Cell Research (2014) :1-4.

www.nature.com/cr

LETTER TO THE EDITOR

Labeling native bacterial RNA in live cells

Cell Research advance online publication 15 April 2014; doi:10.1038/cr.2014.

Dear Editor,

Labeling RNA molecules in their physiological environment is still a technological challenge that should be overcome to study kinetics, localization and protein interactions of these ubiquitous cellular regulators. The only non-invasive method applicable for detecting unmodified RNAs in live cells described so far used two RNA-binding pumilio proteins directly interacting with RNA that triggered complementation of the split fluorescent protein (FP) fused to pumilio proteins [1, 2]. To use this as a universal method, the pumilio proteins should be subjected to mutagenesis to bind each new target RNA. Protein mutagenesis is a time-consuming process and it limits application of this approach. Another method that used RNA-templated protein complementation has been developed but used for *in vitro* RNA detection only [3]. Therefore, detection of unmodified RNAs *in vivo* remains a daunting problem.

Here, we describe a universal method for labeling unmodified RNAs in live cells based on combination of the split aptamer approach and protein complementation. The principle of this method consists of sequence-specific binding of two RNA probes complementary to two adjacent sites on an unmodified RNA target. Each RNA probe is made of two modules connected by a flexible linker: one module is a sequence complementary to the target; the second is a fragment of the split aptamer. When target RNA is present in the cell, the head-to-tail binding of the two RNA probes to the target brings the two fragments of the split aptamer into close proximity, triggering its reassembly. The reassembled aptamer then initiates the association of two split fusion proteins, each containing a fragment of a split FP and a fragment of an RNA-binding protein. As a result, the two non-fluorescent fragments of FP reassociate and become fluorescent (Figure 1A). In all experiments, we used the split RNA-binding protein eIF4A fused with split EGFP and the eIF4A-specific aptamer, the same system we used in our previous RNA-labeling methods [4-6]. We tested two designs for RNA probes: in one case, the probes were expressed as two separate transcripts, in another they were expressed within one long transcript with an unrelated intervening sequence. We found that the second design significantly reduced the fluorescent background. This effect has been studied in detail (see Supplementary information, Data S1) and explained by the interactions of the split fusion proteins with the RNA probes containing split aptamer sequence (Supplementary information, Figure S1). Apparently, the sequestration of split fusion proteins keeps them apart and decreases their spurious association (Supplementary information, Figure S2). Modeling experiments supported the suggested mechanism (Supplementary information, Figure S3).

To test the feasibility of this approach, we targeted a 22 nt-long accessible site in rabbit β-globin mRNA [7] expressed from a plasmid (Supplementary information, Figure S4). The *E. coli* BL21(DE3) cells were transformed with three compatible plasmids expressing mRNA, RNA probes and fusion proteins. All components of the complex were expressed from the T7 promoters upon simultaneous induction with IPTG (for the details see Supplementary information, Data S1). The results showed that β-globin mRNA or its fragment was detected sequence-specifically in 20% of cells (Supplementary information, Figure S4D). It should be noted that the concentration of this plasmid-expressed transcript was ~75 molecules per cell (Supplementary information, Figure S5), which is at least one order of magnitude higher than the average concentrations of endogenous bacterial mRNAs. Therefore, the endogenous *PstC* mRNA expressed from its natural chromosomal site was chosen as the next target. PstC protein is an integral transmembrane transporter. It is a part of the *pst* operon consisting of 5 genes that mediate translocation of inorganic phosphorus, P_i, through the inner membrane [8, 9]. At normal P_i concentrations, genes from the *pst* operon are expressed at a low level from the internal promoters, but under phosphate shortage transcription is induced and is initiated at a promoter located upstream of the first gene in the operon, *PstS* [8]. Based on Mfold analysis of the secondary structure of *PstC* mRNA, the 5′ end has been chosen as a most accessible binding site (Figure 1B). The probe containing corresponding antisense sequences was

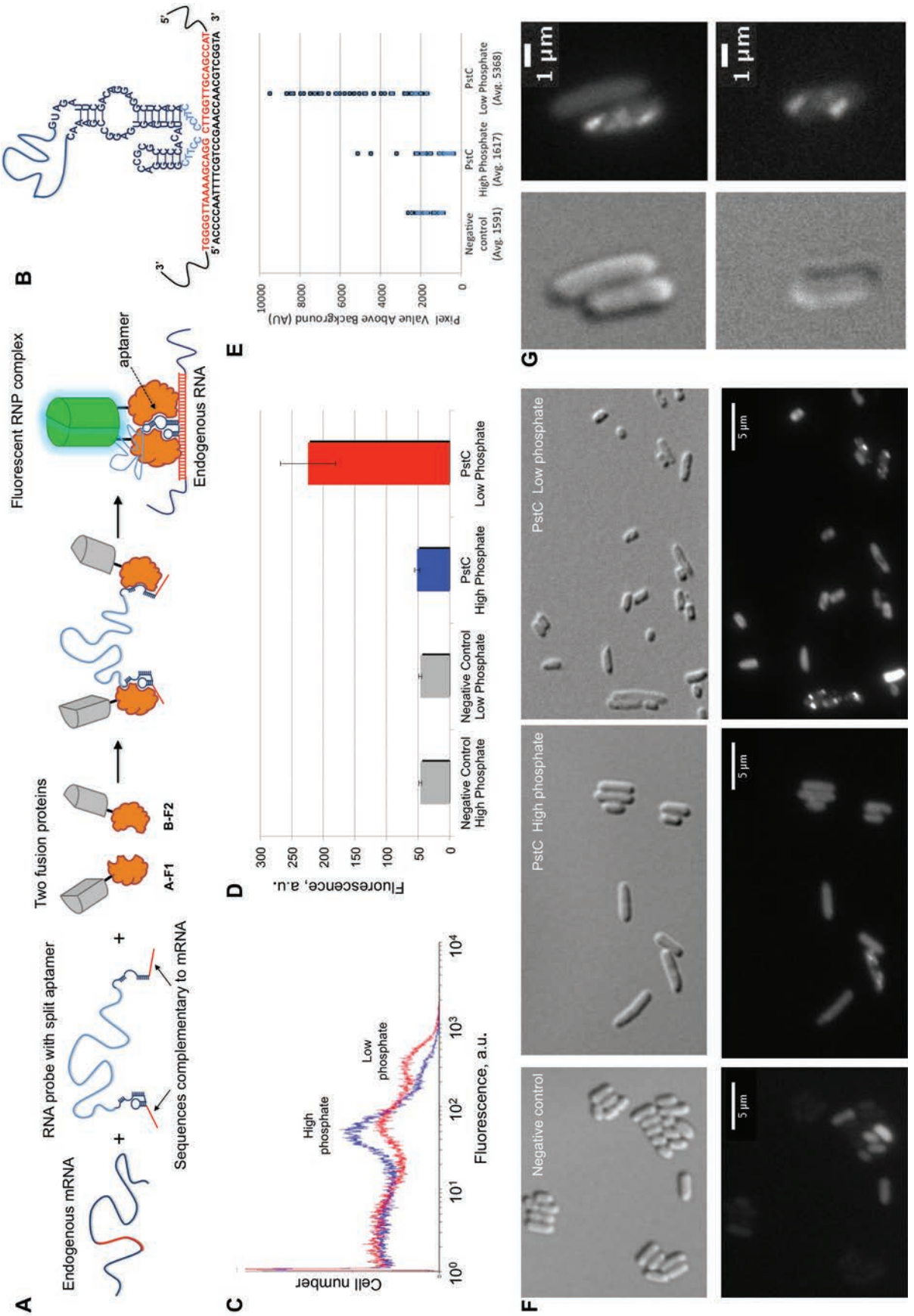
A
Endogenous mRNA
RNA probe with split aptamer
Sequences complementary to mRNA
Two fusion proteins
A-F1
B-F2
Fluorescent RNP complex
aptamer
Endogenous RNA
B
5'
3'
TGGGGTTAAAAGCAGG CTTGGTTGCAGCCAT
5' ACCCCAATTTTCGTCCGAACCAACGTCGGTA 3'
C
Cell number
High phosphate
Low phosphate
Fluorescence, a.u.
D
Fluorescence, a.u.
Negative Control High Phosphate
Negative Control Low Phosphate
PstC High Phosphate
PstC Low Phosphate
E
Pixel Value Above Background (AU)
Negative control (Avg. 1591)
PstC High Phosphate (Avg. 1617)
PstC Low Phosphate (Avg. 5368)
F
Negative control
PstC High phosphate
PstC Low phosphate
5 µm
G
1 µm

