# Design and characterization of a fluorescent protein-based imaging system for RNA in live cells

**Research Thesis**

**In Partial Fulfillment of the Requirements for the Degree of Master of Science in Biotechnology and Food Engineering**

**Aya Friedman**

**Submitted to the Senate of the Technion - Israel Institute of Technology**

**Nisan, 5774 Haifa April 2014**

The research thesis was done under the supervision of Prof. Roee Amit at the Faculty of Biotechnology and Food Engineering.

The generous financial help of the Technion is gratefully acknowledged.

## Table of contents

**Table of Contents- continued**

## List of Figures



## List of Tables

## Abstract

Recently, the single-cell and single-molecule approaches have become increasingly popular. These approaches provide large amounts of data at the highest possible resolution in biology. However, they require expertise in advanced experimental techniques in order to obtain the desired data. One area that still requires development of such techniques is the study of RNA. The fusion of a fluorescent protein to an RNA binding protein is a means of visualizing specific RNA molecules inside living cells. In order for the fusion protein to bind, the target RNA must include a sequence which the binding protein can recognize and attach to. Gathering large amounts of data requires the use of multiple combinations of fusion proteins, which can recognize different binding sites and thus follow multiple RNA species simultaneously. For this purpose, we created a library of phage-derived RNA binding proteins fused to fluorescent proteins, and set out to characterize it.

The fluorescent function of the fusions was analyzed and compared to cells lacking them. This assay was examined on combinations including four different RNA binding proteins, but only three showed significant fluorescence levels. These three were the coat proteins of MS2, Qβ and PP7. The fluorescent genes used were mCherry, mDendra2 and cerulean.

A key characteristic of these proteins is to function as translational repressors of the phage's *replicase* gene, via the gene's translation initiation site. By placing a reporter gene after this site, scientists have studied the specificity of binding of these proteins. In this study, we placed these sequences upstream to genes encoding for fluorescent proteins, and observed the fluorescence levels in cells containing different combinations of fusion proteins and binding sites. In this manner, we were able to assess the binding affinities of each fusion protein to each RNA operator *in vivo*. We discovered that some of the coat proteins are not necessarily orthogonal in their binding site recognition as previously thought. Furthermore, we found out that one of our fusions, MS2-dendra, amplifies rather than represses expression of the reporter, regardless of the structure of the translation initiation site. This finding raises questions regarding the nature of the interaction between MS2-dendra and the reporter's RNA, and concerning the ability of synthetic biology to be a reliable platform for the study of natural processes.

## List of abbreviations and symbols

| | |
|---|---|
| aTc | anhydrous tetracycline |
| a.u. | arbitrary units |
| BS | Binding site |
| DTT | Dithiothreitol |
| eGFP | Enhanced green fluorescent protein |
| FISH | Fluorescence in situ hybridization |
| FP | Fluorescent protein |
| FRET | Förster resonance energy transfer |
| GFP | Green fluorescent protein |
| LB | Luria Betani |
| MCS | Multiple cloning site |
| NAD | β-Nicotinamide adenine dinucleotide |
| O.D. | Optical density |
| pCP | Phage coat protein (name of vector) |
| PCR | Polymerase Chain Reaction |
| RBP | RNA binding proteins |
| RBS | Ribosome binding site |
| ORF | Open reading frame |
| UTR | Untranslated region |

# 1. Introduction

## 1.1 RNA imaging systems

In recent years, more and more studies in biology have been dedicated to the development of techniques that allow observations of live cells. As data concerning the complexity of cellular processes accumulates, it is becoming crucial to be able to distinguish between events occurring at different times as well as space. The use of fluorescent proteins in tracking gene expression has been demonstrated almost twenty years ago (7) and has been utilized ever since to follow protein dynamics in living cells. It is only in the last decade, however, that scientists have been following RNA transcripts in a similar fashion.

Previous techniques to study RNA were based on cell fixation, for example, fluorescence in situ hybridization (FISH). This technique, which was first put to use in the early 1980s, uses sequence specific probes consisting of complementary oligonucleotides attached to fluorescent dyes. The original method could not distinguish between bound and unbound probes in living cells, and required immobilization of hybrid molecules (probes attached to target RNAs) on solid surfaces and removal of the unbound probes. It was only in 1996 that Tyagi and Kramer proposed a new principle for constructing probes that eliminated the need to remove unbound probes (38). This was done by attaching two complementary arm sequences to either end of the probe, one of which was attached to a fluorescent moiety and the other to a quenching one. Before hybridization, the probe is folded up so that the fluorophore is close to the quencher, causing its fluorescence to be quenched when illuminated. Once the probe hybridizes to its target, the result is a double stranded molecule of nucleic acids that is much stronger than the one consisting of the two arms of the probe, and therefore the transition between the states is spontaneous. While the two moieties are apart, the fluorophore can fluoresce when illuminated with the proper light (38). Further developments led to the creation of side-by-side probes: two different probes that target adjacent sequences and when close enough are designed to deliver a signal based on Förster resonance energy transfer (FRET), that is, when both are hybridized to their RNA targets, excitation of one can induce emission of fluorescence in the other (37). The main drawback of using probes in comparison to GFP is the need to disturb the cell membrane in order

to insert them inside (24). Other limitations of the probes include the need to modify their backbones in order to avoid rapid degradation and prevention of probe binding by secondary structures of RNA (37). Another, much newer technique involves aptamers that stimulate fluorescence by binding to existing chromophores and strengthening their signal. The most promising one, a RNA species bound to the chemically-synthesized fluorophore of GFP, is named Spinach, and was used to visualize the transportation of ribosomal RNA out of the nucleus (31). This technique is still in development, as its inventors are looking to enhance the signal derived from Spinach.

The first report of a system to follow RNA in living cells came out in 1998 (2), describing two constructs: the first, an RNA bacteriophage capsid protein fused to a GFP sequence and the second, multiple copies of the capsid protein's binding sites fused into the mRNA of a reporter gene. The reason for inserting multiple copies of these sites rather than a single one was to provide increased intensity of fluorescence at the binding sites, in order to distinguish between bound and unbound fusion proteins. The first application of the system was in yeast cells, but since then the system has been shown to work in bacteria (18), amoebae (9) mammalian cells (41) and *Drosophila* embryos (11). The use of a phage derived RNA binding protein prevents the possibility of attachment to non-specific DNA, as these RBPS recognize a unique structure as their binding site.

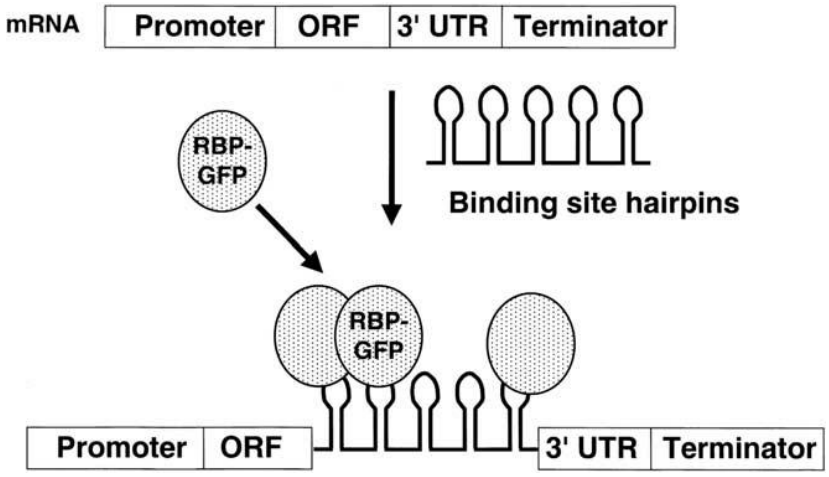



***Figure 1.1:*** *Scheme for GFP imaging of mRNA in living cells (6)**.*** *Two components are required: an mRNA with binding sites for an RNA-binding protein (RBP), shown here as hairpins, and the RBP–GFP fusion. The hairpins can be positioned within the*

*mRNA so as not to affect mRNA function. ORF, open reading frame; UTR, untranslated region.*

In order to reduce the background noise of fluorescence, a major limitation of this system, some groups have chosen to use RNA binding domains fused to split eGFP fragments, which do not fluoresce unless they're united. A union occurs only when the binding domains both interact with an RNA molecule, bringing the eGFP fragments close enough to interact (30). This technique proved to greatly reduce noise from unbound fusion proteins, but it suffers from the major disadvantage of a slow disassociation rate of the fluorescent fragments, preventing them from splitting up again when the cognate RNA undergoes degradation. For this reason, it is still necessary to be able to distinguish between bound and unbound fusion proteins. (37, 30).

## 1.2 Single stranded RNA bacteriophages

The first bacteriophage coat protein that was put to use in this capacity was that of the phage MS2. This is a single stranded RNA coliphage, the capsid protein of which recognizes a 19-nucleotide stem loop structure of RNA in the phage genome (29). By binding to this secondary structure, that includes the Shine-Dalgarno sequence and the initiation codon of the phage *replicase* gene, the coat protein causes translational repression. The mechanism of repression is probably stabilization of the secondary structure by the binding of the coat protein (12). These properties of phage MS2 are conserved in all known single stranded RNA phages (35).

MS2 belongs to the *Leviviridae* family of phages. Among this family of coliphages are also Qβ, GA and SP. Its characteristics include a very small genome (3500-4200 nucleotides) and four coding regions: a major coat protein, a maturation protein, a replicase and either a lysis protein in the case of *Levivirus* species, to which MS2 and GA belong, or a read-through protein in the case of the *Allolevivirus* species, to which Qβ and SP belong. (5) The *Pseudomonas* phage species PP7 shares these properties and some others, which has led scientists to believe it shares an evolutionary ancestor with the *Leviviridae* phages (29). Like the *Levivirus* genera, PP7 has a lysis protein, while the lysis in *Allelovirus* is mediated by the maturation protein. Another

characteristic shared by all of these phages is the topology of the coat protein: an N-terminal β hairpin, a five-stranded anti-parallel β-sheet and two C-terminal α helices (8). The RNA binding region is compromised of a ten stranded β-sheet created by antiparallel association of the protomers in the dimer.

## 1.3 Interaction between the RNA and the coat proteins

### 1.3.1 Specificity of the interaction

Numerous papers have discussed the specificity of binding of different phage coat proteins to their cognate binding sites. These studies utilized synthetic constructs of *replicase –lacZ* fusions, which contain the translation initiation site of the phage replicase gene in place of the first nine codons of *lacZ.* This experimental design meant that the translation initiation of *lacZ* was dependent on the site specific to that replicase (32). Thus, when a coat protein was bound to the RNA molecule, translation initiation was inhibited. In this manner, it was possible to measure the affinities of various coat proteins to their wild type RNA operators, each other's wild type operators and various mutant operators. It has been determined that the coat proteins MS2 and Qβ are able to distinguish between their own RNA recognition sites and each other's. In addition, expression of LacZ from the translation initiation sites of both phages decreased significantly in the presence of the other coat protein, indicating binding of the coat proteins to each other's operators, albeit at a lower affinity than their own. MS2 was shown to repress *lacZ* expression from the Qβ binding site to around 30%, and Qβ repressed expression from the MS2 operator to around 75%, while repression of each coat protein on its own operator was by almost 100%. The PP7 coat protein reduced expression from its own binding site to less than 10%, from the Qβ site to 20% and to 70% from the MS2 site. The coat protein of MS2 reduced expression from the PP7 operator only by about 10% and Qβ did not repress that one at all, according to Lim et al. (25). Other single stranded RNA phages that have been tested in a similar manner are GA and SP. GA showed a high binding affinity for the MS2 operator, as well as for its own (27), while SP, whose amino acid sequence is 80% identical with that of Qβ, seemed to share Qβ's specificity to its binding site as well.

