *1*, 65–72.

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.

Geary, C., Chworos, A., and Jaeger, L. (2011). Promoting RNA helical stacking via A-minor junctions. Nucleic Acids Res. *39*, 1066–1080.

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.

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.

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.

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

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

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.

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.

Holbrook, S.R. (2005). RNA structure: the long and the short of it. Curr. Opin. Struct. Biol. *15*, 302–308.

Huang, L., and Lilley, D.M.J. (2016). The Kink Turn, a Key Architectural Element in RNA Structure. J. Mol. Biol. *428*, 790–801.

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

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

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.

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 168625.

Lehnert, V., Jaeger, L., Michele, F., and Westhof, E. (1996). New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme. Chem. Biol. *3*, 993–1009.

Levy, L., Anavy, L., Solomon, O., Cohen, R., Ohayon, S., Atar, O., Goldberg, S., Yakhini, Z., and Amit, R. (2017). Short CT-rich motifs encoded within σ54 promoters insulate downstream genes from transcriptional read-through. BioRxiv 086108.

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.

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

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

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.

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.

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

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.

Ritort, F. (2007). Nonequilibrium Fluctuations in Small Systems: From Physics to Biology. In Advances in Chemical Physics, S.A. Rice, ed. (John Wiley & Sons, Inc.), pp. 31–123.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature *505*, 701–705.

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.

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.

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.

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

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.

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

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.

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

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

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

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

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.

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

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.

Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M., and Pierce, N.A. (2011). NUPACK: Analysis and design of nucleic acid systems. J. Comput. Chem. *32*, 170–173.

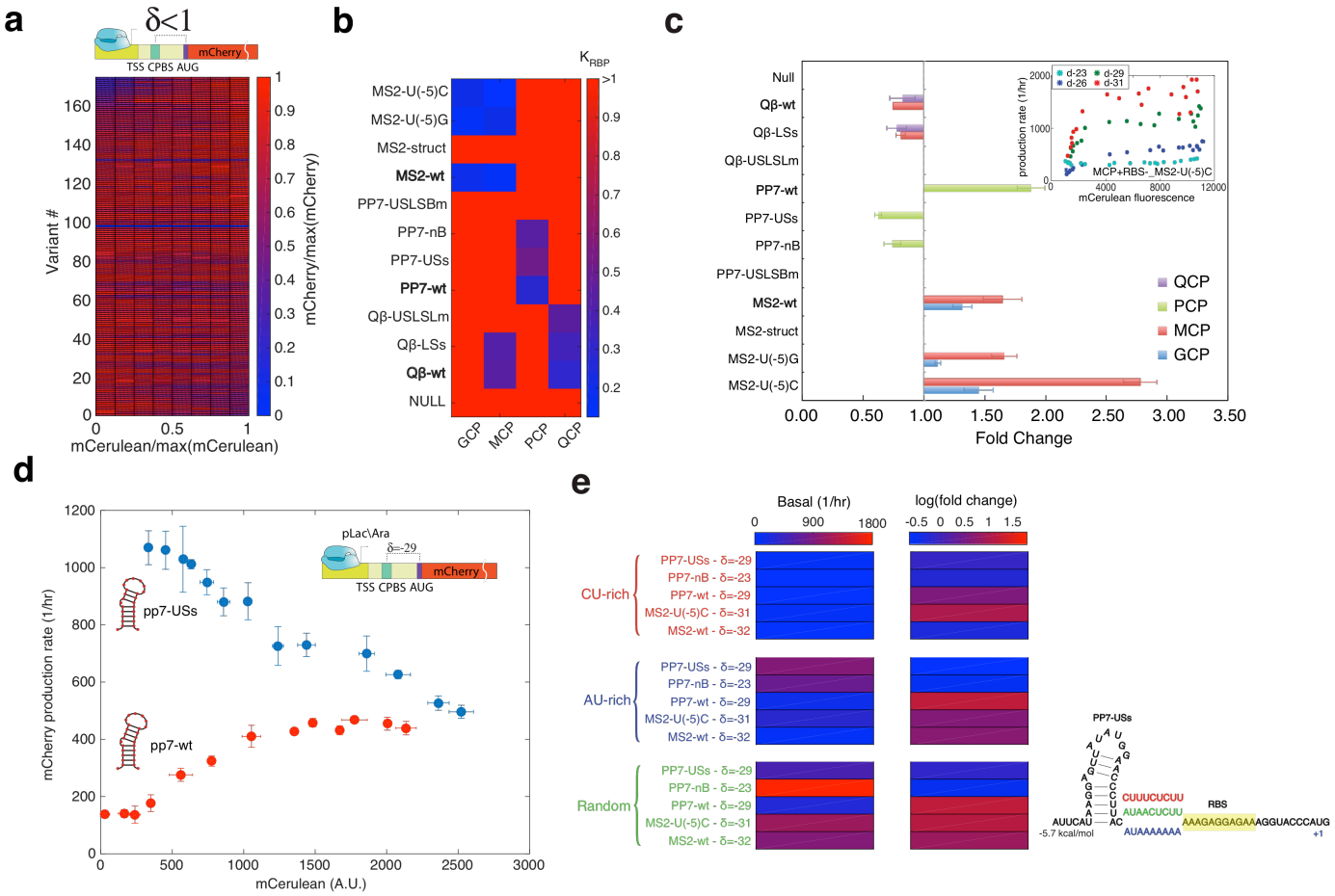
a
δ<1
mCherry
TSS CPBS AUG
Variant #
mCherry/max(mCherry)
mCerulean/max(mCerulean)
b
K_RBP
MS2-U(-5)C
MS2-U(-5)G
MS2-struct
MS2-wt
PP7-USLSBm
PP7-nB
PP7-USs
PP7-wt
Qβ-USLSLm
Qβ-LSs
Qβ-wt
NULL
GCP
MCP
PCP
QCP
c
Null
Qβ-wt
Qβ-LSs
Qβ-USLSLm
PP7-wt
PP7-USs
PP7-nB
PP7-USLSBm
MS2-wt
MS2-struct
MS2-U(-5)G
MS2-U(-5)C
QCP
PCP
MCP
GCP
Fold Change
production rate (1/hr)
mCerulean fluorescence
MCP+RBS-_MS2-U(-5)C
d-23
d-26
d-29
d-31
d
pLac\Ara
δ=-29
mCherry
TSS CPBS AUG
pp7-USs
pp7-wt
mCherry production rate (1/hr)
mCerulean (A.U.)
e
Basal (1/hr)
log(fold change)
CU-rich
AU-rich
Random
PP7-USs - δ=-29
PP7-nB - δ=-23
PP7-wt - δ=-29
MS2-U(-5)C - δ=-31
MS2-wt - δ=-32
PP7-USs
CUUUCUCUU
AUAACUCUU
RBS
AAAGAGGAGAAAGGUACCCAUG
AUUCAU
AC
AUAAAAAAA
-5.7 kcal/mol
+1