Figure 1 Detection and localization of bacterial *PstC* mRNA in live *E. coli* cells. **(A)** Schematics of native RNA recognition by the target-specific RNA probes followed by protein complementation; target RNA hybridizes with RNA probes that have two antisense sequences fused to split aptamer, and this triggers the reassembly of the eIF4A-specific aptamer; aptamer reassembly is followed by reassembly of the eIF4A protein and EGFP; A and B are the fragments of EGFP, F1 and F2 are fragments of the eIF4A protein. **(B)** Structure of the *PstC* mRNA target site and of the eIF4A-specific aptamer split at the central loop; antisense sequences are shown in red, linker sequences are shown in blue, the split site in the aptamer is marked by the curled line. **(C)** FACS analyses of *E. coli* cells with labeled *PstC* mRNA grown in high (blue) and low (red) phosphate media. **(D)** Histograms obtained from the FACS data. Average of 3 independent experiments ± SD is shown. **(E)** Total cell fluorescence obtained from the fluorescent images of single cells using fluorescent microscopy. **(F)** Fluorescent patterns of labeled *PtsC* mRNA in cells in high and low phosphate media. The left panels show cells expressing fusion proteins and the probes antisense to β-globin mRNA (negative control). Scale bar, 5 μm. **(G)** Comparison of *PtsC* mRNA patterns in live (top) and fixed cells (FISH results, bottom). Scale bar, 1 μm.

transformed into *E. coli* cells together with the plasmid pMB53 expressing fusion proteins, A-F1 and B-F2. The cells were grown in media with normal concentration of P_i (2.2 mM) or in media with limited phosphate (0.2 mM). Under normal phosphate conditions, cell fluorescence was only marginally higher than the negative controls (cells expressing fusion proteins and the probes to the β-globin mRNA), while under limited phosphate the average fluorescence increased by about 5-fold as compared to cells grown under normal phosphate concentrations (Figure 1D). This increase in fluorescence correlated with the increase in *PstC* mRNA concentration as determined by RT-PCR (Supplementary information, Figure S5B). At the same time, fluorescence of the negative control cells expressing probes against β-globin mRNA was the same in both high and low phosphate media. Quantitative analyses of single cell fluorescence correlated with FACS results (Figure 1D and 1E).

Microscope images revealed single or several punctate fluorescent signals in about 8% of the cells grown in normal phosphate that were not seen in the control cells (Figure 1F). When the cells were grown in low phosphate media, about 15% of cells revealed signal with a population of cells with oversaturated levels of fluorescence. Inspection of these cells with lower exposure time confirmed similar localization of RNA in small focal points (Supplementary information, Figure S6).

To verify *PstC* mRNA localization by an alternative method, we performed fluorescent *in situ* hybridization (FISH) using 38 TAMRA-labeled probes (Stellaris) specific for this mRNA (Figure 1G and Supplementary information, Table S7, Figure S7). FISH images of the cells grown in normal P_i conditions did not show any signal, while FISH images of the cells grown in phosphate-depleted media revealed characteristic signals in about 25% of the cells very similar to the signals obtained in live cells (Figure 1G).

Thus, our results demonstrate that in live cells we detected signal from *PstC* mRNA with the average RNA concentration of ≤1 molecule per cell. The fluorescence quenching results show abrupt bleaching of the signals (Supplementary information, Figure S8) supporting our hypothesis on single-molecule sensitivity. At the same time, FISH with Stellaris probes was unable to reveal the signal at low *PstC* mRNA concentration. These results underscore the limitations of hybridization methods that use probes with constitutive fluorescent signal, which require washing steps to reveal the specific signal. These limitations of the pre-labeled probes have motivated the efforts to develop molecular sensors that display signal in the presence of the analyte only [3].

We explain the high sensitivity of the new method by the combination of three factors. First, protein complementation reduces fluorescent background by 10-100-fold as compared to the full-size FPs [4, 10]. Second, the mechanism of split protein sequestration by the RNA probes prevents split protein reassociation in the absence of the target RNA, leading to additional 4-5-fold background reduction. Third, *PstC* mRNA is localized, therefore its fluorescence signal is not spread to the entire cell by diffusion during the image acquisition time and is not overwhelmed by cellular autofluorescence. The fact that *PstC* mRNA localization patterns revealed by FISH and our new method look similar suggests that the labeling complex does not interfere with the normal RNA localization. DNA labeling with Hoechst dye showed that *PstC* mRNA does not co-localize with the bulk DNA in both live and fixed cells (Supplementary information, Figure S9). That provides additional assurance that the live cell imaging reflects proper RNA localization. We should emphasize that our method unlike other approaches does not require any RNA modification and it targets short (25-30 nt) RNA sequences. This makes this method applicable not only for mRNA detection but also for other RNA species including short ncRNAs.

Our experiments revealed distinct localization patterns of the *PstC* mRNA in *E. coli* cells, adding more evidence to the growing data on spatial localization of several bacterial mRNAs [11-13]. Further experiments are in progress to address the mechanism of *PstC* mRNA local-

ization.

To the best of our knowledge, this work is the first example of RNA-templated reassembly of split aptamers *in vivo*, which increases the set of tools for the regulated manipulation of a signal depending on the presence of a user-defined RNA target.

Acknowledgments

We are thankful to Meni Wanunu and Andrey Ivankin (Northeastern University, USA) for help with the quenching experiments, and to Maxim Frank-Kamenetskii and Dimitry Beglov for valuable comments on the manuscript. This study was supported by the NSF grant (0244045099) to NEB.

Paul Toran[1], Irina Smolina[2], Harry Driscoll[2], Feng Ding[3], Yingjie Sun[2], Charles R Cantor[4], Natalia E Broude[1, 2]

[1]Cell Biology, Molecular Biology and Biochemistry Program, Boston University, Boston, MA 01225, USA; [2]Department of Biomedical Engineering, Boston University, Boston, MA 01225, USA; [3]Department of Physics and Astronomy, 118 Kinard Lab, Clemson University, Clemson, SC 29634, USA; [4]Sequenom Inc, 3595 John Hopkins Court San Diego, CA 92121-1331, USA

Correspondence: Natalia E Broude[a], Irina Smolina[b]
[a]E-mail: nebroude@bu.edu
[b]E-mail: ismolina@bu.edu

References

1 Ozawa T, Natori Y, Sato M, *et al*. *Nat Methods* 2007; **4**:413-419.
2 Tilsner J, Linnik O, Christensen NM, *et al*. *Plant J* 2009; **57**:758-770.
3 Furman JL, Badran AH, Ajulo O, *et al*. *J Am Chem Soc* 2010; **132**:11692-11701.
4 Valencia-Burton M, McCullough RM, Cantor CR, *et al*. *Nat Methods* 2007; **4**:421-427.
5 Valencia-Burton M, Shah A, Sutin J, *et al*. *Proc Natl Acad Sci USA* 2009; **106**:16399-16404.
6 Borogovac A, Broude NE. *Methods Mol Biol* 2011; **714**:189-199.
7 Allawi HT, Dong F, Ip HS, *et al*. *RNA* 2001; **7**:314-327.
8 Aguena M, Yagil E, Spira B. *Mol Genet Genomics* 2002; **268**:518-524.
9 Spira B, Aguena M, de Castro Oliveira JV, *et al*. *Mol Genet Genomics* 2010; **284**:489-498.
10 Yiu H-W, Demidov VV, Toran P, *et al*. *Pharmaceuticals* 2011; **4**:494-508.
11 Montero Llopis P, Jackson AF, Sliusarenko O, *et al*. *Nature* 2010; **466**:77-81.
12 Nevo-Dinur K, Nussbaum-Shochat A, Ben-Yehuda S, *et al*. *Science* 2011; **331**:1081-1084.
13 dos Santos VT, Bisson-Filho AW, Gueiros-Filho FJ. *J Bacteriol* 2012; **194**:3661-3669.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)

Supplementary information, Data S1

Supplementary text

Design of the RNA probes to target β-globin mRNA expressed from a plasmid

To detect unmodified RNA within live bacterial cell, the cell should be programmed to express two RNA probes; each probe should contain antisense sequence and a fragment of a split aptamer. To detect β-globin mRNA, the probes were designed to contain 11-nt sequences complementary to the target, a 5-nt linker and a sequence corresponding to one fragment of the split eIF4A aptamer. The eIF4A-aptamer was split at the single-stranded loop at the position C35/G36 (**Figure 1**) that is the least likely site to distort the structure of the re-assembled aptamer [5].