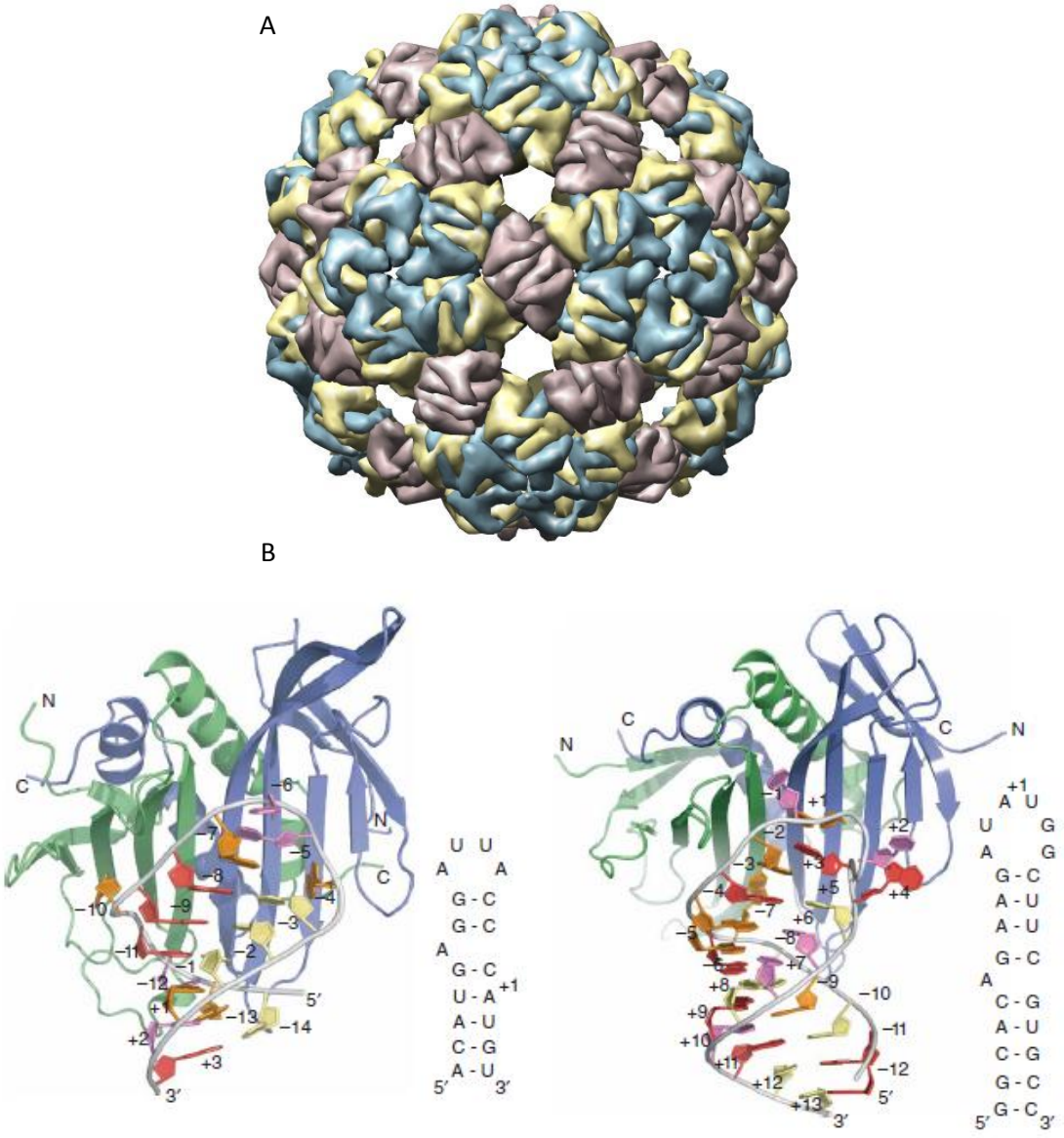
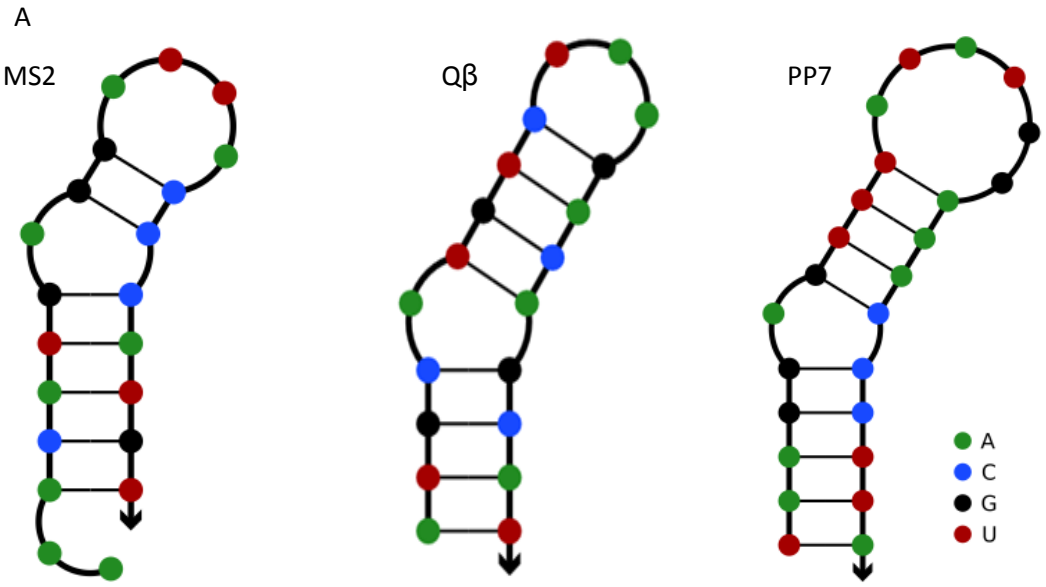



***Figure 1.2:*** *Coat proteins of single stranded RNA bacteriophages.* A: *Qβ capsid, constructed of 180 units of coat protein (19).* B: *Crystal structure of the MS2 (left) and PP7 (right) coat proteins, bound to their cognate RNAs (8). Nucleotides are labeled according to their relative position to the translation initiation site of the replicase gene.*

It is important to note that the identity of only certain nucleotides in the operators is crucial for binding. While the exact structure is needed for recognition by the coat proteins, it is thought that exchanging nucleotides at insignificant positions will still allow binding at a fairly high affinity. Essential nucleotides include the bulged purine in the middle of the base-paired region (figure 1.3) and some of the ones found in the loop. For MS2, for instance, the two adenosines on either end of the loop are critical,

as well as a pyrimidine before the second one. For PP7, it is critical for the sequence of the loop to be purine-U-A-G/U-G-purine in order for it to be recognized by its own wild-type coat protein (26). Structural studies have confirmed that these nucleotides are the ones involved in RNA-protein interaction (8).



B

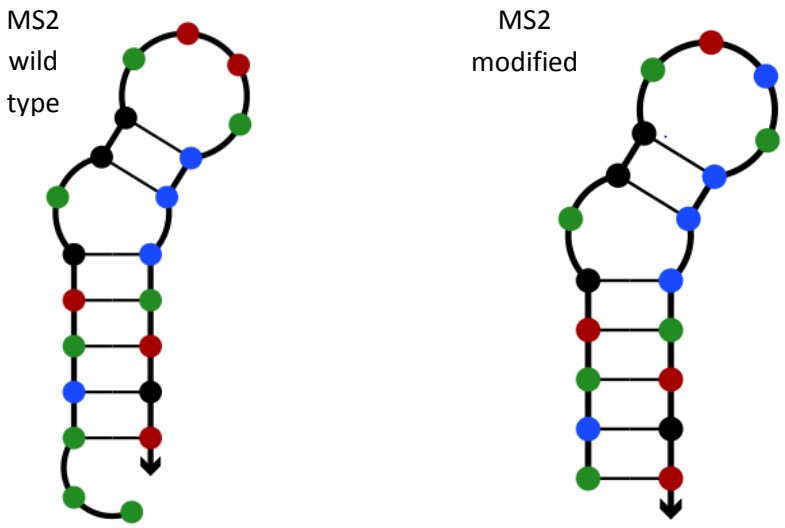



***Figure 1.3:*** *Models of the secondary structure of three RNA operators which the phage coat proteins bind. Images were generated using the software Nupack.* A*: wild type sequences of the structures.* B: *Structure of the wild type binding sequence of MS2 compared to that of a modified sequence, used in a study by Golding et al. (18). It can be seen that the exchange of certain nucleotides does not affect the structure of the sequence.*

1.3.2 Mechanisms of the RNA-protein interaction

The binding mechanisms between the phage coat protein and the RNA structure at the beginning of the *replicase* gene are not the same in different phages. MS2 and PP7 use two symmetrical pockets to bind the adenines in the bulge and in the loop (see figure 1.3). In the PP7 coat protein, these pockets are aligned along the dimer axis and are formed by residues from both protomers. In the MS2 coat protein, these pockets are perpendicular to the dimer axis and are formed by residues from only one protomer. While the residues in both binding pockets are similar, the roles of each one vary between the two phages. In the PP7 coat protein, the aliphatic portion of the side chain of residue Lys58 combines with residue Val83 to form the top wall of the pocket, whereas in the MS2 coat protein, Lys61 forms the bottom wall. The base A(-7) in PP7 makes van der Waals interactions with the valine residue. Hydrogen bonds with A(+1) and G(+3) stabilize the binding between the protein and the RNA at the second adenine binding pocket in the PP7 coat protein. In the MS2 coat protein, Thr45 and Ser47 line the A(–4) and A(–10) binding pockets and recognize the adenine base, while in the PP7 coat protein the residues Ser85 and Thr89 perform the

analogous role. The hydrophobic side chain of residue Val29 in the MS2 coat protein forms the top surface of the binding pockets, but in the PP7 coat protein this residue is replaced by an arginine (Arg24) (8). The corresponding adenine-binding pocket in Qβ is created by the residues Val32, Thr49, Ser51, Gln65 and Lys67.

One of the similarities between the binding mechanisms of the coat proteins of MS2 and Qβ is the position of the nucleotide adjacent to the critical adenine: A(+7) in the Qβ loop and U(-5) in the MS2 loop. The former is stacked with Tyr89, while the latter interacts with Tyr85 in the MS2 coat protein in a similar fashion: interaction between the hydroxyl group of the tyrosine and a phosphate of the RNA backbone. In the PP7 operator, U(-1) is extended away from the loop, into a pocket formed by Thr51, Ala52, Val91 and Thr81, with which it makes hydrogen bonds (8). The nucleotides U(+6) in the Qβ loop and U(-6) in the MS2 loop both point away from the protein and do not make contact with it. (34).

Unlike PP7 and MS2, the coat protein of Qβ recognizes only one nucleotide in the RNA secondary structure as critical for binding: the adenine in the loop. The bulged adenine in the stem loop, a conserved feature of many single stranded RNA phages, was found to be not essential to the binding by Lim et al. (28), who showed that mutating this nucleotide or deleting it altogether resulted only in a minor decrease of affinity. In MS2, the bulged adenine fits into a pocket identical to the one binding the adenine in the loop in a different orientation, but in the Qβ coat protein this pocket is empty. Qβ's bulged adenine (A+1) is stacked with residue Tyr89 of the other protomer, a configuration that has not been observed in any other RNA-protein complex. Other distinctions between Qβ and MS2 are the lengths of both the stems and the loops: the short, three nucleotide hairpin loop of the Qβ secondary structure was found to be critical for the binding of its coat protein (addition of a nucleotide reduced its affinity by over 10-fold), whereas the MS2 wild type hairpin contains four nucleotides (39). The MS2 coat protein was not as discriminatory about the size of the loop, and tolerated three nucleotide loops almost as well as four nucleotides ones, so long as the distance between A-10 and A-4 was preserved. In the Qβ operator, the distance between the analogous adenines is longer by one nucleotide, preventing the bulged one from reaching the binding pocket found in the MS2 coat protein. Furthermore, recognition of the RNA structure by Qβ coat protein required eight base pairs in the stem, while seven were enough for the MS2 coat protein. These discrepancies are probably the reasons why the binding affinities of the two coat

protein to each other's binding sites were found to be significantly lower than to their own. (34).

The importance of base specific interactions varies between the three coat proteins discussed above. The PP7 protein does not make any interactions beside the ones that are specific for four of its nucleotides. In MS2, three bases make direct contact with the protein, as well as the RNA backbone in the stretch between the bulged adenine and the loop. In Qβ, the majority of the interactions between the protein and the RNA involve the sugar-phosphate backbone, and the only nucleotide involved in direct contact with the protein is the adenine in the loop. (34).

## 1.4 The RNA binding peptide λN

A more recent RNA binding protein put to use in live imaging is the phage λ regulatory peptide N (10). This 22-amino-acid peptide has the function of an anti-terminator, regulating transcription of two of the bacteriophage genes. Its binding site is a 15-nucleotide RNA hairpin named BoxB (22, figure 1.4). Like the coat proteins, this peptide shows potential as a useful tool for RNA imaging and has already been used alongside a fusion with MS2 in order to allow tracking of two RNA molecules in a single cell (23).

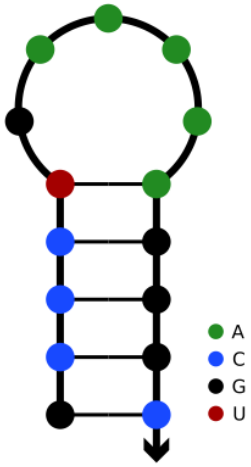



***Figure 1.4:*** *Model of the secondary structure of the BoxB RNA, to which the peptide λN binds.*

## 1.5 Applications of the system

Originally used to follow mRNA trafficking in eukaryotic cells (2), the potential of this method has since then been recognized by various groups of scientists, and it has been put to use in many types of cells for diverse research questions. The first study to make use of it showed that the transportation of *ASH1* RNA occurs via a macromolecular complex, the existence and location of which depends on the 3'UTR of *ASH1* mRNA, and that its speed was consistent with that generated by a cellular motor. In a different study, a fusion of MS2-GFP was used to study the mechanism of localization of *nos* RNA in *Drosophila* embryos, leading to the discovery of two different pathways of transportation for this species of RNA: diffusion and entrapment at the posterior end of the egg by association with germ plasm, in addition to microtubule-dependent transportation (11). A third study, performed in mammalian cells, resulted in the discovery that mRNA molecules can move directly on microtubules and are subjected to continuous cycles of anchoring, diffusion and transport, regulated by specific sequences (13).

Another field that has benefited greatly from this method is that of transcription dynamics. Using one fusion protein has enabled the discovery that transcription from bacterial promoters is bursty (18). In 2011, Suter et al., working in mammalian cells, used this system to measure bursting kinetics of individual genes (36). Usage of two such fusions, PP7-EGFP and MS2-mCherry, simultaneously, allowed discovery that two alleles of the same gene have different rates of transcription in the same cell (20).

Recent developments in this field include putting the phage-based system to study viruses. Boireau et al. used a fusion of MS2-GFP to study the synthesis kinetics of HIV mRNA (4). Viral RNA trafficking inside the cell was also studied with this system (1). Both studies indicated that the addition of a foreign RNA sequence- the coat proteins' binding sites- to the viral genome did not alter its activity, which signifies the system's potential to uncover yet another field of research which has previously been obscure.

## 1.6 Obstacles within the system

The chief obstacle in building this system is the cloning of the multiple binding sites. At least 24 repeat of the MS2 binding site are required in order to detect single

molecules of RNA in mammalian cells (13), and studies in bacteria have used up to 96 repeats (17). The assembly of these constructs requires multiple, repetitive cloning steps. Furthermore, these DNA repeats are prone to deletions by recombination events occurring during DNA replication (33). A significant advantage of the λN system is its requirement for no more than 16 repeats of the BoxB element in order to generate a good signal, and even less according to some studies (10, 23). The currently most common means of avoiding this problem is by inserting slight changes in these sequences (20). By switching the nucleotides which are not crucial for coat protein recognition, such as the ones in the based-pair region, it is possible to facilitate the cloning process and reduce the recombination events.

