© 2016 by Jichuan Zhang. All rights reserved.

DEVELOPING HIGH-RESOLUTION METHODS TO STUDY DNA AND RNA DYNAMICS

BY

JICHUAN ZHANG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Materials Science and Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Jianjun Cheng, Chair
Professor Taekjip Ha, Director of Research, Johns Hopkins University
Assistant Professor Kristopher Kilian
Assistant Professor Thomas E. Kuhlman
Assistant Professor Cecilia Leal

ABSTRACT

Nucleic acids, including DNA and RNA, are one of the most important biomacromolecules inside the cell. DNA stores genetic information, while RNA has more versatile roles including conveying and deciphering genetic information, catalyzing biological reactions, conducting post-transcriptional regulation, and even storing genetic information. To understand the functioning principles of cells and living organisms, it is very important to study DNA- and RNA-involved biochemical reactions as well as the molecular mechanisms behind them, which is quite challenging due to the small size and high metabolism rate of cells. Thanks to the technology advancement in the past several decades, we are able to develop research methods to achieve high-rate, high-resolution and high-throughput study on DNA and RNA. Here we developed several fluorescence-based high-resolution assays to study DNA and RNA dynamics *in vitro* and *in vivo*.

First, we established a single molecule FRET assay to study homodimeric single-stranded binding protein (*Thermus thermophilus* SSB) specifically binding and protecting single-stranded DNA during DNA metabolism. With the help of specifically designed DNA constructs and total internal reflection microscopy, we discovered that homodimeric SSB showed similar one-dimensional diffusion and salt-dependent binding mode transition behavior which was confirmed for intensively studied homotetrameric SSB, suggesting that those behaviors might be universal among SSB homologues from different organisms.

Second, we developed a live cell RNA labeling and imaging method based on an aptamer-fluorogen system called “Spinach”, which contains an RNA sequence that binds a fluorogen DFHBI and induces its fluorescence. We constructed a Spinach array with tandem Spinach repeats and it greatly enhanced the cellular fluorescence signal, and we could easily visualize mRNAs in the cell. We further characterized the Spinach RNA imaging method and found that either single Spinach or Spinach arrays do not affect RNA transcription, protein translation or RNA degradation. Therefore we proposed that aptamer/fluorogen imaging and aptamer array construction could be a generalizable strategy for high performance and low perturbation live cell RNA imaging.

Finally, we expanded the research to gene expression regulation. Our research target, sgrS, is a bacterial small RNA that regulates several target genes at post-transcriptional level in response to sugar-phosphate stress. It is known that sgrS anneals target transcripts via basepairing interaction with the guide of Hfq protein. Nevertheless, how individual nucleotides within sgrS sequence contribute to the regulation process is not clear. Here we utilized the recently developed Sort-Seq method combining fluorescence-activated cell sorting (FACS) and high-throughput DNA sequencing to study sequence-dependent sgrS regulation on its primary target, ptsG. We constructed a target-reporter system with ptsG 5' UTR, which is responsible for sgrS annealing, fused to GFP, whose expression level is indicated by fluorescence level and is capable of being regulated by sgrS. By introducing an sgrS random mutation library, cells show various fluorescence levels due to diverse regulation capabilities of sgrS mutants. By sorting the cells into different groups based on its fluorescence signal followed by extracting the sgrS mutation distribution among sorted groups, we could find out which mutations totally abolished sgrS function. Results showed that 2 nucleotides involved in ptsG/sgrS annealing, G176 and G178, are extremely significant; other 24 nucleotides are equally significant, but they are involved in sgrS binding by Hfq. The study suggested important nucleotides within the sgrS/ptsG annealing region, and emphasized the significance of Hfq in maintaining sgrS function.

To my dear and loving family

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Taekjip Ha sincerely for his guidance and support throughout my entire graduate study. He is a fun person to work with and his creativity, insight and enthusiasm for scientific research continue to humble and inspire me.

I would also like to thank Professors Jianjun Cheng, Kristopher Kilian, Thomas E. Kuhlman and Cecilia Leal for serving on my thesis committee, and providing constructive criticism and thoughtful suggestions for my research, especially for Professor Thomas E. Kuhlman, who is also my research collaborator, and one of the humblest, most patient and most supportive scientists I have ever met.

I would like to thank all the current and former Ha Lab members for their support and consideration. Most of the people are fun to work and talk with, and many of them become my good friends and companions. I always enjoy discussing scientific questions as well as life experience with them, and I am pretty sure my life would have been much more boring and undoubtedly less rewarding without them. Here I also would like to particularly express my deep gratitude to three former postdoctoral fellows from Ha Lab, Ruobo Zhou, Jingyi Fei and Benjamin J. Leslie, who were my close co-workers as well as wonderful mentors during my graduate study. Ruobo was my first mentor and I have seldom seen people as scientifically critical, rigorous and logical, and as mentally firm, tough while empathetic as he is; he was always helpful, supportive and shared with me so much life experience. Jingyi was one of the most energetic, passionate and thoughtful scientist I have seen; she guided me to the RNA biology field and I always enjoyed having discussions with her. Ben was one of the most knowledgeable and helpful person I have ever met; I learned so much about molecular cell biology from him and he was always willing to help me with the tedious trouble-shooting and system optimization; he was really a fun person to work with and his wife was one of the most outgoing and interesting people and his daughter was one of the most adorable babies I have ever known. I genuinely cherish the working experience with all of them during my graduate study.

Olivia Yang, Anustup Poddar, Digvijay Singh, Vasudha Aggarwal, Boyang Hua, Sangwoo Park, Chang-Ting Lin, Jaba Mitra, Myung Hyun Jo, Byung Choul Kim, Jongchan Lee, Aakash Basu,

Matthew Poyton, Yang Liu, Ye Ma, Dmitriy Bobrovnikov and other current Ha Lab members keep inspiring me with their wonderful research while delighting me all the time as friendly and considerate colleagues. I have been working with many of them and will sincerely cherish the experience working together.

Qiucen Zhang, Xuefeng Wang, Isaac Li, Thuy Ngo, Xinghua Shi, Salman Syed, Farhan Chowdhury, Kyu Young Han, Sultan Doganay, Seongjin Park, Mehdi Roeinpeikar, Jungmin Joo, Yuji Ishitsuka, Ankur Jain, Gwangrog Lee, Hajin Kim, Reza Vafabakhsh, Matt Comstock, Jeehae Park, Kyung Suk Lee, Jaya Yodh, Prakrit Jena, Sinan Arslan, Liz Xu, Kaushik Ragunathan, Ankur Jain, Christopher Thibodeaux, Michael Brenner and other former Ha Lab members were truly great colleagues. I have never seen such a group of intelligent, knowledgeable, insightful and passionate people in my life and I got tremendous amount of advice and mentoring from them which I consider to be priceless.

I would like to particularly mention Professors Sua Myong and Jie Xiao, both of whom are wonderful scientists as well as great people, to appreciate their help and concerns on my graduate research and life. I have never seen more supportive and loving groups than their research labs and I want to thank all the current and former lab members from the two labs, including Helen, Peggy, Cong, Younghoon, Hye Ran, Ting, Ram, Alex, Jim, Amir, Chris, Kelsey, Carla, Ryan, Xiaoli, Xinxing, Jason and other members. I would also like to thank Subrata and Andrew from Professor Sarah Woodson Lab.

Very few scientists can be successful without the collaboration with other scientists. I would like to thank Professors Cari Vanderpool, Zan Schulten, Tsutomu Mikawa, James Keck, Thomas E. Kuhlman and Jian Ma and their students Piyush, Max, Shafi, Jin, Katarzyna, Hneil, Cac and Yang for their great input in fulfilling the collaboration. I also would like to thank Professor Erel Levine and his student Neil for useful experimental reagents and helpful scientific suggestions.

For the past six years I am thankful that I have a lot of friends who always support me and try their best to understand me. My life would have been much harder without them as my strong backing. Here I would like to thank Cong Xu, Zheyin Zhao, Zhi Qi, Yang Li, Yang Zhang, Hui Xu, Juan Guan, Yong Wang, Bo Wang, Jingyi Fei, Ruobo Zhou, Shuomeng Guang, Jiu Xu, Zheng Liu, Yi Zhang, Xinlei Wang, Ruiqing Lu, Yin Liu, Ruijuan Xu, Wenjiao Wang, Rui He,

Youzhi Wu, Zhiyi Wang, Wei Dai, Fan Ye, Hanlin Ouyang, Jiong Zhang, Jian Huang, Yuhe Liu, Yifei Yang, Sheng Wang, Yue Ding, Tao Sun, Hong Lin, Qingzhan Zhang and many other friends. Some of them are my labmates and collaborators while some of them understand little about my research, but I would never forget their encouragement and support.

I would like to specially thank Elizabeth “Betty” Ujhelyi who was the technician at Ha Lab, who was so dedicated to lab organization and maintenance and doing so well at making everyone’s research much easier. Betty is also a very nice and responsive person and I couldn’t count how many times I bothered her. Also I have many thanks to Julie, Shawn, Michelle, Sheryl and Sandra, who are current or past secretaries at the Lab, Department Physics or Department of Materials Science and Engineering.

In addition, I would like to express my sadness for late Professors Robert Clegg and Klaus Schulten, who passed away during my graduate study. Both of them are great biophysicsts and have made great contributions to the biophysics field. I had the privilege to attend the courses instructed by the two professors, and the lectures gave me considerable influence. I would remember the great time and treasure the wonderful experience.

Last but certainly the biggest, I sincerely thank my parents and all my family members for loving me, believing me, supporting me and tolerating me. You are the eternal sunshine of my life and I will love you all until the end of the world.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Single-molecule FRET

1.1.1 FRET

Förster (or fluorescence) Resonance Energy Transfer (FRET) is the phenomenon where an excited fluorescent molecule (the donor) transfers its energy to another fluorescent molecule (the acceptor) in close proximity through a dipole-dipole interaction[1]. Due to the non-radiative energy transfer, the acceptor could be excited and emit fluorescence. The extent of energy transfer, usually described by FRET efficiency or *E*, can be given by the formula below and is usually determined by the distance, the spectral overlap and the relative dipole-dipole orientation between the donor and the acceptor molecule.

$$E = \frac{1}{1 + (\frac{R}{R0})^6}$$

In the formula, R indicates the distance between donor and acceptor, R_0 is the Förster distance dependent on donor-acceptor pair and is the distance at which 50% energy transfer takes place. For a typical FRET pair Cy3 and Cy5, R_0 is ~60 Å, and the FRET efficiency shows significant change as a function of Cy3-Cy5 distance within 4 - 7 nm[2]. Since detectable energy transfer for a donor-acceptor pair usually takes place within nanometer range, which is the length scale for a lot of biomolecule interaction or intramolecular conformation change to happen, by labeling the biomolecules of interest with FRET pairs people are able to study inter- and intra-molecular dynamics by measuring the donor and acceptor fluorescence as well as calculating the FRET efficiency.

1.1.2 Single-molecule fluorescence spectroscopy and single-molecule FRET

Despite its usefulness in biological study, people could only obtain limited information by conducting traditional ensemble FRET measurements. Overwhelmed by the big population of

molecules, it is very difficult to observe biomolecular conformational changes as molecules are not synchronized and only averaged behavior could be observed; moreover, short-lived or rare events taking place in the process of conformational change are not easy to be detected. To overcome the obstacle, people combined FRET with single molecule techniques[2,3]. Single molecule observation allows real-time tracking and manipulation on individual molecules and therefore circumvents averaging effect, nor is the molecular synchronization required anymore. In addition, single molecule study provides rich information on understanding molecular heterogeneity, which has been well appreciated in the past decade.

Compared to other single molecule techniques, including single molecule tracking, single molecule force spectroscopy, *etc.*, single molecule FRET has its particular advantages in studying molecular dynamics and interaction[2]. First, single molecule FRET is very useful for measuring relative movements between molecules at an extremely small length scale (usually 4 - 7 nm) while maintaining a very high precision, due to the principle of the energy transfer. Second, it allows detection of rare and short-lived events which will be undoubtedly overwhelmed by the background signal and noise in ensemble measurement. Moreover, single molecule FRET is not as vulnerable to system noise sometimes as other single molecule detection methods are, because what people ultimately measure is the ratio between donor and acceptor fluorescence signal, instead of direct measurement result; in this case some sources of system errors such as excitation light source fluctuation, imaging stage drift are less perturbative. Until now, single molecule FRET has become one of the very useful ways to thoroughly study intramolecular conformation change, molecular dynamics and intermolecular interaction.

1.2 Single-stranded DNA binding protein*

Single-stranded DNA binding (SSB) proteins are ubiquitous and found in prokaryotic and eukaryotic cells, mitochondria, phages and viruses [4-7]. SSB proteins bind specifically to single-stranded (ss) DNA with high affinity in a sequence-independent manner, protect ssDNA from nucleolytic digestion and prevent intrastrand pairing (i.e., hairpin formation), in order to

---

*Chapter 1.2 was excerpted from a published work:

Zhang, J. et al. "Single molecule analysis of *Thermus thermophilus* SSB protein dynamics on single-stranded DNA" Nucl. Acids. Res. doi:10.1093/nar/gkt1316 (2013)

keep ssDNA in a suitable conformation for the action of enzymes involved in DNA replication, repair and recombination[8].

SSB proteins are characterized by the presence of structurally conserved oligosaccharide/oligo-nucleotide binding (OB)-fold domains [9-11] for ssDNA binding. *Escherichia coli* (*Eco*) SSB proteins have a single OB fold per polypeptide and function as homotetramers. In contrast, replication protein A (RPA), the major SSB proteins in eukaryotes, is a heterotrimeric protein complex composed of three distinct subunits and contains six OB.

In addition to ssDNA binding, SSB proteins control the accessibility of ssDNA and can physically interact with a variety of cellular genome maintenance proteins, including nucleases, helicases, polymerases, DNA damage signaling and strand-exchange proteins [6-8], to stimulate their activities. For instance, *Eco*SSB is able to interact directly with at least 14 other proteins, including DNA Polymerase II, III and V, primase, RecQ, RecO, RecJ, RecG, PriA, PriB, Exonuclease I and IX, Uracil DNA Glycosylase and phage N4 RNA polymerase [8], bringing them to their sites of function.

1.3 RNA imaging*

Ribonucleic acids, or RNAs, are biopolymers composed of four types of ribonucleotides, A, U, G and C. RNAs play diverse functions in living cells, including delivering genetic information, catalyzing chemical reactions and regulating gene expression at multiple levels[12,13]. Recent genome-wide analysis has suggested that inhomogeneous RNA localization within different cellular compartments might be more prevalent than previously appreciated[14,15], resulting in highly localized spatio-temporal modulations in gene expression levels within those subcellular compartments. Compared to biochemical approaches, such as northern blot, quantitative reverse transcription PCR (qPCR), high-throughput RNA sequencing, *etc*., direct visualization of RNAs by fluorescence imaging allows spatial and temporal RNA tracking and is capable of correlating transcription, localization, translation and degradation of RNAs and simultaneously revealing

---

*Chapter 3 was excerpted from a published work:

Zhang, J. et al. “Tandem Spinach array for RNA imaging in living bacterial cells” Sci. Rep. doi:10.1038/srep17295 (2015)

cell-to-cell heterogeneity[16-18]. Nevertheless, unlike imaging proteins, there are not many tools for imaging RNAs. Fluorescence *in situ* hybridization (FISH) utilizes fluorescent dye-conjugated oligonucleotides (fluorescent probes) complementary to target RNAs to directly label RNA molecules and has been widely applied to accurately quantify the expression level and to localize distribution of mRNAs in biological samples[19,20]. Introduction of the fluorescence probes usually requires permeabilization of the fixed cell[19]. Alternative delivery methods compatible with live cell imaging include microinjection[21,22], electroporation[14], or transfection using polycationic molecules such as liposomes and dendrimers[23], and membrane permeabilization via cell-penetrating peptides[24] and streptolysin O[25,26]. However, these methods sometimes lead to problems such as cell damage, inhomogeneous probe delivery and inefficient probe annealing to target RNAs[23].

The second type of RNA imaging utilizes indirect labeling and employs fluorescent fusion proteins and specific protein-RNA interactions, such as the RNA bacteriophage MS2 coat protein system[27-29], the PP7 bacteriophage system[30,31], the bacteriophage λ N coat protein system[32], *etc.[33,34]*. In those strategies, RNAs of interest are tagged with a "cognate" RNA sequence, often in a tandem array, recognized and bound by interacting proteins with fluorescent protein fusion. There are several potential limitations associated with this type of methods: (1) usually the overexpressed unbound proteins generate high fluorescent background[35], unless specific measures are taken to reduce the background[33,34,36]; (2) the resulting large ribonucleoprotein complex has been reported to affect the RNA endogenous degradation in some cases[37,38], and may potentially affect trafficking and localization[22]; (3) the conditions required for maturation of the fluorescent proteins prohibit the application of these approaches to certain biological systems, such as anaerobic species[39-41].

More direct ways for RNA labeling in living cells mostly use RNA aptamers that can bind small ligands (or "fluorogens") and activate their fluorescence[42-44]. Aptamers are generally short RNA sequences that can fold into specific tertiary structure and thus bind small molecules called ligands. If the fluorescence of the ligand shows great enhancement upon binding to the aptamer, the binding complex might have the potential to be a labeling reagent for cellular RNA imaging. Due to the much smaller size of ligand (or fluorogen) compared to regular proteins, the

aptamer/fluorogen imaging method might have potential advantages over the RNA binding protein-fluorescent protein labeling strategy.

1.4 Bacterial small non-coding RNA

Small regulatory RNAs (sRNAs) are universal post-transcriptional gene expression regulators in all domains of life. In bacteria, sRNAs are generally 50 - 300 nt non-coding RNAs that play diverse and critical roles in bacterial metabolism and stress response[45,46], pathogen virulence and infection[47], and quorum sensing[48,49]. In a typical case of negative regulation of mRNA by sRNA, the sRNA anneals to its target mRNA through basepairing interaction[46,50,51]. Sometimes the association of the sRNA occludes the ribosome binding site (RBS) of target mRNA and thus prevent translation initiation, resulting in target repression; in some other cases, sRNA annealing influences the stability of target mRNA and promotes its decay by bacterial RNA degradation machinery. In many cases, bacterial sRNA binds to Hfq, a hexameric RNA chaperone, to function[52]. Hfq binding not only enhances the stability of sRNAs *in vivo*, but also promotes interaction between sRNA and its target mRNAs.

In the past several decades, people have discovered hundreds of sRNAs in bacteria. For example, in *Escherichia coli*, more than 100 sRNAs have been identified, many of which play central roles in *E. coli* stress response. Examples include dsrA involved in osmotic and acid stress response[53], oxyS involved in oxidative stress response[54], and sgrS involved in sugar-phosphate stress response[50].

1.5 Flow cytometry and fluorescence-activated cell sorting

Flow cytometry is one of the most common and useful way to study single cells. The instrumentation which conducts flow cytometry, usually called flow cytometer, is made up of three main parts: the fluidics system, the optics system, and the electronics system. When the liquid sample containing cell or particle suspension is sent into the flow cytometer, the fluidics system will generate an ultrathin flow which helps align the cells or particles one by one. The flow will pass a nozzle and become discrete droplets, some of which include a single cell. The droplets will flow through the optics system and illuminated by a laser, and the incorporated cell will scatter light or be excited to emit fluorescence. The signal of the scattered light or the

fluorescence will be collected and a signal conversion will be conducted by the electronics system, and diverse physical properties (size, granularity, *etc*.) and biological properties, usually conveyed by various fluorescence markers, could be acquired. In addition to single cell analysis, the electronics system could also match the obtained information with the initial setting in the program, and charge the droplets with accordingly; the electrical properties are used to decide the future flowing direction of the droplets, and to separate droplets fulfilling different selection criteria to diverse collecting tubes. The great advantages of using flow cytometry for single cell analysis is that advanced flow cytometry can simultaneously detect tens of physical or fluorescence properties, allowing subtle and comprehensive analysis of the sample. Fluorescence-activated cell sorting, or FACS, is the widely appreciated method to identify and sort one or more populations of cells from a collection of mixed cells, according to the fluorescence signal of the cell. The sorted cells can be collected and used for further study and manipulation. Since fluorescence labeling can be used to qualify or quantify cell viability, gene expression, protein marking, *etc*., FACS allows a quick and convenient characterization of cell activities. More powerfully, the sophisticated fluidics system of flow cytometer makes it very easy to pass tens of thousands of droplets through the nozzle, and thus enables very high-throughput analysis of the sample.