We used two designs to express RNA probes (Supplementary **Figure S1**). At the first, each RNA probe was cloned into one of the two multiple cloning sites (MCS) of the pETDuet1 plasmid. This vector has two T7 promoters upstream of each of the MCSs and one T7 terminator downstream of the second MCS enabling the expression of two proteins. Since the vector lacked a transcriptional terminator between the two MCSs, transcription of two RNA probes from this vector would result in a single read-through RNA molecule with two RNA probes separated by the unrelated sequence. Additionally, a shorter transcript from the second T7 promoter would contain the probe cloned in the second MCS (Supplementary **Figure S1A**). In the second design, we cloned a second T7 terminator between the two T7 promoters to express RNA probes within two separate transcripts **(**Supplementary **Figure S1B).** Northern blot analyses showed that the probes expressed from the unmodified pETDuet1 vector were transcribed

preferentially (60%) as one long read-through transcript about 550 nt long. With an additional T7 terminator cloned between the two MCSs, the relative abundance of this long transcript decreased to 25% of the total transcripts and the relative abundance of the shorter transcripts (300 and 250 nt long) increased to 75% **(**Supplementary **Figure S1C).**

To evaluate the fluorescent background of the fusion proteins in the absence of target RNA, the cells expressing RNA probes were transformed with the plasmid pMB53, which encodes two fusion proteins each containing a fragment of a split EGFP and a fragment of the eIF4A protein [3, 6-8]. Unexpectedly, the cells that expressed predominantly RNA probes within one read-through transcript displayed lower fluorescent background than the cells expressing fusion proteins in absence of RNA probes (Supplementary **Figure S1D**). At the same time the cells expressing RNA probes predominantly as separate transcripts displayed two-fold higher fluorescence closer to the background of the fluorescent proteins (Supplementary **Figure S1D**). We also tested control cells that expressed just one RNA probe with half of the aptamer and cells transformed with the empty vector (no aptamer sequences). All of these control cells showed background fluorescence close to the background of proteins alone (Supplementary **Figure S1D**). Based on these results, in all further experiments we used RNA probes that were expressed from the unmodified pETDuet 1 vector with one T7 terminator. It should be emphasized that the background fluorescence in the cells expressing split EGFP (with any variants of the RNA probes) is several fold lower than the cells expressing full-size EGFP (see Supplementary **Figure S1D**). Thus, the

major finding is that the RNA probes expressed within one transcript reduce fluorescent background even further.

Mechanism of the fluorescent RNP complex formation and computational modeling of RNP

The low fluorescent background in cells expressing fusion proteins and mRNA-specific probes as compared with cells expressing fusion proteins alone suggests some specific interactions between probes and fusion proteins. Since RNA probes contain partial sequences of the eIF4A-specific aptamer, they were the most likely candidates for binding the fragments of the eIF4A protein. To test this assumption, several mutant probes were created, each corresponding to a different aptamer mutant known to have lower affinity for the eIF4A protein (Supplementary **Figure S2A**) [5]. These mutants were expressed as split halves within the long read-through transcript and tested for their ability to affect background fluorescence. The results demonstrated that all mutants having lower affinity to the eIF4A protein were less effective at reducing fluorescent background (Supplementary **Figure S2B**), thus supporting the idea that the partial aptamer sequences and fusion proteins do interact. More support for this mechanism has been obtained in the experiments with the fusion proteins that contained the fragments of the split EGFP fused with viral peptides λN and HTLV Rex but lacked fragments of the eIF4A protein [8]. The cells expressing these proteins displayed high background fluorescence in the presence of the RNA probes with eIF4A aptamer sequences similar to the background fluorescence of the proteins alone (Supplementary **Figure S2C**).

To get additional information about the assumed mechanism of the RNP complex formation we preformed *in silico* folding of RNA probes in the presence and absence of the target mRNAs (Supplementary **Figure S2D** and **S2E**). To incorporate target sequences, we fused them with the sequences of the probes through a flexible U_{10} linker. In the absence of target RNA the split aptamer sequences are separated by the secondary structure of the intervening sequences (Supplementary **Figure S2D**). However, the presence of complementary mRNA sequences triggers formation of a stable secondary structure that is practically identical for different mRNAs. We analyzed secondary structures of a handful of RNA probes designed for targeting various bacterial mRNA targets and found similar folds, all of them contained folds similar to the eIF4A aptamer with the exception of a single-stranded loop where the initial split had been introduced (Supplementary **Figure S2E**).

To account for these results, we propose the following mechanism of the RNP complex formation (Supplementary **Figure S2F**). First, we suggest that in the absence of target RNA the fusion proteins, each containing a domain of the eIF4A protein, interact with the fragments of the split eIF4A aptamer. These interactions might be weak, but they might suffice to keep the fusion proteins apart and prevent their re-association. This mechanism explains why the use of the probes expressed within a single transcript is favorable as compared with the probes expressed as separate transcripts. It is likely that the secondary structure of a long transcript containing both RNA probes provides spatial separation of the split aptamer sequences and thus separation of the fusion proteins preventing their re-association. On the contrary, if the two RNA probes with the partial aptamer sequences are expressed as two separate transcripts, their interactions

with the fusion proteins would not prevent their re-association because of the independent movements of the two RNP complexes.

In the presence of target RNA, the formation of the complex is driven by the complementary interactions of the probe's antisense arms with the target mRNA. The formation of the dsRNA places two split aptamer sequences in close proximity and allows proper aptamer folding. Once the aptamer is folded, the fusion proteins come in close proximity which triggers protein complementation and fluorescence development (Supplementary **Figure S2F)**.

To further understand the molecular interactions between the aptamer and the eIF4A protein, we performed computational modeling to predict the 3D structure of the eIF4A aptamer and to reconstruct the RNP complex of eIF4A-aptamer. We first performed *ab initio* folding of the aptamer (58 nt) using an algorithm developed earlier [9], where large-scale benchmark study demonstrated that our approach can accurately predict the 3D structure of short RNAs (<60 nt). In agreement with previous biochemical studies [5] the predicted aptamer structure formed a pseudoknot (Supplementary **Figure S3A)**. Since eIF4A is a RNA helicase, it usually recognizes the sequences between single- and double-strands [10]. Based on the structural models of the aptamer, we postulated that the single-stranded internal loop near the 3' of the aptamer was a good candidate for binding eIF4A (within the dashed-line box of Supplementary **Figure S3A**). Based on a crystallography structure of eIF4A in complex with a single-stranded RNA [11], we reconstructed a model for the RNP complex between eIF4A and aptamer (Supplementary **Figure S3B**). Our RNP model suggests that in addition to the single-stranded sequence there are extensive contacts between the aptamer and eIF4A,

contributing to the strong binding between eIF4A and the aptamer (~10 nM). The 5' stem-loop of the aptamer makes extensive interactions with only one of the sub-domains of eIF4A, while the 3' segment of the aptamer mostly interacts with the other sub-domain (Supplementary **Figure S3C**). Therefore, the split aptamer sequences can bind eIF4A sub-domains separately, and hinder the association of two eIF4A sub-domains when the sequences are separated in the transcript (Supplementary **Figure S2D**). Only when the two split aptamer sequences are brought together by sequence complementation and form their native pseudo-knotted three-dimensional structures, the eIF4A sub-domains re-associate together to bind the high-affinity aptamer (Supplementary **Figure S2E**) and shift the equilibrium toward the fluorescence-competent state.

Detection of the β-globin mRNA in live bacterial cells

E. coli BL21(DE3) expressing fusion proteins and RNA probes were transformed with a third plasmid encoding a 196 nt long fragment of the β-globin mRNA with the 22 nt-long target site. We analyzed fluorescence of the cells after induction of all components of the RNP complex with 1 mM IPTG and overnight growth at 23^0C using flow cytometry (Supplementary **Figure S4B** and **S4C)** and fluorescent microscopy **(**Supplementary **Figure S4D).** The low growth temperature was due to previous studies that showed better folding of the fusion proteins at lower temperatures [8]. The FACS and microscopy data showed that the cells expressing all components of the RNP complex (two fusion proteins, RNA probes and the fragment of the β-globin mRNA) displayed fluorescence that was about 4-5 fold higher than the cells that lacked the RNA

target. Microscope images showed that 17-20% of cells displayed fluorescence above the background **(**Supplementary **Figure S4D).** Using target RNA with a corrupted binding site we observed fluorescence close to the background level (Supplementary **Figure S4D**) that suggested that the complex was assembled sequence-specifically.

We obtained similar results using a full-size β-globin mRNA, although signal-to-background was 2 fold lower than with the mRNA fragment (Supplementary **Figure S4**). Collectively, these results suggest that: (i) the split eIF4A aptamer was reassembled into a proper fold; (ii) the reassembled aptamer was able to initiate complementation of the split fusion proteins, and that (iii) the fluorescent RNP complex was assembled sequence-specifically in the presence of the target mRNA. Thus, these experiments validate the split aptamer approach for detection of unmodified RNAs in live bacterial cells with a 2-5 fold signal-to-background (s/b) ratio.

Quantitation of RNA and protein molecules

To further characterize fluorescent RNP complexes, we quantified the concentrations of the RNA components by RT-PCR. Concentrations of the RNA probes were estimated to be about 5 fold lower than the level of 16S RNA, while the level of the β-globin mRNA was about 675-fold lower compared to 16S RNA (Supplementary **Figure S5A**). Given that the *E. coli* cell contains roughly 50,000 ribosomes [12] these results imply that the concentrations of RNA probes and the β-globin mRNA were about 10000 and 75 mol/cell, respectively. Semi-quantitative protein gel electrophoresis was used to evaluate the concentration of the fusion proteins. It was estimated to be about 5000 molecules of A-F1 and B-F2 per cell (Supplementary **Figure S5C**).