## 1.7 Research significance and objectives

As evidence of the involvement of RNA in processes in single and multicellular organisms accumulates, it is imperative for biological research to have an easy-to-use tool to study it. Tagging the RNA with fluorescent proteins has proved, in the past decade and a half, to be a non-interfering, reliable method of visualization. Current cloning techniques have made the creation of fusion proteins fairly simple, but the production of multiple repeats of binding sites remains a time-consuming setback. Consequently, the goal of this research is to develop a library system for easy implementation of this technology in any RNA-related application. The library system will consist of two libraries, one of fusion proteins and the other of binding sites compatible for each of them. Together, the dual library system will make it simpler to set up the RNA tracking system in different cells and for different purposes.

In order to reach this goal, the specific objectives of this research are as follows:

1. Create and characterize a library of fusion proteins, consisting of bacteriophage- derived RNA binding proteins and fluorescent proteins.
2. Design an assay that will enable easy, rapid and reliable detection of alternative binding sites to the RNA binding proteins.

# 2. Materials and Methods

## 2.1 Materials

2.1.1. Enzymes: all enzymes (restriction enzymes, ligases and polymerases) were purchased from New England Biolabs (NEB).

2.1.2 Bacterial growth media: all reagents in this section were purchased from Sigma-Aldrich Co., unless stated otherwise.

Luria Betani (LB):

1%Bacto™ Tryptone

0.5% Bacto™ Yeast Extract

1%NaCl™

1.8% Bacto™ Agar for plates.

Antibiotics: 0.1% (when relevant) ampicillin/kanamycin/chloramphenicol

For recovery after transformation, the following materials were added to LB: 1% 1M $MgSO_4$, 1% 1M $MgCl_2$ and 2% 1M glucose (all three are products of Merck).

Low growth media for bioassays:

0.05% Bacto™ Tryptone

0.58% NaCl™

0.05M $MgSO_4$

1%Phosphate Buffered Saline (Biological Industries/Beit Haemek)

0.03%glycerol (Gadot)

SeaPlaque™ GTG™ Agarose (Lonza)

2.1.3 Kits:

- GeneJet Plasmid Miniprep Kit (Thermo Scientific) for DNA extraction and purification
- Wizard® SV Gel and PCR Clean-Up system (Promega) for DNA purification from gels and in-vitro enzymatic reactions.

2.1.4 Bacterial strains

- *Escherichia coli* Top10 cells (Genotype: F- *mcrA* Δ( *mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ( *araleu*)7697*gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G). This strain was a gift from Ben Zion Levi's laboratory at the Faculty of Biotechnology and Food Engineering.
- *Escherichia coli* 3.300LG cells (Genotype: ΔGlnL:ΔGlnG) used in the Amit lab for testing synthetic enhancers.

2.1.5 Vectors

- pCP and pc-mCherry/pc-cerulean are versions of the basic pPROLar vector, used in the Amit lab for construction of various libraries. The main difference between the two is the promoter of the MCS: in pCP it is $P_{LtetO}$, and in pc-mCherry it is $P_{Lac.}$ These vectors include a selection marker of kanamycin resistance, and have a low copy number origin of replication, p15a.
- pUC19: a high copy number vector, containing a selection marker of ampicillin resistance. This vector was a gift from Ben Zion Levi's laboratory at the Faculty of Biotechnology and Food Engineering.
- pACT-Tet: a high copy number vector, containing a selection marker of ampicillin resistance and expressing the repressors TetR and LacI (Amit et al. 2011).

2.1.6 Genes and RNA binding sites

2.1.6.1 Genes:

- MS2dlFG coat protein: isolated from Robert Singer's Addgene plasmid 27121 (13).
- PP7dlFG coat protein: isolated from Kathleen Collins' Addgene plasmid 27548 (21). (Used in the fusion PP7-cerulean)
- PP7 coat protein: purchased from Genscript. (Used in the fusions PP7-dendra and PP7- mCherry).
- Qβ coat protein: purchased from Genscript.
- λN peptide: purchased from Genscript.
- mCherry, cerulean, mDendra2: used in the Amit lab as reporter genes.

2.1.6.2 Binding sites:

Binding sites of the proteins were synthesized by adding them unto primers (all primers were purchased from IDT). Table 2.1 contains the sequences of these sites.

Table 2.1: sequences of the RNA binding sites

| Name | Sequence |
|---|---|
| MS2 (18) | ACAUGAGGAUCACCCAUGU |
| PP7 (Delebeque et al. 2011) | CACAGAAGATATGGCTTCGTG |
| Qβ (28) | ATGCATGTCTAAGACAGCAT |
| Standard RBS | ATTAAAGAGGAGAAA |

2.1.6.3 Linkers:

Fusion proteins were constructed with short linkers of amino acids between the 5' gene (the RNA binding protein) and the 3' gene (the fluorescent protein). Two different linkers were used, not due to any theory-based strategy, but for the simple reason that these constructs were the first to be cloned successfully into the pCP vector. These linkers encoded for seven amino acids (cloned from Singer's Addgene plasmid number 27121) and ten amino acids (cloned from Collins' Addgene plasmid number 27548). The DNA sequences of these linkers are described in table 2.2, as are the fusions they were inserted into.

Table 2.2: DNA sequences of the linkers used to fuse the proteins

| Ggtggcgaccggtggatccgc (Singer) | Ggatccctgcaggcctcagggcccgatcgatgcggccg (Collins) |
|---|---|
| MS2-mCherry | Qβ- mCherry |
| MS2-dendra | Qβ-dendra |
| MS2-cerulean | Qβ-cerulean |
| PP7-dendra | PP7-cerulean |
| PP7-mCherry | |
| λN-mCherry | |
| λN- dendra | |
| λN-cerulean | |

## 2.2 Methods

2.2.1 Cloning

Fusion proteins were created and inserted into pCP vectors using the Gibson assembly method (15). This method was also used to create the pc-mCherry/cerulean vectors used in the binding assay. This method uses a single isothermal step to perform a recombination reaction. It requires DNA fragments with overlapping sequences at their ends and three enzymes: T5 exonuclease, Taq ligase and Phusion polymerase (figure 2.1). The reaction is performed for one hour at 50°C. Each reaction tube (15µl) contains the following:

4µl 5xISO buffer (see below)

0.008µl 10 U/µl T5 exonuclease

0.25µl 2 U/µl Phusion polymerase

2 µl 40 U/µl Taq ligase

8.742 µl $H_2O$

The 5xISO buffer contains the following:

50% 1M Tris-HCl pH 7.5

2.5% 2M $MgCl_2$

4% 100mM dNTP mix (25 mM each: dGTP, dCTP, dATP, dTTP)

5% 1M DTT

1.5gr PEG-8000

5% 100 mM NAD

$H_2O$- completion to 100%

For assembly of two fragments, the linearized vector backbone was added at a concentration of 0.05M and the insert at a concentration of 0.15M.

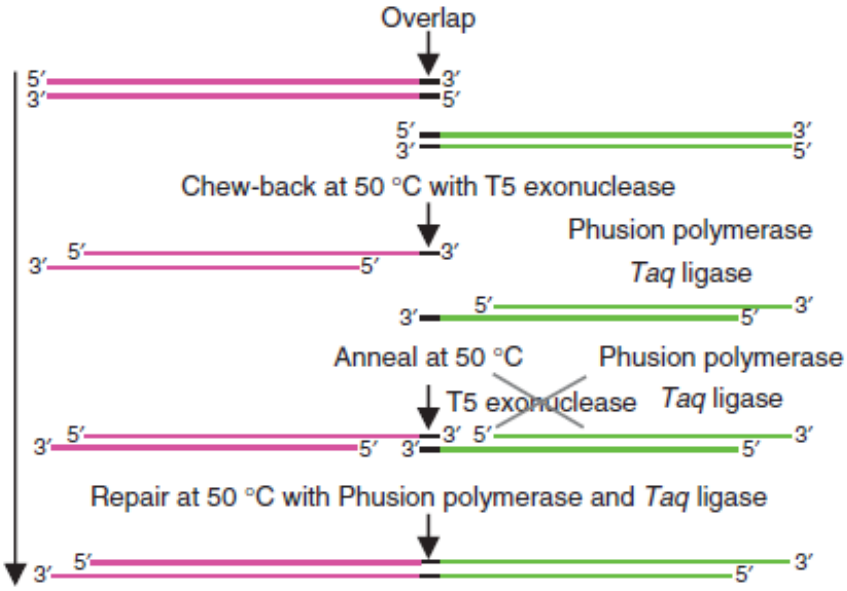



***Figure 2.1:*** *Scheme of the Gibson assembly method (16). T5 exonuclease chews off the 5' ends if the fragments, complementary single strand DNA overhangs anneal, the Phusion polymerase repairs the gaps created by unspecified exonuclease activity and the Taq ligase seals the nicks.*

Enzymatic reactions (PCR, restriction, and ligation) were carried out as recommended in NEB's protocols.

Transformation into empty cells was done using the heat shock method with chemically competent Top10 cells. Recovery was performed using LB medium enriched with 1%$MgSO_4$, 1%$MgCl_2$ and 2% glucose, while shaking at 250 rpm at 37°C.

Transformation of a second plasmid into a cell, or transformation into 3.300LG cells was performed by electroporation, after growing the cells to an $O.D._{600nm}$ of ~ 0.6 and washing the culture repeatedly with deionized water and a 10% glycerol solution. Recovery was performed as mentioned above.

2.2.2 Fluorescent functionality of the fusion proteins

For measuring fluorescence, cells were grown overnight in a flask in LB medium, shaken at 250 rpm at 37°C. The cells were diluted in the morning into fresh LB medium and grown at the same conditions to an $O.D._{600nm}$ of ~ 0.6. Upon arrival at this O.D., the medium was changed to a low growth one suitable for bioassays, and the inducer anhydrous tetracycline (aTc) was added. Relevant antibiotics were present in both media. Fluorescence levels were checked on a plate reader (Tecan i-control infinite 200Pro) at the appropriate wavelengths once an hour, starting with three hours after induction and up to five hours after. Excitation and emission wavelengths for each fluorescent protein are described in table 2.3. Cell density was measured again on the plate reader, at a wavelength of 595nm.

Table 2.3: Excitation and emission wavelengths of fluorescent proteins used in fusions

| **Protein** | **Excitation [nm]** | **Emission [nm]** |
|---|---|---|
| cerulean | 420 | 485 |
| Dendra2 | 495 | 535 |
| mCherry | 560 | 612 |

The plates used in this assay were 96 well, flat bottom black polystyrol plates manufactured by Greiner.

2.2.3 The binding assay

Cells were grown overnight in a flask in LB medium, in the same conditions a described for the fluorescence assay and diluted the following morning to an $O.D._{600nm}$ of roughly 0.2-0.7. The dilution was performed while switching the medium from LB to the same low growth one mentioned in the previous section. In this medium, the cells were suspended at 37°C, 250 rpm for nine hours and their fluorescence was monitored with a plate reader every two hours. The same plates and plate reader that are mentioned in section 2.2.2 were used in this assay.

2.2.4 Microscopy images

The microscopy images presented in this thesis were acquired by Dr. Sarah Goldberg. Cells were grown overnight in a flask in LB medium, shaken at 250 rpm at 37°C. The cells were diluted in the morning by a factor of 1:100 into the low growth medium mentioned previously. After the change of medium, the cells were returned to 37°C and shaken at 250 rpm for at least two more hours.

Slides were prepared according to the protocol described by Young et al., (40). 1.5% (weight/volume) low-melt agarose (SeaPlaque™ GTG™ Agarose) was added to 10 ml of Phosphate Buffered Saline (0.15 g of the agarose) and dissolved by microwaving. After a few minutes of cooling, ~1 ml of agarose was pipetted onto a 22-$mm^2$ cover glass slide. A second cover glass was placed on top of the agarose to create an agarose sandwich in between the two slides. These pads were then left to solidify at room temperature for about an hour.

Once the cells were ready for imaging, agarose pads were uncovered and cut into smaller pieces, roughly around 5mm wide, using a scalpel. One microliter of cells was pipetted on each one and left to dry at room temperature for fifteen minutes. Next, pads were flipped onto a cover glass–bottom dish, with the bacteria sandwiched between the agarose pads and the cover glass. The cover glass dishes were sealed with Parafilm in order to prevent evaporation and taken to the microscope. An Andor Xion-Ultra EMCCD camera was used for obtaining the images presented in this work.

# 3. Results

## 3.1 Fluorescent functionality of the fusion proteins

The first part of the research was to create fusions of phage coat proteins and fluorescent proteins. The coat proteins of MS2, PP7 and Qβ, along with the peptide λN, were cloned upstream of the fluorescent proteins cerulean, Dendra2 and mCherry, with one of two types of linkers between them: GGDRWIR or GSLQASGPDRCG (see Table 2.2). The advantage of using these three fluorescent proteins in particular is that their excitation and emission spectra barely overlap, which makes them promising candidates for imaging of multiple RNA molecules simultaneously. The decision to place the RNA binding proteins at the N-terminus and the fluorescent one at the C-terminus was random, and we believe switching between the components would not affect their ability to function. Fusion proteins were cloned into chemically competent cells, from which DNA was extracted and sequenced in order to determine whether the cloning was successful. This initial step in the research accomplished the creation of twelve fusions: each of the four binding proteins with all three fluorescent proteins.