Figure 1

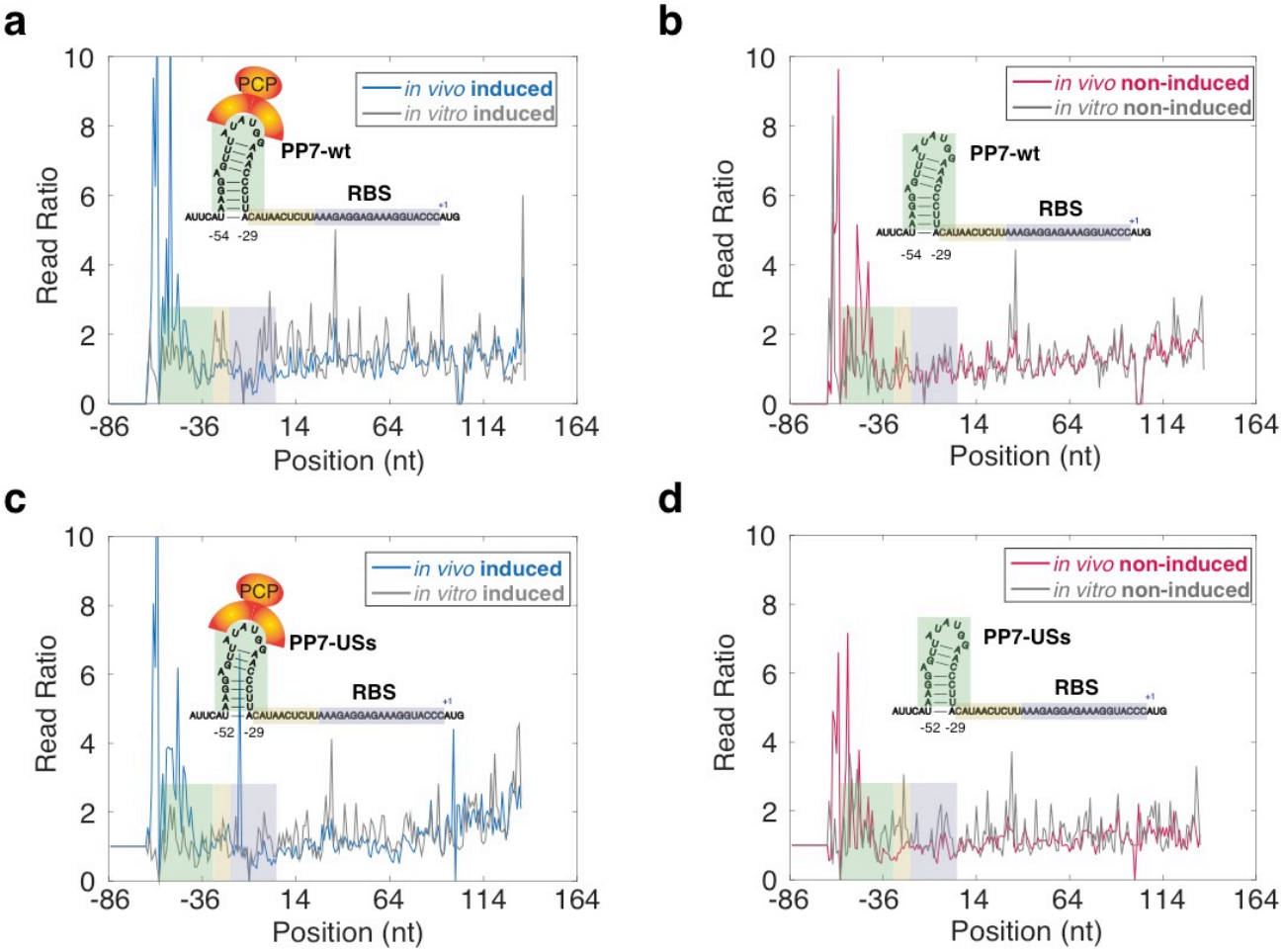
a
b
c
d
Read Ratio
Position (nt)
-86
-36
14
64
114
164
PCP
PP7-wt
PP7-USs
RBS
in vivo induced
in vitro induced
in vivo non-induced
in vitro non-induced
AUUCAU
ACAUAACUCUUAAAGAGGAGAAAGGUACCCAUG
-54 -29
-52 -29
+1

Figure 2

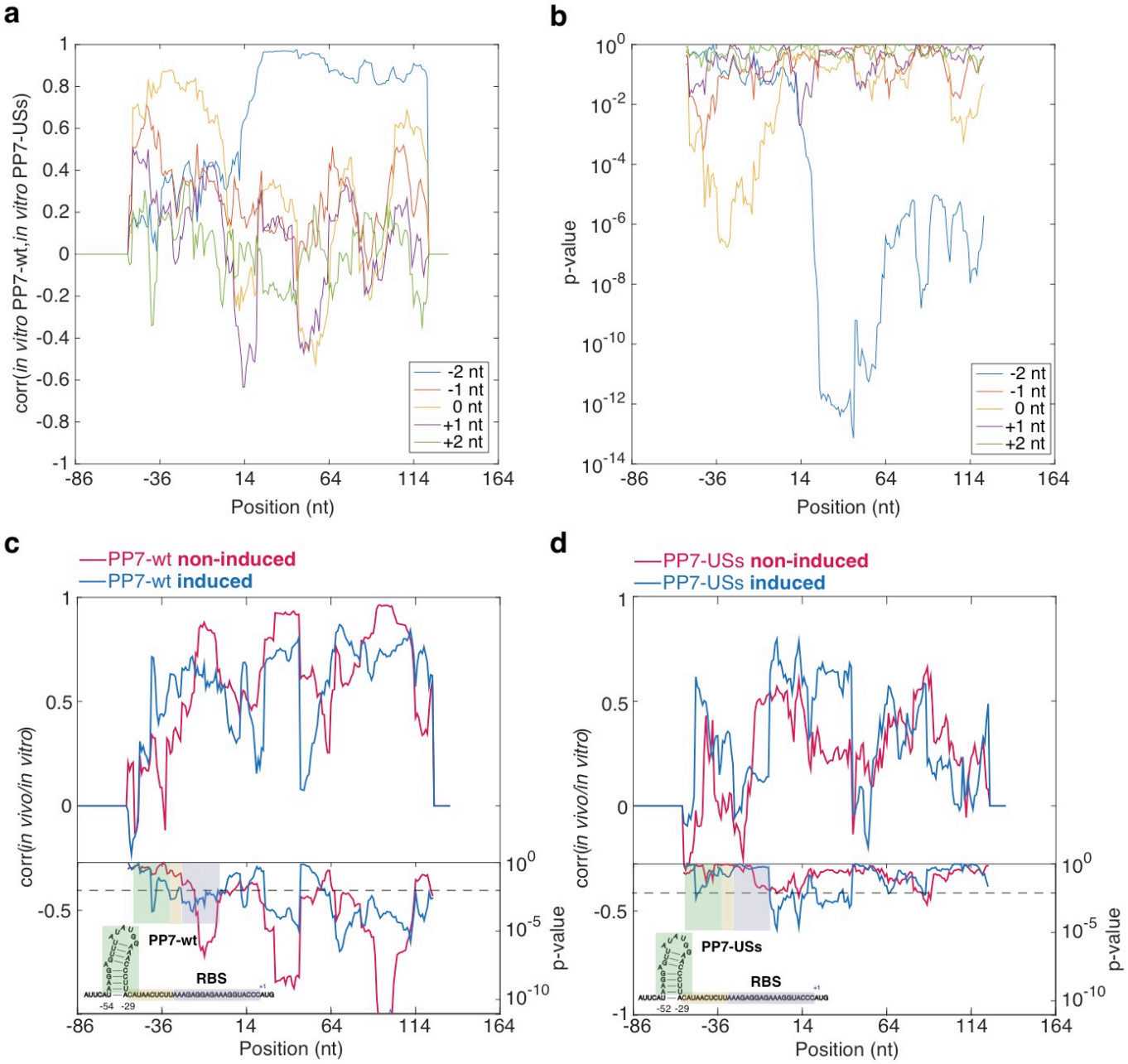
a
corr(in vitro PP7-wt,in vitro PP7-USs)
Position (nt)
-2 nt
-1 nt
0 nt
+1 nt
+2 nt
b
p-value
Position (nt)
-2 nt
-1 nt
0 nt
+1 nt
+2 nt
c
PP7-wt non-induced
PP7-wt induced
corr(in vivo/in vitro)
p-value
PP7-wt
RBS
Position (nt)
d
PP7-USs non-induced
PP7-USs induced
corr(in vivo/in vitro)
p-value
PP7-USs
RBS
Position (nt)