We also used cell's fluorescence to roughly estimate protein concentration. Full-size EGFP was expressed in the same vector pACYCDuet1 as split fusion proteins and induced by 1 mM IPTG for 3 hr. Using standard conversion of OD_{600} for *E. coli* cells in LB, we estimated cell concentration as approximately 0.76 x 10^{10} cells/ml. These cells were spun down and concentrated 100x to get a final concentration of 7.6 x 10^{11} cells/mL. The stock solution of EGFP was diluted 10x and absorbance at 490 nm was measured to be 0.002 over the background. Using the Beer-Lambert law and ε for EGFP as 55000 M^{-1} cm^{-1} we estimated EGFP concentration as 0.363 μM.

We then measured fluorescence by filling a well on a slide with EGFP solution and a well with the cell suspension. The wells were viewed under 10x magnification to give a uniform intensity. Comparative analysis was performed in ImageJ, averaging the intensities over the field of view for cells and EGFP solution, and subtracting the baseline intensity of PBS solution. The results were an average of 5000 a.u for the cells and 300 a.u. for the EGFP solution.

The average number of EGFP molecules per cell was calculated using the equation of Taniguchi *et al*. [13].

Avg. molecules per cell: $$n = \frac{\frac{Fcell}{Fprotein}[Protein]\, N_A}{Cell\ concentration}$$

$$n = \frac{\frac{5000}{300}(0.363\mu M)N_A}{7.16\times10^{14}} = 5087 \text{ EGFP molecules/cell}$$

Thus, both protein gel analysis and fluorescent measurements resulted in very close concentrations of the proteins per cell.

Efficiency and sensitivity of RNA detection

Since all components of the fluorescent RNP complex are internal and are synthesized by the cell it is important to understand their relative concentrations and their effect on efficiency and sensitivity of RNA detection. From one hand, due to the interactions between the RNA probes and fusion proteins, the concentrations of the RNA probes should exceed concentrations of the proteins. Indeed, the excess of RNA probes will result in efficient binding of split proteins and will reduce the background. From the other hand, the large excess of the RNA probes over the proteins will reduce the proportion of the probes with the two binding sites occupied by the proteins and this will decrease the efficiency of RNA detection.

Our data show that the concentration of the RNA probes was about 2-fold over concentration of the fusion proteins: 10000 RNA probe molecules per cell versus 5000 molecules of the fusion proteins per cell **(**Supplementary **Figure S5A** and **S5C).** As our experimental data show, this ratio provides substantial reduction (75%) of the background fluorescence (Supplementary **Figure S1D**). At this ratio due to statistical probability only 25% of the probes have both binding sites occupied by the proteins. As a result, only these probes are efficient in detecting the target. Additionally, due to the non-synchronized cell population different cells have fusion proteins on different stages of expression and protein maturation making some cells non capable of developing fluorescence (see [3]). All these factors explain that only 17-20% of cells displayed signal with β-globin mRNA (Supplementary **Figure S4D**). Thus, even though the average RNA concentration was about 75 mol of target RNA, only 20% of cell displayed signal.

At the same time, our results on endogenous *PstC* mRNA detection imply that the sensitivity of detection is very high. Indeed, we detected signal from *PstC* mRNA in about 8% of cells with the average RNA concentration 1 molecule per cell. Taking into account that the efficiency of our method is about 20% and using the Poisson distribution showing that only 30% of cells have at least 1 mol of *PstC* mRNA (Supplementary **Table S8**), we conclude that the method has sensitivity close to a single molecule.

Our data allow estimating sensitivity of RNA FISH. Hybridization with 38 Stellaris probes showed *PstC* mRNA signal in 25% of cells grown under phosphate shortage (5 molecule per cell) and no signal in the cells grown under normal phosphate concentrations (1 molecule per cell) **(**Supplementary **Figure S7)**. In FISH approach the probes are external and in great excess over the target, therefore each cell and each RNA molecule are equally capable of producing signal. Since we detected signal in 25% of cells with average 5 molecules per cell we can conclude that FISH sensitivity is 7 or more RNA molecules per cell based on Poisson distribution (Supplementary **Table S8**). Poisson distribution also explains why in the cells grown under normal phosphate concentration with average *PstC* mRNA concentration 1 mol/cell we did not observe any signal. Indeed, in this case there are no cells that have 7 or more RNA copies/cell (Supplementary **Table S8).** Relatively low FISH sensitivity is not surprising. It is known that the probes with constitutive fluorescent signal display reduced sensitivity because of the high background stemming from the non-specific probe binding. This relatively low sensitivity is the major motivation behind all efforts to develop biosensors that

display signal only in response to the presence of a user-defined target, such as molecular beacons or other target templated biosensors.

Materials and Methods

Cell Lines

All cloning reactions have been performed using *E. coli* XL-10 cells, while all expression used *E*. *coli* BL21(DE3) cells.

Plasmid Construction

Two RNA probes were designed to contain one half of the eIF4A-binding aptamer and 11 nucleotides complementary to a 22 nt site within the β-globin mRNA. These two parts were linked through the linker CCTCC. The corresponding oligonucleotides were purchased from Integrated DNA Technologies (IDT) (Supplementary **Table S1**) and cloned into the two multiple cloning sites of the vector pETDuet1 (Novagen). Probe #1 was cloned between the *BamH*I and *Not*I restriction sites in the first MCS. Probe #2 was cloned between the *Bgl*II and *Xho*I restriction enzyme sites in the second MCS (plasmid pPTT05) using a cloning method descried by Vasl and co-workers [1]. The negative controls contained Probe #1 or Probe #2 cloned separately in pETDuet1 (Supplementary **Table S2**, plasmids pPTT01 and pPTT20). The probe that targeted the antisense strand of the β-globin mRNA was cloned in the plasmid pPTT66. All plasmids were isolated using a QIAgen® Miniprep Kit and confirmed by sequencing.

A 198 bp oligonucleotide corresponding to the β-globin gene was purchased from IDT (Supplementary **Table S1**) and cloned into the pRSFDuet vector (Novagen®) between the *BamH*1 and *Not*1 restriction sites in the first multiple cloning site (plasmid pPTT07). A negative control contained the same fragment of the β-globin mRNA with a scrambled 22-nt target site (plasmid pPTT10).

A full-length rabbit β-globin gene was reverse transcribed from the β-globin mRNA (Sigma-Aldrich®) using an Invitrogen™ SuperScript® III First-Strand Synthesis Kit. The cDNA template was cloned using Overlap Extenstion PCR with primers 15 and 16 (Supplementary **Table S1**) into the pRSFDuet (Novagen®) vector (plasmid pPTT11). The negative control, a full-length β-globin mRNA with the scrambled target site, was created from the full-length β-globin mRNA plasmid (Table 2, plasmid pPTT11) using PCR mutagenesis with the primers 3 and 4 resulting in a plasmid pPTT12 (Supplementary **Table S2**).

All cell lines are listed in Supplementary **Table S3**.

Mutagenesis of the aptamer sequences

To introduce the point mutation G21A (M12) into the split aptamer sequences within plasmid pPTT05 we used PCR mutagenesis with the primers 5 and 6 (plasmid pTT32, Supplementary **Table S2**). The mutation U20A (M7) was obtained accidentally during mutagenesis and it is contained within the plasmid pPTT31. Lastly, a mutation/insertion mutant A50G/ins55A was obtained as a byproduct during cloning of the RNA probes targeting 16S RNA (this target has not been used in this study). This

mutant is expressed in plasmid pPTT38. All mutations were confirmed by sequencing and are listed in Supplementary **Table S4.**

Preparation of probes targeting bacterial mRNAs

All plasmids expressing RNA probes targeted at different mRNAs targets were obtained by one step mutagenesis of the plasmid pPTT05 using protocol described by Unger *et al.* [2]. The target sites are listed in Supplementary **Table S5** and the primers for mutagenesis in Supplementary **Table S6.** All sequences were confirmed by sequencing. The plasmids expressing mRNA-specific probes are listed in Supplementary **Table S2**.

Transformation and cell culturing

To label β-globin mRNA *E. coli* cells BL21(DE3) were transformed with three compatible plasmids, pPTT05 (express RNA probes, pETDuet1 background), pPTT07 (express RNA target β-globin mRNA, pRSFDuet background), and pMB53 (express eIF4A-EGFP fusion proteins, pACYCDuet 1 background), a plasmid described earlier [3]. All three plasmids have T7 promoters and were induced simultaneously by 1 mM IPTG. To detect endogenous bacterial mRNAs two plasmids were necessary, a plasmid expressing two RNA probes and the plasmid pMB53. LB media was supplied with the necessary antibiotics and the cells starting from one colony were grown at 37 ^{0}C for 3-4 hrs. The concentrations of antibiotics were: chloramphenicol, 34 μg/ml, kanamycin, 50 mg/ml, and ampicillin, 100 μg/ml. The culture was then diluted (1:10 to 1:100) into 3 ml, induced with 1 mM IPTG and grown overnight in LB at 20-25 ^{0}C. Cells were collected at

equal optical densities between OD_{600} of 0.4 – 0.6. 500 μL of cells were washed twice in 500 μl 1X PBS and used for FACS analyses and fluorescent microscopy.

