An *in vivo* binding assay for RNA-binding proteins based on repression of a reporter gene

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 employ a reporter assay and SHAPE-seq to study translational regulation by RNA-binding proteins, in bacteria. For the reporter assay, we designed 50 constructs, each with a single hairpin based on the binding sites of the RNA-binding coat proteins of phages MS2, PP7, GA, and Qβ, at various positions within the N-terminus of a reporter gene. In the absence of RNA-binding proteins, the translation level depends on the position of the hairpin, and exhibits three-nucleotide periodicity. For hairpin positions within the initiation region, in the presence of cognate RNA-binding protein, we observe strong translational repression. *In vivo* SHAPE-seq results for a representative construct indicate that repression correlates with protection of both the hairpin and the ribosome binding site. Our data suggest that the RBP-hairpin complex entraps the 30S subunit, thereby stalling initiation. We utilize the repression phenomenon in a high-throughput assay for quantitative study of protein-RNA binding *in vivo*.

Keywords

Protein-RNA interaction, phage coat protein, MS2, PP7, translational repression, hairpin.

The regulation of gene expression is a process central to all biological life-forms. It is a process thought to be mediated largely by proteins, which interact with either chromatin or its RNA product. The best-known form of regulation is mediated by transcription factors, which control RNA levels by their sequence-specific interaction with DNA. Gene expression regulation based on protein-RNA interactions, however, is quite different, due to the nature of RNA. Unlike DNA which is a long, chromatinized, replicated, and for the most part exists as a double stranded molecule, RNA is a short, transient (i.e. constantly manufactured and degraded), exists in multiple copies, has particular modifications [1,2], and folds into functional secondary and tertiary structures. RNA structure is thought to be highly dynamic, and is dependent on many factors such as temperature, cellular RNA-binding protein (RBP) content, presence or absence of translating ribosomes, and interaction with other RNA molecules [3]. Thus, a typical RBP-RNA interaction is likely to depend not only on the presence of a specific binding sequence, but also on many other factors.

In bacteria, post-transcriptional regulation has been studied extensively in recent decades. There are well-documented examples of RBPs that either inhibit or directly compete with ribosome binding via a variety of mechanisms. These include direct competition with the 30S ribosomal subunit for binding via single stranded recognition [4], entrapment of the 30S subunit in an inactive complex via a nested pseudoknot structure [5], and ribosome assembly inhibition when the RBP is bound to a structured RBP binding site, or hairpin [6–9]. These hairpins have been studied in three distinct positions: either immediately downstream and including the AUG [7], upstream of the Shine-Dalgarno sequence [8], or as structures that entrap Shine-Dalgarno motifs, as for the PP7 and MS2 phage coat-protein binding sites. There is also a well-characterized example of translation stimulation: binding of the phage Com RBP was shown to destabilize a sequestered ribosome binding site (RBS) of the Mu phage *mom* gene, thereby facilitating translation [10,11]. While these studies indicate a richness of RBP-RNA-based regulatory mechanisms, a systematic understanding of the relationship between RBP binding, sequence specificity, the underlying secondary and tertiary RNA structure, and the resultant regulatory output is still lacking.

In recent years, advances in next generation sequencing technology combined with selective nucleic acid labeling approaches have facilitated focused study of specific RNA structures *in vivo*. These chemical-modification approaches [12–16] can generate a “footprint” of the dynamical structure of a chosen mRNA molecule *in vivo*, while in complex with ribosomes and/or other RBPs. In parallel, synthetic biology approaches that simultaneously

characterize large libraries of synthetic regulatory constructs have been increasingly used to complement the detailed study of single mRNA transcripts. While these synthetic approaches have been mostly applied to transcriptional regulatory platforms [17–20], their potential for deciphering post-transcriptional regulatory mechanisms have been demonstrated in a recent study that interrogated IRES sequences in mammalian cells [21].

Building on these advances and on a smaller-scale demonstration of translational repression by the RBP L7Ae in both bacteria and mammalian cells [22], we measured the regulatory output of a small library of synthetic constructs in which we systematically varied the position and type of RBP binding sites. In addition, we selectively applied SHAPE-seq [23,24,15] to a single variant, to further identify and characterize RBP-based regulatory mechanisms in bacteria. Our findings indicate that the chosen structure-binding RBPs - coat proteins from the bacteriophages GA [25], MS2 [26], PP7 [9], and Qβ [27] - generate a strong repression response when bound in the translation initiation region. We employed this repression phenomenon as an *in vivo* binding assay for RBP-RNA interactions. Using this assay, we can quantitatively characterize RBP binding to a set of mutated binding sites in a high-throughput fashion, thereby increasing our understanding of RBP-RNA binding *in vivo*, and enabling the engineering of more complex RNA-based applications.

Results

Hairpins inside the coding region do not inhibit translation

We first studied the translational regulatory effects of hairpins located within the 5' end of a transcribed gene. We chose to work with RBP binding-site sequences encoding hairpin structure (Fig. 1a, all structures herein were generated using NUPACK [28]), in the absence of their cognate RBPs [the phage coat proteins of bacteriophages GA (GCP), MS2 (MCP), PP7 (PCP), and Qβ (QCP)]. The four wild-type binding sites are characterized by two stems of varying length, which are separated by a single unpaired nucleotide or "bulge", and a loop of differing size that varies from 3 to 6 nucleotides. We measured the effect of hairpin position using a reporter system comprising a plasmid with a fluorescent mCherry gene under a constitutive pLac/Ara promoter (Fig. 1b-top right), with a single binding site inserted into the mCherry coding region (additional nucleotides were inserted downstream of the binding site to maintain the reading frame). Hairpin distance δ from the first nucleotide of the mCherry AUG was varied from 4 (immediately adjacent to the AUG) to 11 nt, at 1 nt resolution (see Supplementary Table 1 for sequences), and from 4 to 21 nt for PP7-wt. In Fig.

1b-right, we plot the mean rate of production of each construct as a function of δ. The data show that mCherry expression depends strongly on the position of the hairpin relative to the AUG, varying between inhibition to strong expression, where the black dashed line corresponds to the expression level for unstructured mRNA, averaged over two variants lacking apparent hairpins at the N-terminus (see Supplementary Table 1 – one with only the natural mCherry reporter, and another with a codon for Glycine [GGC] added after the AUG). In particular, constructs containing the Qβ-wt and GA-wt hairpins exhibit approximately a two-order-of-magnitude jump in expression when δ is shifted from 8 to 9, and 7 to 8, respectively.

Interestingly, mCherry expression for all four hairpins exhibited 3-nt periodicity, indicating that the basal expression levels depend on the position of the hairpin within the reading frame of the translating ribosome. To explore the effect of the in-frame positions (δ=4,7, etc.), which are inaccessible using the PP7-wt, MS2-wt, or Qβ-wt sequence (due to in-frame stop codons), we studied both the GA-wt binding site (out-of-frame stop codon for δ=6,9, etc.) and two Qβ binding sites for which the stop codon was mutated (see Supp. Fig. 1). Basal expression levels for both the GA-wt (green line) and the mutated Qβ binding sites (Fig. 1b-inset) show that there is no observed special regulatory role for the in-frame (δ=4,7, etc.) positions.

We propose that the ribosome unwinds the hairpin structures at different rates during initiation (depending on AUG-hairpin distance δ), thus providing a plausible explanation for the dependence of expression on the position of the binding site with respect to the start codon of mCherry. Such a position-dependent translation process can be modeled, for example, as a hopping process (see Supplementary Methods). Here (Fig. 1c - left), the transition between two translation steps can occur over one of three time scales. The first, fast time scale (T_{fast}) corresponds to the time it takes the ribosome to pull in the next three unpaired nucleotides, and to shift the tRNA one position (from the P position to E, or from the A position to P). The second, slower kinetic process (T_{int}) is carried out when the ribosome encounters a structured trinucleotide whose first base is single-stranded (top schematic left). Finally, a slow kinetic process (T_{slow}) unfolds when the ribosome encounters a structured trinucleotide whose first base is double-stranded. Recent single-molecule studies [29] showed that the ribosome is capable of forcing open a double-stranded trinucleotide, but the rate at which this happens is significantly slower than a typical fast translation step (0.2

aa/sec vs 10-20 aa/sec – see Supplementary Methods). As a result, the number of "slow" steps (Fig. 1c-right) that a ribosome needs to take in order to clear the initiation region determines the initiation rate, and could potentially also dominate the overall translation rate.