CHAPTER 2

SINGLE MOLECULE STUDY ON *Thermus thermophilus* SINGLE-STRANDED BINDING PROTEIN*

2.1 Introduction

Single-stranded DNA binding (SSB) proteins are ubiquitously extant in all kinds of organisms from virus to bacteria to eukaryotes. It specifically binds and protects single-stranded (ss) DNA with high affinity via oligosaccharide/oligonucleotide binding (OB)-fold domains [9-11]. The majority of bacterial SSB proteins have a single OB fold per polypeptide and function as homotetramers, including the best studied example, *Escherichia coli* (*Eco*) SSB [4,55,56]. Eukaryotic human mitochondrial SSB, like *Eco*SSB, also encodes a single OB fold per monomer and functions as stable homotetramers [57]. In contrast, replication protein A (RPA), the major SSB proteins in eukaryotes, is a heterotrimeric protein complex composed of three distinct subunits and contains six OB folds per heterotrimer [5,58]. Interestingly, SSB proteins from the Thermus-Deinococcus genera of bacteria were recently identified as stable homodimers in solution with two OB folds per monomer, contrary to the homotetrameric form so far found in bacteria [59-64]. *Thermus thermophilus* (*Tth*) SSB is a representative homodimeric SSB in the Thermus-Deinococcus group. Structural analysis of homodimeric SSBs has indicated that although homodimeric SSBs share a similar tertiary arrangement within each OB fold as is seen in homotetrameric SSBs, the quaternary arrangement of the four OB-folds in the protein's active form is considerably different from that in homotetrameric SSBs [61,63,64]. *Escherichia coli* maintains 200-3,000 *Eco*SSB tetramers per cell and do not increase SSB levels significantly in response to DNA damaging conditions, whereas in *Deinococcus radiodurans* the 20,000 *Dra*SSB dimers per cell increase to 56,000 dimers in response to ionizing radiation, indicating that the homodimeric SSBs found in extremeophiles may have an expanded role in DNA metabolism [61].

---

*Chapter 2 was excerpted from a published work:

Zhang, J. et al. "Single molecule analysis of *Thermus thermophilus* SSB protein dynamics on single-stranded DNA" Nucl. Acids. Res. doi:10.1093/nar/gkt1316 (2013)

Using *Eco*SSB as the prototypical SSB protein, we previously showed that *Eco*SSB can diffuse (or slide) on ssDNA spontaneously with an estimated diffusion coefficient of 300 nucleotide^2/s at 37 ℃, and that diffusional migration of *Eco*SSB transiently melts DNA hairpin structures and stimulates RecA filament elongation [65]. We also showed that *Eco*SSB interacting proteins slow *Eco*SSB diffusion on ssDNA by physically interacting with the last 8-10 amino acids within the conserved SSB C-terminal tail, raising the possibility that SSB acts as a mobile platform on ssDNA for the replication, repair and recombination machinery [66]. However, it is not clear whether the diffusion activity observed for homotetrameric SSBs is general for other types of SSBs.

SSB binds ssDNA in different binding modes characterized by the length of ssDNA that it contacts and the number of ssDNA-binding domains involved [4,58,67,68]. For *Eco*SSB, the relative stabilities of these binding modes largely depend on salt concentration and protein binding density: 1) At low monovalent salt concentrations (< 20mM NaCl) and high protein to DNA ratios, an *Eco*SSB tetramer occlude ~35 nucleotides (nt) with high inter-tetramer cooperativity using only two out of four OB-folds (termed the $(SSB)_{35}$ mode); 2) At high salt concentrations (≥ 200mM NaCl), an *Eco*SSB tetramer occlude ~65 nt with low cooperativity using all four OB-folds (termed the $(SSB)_{65}$ mode) [4,67]. Furthermore, we previously observed direct transitions between the $(SSB)_{35}$ and $(SSB)_{65}$ binding modes in the presence of 10-100 nM *Eco*SSB and at low or intermediate salt concentrations (≤200 mM NaCl), and that 70-nt ssDNA with two *Eco*SSB tetramers bound can undergo dynamic structural rearrangement between two conformations (termed $(SSB)_{35}$and $(SSB)_{35b}$ modes) while two *Eco*SSB tetramers remain bound to the 70-ntssDNA [69]. In contrast, the occluded site size of a homodimeric SSB from the Thermus-Deinococcus group shows much more reduced dependence on salt concentration than that of an *Eco*SSB tetramer [59,67,70-76]. For example, the occluded site size of a *Dra*SSB dimer is 45-47 nt at low salt concentrations (< 20mM NaCl) and increases to 50-55 nt at high salt concentrations (≥ 200mM NaCl) [67,75,76]. Therefore, it is unclear whether homodimeric SSBs can undergo similar binding mode transitions as seen in *Eco*SSB.

Here we use single molecule fluorescence resonance energy transfer (FRET) [2,77] to study *Tth*SSB binding, diffusion and binding mode transition on ssDNA.

2.2 Results

2.2.1 *Tth*SSB binding site size estimation using smFRET

We used single-molecule FRET, a single-molecule method to sensitively monitor the redistribution and changes of distance between a donor and an acceptor fluorophore in the range of 3–8 nm [2],to examine the binding of a single *Tth*SSB dimer to a short stretch of ssDNA. A partial duplex DNA substrate containing a 3'60-nt Poly(T) overhang (referred to as $(dT)_{60}$) was immobilized on a surface passivated with polyethylene glycol (PEG). A donor (Cy3) and an acceptor (Cy5) were attached near the two ends of the 60-nt ssDNA overhang respectively so that the FRET efficiency between them reports on the conformations of the 60-nt ssDNA (Figure 2.1A). The occluded site size of a *Tth*SSB dimer should be around 45-55 nt with a slight dependence on salt concentration, similar to other homodimeric SSBs in the Thermus-Deinococcus group [67,74,75]. The 60-nt ssDNA overhang can hence accommodate only one *Tth*SSB dimer in its fully wrapped binding mode. Indeed, previous studies using gel mobility shift assays have shown that $(dT)_{60}$ can efficiently accommodate only one *Tth*SSB dimer [60].

Before adding the proteins, a low FRET peak was observed for $(dT)_{60}$ in the single-molecule (sm) FRET histograms (Figure 2.1B) due to the large end-to-end distance of the 60-nt ssDNA region. As we increased salt concentration from 20 to 500 mM NaCl, the FRET peak position for $(dT)_{60}$ alone moved from ~0.05 to ~0.2 due to the increased ssDNA flexibility (or compaction) at higher salt concentrations[78].1 nM *Tth*SSB dimers were then added into the sample chamber and incubated with the surface-tethered DNA substrates for 5 min at 500 mM NaCl, followed by a buffer wash to flush out the excess unbound *Tth*SSB proteins. Upon *Tth*SSB binding, a much higher FRET peak centered at ~0.8 was observed at 500 mM NaCl even after the buffer wash (Figure 2.1B). As *Tth*SSB can remain bound to $(dT)_{60}$ for 2 hours in the absence of unbound proteins ($k_{off} < 10^{-4}$ s^{-1}; data not shown), we subsequently flowed in buffers containing different salt concentrations (with no proteins) for data acquisition. The observed FRET peak values were ~0.74 and ~0.63 for 100 and 20 mM NaCl respectively (Figure 2.1B). The significantly increased FRET efficiencies observed upon *Tth*SSB binding in a broad range of salt concentrations (20-500 mM NaCl) indicate that the binding of a *Tth*SSB dimer brings the two ends of the *Tth*SSB -bound ssDNA region into close proximity, as seen in homotetrameric SSBs

[56,69,79].The slight decrease in FRET at decreased salt concentrations may result from either a difference in the *Tth*SSB occluded site size [67,74,75], a difference in the flexibility of unbound ssDNA regions [78], or both.

2.2.2 Probing *Tth*SSB diffusion using smFRET

2.2.2.1 *Tth*SSB is capable of diffusion along ssDNA

To test whether *Tth*SSB can diffuse along ssDNA, we used the single-molecule diffusion detection assay previously developed for *Eco*SSB based on DNA hybridization [65,80]. In this assay, we used a partial duplex DNA with a 5' 72-nt ssDNA overhang which consists of a $(dT)_{60}$ region and a 12-nt extension of mixed sequence (referred to as $(dT)_{60+12m}$). Cy3 and Cy5 were attached near the two ends of the $(dT)_{60}$ region (Figure 2.2A). Before adding the proteins, a low FRET peak centered at ~0.15 was observed for $(dT)_{60+12m}$ in the smFRET histogram (Figure 2.2B). Upon *Tth*SSB binding to $(dT)_{60+12m}$, a broad FRET peak centered at ~0.43 was observed (Figure 2.2B) and FRET-time traces obtained from single *Tth*SSB-DNA complexes displayed fast FRET fluctuations beyond measurement noise (Figure 2.2A), as previously observed for *Eco*SSB diffusion on $(dT)_{69+12m}$ [65]. These FRET fluctuations were markedly suppressed and a steady FRET value (~0.7) was observed when the 12-nt mixed sequence within the 72-nt ssDNA overhang was hybridized to its complementary DNA strand (Figures 2.2A and 2.2B). To exclude binding and dissociation of additional *Tth*SSB molecules as the cause of fluctuations, excess unbound *Tth*SSB was removed by a buffer wash step as described above for $(dT)_{60}$. We also ruled out local melting of the duplex portion as a source of FRET fluctuations (Figure 2.3). Therefore, the FRET fluctuations are likely caused by transient excursions of *Tth*SSB from the $(dT)_{60}$ region to the 12-nt extension, i.e. one-dimensional *Tth*SSB diffusion.

To further quantify the time scale of FRET fluctuations, we calculated the average cross-correlation of the donor and acceptor fluorescence intensities from >100 fluorescence intensity time traces of single $(dT)_{60+12m}$ molecules in each condition [66,81]. Anti-correlation between Cy3 and Cy5 intensities was only observed for *Tth*SSB-bound $(dT)_{60+12m}$ without hybridization of the 12-nt extension, and the single exponential fit to the average cross-correlation curve yields a time scale of FRET fluctuations, τ = 61±11 ms (Figure 2.2C). Cross-correlation analysis

showed no significant anti-correlation between Cy3 and Cy5 fluorescence intensities for either unbound $(dT)_{60+12m}$ or *Tth*SSB-bound hybridized $(dT)_{60+12m}$ (Figure 2.2C).

2.2.2.2 *Tth*SSB appears to be positioned at the center of short Poly(T) ssDNA through rapid diffusion

To further test *Tth*SSB diffusion on ssDNA, we designed a series of partial duplex DNA substrates containing 3' Poly(T) overhangs of different lengths (referred to as $(dT)_{60+n}$). The ssDNA overhang is slightly longer than the occluded site size of a *Dra*SSB dimer, and consists of a $(dT)_{60}$ region with Cy3 and Cy5 attached near its ends (Figure 2.4A) and a $(dT)_n$ extension (n = 0, 4, 8, 12 or 16). When n=0, $(dT)_{60+n}$ is the substrate we used in Figure 2.1. Before adding the protein, all the $(dT)_{60+n}$ substrates showed a single low FRET peak centered at ~0.2 in the smFRET histograms at 500 mM NaCl (Figure 2.4B). After incubation with 1 nM *Tth*SSB dimer in 500 mM NaCl and the buffer wash step, a single peak centered at a higher FRET value was observed in the smFRET histograms for all five DNA substrates (Figure 2.4C). Interestingly, the average cross-correlation curves determined from >100 fluorescence intensity time traces of single *Tth*SSB-bound $(dT)_{60+n}$ molecules showed no significant anti-correlation between Cy3 and Cy5 fluorescence intensities (Figure 2.4D), indicating that FRET fluctuations, which are expected from *Tth*SSB diffusion, could not be well detected with our experimental time resolution (~30 ms). We considered two possible mechanisms to account for this observation: 1) After its initial binding to a random location of the ssDNA overhang, *Tth*SSB is unable to move along$(dT)_{60+n}$, resulting in a time-independent single FRET state for each *Tth*SSB-bound $(dT)_{60+n}$ molecule; 2) After its initial binding to a random location, *Tth*SSB is able to move along $(dT)_{60+n}$, but the movement is too fast to be detected within our 30 ms time resolution, resulting in a time-averaged FRET value for each *Tth*SSB-bound $(dT)_{60+n}$ molecule. For the first mechanism, one would expect that the FRET efficiency distribution of $(dT)_{60+n}$ with a larger n value should be broader with *Tth*SSB bound and should include all the FRET efficiencies observed for *Tth*SSB-bound $(dT)_{60+n}$ with a smaller n value, because a longer overhang would provide more possible binding positions for *Tth*SSB and hence result in a broader distribution of FRET efficiency. However, the single peak of the smFRET histogram, instead of becoming broader, is shifted towards a lower FRET value as n increases (Figure 2.4C), disfavoring the first mechanism. Therefore, we favor the second mechanism in which *Tth*SSB rapidly diffuses along the entire

length of the ssDNA overhang and appears to be positioned at the center of the overhang within our time resolution. This time-averaging effect has also been observed in other single molecule FRET studies [82,83] .

2.2.2.3 *Tth*SSB diffusion on Poly(T) ssDNA could be directly observed at lower temperature

All of the measurements above for $(dT)_{60+n}$ were conducted at room temperature (23 ℃) and in 500 mM NaCl. As a further test for the second mechanism, we repeated the experiments at room temperature (23 ℃) and lower temperatures (18, 14 and 11 ℃) and in 100 mM NaCl for $(dT)_{60+4}$ (Figure 2.5A). Previously, we demonstrated that lowering the temperature slows *Eco*SSB diffusion on ssDNA [65]. Indeed, we were able to observe FRET fluctuations beyond measurement noise in the FRET time traces of *Tth*SSB-bound $(dT)_{60+4}$ at 18 and 11 ℃ (Figures 2.5A and 2.5B). We excluded the possibility of local wrapping-unwrapping of ssDNA on *Tth*SSB by conducting the experiment at the same condition (100 mM NaCl, 11℃) on the $(dT)_{60}$ substrate with Cy3 and Cy5 attached near the two ends of the ssDNA overhang. FRET fluctuations were due to anti-correlated fluctuations of the donor and acceptor intensities, as confirmed by the cross-correlation analysis (Figure 2.5C), which yields the fluctuation time scales τ(23 ℃) = 67±28 ms, τ(18 ℃) = 77±14 ms and τ(11 ℃) = 88±13 ms. The Arrhenius fit of $\ln(1/\tau)$ vs. 1/T (Figure 2.5D) gave an apparent activation energy of E_a = 15±3 kJ/mol for *Tth*SSB diffusion, which is smaller than the E_a determined for *Eco*SSB diffusion on ssDNA (~81±7 kJ/mol) [65].

Having demonstrated that *Tth*SSB is capable of diffusion on $(dT)_{60+n}$ under high and intermediate salt condition (500 and 100 mM NaCl), we next tested *Tth*SSB diffusion at low salt concentrations. After loading a *Tth*SSB dimer onto surface-tethered $(dT)_{60+n}$ in 500 mM NaCl, we flushed out unbound proteins with a buffer also containing 500 mM NaCl. A buffer containing either 100 or 20 mM NaCl was subsequently flowed into the sample chamber before data acquisition. At both 100 and 20 mM NaCl, we observed similar results as at 500 mM NaCl: a single narrow peak was observed in the smFRET histogram (data not shown) and is shifted towards a lower FRET value as *n* increases (Figure 2.4E), suggesting that *Tth*SSB diffuses on $(dT)_{60+n}$ in a wide range of salt concentrations (20-500 mM NaCl). As discussed above for $(dT)_{60}$, the slight decrease in FRET efficiency at decreased salt concentrations for each $(dT)_{60+n}$ substrate

(Figure 2.4F) may result from either a difference in the *Tth*SSB occluded site size [67,74,75], a difference in the flexibility of unbound ssDNA regions [78], or both.

2.2.2.4 *Tth*SSB diffusion transiently melts short DNA hairpin structures

Single molecule studies have shown homotetrameric SSBs are able to transiently disrupt short DNA hairpin structures [65,84].To further demonstrate the functional role of *Tth*SSB diffusion, we examined whether *Tth*SSB can use its diffusion activity to transiently melt DNA hairpin structures. We designed a partial duplex DNA containing a 3' $(dT)_{65}$ ssDNA overhang with a 7 base-pair internal hairpin located at the 3' end of the overhang (referred to as $(dT)_{65+hp+3}$).Cy3 and Cy5 were attached to the ends of the hairpin sequence such that FRET reports the formation or melting of the hairpin (Figure 1.6A). Before adding the proteins, $(dT)_{65+hp+3}$ exhibited steady high FRET (~0.9) in the single molecule FRET time traces obtained at 500 mM NaCl, suggesting that the hairpin itself is stably formed at room temperature (T_m > 60 ℃, 500 mM NaCl) (Figure 2.6A). After incubation with 1 nM *Tth*SSB dimer in 500 mM NaCl and the buffer wash step, a number of brief excursions to a low FRET state (~0.5) were observed in FRET-time traces, indicating transient unzipping and reformation of the hairpin. Cross-correlation analysis yielded the time scale of FRET fluctuations τ = 420±36 ms (Figure 2.6B). As the $(dT)_{65}$ region is slightly larger than the occluded site size of a *Tth*SSB dimer, one *Tth*SSB dimer is expected to remain bound to the $(dT)_{65}$ region after the buffer wash. Since the unbound proteins were flushed out before the data acquisition, the transient melting of the hairpin must be due to *Tth*SSB invasion to the hairpin segment.

2.2.3 *Tth*SSB shows binding mode transitions in the presence of proteins in solution

Previous ensemble studies have suggested that homodimeric SSBs, like homotetrameric SSBs, may undergo a transition between different binding modes, although the occluded site size of a homodimeric SSB showed much more reduced salt concentration-dependence than that of a homotetrameric SSB [67,74]. Using smFRET, we previously showed that a partial duplex DNA substrate containing a $(dT)_{70}$ overhang may accommodate either one *Eco*SSB tetramer in the $(SSB)_{65}$ mode or two *Eco*SSB tetramers in the $(SSB)_{35}$ mode, and that a $(dT)_{70}$ molecule undergoes transitions between these two scenarios in the presence of *Eco*SSB proteins in solution [69]. Given that *Tth*SSB dimers have a smaller occluded size (45-55 nt), we asked

whether similar results can be obtained for *Tth*SSB on $(dT)_{60}$. Instead of flushing out the unbound proteins, we kept 1 or 10 nM *Tth*SSB and acquired the data at 20, 100 or 500 mM NaCl accordingly. Figure 2.7A shows representative FRET-time traces obtained from single *Tth*SSB-bound $(dT)_{60}$ molecules. Clear fluctuations between two discrete FRET states (referred to as the high/low FRET state) were observed at 20 and 100 mM NaCl, but not at 500 mM NaCl. We determined the average transition rates between the two FRET states. Previous biochemical studies have shown that a homodimeric SSB binds up to two ssDNA molecules of $(dT)_{25}$ with a salt concentration-dependent negative cooperativity: at both 20 and 200 mM NaCl, the second molecule of $(dT)_{25}$ has a weaker affinity to *Dra*SSB than the first $(dT)_{25}$ molecule, but such negative cooperativity (i.e. the difference in the affinities of the first and the second bound $(dT)_{25}$) is much larger at 20 mM NaCl than at 200 mM NaCl [67]. The salt concentration-dependent negative cooperativity observed for homodimeric SSBs predicts that the decrease in salt concentration would favor the transition from single bound protein to two bound proteins per $(dT)_{60}$ because larger negative cooperativity would make it easier for partial unwrapping of $(dT)_{60}$ from a single *Tth*SSB dimer and thus a second TthSSB dimer has the chance to bind the released ssDNA region. Indeed, the transition rate from high to low FRET states ($k_{H\rightarrow L}$) is larger at higher *Tth*SSB concentrations and at lower salt concentrations, whereas the transition rate from low to high FRET states ($k_{L\rightarrow H}$) does not depend on *Tth*SSB concentration but is smaller at lower salt concentrations (Figure 2.7A and 2.7B). We hence assigned the high FRET state to $(dT)_{60}$ with one *Tth*SSB dimer bound and the low FRET state to $(dT)_{60}$ with two *Tth*SSB dimers bound. In the low FRET state, two *Tth*SSB dimers bind to the same $(dT)_{60}$ molecule, with each *Tth*SSB dimer occupying 30-nt ssDNA region on average.

Next, we tested another DNA substrate with a longer ssDNA overhang $(dT)_{60+16}$ in the presence of 0.5-4 nM *Tth*SSB dimers in solution. Two-state transitions were also observed in the single molecule FRET-time traces at 20 and 100 mM NaCl, but not at 500 mM NaCl (Figures 2.8A and Figure 2.9). However, one of the two states in the FRET-time traces displayed fast FRET fluctuations (referred to as the 'dynamic' state), whereas the other state displayed a steady FRET value (referred to as the 'static' state). We determined the average transition rates between the 'dynamic' and 'static' states, $1/\Delta T_{dynamic}$ and $1/\Delta T_{static}$ ($\Delta T_{dynamic}$ and ΔT_{static} represent average durations in the 'dynamic' and 'static' states, respectively). We found $1/\Delta T_{static}$, similar to $k_{H\rightarrow L}$, displays a linear dependence on *Tth*SSB concentration, whereas $1/\Delta T_{dynamic}$, similar to $k_{L\rightarrow H}$, is

independent of *Tth*SSB concentration (Figure 2.8B). Therefore, we assigned the 'static' state to $(dT)_{60+16}$ with one *Tth*SSB dimer bound and the 'dynamic' state to $(dT)_{60+16}$ with two *Tth*SSB dimers bound. $1/\Delta T_{static}$ and $1/\Delta T_{dynamic}$ hence gave an association rate of $(2.1\pm0.1)\times10^7\ M^{-1}\cdot s^{-1}$ and a dissociation rate of $0.26\pm0.07\ s^{-1}$ for the second *Tth*SSB dimer binding and dissociation to/from $(dT)_{60+16}$, respectively. Furthermore, we quantified the FRET fluctuations within the 'dynamic' state using a hidden Markov model (HMM) based statistical approach that determines the most likely time sequence of FRET states (Figure 2.8C) [85]. FRET-efficiency distributions within the 'dynamic' state can be fit well to a double-Gaussian function (Figure 2.8D), indicating the existence of two distinct FRET substates (~0.3 and 0.6 FRET, respectively). The transition rate from the 0.3 FRET to 0.6 FRET states was determined to be $5.3\pm1.7s^{-1}$, and the transition rate from the 0.6 FRET to 0.3 FRET states was determined to be $3.7\pm1.5s^{-1}$, both of which are independent of *Tth*SSB concentration (Figure 2.8E). We therefore assigned the two FRET states to two distinct structural arrangements of $(dT)_{60+16}$ with two *Tth*SSB dimers bound. The association of the second *Tth*SSB dimers results in a fast 'dynamic' FRET state for $(dT)_{60+16}$ but not for $(dT)_{60}$, probably because the dynamic structural rearrangement from the DNA molecule with two *Tth*SSB dimers bound requires the $(dT)_{16}$ extension.