Figure 3

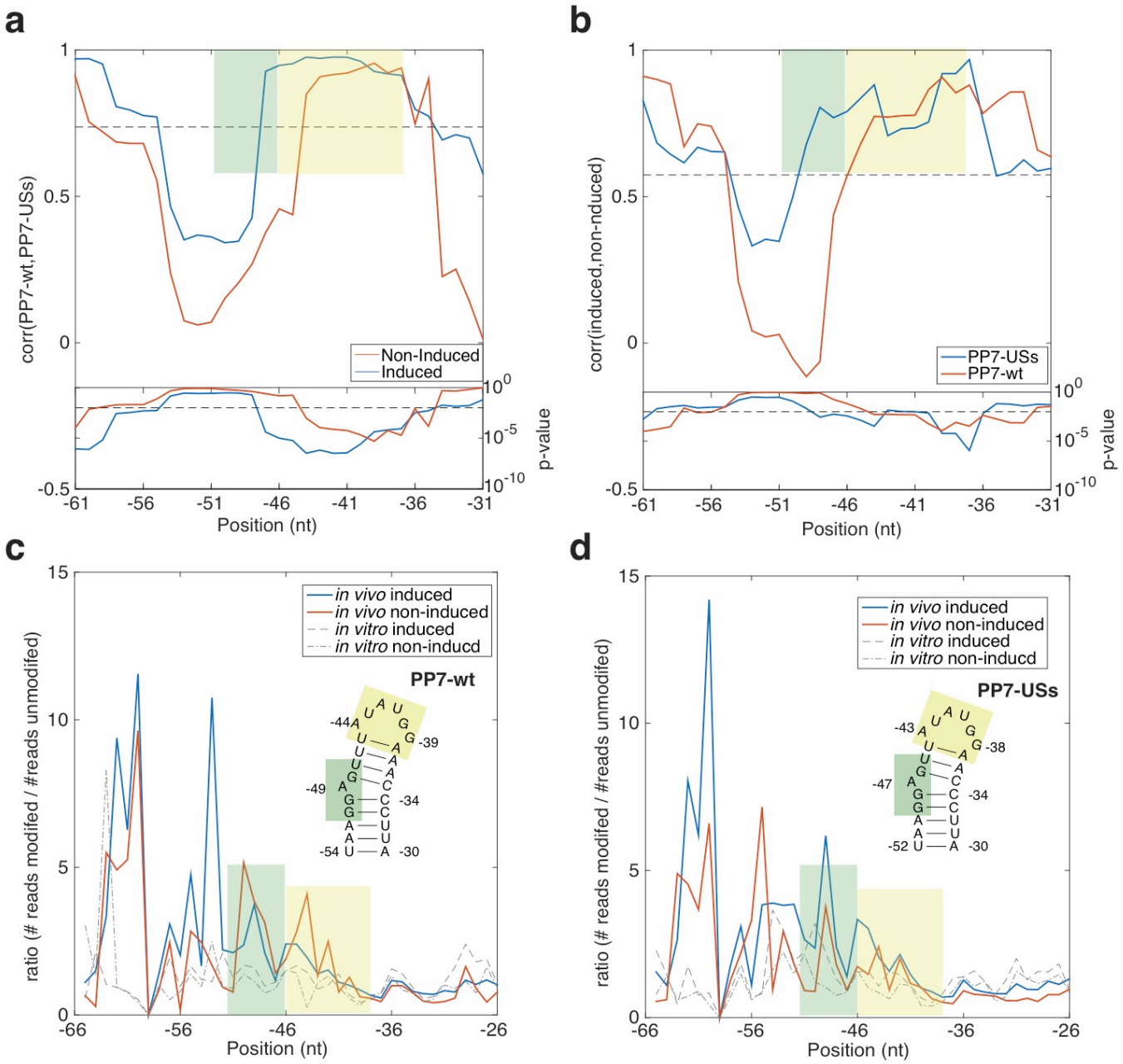
a
b
c
d
corr(PP7-wt,PP7-USs)
corr(induced,non-nduced)
Non-Induced
Induced
PP7-USs
PP7-wt
p-value
Position (nt)
ratio (# reads modifed / #reads unmodifed)
in vivo induced
in vivo non-induced
in vitro induced
in vitro non-inducd
PP7-wt
PP7-USs

Figure 4

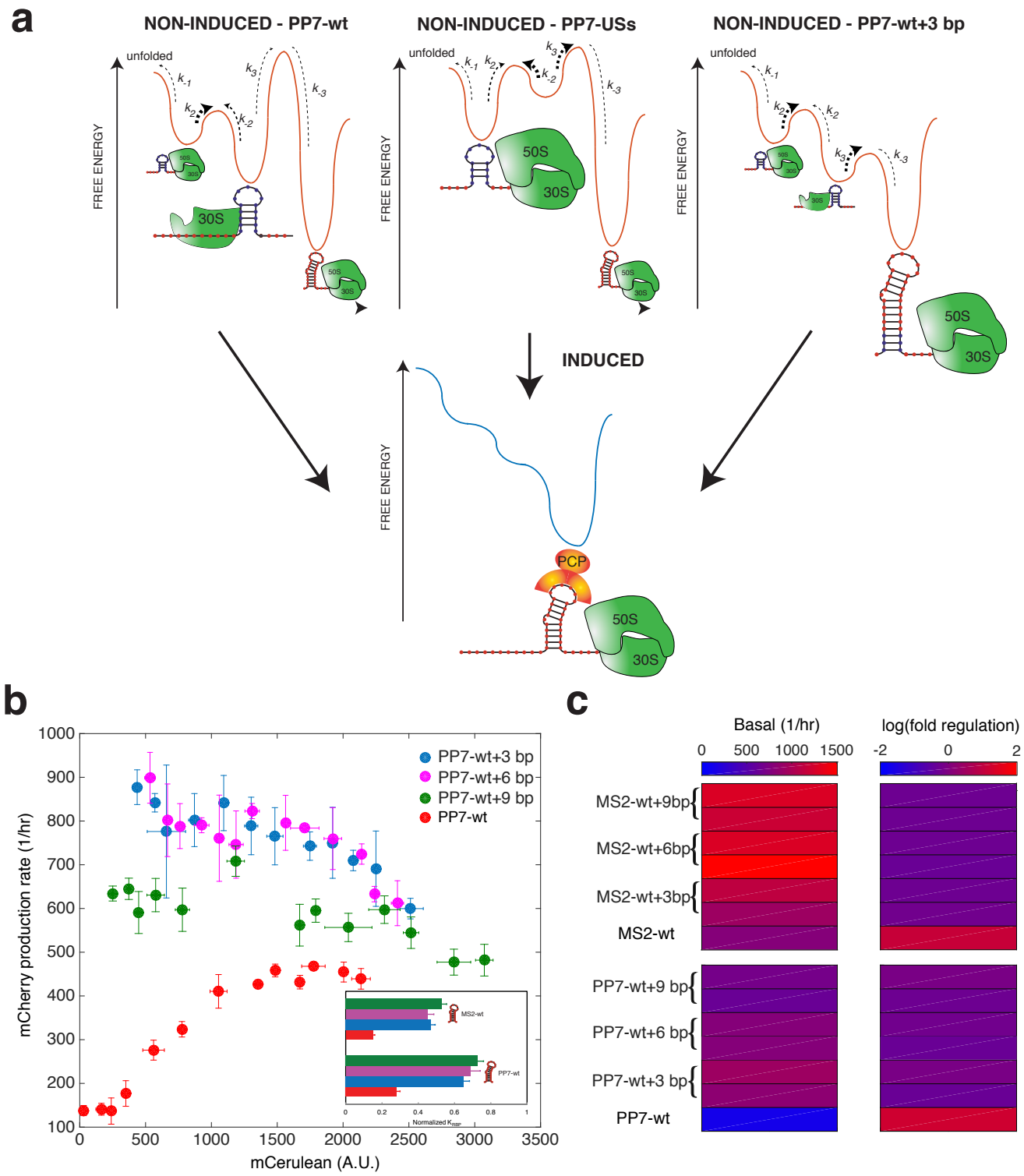
a
NON-INDUCED - PP7-wt
NON-INDUCED - PP7-USs
NON-INDUCED - PP7-wt+3 bp
unfolded
FREE ENERGY
50S
30S
INDUCED
PCP
b
mCherry production rate (1/hr)
mCerulean (A.U.)
PP7-wt+3 bp
PP7-wt+6 bp
PP7-wt+9 bp
PP7-wt
MS2-wt
Normalized K_REP
c
Basal (1/hr)
log(fold regulation)
MS2-wt+9bp
MS2-wt+6bp
MS2-wt+3bp
MS2-wt
PP7-wt+9 bp
PP7-wt+6 bp
PP7-wt+3 bp
PP7-wt