FACS Analysis and Microscopy

To measure fluorescence of a cell population, we used a Becton-Dickinson FACSCalibur flow cytometer with a 488-nm argon excitation laser and a 515–545 nm emission filter (FL1). We measured fluorescence of 100,000 cells in each sample. For fluorescent microscopy we immobilized bacterial cells between a cover slip and a thin slab of 0.8% agarose in 1X PBS. The samples were imaged at room temperature on a Nikon inverted microscope Eclipse Ti-E with Nikon CF160 optical system and NIS Elements software. All experiments were carried out on a 100X oil immersion objective with 0.2-1 sec exposure depending on experiment. Differential contrast (DIC) and fluorescent (FITC) images were taken simultaneously.

Quenching experiments

E. coli cells were immobilized between a cover slip and a thin slab of 0.8% agarose in 1X PBS. Samples were excited by a 488-nm laser beam (15 mW) and imaged by an EMCCD camera (iXon DV 897, Andor Technology) at 33 frames/s. The image acquisition program was written using LabVIEW and the movies were analyzed using programs written in MATLAB.

qRT-PCR

Real-time quantitative PCR analysis was preformed on RNA samples isolated from 1 ml of cultures that were harvested at equal OD_{600} 0.4-0.6. RNA was isolated using Qiagen RNeasy® Mini Kit and was turned into cDNA using an Invitrogen SuperScript® III First-Strand Synthesis System. Each cDNA reaction used 2.0 μg of RNA from each cell line. qRT-PCR was performed using about 200 ng cDNA and the Roche RT-PCR Kit with SYBR Green, according to the manufacturer's guidelines. A reaction lacking reverse transcriptase was included for each sample, which served as a control for DNA contamination. 16S RNA was used as a reference gene. Amplification was performed using Roche LightCycler 480 under the following conditions: 95 ^{0}C – 5 min, (95 ^{0}C – 10 s, 58 ^{0}C – 10 s, 72 ^{0}C-10 sec) for 45 cycles. The PCR product identities were confirmed by electrophoresis on 1% agarose gels and product uniformity was tested using melting curves. RT-PCR for each RNA target was performed on at least two independently isolated RNA samples in triplicates, and the mean C_t values were calculated.

Analysis of protein expression

The cells expressing fusion proteins were induced with 1 mM IPTG and grown to density OD_{600} 1.3-1.4. Cells from1 ml of cell culture were precipitated, washed and proteins were fractionated using PeriPreps^{TM} Periplasmic kit and separated using SDS PAGE with BenchMark protein ladder (Invitrogen). The gel was stained by SimplyBlue and the amount of the overexpressed proteins was evaluated by comparison the intensities of the corresponding bands with the amount of the marker protein **(**Supplementary **Figure S4).** M- BenchMarkTM protein ladder (Invitrogen), 1, 2, 3, -

extracellular proteins, 4, 5, 6 – periplasmic proteins; 2 and 5 - uninduced, 1,3, 4 and 6 induced with 1 mM IPTG.

Northern blotting

Northern blotting was performed using total bacterial RNA isolated with a Qiagen RNeasy® Mini Kit. RNA separation using formaldehyde-agarose gel, transfer to the Hybond-H membrane and RNA labeling was performed as recommended by the DIG High Prime DNA labeling and Detection Starter Kit II (Roche).

***In silico* analyses of secondary structures of RNA targets and RNA probes**

Secondary structure predictions were performed using the mFold RNA package algorithms (http://mfold.rna.albany.edu).

RNA FISH with multiple singly labeled (TAMRA) Stellaris probes (see Supplementary Table S8 for the probe sequences).

To do RNA FISH we used the protocol described in Montero-Llopis *et al.* [4]. Briefly, the cells were fixed in 4% formaldehyde, 30 mM Na_2HPO_3, pH 7.5, 15 min, at room temperature for 30 min on ice. Then the cells were spun down at 6,000 rpm for 3 min and washed in 1x DEPC-treated PBS. This step was repeated 3 times. The pellet was re-suspended in 100 μl of GTE buffer (50 mM glucose, 20 mM Tris-HCl pH 7.5, 10mM EDTA pH 8) and 4 μl of lysozyme solution was added to 12 μl of cell suspension (10 μg/ml in GTE buffer plus 4mM vanadyl ribonucleoside complex (VRC)). The cell

suspension was placed on the poly-L-lysine coated slides (PolySciences) and incubated 10 min at room temperature. Excess liquid was aspirated and slides were dried (1 min). The slides were put in -20 ^{0} C methanol for 10 min, dipped into -20 ^{0} C acetone for 30 sec and dried on air.

The slides were Incubated at 37 ^{0}C for 30-60 min in 40% formamide and 2 x SSC treated with DEPC. Hybridization probe (final concentration 250 nM) was added to hybridization solution I (80% formamide, 1mg/ml E. coli tRNA, 2x SSC DEPC-treated, 70 μg/ml calf thymus DNA) and incubated at 80 ^{0}C for 5 min before mixing with the hybridization solution II (20% dextran sulphate, 4 mM VRC, 40 U RNase inhibitor, 0.2% RNase free BSA, 2x DEPC-treated SSC) in 1:1 ratio. 25-50 μl of hybridization mixture was added to each well and hybridized overnight. Slides were washed twice in 50% formamide, 2x SSC treated with DEPC for 30 min, and briefly rinsed 5 times in DEPC-treated PBS. 4 μl of 1.5 μg/ml DAPI solution was added. The slides were covered, sealed with the nail polish, and visualized immediately.

Supplementary References

1. Vasl J, Panter G, Bencina M, Jerala R. Preparation of chimeric genes without subcloning. *BioTechniques* 2004; **37**:726, 728, 730.

2. Unger T, Jacobovitch Y, Dantes A, Bernheim R, Peleg Y. Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *J Struct Biol* 2010; **172**:34-44.

3. Valencia-Burton M, McCullough RM, Cantor CR, Broude NE. RNA visualization in live bacterial cells using fluorescent protein complementation. *Nature methods* 2007; **4**:421-427.

4. Montero Llopis P, Jackson AF, Sliusarenko O *et al.* Spatial organization of the flow of genetic information in bacteria. *Nature* 2010; **466**:77-81.

5. Oguro A. RNA aptamers to initiation factor 4A helicase hinder cap-dependent translation by blocking ATP hydrolysis. *RNA* 2003; **9**:394-407.

6. Valencia-Burton M, Shah A, Sutin J *et al.* Spatiotemporal patterns and transcription kinetics of induced RNA in single bacterial cells. *Proceedings of the National Academy of Sciences of the United States of America* 2009; **106**:16399-16404.

7. Borogovac A, Broude NE. Visualization of induced RNA in single bacterial cells. *Methods Mol Biol* 2011; **714**:189-199.

8. Yiu H-W, Demidov VV, Toran P, Cantor CR, Broude NE. RNA Detection in Live Bacterial Cells Using Fluorescent Protein Complementation Triggered by Interaction of Two RNA Aptamers with Two RNA-Binding Peptides. *Pharmaceuticals* 2011; **4**:494-508.

9. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. *Ab initio* RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA* 2008; **14**:1164-1173.

10. Abramson RD, Dever TE, Lawson TG, Ray BK, Thach RE, Merrick WC. The ATP-dependent interaction of eukaryotic initiation factors with mRNA. *The Journal of biological chemistry* 1987; **262**:3826-3832.

11. Montpetit B, Thomsen ND, Helmke KJ, Seeliger MA, Berger JM, Weis K. A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature* 2011; **472**:238-242.

12. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol Microbiol* 2012; **85**:21-38.

13. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, and Xie XS. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 2010; **329**: 533-538.