2.3 Discussion

In this study, we employed smFRET to study the dynamics of *Tth*SSB, a representative homodimeric SSB, binding to short ssDNA molecules ranging from 60 to 81 nt. Although their quaternary arrangements of the OB-folds are different [61,63,64], both homotetrameric and homodimeric SSBs bind ssDNA in a way that brings the two ends of bound DNA in close proximity under high salt conditions, and they are both capable of rapid diffusion along ssDNA. Previously, we have shown that even when the SSB-ssDNA structure is not fully wrapped and a small force (1-5 pN) is used to unwrap some ssDNA from the protein surface, *Eco*SSB diffusion on ssDNA still persists. Our results imply that the diffusion activity may be a shared property for different types of ssDNA binding proteins.

Similar to homotetrameric SSBs, *Tth*SSB can diffuse continually as long as there is an available extension of ssDNA beyond its occluded site size. The activation energy for *Tth*SSB diffusion is ~ 15±3 kJ/mol, smaller than that of *Eco*SSB diffusion (~81±7 kJ/mol) under the same condition

[65]. Additionally, *Tth*SSB diffusion appears to be faster than that of *Eco*SSB, and is too fast to be detected for some DNA substrates (e.g. $(dT)_{60+n}$) within the 30 ms time resolution. This is consistent with the previous observation that binding affinity of ssDNA for homodimeric SSBs is weaker than for *Eco*SSB [67], because SSB diffusion requires breakage and reformation of SSB-DNA interactions [66].

SSB diffusion may have multiple functional roles. First, rapid SSB diffusion along DNA should be important in redistributing SSB on ssDNA after its initial binding to a random DNA location, because for proteins with such high affinities, redistribution would be difficult if it required complete dissociation and reassociation. Second, SSB diffusion can transiently melt short DNA hairpin structures. We have shown that the hairpin removal by *Eco*SSB diffusion is responsible for the facilitated RecA filament growth [65]. For *Tth*SSB, it has been shown that *Tth*SSB stimulates the synthesis rate of DNA polymerases from *Tth* and *Pyrococcus* [86], and many different types of SSBs have been used to increase the amplification efficiency for the polymerase chain reaction (PCR) [87]. Because specific interactions between polymerases and SSBs are not required for the stimulated activity, the observation was presumably due to SSB diffusion that removes DNA hairpin structures. Third, SSB diffusion may provide a mechanism for how SSBs recruit other SSB-interacting enzymes onto ssDNA for subsequent DNA processing. We previously showed that the specific interactions between *Eco*SSB and *Eco*RecO do not abolish but moderately slows *Eco*SSB diffusion on ssDNA [66].

Furthermore, we presented direct evidence that *Tth*SSB undergoes rapid binding mode transitions in the presence of free proteins in solution. Although the occluded site size of a homodimeric SSB displays a more reduced dependence on the salt concentration than that of an *Eco*SSB tetramer and ranges from 45 to 55 nt with increase in salt concentration [59,67,70-76], we observed transitions between a single bound protein and two bound proteins per ssDNA on $(dT)_{60}$ and $(dT)_{60+16}$ at low and moderate salt concentrations (20-100 mM NaCl). When two *Tth*SSB dimers bind to the same 60-nt or 76-nt ssDNA molecule, the second *Tth*SSB has a weaker affinity to DNA than the first bound *Tth*SSB.Additionally,76-nt ssDNA with two *Tth*SSB dimers bound may undergo dynamic structural rearrangement between two conformations while the two *Tth*SSB dimers remain bound to the 76-ntssDNA (Figure 2.10). Similar observations were made previously for *Eco*SSB binding to $(dT)_{70}$ [69].

In summary, the data presented in this study extends our previous observations on homotetrameric *Eco*SSBs to homodimeric SSBs, and suggest that *Tth*SSB/DNA complexes are highly dynamic (Figure 2.10). Importantly, the covalent linkage between multiple, non-identical OB-fold domains in *Tth*SSB dimers, which is also the signature of eukaryotic RPAs, does not preclude SSB diffusion or binding mode transitions. When only one *Tth*SSB dimer binds to an ssDNA region that is slightly longer than its occluded site size, it diffuses rapidly while remaining stably bound, which should be useful to cover and protect even small ssDNA gaps between two bound *Tth*SSBs and remove DNA hairpin structures. Additionally, *Tth*SSB/DNA complexes undergo dynamic interconversions among different binding modes and/or structural rearrangements if there are unbound proteins nearby. These dynamic features of SSB/complex should be important for the functional roles of SSBs in genome maintenance and arepossibly shared among different types of SSBs.

2.4 Experimental procedure

2.4.1 DNA sequences and annealing procedures

DNA oligonucleotides used for the single molecule experiments are listed below:

(1) 5'-/Cy5/ GCC TCG CTG CCG TCG CCA -/biotin/-3'.

(2) 5'-TGG CGA CGG CAG CGA GGC $(T)_{60}$-/Cy3/-3'.

(3) 5'-TGG CGA CGG CAG CGA GGC $(T)_{59}$ / iAmMC6T/$(T)_n$ -3', where n=4, 8, 12,or 16.

(4) 5'- TGG CGA CGG CAG CGA GGC-/Cy3/-$(T)_{58}$ -3'

(5) 5'- /biotin/ TGG CGA CGG CAG CGA GGC /Cy5/-3'

(6) 5'-/5Phos/ GGG CGG CGA CCT /iAmMC6T/ $(T)_{60}$ GCC TCG CTG CCG TCG CCA-3'

(7) 5'-AGG TCG CCG CCC - 3'

(8) 5'-TGG CGA CGG CAG CGA GGC $(T)_{65}$ / iAmMC6T/ TGT GAC TGA GAC AGT CAC TT-/Cy5/-3'

‘/iAmMC6T/’ represents amine-modified thymine that is used to label the DNA with Cy3. ‘/biotin/’ represents biotin modification which was used for the surface immobilization. /5Phos/ represents 5’ phosphorylation. ‘/Cy3/’and ‘/Cy5/’represents the Cy3 and Cy5 fluorophores respectively that were attached directly to the DNA backbone using phosphoramidite chemistry. The partial duplex DNA substrates carrying Cy3 and Cy5 were annealed by mixing ~5μM of biotinylated strand and ~7 μM of non-biotinylated strand in 10 mM Tris:HCl (pH 8.0) and 50mM NaCl, followed by slow cooling from 90°C to room temperature for ~ 2 hours.

2.4.2 Protein expression and purification

Thermus thermophilus (*Tth*) HB8 SSB proteins were expressed and purified as previously described [88,89]. All protein concentrations cited in the text refer to *Tth*SSB dimers.

2.4.3 Sample assembly and data acquisition

Single molecule FRET experiments were performed at 23 ± 1℃ unless specified otherwise. 50-100 pM of partial duplex DNA substrates were immobilized on a quartz slide surface which is coated with polyethylene glycol (mPEG-SC, Laysan Bio) in order to eliminate nonspecific surface adsorption of proteins[2]. The immobilization was mediated by biotin-Neutravidin binding between biotinylated DNA, Neutravidin (Pierce), and biotinylated polyethylene glycol (Bio-PEG-SC, Laysan Bio). For the experiments shown in Figures 2.1 - 2.6, 1 nM *Tth*SSB were added into the sample chamber and incubated with the surface-tethered DNA substrates for 5 min in Buffer A containing 20 mM Tris:HCl (pH 8.0), 500 mM NaCl and 0.1 mg/ml BSA. A buffer wash step was then performed using Buffer A to flush out the excess unbound proteins. Finally, Buffer B containing 20 mM Tris:HCl (pH 8.0), 0.1 mg/ml BSA, 2% (v/v) glycerol, 0.5 % (w/v) D-glucose, 165 U/ml glucose oxidase, 2170 U/ml catalase, 3 mM Trolox and indicated amounts of NaCl was injected into the sample chamber before data acquisition. For the experiments shown in Figures 2.7 - 2.9, buffer B with desired *Tth*SSB concentration was directly injected into the sample chamber for data acquisition after immobilizing the DNA substrates. Single molecule data were acquired using total internal reflection fluorescence (TIRF) microscope [80] with a time resolution of ~ 30 ms. In some circumstances where the experiment needs to be performed at lower temperatures, a water-circulating bath circulator (NESLAB RTE-7 Digital One, Thermo Scientific) was used to cool down the sample stage, objective and prism-

holder all together. The temperature of the outer surface of the sample chamber was measured by a handhold digital thermometer (Omega) as the best estimation of the experimental temperature inside the sample chamber.

2.4.4 FRET efficiency calculation

Apparent FRET efficiency was calculated from the fluorescence intensities of the donor (I_D) and acceptor (I_A) using the formula $E_{FRET}=I_A / (I_A+ I_D)$. The background and the cross-talk between the donor and acceptor were corrected as described previously[2].Single-molecule FRET histograms were built from more than 5000 surface-tethered DNA molecules.

2.4.5 Cross-correlation analysis

The cross-correlation analysis was performed as described previously [81,90]. The cross-correlation curves were calculated between donor and acceptor fluorescence intensity time traces obtained from each DNA molecule carrying a donor and an acceptor, and were then averaged over more than 100 DNA molecules. All cross-correlation curves presented are average curves. We determined the characteristic time, τ, by fitting the average cross-correlation curve to a single exponential function.

2.5 Figures

Figure 2.1

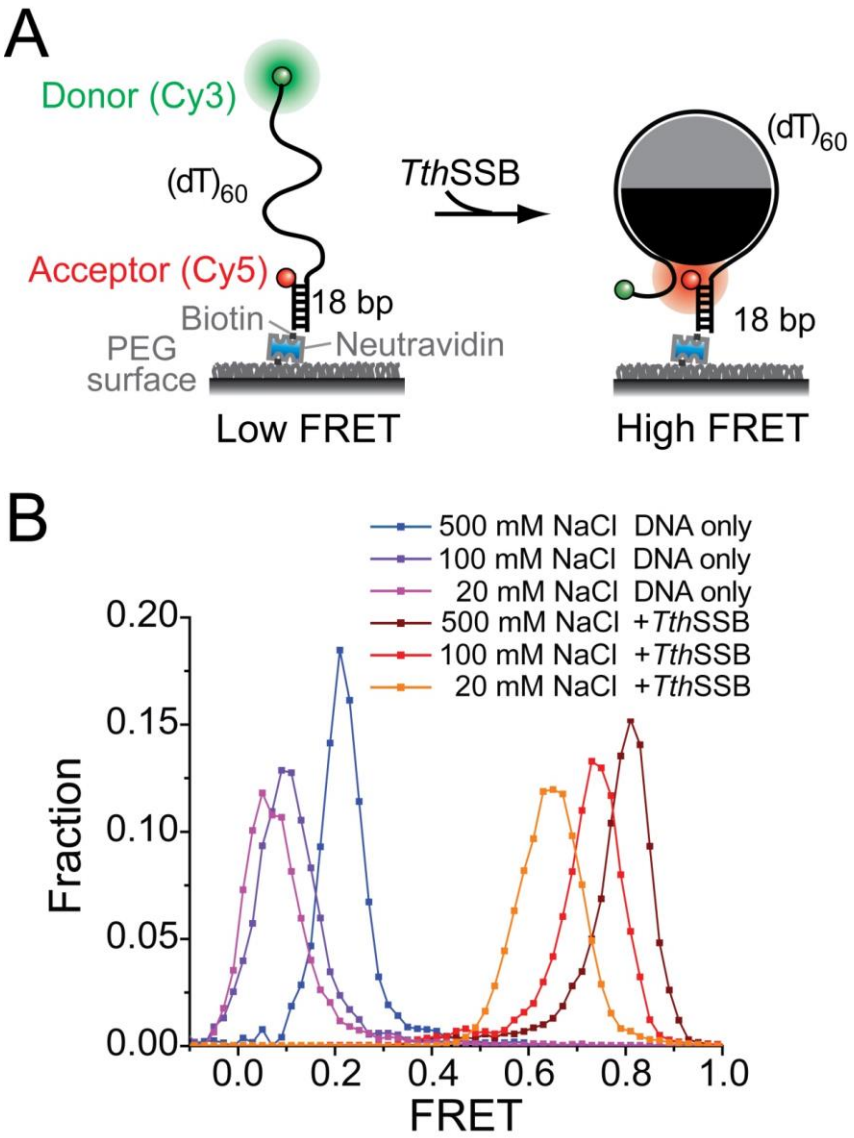

Figure 2.1. Single molecule FRET assays that report the conformations of *Tth*SSB-bound ssDNA. (A) A schematic illustration of our single molecule FRET experimental design for *Tth*SSB. A partial duplex DNA substrate ($(dT)_{60}$) was immobilized on a PEG-coated surface. Cy5 and Cy3 were attached to the ss-dsDNA junction and the end of the ssDNA overhang, respectively. (B) Single molecule (sm) FRET histograms for $(dT)_{60}$ DNA alone and *Tth*SSB binding to $(dT)_{60}$ at different salt concentrations. *Tth*SSB proteins were loaded onto $(dT)_{60}$ under high salt condition (500 mM NaCl) and unbound proteins were removed before data acquisition.

Figure 2.2

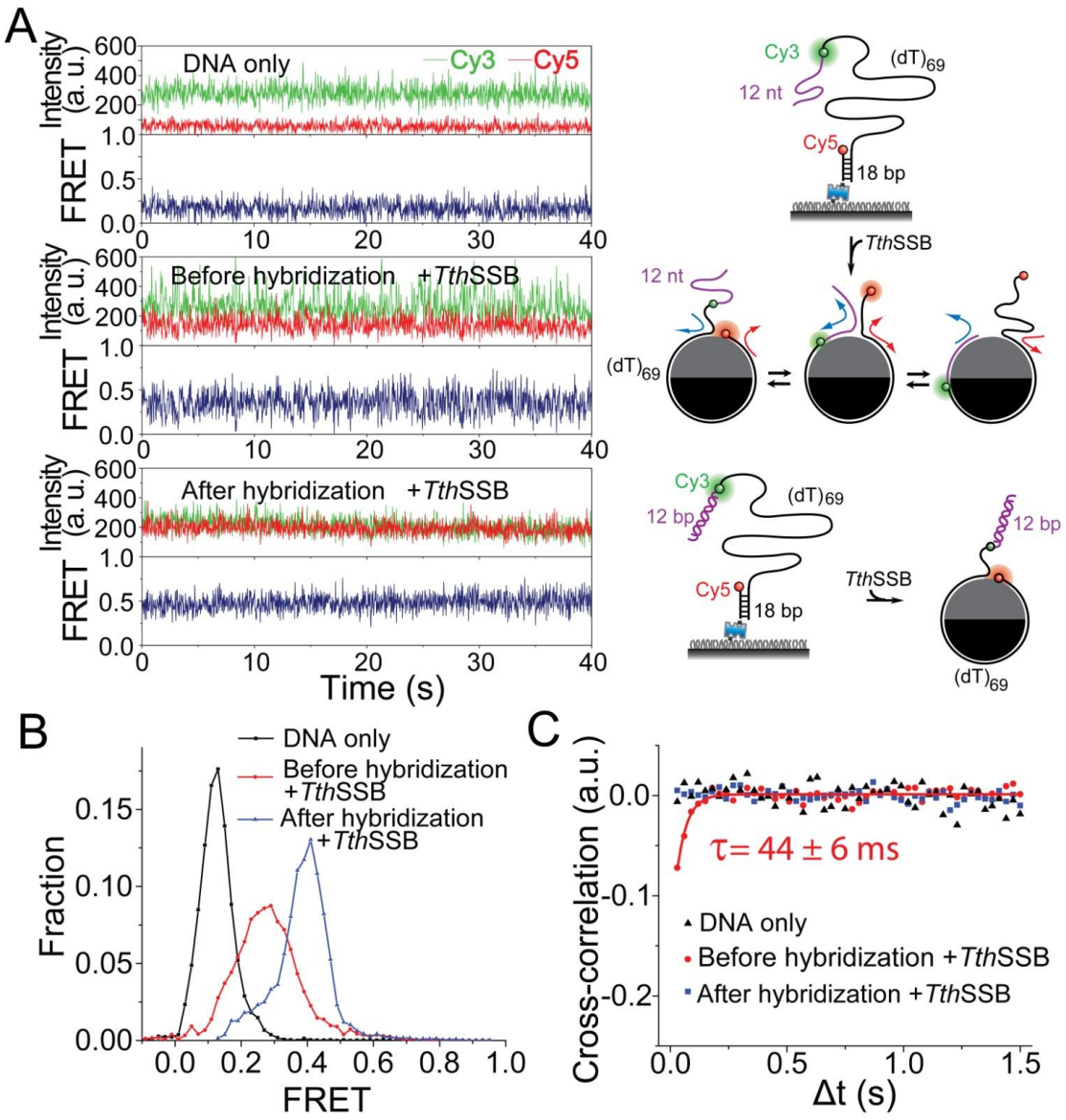

Figure 2.2. Protein diffusion detection assays based on DNA hybridization. (A-C) Representative single molecule time traces (A), smFRET histograms (B), and average cross-correlation curves (C), for$(dT)_{69+12m}$ DNA alone and *Tth*SSB-bound $(dT)_{69+12m}$ with and without the hybridization to the 12-nt mixture sequence ssDNA region. FRET fluctuations beyond measurement noise were detected only when the 12-nt extension is available for *Tth*SSB binding. Unbound proteins were removed before data acquisition such that the FRET flucatuations reflect only the repositioning of the bound *Tth*SSB along $(dT)_{69+12m}$. For simplification, the 18-bp duplex DNA region is not shown for *Tth*SSB-bound $(dT)_{69+12m}$ in (A). The solid line in (C) is a fit to a single exponential function.

Figure 2.3

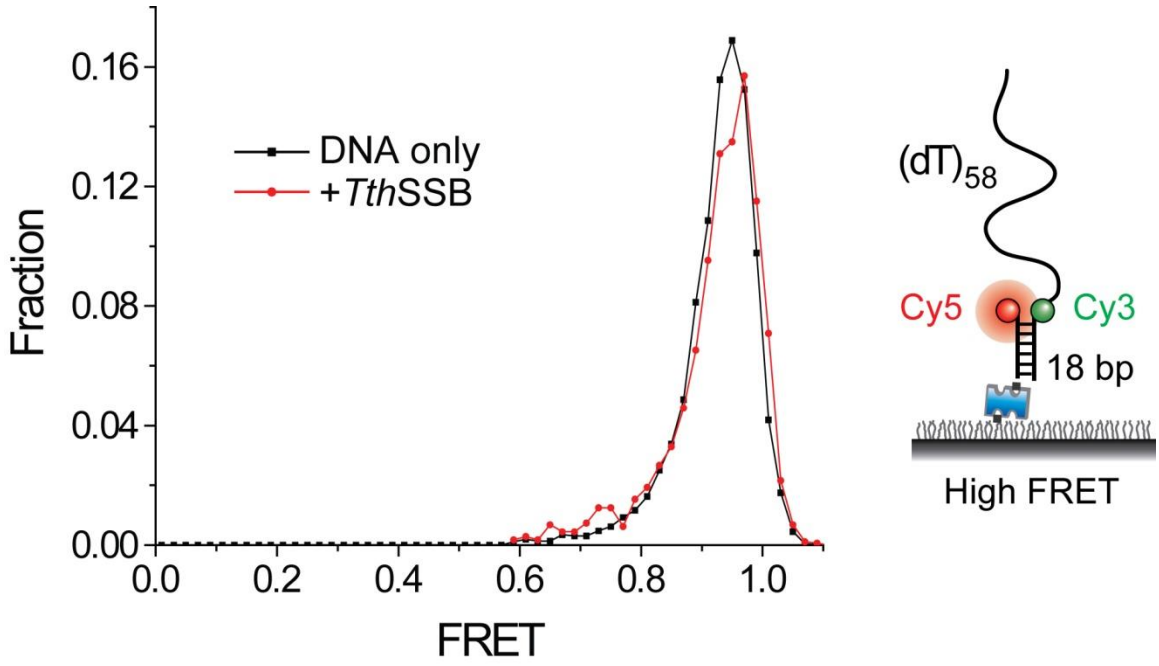

Figure 2.3. Test for duplex destabilization by *Tth*SSB. A partial duplex DNA substrate containing a 3' $(dT)_{58}$ overhang was used. Cy3 and Cy5 were attached near the ss-dsDNA junction such that FRET reports the destabilization (or unwinding) of the duplex DNA region. No change was detected in the single molecule FRET histogram before and after adding 10 nM *Tth*SSB at 500 mM NaCl, indicating that *Tth*SSB binding does not destabilize the 18-bp duplex region.