Figure 5

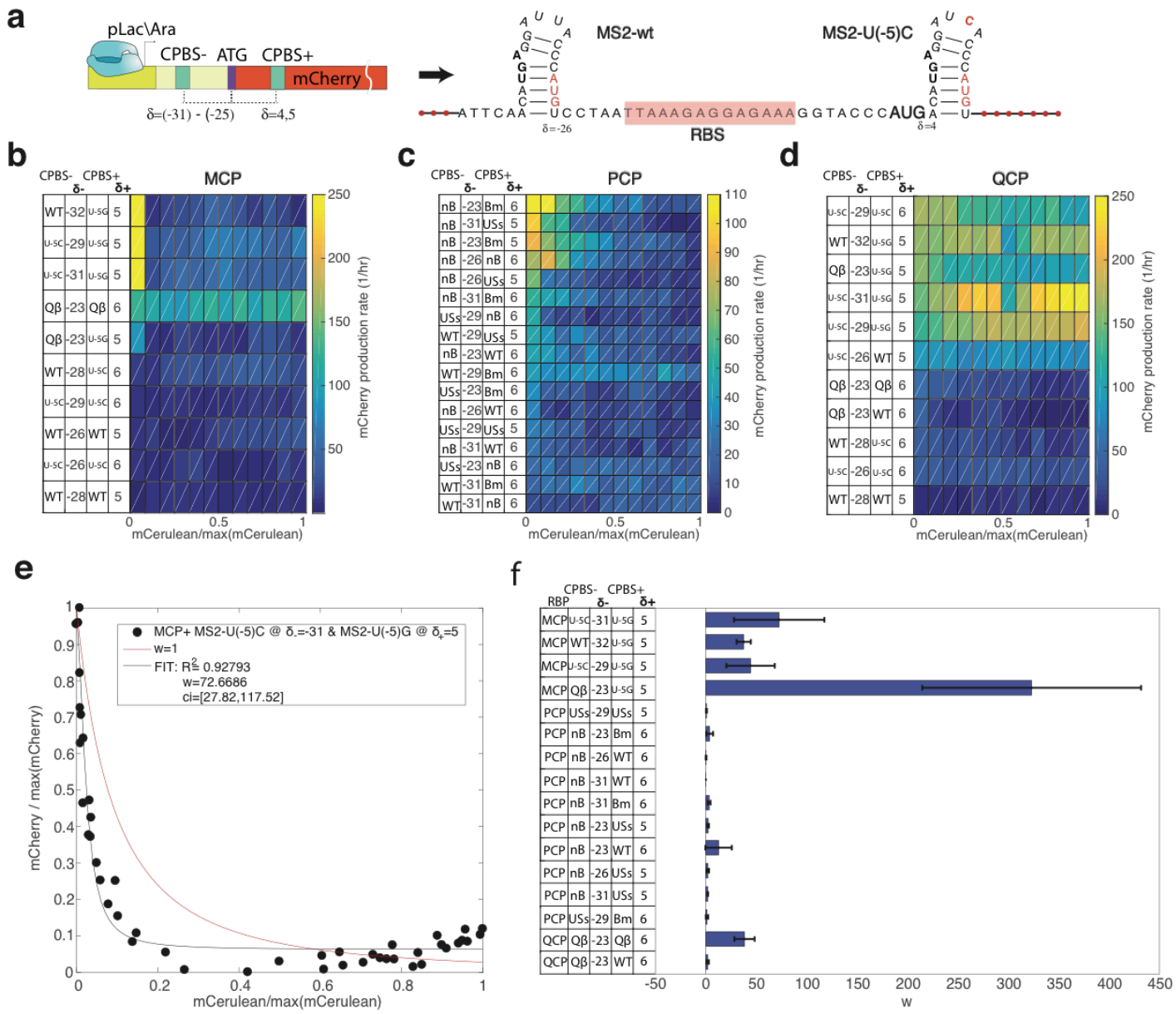
a
pLac\Ara
CPBS- ATG CPBS+
mCherry
δ=(-31) - (-25)
δ=4,5
MS2-wt
MS2-U(-5)C
ATTCAA—UCCTAATTAAAGAGGAGAAAGGTACCCAUG
δ=-26
δ=4
RBS
b
MCP
c
PCP
d
QCP
CPBS- CPBS+
δ- δ+
mCherry production rate (1/hr)
mCerulean/max(mCerulean)
e
MCP+ MS2-U(-5)C @ δ-=-31 & MS2-U(-5)G @ δ+=5
w=1
FIT: R2= 0.92793
w=72.6686
ci=[27.82,117.52]
mCherry / max(mCherry)
mCerulean/max(mCerulean)
f
RBP
w

Figure 6

Supplementary Information for synthetic protein-sensing riboswitches

Noa Katz, Beate Kaufmann, Roni Cohen, Oz Solomon, Orna Atar,
Zohar Yakhini, Sarah Goldberg, and Roee Amit

Contents

1 Fluorescence expression assay **31**
1.1 Design and construction of binding-site plasmids 31
1.2 Design and construction of fusion-RBP plasmids 31
1.3 Transformation of binding-site plasmids 31
1.4 Single-clone expression level experiments 31
1.5 Dose response fitting routine and K_d extraction 32

2 SHAPE-seq **36**
2.1 Experimental setup 36
2.2 Library preparation and sequencing 37
2.3 Bioinformatics analysis 37
2.4 Correlation analysis 38

3 Cooperativity model **40**

1 Fluorescence expression assay

1.1 Design and construction of binding-site plasmids

Binding-site cassettes (see Supp. Table S1) 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 promoter, ribosome binding site (RBS), and a KpnI restriction site. Minigenes for the experiments in which the hairpin was encoded in the $\delta > 0$ region, also contained 80 bases of the 5' end of the mCherry gene, and an ApaLI 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 Table S1), at varying distances. All binding sites were derived from the wild-type binding sites of the coat proteins of one of the four bacteriophages MS2, PP7, Qbeta, and GA. 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). About 10% of 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 five fusion-RBP plasmids (see below).

1.2 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 S2). All RBPs presented (PCP, MCP, QCP, and GCP) 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) [1]. 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.

1.3 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).

1.4 Single-clone expression level experiments

Dose-response fluorescence experiments were performed using a liquid-handling system (Tecan Freedom EVO 100, MCA 96-channel) in combination with a liconic incubator and a TECAN Infinite F200 PRO platereader. Each measurement was carried out in duplicates. Double-transformant strains were grown at 37°C and 250 rpm shaking in 1.5 ml LB in 48-well plates with appropriate antibiotics (Kan and Amp) over a period of ∼16 hours (overnight). In the morning, the inducer for the rhlR promoter C_4-HSL (N-butyryl-L-Homoserine lactone, Cayman Chemical) was pippetted manually to 4 wells in an inducer plate, and then diluted by the robot into 24 concentrations ranging from 0 to 218 nM. While the inducer dilutions were being prepared, semi-poor medium consisting of

95% bioassay buffer (for 1 L: 0.5 g Tryptone [Bacto], 0.3 ml Glycerol, 5.8 g NaCl, 50 ml 1M $MgSO_4$, 1ml 10xPBS buffer pH 7.4, 950 ml DDW) and 5% LB was heated in the incubator, in 96-well plates (Perkin Elmer). The overnight strains were then diluted by the liquid-handling robot by a factor of 100 into 200 µL of pre-heated semi-poor medium, in 96-well plates suitable for fluorescent measurement (Perkin Elmer). The diluted inducer was then transferred by the robot from the inducer plate to the 96-well plates containing the strains. The plates were shaken at 37°C for 6 hours.

Measurement of OD, and mCherry and mCerulean fluorescence were taken via a platereader (Tecan, F200) every 30 minutes. Blank measurements (growth medium only) were subtracted from all fluorescence measurements. For each day of experiment (16 different strains), a time interval of logarithmic growth was chosen (T_0 to T_{final}) according to the measured growth curves, between the linear growth phase and the stationary (T_0 is typically the third measured time point). Six to eight time points were taken into account, discarding the first and last measurements to avoid errors derived from inaccuracy of exponential growth detection. Strains that showed abnormal growth curves or strains where logarithmic growth phase could not be detected, were not taken into account and the experiment was repeated.

The average normalized fluorescence of mCerulean, and rate of production of mCherry were calculated for each inducer concentration using the routine developed in [2, Supplementary Information], as follows :

mCerulean average normalized fluorescence: for each inducer concentration, mCerulean measurements were normalized by OD. Normalized measurements were then averaged over the N logarithmic-growth timepoints in the interval $[T_0, T_{final}]$, yielding:

$$m\hat{Cerulean} = \frac{1}{N} \sum_{t=T_0}^{T_{final}} \frac{mCerulean(t)}{\mathrm{OD}(t)} \,. \tag{1}$$

mCherry rate of production: for each inducer concentration, mCherry fluorescence at T_0 was subtracted from mCherry fluorescence at T_{final}, and the result was divided by the integral of OD during the logarithmic growth phase:

$$mCherry\, rate\, of\, production = \frac{mCherry(T_{final}) - mCherry(T_0)}{\int_{To}^{T_{final}} dt\, \mathrm{OD}(t)} \,. \tag{2}$$

Finally, we plotted mCherry rate of production [3] as a function of averaged normalized mCerulean expression, creating dose response curves as a function of RBP-mCerulean fluorescence. Data points with higher than two standard deviations calculated over mCerulean and mCherry fluorescence at all the inducer concentrations of the same strain) between the two duplicates were not taken into account and plots with 25% or higher of such points were discarded and the experiment repeated.