In order for the system to work, both components of the fusion proteins were required to function just as they did outside of the fusions, which means the phage proteins needed to maintain their ability to recognize and interact with their cognate RNA species, and the fluorescent ones needed to be able to fluoresce at the expected wavelengths. In order to test their capabilities, we decided to first examine their fluorescence alone, using a plate reader to observe excitation and emission from cells that contained the fusions versus cells that lacked them. This assay was intended to ascertain whether the fluorescent proteins, placed at the C-terminus of the fusion and lacking the initiation codon, could be detected via their unique feature. Inability to fluoresce would render the imaging system useless.

Fusion proteins were tested for fluorescence after being cloned into pCP vectors (p15a origin of replication, kanamycin resistance), under the regulation of the promoter $P_{LtetO}$ (see figure 3.1). Fluorescence assays were performed in 3.300LG cells. The repressor TetR, which regulates the $P_{LtetO}$ promoter, was expressed from an additional plasmid, pACT-Tet (pMB1-derived origin of replication, ampicillin resistance). Cells were diluted from overnight starters and grown in LB medium until culture reached

O.D.$_{600nm}$ 0.6. After that it was transferred to a low growth buffer and the inducer anhydrous tetracycline (aTc) was added in rising concentrations.

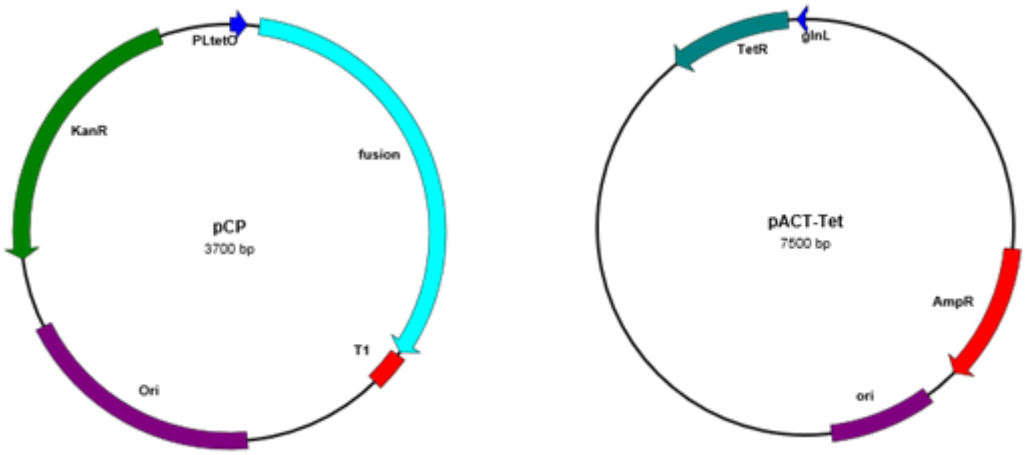
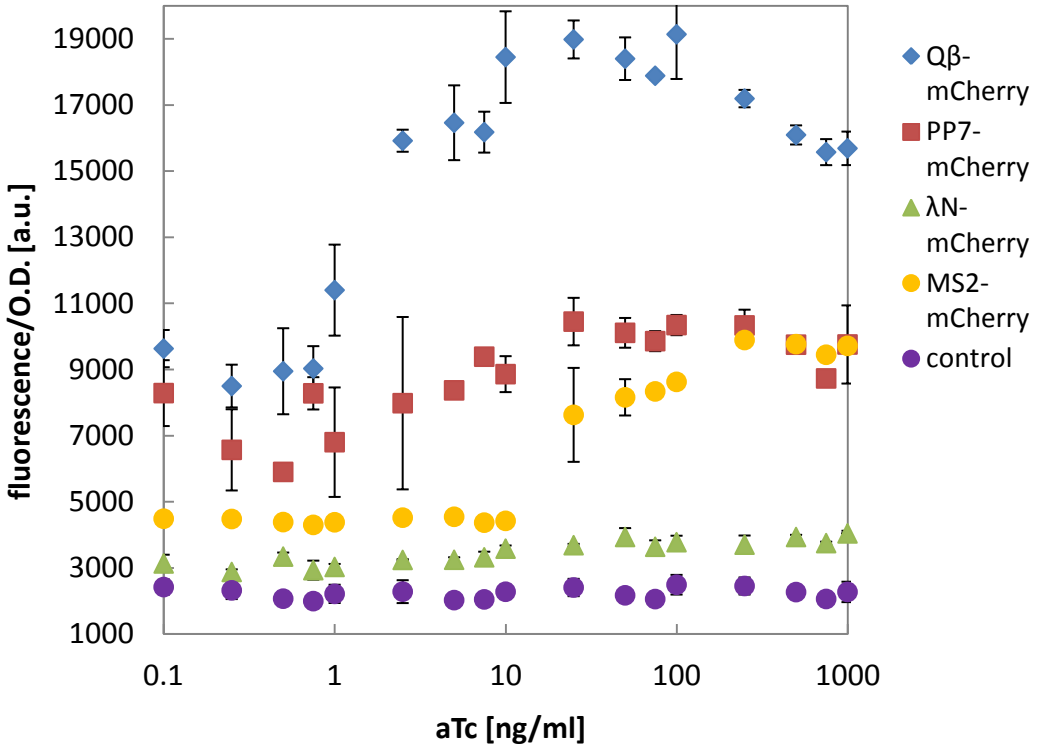


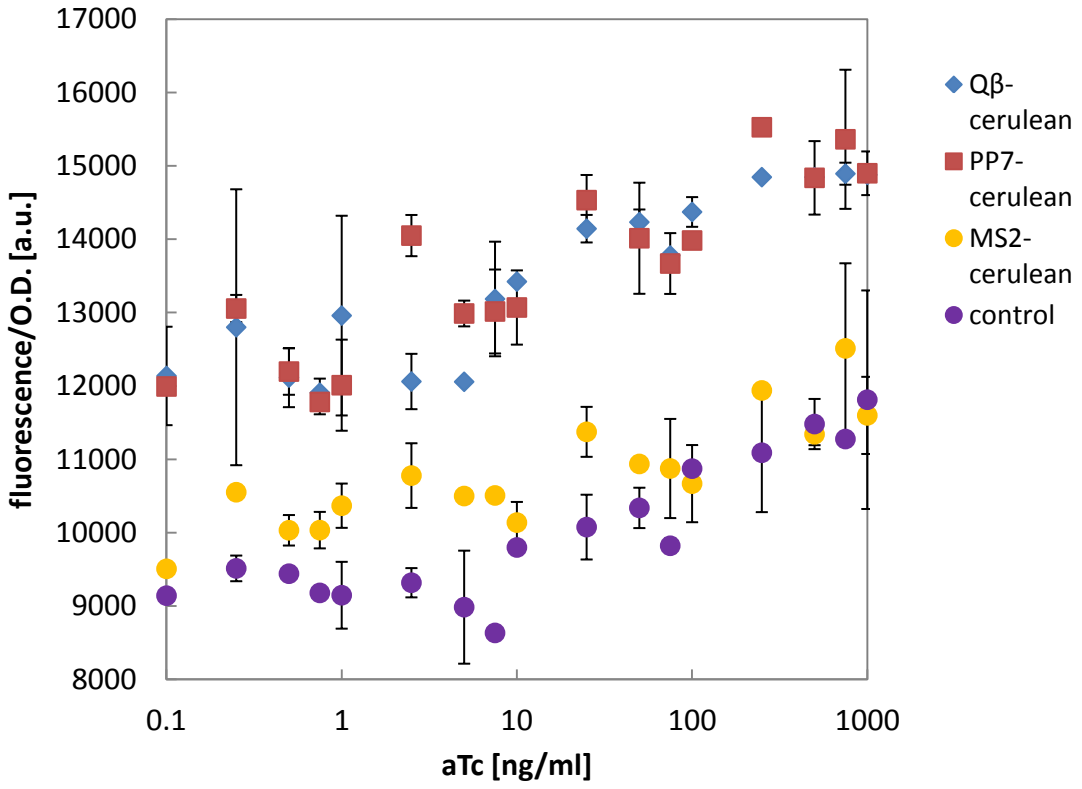

***Figure 3.1***: *The two plasmids constituting the system of the fluorescence functionality assay, pCP and pACT-Tet. The fusions all had a phage coat protein at the N-terminus, lacking a stop codon, a linker of either seven or ten amino acids and then a fluorescent protein lacking a start codon at the C-terminus.*

Figure 3.2 shows the levels of normalized fluorescence (raw fluorescence measurements divided by the optical density measured at absorption wavelength 595nm) as affected by the concentration of the inducer, five hours after its addition to the culture. The control strain lacked the plasmid expressing the fusions (pCP), which means its measured fluorescence is the result of the autofluorescence of the cell and from the proteins expressed from pACT-Tet (mainly TetR), which was expected to be distinctly trivial compared to the fluorescent proteins that are part of the fusions.

A


19000
17000
15000
13000
11000
9000
7000
5000
3000
1000
fluorescence/O.D. [a.u.]
0.1
1
10
100
1000
aTc [ng/ml]
Qβ-mCherry
PP7-mCherry
λN-mCherry
MS2-mCherry
control

B


17000
16000
15000
14000
13000
12000
11000
10000
9000
8000
fluorescence/O.D. [a.u.]
0.1
1
10
100
1000
aTc [ng/ml]
Qβ-cerulean
PP7-cerulean
MS2-cerulean
control

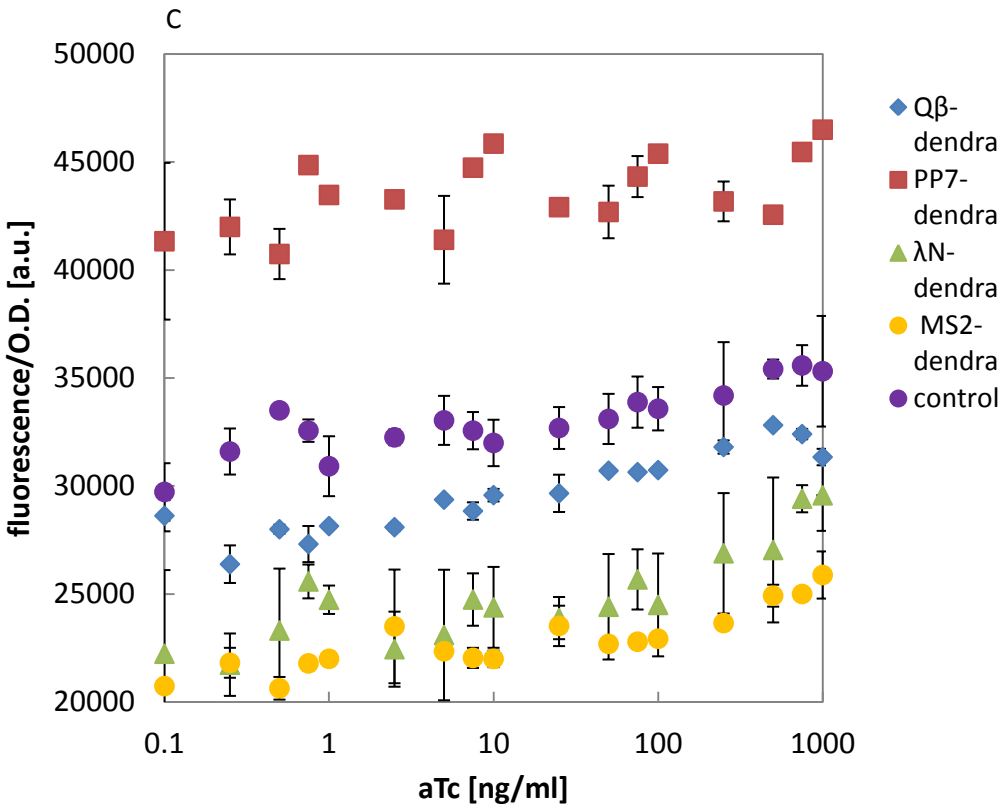



***Figure 3.2:*** *Fluorescence measurement of the fusion proteins.* A: *fusions with the fluorescent protein mCherry, the excitation wavelength of which is 560nm and the emission 612nm.* B: *fusions with the fluorescent protein cerulean, the excitation wavelength of which is 420nm and the emission 485nm.*C: *fusions with the fluorescent protein dendra2, the excitation wavelength of which is 495nm and the emission 535nm.*

The fluorescent measurements shown in figure 3.2A are for all the fusions in which the fluorescent protein is mCherry, measured at excitation and emission wavelengths 560 and 612nm respectively. The blue diamonds, corresponding to Qβ-mCherry, show an initial low level of fluorescence at low concentrations of aTc, and then a sharp increase at aTc values of 1-10 ng/ml, to a higher level (roughly two-fold of the base line). A similar behavior is also observed for PP7-mCherry (red squares) and MS2-mCherry (yellow circles), although not to the extent shown with Qβ. The clearly identifiable response to aTc indicates that the levels of the fluorescent proteins are proportional to the number of active TetR proteins in the cells, which is consistent with our expectations that a fusion protein, containing the fluorescent protein at the C-terminus, was expressed from the $P_{LTetO}$ promoter. Interestingly, the high levels of base line fluorescence demonstrated by these three fusions, despite the presence of a

strong repressor, indicate leakage in the circuit. Furthermore, the levels of induction observed in these experiments were lower than normal induction, which can raise expression levels by at least ten-fold. This result might be improved by changing promoters to one which is less leaky and is able, when active, to induce higher transcription rates as well. λN-mCherry (green triangles) did not rise above basal level, which was extremely close to the level of the control strain (violet circles).

Figure 3.2B depicts fluorescence measurements of the coat proteins-cerulean fusions. The λN-cerulean fusion was not tested for induction because earlier experiments showed its fluorescence was below the control level (data not shown). MS2-cerulean (the yellow circles) showed no significant fluorescence above the control level, while the Qβ-cerulean and PP7-cerulean (blue diamonds and red squares, respectively) fusions did show fluorescence, but the rise appeared to be linear, and a similar rise was spotted in the control strain. This is in contrast to the mCherry fusions, where the control strain's fluorescence stayed flat and the increase in the coat proteins-mCherry fusions was sharp.