Figure 2.4

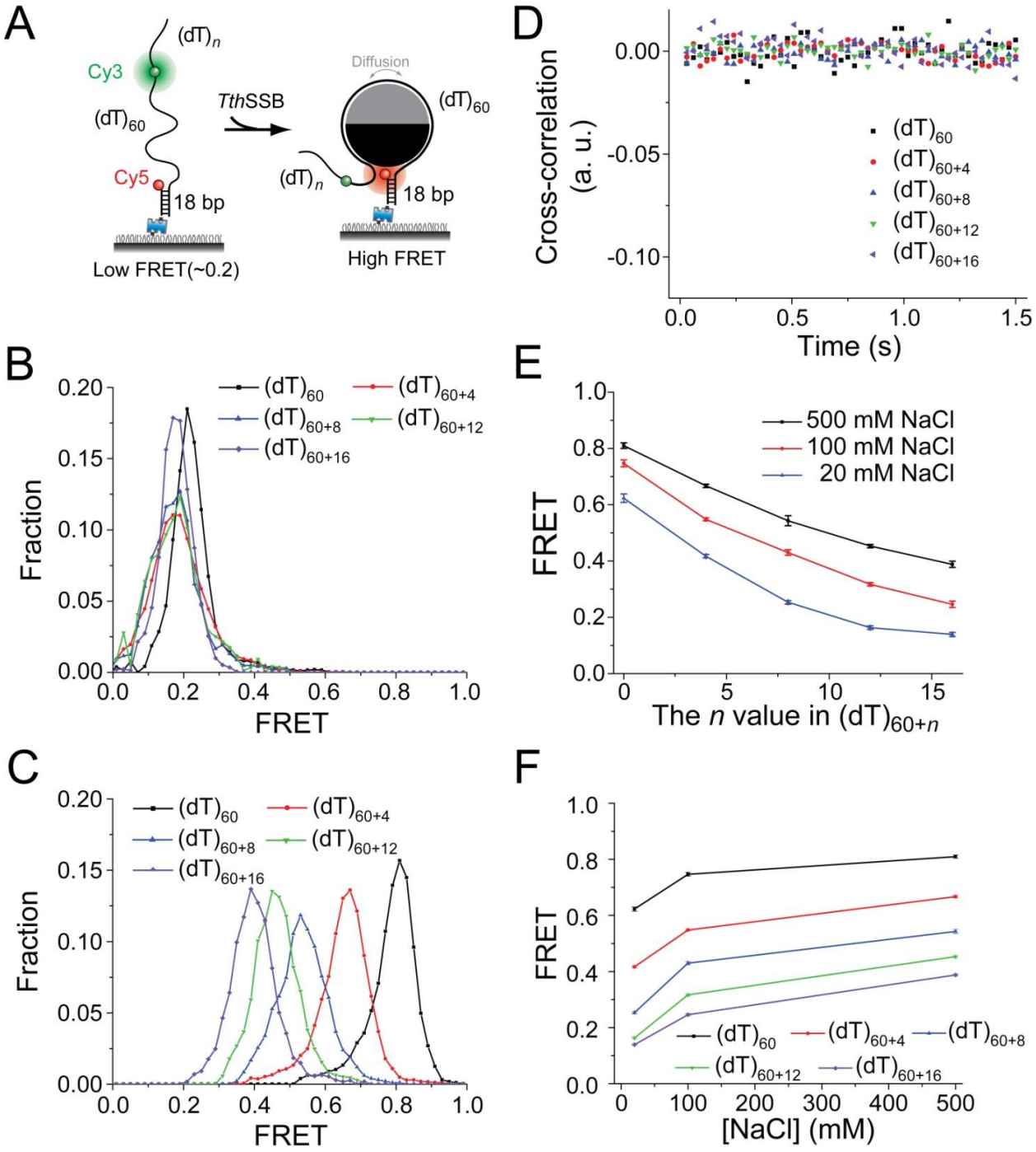
A
(dT)n
Cy3
(dT)60
Cy5
18 bp
Low FRET(~0.2)
TthSSB
Diffusion
(dT)60
(dT)n
18 bp
High FRET
D
Cross-correlation (a. u.)
0.00
-0.05
-0.10
(dT)60
(dT)60+4
(dT)60+8
(dT)60+12
(dT)60+16
0.0
0.5
1.0
1.5
Time (s)
B
Fraction
0.20
0.15
0.10
0.05
0.00
(dT)60
(dT)60+4
(dT)60+8
(dT)60+12
(dT)60+16
0.0
0.2
0.4
0.6
0.8
1.0
FRET
E
FRET
1.0
0.8
0.6
0.4
0.2
0.0
500 mM NaCl
100 mM NaCl
20 mM NaCl
0
5
10
15
The n value in (dT)60+n
C
Fraction
0.20
0.15
0.10
0.05
0.00
(dT)60
(dT)60+4
(dT)60+8
(dT)60+12
(dT)60+16
0.0
0.2
0.4
0.6
0.8
1.0
FRET
F
FRET
1.0
0.8
0.6
0.4
0.2
0.0
(dT)60
(dT)60+4
(dT)60+8
(dT)60+12
(dT)60+16
0
100
200
300
400
500
[NaCl] (mM)

Figure 2.4 (continued)

Figure 2.4. *Tth*SSB appears to be positioned at the center of the $(dT)_{60+n}$ ssDNA overhangs though rapid diffusion. (A) A schematic illustration of our experimental design for $(dT)_{60+n}$ (n=0, 4, 8, 12 or 16). Cy3 and Cy5 were attached to the ss-dsDNA junction and the middle of the ssDNA overhang, respectively, separated by $(dT)_{60}$. (B) Single molecule FRET histograms of $(dT)_{60+n}$ DNA in the absence of proteins, obtained at 500 mM NaCl. (C) smFRET histograms for *Tth*SSB-bound $(dT)_{60+n}$, obtained at 500 mM NaCl, showing a single narrow FRET peak. (D) Average cross-correlation curves for *Tth*SSB-bound $(dT)_{60+n}$, obtained at 500 mM NaCl, indicating no significant FRET fluctuations. (E) The FRET value at the FRET peak position versus the *n* value in $(dT)_{60+n}$, obtained at 20, 100, or 500 mM NaCl. (F) The FRET value at the FRET peak position versus the salt concentration for different $(dT)_{60+n}$ substrates. *Tth*SSB proteins were loaded onto $(dT)_{60}$ under high salt condition (500 mM NaCl) and unbound proteins were removed before the buffer containing a lower NaCl concentration was added to the sample chamber.

Figure 2.5

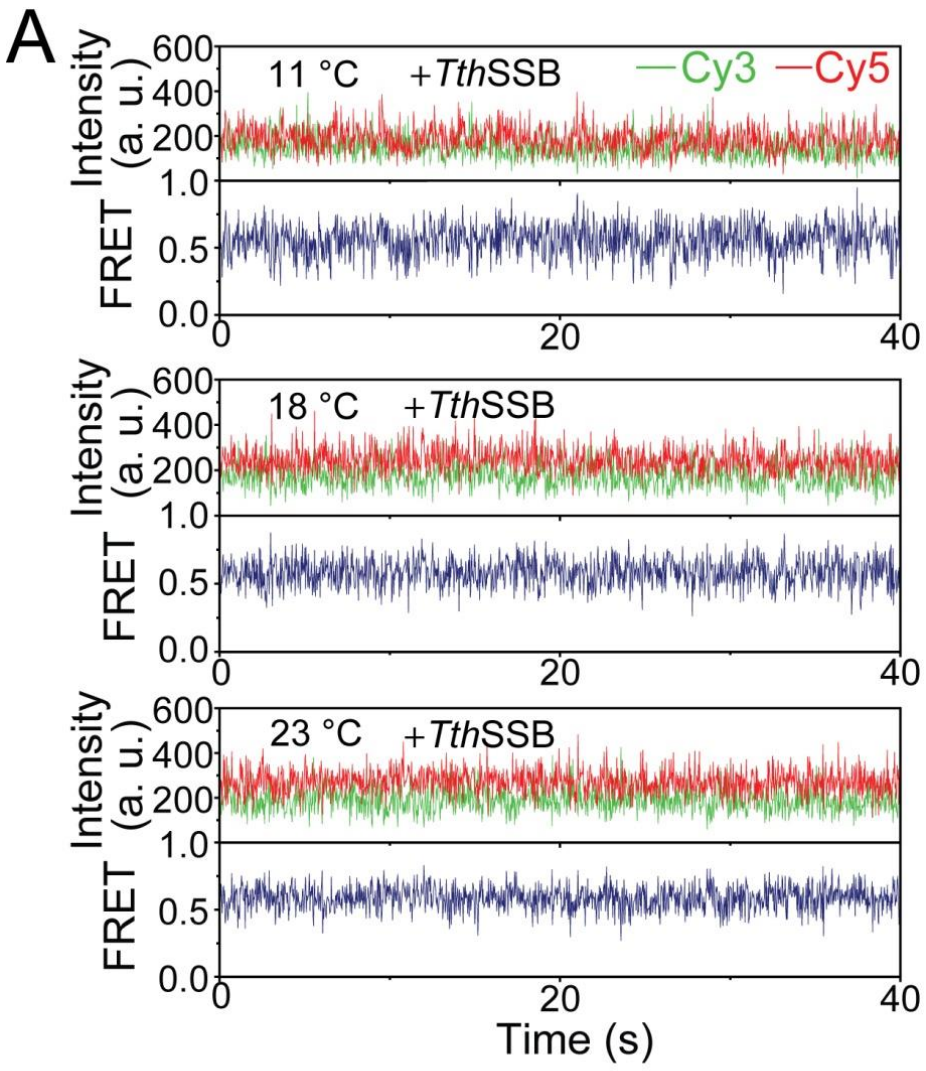
A
11 °C +TthSSB
Cy3
Cy5
18 °C +TthSSB
23 °C +TthSSB
Intensity (a. u.)
FRET
600
400
200
1.0
0.5
0.0
0
20
40
Time (s)

Figure 2.5 (continued)

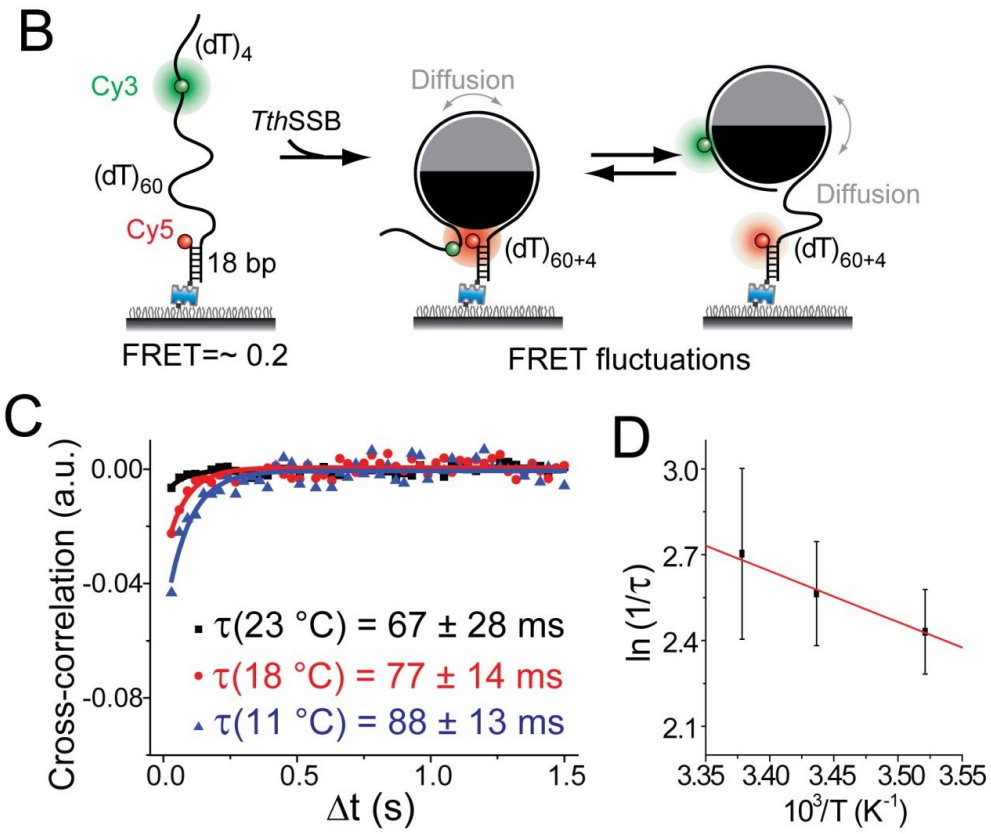

Figure 2.5. Lower temperatures slow *Tth*SSB diffusion on $(dT)_{60+4}$. (A) Representative single molecule time traces for *Tth*SSB-bound $(dT)_{60+4}$ obtained at different temperatures (23, 18 and 11 ℃) and at 100 mM NaCl. (B) A schematic illustration of our experimental design, showing how *Tth*SSB diffusion may result in FRET fluctuations. (C) Average cross-correlation curves for *Tth*SSB-bound $(dT)_{60+4}$ obtained at different temperatures and at 100 mM NaCl. More significant FRET fluctuations beyond measurement noise were detected at lower temperatures. The time scales of the FRET fluctuations were determined from fits to a single exponential function (solid lines). Unbound proteins were removed before data acquisition. (D) Arrhenius plot of apparent rates as a function of 1/T.

Figure 2.6

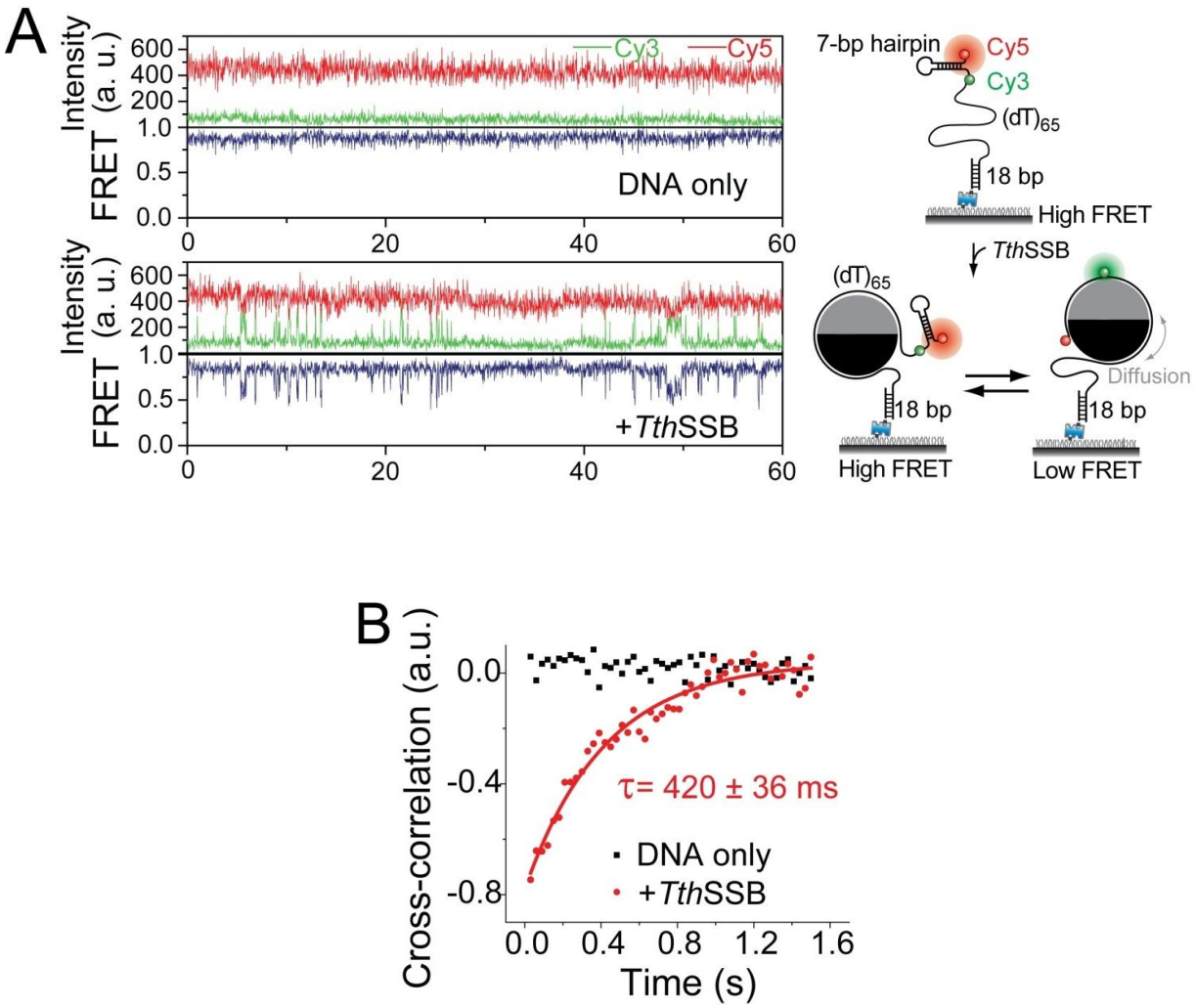

Figure 2.6. *Tth*SSB diffusion can transiently melt short DNA hairpin structures. (A and B) Representative single molecule time traces (A) and average cross-correlation curves (B), for $(dT)_{65+hp+3}$ alone and *Tth*SSB-bound $(dT)_{65+hp+3}$, obtained at 500 mM NaCl. Unbound proteins were removed before data acquisition. The solid line in (B) is a fit to a single exponential function.

Figure 2.7

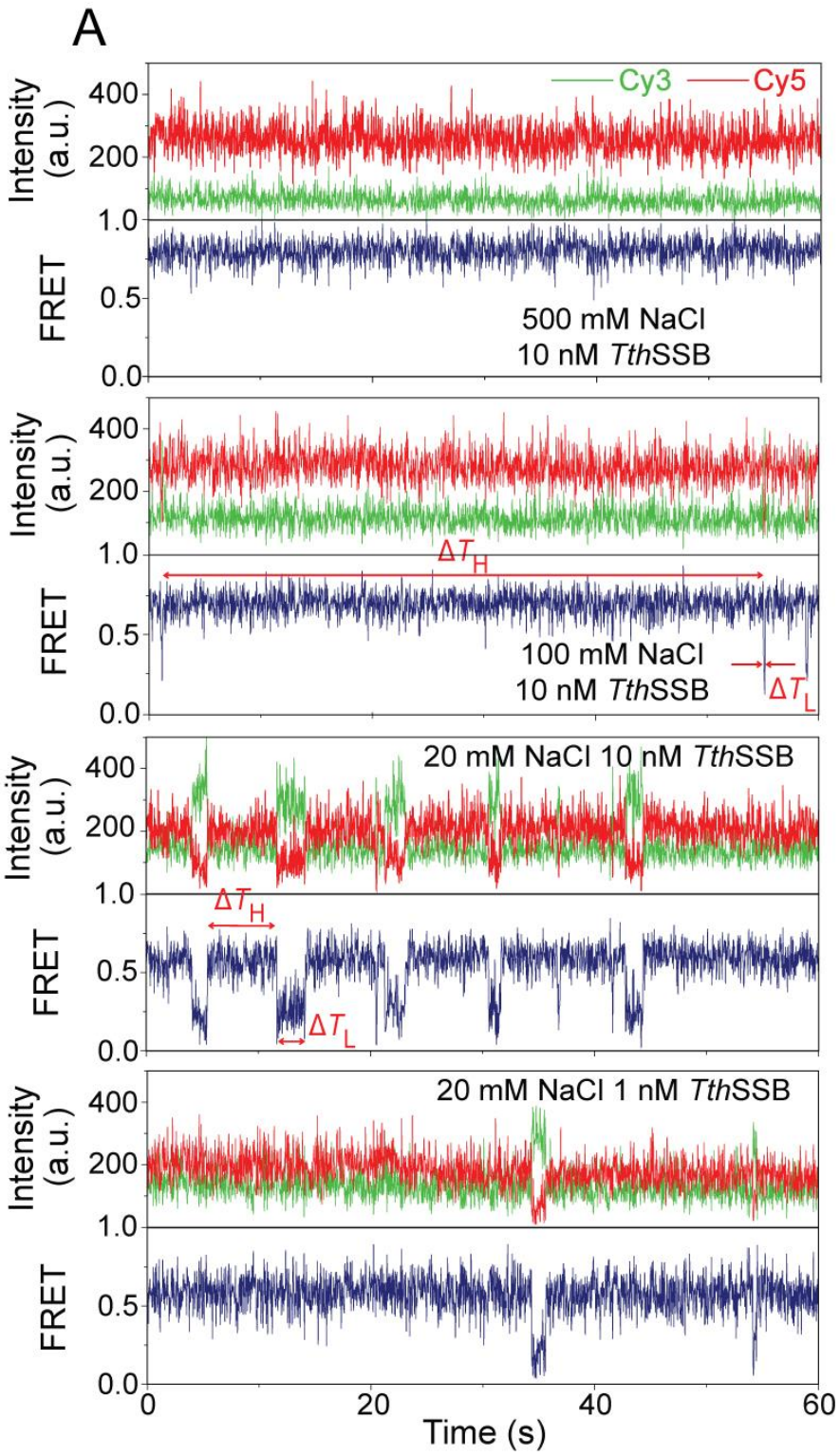
A
Cy3
Cy5
Intensity (a.u.)
FRET
500 mM NaCl
10 nM TthSSB
100 mM NaCl
10 nM TthSSB
ΔT_H
ΔT_L
20 mM NaCl 10 nM TthSSB
20 mM NaCl 1 nM TthSSB
Time (s)