Supplementary information, Figure S1

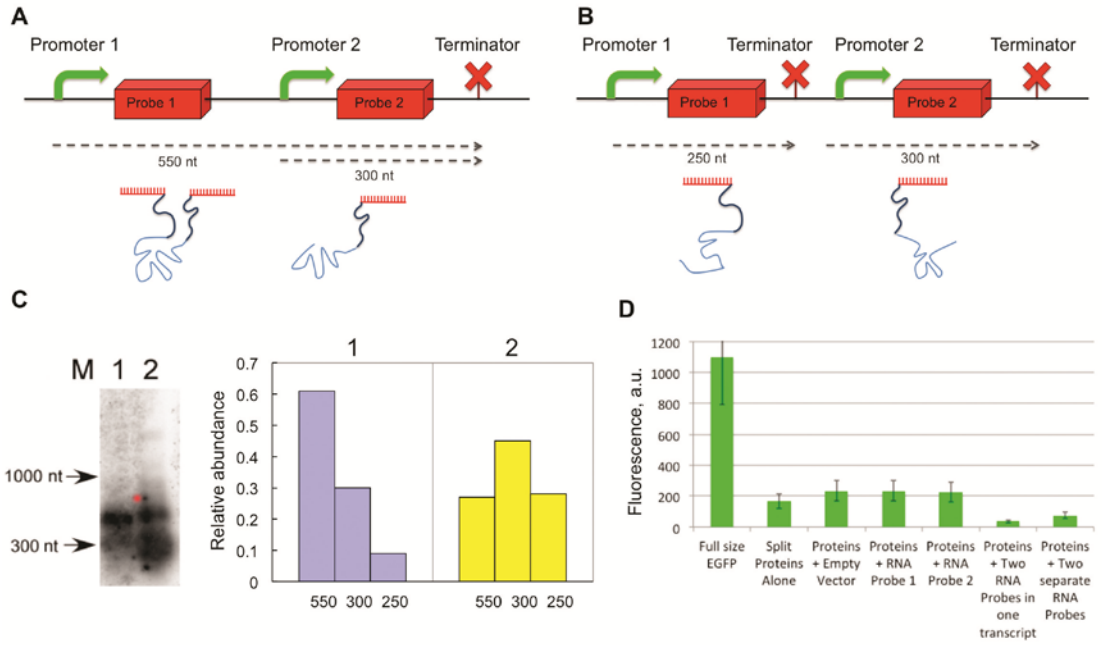

Figure S1 Two different designs of the RNA probes. **(A)** RNA probes are expressed from the unmodified pETDuet 1 vector as two transcripts, 550 and 300 nt long; **(B)** RNA probes are expressed from a modified pETDuet vector with an additional T7 terminator as two separate transcripts, 250 and 300 nt long; **(C)** Northern blot analysis and quantitation of the transcripts expressed from the plasmids shown in **A and B; 1** corresponds to the design shown in **A, 2** corresponds to the design in **B**; **M**, markers**;** numbers under the histogram indicate the length of the transcripts in nucleotides (nt). **(D)** Background fluorescence of the *E. coli* cells expressing full-size EGFP or two fusion proteins, A-F1 and B-F2, and different RNA probes. Note that RNA probes expressed within one long transcript (design A) display the lowest background fluorescence. Data are given as average of 4 independent experiments ±standard deviation.

Supplementary information, Figure S2

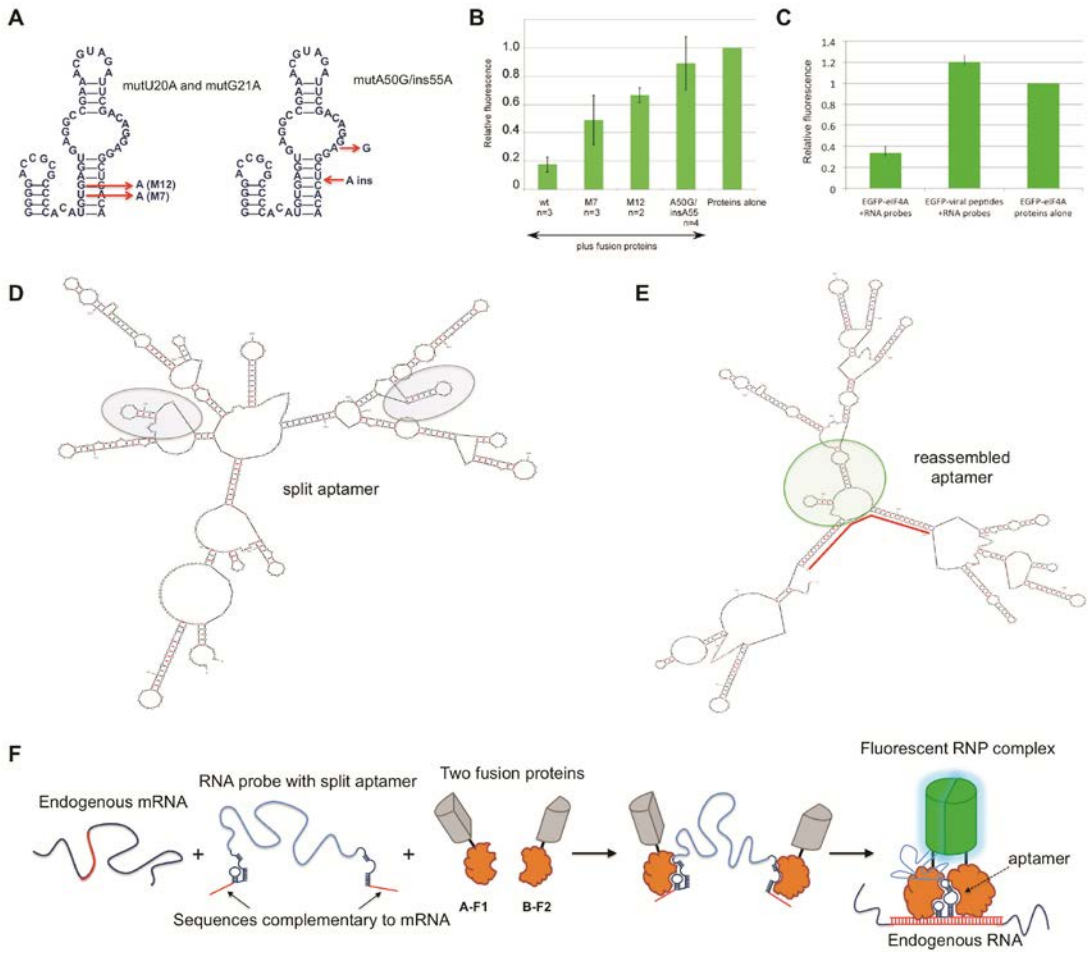

Figure S2 Suggested mechanism of the fluorescent RNP complex formation. **(A)** Structures of the eIF4A-aptamer mutants that bind fusion proteins with decreased efficiency. (**B**) Fluorescence of the *E. coli* cells expressing these eIF4A-aptamer mutants within the RNA probes and fusion proteins, A-F1 and B-F2. Note that all RNA probes with mutant aptamer sequences are less capable of reducing fluorescent background of fusion proteins. **(C)** Fusion proteins containing split EGFP and viral peptides that do not bind eIF4A aptamer display high fluorescence in the presence of the RNA probes with fragments of eIF4A-aptamer. **(D)** MFOLD structure of the transcript containing two RNA probes with the split aptamer

fragments in absence of the β-globin mRNA (the length of the transcript is from the T7 promoter to the T7 terminator); **(E)** MFOLD structure of the transcript containing two RNA probes in the presence of the target mRNA, target mRNA sequences are in red. Note the structure in oval that resembles eIF4A-aptamer fold; (**F**) Suggested mechanism of fluorescent complex formation with an intermediate step when the fusion proteins bind RNA probes. Apparently, this step is responsible for low fluorescent background.

Supplementary information, Figure S3

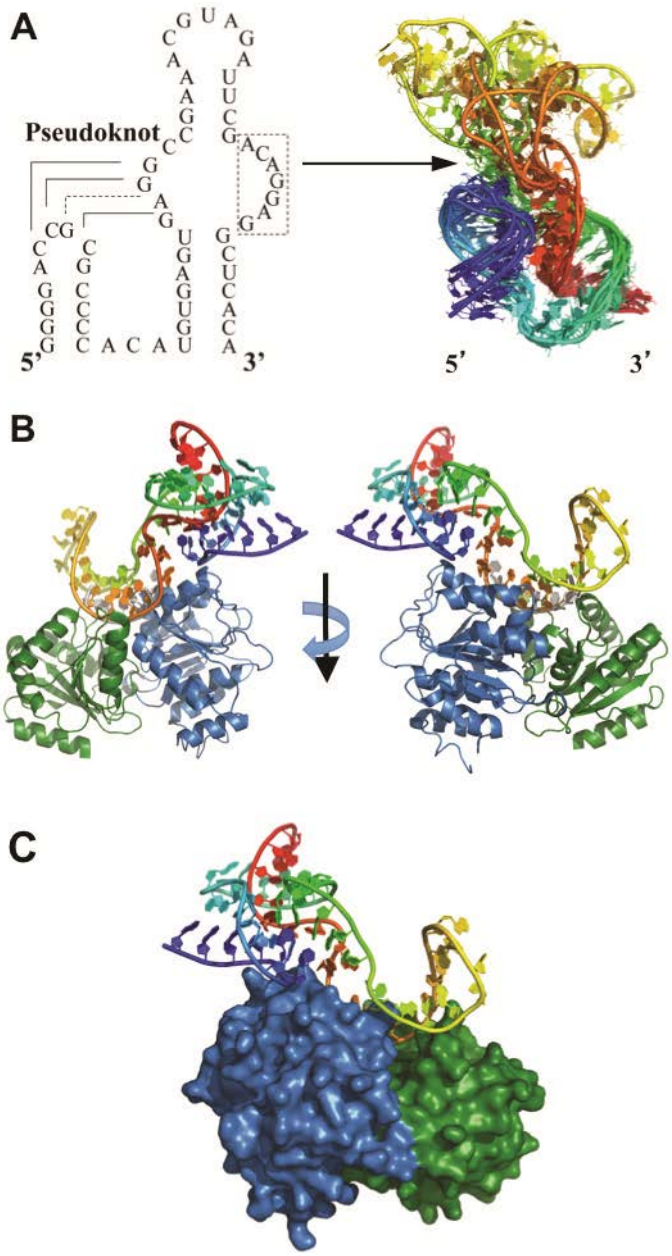

Figure S3 Computational modeling of the aptamer/eIF4A RNP complex. **(A)** The secondary and tertiary structure of the aptamer. The dashed-line box indicates the putative eIF4A binding sequence. An ensemble of computationally predicted three-dimensional structures with cartoon representation is shown (right panel) in rainbow color from blue (5') to red (3'). **(B)** The eIF4A-aptamer RNP structure were reconstructed based on

the complex structure of eIF4A-ADP-RNA segment (PDB: 3PEY) by aligning the eIF4A binding sequence with RNA segment in the known structure. The two sub-domains of eIF4A in cartoon are colored differently. **(C)** The surface of the protein is shown to illustrate the binding between aptamer and eIF4A sub-domains.