The fusions with dendra2 as a fluorescent protein are shown in figure 3.2C. Three of the dendra2 fusions, MS2, Qβ and λN, represented by yellow circles, blue diamonds and green triangles respectively, gave fluorescence which was below the control level (violet circles). The fourth, PP7-dendra (red squares), gave about 30% more fluorescence in those wavelengths than the control, but like the cerulean fusions, did not seem to be affected by the inducer. Upon cloning of the PP7, MS2 and Qβ-dendra2 fusions into plasmids with a high copy number (pUC19) and under the regulation of the promoter $P_{lac}$, the ratio between the fluorescence of the fusions compared to the autofluorescence of the control strain (Top10 cells lacking the fusion) improved to a high extent (Figure 3.3).

All three λN fusions (mCherry, cerulean and dendra) were not taken to further tests since they did not appear to be functional fluorescence-wise.

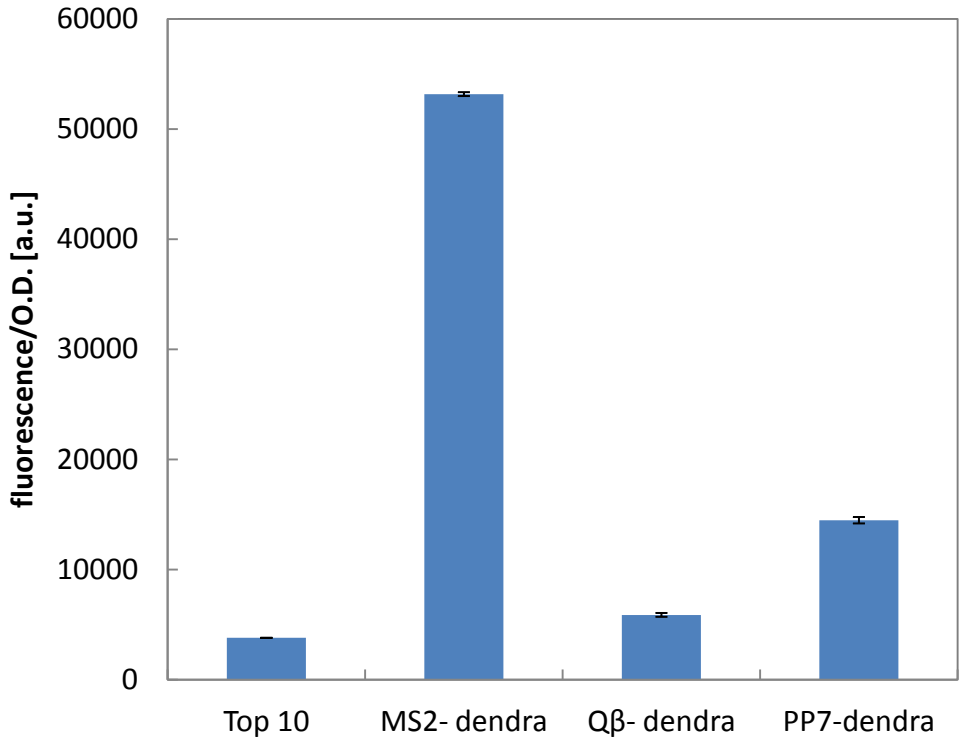



***Figure 3.3:*** *Fluorescence measurement of the fusion proteins with dendra2 on a high copy number plasmid. Top10 is the control strain, which lacked the plasmid from which the gene encoding for a fusion protein was expressed.*

3.2 The binding assay

In order to confirm the functionality of the fusions, it was necessary to test their RNA-binding capabilities in parallel to measuring their nascent fluorescence. To that end, a second fluorescent assay, termed the binding assay, was constructed. This assay exploits the phage coat proteins' function as natural translational repressors by binding their cognate RNA operators. In order to show that the fusions are also fully functional RNA binding proteins, the operators were placed upstream of reporter genes, under the assumption that the binding of the proteins to the operators would affect the expression of the genes.

Eight of the cloned fusions, the ones that showed they were able to function as fluorescent proteins, were taken for further experiments, intended to determine their specificity of binding. The binding sites of PP7, MS2 and Qβ were inserted in place of a typical RBS found in *E. coli* cloning vectors, upstream of a reporter gene encoding a fluorescent protein, either mCherry or cerulean. This construct was encoded from a

low copy number plasmid (p15a origin of replication, named pc-mCherry or pc-cerulean, according to the reporter), while a high copy number plasmid (pUC19, pMB1-derived origin of replication) expressed a fusion protein (figure 3.4). Different combinations of fusion proteins and binding sites were transformed into the same cells, and fluorescence measurements were used to assess the effect of each fusion protein on each binding site (figure 3.5). Cells that contained mCherry as a reporter included fusions of either cerulean or dendra, while the mCherry fusions were transformed into cells that had cerulean as the reporter. Binding affinities can be compared quantitatively by measuring the translation regulatory effect reflected by the fluorescence level. Full (100%) expression was defined as the expression of the reporter in the absence of any fusion, i.e., in the presence of the original pUC19 plasmid.

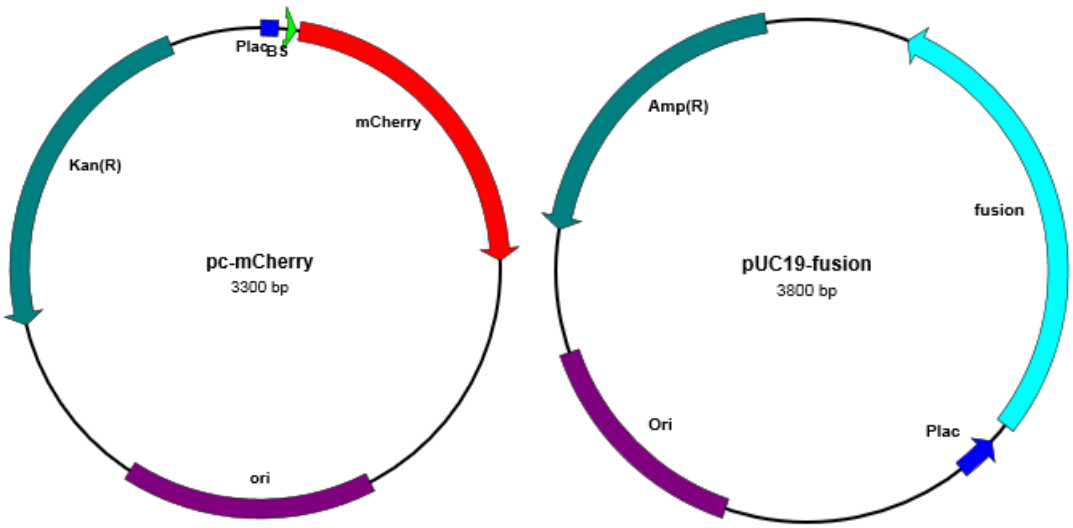



***Figure 3.4***: *The two plasmids constituting the system of the binding assay, pc-mCherry and pUC19-fusion.*

3.2.1 Effects of the fusion proteins on the various RBSs

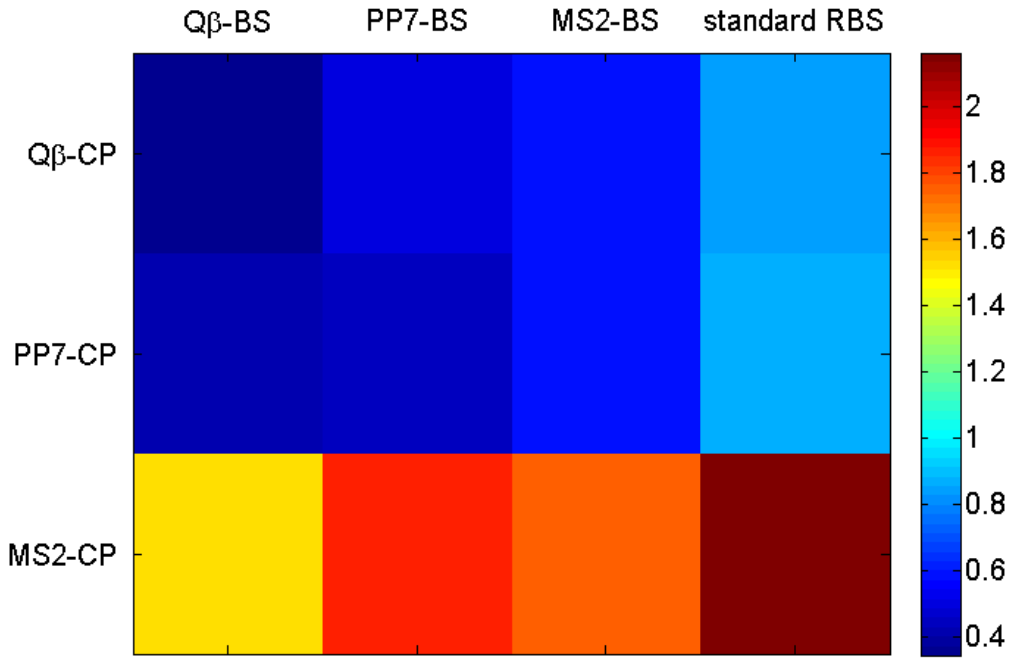
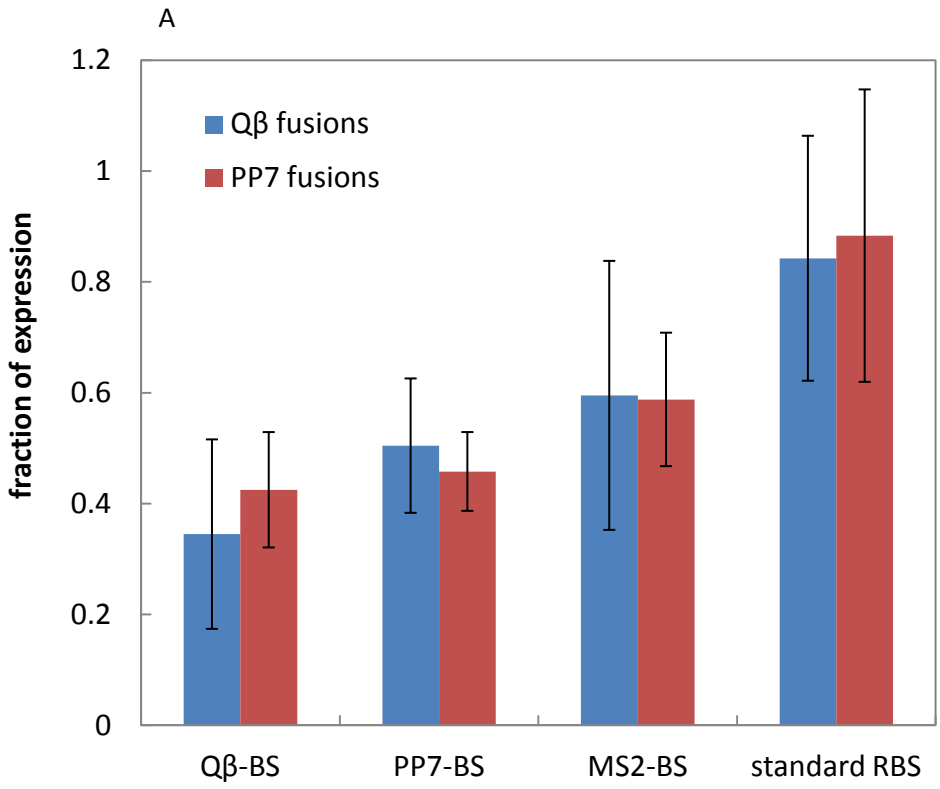



***Figure 3.5***: *Heat map of the effects of each coat protein (as part of the fusion) on three different binding sites and on the standard RBS, using mCherry as the reporter. The numbers on the right signify the fraction of expression.*

Figure 3.5 is a heat map that summarizes all the data from the binding assay experiments. The dark blue squares show that both Qβ and PP7 generate translation repression as expected, while MS2 promotes translation (yellow, orange and brown squares) via some unknown mechanism. In figure 3.6, which zooms closer on the data, one can see that both PP7-FP and Qβ-FP fusions repress at about the same level constructs that contain the other protein's binding sites as compared with constructs containing their cognate site. In addition, their binding to an MS2 site yields reduced repression (to about 70%), which is consistent with previous experiments (25). This implies that both Qβ and PP7 can bind a multitude of binding sites, and are not orthogonal as previously thought. Finally, expression from the standard RBS was not significantly affected by the fusions, which serves as a control.

In contrast to the PP7 and Qβ results, the MS2-dendra fusion did not repress expression from any of the binding sites, or the standard RBS, but seemed to amplify it instead. In figures 3.5 and 3.6B, it is easy to see that all combinations including this

fusion gave expression which was above the defined hundred percent. The percentage of expression varied from 150% on the Qβ BS to over 200% seen on the standard RBS.