Figure 2.7 (continued)

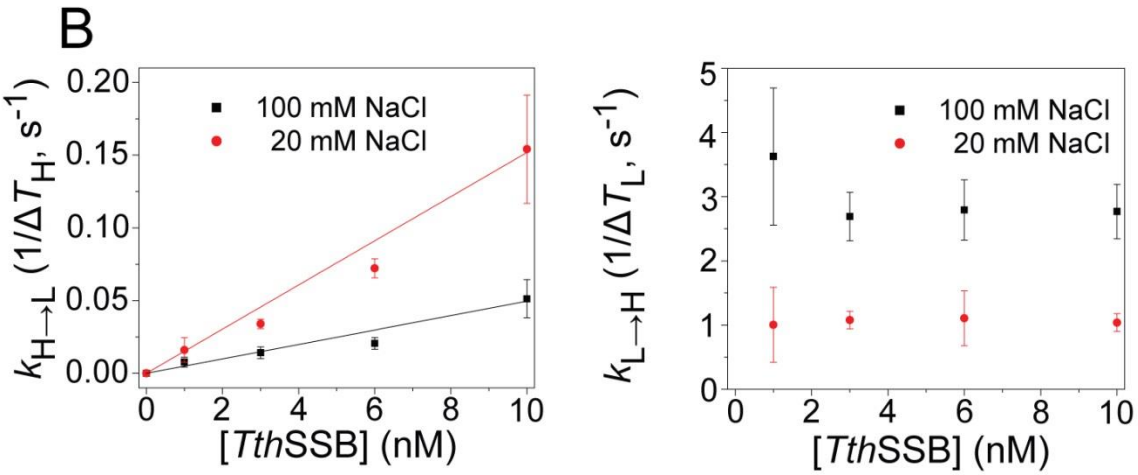

Figure 2.7. Transitions between different binding modes for *Tth*SSB binding to $(dT)_{60}$ in the presence of proteins in solution. (A) Representative single molecule FRET-time traces for *Tth*SSB-bound $(dT)_{60}$, obtained at 500 and 100 mM NaCl when 10 nM *Tth*SSB were present, and those obtained at 20 mM NaCl and when 10 or 1 nM *Tth*SSB were present in the sample camber. 1 nM *Tth*SSB were first incubated with the surface-tethered $(dT)_{60}$ in 500 mM NaCl, followed by a buffer wash to remove the unbound *Tth*SSB. The buffer containing the indicated *Tth*SSB and NaCl concentration was then injected into the chamber for data acquisition. ΔT_H and ΔT_L represent the durations for the high and low FRET states, respectively. (B) Transtion rates of the interconversions between high and low FRET states ($1/\Delta T_H$ and $1/\Delta T_L$ represent the high to low FRET state transiton and the low to high FRET state transition, respectively), as a function of the *Tth*SSB concentration. Solid lines are the linear fits for high to low FRET state transition rate as a function of *Tth*SSB concentration.

Figure 2.8

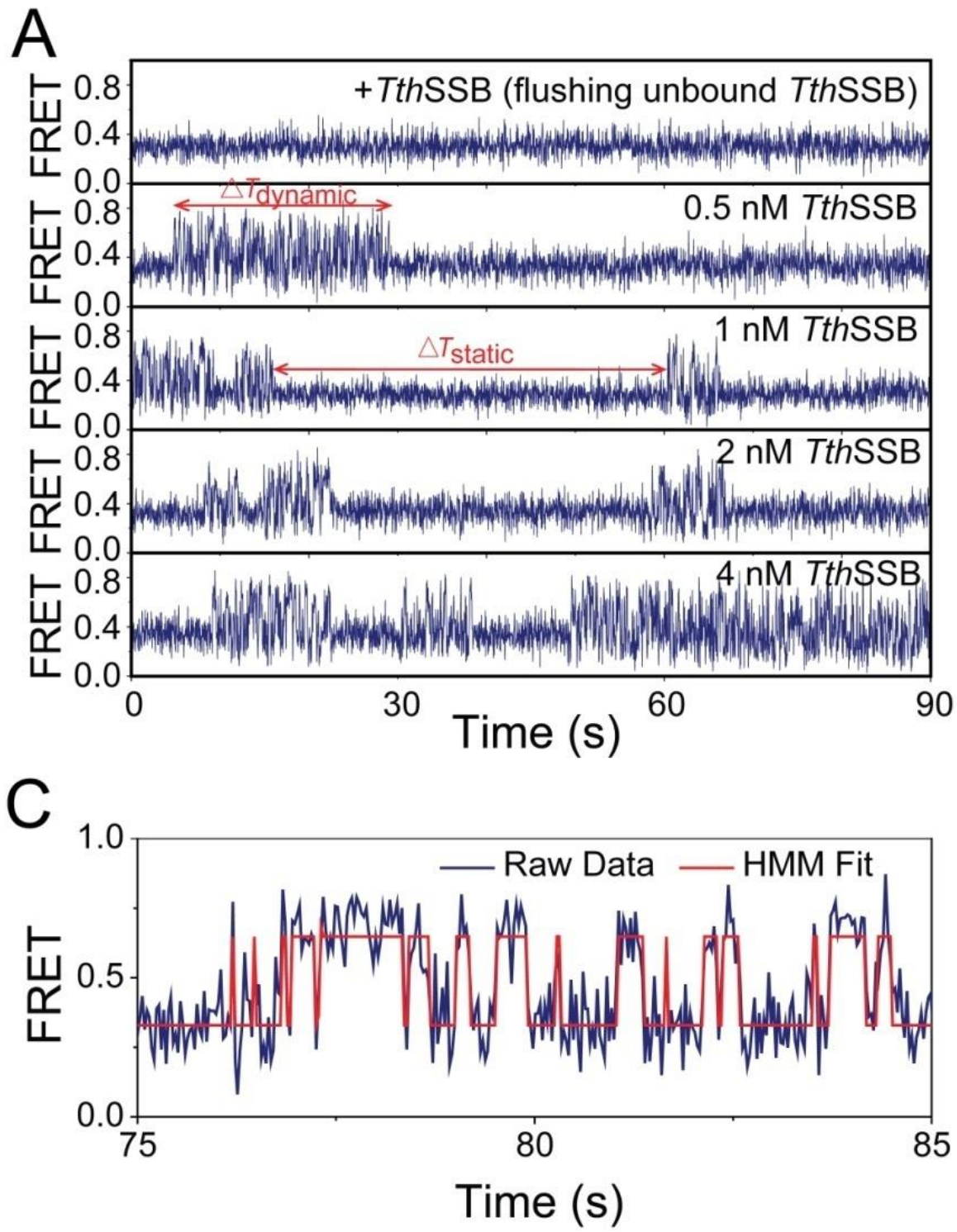
A
+TthSSB (flushing unbound TthSSB)
0.5 nM TthSSB
1 nM TthSSB
2 nM TthSSB
4 nM TthSSB
△Tdynamic
△Tstatic
FRET
0.0
0.4
0.8
0
30
60
90
Time (s)
C
Raw Data
HMM Fit
FRET
0.0
0.5
1.0
75
80
85
Time (s)

Figure 2.8 (continued)

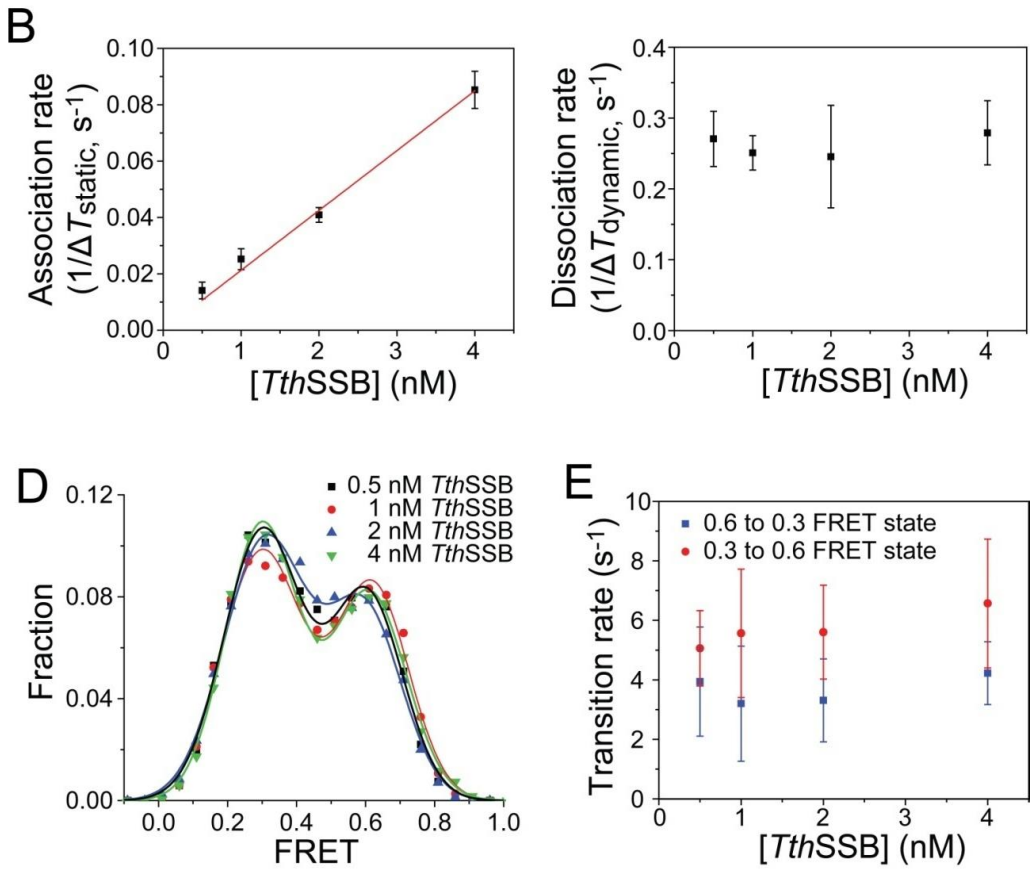

Figure 2.8. Transitions between different binding modes for *Tth*SSB binding to $(dT)_{60+16}$ in the presence of proteins in solution. (A) Representative single molecule FRET-time traces for *Tth*SSB-bound $(dT)_{60+16}$, obtained at 100 mM NaCl and when 0, 0.5, 1, 2 or 4 nM *Tth*SSB were present in the sample camber. $\Delta T_{dynamic}$ and ΔT_{static} represent the durations for the 'dynamic' and 'static' states, respectively. (B) Association and dissociation rates of the second *Tth*SSB to/from $(dT)_{60+16}$ ($1/\Delta T_{static}$ and $1/\Delta T_{dynamic}$, respectively), as a function of the *Tth*SSB concentration. (C) Hidden Markov Model (HMM)-derived idealized FRET trajectory (red) superimposed on the FRET-time trace (blue), for a selected time period during which the molecule is in the 'dynamic' state. Two FRET states were determined from the HMM fit (~0.3 and ~0.6 FRET). (D) FRET-efficiency distributions within the 'dynamic' state at different *Tth*SSB concentrations (averaged from >50 molecules).Solid lines are the fits to a double-Gaussian function. (E) Transition rates between the 0.3 and 0.6 FRET states, determined from the HMM fit at different *Tth*SSB concentrations.

Figure 2.9

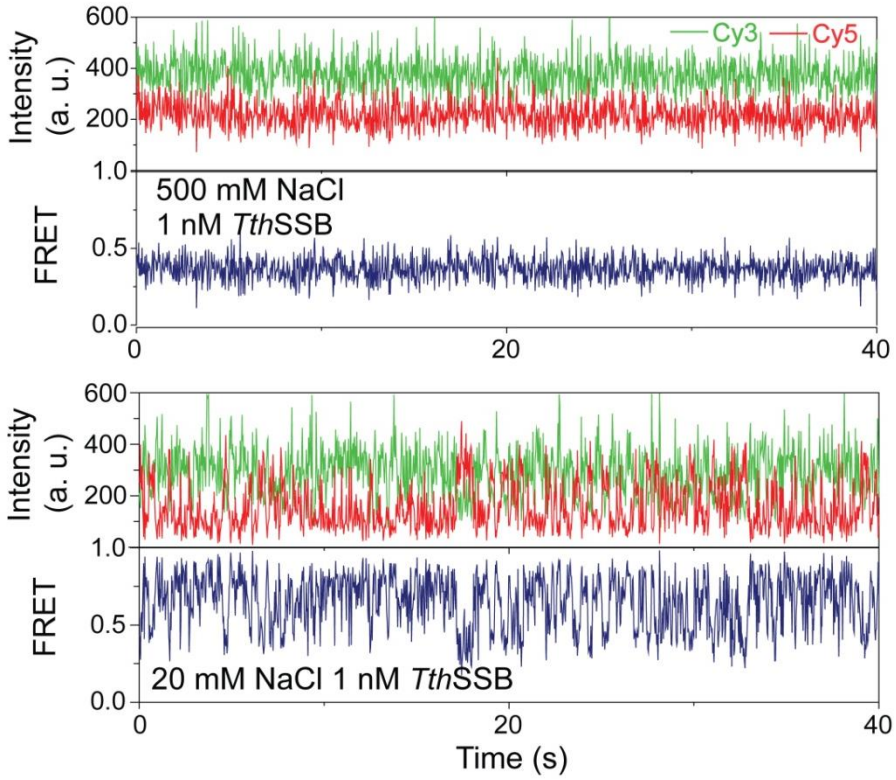

Figure 2.9. Representative single molecule FRET-time traces for *Tth*SSB-bound $(dT)_{60+16}$, obtained at 500 or 20mM NaCl and when 1 nM *Tth*SSB were present in the sample chamber. Fast FRET fluctuations were observed at 20 mM but not at 500 mM NaCl, indicating that binding mode transition takes place at 20 mM but not at 500 mM NaCl for $(dT)_{60+16}$ DNA substrate.

Figure 2.10

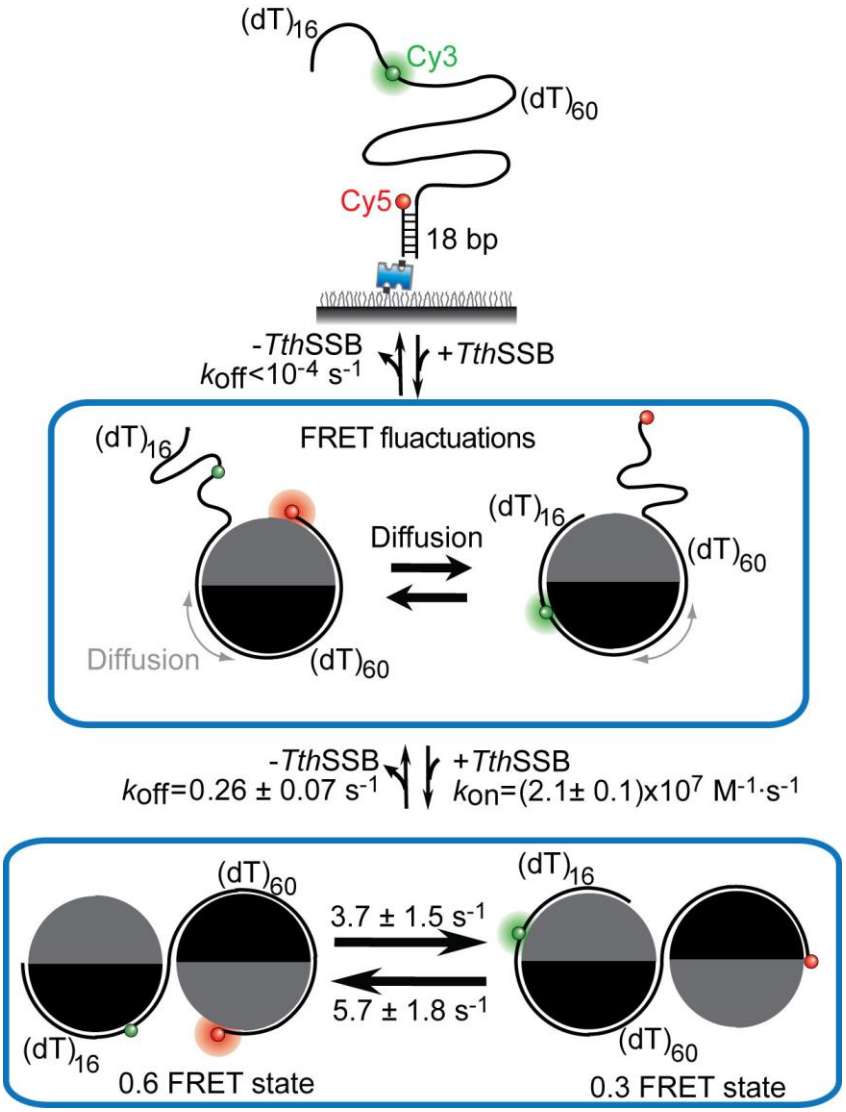

Figure 2.10. Model of *Tth*SSB dynamics on ssDNA. $(dT)_{60+16}$ is illustrated as the ssDNA template for *Tth*SSB binding. For simplification, only the ssDNA region is shown in the cartoon for *Tth*SSB-bound $(dT)_{60+16}$.

CHAPTER 3

CONSTRUCTION AND IN VITRO CHARACTERIZATION OF SPINACH ARRAY*

3.1 Introduction

RNAs play diverse important roles inside the cell. To study RNA biology, one of the most straightforward and useful ways is to directly visualize RNA molecules by fluorescence imaging. As we mentioned before, one of the most commonly used method is fluorescence *in situ* hybridization (FISH) targets RNAs with fluorescence dye-labeled complementary DNA oligos. *FISH* method does not require target RNA modification and performs greatly for *in situ* RNA imaging and quantification but is usually not compatible with live cell RNA imaging. An alternative RNA imaging method utilizes RNA binding protein-fluorescent protein fusion and target RNAs were recognized and labeled after modification with specific "cognate" sequence. One of the most widely applied systems is the RNA bacteriophage MS2 coat protein system, and this imaging strategy allows live cell visualization and tracking of RNAs. However, protein association to RNAs might affect original RNA localization and metabolism.

More direct ways for RNA labeling in living cells mostly use RNA aptamers that can bind small ligands (or "fluorogens") and activate their fluorescence. Among several aptamer-ligand combinations, "Spinach" and analogous systems (Spinach[91], Spinach 2[92], RNA Mango[93], Broccoli[94], *etc*.) have shown the greatest potential in biochemical assay and live cell imaging[91-95]. Spinach uses a short RNA aptamer (24-2 and 24-2-min RNA sequences, reported by Paige and coworkers[91]; ~100 nucleotides) that exhibits EGFP-like green fluorescence upon binding of 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), a fluorogenic ligand that is structurally similar to the EGFP chromophore and is membrane-permeable and nontoxic. Spinch RNA folds into an RNA G-quadruplex structure providing a

---

*Chapter 3 was excerpted from a published work:

Zhang, J. et al. "Tandem Spinach array for RNA imaging in living bacterial cells" Sci. Rep. doi:10.1038/srep17295 (2015)

binding site for the fluorogen[96-98], and structural stabilization of the fluorogen bound to the G-quadruplex structure is likely responsible for its strong fluorescence enhancement. Although the Spinach system has been used for imaging highly abundant nontranslated RNAs (tRNAs[91], rRNAs[91,92,94] and trinucleotide repeats[92]) and further applied to detect cellular metabolites and proteins[99,100], there have been very few studies on the utility of Spinach in imaging cellular mRNA[92], and the fluorescence signal was only barely above cellular autofluorescence level, likely due to low abundance of mRNA and the low brightness of a single fluorogen-bound RNA aptamer. To amplify the signal, making a tandem array of these RNA aptamers, as was used for MS2 and PP7 systems[38,101], could be a general solution to improve the brightness of the aptamer-fluorogen systems. However, it remains to be tested whether this strategy is suitable for RNA imaging because making tandem arrays of these RNA aptamers would significantly increase the RNA length which may potentially affect mRNA metabolism.

Here we use Spinach as our model system and constructed Spinach arrays where we fused multiple repeats of the Spinach aptamer on a single RNA molecule. In this chapter we measured the *in vitro* fluorescence of Spinach arrays, characterized its folding efficiency and fluorogen-binding binding kinetics, before applying Spinach arrays to RNA imaging.