RBPs repress translation when bound within the ribosomal initiation region of mRNA

We next explored the regulatory effect of an RBP-hairpin complex, as a function of hairpin position downstream to the AUG. To do so, we constructed 22 two-plasmid strains (Fig. 2a): the first was a subset of the plasmids that were used for the hairpin experiments in the N-terminus region (see Fig. 1b-right for schematic and Supplementary Table 1), while the second encoded an RBP fused to mCerulean, under a pRhlR promoter inducible by N-butanoyl-L-homoserine lactone (C_4-HSL). We used two different RBPs: the phage coat proteins QCP and PCP of bacteriophages Qβ and PP7, respectively (see Supplementary Table 2). Each RBP—binding-site plasmid pair was transformed into *E. coli* TOP10 and grown in 24 different C_4-HSL concentrations, in duplicate. Optical density, mCherry, and mCerulean fluorescence levels were measured at multiple timepoints for each inducer concentration. mCherry production rates [30,31] were computed over a 2-3 hr window (see Supplementary Methods and Supp. Fig. 2) for each inducer level, and mCerulean levels were averaged over the same time-frame. In Fig. 2b, we plot a series of dose-response curves obtained for PCP on three constructs containing the PP7-wt binding sites, positioned at δ=7 (red), 11 (blue), and 16 (green) nt, respectively. For the hairpin located at δ=7, the mCherry rate of production is reduced by nearly two orders of magnitude. For the hairpin positioned at δ=11, a weakly repressing dose-response function was detected, while no repression was observed at δ=16.

We hypothesized that the RBP-bound hairpin complex interferes with the initiation process, thus preventing the 70S elongating subunit from forming (see Fig. 2c). To test this model, we measured the mCherry rate of production for the PP7-wt and Qβ-wt sites in the presence of PCP and QCP, respectively, for δ varying from 5 to 21 nt. Since the initiation region is thought [32] to extend from the RBS to about 8-10 nucleotides downstream of the AUG (+11 to +13 in our coordinates), the chosen range of δ covered both the initiation and elongation regions. The data (Fig. 2d) show that while strong repression is observed for both RBPs within the initiation region, the RBP inhibition effect completely disappears for $\delta \geq 16$, for both binding sites. This stands in contrast to the regulatory effect generated by the PP7-wt hairpin by itself (see Fig. 1b), which persists within both the ribosomal initiation and elongation regions.

SHAPE-seq measurement suggest the bound RBP entraps the 30S subunit in an inactive mode

To provide additional experimental support for the inhibition mechanism proposed above, we employed SHAPE-seq (see Supplementary Methods and Supp. Fig. 3) using the acylimidazole reagent 2-methylnicotinic acid imidazolide (NAI), which modifies the 2' OH of non- or less-structured, accessible RNA nucleotides as found in single-stranded RNA molecules [23]. We focused on the PP7-wt, δ=5 construct for which we observed down-regulation in the presence of PCP, and on the 5S rRNA as a previously-characterized control [24] (Supplementary Fig. 4a-b). The protocol was applied both *in vivo* [following [24]] and *in vitro* [following [15]] for the two RNA targets. As previously shown [33,34], RBP-RNA interactions *in vivo* limit the 2'-hydroxyl acylation, whereas the chemical probing is uninhibited *in vitro*, where RBPs are absent. To compare to gene expression results, we carried out SHAPE-seq on the two induction states 0 nM of C_4-HSL (i.e no PCP present), and 250 nM of C_4-HSL (PCP fully induced), after ensuring that the RNA levels at both induction states are similar using quantitative PCR (Fig. 3b – inset). In addition, we applied SHAPE-seq to purified and refolded transcripts *in vitro*. To quantify the propensity of a particular position to be modified in SHAPE-seq, we computed the ratio of the number of counts obtained for the modified sample to the unmodified sample, for each condition. We validated this analysis approach by first comparing the ratio computation to a previously published "reactivity" algorithm [35] for the 5S rRNA samples, and found them to be comparable (see Supplementary Fig. 4a-b and Supplementary Methods).

We next examined the read ratio (Fig. 3a) and reactivity (Supplementary Fig. 4c) for the *in vitro* PP7-wt, δ=5 samples. For both analysis approaches, we observed three peaks of high ratio/reactivity overlapping or adjacent to the binding-site's encoded region, which can be attributed to the RBS (50-51 nt), the small A-bulge in the structure (64 nt), and the large 6-nucleotide loop (69-74 nt) of the PCP binding site. The rest of the RNA nucleotides showed varying ratios/reactivities, consistent with a structured RNA molecule *in vitro* with various single-stranded bases embedded within.

In vivo, in the absence of RBP (i.e. where gene expression is observed), we obtained a read ratio signal for PP7-wt, δ=5 (Fig. 3b-red line) whose baseline was higher than that observed for the *in vitro* case. Moreover, the signal was characterized by multiple peaks of both high ratio and reactivity (Supplementary Fig. 4c-red line), at positions that were both

reactive (e.g. 60-75 nt range corresponding to the binding site) and non-reactive in the *in vitro* samples (Fig. 3a-black line and Supplementary Fig. 4c-blue line). This signal is consistent with an interpretation whereby a translationally-active RNA molecule is less structured due to ribosome binding, initiation, and elongation, and thus prone to modifications.

For the fully-induced *in vivo* case (Fig. 3b-blue line, and Supplementary Fig. 4d-red line), the ratio signal baseline for PP7-wt, δ=5 is reduced to levels comparable to those observed for the *in vitro* molecule. In addition, the peaks for the RBS (50-51 nt) and the one attributed to the large PP7-wt loop (69-77 nt) *in vitro* are strongly diminished, while the small peak attributed to the A-bulge seems to be partially attenuated (64 nt). This analysis is consistent with an interpretation of a structured, translationally-inactive RNA molecule, with a protein bound to the large PP7-wt loop [36], and another protein (possibly an inactive 30S subunit) bound to the RBS and the initiation region (schematically shown in Fig. 2c). This interpretation is further supported by the extended hypo-modified region downstream spanning to the ~80-90 nt range (i.e. bound 30S protection spans a wider region than that of the structured binding sites), and the recovery of the *in vitro* ratio and reactivity scores starting at the ~91-92 nt position.

Effective dissociation constant is insensitive to binding-site position

Given the strong RBP-induced repression phenomenon observed in the ribosomal initiation region, we wanted to use this effect to further characterize the binding of the RBPs to structured binding sites. To do so, we first constructed a set of mutated binding sites with various structure-modifying and non-structure-modifying mutations [compare Fig. 4a - red highlighting to the native sites for MCP (MS2-wt - top left), PCP (PP7-wt middle-left), and QCP (Qβ-wt bottom-left)]. The mutated binding sites for MCP and PCP were taken from [9] and [9], respectively (Fig. 4a). We then constructed two to four new constructs for each mutated binding site that differed in binding-site position downstream to the AUG. In addition, we constructed a set of control plasmids that lacked a hairpin within the N-terminus of the mCherry reporter gene. Altogether, we constructed 50 reporter plasmids and 5 no hairpin controls. The new constructs, and the ones previously tested (Figs. 1b and 2d), were co-transformed with all four RBP plasmids to yield 220 RBP—binding-site strains. Our goal with this design was to test not only the binding affinity to the native RBP, but also the

relative affinity to the other RBPs, thus obtaining an estimate for the selectivity of the RBP binding.

We plot the dose-response curves of 175 out of the 220 as a heat-map in Fig. 4b (strains with basal mCherry rate of production <50 a.u./hr were excluded). 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 accordance with fold-regulation of the response, with the most repressive variants positioned at the bottom, and least repressive at the top. The data show that that there is a substantial subset of strains, which exhibit strong repression for at least one position (~50 variants), with the strongest mCherry signal occurring at the lowest mCerulean level. To obtain an estimate for the effective binding affinity for each down-regulating variant, we fitted each dose-response curve that exhibited a typical repression response (see Supplementary Fig. 2) with a Hill-function-based model (see Supplementary Methods), which assumes a simple relationship between the concentration of RBP, the dissociation constant, and the output expression rate. Finally, we normalized the resulting dissociation constant by the maximal mCerulean expression for the matching RBP to facilitate comparison of the results for the different proteins, yielding an effective dissociation constant (K_{RBP}). Typical error in estimation of the effective dissociation constant was 5-20%, and by averaging K_{RBP} of each RBP—binding-site pair over multiple positions (values of δ) we obtained estimated errors of ~10%.

In Fig. 4c, we plot the averaged K_{RBP} for different RBP—binding-site combinations as a heat-map only for those sites (Fig. 4a) for which all four RBPs were tested (``null'' corresponds to an average K_{RBP} computation made on several of the non-binding-site controls). The data show that the effective dissociation constants measured for native sites with their cognate RBPs were low and approximately equal, indicating that native sites are optimized for binding (blue squares). Mutated sites which retained binding affinity displayed slightly larger dissociation constants (blue-purple), while the K_{RBP} values of RBP-binding site combinations that did not generate a binding signature were set to the maximum normalized value 1 (K_{RBP}=1, red). When examining the data more closely we find that PCP is completely orthogonal to the MCP/QCP/GCP group, with no common binding sites. Conversely, we observed crosstalk between the different members of the MCP/QCP/GCP group, with increased overlap between MCP and GCP, which is consistent with previous studies [25].