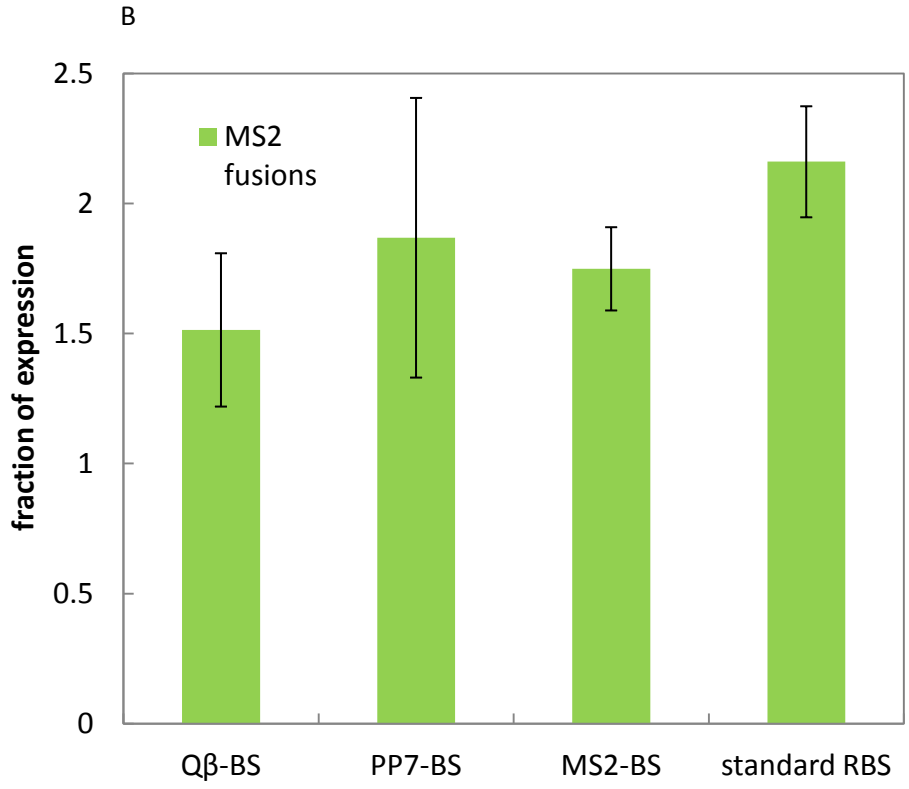



***Figure 3.6***: *The effects of the fusions on expression from different RBSs with mCherry as the reporter protein.* A: *PP7 and Qβ fusions,* B: *MS2-dendra fusion.*

3.2.2 Strengths of various RBSs

When quantifying the effect of the fusions in each of the binding site, each result was normalized by the values obtained from the binding sites alone. This was because each binding site's capacity as a ribosome binding site is different, leading to different expression levels even without repression. These differences suggest that even with similar levels of repression, which is the case with Qβ and PP7 on each of their binding sites, the resulting level of reporter protein in the cell is different. The ability of each binding site to function as a translation initiation site was assessed, and their relative strengths compared in figure 3.7. It is clear that all three phage operators are weak RBSs compared with the standard RBS used for optimal gene expression in various vectors. The variance between each of the three operators is also significant, with the one from phage Qβ at least twice as strong as the other two.

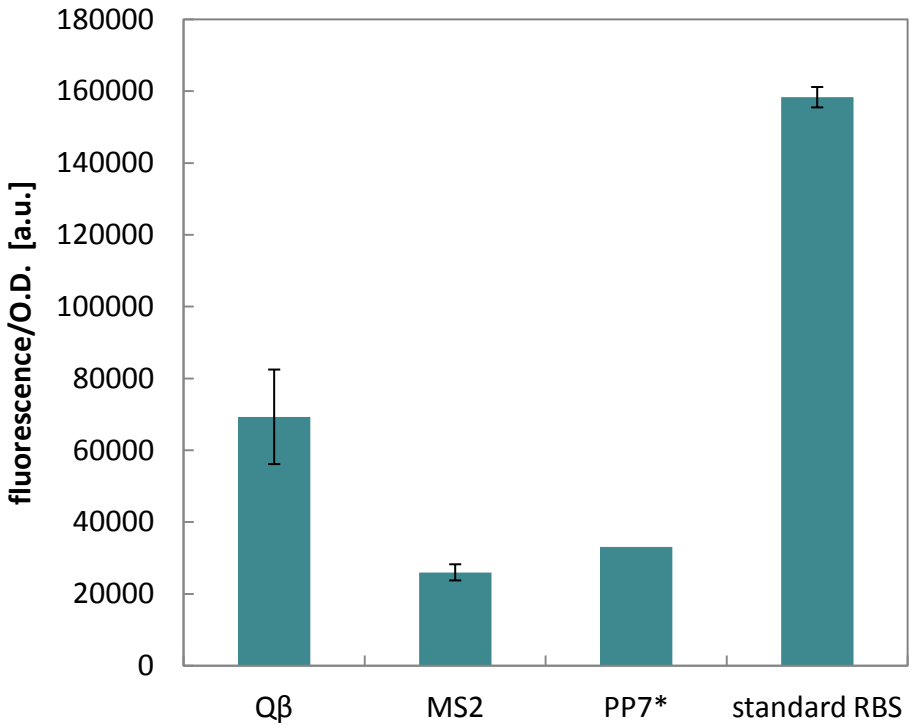
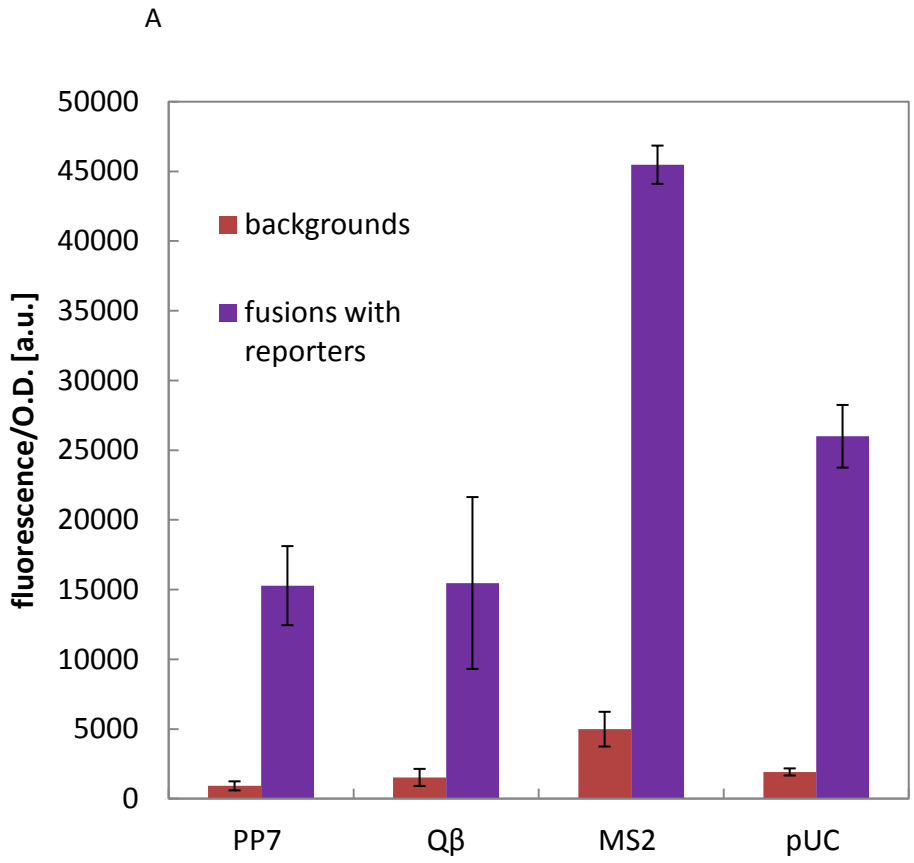



***Figure 3.7***: *Relative strengths of the coat protein operators as ribosome binding sites. (*Not enough data was obtained for assessment of the standard deviation of the PP7 site).*

### 3.2.3 Background measurements

The background fluorescence of the cells and fusions was measured in order to make certain the fluorescence measured came indeed from the level of the reporter protein, and was not a byproduct of other components in the cell. Figure 3.8 shows the levels of background fluorescence in the mCherry wavelengths (A) and cerulean wavelengths (B). These measurements were done in strains in which the pc-reporter plasmid lacked both reporter and any ribosome binding site after the promoter Plac.



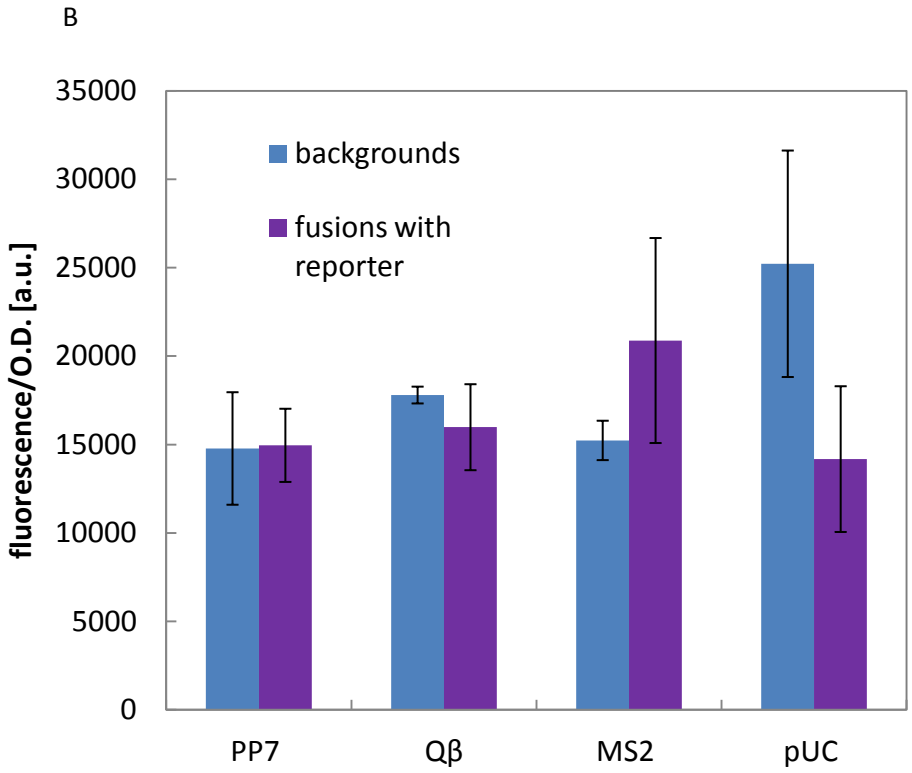



***Figure 3.8***: *Background fluorescence of the cells with pUC vectors* inside them. A: *fluorescence in the mCherry channel (excitation 560nm, emission 612nm);* B: *fluorescence in the cerulean channel (excitation 420nm, emission 485nm).*

The background fluorescence was compared to the fluorescence of cells with the same fusions, as well as with reporters following a MS2 binding site, which was discovered to emit the weakest reporter expression of all the different sites (figure 3.7).

Background fluorescence in the mCherry channel was much lower than the values measured with the reporter, constituting about 10-15% of the signal (figure 3.8A). This means the differences of levels of fluorescence in the actual assay can be relied upon to derive from the reporter. However, the background in the cerulean channel was higher than or equal to the levels detected in cells containing the reporter (figure 3.8B), suggesting that the results of the experiments using cerulean as a reporter cannot be relied upon not to be an artifact of the system. For this reason, cerulean was ruled out as a reporter in this particular scheme.

## 3.3 Microscopy experiments

One of the main goals of this research is to develop tools that will enable the scientific community to visualize RNA molecules inside cells, using microscopy. In order to verify the ability of the fusions to do this, a Bacterial Artificial Chromosome containing 96 binding sites of MS2 (17) was inserted into a cell containing the plasmid pCP-MS2-mCherry. When these cells are observed under the microscope, as shown in figure 3.9, it is possible to see bright spots inside them, indicating areas where multiple MS2-mCherry fusions are bound to a single RNA molecule.

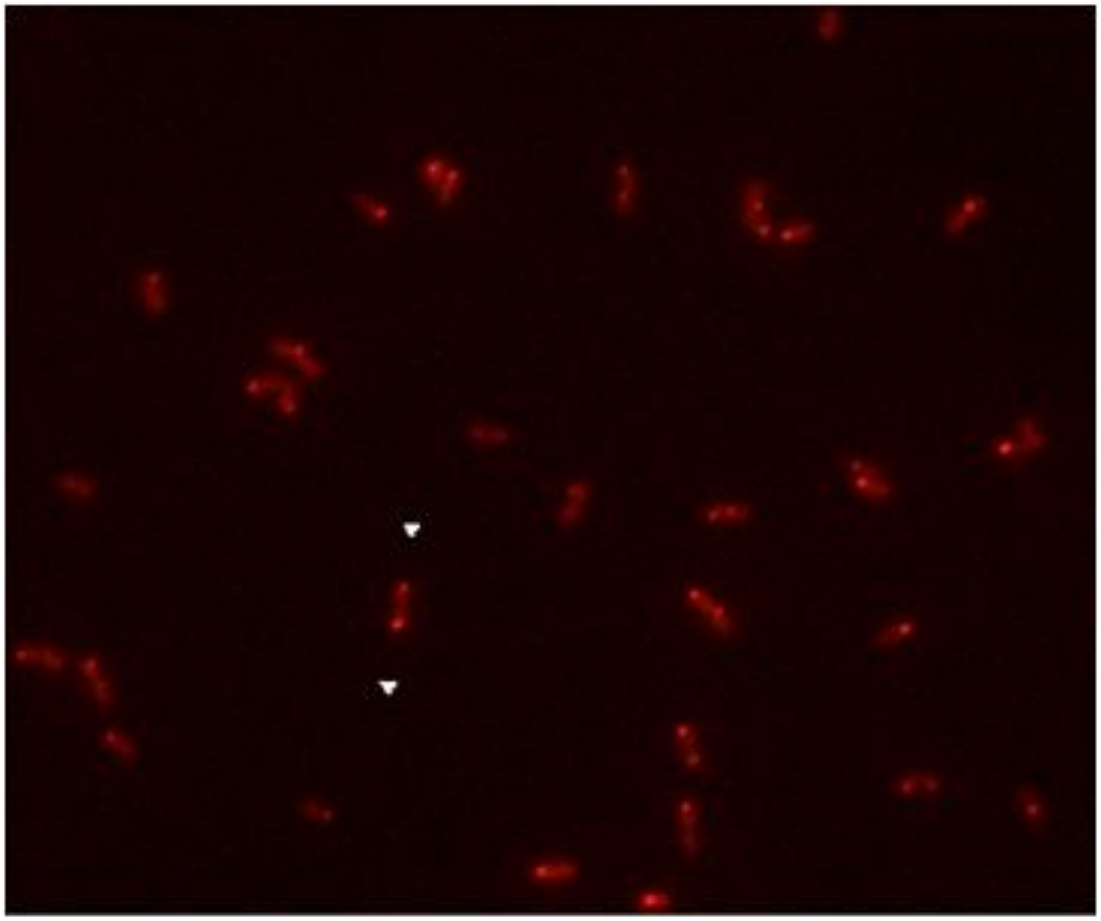


***Figure 3.9***: *Microscopy image of cells containing the fusion protein MS2-mCherry and a cassette of 96 binding sites of MS2. Spots indicate binding regions (i.e., multiple fusions bound to one RNA molecule) inside the cells. The arrows denote a single cell in which there are two such spots.*

# 4. Discussion

MS2-GFP and its counterparts were first used to track RNA fifteen years ago, but this system is still in the process of development. The background fluorescence, increased by the free fusion proteins in the cell, is a major limitation, creating a requirement for up to dozens of binding sites on the target RNA. Recent advances, such as tracking more than one target simultaneously, or observing non-coding RNA, indicate even greater potential for this method, but its set up is far from trivial. Facilitating the assembly of these constructs will make it much more accessible.