3.2 Results

3.2.1 Construction of Spinach array

We designed a series of Spinach arrays containing different numbers of tandem Spinach aptamers (Spi-*n*R, n = 8, 16, 32, 64). Between two adjacent aptamer repeats we inserted 17-nt randomized spacer sequence, which was used for constructing tandem repeats of MS2 RNA sequence[37,38]. The 8-repeat Spinach array (Spi-8R) DNA sequence was ordered from Genscript, and 16-, 32- and 64-repeat Spinach arrays (Spi-16R, -32R and -64R) were constructed based on Spi-8R (Methods and Materials). The Spinach array DNA sequence was inserted into pET28a plasmid with the embedded T7 promoter and *lac* operator system which can be used for either *in vitro* transcription or cellular expression of the Spinach array RNA (Figure 3.1A). The Spinach array RNA used for *in vitro* measurement and test was obtained via *in vitro* transcription using T7 RNA polymerase (Methods and Materials).

3.2.2 Spinach arrays significantly enhance fluorescence intensities of single Spinach

Fluorescence intensities of *in vitro* transcribed Spinach arrays (100 nM RNA) were measured by fluorometer after RNA folding and incubation with DFHBI (20 μM) (Methods and Materials). Compared to single Spinach aptamer (Spi), Spi-*n*R shows greatly enhanced fluorescence signal with the same maximum excitation/emission wavelengths at 460/505 nm (Figure 3.1B). The fluorescence intensity increases with the repeat number: it changes approximately 17-fold when the aptamer repeat number increases 64-fold (Figure 3.1C). We calculated how the fluorescence intensity of Spi-*n*R increases upon duplication of the aptamer number within the array, described by fluorescence enhancement efficiency (Methods and Materials). We found that for Spinach arrays, upon aptamer repeat number duplication, the fluorescence intensity showed ~1.59-fold enhancement on average (Figure 3.2).

3.2.3 Spinach arrays show relatively low folding efficiency

We noticed that the duplication of the number of Spinach aptamers within the Spinach array does not lead to doubling of the fluorescence intensity. There are two possible explanations for the less than 2-fold increase in the brightness when the aptamer number was doubled: fluorescence quenching between adjacent aptamers, and misfolding/incomplete folding of multiple aptamers[38]. To test inter-aptamer quenching, we measured the fluorescence lifetimes of single Spinach and Spinach arrays. The average fluorescence lifetime of Spi, Spi-*8*R and Spi-*32*R was 4.0, 3.91 and 3.63 ns, respectively. By calculating the extent of quenching using the measured average fluorescence lifetime (Methods and Materials), we estimated that quenching can account for only up to 10% in fluorescence reduction even in the worst case. To test whether Spinach arrays have folding problems, we used a previously reported assay to estimate the folding efficiency[92,94] and found that the relative folding efficiency of Spi-*8*R compared to Spi is 43.5 ± 1.6 % and the value drops further to 34.6 ± 2.0 % for Spi-*32*R (Table 3.1). In previous literature people have reported that single Spinach is subject to misfolding[92]. Spinach arrays may misfold more extensively since in addition to the relatively low thermostability of Spinach aptamer which brings about its susceptibility to misfolding, the tandem aptamer sequence within a Spinach array might encounter inter-aptamer misfolding, similar to the example where proteins consisting repetitive domains are known to misfold due to interdomain interactions[102]. Our

measurement showed that the folding efficiencies of Spinach arrays are significantly lower than that of single Spinach, suggesting that inter-aptamer misfolding possibly takes place. Incorporating the single Spinach aptamer into a tRNA scaffold could slightly increase its fluorescing likely through improving the folding of the Spinach sequence (Table 3.1). Nevertheless this strategy might not be applicable to Spinach arrays: on one hand, it does not solve the problem of inter-aptamer misfolding; on the other hand, recent study showed that the RNA aptamer incorporated into tRNA scaffold suffered endonucleolytic cleavage due to tRNA sequence recognition by RNases in bacteria and mammalian cells[94], disfavoring the extensive use of tRNA scaffold in Spinach imaging.

3.2.4 Kinetics measurement of DFHBI binding onto Spinach arrays

We measured the binding kinetics of DFHBI onto Spinach aptamer within a Spinach array and tested whether incorporating Spinach aptamers into an array significantly influences the binding kinetics. During our measurement, certain concentrations of pre-folded single Spinach (Spi) or Spinach arrays (Spi-*n*R) were quickly mixed with certain concentrations of DFHBI to measure the fluorescence intensity of the solution as a function of time, which can be further analyzed to resolve the binding kinetics.

Compared to the binding kinetics of DFHBI binding to Spi which is relatively easy to resolve and has been discussed in previous literature (Kyu Young, Samie Jaffrey and French group), the binding kinetics of DFHBI binding to Spi-*n*R is much more complex, not only because Spi-*n*R has multiple DFHBI binding sites, but also due to the misfolding of Spinach arrays. In order to simplify the binding kinetics resolving, we only consider the binding reaction between DFHBI and Spinach aptamers, regardless of single Spinach or Spinach arrays. In this case the binding reaction would be described as below:

$$\mathrm{Spinach} + \mathrm{DFHBI} \rightleftharpoons \mathrm{Spinach} - \mathrm{DFHBI}$$

To measure the binding kinetics between DFHBI and Spinach aptamers, we transformed the concentrations of single Spinach or Spinach arrays into that of Spinach aptamers. For single Spinach, we assumed that the Spinach aptamer concentration is identical to that of Spi, and then we used the previously measured relative folding efficiency of Spinach arrays (43.5% for Spi-*8*R

and 34.6% for Spi-*32*R, compared to Spi) to calculate the Spinach aptamer concentration for Spi-*8*R and Spi-*32*R. To estimate the concentration of Spinach/DFHBI binding complex concentration, we compared the acquired fluorescence intensity of the sample to the equilibrium fluorescence intensity of the mixed solution of 100 nM Spi and 20 μM DFHBI, where we considered 100 nM Spinach/DFHBI complex formed upon reaction equilibrium.

In our experiments, the initial concentrations of the RNA, the effective Spinach aptamers and DFHBI in the experiments were: Spi, 40 nM, 40 nM Spinach aptamer; Spi-*8*R, 10 nM RNA, 34.8 nM Spinach aptamer; Spi-*32*R, 5 nM, 55.4 nM Spinach aptamer; DFHBI, 200 - 700 nM. We recorded the fluorescence intensity within 1800 s upon mixing, and transform the fluorescence intensity into binding complex concentration, to describe the binding complex formation as a function of time. Here we showed representative binding curves showing real-time binding reaction between DFHBI and Spi, Spi-*8*R and Spi-*32*R. We could notice that the binding complex formation, due to the binding reaction between DFHBI and either single Spinach or Spinach arrays, showed a fast increase state followed by a much slower increase state, before reaching the plateau, if there is any.

In order to analyze the data and retrieve binding kinetics information (k_{on}, k_{off} and K_D), we used a Matlab code to simulate our experimental data and determine the most probable k_{on} and k_{off} value (Methods and Materials) based on a simple kinetic model that could be described as below:

$$\mathrm{Spinach} + \mathrm{DFHBI} \rightleftharpoons \mathrm{Spinach} - \mathrm{DFHBI}$$

$$\frac{\mathrm{d}[\mathrm{Spinach} - \mathrm{DFHBI}]}{\mathrm{dt}} = k_{on} \times [\mathrm{Spinach}] \times [\mathrm{DFHBI}] - k_{off} \times [\mathrm{Spinach} - \mathrm{DFHBI}]$$

$$\frac{\mathrm{d}[\mathrm{Spinach}]}{\mathrm{dt}} = \frac{\mathrm{d}[\mathrm{DFHBI}]}{\mathrm{dt}} = k_{off} \times [\mathrm{Spinach} - \mathrm{DFHBI}] - k_{on} \times [\mathrm{Spinach}] \times [\mathrm{DFHBI}]$$

Initially we simulated the whole binding curve (1800 s), obtained the most probable k_{on} and k_{off} value for the specific experimental condition, and applied the predicted kinetic information to draw a simulated binding curve. We found that the result always showed great deviation from the experimental data at a very early stage (within 100 s, data not shown), suggesting that the binding curve could not be well described by our simplified kinetic model. Perhaps the existence

of the slow increase state of the binding complex formation shown in the binding curve could not be explained by our proposed kinetic model.

In this case we selectively simulated the first 100 s of the binding curve (Figure 3.4A), which is part of the fast increase state, and determined the corresponding most probable k_{on}, k_{off} and K_D value. We found that for the first 100 s of the binding curve, the simulated data fitted the experimental data quite well (Figure 3.4A), suggesting that our proposed kinetic model might be a good approximation to describe the fast increase state. We then used the predicted k_{on}, k_{off} and K_D value we obtained from the simulation of the first 100 s data from the binding curve to draw a predicted binding curve for the whole time window (1800 s) for our measurement (Figure 3.4B). We found that the simulated binding curve predicted a smaller DFHBI/Spinach concentration than the experimentally measured value at the plateau region (Figure 3.4B). Nevertheless, the predicted equilibrium binding complex concentration is still as high as 85% of measured equilibrium DFHBI/Spinach concentration for Spi and 80% for Spi-*8*R and Spi-*32*R. We hypothesized that additional conformational change of the aptamer resulted in a second slower binding state with potentially altered k_{on} and k_{off}. However, with the single observable in our experiment, we were unable to determine the parameters for the additional kinetic steps.

We found from our simulation result that for Spi, Spi-*8*R and Spi-*32*R, the k_{on}, k_{off} and K_D value were quite consistent at various DFHBI concentrations (200 - 700 nM). The average value of the k_{on} and k_{off} we calculated (Spi, $k_{on} = (8.3 \pm 0.1)\times 10^4\ M^{-1}\cdot s^{-1}$, $k_{off} = (2.5 \pm 0.4)\times 10^{-2}\ s^{-1}$; Spi-*8*R, $k_{on} = (9.3 \pm 0.6)\times 10^4\ M^{-1}\cdot s^{-1}$, $k_{off} = (2.3 \pm 0.6)\times 10^{-2}\ s^{-1}$; Spi-*32*R, $k_{on} = (6.6 \pm 0.8)\times 10^4\ M^{-1}\cdot s^{-1}$, $k_{off} = (3.3 \pm 0.3)\times 10^{-2}\ s^{-1}$) also indicated that the binding kinetics between DFHBI and the Spinach aptamer did not alter significantly between single Spinach (Spi) and tandem Spinach arrays (Spi-*8*R and Spi-*32*R).

We have included the calculated k_{on}, k_{off} and K_D value in Table 3.2.

3.3 Discussion

We constructed Spinach arrays and conducted *in vitro* characterization. We found that Spinach arrays could significantly increase fluorescence intensity of single Spinach aptamer. According to our result Spinach arrays could be useful for live cell RNA imaging.

3.4 Experimental procedure

3.4.1 Chemical synthesis of DFHBI

DFHBI was synthesized according to the procedure introduced by Paige and coworkers[91], and was dissolved in DMSO to 10 mM final concentration and the solution was stored at -20 °C.

3.4.2 Spi-tRNA and Spi sequence for *in vitro* fluorescence measurement

Spi-tRNA, containing one repeat of the 24-2 aptamer inside a $tRNA^{Lys}$ scaffold in the pET28c vector was given as a kind gift by Prof. Samie Jaffrey (Weill Cornell Medical College)[91]. The 24-2 sequence was amplified by PCR from pET28c-Spi-tRNA and inserted into pET28a plasmid (Novagen) between XbaI and HindIII restriction sites to create plasmid pET28a-Spi. Plasmids were transformed into *E. coli* strain BL21-DH5α (Promega) for DNA cloning.

3.4.3 Design and construction of Spinach arrays

The Spi-*8*R sequence was synthesized by Genscript as double-stranded DNA and was inserted into pUC57 plasmid (Genscript) between the EcoRI and HindIII restriction sites to generate the plasmid pUC57-Spi-*8*R. Each repeat sequence contains the Spinach (minimal) sequence, or 24-2-min sequence reported by Paige and coworkers[91], plus a 17 nt randomized linker sequence. The plasmid was digested with restriction endonuclease XbaI and HindIII to generate Spi-*8*R sequence, which was then inserted into pET28a plasmid between the XbaI and HindIII restriction sites to generate the plasmid pET28a-Spi-*8*R.

The construction method of Spi-*16*R, -*32*R and -*64*R was previously described by Golding and coworkers[91]. Briefly, we digested the pET28a-Spi-*8*R plasmid in two separate reactions- XbaI and HindIII were used to generate the insert fragment Spi-*8*R, and it was inserted into the pET28a-Spi-*8*R vector generated from NheI and HindIII digestion of pET-Spi-*8*R plasmid. As XbaI and NheI are isocaudomers and would generate the same sticky ends during restriction digestion, the Spi-*8*R fragment could be ligated with the pET28a-Spi-*8*R vector to obtain pET28a-Spi-*16*R. The similar cycle was repeated to create pET28a-Spi-*32*R and -*64*R.

3.4.4 *In vitro* transcription and RNA folding

Spi and Spi-*n*R (n = 8, 16, 32, 64) RNA were synthesized by *in vitro* transcription using MEGAshortscript T7 kit (Life Technologies) and MEGAscript T7 Kit (Life Technologies), respectively, following the protocol. pET28a-Spi and pET28a-Spi-*n*R DNA sequence linearized by HindIII were used as *in vitro* transcription template. The transcribed RNA underwent buffer exchange twice using P-6 micro bio-spin column (Bio-Rad) into RNA storage buffer (10 mM Tris acetate (pH 8.0), 0.1 mM EDTA and 10 mM KCl) to remove unreacted nucleotides. The RNA was folded in selection buffer (40 mM K-HEPES (pH 7.5) and 125mM KCl[91]) by incubation at 90 °C in water bath for 2 min, followed by slow cooling down to 65 °C, then supplemented with 5 mM $MgCl_2$ to assist RNA folding and further cooling down to room temperature.

3.4.5 Fluorescence measurement of *in vitro* transcribed RNA

Folded Spi and Spi-*n*R (*n* = 8, 16, 32, 64) RNA was diluted in selection buffer (40 mM K-HEPES (pH 7.5), 125 mM KCl) supplemented with 5 mM $MgCl_2$ and incubated with 20 µM DFHBI at room temperature (23 °C) for 5 min. Fluorescence measurement of the RNA/DFHBI complex was performed with a fluorometer (Cary Eclipse Fluorescence Spectrophotometer, Agilent Technologies) using the following instrument parameters: “Scan” mode; excitation wavelength 460 nm, excitation slit width 5 nm; emission wavelength 480 - 600 nm, emission slit width 5 nm.

To obtain the excitation and emission spectrum of Spi and Spi-*n*R RNA, the following instrument parameters were applied: (1) excitation spectrum, emission wavelength 505 nm and emission slit width 5 nm, excitation wavelength 400 - 480 nm and excitation slit width 5 nm; (2) emission spectrum, excitation wavelength 460 nm and excitation slit width 5 nm, emission wavelength 480 - 580 nm and emission slit width 5 nm.

3.4.6 Fluorescence enhancement efficiency measurement for Spinach arrays

We defined the fluorescence enhancement efficiency as the relative fluorescence intensity increment when the aptamer repeat number is doubled within the Spinach array. For example, upon aptamer number duplication, the value of the fluorescence enhancement efficiency for the Spinach array would be 0 if the fluorescence intensity remains unchanged and would be 1 if the

fluorescence intensity is doubled. The fluorescence intensity of Spi-*n*R, $I_{Spi\text{-}nR}$, as a function of the fluorescence enhancement efficiency, can be formulated as:

$$I_{Spi\text{-}nR} = I_{Spi} \times (1+eff)^{\log2(n)}$$

In the equation $I_{Spi\text{-}nR}$ is the *in-vitro* fluorescence of Spi-*n*R (Figure 3.1C), I_{Spi} is the *in-vitro* fluorescence of Spi (Figure 3.1C), "eff" is the fluorescence enhancement efficiency, and $\log_2(n)$ is the value of logarithms base 2 of *n*, which we further defined as duplication round, or the number of repeat duplication required to construct Spi-*n*R from Spi. We can transform the formula into:

$$\log_2(I_{Spi\text{-}nR}) = \log_2(1+eff) \times \log_2(n) + \log_2(I_{Spi})$$

The enhancement efficiency value can be obtained by plotting $\log_2(I_{Spi\text{-}nR})$ as a function of $\log_2(n)$ and conducting linear fitting to obtain the slope value, which is the $\log_2(1+eff)$ in the equation. Then the enhancement efficiency can be calculated with the slope value. We put the data plot and linear fitting in Figure 3.2, and had the slope value 0.67 and the fluorescence enhancement efficiency value 0.59.

3.4.7 Relative aptamer folding efficiency measurement of Spi-*n*R compared to Spi

To measure the relative folding efficiencies of Spi, Spi-*8*R and Spi-*32*R compared to Spi-tRNA, we kept the aptamer amount identical for Spi-tRNA, Spi, Spi-8R and Spi-32R, and incubated the RNA with excess amount of DFHBI. The solution was sent for fluorescence measurement and the fluorescence of Spi, Spi-*8*R and Spi-*32*R was compared with that of Spi-tRNA to calculate the relative folding efficiencies of Spi, Spi-*8*R and Spi-*32*R compared to Spi-tRNA. In the experiment we incubated 100 nM Spi-tRNA, 100 nM Spi, 12.5 nM Spi-*8*R or 3.125 nM Spi-*32*R with 10 μM DFHBI, respectively, making the aptamer concentration in each solution 100 nM and DFHBI 100 times the concentration of the aptamer.

3.4.8 Spinach binding kinetics measurement

The fluorescence increase was recorded with a fluorometer (Cary Eclipse) after rapid mixing of certain concentrations of Spi or Spi-*n*R with different concentrations of DFHBI to obtain a total volume of 100 μL in a cuvette for fluorescence measurement. The fluorescence intensity was

recorded as a function of time using excitation and emission wavelengths of 460 and 510 nm, respectively. The illumination intensity was estimated to be 1−5 W/cm^2.

3.4.9 Time-resolved fluorescence measurement and fluorescence lifetime calculation

The fluorescence lifetime measurements were performed by a custom built confocal microscope as described elsewhere[103]. An excitation light (473 ± 5 nm; ~ 10 ps pulse duration) was generated from an ultrafast laser (MaiTai HP, Spectra Physics) with the use of a photonic crystal fiber (FemtoWhite 800, NKT Photonics) and a pre-stretcher (N-SF57, CASIX). The fluorescence signal was detected by an avalanche photo diode (Micro Photon Devices) and registered by a time correlated single photon counting module (SPC630, Becker Hickl).

The extent of quenching (η) of Spi-*n*R compared with Spi-tRNA was calculated as follows. η = 1 - QY(Spi-*n*R)/QY(Spi-tRNA) = 1- τf(Spi-tRNA)/τf(Spi-*n*R), where QY and τf denote quantum yield and fluorescence lifetime, respectively.