Finally, we note that structure-conserving mutations to native binding sites in the loop area [MS2-U(-5)C and MS2-U(-5)G] or stem (PP7-USLSBm) did not seem to affect binding. However, putative structure-altering (MS2-struct – where the lower stem is abolished) and destabilizing (Qb-USLSLm – where the GC base-pairs are converted to UA base-pairs in the lower stem) mutations significantly affected binding. Interestingly, structure-altering mutations, which retain apparent binding site stability (PP7-nB and PP7-USs), also seemed to retain a strong binding affinity to the native RBP. Altogether, these results suggest that binding sites positioned within the initiation region can tolerate multiple mutations as long as certain key structural features necessary for binding and hairpin stability (e.g. loop size) are conserved.

Discussion

Synthetic biology approaches have been increasingly used in recent years to map potential regulatory mechanisms of transcriptional and translational regulation, in both mammalian and bacterial cells. In this work we built on the work of [22] to quantitatively study RBP-based regulation in bacteria using a combined synthetic biology and SHAPE-seq approach. Using our library of RNA regulatory variants, we were able to identify and characterize a position-dependent repression of translational initiation and elongation by a hairpin structure, and strong repression of initiation when the hairpin was bound by an RBP. Characterizing these mechanisms required multiple lines of evidence.

For the translation repression mechanism, we first noted the intriguing observation of 3-nt periodicity in reporter fluorescence levels as a function of the position of the structured hairpin within the N-terminus of the reporter gene. Since other studies have suggested that ribosomes can unwind hairpin structures [37,38], we tentatively attribute the observed periodic effect to position-dependent resistance to unwinding of the mRNA hairpin. Interestingly, the regulatory effect of hairpins inside the ribosomal initiation region has been a source of controversy, with some studies [39,38,40,41,32] suggesting that such structured regions should be detrimental to expression, while others have shown that hairpins do not seem to affect expression in the absence of an RBP [7,42,43] or if the stem is shorter than 6 bp. Our structures are characterized by two stems that are separated by a small 1 nt bulge in the middle that reduces structural stability (e.g. compare PP7-wt and PP7-nb in Fig. 4a), which may partially account for the differences from previous observations. Our 3-nt periodicity finding, which persists to at least +21 from the AUG, suggests that there is still much to be discovered about

the regulatory function of hairpins within coding regions, and their relationship to translation initiation and elongation.

When a hairpin exhibiting the 3-nt periodicity signal is located inside the initiation region and its cognate RBP is present at sufficiently high levels, complete repression is observed. However, this repression is sharply attenuated at +10-11 nt from the AUG, and by +16 no RBP-based repression is observed. This is despite the fact that the underlying 3-nt periodicity persists, indicating that the hairpin structure is still stable. This observation, combined with the *in vivo* SHAPE-seq data showing no detectible modification of the loop-region on the induced PP7-wt δ=5 variant, support a model [5,44–47] in which the bound RBP interrupts initiation (e.g. by preventing binding of one or more of the initiation factors to the small 30S subunit, inhibiting hairpin unwinding, etc.) leaving the 30S subunit bound and in an inactive state. However, once the 70S subunit is fully assembled, a bound RBP-hairpin complex located outside the initiation region is unable to inhibit elongation. These results are consistent with the observations of [42,32], who both showed that structured stems of 6 bps or longer in the N-terminus can silence expression up to +11-13 from the AUG, but show negligible silencing when positioned further downstream. Thus, both our results and previous observations support a model whereby the 30S subunit protects a region up to +11-13 nucleotides from the AUG, and that a combination of a short hairpin and a bound RBP is able to inhibit the initiation process.

Finally, our results open the door for biotechnological applications. We found that GCP, QCP, and PCP are capable of binding mutually-orthogonal binding sites that differ in structure, while MCP is generally more promiscuous. In addition, we characterized the specific *in-vivo* interaction of each RBP—binding-site pair by an effective K_{RBP}, which we found to be independent of binding site location. Interestingly, the *in vivo* K_{RBP} measured for some of the binding sites relative to their native site, differ from past *in vitro* measurements. In particular, PP7-nB and PP7-USs exhibited little or no binding in the *in vitro* setting [9], while displaying strong binding in our assay. Our work thus establishes a blueprint for an *in vivo* assay for measuring the dissociation constant of RBPs with respect to their candidate binding sites. This assay can be used to discover additional binding sites for known RBPs, which could be utilized in synthetic biology applications where multiple non-identical or orthogonal binding sites are needed.

Supplementary Information

Supplementary_Information.pdf contains a description of Supplementary Methods, and Supplementary Figures 1-4.

Supplementary_Table_1.xlsx contains the sequences of all binding-site variants.

Supplementary_Table_2.xlsx contains the sequences of the RBPs used in this work.

Supplementary_Table_3.xlsx contains the sequences of all primers used in the SHAPE-seq protocol.

Supplementary_Data.xlsx contains the SHAPE-seq read-end counts.

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.

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 helped analyze 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.

References

1. Dominissini, D. *et al.* The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530,** 441–446 (2016).
2. Lewis, C. J. T., Pan, T. & Kalsotra, A. RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* **18,** 202–210 (2017).
3. Dethoff, E. A., Petzold, K., Chugh, J., Casiano-Negroni, A. & Al-Hashimi, H. M. Visualizing transient low-populated structures of RNA. *Nature* **491,** 724–728 (2012).
4. Brown, D., Brown, J., Kang, C., Gold, L. & Allen, P. Single-stranded RNA Recognition by the Bacteriophage T4 Translational Repressor, RegA. *J. Biol. Chem.* **272,** 14969–14974 (1997).

5. Schlax, P. J., Xavier, K. A., Gluick, T. C. & Draper, D. E. 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 (2001).
6. Romaniuk, P. J., Lowary, P., Wu, H. N., Stormo, G. & Uhlenbeck, O. C. RNA binding site of R17 coat protein. *Biochemistry (Mosc.)* **26,** 1563–1568 (1987).
7. Cerretti, D. P., Mattheakis, L. C., Kearney, K. R., Vu, L. & Nomura, M. Translational regulation of the spc operon in Escherichia coli. *J. Mol. Biol.* **204,** 309–325 (1988).
8. Sacerdot, C. *et al.* 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 (1998).
9. Lim, F. & Peabody, D. S. RNA recognition site of PP7 coat protein. *Nucleic Acids Res.* **30,** 4138–4144 (2002).
10. Hattman, S., Newman, L., Murthy, H. M. & Nagaraja, V. 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 (1991).
11. Wulczyn, F. G. & Kahmann, R. Translational stimulation: RNA sequence and structure requirements for binding of Com protein. *Cell* **65,** 259–269 (1991).
12. Lucks, J. B. *et al.* Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-seq). *Proc. Natl. Acad. Sci.* **108,** 11063–11068 (2011).
13. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505,** 701–705 (2014).

14. Ding, Y., Kwok, C. K., Tang, Y., Bevilacqua, P. C. & Assmann, S. M. Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat. Protoc.* **10,** 1050–1066 (2015).

15. Flynn, R. A. *et al.* Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat. Protoc.* **11,** 273–290 (2016).

16. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **14,** 75–82 (2017).

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

18. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30,** 521–530 (2012).

19. Patwardhan, R. P. *et al.* Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30,** 265–270 (2012).

20. Levy, L. *et al.* Short CT-rich motifs encoded within σ54 promoters insulate downstream genes from transcriptional read-through. *bioRxiv* 086108 (2017). doi:10.1101/086108

21. Weingarten-Gabbay, S. *et al.* Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351,** aad4939 (2016).

22. Saito, H. *et al.* Synthetic translational regulation by an L7Ae–kink-turn RNP switch. *Nat. Chem. Biol.* **6,** 71–78 (2010).

23. Spitale, R. C. *et al.* RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9,** 18–20 (2013).

24. Watters, K. E., Abbott, T. R. & Lucks, J. B. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-seq. *Nucleic Acids Res.* gkv879 (2015). doi:10.1093/nar/gkv879

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

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

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

28. Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32,** 170–173 (2011).

29. Qu, X. *et al.* The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* **475,** 118–121 (2011).

30. Zeevi, D. *et al.* Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Res* **21,** 2114–2128 (2011).

31. Keren, L. *et al.* Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.* **9,** 701 (2013).

32. Espah Borujeni, A. *et al.* Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Res.* **45,** 5437–5448 (2017).

33. McGinnis, J. L. *et al.* In-cell SHAPE reveals that free 30S ribosome subunits are in the inactive state. *Proc. Natl. Acad. Sci.* **112,** 2425–2430 (2015).

34. Spitale, R. C. *et al.* Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519,** 486–490 (2015).

35. Aviran, S. *et al.* Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Natl. Acad. Sci.* **108,** 11069–11074 (2011).

36. Chao, J. A., Patskovsky, Y., Almo, S. C. & Singer, R. H. Structural basis for the coevolution of a viral RNA–protein complex. *Nat. Struct. Mol. Biol.* **15,** 103–105 (2008).

37. Yusupova, G. Z., Yusupov, M. M., Cate, J. H. D. & Noller, H. F. The Path of Messenger RNA through the Ribosome. *Cell* **106,** 233–241 (2001).

38. Gu, W., Zhou, T. & Wilke, C. O. A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes. *PLOS Comput. Biol.* **6,** e1000664 (2010).

39. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science* **324,** 255–258 (2009).