1.5 Dose response fitting routine and K_d extraction

Final data analysis and fit were carried out on plots of rate of mCherry production as a function of averaged normalized mCerulean fluorescence at each inducer concentration. Such plots represent production of the reporter gene as a function of RBP presence in the cell. The fitting analysis and K_d extraction were based on the following two-state thermodynamic model:

$$mCherry\, rate\, of\, production = P_{bound} k_{bound} + P_{unbound} k_{unbound} \,. \tag{3}$$

Here, the mCherry mRNA is either bound to the RBP or unbound, with probabilities P_{bound} and $P_{unbound}$ and ribosomal translation rates k_{bound} and $k_{unbound}$, respectively. The probabilities of the two states are given by:

$$P_{boound} = \frac{([x]/K_d)^n}{1 + ([x]/K_d)^n}$$

and

$$P_{unboound} = \frac{1}{1 + ([x]/K_d)^n} \,,$$

where $[x]$ is RBP concentration, K_d is an effective dissociation constant, and n is a constant that quantifies RBP cooperativity; it represents the number of RBPs that need to bind the binding site simultaneously for the regulatory effect to take place. Substituting the probabilities into Eq. 3 gives:

$$mCherry\, rate\, of\, production = \frac{([x]/K_d)^n}{1 + ([x]/K_d)^n} k_{bound} + \frac{1}{1 + ([x]/K_d)^n} k_{unbound} \,.$$

For the case in which we observe a downregulatory effect, we have significantly less translation for high $[x]$, which implies that $k_{bound} \ll k_{unbound}$ and that we may neglect the contribution of the bound state to translation. For the case in which we observe an upregulatory affect for large $[x]$, we have $k_{bound} \gg k_{unbound}$, and we neglect the contribution of the unbound state.

The final models used for fitting the two cases are summarized as follows:

$$m\,Cherry\, rate\, of\, production \simeq \begin{cases} \frac{k_{unbound}}{1+([x]/K_d)^n} + C & \text{downregulatory effect} \\ \\ \frac{([x]/K_d)^n k_{bound}}{1+([x]/K_d)^n} + C & \text{upregulatory effect} \end{cases} \,,$$

where C is the fluorescence baseline. Only fit results with $R^2 > 0.6$ were taken into account, For those fits, K_d error was typicaly in the range of 0.5-20%, for a 0.67 confidence interval.

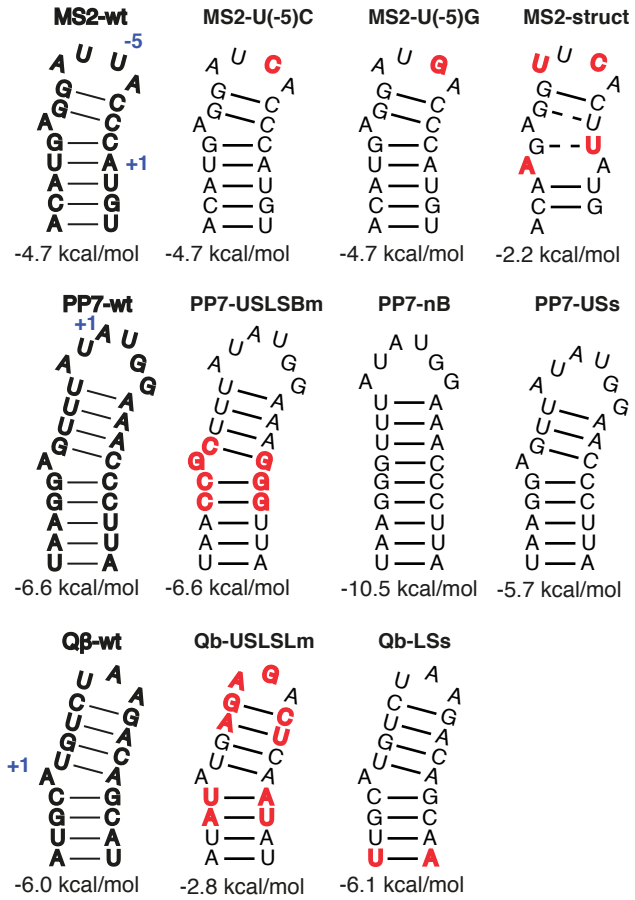

Figure S1: Binding sites used in the study. Structure schematic for the 11 binding sites used in the binding affinity study. Red nucleotides indicate mutations from the original wt binding sequence. Abbreviations: US/LS/L/B = upper stem/ lower stem/ loop/ bulge, m = mutation, s = short, struct = significant change in binding site structure. .

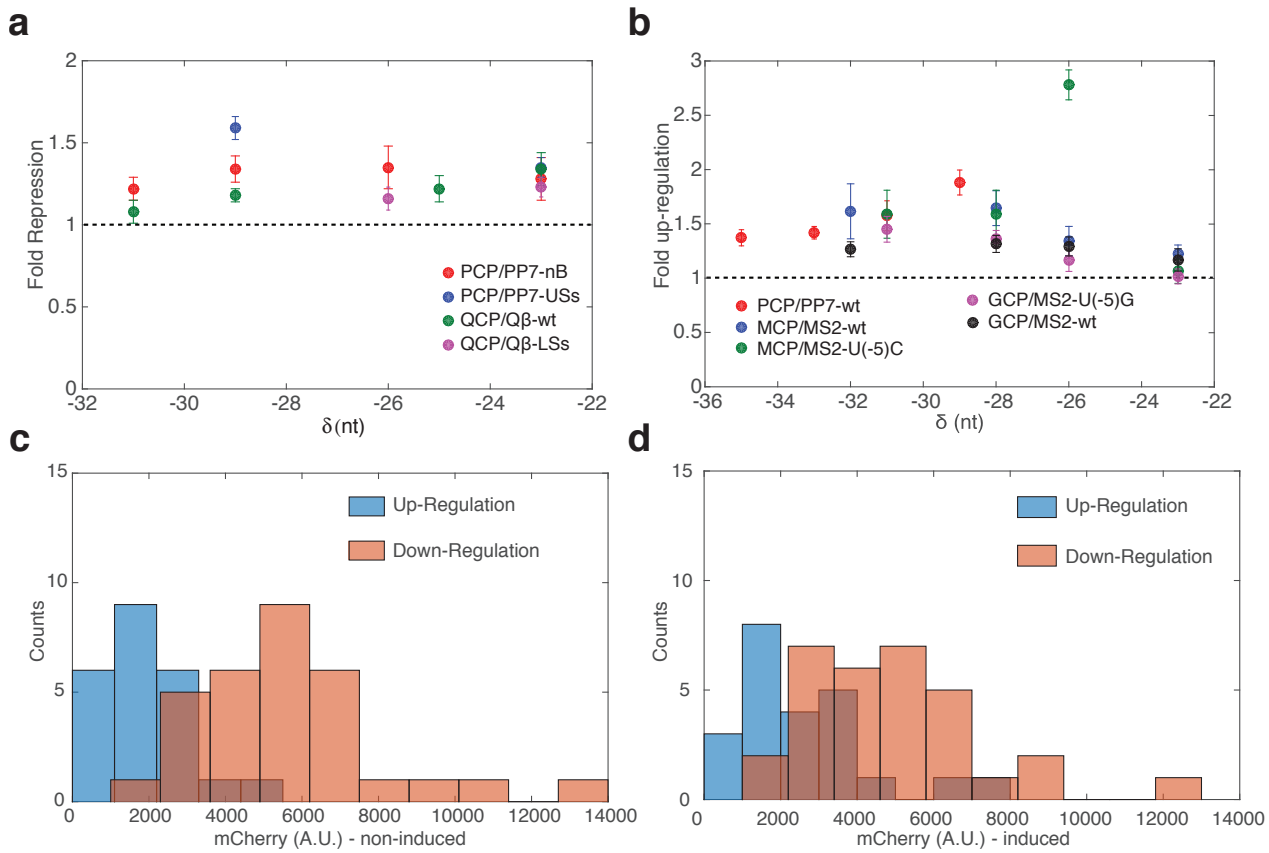

Figure S2: Translation regulation in the 5' UTR. (a) Fold repression effects as a function of position on the 5' UTR for the down regulating sites, showing no position-dependent effect. (b) Fold up-regulation effects as a function of position on the 5' UTR for the up regulating sites, showing no position-dependent effect. (c) Comparison of basal mCherry production rate for the up- (blue) and down-(red) regulating sites showing a distinct lower mean for the up-regulating sites. (d) Comparison of full induction mCherry production rate for the up- (blue) and down-(red) regulating sites showing a two relatively over-lapping distributions.

2 SHAPE-seq

2.1 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 (Section 1.1) and the RBP-fusion plasmid (see Section 1.2 and Supp. Table 1-2S2), and grown at 37°C for 16 hours while shaking at 250 rpm. Overnight cultures were diluted 1:100 into semi-poor medium (see Section 1.4). Each bacterial sample was divided into a non-induced sample and an induced sample in which RBP protein expression was induced with 250 nM C_4-HSL, as described above.