3.4.10 Sequence information

Spi

gacgcaactgaatgaaatggtgaaggacgggtccaggtgtggctgcttcggcagtgcagcttgttgagtagagtgtgagctccgtaactag
tcgcgtc

Spi-tRNA

gcccggatagctcagtcggtagagcagcggccggacgcaactgaatgaaatggtgaaggacgggtccaggtgtggctgcttcggcagtg
cagcttgttgagtagagtgtgagctccgtaactagtcgcgtccggccgcgggtccagggttcaagtccctgttcgggcgcca

(red letters refer to tRNA sequence; green letters refer to Spinach aptamer sequence, or 24-2 sequence)

Spi-*8*R

tagacggcatggggagacgcgaccgaaatggtgaaggacgggtccagtgcttcggcactgttgagtagagtgtgagctccgtaactggtc
gcgtcacgtaagatgctccggttagggagacgcgaccgaaatggtgaaggacgggtccagtgcttcggcactgttgagtagagtgtgagc
tccgtaactggtcgcgtcactgatgtaccgttgagcagggagacgcgaccgaaatggtgaaggacgggtccagtgcttcggcactgttga
gtagagtgtgagctccgtaactggtcgcgtcactcgctagagcatggtttgggagacgcgaccgaaatggtgaaggacgggtccagtgctt
cggcactgttgagtagagtgtgagctccgtaactggtcgcgtcactggggcacgccgtctgggggagacgcgaccgaaatggtgaagga
cgggtccagtgcttcggcactgttgagtagagtgtgagctccgtaactggtcgcgtcacttactgcgaccgcaatagggagacgcgaccga
aatggtgaaggacgggtccagtgcttcggcactgttgagtagagtgtgagctccgtaactggtcgcgtcacgcgcgcaaccgggtagagg

gagacgcgaccgaaatggtgaaggacgggtccagtgcttcggcactgttgagtagagtgtgagctccgtaactggtcgcgtcacgtaactc
acggcgctatgggagacgcgaccgaaatggtgaaggacgggtccagtgcttcggcactgttgagtagagtgtgagctccgtaactggtcg
cgtc

(black letters refer to linker sequence between Spinach aptamers; green letters refer to Spinach aptamer (minimal) sequence, or 24-2-min sequence)

T7 promoter-*lac* operator

taatacgactcactatagggaattgtgagcggataacaattc

(red letters refer to T7 promoter sequence, and green letters refer to *lac* operator sequence)

3.5 Figures and tables

Figure 3.1

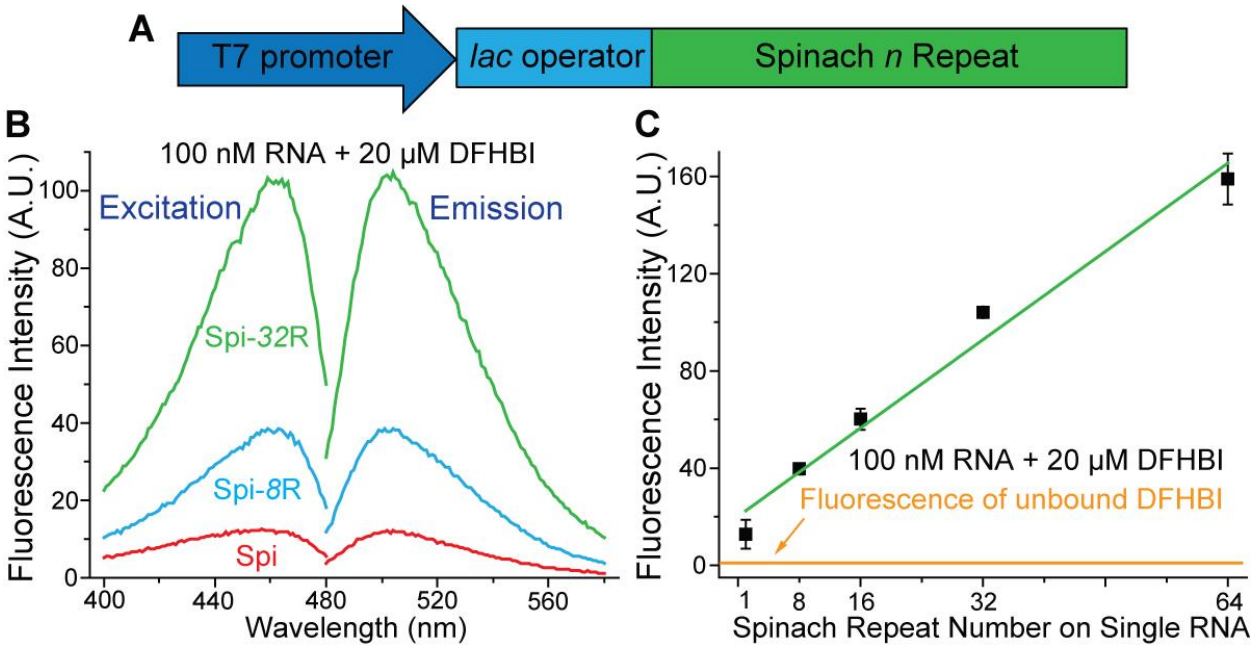

Figure 3.1. (A) Sketch of in vitro transcription system for Spinach arrays (Spi-nR). (B) Excitation and emission spectra of single Spinach aptamer and Spinach arrays (100 nM RNA + 20 μM DFHBI). (C) Fluorescence intensity of Spi and Spi-nR, measured by fluorometer.

Figure 3.2

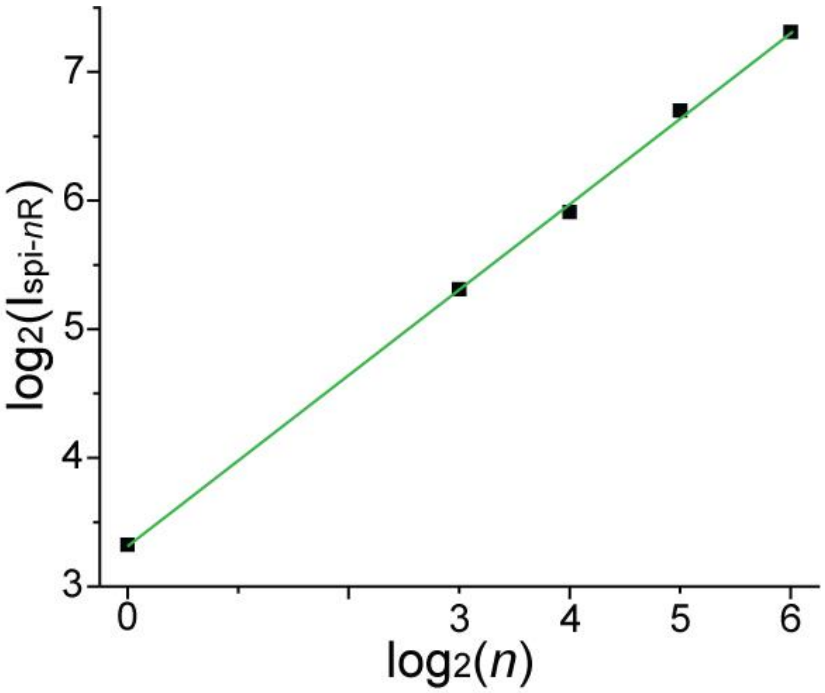

Figure 3.2. The plot of the in-vitro Spi-nR (ISpi-nR) fluorescence intensity as a function of the aptamer repeat number (n), after taking base 2 logarithm of both the fluorescence intensity and the repeat number value. The green line is the linear fitting of the data. The slope of the fitting curve, which is 0.67 here, is used to calculate the fluorescence enhancement efficiency value, which is used to describe the in-vitro fluorescence fold enhancement upon aptamer repeat number duplication within the Spinach array. The calculated efficiency value is 0.59 here.

Figure 3.3

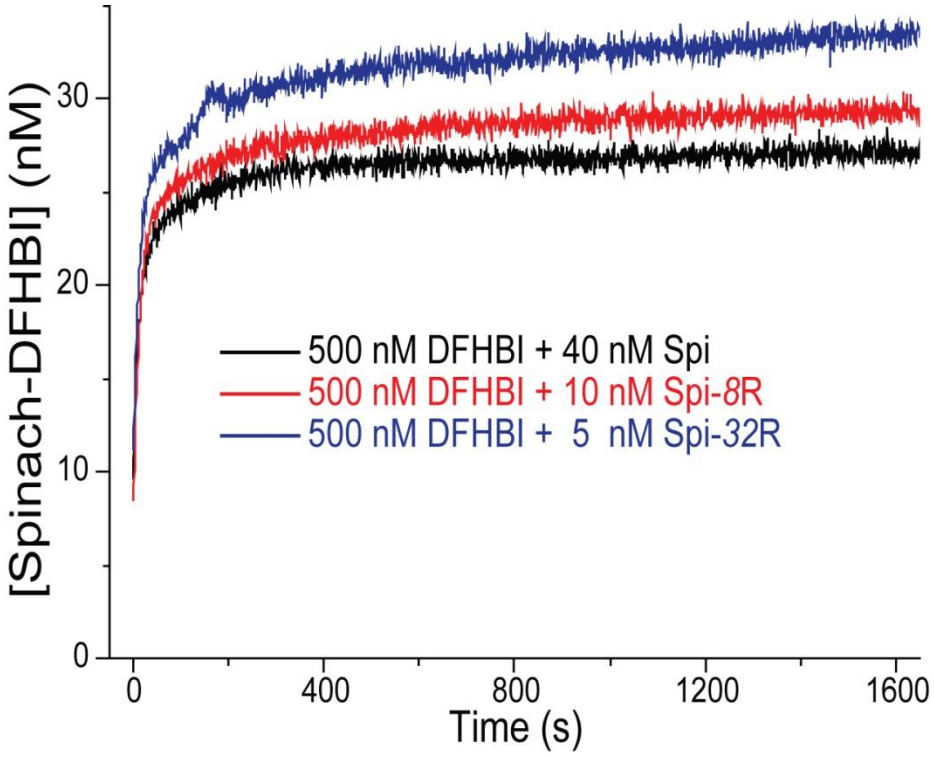

Figure 3.3. Representative Spinach/DFHBI binding curve for single Spinach aptamer (Spi) and tandem Spinach arrays (Spi-8R and Spi-32R).

Figure 3.4

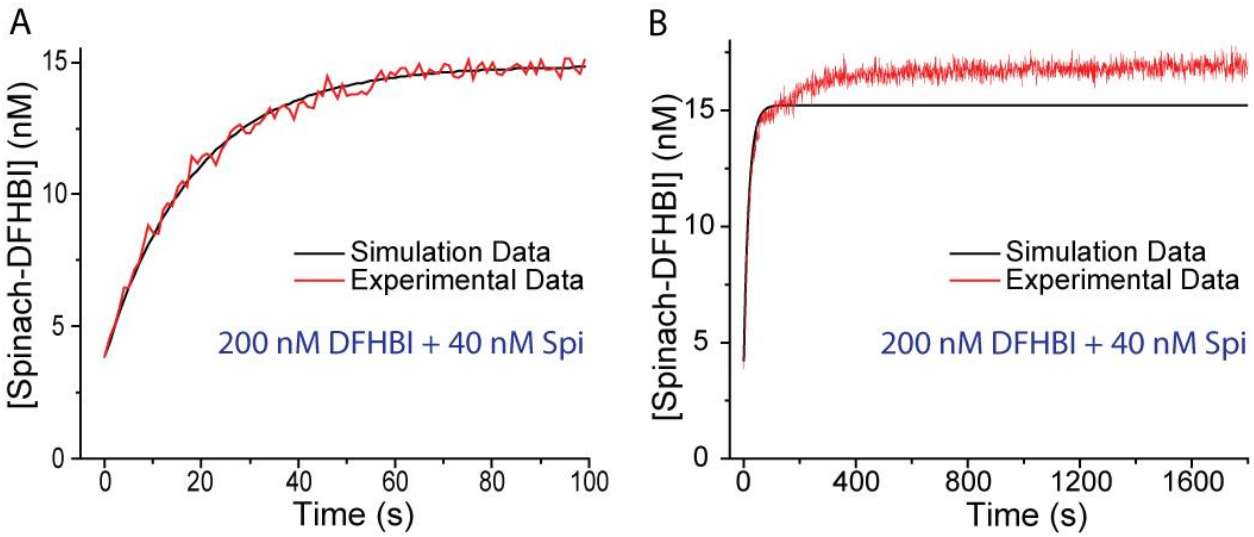

Figure 3.4. Representative Spinach/DFHBI binding curve (200 nM DFHBI + 40 nM Spi; red line) and the fitting curve using the estimated most probable kon and koff value (black line). (A) First 100 s of the binding curve and the fitting curve generated by simulating the first 100 s of the binding curve. (B) The full binding curve (1800 s) and the fitting curve using the estimated most probable kon and koff value generated from simulating the first 100 s of the binding curve.

Table 3.1

	RNA concentration (nM)	Aptamer concentration (nM)	Fluorescence (A. U.)	Relative folding efficiency
Spi	100	100	10.6 ± 0.9	100%
Spi-*8*R	12.5	100	4.61 ± 0.17	43.5 ± 1.6%
Spi-*32*R	3.125	100	3.67 ± 0.21	34.6 ± 2.0%

Table 3.1. The relative folding efficiencies of Spi and Spi-nR compared to Spi-tRNA.

Table 3.2

	k_{on} ($M^{-1}\cdot s^{-1}$)	k_{off} (s^{-1})	K_D (nM)
5 nM Spi-32R, 200 nM DFHBI	7.9×10^{4}	3.1×10^{-2}	392
5 nM Spi-32R, 300 nM DFHBI	8.7×10^{4}	2.7×10^{-2}	310
5 nM Spi-32R, 400 nM DFHBI	6.7×10^{4}	2.1×10^{-2}	313
5 nM Spi-32R, 500 nM DFHBI	7.7×10^{4}	3.5×10^{-2}	455
5 nM Spi-32R, 600 nM DFHBI	5.7×10^{4}	3.5×10^{-2}	614
5 nM Spi-32R, 700 nM DFHBI	6.9×10^{4}	2.9×10^{-2}	420
Mean ± SD	$(7.3 \pm 1.1)\times10^{4}$	$(3.0 \pm 0.5)\times10^{-2}$	417 ± 112
10 nM Spi-8R, 200 nM DFHBI	10.1×10^{4}	2.7×10^{-2}	267
10 nM Spi-8R, 300 nM DFHBI	9.8×10^{4}	2.8×10^{-2}	286
10 nM Spi-8R, 400 nM DFHBI	8.9×10^{4}	2.9×10^{-2}	326
10 nM Spi-8R, 500 nM DFHBI	9.3×10^{4}	1.7×10^{-2}	183
10 nM Spi-8R, 600 nM DFHBI	8.8×10^{4}	1.5×10^{-2}	170
Mean ± SD	$(9.4 \pm 0.6)\times10^{4}$	$(2.3 \pm 0.7)\times10^{-2}$	246 ± 67
40 nM Spi, 200 nM DFHBI	9.9×10^{4}	3.1×10^{-2}	313
40 nM Spi, 400 nM DFHBI	7.7×10^{4}	2.3×10^{-2}	299
40 nM Spi, 500 nM DFHBI	7.7×10^{4}	2.5×10^{-2}	325
40 nM Spi, 600 nM DFHBI	7.7×10^{4}	2.1×10^{-2}	273
Mean ± SD	$(8.3 \pm 1.1)\times10^{4}$	$(2.5 \pm 0.4)\times10^{-2}$	303 ± 22

Table 3.2. The kon, koff and KD of DFHBI binding onto the Spinach aptamer within Spi, Spi-8R and Spi-32R, measured under various DFHBI concentration conditions.

CHAPTER 4

LIVE CELL IMAGING USING SPINACH ARRAY*

4.1 Introduction

In this chapter we continued our study on Spinach systems by applying the Spinach arrays we constructed and tested *in vitro* to live cell RNA imaging. Upon introduction of Spinach arrays into the cell and fusing Spinach arrays to a protein-coding sequence, we also characterized whether Spinach arrays affect RNA transcription, protein translation or RNA degradation.

4.2 Results

4.2.1 Construction of RNA expression system in *E. coli*

In order to measure the fluorescence enhancement of Spinach array-tagged mRNA in living cells and characterize its influence on mRNA transcription, translation and degradation, we inserted Spinach arrays in the 3' UTR (untranslated region) of the monomeric red fluorescent protein, mRFP1, coding sequence (henceforth called RFP-Spi-*n*R (*n* = 8, 16, 32, 64) (Figure 4.1A). For comparison, we also prepared two constructs: RFP-Spi in which Spinach aptamer was directly linked to RFP sequence, and RFP-Spi-tRNA in which Spinach aptamer was incorporated into a human $tRNA^{Lys}$ scaffold as previously reported[91,104] and then linked to RFP sequence. The mRFP1 coding sequence without any Spinach tag, RFP, was used as a negative control.

4.2.2 Spinach arrays significantly enhance live cell RNA imaging fluorescence

We then measured how much brighter Spinach arrays are compared to single Spinach in living cells. Live cell imaging of *E. coli* transcribing RFP-Spi showed no fluorescence enhancement in the Spinach fluorescence channel over the background autofluorescence level of cells expressing

---

*Chapter 4 was excerpted from a published work:

Zhang, J. et al. "Tandem Spinach array for RNA imaging in living bacterial cells" Sci. Rep. doi:10.1038/srep17295 (2015)

untagged RFP (Figure 4.2). In comparison, cells transcribing RFP-Spi-tRNA showed homogenously distributed Spinach fluorescence with a slight enhancement over the autofluorescence level (Figures 4.1C, 4.1D and 4.2). The fluorescence deficiency in cells transcribing RFP-Spi probably results from misfolding of the Spinach aptamer when fused to an mRNA. In contrast, *E. coli* transcribing RFP-Spi-*n*R showed strong Spinach fluorescence. The increase in average fluorescence intensity per cell as a function of *n* (~17 fold enhancement from RFP-Spi-tRNA to RFP-Spi-*64*R, after autofluorescence subtraction) generally reflects the *in vitro* trend (~17 fold enhancement from Spi to Spi-*64*R).

Interestingly, we found that unlike the homogenously distributed cellular fluorescence observed for *E. coli* expressing RFP-Spi-tRNA, the fluorescence for *E. coli* expressing RFP-Spi-*n*R preferentially accumulated at cell poles. In order to determine whether the apparent difference in cellular fluorescence distribution is caused by Spinach arrays, we conducted an RNA fluorescence *in situ* hybridization (FISH) experiment, which is a widely applied method to characterize target RNA localization and expression level, on cells expressing RFP, RFP-Spi-tRNA and RFP-Spi-*n*R (*n* = 8, 32). We used Cy5-labeled oligonucleotide probes targeting the mRFP1 (Figure 4.3) or the Spinach array (Figure 4.4) in different sets of FISH experiments. Fluorescence images of the FISH samples indicated that upon IPTG induction, target mRNAs preferentially accumulated at the cell poles where mRNAs transcribed from plasmid DNA are typically observed in *E.coli*[105], regardless of the presence or the types of the Spinach tag at the 3' UTR (Figure 4.3). Therefore, we conclude that the low fluorescence from single Spinach failed to reveal accurate mRNA localization (Figures 4.1 and 4.5) and the Spinach array could address the problem by greatly enhancing fluorescence signal for cellular mRNA imaging.

4.2.3 Spinach array does not affect mRNA transcription and protein translation

Quantitative reverse transcription PCR (qPCR) was conducted to quantify the expression level of mRFP1-coding mRNA for cells transcribing RFP, RFP-Spi-tRNA and RFP-Spi-*n*R. We found in qPCR data that the cellular level of Spinach-tagged mRNA under induction (1 mM IPTG, 60 min), either by a single Spinach (RFP-Spi-tRNA) or by an array (RFP-Spi-*n*R), is similar to that of untagged mRNA (Figure 4.1B), suggesting that mRNA transcription is unperturbed by the Spinach tag even with the longest array (Spi-*64*R) tested.

In addition, we measured the average fluorescence intensities of mRFP1 protein per cell, translated from Spinach-tagged or untagged mRNAs, in cells expressing RFP, RFP-Spi-tRNA and RFP-Spi-*n*R, and all types of cells showed comparable mRFP1 protein fluorescence (Figures 4.1C and 4.1D), indicating that mRNA translation is not affected by Spinach tag either as a single aptamer or as an array.

4.2.4 Spinach array does not affect mRNA degradation

Here we examined the effect of Spinach arrays on RNA degradation. We conducted an mRNA decay assay for *E. coli* cells expressing RFP-Spi-*32*R or untagged RFP. We monitored the fluorescence of RFP-Spi-*32*R mRNA and mRFP1 protein after IPTG removal which would stop synthesis of new RNA. Spinach fluorescence levels greatly decreased within 30 min and were depleted by 90 min after withdrawal of IPTG, suggesting mRNA degradation (Figures 4.6A and 4.6B). We also compared mRFP1 fluorescence between *E. coli* expressing RFP-Spi-*32*R and untagged RFP (Figures 4.6A and 4.7). In both cases the fluorescence drop showed a significant time lag behind Spinach fluorescence change, which is possibly resulted from the combination of the following: (1) continuous translation of mRFP1 from the remaining mRNA after IPTG removal, (2) much longer lifetime of protein compared to that of mRNA[106-108], and/or (3) mRFP1 maturation[109,110] before fluorescing. We observed very similar trends of mRFP1 fluorescence intensity change in cells expressing RFP-Spi-*32*R vs. RFP (Figure 4.7), indicating that the Spinach tag does not affect mRNA degradation or translation kinetics. To further confirm the effect of Spinach array on mRNA degradation, we conducted qPCR experiment to measure the abundance of untagged RFP mRNA and RFP-Spi-*n*R (n = 8, 32) as a function of time after IPTG removal (Figure 4.6C). qPCR data showed that RFP-Spi-*n*R mRNA level was decreased by approximately 60% and 90% 30 and 120 min after IPTG removal, respectively, with the same trend observed for the untagged RFP. We therefore conclude that Spinach tag even with multiple repeats does not influence mRNA decay.

4.2.5 Image cellular mRNA at low expression level using laser excitation and pulsed illumination strategy

As the Spinach array effectively enhanced the fluorescence signal compared to single Spinach tag and performed well in quantitatively reporting the mRNA abundance in live cells through imaging, we then tuned the mRNA transcription level to further characterize the imaging sensitivity of the Spinach array. We replaced the T7 promoter (P_{T7}-RFP-Spi-*32*R) with a native *lacZYA* promoter (P_{lac}) for RFP-Spi-32R transcription (P_{lac}-RFP-Spi-*32*R). RNA synthesis by endogenous *E. coli* RNA polymerase instead of T7 polymerase reduced the mRNA expression level by approximately two orders of magnitude according to qPCR quantification (Figure 4.8B). Using 16S ribosomal RNA (16S rRNA) (~20,000 - 70,000 copies per *E. coli* cell[38,111]) as a reference, we estimated the copy number of RFP-Spi-32R mRNA transcribed under the control of P_{lac} and P_{T7} to be ~50 - 180 and ~3,000 - 11,000 per cell (Figure 4.8B), respectively, which is consistent with the transcription levels for the two expression systems[112,113]. We first conducted epifluorescence microscopy on cells expressing P_{lac}-RFP-Spi-*32*R with the same imaging condition applied for P_{T7}-RFP-Spi-*32*R. P_{lac}-RFP-Spi-*32*R did not show any fluorescence signal in Spinach fluorescence channel beyond background autofluorescence level of uninduced P_{lac}-RFP-Spi-*32*R (Figure 4.9), indicating that the epifluorescence microscope we used here is not sensitive enough to detect fluorescence signal from Spinach array with ~50 - 180 RNA copies per cell.