40. Goodman, D. B., Church, G. M. & Kosuri, S. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science* **342,** 475–479 (2013).

41. Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z. & Bluthgen, N. Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* **9,** 675–675 (2014).

42. Paulus, M., Haslbeck, M. & Watzele, M. RNA stem–loop enhanced expression of previously non-expressible genes. *Nucleic Acids Res.* **32,** e78–e78 (2004).

43. Kelsic, E. D. *et al.* RNA Structural Determinants of Optimal Codons Revealed by MAGE-Seq. *Cell Syst.* **3,** 563–571.e6 (2016).

44. Spedding, G. & Draper, D. E. Allosteric mechanism for translational repression in the Escherichia coli alpha operon. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 4399–4403 (1993).

45. Zengel, J. M. & Lindahl, L. Diverse Mechanisms for Regulating Ribosomal Protein Synthesis in Escherichia coli. *Prog. Nucleic Acid Res. Mol. Biol.* **47,** 331–370 (1994).

46. Marzi, S. *et al.* Structured mRNAs Regulate Translation Initiation by Binding to the Platform of the Ribosome. *Cell* **130,** 1019–1031 (2007).

47. Babitzke, P., Baker, C. S. & Romeo, T. *Regulation of translation initiation by RNA binding proteins*. **63,** (2009).

Figure Captions

Figure 1: Dependence of mCherry expression level on hairpin position

(a) The four hairpins used in this experiment were the native (wt) binding sites for the MS2, PP7, GA, and Qβ coat proteins. Stop codons and start codons inside binding sites are highlighted, in bold and red. (b) Mean mCherry basal expression levels for 26 variants as a function of distance from the A of the AUG. Data for PP7-wt (red), GA-wt (green), Qβ-wt (blue), and MS2-wt (magenta) binding sites as function of hairpin position δ, in the absence of RBPs (non-induced case). Dashed line represents the average basal level for two constructs with a non-structured sequence replacing the binding site downstream of the AUG. Each of the wt binding sites has a stop codon encoded within its sequence [see (a)]. As a result, every binding site has a set of positions (δ=4,7, etc. for MS2-wt, PP7-wt, and Qβ-wt, and δ=6,9, etc. for GA-wt), for which a stop codon in frame with the upstream AUG prevents proper readout of the structure's effect on expression. For this same frame, the start codons inside the hairpins did not result in mCherry expression. CPBS stands for coat-protein binding site (schemas, top). (c) A simple kinetic scheme that can explain the 3-nt periodicity observed for the mCherry expression. The reading frame in which the ribosome encounters the hairpin determines the ultimate rate of mRNA unwinding, thus leading to three possible timescales. See supplementary methods.

Figure 2: Translational regulation by a RBP-hairpin complex in the ribosomal initiation region

(a) Experimental schematic. Top: plasmid expressing the RBP-CP fusion from a pRhlR inducible promoter. Bottom: a second plasmid expressing the reporter plasmid with the RBP binding site encoded within the 5' end of the gene (at position δ>0). (b) Dose-response functions for PCP with a reporter mRNA encoding PP7-wt at three positions: δ=7 (red), δ=11 (blue), and δ=16 (green) nt. (c) A schematic for the mechanistic repression model. The bound RBP (middle) is able to disrupt the formation of the elongating 70S subunit, leaving the bound 30S subunit inactive. If the hairpin is positioned downstream of the initiation region (bottom), the 70S subunit is able to assemble and subsequently unwind the RBP-bound structure at a rate T_{slow}, which is substantially slower than the rate for a standard elongation

step T_{fast}. (d) Fold-repression measurements for PCP (red) and QCP (blue) as a function of hairpin position δ. Fold repression is computed by the ratio of the mCherry rate of production at no induction to the rate of production at full induction. Note, for two constructs (QCP with δ=4 and δ=8) the basal levels without induction were too low for fold-repression measurements.

Figure 3: SHAPE-seq analysis of the PP7-wt binding site in absence and presence of RBP

(a) *In vitro* SHAPE-seq read ratio for the construct containing PP7-wt. (Inset) Structure of the PP7-wt binding site, with yellow and green boxes highlighting regions that are prone to NAI modification. (b) Comparison of *in vivo* SHAPE-seq ratios for induced (blue) and non-induced (red) strains containing a PP7-wt binding site in the δ=5 position. (Inset) Quantitative PCR measurements of the induced (blue) and non-induced (red) strains. See Supplementary Methods for definition of SHAPE-seq read ratio.

Figure 4: Repression effect can be used to estimate an effective dissociation constant K_{RBP}

(a) 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. (b) Dose responses for 175 variants whose basal rate of production levels were > 50 a.u./hr. Each response is divided by its maximal mCherry level, for easier comparison. Variants are arranged in order of increasing fold up-regulation. (c) Normalized K_{RBP} for variants that generated a detectible down-regulatory effect for at least one position. Blue corresponds to low $K_{RBP,}$ while red indicates high $K_{RBP.}$ If there was no measureable interaction between the RBP and binding site, K_{RBP} was set to 1.