Bacterial cells were grown until OD_{600}=0.3, 3 ml of cells were centrifuged and gently resuspended in 0.5 ml semi-poor medium supplemented with a final concentration of 30 mM 2-methylnicotinic acid imidazole (NAI) suspended in anhydrous dimethyl sulfoxide (DMSO, Sigma Aldrich) [4], or 5% (v/v) DMSO. Cells were incubated for 5 min at 37°C while shaking and subsequently centrifuged at 6500 g for 5 min. Silica-membrane–based RNA isolation (RNAeasy Mini Kit, Qiagen) was used for strains containing RBS+_PP7_USs_d-29 and RBS-_PP7_wt_d-29. Samples were divided into the following sub-samples:

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 [5]. 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 [6, Supplemental Information], 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 S3) for mCherry that are specific for either the mCherry transcript (RBS-_PP7_USs_d-29, RBS-_PP7_wt_d-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 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 Table S4) 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, DNA was resuspended in 20 µl ultra-pure water and the dsDNA library was prepared.

2.2 Library preparation and sequencing

To add the TruSeq Universal Adapter (#6) sequence as well as the TruSeq Illumina indexes (#7-30 for next generation sequencing, see Supp. Table 3), the PCR reaction mix (1x Q5 HotStart reaction buffer, 0.1 mM dNTPs, 1 U Q5 HotStart Polymerase [NEB]) was added to 6 µl ssDNA, 4 nM selection primer #4 (primer extends 4 nucleotides into mCherry transcript to avoid the enrichment of ssDNA-adapter products), 0.5 µM TruSeq Universal Adapter #6 and 0.5 µM Illumina Index primer (one of #7-30, see Table S4). 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 sequenced using the Illumina HiSeq 2500 sequencing system using 2x50 bp paired-end reads.

2.3 Bioinformatics analysis

Illumina reads were first adapter-trimmed using cutadapt [7] and were aligned against a composite reference built from PP7-wt and PP7-USs 5' UTR-mCherry sequences, and PhiX genome (PhiX is used as a control sequence in Illumina sequencing). Alignment was performed using bowtie2 [8] in local alignment mode (bowtie2 --local). For full alignment details see Supplementary Data 1.

Reverse transcriptase (RT) drop-out positions were indicated by the end position of Illumina Read 2 (the second read on the same fragment). Drop-out positions were identified using an in-house Perl script (can be provided upon request). Reads that were aligned only to the first 19 bp were eliminated from downstream analysis, as these correspond to the RT primer sequence. In addition, reads that covered the full length (i.e. reads that spanned from the RT-primer to the TSS) of the sequences were also removed, as they are non-informative for SHAPE-seq RT drop-out calculation. For each position upstream of the RT-primer, the number of drop-outs detected was summed. This process was repeated for both the mCherry mRNA and the 5S rRNA sequence

After the trimming, the reporter libraries averaged ∼500,000± 85,000 total reads per reporter library (31 total), or ∼13.204 ± 7.002% of the initial reads number for each library (see Supplemental Data 1 for reads per position on all SHAPE-Seq libraries). The relative low number of trimmed

reads may have been due to the poor choice of the RT primer for mCherry, which also binds to *E. coli* 23S rRNA.

To facilitate proper signal comparison, all libraries of each construct (8 per construct) were normalized to have the same total number of reads. For each library j and position i=1,...,L, we normalized the number of drop-outs $D_j(i)$ according to:

$$\hat{D}_j(i) = \frac{D_j(i)}{\sum_{i=1}^{L} D_j(i)} \times M \ ,$$

where L is the length of the sequence under investigation after RT primer removal, and M is the maximal total number of drop-outs over all samples:

$$M = \max_j \{ \sum_{i=1}^{L} D_j(i) \} \ .$$

Note that M enables comparison between samples with different numbers of drop-outs, and also serves as a large constant in order to avoid dividing two very small numbers in the next step.

The SHAPE-seq *read ratio* was calculated for each strain s at each position i, as follows:

$$R_s(i) = \frac{\hat{D}_{s,mod}(i)}{\hat{D}_{s,non-mod}(i)} \ , \tag{4}$$

where $\hat{D}_{s,mod}(i)$ and $\hat{D}_{s,non-mod}(i)$ are the normalized drop-outs of the NAI-modified and the non-modified (DMSO) libraries, respectively (see Section 2.1). The normalized reads as a function of position from the TSS are supplied in Supplemental Data 1.

2.4 Correlation analysis

To provide a more quantitative assessment of the SHAPE-seq data, we employed a running Pearson correlation window approach. We reasoned that by correlating sequence segments of lengths 11 nt and above, segments that have a similar modification pattern, which reflect an underlying similar structure, will be characterized by correlated *read ratio* patterns. To determine the statistical significance of such correlation, we set a threshold of correlation coefficient whose p-value will be sufficently low. In our case, each SHAPE-Seq read-ratio library is approximately 200 nucleotides long. Therefore, we chose a correlation coefficient threshold whose p-value is equivalent to 0.01. This implies that given a null model of no-correlation between any two segments. In a run that consists of ~150-200 such segments, 1-2 events with a high correlation coeficient value will emerge by random.

Based on previous knowledge on our sequence: (e.g. binding site structure and location), we can now determine structually similar segments that are then likely to be consist of particular structural elements. This, in turn, can allow us to analyze the entire read-ratio segments in such a fashion, and isolate those segments which are unique and do not correlate with any of the other read ratio signals. Such unique segments, which are uncharacterized structurally, form good candidates for the elements which uncode the up-regulatory response.

To demonstrate the validity of this approach, we applied it to the *in vitro* read ratio results. In this case, since the *in vitro* modifications take place after the RNA is extracted from the cells, denatured, and only then refolded for modifications, the original cellular information is lost. Consequently, the structure of the RNA molecules in the two samples should be the same. To check this we applied two running correlation windows on the entire read ratio segements of 21 nt and 41 nt respectively. In both case (Supp. Fig. 3 and 4), a significant correlation with p-values <0.01 was

detected for the entire read ratio signal. In addition, the lower confidence intervals (dashed line) for both signals are also significantly larger than 0 for the entire read length. Therefore, using both statistical tests, the read ratio signals can be siad to be correlated, which is consistent with our expectations.

Running correlation windows and other statistical measurements were calculated using MATLAB.