## 4.1 Fluorescent functionality of the fusion proteins

Figure 3.2 shows the results of the fluorescence assay, which examined the effect of being fused to a RNA binding protein to a fluorescent protein's original function. Based on this assay, eight fusions were taken to the binding assay, the ones that showed fluorescence above the base line of the control strain. However, it is unclear whether lower levels of fluorescence were the results of incompatible folding of the fusion or simply high background levels of autofluorescence. It is known that autofluorescence often eclipses the expression of fluorescent proteins used as reporters (3; 14). This is demonstrated by comparing the levels of the mDendra2 fusions on the high copy vector and the low copy (figure 3.3 with figure 3.2C). MS2-dendra and Qβ-dendra both appear to rise above the control level when expressed from high copy plasmid, indicating that the copy number of the plasmid limited their expression in the previous test, rather than the structure of the fusion. It is highly probable that other fusions that appeared not to fluoresce when expressed from pCP, such as MS2-cerulean, are capable of giving high fluorescence in other expression systems.

## 4.2. The binding assay

This assay was based on a similar one performed in studies that have explored the relationship between the bacteriophage RNA binding proteins and their operators (32, 25). By inserting a reporter gene that leads to a quantifiable product, this assay gives quantitative data about *in vivo* binding affinities.

4.2.1 Effects of the fusion proteins on the various RBSs

In regards to binding affinities, the PP7 fusions and the Qβ fusions seemed to have almost identical effects on each of the RBSs. This result was surprising as no previous experiment detected such crosstalk between these two coat proteins. A previous study reported that the PP7 coat protein can repress expression from the Qβ binding site to about 20%, while reducing its own to less than 10% (25). These numbers are close enough to the results of this study for us to assume the differences are due to the variance. It also reported a 30% repression of the MS2 binding site by the PP7 and Qβ proteins, another result which was seen in this assay (figure 3.6A) The same study reported no effect whatsoever of the Qβ coat protein on the PP7 operator. One possible explanation for this discrepancy is that the PP7 operator in this research is not the wild type, but a slightly different sequence. A more significant difference between the assays is the reporter gene. While LacZ is an enzyme in need of a substrate, without which there is no color, fluorescent proteins have the advantage of being independent of other factors. Every protein produces only one fluorophore, giving a more linear response of fluorescence versus reporter level. This means that once a LacZ protein is created, it can catalyze the cleavage of multiple substrate molecules. Slight changes in reactions times might lead to high variations in colorimetric measurements.

Based on the fluorescence experiments alone, it is difficult to determine why the MS2 fusions seem to amplify the expression of the reporter. This phenomenon was not previously described in the context of the binding assay. In all the experiments containing the MS2 fusions, the amplification effect was observed regardless of the type of binding site present, including the standard RBS (figures 3.5 and 3.6). As a result, it seems unlikely that the amplification effect is due to the interaction between the MS2 fusions and the binding sites. The mechanism of action in this case is probably related to one of the uniform areas of the gene. It is not certain that the up regulation in this case is even at the posttranscriptional level. It is interesting to note that the first group that used a MS2-GFP fusion in bacteria (18) reported an almost absolute lack of degradation of the target RNA, the fluorescent protein variant named mRFP1 fused to a cassette of 96 MS2 binding sites. Although the two constructs are very different – a single binding site versus 96, 5'UTR versus 3'UTR, an increase in expression via reduction of degradation of the *mCherry* mRNA might possibly

explain this trend. Additional tests, such as RNA immunoprecipitation and Real Time PCR, are required in order to better understand the mechanism which underlies this phenomenon.

Perhaps the major difference between the originally described binding assay of the phage coat proteins and this particular assay is the placement of the binding sites. Whereas the former placed the binding sites so that the AUG codon inside them will replace that of the *LacZ* and subsequent nucleotides will become part of the translated region, the binding sites in this assay were placed in the exact position of the standard RBS, ending eight nucleotides upstream of the *mCherry's* start codon. This design, which did not seem to prevent repression caused by the PP7 and Qβ coat proteins, might be severely flawed when it comes to the MS2. If reconstruction of the mRNA changes the results of this assay, it will mean the system is more confined to it natural limits than we first realized. This is another experiment to be taken into account when examining possible explanations.

### 4.2.2 Strengths of various RBSs

Different binding sites give different levels of reporter, even without the presence of the coat proteins (32). In order to normalize the effects of the coat proteins properly, expression levels were measured in the presence of the high copy number plasmid pUC19, without any of the fusions on it. Unsurprisingly, the normal RBS, used for standard cloning and lacking a secondary structure, gave a level of expression at least twice as high as the phage binding sites (figure 3.7).

## 4.3 Microscopy experiments

The microscopy images of cells with MS2-mCherry and its cognate binding site (figure 3.9) showed their binding as bright spots inside the cell. The appearance of the bright spots is due to the localization of the MS2-mCherry proteins to single mRNA molecules. As a result, this image provides further evidence that the MS2-mCherry protein constructed in this study is fully functional. It is important to note that the lack of images with other fusions and binding sites cassettes is due to the extreme challenge of creating a cassette of multiple repeats. One approach to overcoming this challenge is to increase the repertoire of binding sites, which was the primary goal of this thesis.

## 4.4 Future work

The initial results of the binding assay, which suggested PP7 and Qβ may bind to each other's RNA binding sites with a high affinity, indicated that the binding sites can be used as variants in a cassette for either one of the proteins. If the current common practice makes use of two variants in each cassette (20), it is now possible to use at least three. Additional experiments must be conducted to verify that the binding of the coat proteins to their operators is indeed the mechanism of repression in these cells. This can be done with RNA immunoprecipitation or electrophoretic mobility shift assays. Real Time PCR can be performed in order to ascertain that the reporter protein levels are not the consequences of changes in the mRNA levels, as is suspected in the case of the cells expressing the MS2 fusions. As mentioned above, it is important to rule out the possibility that some or all of its results are caused by the construction of the synthetic system, such as the reporter type or the position of the binding site, rather than by the natural elements involved. If this *in vivo* assay does indeed measure binding affinities, it can be used to discover more high affinity binding sites that will be incorporated into the binding cassettes, facilitating the cloning procedure. By inducing changes to turn it into a high throughput assay, it might prove to be an exceedingly useful tool in studying regulation at the RNA level.

## References


1. Basyuk, E., Galli, T., Mougel, M., Blanchard, J.M., Sitbon, M. and Bertrand, E. *Retroviral Genomic RNAs Are Transported to the Plasma Membrane by Endosomal Vesicles.* Developmental Cell **5**: 161-174 (2003).
2. Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H. and Long, R.M. *Localization of* ASH1 *mRNA Particles in Living Yeast.* Molecular Cell **2**:437-445 (1998).
3. Billinton, N. and Knight, A.W. *Seeing the Wood through the Trees: A review of Techniques for Distinguishing Green Fluorescent Protein from Endogenous Autofluorescence.* Analytical Biochemistry **291**: 175-197 (2001).
4. Boireau, S., Maiuri, P., Basyuk, E., de la Mata, M., Knezevich, A., Pradet-Balade, B., Bäcker, V., Kornblihtt, A., Marcello, A. and Bertrand, E. *The Transcriptional Cycle of HIV-1in Real-Time and Live Cells.* Journal of Cell Biology **179** (2):291-304 (2007).
5. Bollback, J.P. and Huelsenbeck, J.P. *Phylogeny, Genome Evolution, and Host Specificity of Single-Stranded RNA Bacteriophage (Family Leviviridae).* Journal of Molecular Evolution **52**:117-128 (2001).
6. Brodsky, A.S. and Silver, P. A. *Identifying Proteins that Affect mRNA Localization in Living Cells.* Methods **26**: 151-155 (2002).
7. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. and Prasher, D. C. *Green Fluorescent Protein as a Marker for Gene Expression.* Science **263**: 802-805 (1994).
8. Chao, J.A., Patskovsky, Y., Almo, S.C. and Singer, R.H. *Structural Basis for the Coevolution of a Viral RNA-Protein Complex.* Nature Structural and Molecular Biology **15** (1): 103-105 (2008).
9. Chubb, J.R., Trcek, T., Shenoy, S.M. and Singer, R. H. *Transcriptional Pulsing of a Developmental Gene.* Current Biology, **16**: 1018–1025 (2006).
10. Daigle, N. and Ellenberg, J. $\lambda_N$*-GFP: an RNA Reporter System for Live Cell Imaging.* Nature Methods **4** (8): 633-636 (2007).
11. Forrest, K.M. and Gavis, E.R. *Live Imaging of Endogenous RNA Reveals a Diffusion and Entrapment Mechanism for* nanos *mRNA Localization in* Drosophila. Current Biology, **13**: 1159–1168 (2003).

12. Fouts, D. E., True, H. L. and Celander, D.W. *Functional recognition of fragmented operator sites by R17/MS2 coat protein, a translational repressor*. Nucleic Acids Research **25** (22): 4464-4 4473 (1997).
13. Fusco, D., Accornero, N. Lavoie, B, Shenoy, S.M., Blanchard, J.M., Singer, R.H. and Bertrand, E. *Single mRNA Molecules Demonstrate Probabilistic Movements in Living Mammalian Cells.* Current Biology **13**: 161-167 (2003).
14. Garcia, H.G., Lee, H.J., Boedicker, J.Q. and Phillips, R. *Comparison and Calibration of Different Reporters for Quantitative Analysis of Gene Expression.* Biophysical Journal **101**: 535-544 (2011).
15. Gibson, D.G., Smith, H.O., Hutchison, C.A., Venter, J.C. and Merryman, C. *Chemical Synthesis of the Mouse Mitochondrial Genome.* Nature Methods **7** (11): 901-903 (2010).
16. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C, Hutchison, C.A. and Smith, H.O. *Enzymatic Assemblies of DNA Molecules Up to Several Hundred Kilobases.* Nature Methods **6** (5): 343-345 (2009).
17. Golding, I. and Cox, E. C. *RNA Dynamics in Live Escherichia coli Cells.* PNAS **101** (31): 11310-11315 (2004).
18. Golding, I., Paulsson, J., Zawilski, S.M. and Cox, E. C. *Real-Time Kinetics of Gene Activity in Individual Bacteria.* Cell **123**: 1025-1036 (2005).
19. Golmohammadi, R., Fridborg, K., Bundule, M. Valegard, K. and Liljas, L. *The Crystal Structure of Qβ at 3.5Å resolution.* Structure **4**: 543-554 (1996).
20. Hocine, S., Raymond, P., Zenklusen, D., Chao, J.A. and Singer, R.H. *Single-Molecule Analysis of Gene Expression Using Two-Color RNA Labeling in Live Yeast.* Nature Methods **10**(2): 119-121 (2013).
21. Hogg, J.R. and Collins, K. *RNA- Based Affinity Purification Reveals 7SK RNPs with Distinct Composition and Regulation.* RNA **13** (6): 868-880 (2007).
22. Keryer-Bibens, C., Barreau, C. and Osborne, H. B. *Tethering of Proteins to RNA by Bacteriophage Proteins.* Biology of the Cell **100**: 125-138 (2008).
23. Lange, S., Katayama, Y., Schmid, M., Burkacky, O., Bräuchle, C., Lamb, D.C. and Jansen, R.P. *Simultaneous Transport of Different Localized mRNA Species Revealed by Live Cell Imaging.* Traffic **9**: 1256-1267 (2008).
24. Levsky, J. M. and Singer, R.H. *Fluorescence In Situ Hybridization: Past, Present and Future.* Journal of Cell Science **116** (14): 2833-2838 (2003).