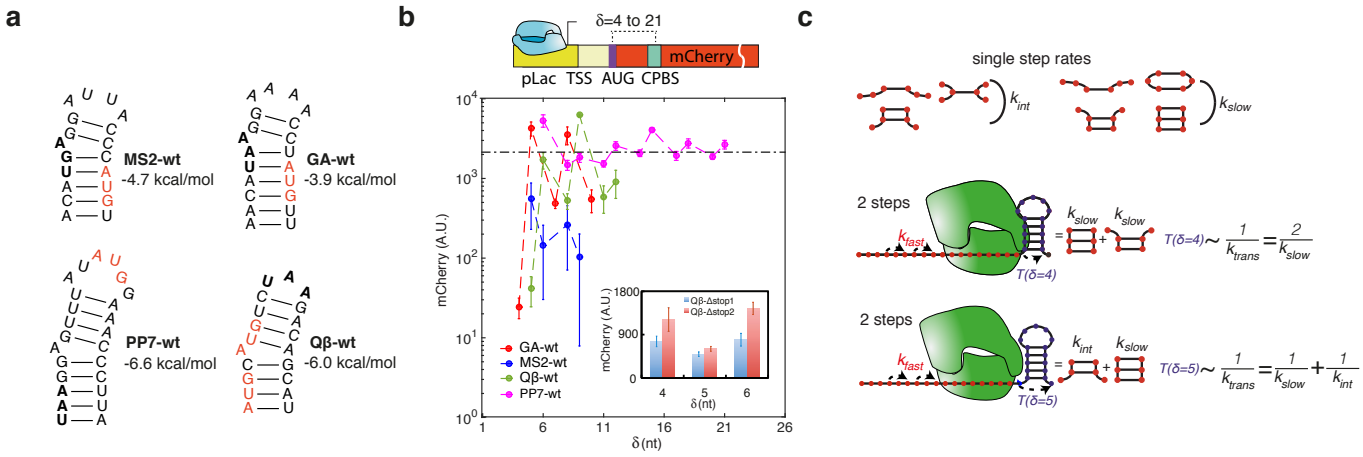


Figure 1

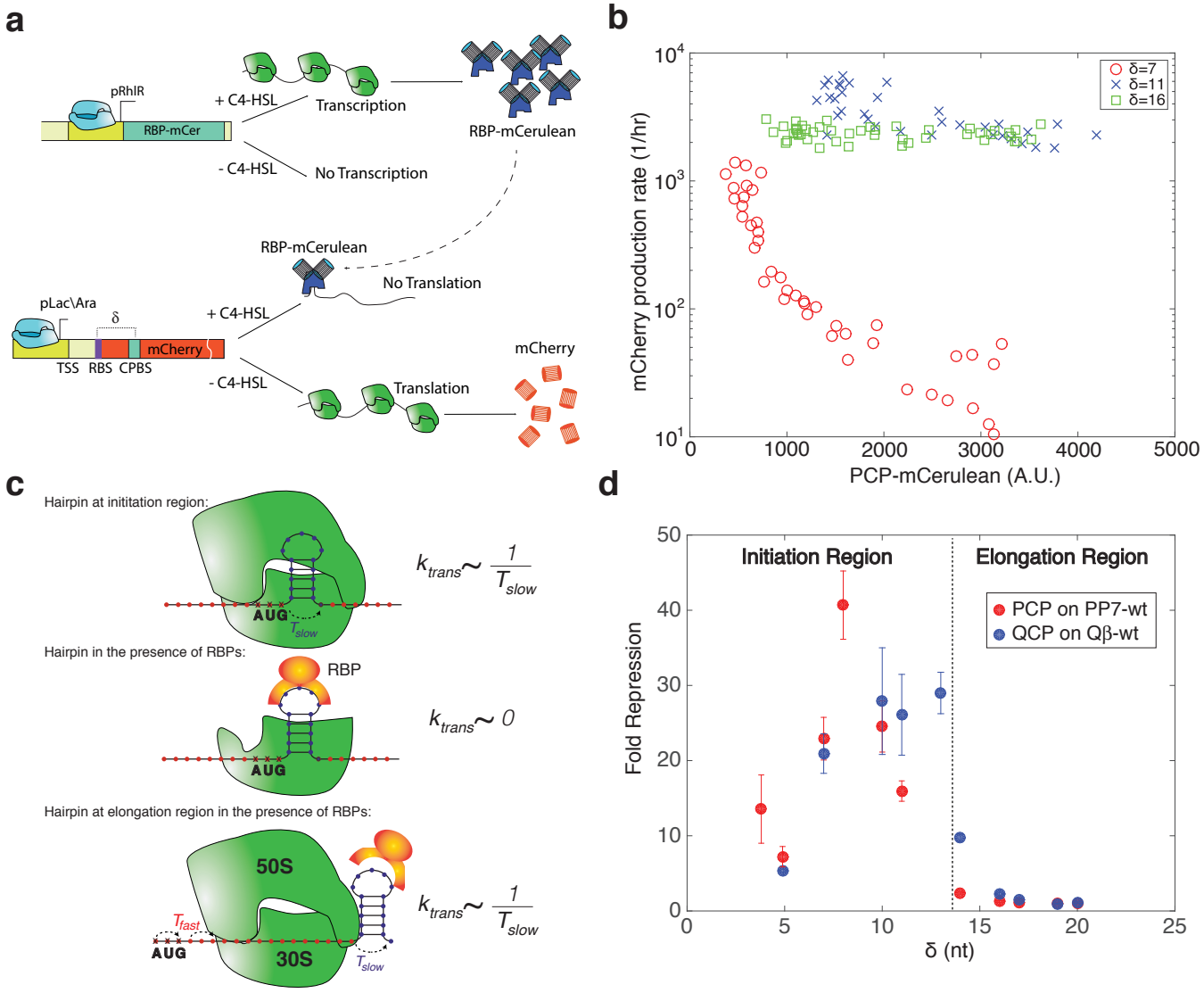


Figure 2

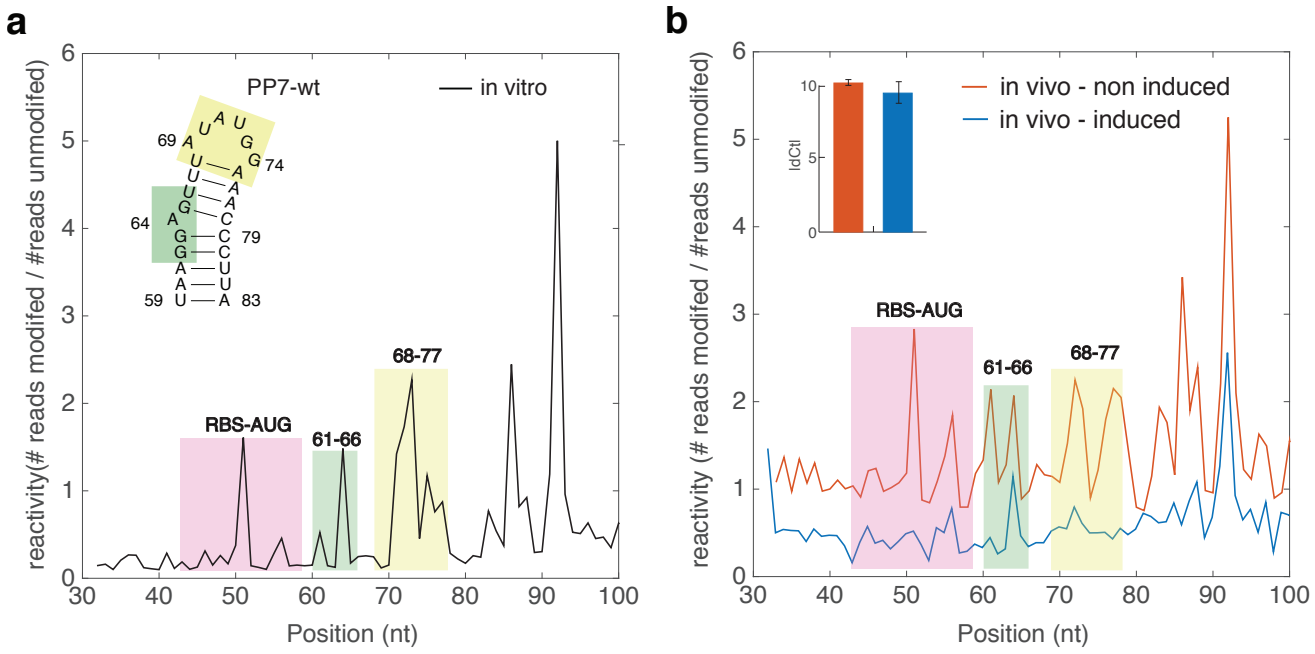


Figure 3

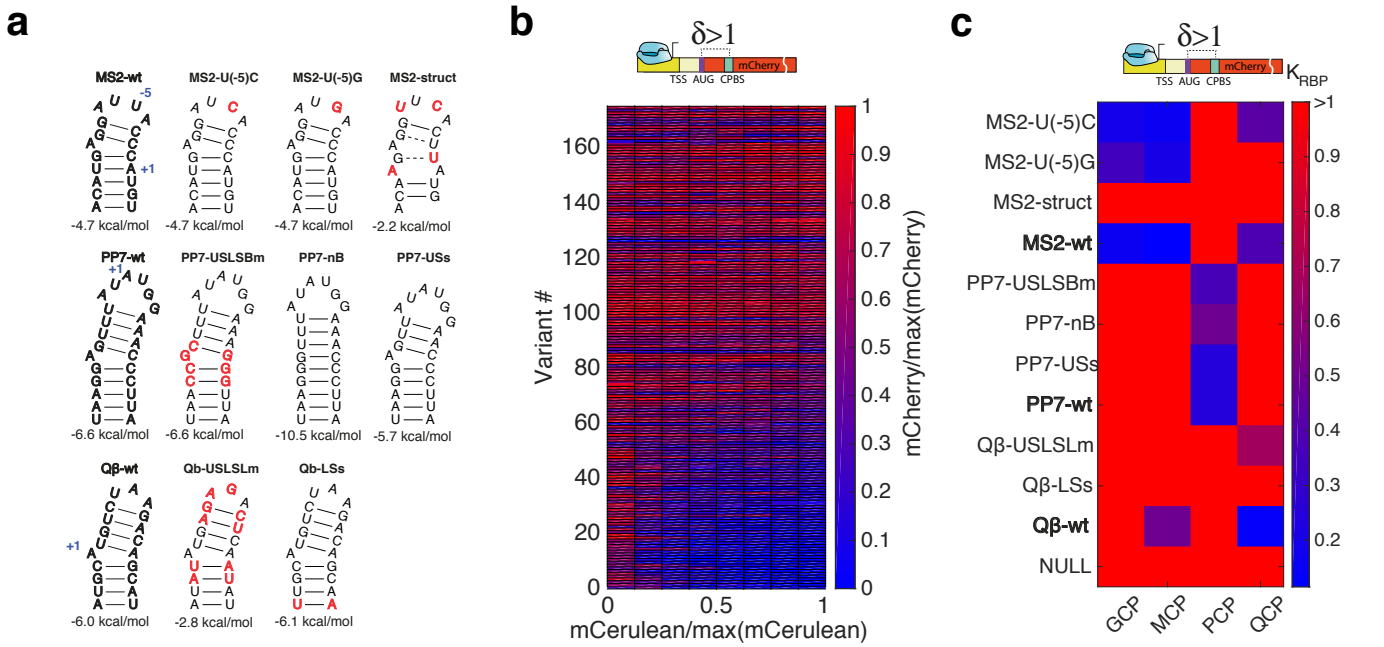


Figure 4

Supplementary Information for paper entitled: An in vivo binding assay for RNA-binding proteins based on repression of a reporter gene

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

Contents

1 Modelling

1.1 Modelling a hairpin inside a coding region

To model translation in the presence of a hairpin, we build on previous single-molecule experiments.[1] The authors proposed a model in which a ribosome has two modes of unwinding an mRNA hairpin: destabilization of the hairpin by biasing towards an open or unstructured state, or mechanically pulling apart the closed RNA strands. The former, faster mode arises when a closed structure opens spontaneously, while the slower mode provides a way for the ribosome to read through highly stable RNA structures. The authors were not able to detect single-nucleotide steps of the ribosome, but did observe clear single-codon steps. They found that the translation rate for the double-stranded RNA (mechanical unwinding mode) was 0.23 a.a./s, while the translation rate for the single-stranded RNA (biasing toward open mRNA mode) was 0.44 a.a./s. Both of these rates are considerably slower than the *in vivo* translation rate for the ribosome on an unstructured mRNA, estimated to be 10-20 a.a./s.[2]

These experimental observations motivated us to model translation of hairpin structure as a set of discrete pair-opening and translation steps.