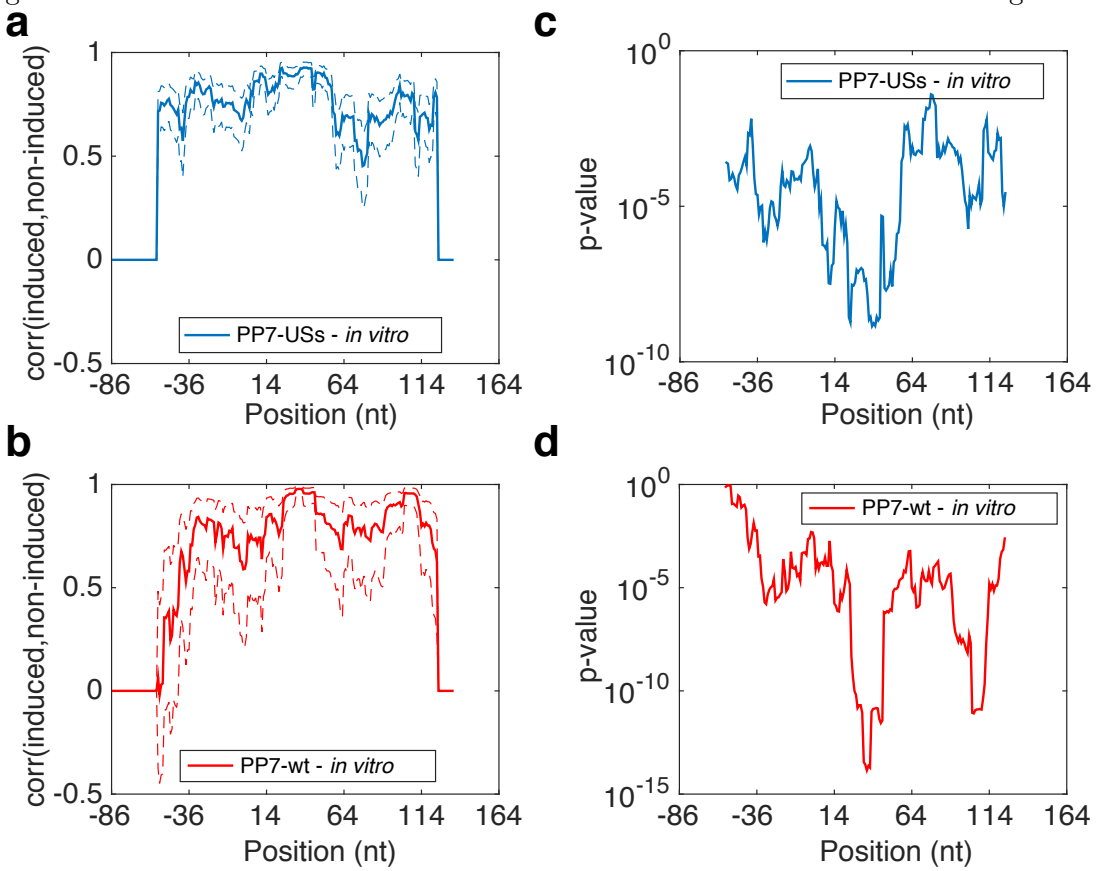

Figure S3: SHAPE-seq in vitro correlation analysis - 21 nt window (a-b) Pearson correlation of the in vitro induced and non-induced read-ratio librarries for the PP7-USs (a) and PP7-USs (b) binding sites. Dashed lines correspond to confidence intervals at 1 σ. (c-d) Positioned based two-tail p-value computation for the PP7-USs (c) and PP7-wt (d) correlation computations. P-value corresponds to the statistical significance of the correlation coefficient as compared with a null model of correlation coefficient zero.

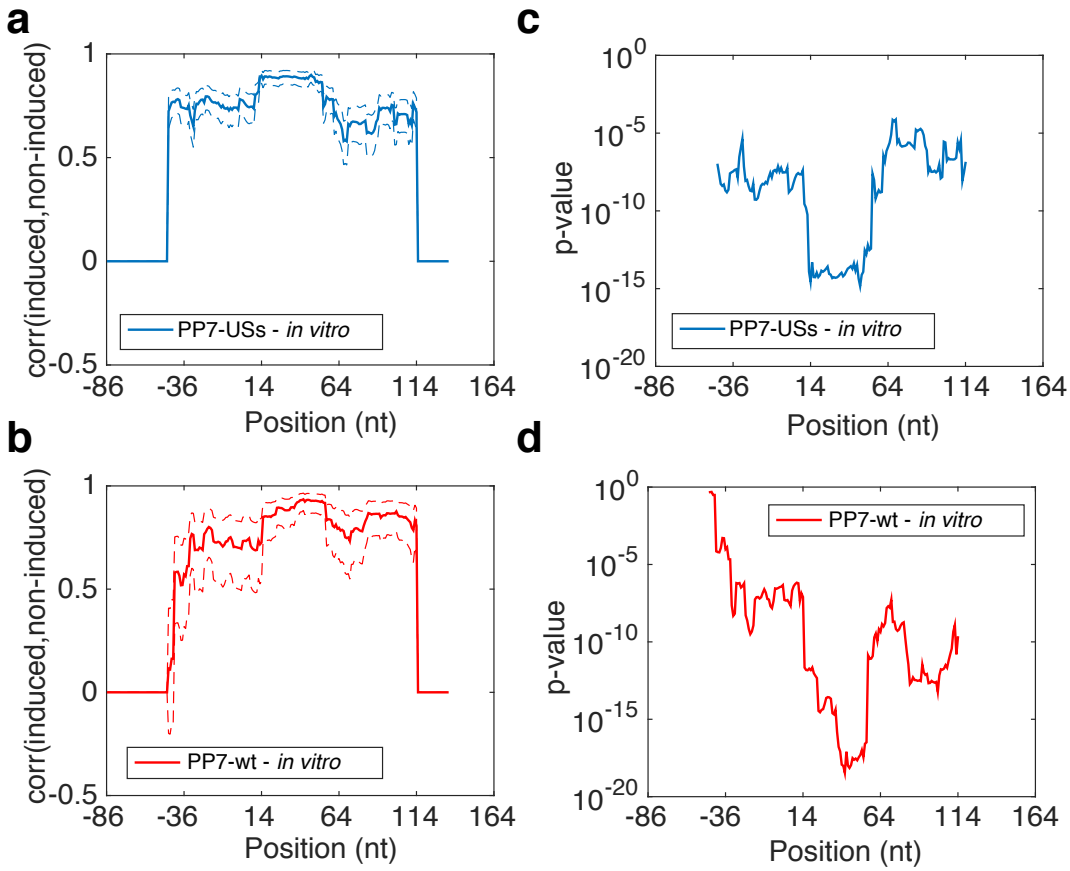

Figure S4: SHAPE-seq in vitro correlation analysis - 41 nt window (a-b) Pearson correlation of the in vitro induced and non-induced read-ratio librarries for the PP7-USs (a) and PP7-USs (b) binding sites. Dashed lines correspond to confidence intervals at 1 σ. (c-d) Positioned based two-tail p-value computation for the PP7-USs (c) and PP7-wt (d) correlation computations. P-value corresponds to the statistical significance of the correlation coefficient as compared with a null model of correlation coefficient zero.

3 Cooperativity model

To estimate the degree of cooperativity in RBP binding to the tandem binding site, we used the following 4-state thermodynamic model:

$$Z = 1 + \frac{[RBP]}{K_{RBP1}} + \frac{[RBP]}{K_{RBP2}} + \left(\frac{[RBP]^2}{K_{RBP1}K_{RBP2}}\right) w, \tag{5}$$

where K_{RBP1} and K_{RBP2} are the dissocation constants measured for the two single-binding-site variants, $[RBP]$ is the concentration of the RN binding proteins, and w is the cooperativity factor. In a four state model, we assume four potential RNA occupancy states. No occupancy - receiving the relative weight 1. A state with single hairpin bound by an RBP receiving either the weight $\frac{[RBP]}{K_{RBP1}}$ or the weight $\frac{[RBP]}{K_{RBP2}}$ depending on whether the 5' UTR or N-terminus states are occupied respectively. Finally, for the state where both hairpins are occupied we have the generic weight $\left(\frac{[RBP]^2}{K_{RBP1}K_{RBP2}}\right) w$ which takes into account also a potential interaction between the two occupied states, which can be cooperative if $w > 1$ or anti-cooperative if $w < 1$. No interaction is the case where $w = 1$.

Next, we compute the relevant probabilities for translation for each weight. We know that when the N'terminus hairpin is occupied translation cannot proceed, however, some translation can result (albeit via a lower rate), when the 5' UTR hairpin is occupied. This leads to the following rate equation for protein translation:

$$\begin{aligned}\frac{d\,[P]}{dt} &= k_{basal}\,[mRNA]\left(\frac{1}{1+\frac{[RBP]}{K_{RBP1}}+\frac{[RBP]}{K_{RBP2}}+\left(\frac{[RBP]^2}{K_{RBP1}K_{RBP2}}\right)w}\right)\\ &+ k_{utr}\,[mRNA]\left(\frac{\frac{[RBP]}{K_{RBP1}}}{1+\frac{[RBP]}{K_{RBP1}}+\frac{[RBP]}{K_{RBP2}}+\left(\frac{[RBP]^2}{K_{RBP1}K_{RBP2}}\right)w}\right)-\gamma\,[P]\,, \qquad (6)\end{aligned}$$

where γ is the protein degradation rate. When meauring rate of production and given the stability of mCherry, the degradation rate of mCherry is negligible over the 1-2 hr range of integration that was used in 2. Since we normalized the basal levels of mCherry rate of production, 6reduced to the following fitting formula for the data:

$$\begin{aligned}Normalized\ mCherry\ rate\ of\ production &= \left(\frac{1}{1+\frac{[RBP]}{K_{RBP1}}+\frac{[RBP]}{K_{RBP2}}+\left(\frac{[RBP]^2}{K_{RBP1}K_{RBP2}}\right)w}\right)\\ &+ \frac{k_{utr}}{k_{basal}}\left(\frac{\frac{[RBP]}{K_{RBP1}}}{1+\frac{[RBP]}{K_{RBP1}}+\frac{[RBP]}{K_{RBP2}}+\left(\frac{[RBP]^2}{K_{RBP1}K_{RBP2}}\right)w}\right).\end{aligned}$$

Finally, given our previous measurements for K_{RBP1} and K_{RBP2}, this formula reduces to a two-parameter fit for $w$and $\frac{k_{nterm}}{k_{basal}}$. See Supp. Fig. 6 and Supp. Table 4 for the fits and associated fitting parameter details for 14 of 16 dose-response down-regulatory tandem data sets that were used in the analysis.