25. Lim, F., Downey, T.P. and Peabody, D.S. *Translational Repression and Specific RNA Binding by the Coat Protein of the Pseudomonas Phage PP7.* The Journal of Biological Chemistry **276** (25): 22507-22513 (2001).
26. Lim, F. and Peabody, D.S. *RNA Recognition Site of the PP7 Coat Protein.* Nucleic Acids Research **30** (19): 4138-4144 (2002).
27. Lim, F., Spignola, M. and Peabody, D.S. *Altering the RNA Binding Specificity of a Translational Repressor.* The Journal of Biological Chemistry **269** (12): 9006-9010 (1994).
28. Lim, F., Spignola, M. and Peabody, D.S. *The RNA Binding Site of Bacteriophage Qβ Coat Protein.* The Journal of Biological Chemistry **271** (50): 31839-31845 (1996).
29. Olsthoorn, R.C.L., Garde, G., Dayhuff, T., Atkins, J.F. and Van Duin, J. *Nucleotide Sequence of a Single-Stranded RNA Phage from* Pseudomonas aeruginosa: *Kinship to Coliphages and Conservation of Regulatory RNA Structures.* Virology **206**: 611-625 (1995).
30. Ozawa, T., Natori, Y., Sato, M. and Umezawa Y. *Imaging Dynamics of Endogenous Mitochondrial RNA in Single Living Cells.* Nature Methods **4** (5): 413-419 (2007).
31. Paige, J.S., Wu, K and Jaffrey, S.R. *RNA Mimics of Green Fluorescent Protein.* Science **333**: 642-646 (2011).
32. Peabody, D.S. *Translational Repression by Bacteriophage MS2 Coat Protein Expressed from a Plasmid.* The Journal of Biological Chemistry **265** (10): 5684-5689 (1990).
33. Querido, E. and Chartrand, P. *Using Fluorescent Proteins to Study mRNA Trafficking in Living Cells.* Methods in Cell Biology **85**: 273-292 (2008).
34. Rumnieks, J. and Tars, T. *Crystal Structure of the Bacteriophage Qβ Coat Protein in Complex with the RNA Operator of the Replicase Gene.* Journal of Molecular Biology **426** (5): 1039-1049 (2014).
35. Spingola, M., Lim, F. and Peabody, D.S. *Recognition of Diverse RNAs by a Single Protein Structural Framework.* Archives of Biochemistry and Biophysics **405**: 122-129 (2002).
36. Suter, D.M., Molina, N., Gatfield, D. Schneider, K., Schibler, U. and Naef, F. *Mammalian Genes are Transcribed with Widely Different Bursting Kinetics.* Science **332**: 472-474 (2011).

37. Tyagi, S. *Imaging Intracellular RNA Distribution and Dynamics in Living Cells.* Nature Methods **6** (5): 331-338 (2009).
38. Tyagi, S. and Kramer, F.R. *Molecular Beacons: Probes that Fluoresce Upon Hybridization.* Nature Biotechnology, **14**: 303-308 (1996).
39. Witherell, G. W. and Uhlenbeck, O.C. *Specific RNA Binding by Qβ Coat Protein.* Biochemistry **28**: 71-76 (1989).
40. Young, J.W., Locke, J. C.W., Altinok, A., Rosenfeld, N., Bacarian, T., Swain, P.S., Mjolsness, E. and Elowitz, M. *Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy.* Nature Protocols **7** (1): 80-88 (2012).
41. Yunger, S., Rosenfeld, L., Garini, Y. and Shav-Tal, Y. *Single-Allele Analysis of Transcription Kinetics in Living Mammalian Cells.* Nature Methods, **7**: 631-633 (2010).

**תקציר**

בשנים האחרונות התפיסה הדוגלת במחקר ביולוגי ברמת התא הבודד או המולקולה הבודדת צוברת תאוצה. גישה זו מאפשרת להשיג מידע רב ברמות הנמוכות ביותר הקיימות בתחום הביולוגי. על מנת להשיג מידע זה, יש צורך בשיטות ניסויות מתקדמות. אחד התחומים בהם יש עניין רב להגיע לרזולוציה של מולקולות בודדות הינו חקר ה- RNA. השיטות שהיו מקובלות במשך מספר עשורים לחקר ה-RNA עשו שימוש בחומרים שקיבעו את התאים, והתוצאות שהתקבלו היו תמונות סטטיות של מצב התא בעת הוספת החומר. עם הצטברות המידע על מורכבות התהליכים התוך תאיים, התעוררה ההבנה שיש צורך בשיטת ניסוי שלא תפריע לתהליכים המתרחשים התא, ותיתן מימד נוסף של זמן לתוצאות. אחת השיטות לעשות זאת, שהתפתחה בעקבות השימוש בחלבונים פלורוסנטים כגנים מדווחים, הינה איחוי של חלבון פלואורסנטי לחלבון קושר RNA. כדי שהאחרון ייקשר, המולקולה שאחריה רוצים לעקוב חייבת להכיל בתוכה רצף שאותו מזהה החלבון ואליו הוא נקשר. על מנת לחקור מספר מולקולות בו זמנית יש צורך בשימוש מספר קומבינציות של חלבונים מאוחים, שיזהו רצפים שונים שימוקמו על מולקולות המטרה. לצורך זה יצרנו ספרייה של חלבונים קושרי RNA אשר מאוחים לחלבונים פלואורסנטים.

החלבונים הנפוצים לשימוש בשיטה זו הינם חלבונים אשר מקורם בפאג׳ים, שם הם מהווים את המעטפת החלבונית של החומר הגנטי, שבמקרה הזה הינו RNA חד גדילי. חלבוני המעטפת נקשרים למבנה מסוים ב-RNA שיוצר רצף של 19 נוקליאוטידים. רצף זה ממוקם בגנום הויראלי באזור תחילת התרגום של הגן המקודד לרפליקאז. כל עוד קשור אליו חלבון המעטפת, אין אפשרות לריבוזום להתחבר לאזור ולתרגם את הגן לחלבון. חוקרים גילו כי הקישור בין חלבון המעטפת ל-RNA תלוי במבנה השניוני של ה-RNA יותר מאשר ברצף שלו. החלפת נוקליאוטידים בצורה שלא תשנה את המבנה השניוני לא תמנע קישור של החלבון.

מערכת ההדמיה של ה- RNA בנויה כך שיהיו בתא יותר חלבונים מאוחים מאשר מולקולות מטרה. התוצאה הינה רמה גבוהה של פלואורסנציה בתוך התא. על מנת להבדיל בין חלבונים קשורים לבין אלה שלא קשורים, יש צורך להגביר את הסיגנל על מולקולת היעד. ניתן לעשות זאת ע״י החדרת מספר רב של אתרים למולקולת היעד, שירכזו אליהם מספר רב של חלבונים מאוחים אשר יצרו סיגנל חזק יחסית לרקע. הקושי העיקרי בהרכבת המערכת הינו ביצירת רצף כזה. השיטות הקיימות לסינתזת דנ״א לא מצליחות להתמודד עם האתגר שבהחדרת מספר גדול של רצפים חוזרים. פה נכנס היתרון של אופן הקישור של החלבון למבנה מסוים, ולא לרצף מסוים. היכולת להחליף נוקליאוטידים מבלי לפגוע באפיניות של החלבון ל-RNA מאפשרת להכניס שינויים ברצפים אשר יפחיתו מהחזרות ויקלו על תהליך השיבוט.

בשלב הראשון של מחקר זה התמקדנו באפיון החלבונים המאוחים מבחינת יכולת הפלואורסנציה שלהם. קומבינציות של ארבעה חלבונים שמקורם בפאג׳ים: MS2, PP7, Qβ ו-λN, ושלושה חלבונים פלואורסנטים: mCherry, mDendra2 ו-cerulean הוחדרו לתוך פלסמידים בעלי origin of replication הנותן מספר עותקים נמוך בתא, ולאחר השיבוט עברו בדיקה למדידת יכולתם

לפלוט פלואורסנציה באורכי הגל המתאימים כחלבונים מאוחים. כביקורת נמדדה הפלואורסנציה של חיידקים אשר לא הכילו את הפלסמיד עם הגן המאוחה. ניסויים אלה הראו כי כל החלבונים אשר בהם החלבון הפלואורסנטי היה mCherry, אשר פעיל באורכי גל בתחום האדום (ערעור ב-560 ננומטר ופליטה ב-612), רמת הפלואורסנציה הייתה מעל לרמה של קו הביקורת. מתוך החלבונים אשר הכילו את ה-cerulean (חלבון אשר פעיל בתחום האור הכחול), רק PP7 ו-Qβ נתנו רמה גבוהה מהביקורת. באלה בהן שימש mDendra2 כחלבון הפלואורסנטי, רק האיחוי עם PP7 הביא לפליטת פלואורסנציה מעל לרמת הביקורת. יחד עם זאת, כאשר החלבונים הללו הועברו לפלסמיד בעל מספר עותקים רב בתא, התקבלה רמת פלואורסנציה גבוהה מכל החלבונים. תוצאה זו מעידה כי אין בעיה ביכולת החלבונים לפלוט אור פלואורסנטי בעקבות ערעור, אלא שרמת ביטויים במערכת הראשונית נמוכה מרמת רעש הרקע הנמדד באותם אורכי גל אשר בהם ה- mDendra2 פעיל (ערעור ב-495 ננומטר, פליטה ב-535).

בעקבות תוצאות ניסויים אלה הוחלט להפסיק את השימוש בחלבון λN ולהתמקד ב-MS2, PP7 ו- Qβ.

בשלב השני של המחקר נבדק הקישור בין החלבונים לאתרי ההכרה שלהם, באמצעות שיטה מתחילת שנות ה-90 המנצלת את תפקידם השני של חלבוני המעטפת כרפרסורים לתרגום *הרפליקאז*. על ידי החדרת האתר באזור תחילת התרגום של גן מדווח, חוקרים אספו נתונים לגבי האפיניות של חלבוני המעטפת השונים לאתרי הקישור שלהם, אתרי הקישור של החלבונים האחרים ואתרים המכילים מוטציות שונות. במחקר זה נעשה שימוש בגנים פלואורסנטים בתור הגנים המדווחים. האתרים שנבחרו לבדיקה היו האתר הטבעי של Qβ ועוד שני אתרים, אחד של MS2 והשני של PP7, אשר אינם האתרים המקוריים שלהם אך כבר שימשו בהצלחה כאתרי הקישור במערכות הדמייה דומות של RNA.

החלבונים המאוחים הוחדרו לפלסמיד אחד והגן המדווח עם אתרי הקישור השונים לפלסמיד אחר, וכל שניים שובטו לחיידק. רמת הביטוי של הגן הפלואורסנטי נמדדה בנוכחות של אחד מהחלבונים, וכן בנוכחות פלסמיד ריק אשר נתן מידע על רמת הביטוי המקסימלית אליה ניתן היה להגיע במערכת זו. התוצאות הראו בבירור כי נוכחות החלבונים המאוחים אכן משפיעה על רמת הביטוי של הגן אשר תרגומו היה תלוי באתר הקישור ל- RNA, אך ההשפעות תאמו את המתואר בספרות באופן חלקי בלבד. החלבונים של Qβ ו-PP7 אכן הורידו משמעותית את רמת הביטוי מפלסמידים אשר הכילו את הרצפים של כל אחד מהם בהתאמה, אך הורידו ברמה דומה את הביטוי מהרצפים אחד של השני. שניהם אף הורידו ב-30% את רמת הביטוי שהתקבל מרצף ההכרה של MS2, תוצאה שגם תוארה בעבר בספרות. החלבונים המאוחים עם MS2, לעומת זאת, לא הורידו את רמת הביטוי באף אחד מהמקרים, אלא רק העלו אותם. תוצאה זו מאוד הפתיעה אותנו, ובעקבותיה החלטנו לבדוק את השפעת החלבונים גם על רצף סטנדרטי לקישור לריבוזום, אשר נועד לשימוש במערכות ביטוי גנים. כאשר הוא הוחדר לאזור תחילת התרגום של הגן המדווח, התברר כי הוא אכן לא מושפע מנוכחות Qβ ו-PP7 , תוצאה המחזקת את ההנחה כי השפעתם נובעת מקישורם למבנים השניוניים הייחודים של הרצפים שלהם. מנגד, התברר כי

MS2 כן מגביר את רמת הביטוי גם מרצף זה ואפילו באופן חזק יותר מאשר חלק מהאתרים הויראליים (תגבור פי 2 באתר הקישור לריבוזום ובין 1.5 ל-2 באתרים הויראליים). תוצאה זו מרמזת כי השפעת MS2 על ביטוי הגן הפלואורסנטי במערכת זו אינה תלויה באתר תחילת התרגום, אך מנגנון הפעולה של הגברת הביטוי איננו ברור בשלב זה.

השימוש בחלבונים קושרי RNA שמקורם בפאג׳ים נועד למנוע את קישורם לרצף בתא שאינו הרצף המיועד להם. התגלית שהחלבון MS2 מגביר את ביטוי הגן המדווח בדרך שאינה ברורה מעלה את השאלה האם ההנחה הזו בהכרח נכונה תמיד, ושמא המערכת הסינטטית שנועדה לאפשר מעקב אחרי ה- RNA תוך כדי תהליך חיים טבעי של התא לא תמיד מהימנה. לצורך בחינת שאלות אלה, יש צורך בביצוע מגוון ניסויים, שיאפשרו לקבוע האם מערכת הניסוי למדידת אפיניות הקישור איננה מתאימה, או שמא החלבון המאוחה אכן משפיע על ביטוי הגן, ואם כן, מהו המנגנון בו הוא עושה זאת. שאלות אלו עשויות להוביל לתפיסות חדשות לגבי השימוש בביולוגיה סינטטית לחקר תהליכים טבעיים.

המחקר נעשה בהנחיית פרופ׳ רועי עמית בפקולטה להנדסת ביוטכנולוגיה ומזון.

אני מודה לטכניון על התמיכה הכספית הנדיבה בהשתלמותי.

# תכנון ואפיון מערכת הדמייה של RNA בתאים חיים המבוססת על חלבונים פלואורסנטיים

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת התואר מגיסטר למדעים בהנדסת ביוטכנולוגיה ומזון

איה פרידמן

הוגש לסנט הטכניון- מכון טכנולוגי לישראל

ניסן תשע״ד חיפה אפריל 2014