A toy model for hairpin unwinding

When a ribosome translates, it normally pulls a single-stranded nucleic acid chain through a tunnel as the tRNA is shifted from the aminoacyl ('A') to the peptidyl ('P') site, and from the peptydyl site to the exit ('E') site, by an as-yet-poorly-understood mechanism. However, when it encounters a hairpin structure, one of two things may happen: either the single-strand/double-strand junction opens up spontaneously, or the ribosome uses a forcing mechanism to directly open the RNA molecule. Since each translation step is three nucleotides long, the force-pulling must unwind three base-pairs. This implies that the typical translation time for a trinucleotide encoding an amino acid depends on its precise structure. We propose the following translation times:

$$T_{\text{trinucleotide}} = \begin{cases} T_{\text{ssRNA}} & \text{fully} - \text{unpaired trinucleotide} \\ T_{\text{pos1}-\text{ss}} & \text{partially} - \text{structured trinucleotide, 1st base unpaired} \\ T_{\text{pos1}-\text{ds}} & \text{partially} - \text{structured trinucleotide, 1st base paired} \\ T_{\text{dsRNA}} & \text{fully} - \text{paired trinucleotide} \end{cases}, \quad (1)$$

where $k_{\text{ssRNA}} = 1/T_{\text{ssRNA}}$ and $k_{\text{dsRNA}} = 1/T_{\text{dsRNA}}$ correspond to the translation rates for fully single-stranded and double-stranded trinucleotides, respectively, and we introduce two intermediate rates $k_{\text{pos1}-\text{ss}} = 1/T_{\text{pos1}-\text{ss}}$ and $k_{\text{pos1}-\text{ds}} = 1/T_{\text{pos1}-\text{ds}}$.

The rate for translation of the hairpin is then

$$\begin{aligned} k_{\text{hairpin}}(\delta) &= \frac{1}{\sum_{\text{hairpin trinucleotides}} T_{\text{trinucleotide}}} \\ &= \frac{1}{lT_{\text{ssRNA}} + mT_{\text{pos1}-\text{ss}} + nT_{\text{pos1}-\text{ds}} + pT_{\text{dsRNA}}} \\ &= \frac{1}{l/k_{\text{ssRNA}} + m/k_{\text{pos1}-\text{ss}} + n/k_{\text{pos1}-\text{ds}} + p/k_{\text{dsRNA}}}, \end{aligned} \quad (2)$$

where where $l(\delta), m(\delta), n(\delta)$, and $p(\delta)$ correspond to the number of fully-single-stranded, partially-structured with first-position unpaired, partially-structured with first-position paired, and fully-double-stranded trinucleotides encoding the hairpin, respectively. We assume a separation of timescales

$T_{\mathrm{dsRNA}} \gtrsim T_{\mathrm{pos1-ds}} \gg T_{\mathrm{pos1-ss}} \gg T_{\mathrm{ssRNA}}$. We note that a previous experiment[1] studied only the fully single- and double-stranded cases, so this assumption has to be verified experimentally.

We show the various hairpin reading frames cases that correspond to translation rates a function of hairpin position in Fig. 1c. By varying the position of the hairpin relative to the reading frame, different values of n are encountered, resulting in different mean translation rates:

$$\begin{aligned} k_{\mathrm{hairpin}}(\delta=4) &= \frac{1}{T_{\mathrm{pos1-ds}}+T_{\mathrm{dsRNA}}} \simeq \frac{1}{2T_{\mathrm{dsRNA}}} \equiv \frac{k_{\mathrm{slow}}}{2} \\ k_{\mathrm{hairpin}}(\delta=5) &= \frac{1}{T_{\mathrm{pos1-ss}}+T_{\mathrm{dsRNA}}} \equiv \frac{1}{1/k_{\mathrm{int}}+1/k_{\mathrm{slow}}} \quad . \end{aligned} \tag{3}$$

The model assumes that the translation rate of the entire mRNA molecule is dominated by the translation rate of a single hairpin, located not too far from the 5' end of the transcript in our experiments, and that the number of mechanical-opening steps n determines the translation rate. The effect of the hairpin dominates if $k_{\mathrm{dsRNA}} \ll k_{\mathrm{ssRNA}}$, which is consistent with the value for $k_{\mathrm{dsRNA}} \sim 0.23$ a.a./s quoted for the fully-paired trinucleotide,[1] and a value of k_{ssRNA} ∼10-20 a.a./s.[2] We note that the *in vitro* value of 0.44 a.a./s measured for k_{ssRNA}[1] is much lower than the typical *in vivo* values. This is not that surprising given the many differences between the *in vivo* and *in vitro* settings. We expect the *in vitro* estimate of k_{dsRNA} to be more accurate because it should depend mainly on the binding energy of the mRNA structure, which was probed directly in the *in vitro* experiment by applying force using optical tweezers to open the mRNA. We do not know why the hairpin dominates over other possible RNA structures encoded in the transcript. It may be more stable, or it may dominate because the mRNA structure has more time to form at positions close to the initiation region, unlike regions far from initiation for which elongation and transcription are known to be coupled.[3]

An alternative explanation for the position-dependence of the translation rate is that the different rates are the result of the use of 3 different codon sets for the hairpin, that arise as the unchanged hairpin sequence is shifted through the possible reading frames. It is hard to compensate for the change in codon set since the hairpin sequence encodes structure that we wish to preserve. Determining which of these models is correct would require the detailed analysis of many more hairpin sequences, and possibly single-molecule measurements.

1.2 Translation in the presence of an RBP-bound hairpin complex in the coding region

In order to describe the effects of a hairpin bound by an RNA binding protein, we will make the following assumptions:

1. If the hairpin is outside of the initiation region, the bound RBP has no regulatory effect and the rate of translation is determined by the structural characteristics of the hairpin as described above.

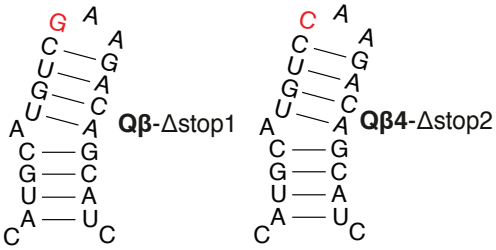
2. If the hairpin is within the initiation region (i.e $\delta \leq 10$) then a bound RBP can abolish translation entirely. The effect of the RBP reduces to preventing the large 50S ribosomal subunit from binding to the 30S subunit, thus inhibiting the formation of the translationally active 70S ribosomal unit.

Given these assumptions, we have a two-state model for translation in the initiation region:

$$k_p = \begin{cases} k_{\mathrm{hairpin}} & \text{RBP unbound} \\ 0 & \text{RBP bound} \end{cases} , \tag{4}$$

where k_{hairpin} is given by Eq. 2.

Note, that a recent paper[2] proposed that fully-formed 70S subunits may also be active in translation by scanning from the 5' end of the mRNA (even in the absence of an available RBS). This would imply that some residual translation should be always present depending on the percentage of alternate 70S scanning events ($k_p > 0$ even for RBS-bound case), since we showed that RBPs that are bound inside the elongation region cannot inhibit translation. The very large fold-repression observations (>10) combined with the extremely low expression levels that we observed at full induction indicate that we are not observing such scanning events in our experiments, or, if they do exist, that they contribute negligibly to the total translation process.



Supplementary Figure 1: Qβ binding sites with STOP codon mutated (mutated nucleotide in red) so that all 3 reading frames yield reporter product.

2 Fluorescence expression assay

2.1 Design and construction of binding-site plasmids

Binding-site cassettes (see 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).

2.2 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 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) .[4] 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.

2.3 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).

2.4 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 [5, Supplementary Information], as follows (see Supp. Fig. 2):

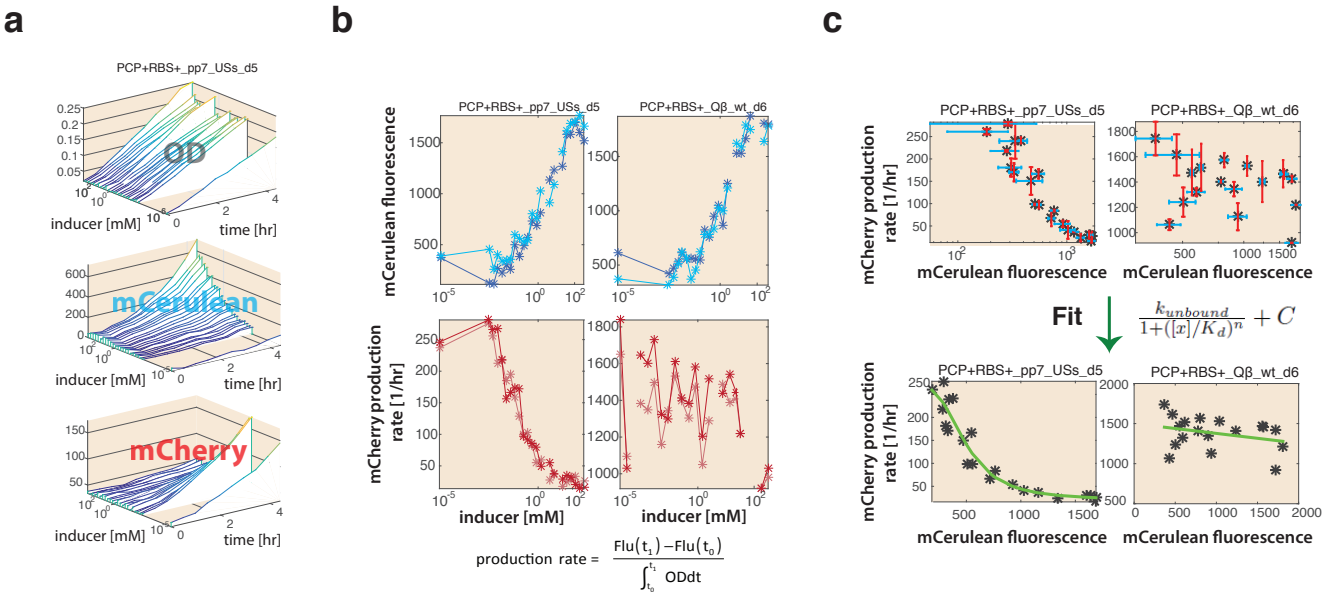
mCerulean average normalized fluorescence: for each inducer concentration, mCerulean measurements were normalized by OD. Normalized measurements were then averaged over the N logarithmic-growth timepoints in the interval $[T_0, T_{final}]$, yielding:

$$\hat{mCerulean} = \frac{1}{N} \sum_{t=T_0}^{T_{final}} \frac{mCerulean(t)}{\mathrm{OD}(t)} \,. \tag{5}$$

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:

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

Finally, we plotted mCherry rate of production[6] 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.



Supplementary Figure 2: Estimation of K_d from fluorescence data. (a) 3D plots depicting OD levels (top), mCerulean fluorescence (middle), and mCherry fluorescence (bottom) as a function of time and inducer concentration. (b) (top) mCerulean steady state expression levels is computed by dividing each fluorescence level by the respective OD and averaging all values in the 3-4 hr range for two strains. (bottom) mCherry production rate is computed according to the formula between 2-4 hrs after induction. (c) (top) mCherry production rate is plotted as a function of mean mCerulean fluorescence averaged over two biological duplicates for two strains. (bottom) Fit for K_d using the fitting formula is shown for both strains, with only the one on the left exhibiting a specific binding response.

2.5 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:

$$mCherry\,rate\,of\,production = P_{bound}k_{bound} + P_{unbound}k_{unbound}\,. \tag{6}$$

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. 6 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.

The final model used for fitting was:

$$m\,Cherry\,rate\,of\,production \simeq \frac{k_{unbound}}{1 + ([x]/K_d)^n} + C\,,$$

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.

3 Quantitative PCR (qPCR)

3.1 RNA extraction and reverse-transcription

Starters of *E. coli* TOP10 containing the relevant constructs on plasmids were grown in LB medium with appropriate antibiotics overnight (16 hr). The next morning, the cultures were diluted 1:100 into fresh semi-poor medium (see Section 2.4), and grown for five hours. For each isolation, RNA was extracted from 1.8 ml of cell culture using 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. RNA quality was assessed by running 500 ng on 1% agarose gel. After extraction, RNA was subjected to DNAse (Ambion/Life Technologies) and then reverse-transcribed using MultiScribe Reverse Transcriptase and random primer mix (Applied Biosystems/Life Technologies). For qPCR experiments, RNA was isolated from three individual colonies for each construct.

3.2 qPCR measurements

Primer pairs for mCherry and normalizing gene *idnT* were chosen using the Primer Express software, and BLASTed (NCBI) with respect to the *E. coli* K-12 substr. DH10B (taxid:316385) genome (which is similar to TOP10) to avoid off-target amplicons. qPCR was carried out on a QuantStudio 12K Flex machine (Applied Biosystems/Life Technologies) using SYBR-Green. Three technical replicates were measured for each of the three biological replicates. A C_T threshold of 0.2 was chosen for all genes.

4 SHAPE-seq

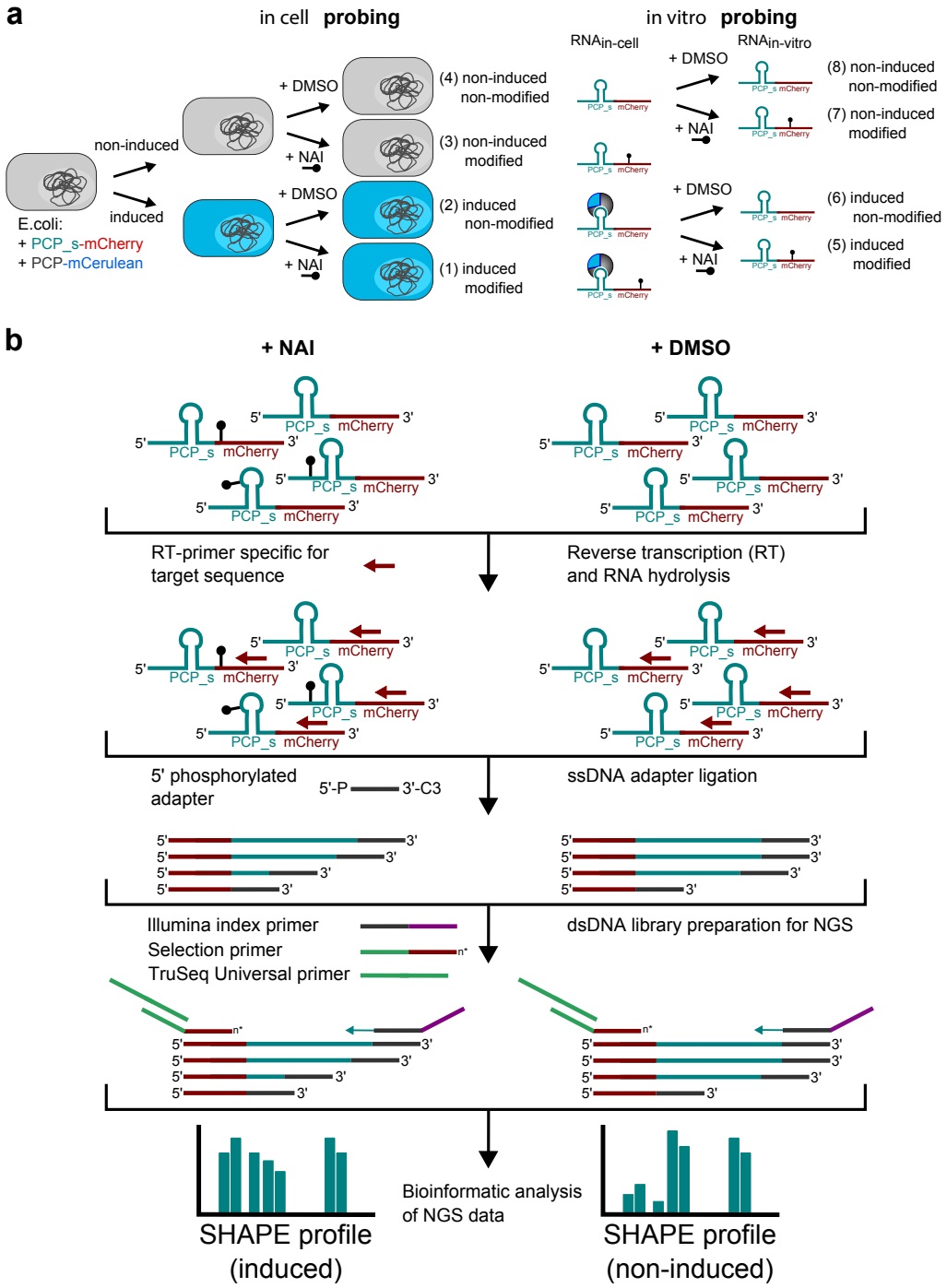
4.1 Experimental setup

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 2.1) and the RBP-fusion plasmid (see Section 2.2 and Table S3 for details of strains, and Supp. Fig. 3), 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 2.4). 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),[7] 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. TRIzol-based RNA isolation was performed as described above and used for the strain harboring RBS+_PP7_wt_d4, while the 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 (Supp. Fig. 3b):

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 .[8] 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.



Supplementary Figure 3: Schematic overview of SHAPE-seq experiment. (A) Overnight-grown bacterial strains harboring both the RBP-binding site plasmid (PP7-wt, δ=5 containing the mCherry reporter) and the RBP-fusion plasmid (PCP-mCerulean) are split into two samples and PCP-mCerulean expression are induced (using C_4-HSL) for one of them. Following protein expression, each bacterial sample is further split and treated with either DMSO (as the non-modified control) or NAI. Subsequently, RNA is isolated and either further chemically probed (samples 2+4) or directly used for subsequent steps of SHAPE-seq (samples 1+3). (B) Following 2' hydroxyl acylation and subsequent RNA isolation, RNA samples are reverse-transcribed using a gene-specific primer that binds in the target transcript. During reverse transcription, reverse transcriptase is stalled one nucleotide before the modification. Subsequently, a single-stranded 5' phosphorylated (5'P) and 3-carbon spacer (3'C) adapter sequence is ligated to the obtained cDNAs, which serves in the next step as a primer-binding site for the Illumina index primers to prepare double-stranded DNA for Illumina next generation sequencing.