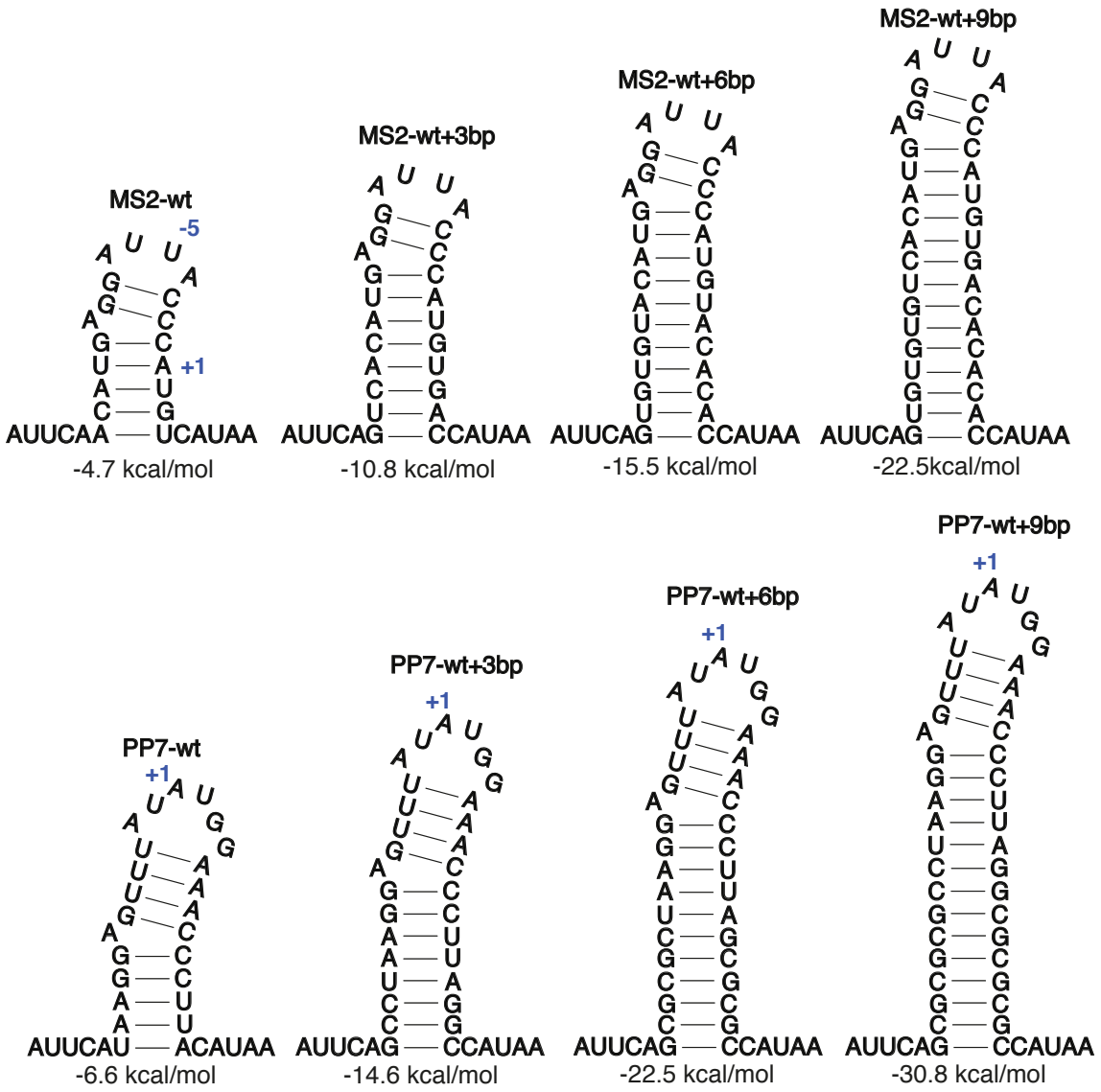

Figure S5: Binding site schematic and free-energies of the PP7-wt and MS2-wt bindings that are augmented by longer stems.

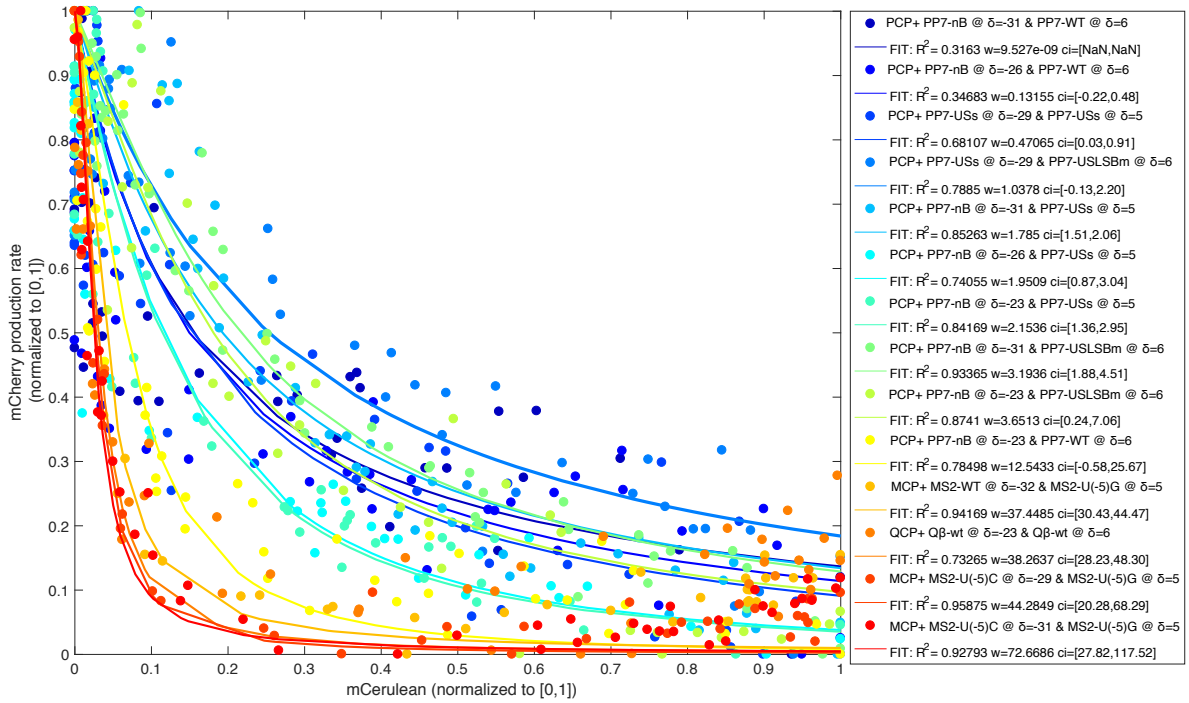

Figure S6: Tandems dose response and associated fits. Dose responses and model fits for 14 of the 16 tandems whose cooperativity parameter was plotted in Fig. 6F.

References

[1] Gerardo Medina, Katy Juárez, Brenda Valderrama, and Gloria Soberón-Chávez. Mechanism of Pseudomonas aeruginosa RhlR Transcriptional Regulation of the rhlAB Promoter. *Journal of Bacteriology*, 185(20):5976–5983, October 2003.

[2] Leeat Keren, Ora Zackay, Maya Lotan-Pompan, Uri Barenholz, Erez Dekel, Vered Sasson, Guy Aidelberg, Anat Bren, Danny Zeevi, Adina Weinberger, Uri Alon, Ron Milo, and Eran Segal. Promoters maintain their relative activity levels under different growth conditions. *Molecular Systems Biology*, 9(1), January 2013.

[3] Danny Zeevi, Eilon Sharon, Maya Lotan-Pompan, Yaniv Lubling, Zohar Shipony, Tali Raveh-Sadka, Leeat Keren, Michal Levo, Adina Weinberger, and Eran Segal. Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Res.*, 21(12):2114–2128, December 2011.

[4] Robert C. Spitale, Pete Crisalli, Ryan A. Flynn, Eduardo A. Torre, Eric T. Kool, and Howard Y. Chang. RNA SHAPE analysis in living cells. *Nature Chemical Biology*, 9(1):18–20, January 2013.

[5] Ryan A. Flynn, Qiangfeng Cliff Zhang, Robert C. Spitale, Byron Lee, Maxwell R. Mumbach, and Howard Y. Chang. Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nature Protocols*, 11(2):273–290, February 2016.

[6] Kyle E. Watters, Timothy R. Abbott, and Julius B. Lucks. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic Acids Research*, 44(2):e12–e12, January 2016.

[7] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):pp. 10–12, May 2011.

[8] Ben Langmead and Steven L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, April 2012.