Supplementary information, Figure S4

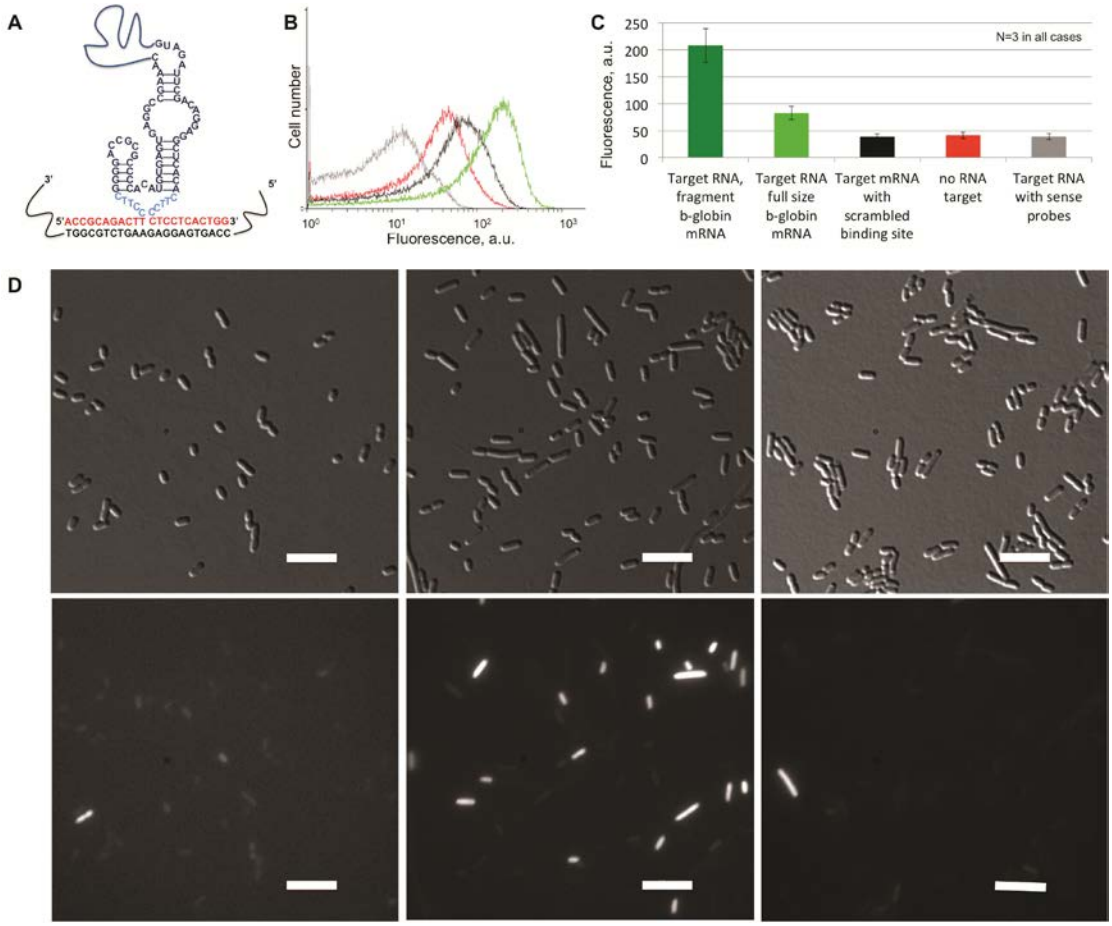

Figure S4 Detection of β-globin mRNA in live *E. coli* cells. **(A)** Structure of the β-globin mRNA target site and of the eIF4A-specific aptamer split at the central loop; antisense sequences are shown in red, linker sequences are shown in blue, the intervening sequence and the split site in the aptamer are marked by a curving line; (**B**) An example of FACS analysis of *E. coli* cells expressing different RNP complexes. Red curve, cells express two protein fusions and two RNA probes, no target RNA is present. Black curve**,** the cells express two protein fusions and two RNA probes plus a fragment of β-globin mRNA with corrupted target sequence. Green curve**,** the cells express protein fusions, two RNA probes and the β-globin mRNA fragment with the correct binding site. Gray curve, *E. coli* BL21(DE3) cells;

(C) Histograms obtained from data similar to as in **(B)**, average of 3 independent experiments ±standard deviation is shown; **1,** average fluorescence of the cells expressing fragment of β-globin mRNA; 2, cells express full-size β-globin mRNA; 3, β-globin mRNA with the corrupted binding site; 4, on target RNA is present; 5, the probes are targeted at the antisense strand. **(D) DIC (**top panels**) and fluorescent images of the cells (**bottom panels**)**. **(Left),** cells express two fusion proteins and two RNA probes in RNA absence; **(middle)**, cells express β-globin mRNA, protein fusions and two RNA probes **(right),** cells express β-globin mRNA with scrambled target sequence plus fusion proteins and RNA probes**.** All images were taken in similar settings, scale bar 10 μm.

Supplementary information, Figure S5

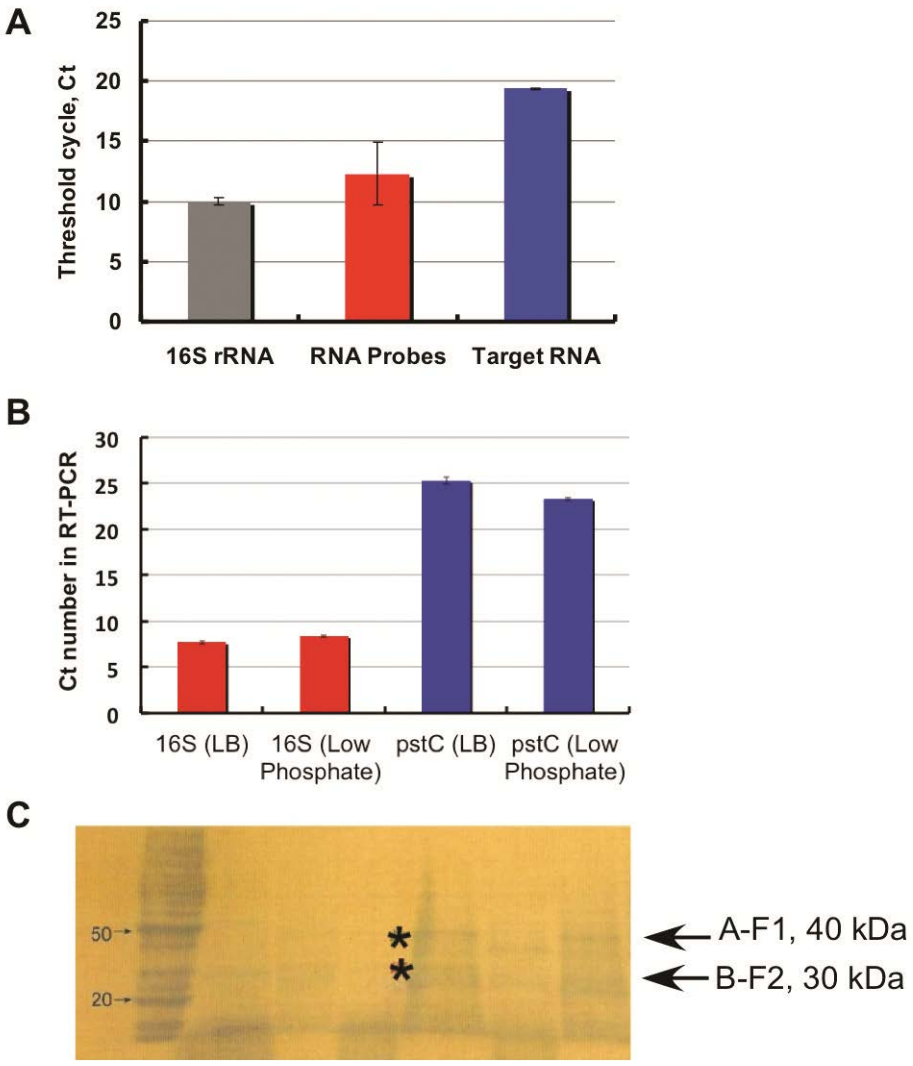

Figure S5 Quantification of RNA and protein components of the RNP complexes. **(A) Ct numbers obtained in RT-PCR for β-globin mRNA, RNA probes, and 16S RNA.** Average of 3 experiments is shown**. (B) RT- PCR analyses of *PstC* mRNA concentrations in Bl21(DE3) *E. coli* cells grown in normal and low phosphate.** 16S RNA was used as a control**.** Average C_t of 3 experiments is shown. **(C) SDS-PAGE of over-expressed fusion proteins A-F1 and B-F2 in BL21(DE3) *E. coli* cells. M-** BenchMark™ protein ladder (Invitrogen), 1, 2, 3, - extracellular proteins, 4, 5, 6 – periplasmic proteins; 2 and 5- uninduced, 1,3, 4 and 6 induced with 1 mM IPTG.