Subsequent steps of the SHAPE-seq protocol, that were applied to all samples, have been described elsewhere [9, 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 Table S4) for mCherry (#1) or 5S rRNA (#2) that are specific for either the mCherry transcript (RBS+_PP7_wt_d4, RBS-_PP7_USs_d-29, RBS-_PP7_wt_d-29) or 5S rRNA transcript (used as a positive control for SHAPE-seq). 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.

4.2 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 Table S4), 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) or #5 (primer extends 3 nucleotides into 5S rRNA transcript), 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.

4.3 Bioinformatics analysis

Illumina reads were first adapter-trimmed using cutadapt [10] and were aligned against a composite reference built from mCherry, *E. coli* 5S rRNA sequences, and PhiX genome (PhiX is used as a control sequence in Illumina sequencing). Alignment was performed using bowtie2 [11] 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). Interestingly, the 5S rRNA library which was used as a positive control yielded ∼6,000,000 reads after trimming. The difference between the two targets 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 ratio was calculated for each strain s at each position i, as follows:

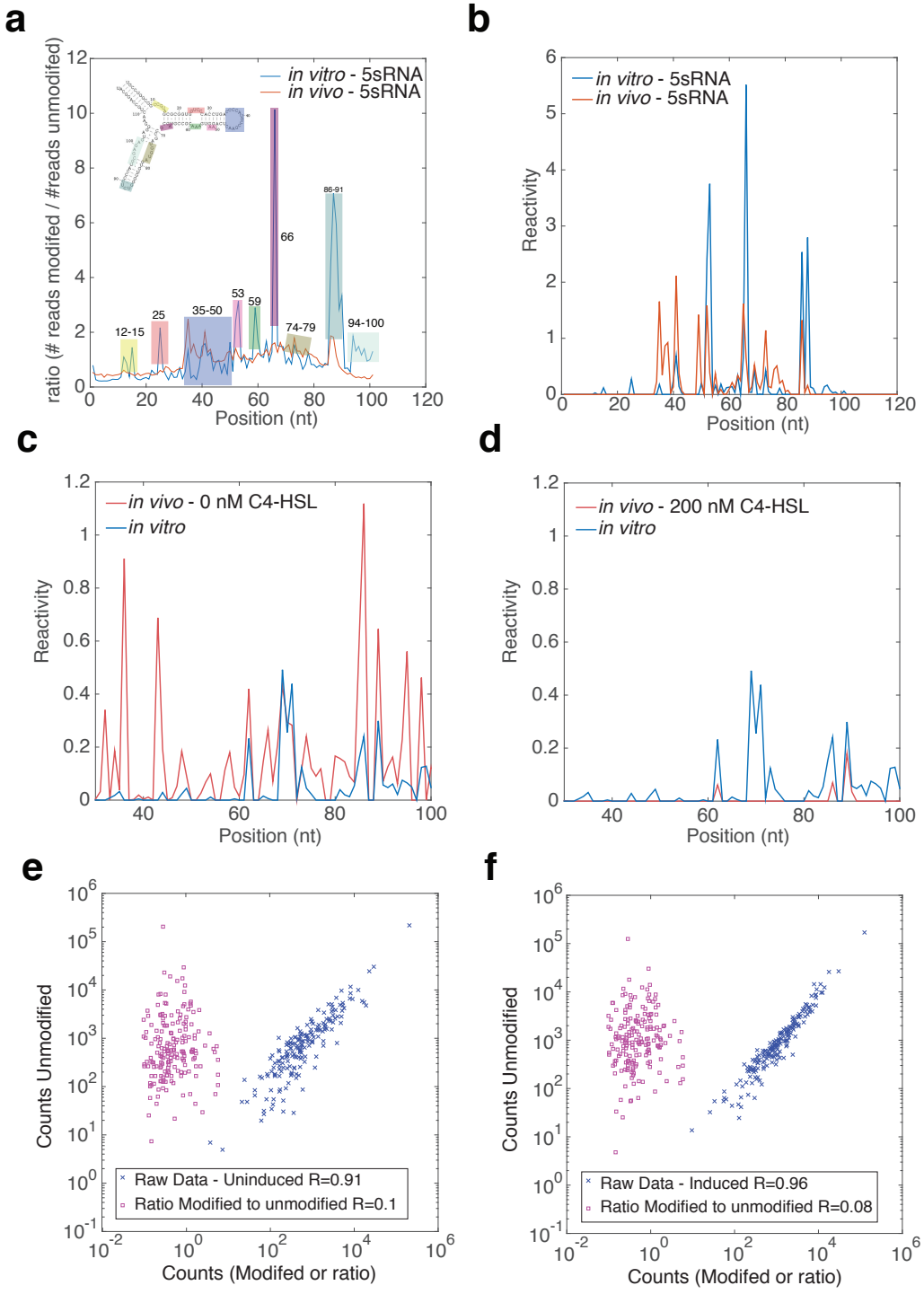
$$R_s(i) = \frac{\hat{D}_{s,mod}(i)}{\hat{D}_{s,non-mod}(i)} \,, \tag{7}$$

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 4.1). The normalized reads as a function of position from the TSS are supplied in Supplemental Data 1.

4.4 Reactivity analysis

SHAPE-seq reactivity scores were calculated using SEQualyzer .[12] Briefly, SEQualyzer calculates SHAPE-seq reactivity scores by first modelling RT drop-out at a particular position as Poisson distribution, and then attempting to infer the probability for a base to be single-stranded . High reactivity score corresponds to high probability for single-stranded base, and low reactivity means

high probability for the base to reside in a double-stranded region of the RNA molecule .[9,13,14] SEQualyzer uses $\hat{D}_{s,mod}(i)$ and $\hat{D}_{s,non-mod}(i)$ as inputs. We compared the performance of the reactivity score to our simpler ratio analysis (Eq. 7). We assumed that in general modifications and RT drop-out are not correlated, thus dividing the modified signal by the unmodified removes the RT-drop out signal, while retaining the reactivity information. In Supp. Fig. 4, we show that for the case of 5S RNA and the PP7wt binding site with δ=5 construct, this operation indeed led to complete decorrelation, thus validating our approach. While the reactivity score based on SEQualizer has been used to analyze previous data, we are of the opinion that an unfiltered approach is more beneficial for the type of analysis that we carried out in this paper.



Supplementary Figure 4: SHAPE-seq ratio vs reactivity computations. (a). Comparison of SHAPE-seq ratio for *E. coli* 5S RNA for *in vitro* (blue) and *in vivo* (red) modified samples. (b) Reactivity score computation for the 5S RNA modification data showing a similar pattern. (c-d) Reactivity scores for the non-induced (c) and induced (d) PP7-wt-δ=5 samples. Reactivity scores for *in vitro* (blue) and *in vivo* (red). (e-f) Plots depicting the the unmodified counts (y-axis) as a function either the modified counts (blue x symbols) or modified to unmodifed ratio (magenta squares) for both the non-induced (e) and induced (f) PP7-wt-δ=5 samples. Data show no correlation between the unmodified library score to our ratio score providing support for our analysis method. All reactivity computations were based on based on Aviran et al. 2011.[13]

References

[1] Xiaohui Qu, Jin-Der Wen, Laura Lancaster, Harry F. Noller, Carlos Bustamante, and Ignacio Tinoco. The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature*, 475(7354):118–121, July 2011.

[2] Hiroshi Yamamoto, Daniela Wittek, Romi Gupta, Bo Qin, Takuya Ueda, Roland Krause, Kaori Yamamoto, Renate Albrecht, Markus Pech, and Knud H. Nierhaus. 70s-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria. *Proceedings of the National Academy of Sciences*, 113(9):E1180–E1189, March 2016.

[3] Katelyn McGary and Evgeny Nudler. RNA polymerase and the ribosome: the close relationship. *Current Opinion in Microbiology*, 16(2):112–117, April 2013.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

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

[12] Krishna Choudhary, Luyao Ruan, Fei Deng, Nathan Shih, and Sharon Aviran. SEQualyzer: interactive tool for quality control and exploratory analysis of high-throughput RNA structural profiling data. *Bioinformatics*, 33(3):441–443, February 2017.

[13] Sharon Aviran, Cole Trapnell, Julius B. Lucks, Stefanie A. Mortimer, Shujun Luo, Gary P. Schroth, Jennifer A. Doudna, Adam P. Arkin, and Lior Pachter. Modeling and automation of sequencing-based characterization of RNA structure. *Proceedings of the National Academy of Sciences*, 108(27):11069–11074, July 2011.

[14] Julius B. Lucks, Stefanie A. Mortimer, Cole Trapnell, Shujun Luo, Sharon Aviran, Gary P. Schroth, Lior Pachter, Jennifer A. Doudna, and Adam P. Arkin. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proceedings of the National Academy of Sciences*, 108(27):11063–11068, July 2011.