Supplementary information, Figure S6

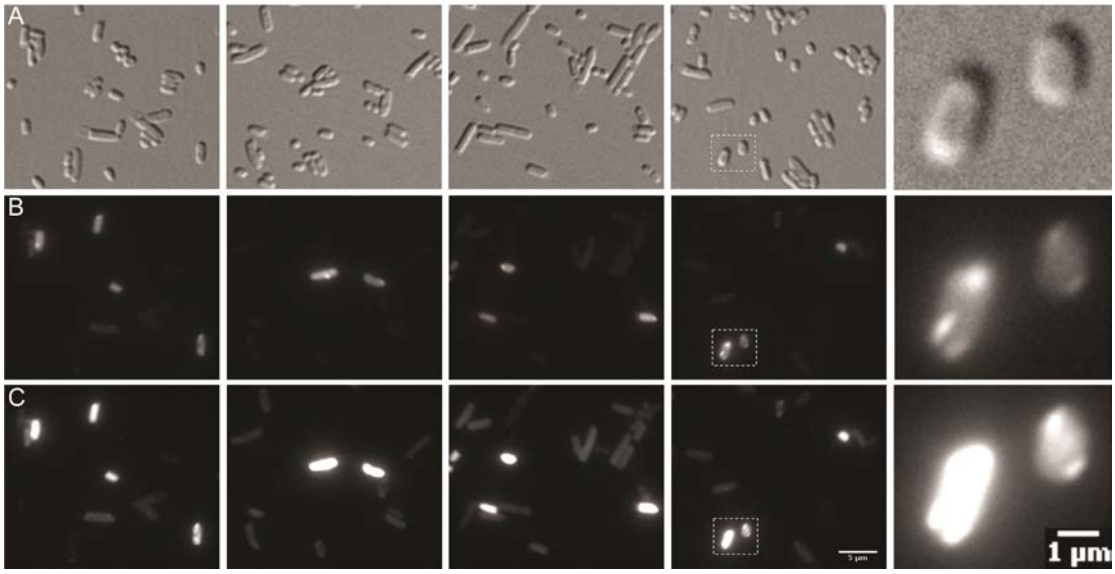

Figure S6 Oversaturated fluorescent signal masks localization of the *PstC* mRNA in ***E. coli*** cells. **(A)** DIC images, **(B)** fluorescent images with 200 msec exposure**; (C)** fluorescent images with 800 msec exposure**.** The most right panel shows the enlarged cells marked by the square.

Supplementary information, Figure S7

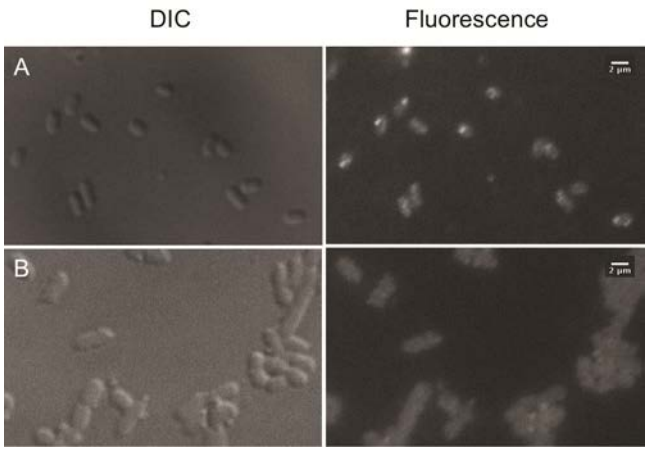

Figure S7 RNA FISH of the *PstC* mRNA with 38 Stellaris probes labeled with TAMRA. **(A)** BL21(DE3) *E. coli* cells grown in the low phosphate (0.2 mM) medium A (15) (0.12 M Tris; 80 mM NaCl; 20 mM KCl; 20 mM NH_4Cl; 3 mM Na_2SO_4; 1 mM $MgCl_2$; 0.2 mM $CaCl_2$; 2 mM ZnCl2; 0.5% bactopepton; 0.5% glucose; pH7.5. **(B)** BL21(DE3) *E. coli* cells grown in LB medium (high phosphate, 2.2 mM).

Supplementary information, Figure S8

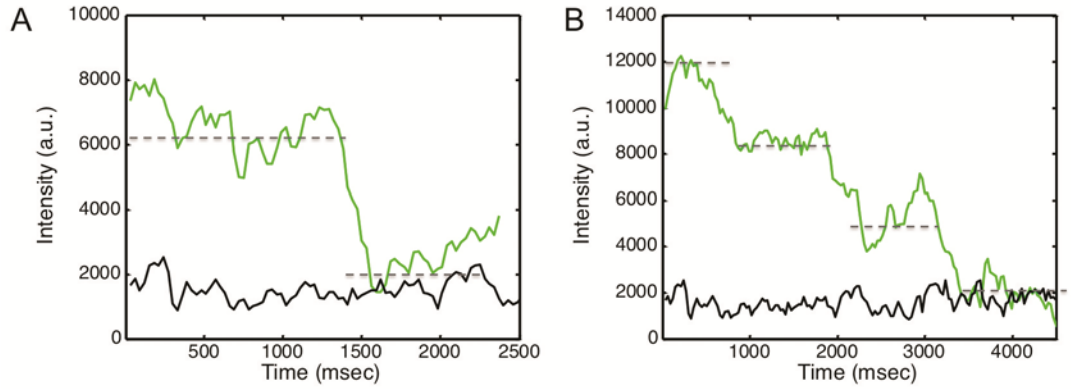

Figure S8 Spot fluorescence traces show stepwise bleaching. We show two characteristic fluorescence traces of the punctate signals in the cells with labeled *PstC* mRNA. The X-axis shows the time of exposure. The Y-axis shows the fluorescence in arbitrary units. **(A)** One-step photobleaching is characteristics for single fluorescent molecule observed in 5% of cells. **(B)** Multi-step photobleaching is characteristics for a few fluorescent molecules observed in 2% of cells; background line was taken from the cells with now signal (black curve). 10% of cell with the punctate signals did not show well-defined abrupt photobleaching.

Supplementary information, Figure S9

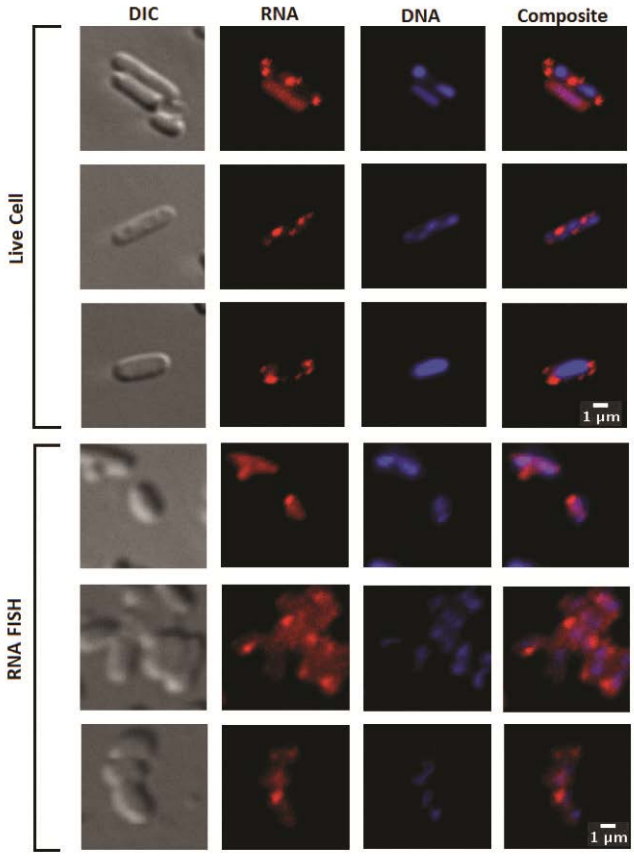

Figure S9 *PstC* mRNA localization in live cells and in FISH look similar and the signal does not overlap with nucleoid DNA. DNA was labeled with Hoechst33342 dye. Note the strong background fluorescence in RNA FISH masking DNA localization.