In order to obtain higher fluorescence signal from the Spinach array, we used a 473 nm laser instead of the lamp light as the excitation source. Previous study showed that the DFHBI bound to the Spinach aptamer quickly dissociates (within ~ 100 ms) upon strong excitation, causing the loss of fluorescence[103]. To address the problem, we utilized pulsed laser excitation to allow for Spinach to rebind DFHBI after light-induced DFHBI dissociation and regain fluorescence (Figure 4.8A). An automatically controlled mechanical shutter was applied to the generate a 0.2 Hz repetitive laser pulse with 50 ms pulse width[103] (Experimental Procedure).

During sample imaging, we first used continuous-wave (CW) laser to illuminate the sample for 10 s to reduce the cellular autofluorescence level (pre-photobeaching, or pre-PB) (Figure 4.8A and 4.10)[114]. Afterwards we used the pulsed laser to illuminate the sample and acquire the imaging data. With the help of pulsed illumination, we could observe clear distinction between uninduced P_{lac}-RFP-Spi-*32*R (~4 - 15 mRNA copies per *E. coli* cell), where we hardly observed any fluorescence signal (Figure 4.8D), and induced P_{lac}-RFP-Spi-*32*R (~ 50 - 180 mRNA copies

per *E. coli* cell), where we found bright spots in many cells residing at cell poles (Figure 4.8C and 4.8D). We attribute the bright spots to transcription sites containing multiple mRNAs. The bright spots disappeared within a very short illumination time (< 500 ms) and reemerged upon illumination reinstatement after its withdrawal (> 5 s) (Figure 4.8C), which is a typical optical characteristics of Spinach fluorescence[103], suggesting that the fluorescence originated from the Spinach array. If we superposed multiple cell images from repeated cycles to achieve stronger fluorescence, we could clearly find that fluorescence signal showed preferential localization at cell poles (Figure 4.8D). The fluorescence consistent with the observed mRNA localization of induced P_{lac}-RFP-Spi-*32*R verified by RNA FISH (Figure 4.11). Overall, with the help of the pulsed illumination method, we further enhanced the performance of the Spinach array and showed the potential to apply the system to image lower abundance cellular RNA.

4.3 Discussion

The Spinach system is a recently developed RNA labeling and imaging method based on aptamer binding and fluorescence induction of the fluorogenic small molecule DFHBI**[91]**. It has many potential advantages over widely applied RNA labeling methods using fluorescent protein-fused RNA binding proteins, such as low fluorescence background, elimination of separate introduction of RNA binding proteins, and evasion of perturbation on target RNAs by protein binding. However, there were few reports on single Spinach aptamer labeling for cellular mRNA imaging**[95]**, likely due to its low fluorescence brightness suggested by our experimental data (Figure 4.1D). To address this problem, we employed a tandem Spinach array to tag a single mRNA molecule and demonstrated that the Spinach array containing 64 aptamer repeats can enhance the fluorescence by 17-fold in live cells compared to a single Spinach aptamer. The fluorescence enhancement by the Spinach array allowed us to greatly improve mRNA imaging quality comparing to using the single Spinach tag. In particular, we observed inhomogenous RNA distribution and distinct RNA loci in E. *coli* (Figures 4.5D and 4.8D) using the Spinach array, whereas single Spinach tag suffered from low fluorescent signal and failed to report correct RNA localizations in cells. For cells with a lower mRNA level (~120 - 180 mRNAs per E. *coli*.), we further applied pulsed illumination strategy to effectively boost the fluorescence of the Spinach array such that we could observe mRNA localizations. Despite the 17-fold fluorescence enhancement achieved by constructing the aptamer array, we also noted that the average

efficiency for an aptamer to correctly fold seems to decrease with the increase of the aptamer repeat number. One possible reason for this is the crosstalk and mispairing between adjacent or spatially close aptamers in a tandem array, which were reported and discussed for tandem arrays in previous RNA and protein folding studies**[115-118]**. In addition, the intrinsic instability of the Spinach aptamer may also play a role (~32% folded, 25°C**[92]**). Future introduction of several recently reported aptamers with improved folding efficiency (for example, Spinach 2, ~58%**[92]**; Broccoli, ~60%**[94]**) may enhance the fluorescence of the aptamer array. It is also possible that optimizing linker sequences may improve the performance although mispairing-based misfolding may not be avoided by changing linker sequences alone.

With a series of characterizations on the aptamer tandem array, we demonstrated that in *E. coli*, Spinach array-tagged mRNA had no significant alterations on transcription, translation or degradation. This may be attributed to the small size and high dissociation constant (K_D) of the fluorogen. In contrast, the MS2/PP7 coat protein labeling method has been reported to impede RNA degradation in bacteria in previous studiess[37,38], likely due to the stable association of many MS2/PP7 proteins to mRNA, which prevents the bacterial RNA degradation machinery from functioning. In addition, the Spinach array did not alter mRNA localization. When applied to eukaryotic cell imaging, in order to decrease the fluorescence background introduced by the MS2 coat protein-fluorescent protein fusion (MS2-FP) in cytosol, the MS2-FP proteins are usually fused with a nuclear localization sequence (NLS) to guide excess unbound proteins to the nucleus[22,98]. Although the strategy increases the signal-to-noise ratio, it might potentially perturb the endogenous RNA localization. In contrast, DFHBI remains non-fluorescent until binding to the Spinach aptamer, which circumvents extra modifications that decrease background fluorescence but could possibly affect target RNA localization.

In conclusion, we constructed a tandem Spinach aptamer array that could enhance fluorescence imaging quality of the Spinach/DFHBI system for live cell RNA imaging while introducing minimal perturbation to the target RNA. Nevertheless, Spinach arrays still show several limitations, including relatively low brightness and large size, making them ill-suited for studying mammalian cells, where autofluorescence is stronger and RNAs are more dispersedly distributed and undergo intensive motor-driven transport. Several recently developed fluorogen/aptamer systems, including RNA Mango, Spinach 2, Broccoli, *etc*, showed the

potential to further enhance the performance. All of those newly discovered fluorogen/aptamer systems have similar fluorescing mechanisms as that of the Spinach aptamer while showing higher brightness, shorter length or more robust folding behavior. We envision that incorporating the newly reported aptamers into a tandem array could also enhance their fluorescence signals per RNA molecule. Furthermore, we expect the tandem arrays derived from those aptamers would be very likely to bring minimal perturbation to their target RNAs, due to their similarities to the Spinach aptamer.

4.4 Experimental procedure

4.4.1 Design and construction of Spinach arrays and RFP-Spi, -Spi-tRNA and -Spi-*n*R

The mRFP1-coding sequence was amplified by PCR from pTRUEBLUE-BAC2-$P_{lac/ara}$-mRFP1-96BS constructed by Golding and coworkers[38] , and was inserted into pET28a between NcoI and NdeI to generate pET28a-RFP plasmid. The further construction of pET28a-RFP-Spi, -Spi-tRNA, -Spi-*n*R plasmid was accomplished by inserting the Spinach sequence flanking XbaI and HindIII sticky ends into pET28a-RFP plasmid between NheI and HindIII restriction sites. Plasmids were transformed into BL21-DH5α (Promega) and BL21-DE3-Rosetta (Novagen) *E. coli* strains for DNA cloning and live cell imaging, respectively.

To replace T7 promoter in pET28a-RFP-Spi-*32*R plasmid, a *lacZYA* (*lac*) promoter-*lac* operator sequence, amplified by PCR from pUC57-Spi-*8*R (Genscript), flanking BglII and XbaI sticky ends was inserted into the plasmid between BglII and XbaI restriction sites.

4.4.2 Bacteria growth and induction

E. coli cells were grown at 37 °C with antibiotics according to the plasmid selection markers (100 μg/mL ampicillin (Gold Biotechnology, Inc) for pUC57 and pUC57-Simple plasmid, 50 μg/mL Kanamycin (Roche Diagnostics) for pET28a and pET28c plasmid, 30 μg/mL Chloramphenicol (Sigma-Aldrich) for pTRUEBLUE-BAC2 plasmid, and 30 μg/mL Chloramphenicol for *E. coli* Rosetta strain) in Lysogeny Broth (LB) (LB Broth Miller, EMD Millipore) liquid and solid media. To gauge cell density, optical density (OD) of the medium was assayed at 600 nm using a plastic cuvette in a Spectramax Plus 384 Microplate Reader

(Molecular Devices, Inc.). For RNA and protein induction, Rosetta cells transformed with a given plasmid were grown at 37 °C in LB medium overnight from a single colony, and diluted 1000-fold into fresh LB medium and kept growing until at OD_{600} = 0.2. IPTG (Sigma-Aldrich) was then supplemented with 1 mM final concentration to induce RNA and protein production under T7 promoter-*lac* operator control. To maintain exponential growth of the cells, pre-warmed medium was added to dilute the cell culture to OD_{600} = 0.3 whenever OD_{600} of the culture exceeded 0.5.

4.4.3 RNA decay assay

Cells were grown in LB medium and induced by IPTG as described above. After 60 min induction by 1 mM IPTG, the cells were centrifuged to a pellet and the supernatant was carefully aspirated and the cell pellet was resuspended in fresh pre-warmed LB medium without IPTG. The centrifugation and resuspension process was performed twice to remove the remaining IPTG, and the cell pellet was finally resuspended in LB medium without IPTG. The cells were grown at 37 °C, and at different time points after IPTG removal, a bit of cell culture was taken out for imaging.

4.4.4 Epifluorescence microscopy and image analysis

To prepare the imaging sample, 1 mL of cell culture was supplemented with DFHBI to 100 μM final concentration 10 min before imaging, and was kept growing at 37 °C for DFHBI permeation and binding to Spinach aptamer. The cells were then centrifuged and supernatant was removed. After cell resuspension in pre-warmed M9 minimal medium (M9, Minimal Salts, 5X; Sigma-Aldrich) supplemented with 2mM $MgCl_2$ and 100 μM DFHBI, a few μL were sandwiched between a glass coverslip (No. 1.5) and a thin slab of 1.5 % (w/v) agarose gel. M9 minimum medium containing 2mM $MgCl_2$ and 100 μM DFHBI or 1x PBS medium was used to dissolve the agarose and make the gel for live cell imaging or FISH imaging, respectively.

All epifluorescence images were taken with a Nikon Eclipse (TE-2000-U, Nikon) microscope equipped an oil immersion objective (1.3 NA 100x) and an epifluorescence system. The epifluorescence system used a lamp light source (X-Cite Series 1200, Excelitas Technologies) to illuminate the sample and an emCCD camera (iXon3 897, Andor Technology) to acquire the

fluorescence image. The filter sets applied were: Brightfield (exposure time 100 ms), Spinach (Ex 450-490 nm, Em 500-550 nm, exposure time 200 ms), mRFP1 protein (Ex 540-580 nm, Em 593-668 nm, exposure time 10 ms), Cy5 FISH probe (Ex 590-650 nm, Em 663-738 nm, exposure time 100 ms). The images were processed and analyzed using a MATLAB code reported by Golding and coworkers[38], which is able to identify single cells and measure the average fluorescence level of each cell in different fluorescence channels.

4.4.5 Pulsed illumination microscopy and image analysis

Pulsed illumination imaging was conducted by a home-built objective-TIRF microscope with an oil immersion objective (1.4 NA 100x, Olympus) equipped with an emCCD camera (iXon DU-887, Andor Technology). Illumination with a 473 nm laser (MLL-III-473, Opto Engine LLC) was controlled by a mechanical shutter (UniBlitz VMM-D3) through a National Instruments NI-6503 digital I/O controller card and synchronized to CCD via home-built software (cplc.illinois.edu/software)[2], and the data acquisition and analysis procedures conducted. The recorded movie was processed by a MATLAB code to generate a collection of fluorescence images of consecutive single frames. The superposed image was conducted by stacking selected frames in ImageJ.

4.4.6 Total RNA extraction and purification

For each sample, the cell culture was measured OD_{600}. We took out appropriate volume of the cell culture which contained the same total cell number of 1 mL 0.5 OD_{600} cells. The cells were centrifuged at 5,000 g and 4 °C for 5 min, and the supernatant was carefully aspirated. The cell pellet was lysed by 1 mM lysozyme/TE buffer (10 mM Tris:HCl, 1 mM EDTA, pH = 8.0) and total RNA was extracted and purified from the cell lysate through RNeasy Mini Kit (Qiagen) according to the protocol. 10 µg extracted RNA were further treated with DNase using Turbo DNA-free kit (Life Technologies) to remove remaining DNA which interfered with qPCR experiments. Afterwards reverse transcription reaction was conducted with 100 ng RNA in a 20 µL reaction volume to synthesize cDNA required for qRT-PCR experiments using iScript cDNA Synthesis Kit (Bio-Rad) from the DNAse-treated RNA, according to the protocol.

4.4.7 Quantitative reverse transcription PCR (qPCR) and transcript number estimation

1 µL reverse-transcribed cDNA were taken out of the 20 µL total volume of each reverse transcription reaction, and were diluted to 10 µL. Regular PCR reactions were first conducted using the diluted cDNA and designed qPCR primers to confirm that proper cDNA products were generated. After that 1 µL of the diluted cDNA were supplemented with qPCR primer and reaction and detection SsoAdvanced SYBR Green Supermix (Bio-Rad) to a 20 µL reaction volume and qPCR reactions were assembled in a 96-well PCR plate (Bio-Rad). qPCR primers targeting the mRFP1-coding sequence were used to quantify RFP-Spi-*n*R RNA, and primers targeting 16S ribosomal RNA were used to quantify 16S ribosomal RNA as the internal standard. The qPCR reactions were conducted and monitored by Bio-Rad CFX96 Touch Real-Time PCR Detection System.

The expression level and the cellular transcript number of RFP or RFP-Spi-*n*R mRNA were roughly estimated by calculating the relative RNA expression level compared to that of 16S rRNA, using ΔC_T Method[119]. We have made a series of known dilutions of the cDNA samples and created standard curves of mRFP1 and 16S rRNA qPCR primers by plotting the threshold-crossing cycle number (C_T) of the amplification curves, to estimate the amplification efficiencies of both primer pairs, and to confirm that their amplification efficiencies are similar and reliable (between 90% and 105%). The relative expression level of a specific sample between RFP (-Spi-*n*R) mRNA compared to that of 16S rRNA can be calculated simply with the primer amplification efficiencies extrapolated from the standard curves and the C_T values measured by the qPCR experiments. The approximate RNA copy numbers per cell for RFP (-Spi-*n*R) mRNA were estimated by translating the relative value into absolute copy number using the value of 20,000 - 70,000 16S rRNA molecules reported by previous study[38,111,120].

4.4.8 RNA fluorescence *in situ* hybridization (FISH)

The FISH probes, which are DNA oligonucleotides with 3’ amine modification, were designed and ordered from Biosearch Technologies, and labeled with Cy5 NHS (GE Healthcare). The protocol of RNA FISH, including probe design and labeling, cell fixation and permeabilization, and probe hybridization, were reported by So and coworkers[19,121]. Cell preparation and sample hybridization are briefly described below.

After cell harvest (equivalent to 15 mL OD_{600} = 0.4) and centrifugation (4 °C, 4500x *g*, 8 min), supernatant was removed and cells were resuspended in 1 mL freshly prepared fixation solution (1x PBS, 3.7% (w/w) formaldehyde) and gently shaken at room temperature for 30 min. Centrifuge (400x *g*, 8 min) the cell suspension, remove supernatant, and then wash twice in 1 mL 1x PBS. Resuspend the cells in 70% (w/v) Ethanol, and leave the cell suspension at room temperature for at least 1h to permeate the cell. Afterwards centrifuge (600x *g*, 7 min) and remove supernatant. Resuspend cells in 1 mL 40% wash buffer (40% formamide (v/v), 2x SSC) and leave at room temperature for a few minutes, and then centrifuge (600x *g*, 7 min) and remove supernatant. Resuspend the cells well in 50 µL hybridization solution, which is mixed with 195 ng probes (15 ng for each probe, 13 probes in total) and 50 µL 40% hybridization buffer (10% dextran sulfate (w/v), 40% formamide (v/v), 1 mg/mL *E. coli* tRNA, 2x SSC, 0.2 mg/mL BSA, 2 mM Ribonucleoside Vanadyl Complex), and leave at 30 °C overnight. On the next day, take a few µL of hybridization sample, add 20 volumes of 40% wash buffer, followed by mixing and centrifugation (600x *g*, 7 min). Afterwards repeat the following steps 3 times: resuspend the cells with 20 volumes 40% wash buffer after removing supernatant, incubate for 30 min at 30 °C, and centrifuge and remove supernatant. Finally resuspend the cells in 1 volume of 2x SSC and the cell resuspension are ready to image. Prepare 1.5% (w/v) agarose/PBS gel for imaging sample preparation, as described before. Cell sample was imaged using epifluorescence microscopy with the filter set mentioned above for Cy5 FISH probe.

4.4.9 Sequence information

Spi

gacgcaactgaatgaaatggtgaaggacgggtccaggtgtggctgcttcggcagtgcagcttgttgagtagagtgtgagctccgtaactag tcgcgtc

Spi-tRNA

gcccggatagctcagtcggtagagcagcggccggacgcaactgaatgaaatggtgaaggacgggtccaggtgtggctgcttcggcagtg cagcttgttgagtagagtgtgagctccgtaactagtcgcgtccggccgcgggtccagggttcaagtccctgttcgggcgcca

(red letters refer to tRNA sequence; green letters refer to Spinach aptamer sequence, or 24-2 sequence)

Spi-*8*R

tagacggcatggggagacgcgaccgaaatggtgaaggacgggtccagtgcttcggcactgttgagtagagtgtgagctccgtaactggtc
gcgtcacgtaagatgctccggttagggagacgcgaccgaaatggtgaaggacgggtccagtgcttcggcactgttgagtagagtgtgagc
tccgtaactggtcgcgtcactgatgtaccgttgagcagggagacgcgaccgaaatggtgaaggacgggtccagtgcttcggcactgttga
gtagagtgtgagctccgtaactggtcgcgtcactcgctagagcatggtttgggagacgcgaccgaaatggtgaaggacgggtccagtgctt
cggcactgttgagtagagtgtgagctccgtaactggtcgcgtcactggggcacgccgtctgggggagacgcgaccgaaatggtgaagga
cgggtccagtgcttcggcactgttgagtagagtgtgagctccgtaactggtcgcgtcacttactgcgaccgcaatagggagacgcgaccga
aatggtgaaggacgggtccagtgcttcggcactgttgagtagagtgtgagctccgtaactggtcgcgtcacgcgcgcaaccgggtagagg
gagacgcgaccgaaatggtgaaggacgggtccagtgcttcggcactgttgagtagagtgtgagctccgtaactggtcgcgtcacgtaactc
acggcgctatgggagacgcgaccgaaatggtgaaggacgggtccagtgcttcggcactgttgagtagagtgtgagctccgtaactggtcg
cgtc

(black letters refer to linker sequence between Spinach aptamers; green letters refer to Spinach aptamer (minimal) sequence, or 24-2-min sequence)

mRFP1

atggcctcctccgaggacgtcatcaaggagttcatgcgcttcaaggtgcgcatggagggctccgtgacacgagttcgagatcagagggcg
agggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggcggccccctgcccttcgcctgggacatcctgtcccc
tcagttccagtacggctccaaggcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgg
gagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaa
gctgcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctccaccgagcggatgtacccc
gaggacggcgccctgaagggcgagatcaagatgaggctgaagctgaaggacggcggccactacgacgccgaggtcaagaccaccta
catggccaagaagcccgtgcagctgcccggcgcctacaagaccgacatcaagctggacatcacctcccacaacgaggactacaccatcg
tggaacagtacgagcgcgccgagggccgccactccaccggcgcctaa

T7 promoter-*lac* operator

taatacgactcactataggggaattgtgagcggataacaattc

(red letters refer to T7 promoter sequence, and green letters refer to *lac* operator sequence)

***lac* promoter-*lac* operator**

tttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttc

(blue letters refer to *lacZYA* promoter sequence, and green letters refer to *lac* operator sequence)

RNA FISH probes

mRFP1 coding sequence

5’-tgatgacgtcctcggaggag-3’
5’-accttgaagcgcatgaactc-3’
5’-tgatctcgaactcgtgtcac-3’
5’-tactggaactgaggggacag-3’
5’-gggaaggacagcttcaagta-3’

5’-cgtcctcgaagttcatcacg-3’
5’-gcttcaccttgtagatgaac-3’
5’-atggtcttcttctgcattac-3’
5’-ttcagcctcatcttgatctc-3’
5’-atgtaggtggtcttgacctc-3’
5’-cttgatgtcggtcttgtagg-3’
5’-tcgttgtgggaggtgatgtc-3’
3’-atgtaggtggtcttgacctc-3’

Spinach aptamer

5’-aagcactggacccgtccttc-3’
5’-cacactctactcaacagtgc-3’
5’-tgacgcgaccagttacggag-3’

qPCR primers

mRFP1

Forward: 5’-tgaggctgaagctgaaggac-3’
Reverse: 5’-tgtccagcttgatgtcggtc-3’

16S rRNA

Forward: 5’-aggccttcgggttgtaaagt-3’
Reverse: 5’-attccgattaacgcttgcac-3’

4.5 Figures

Figure 4.1

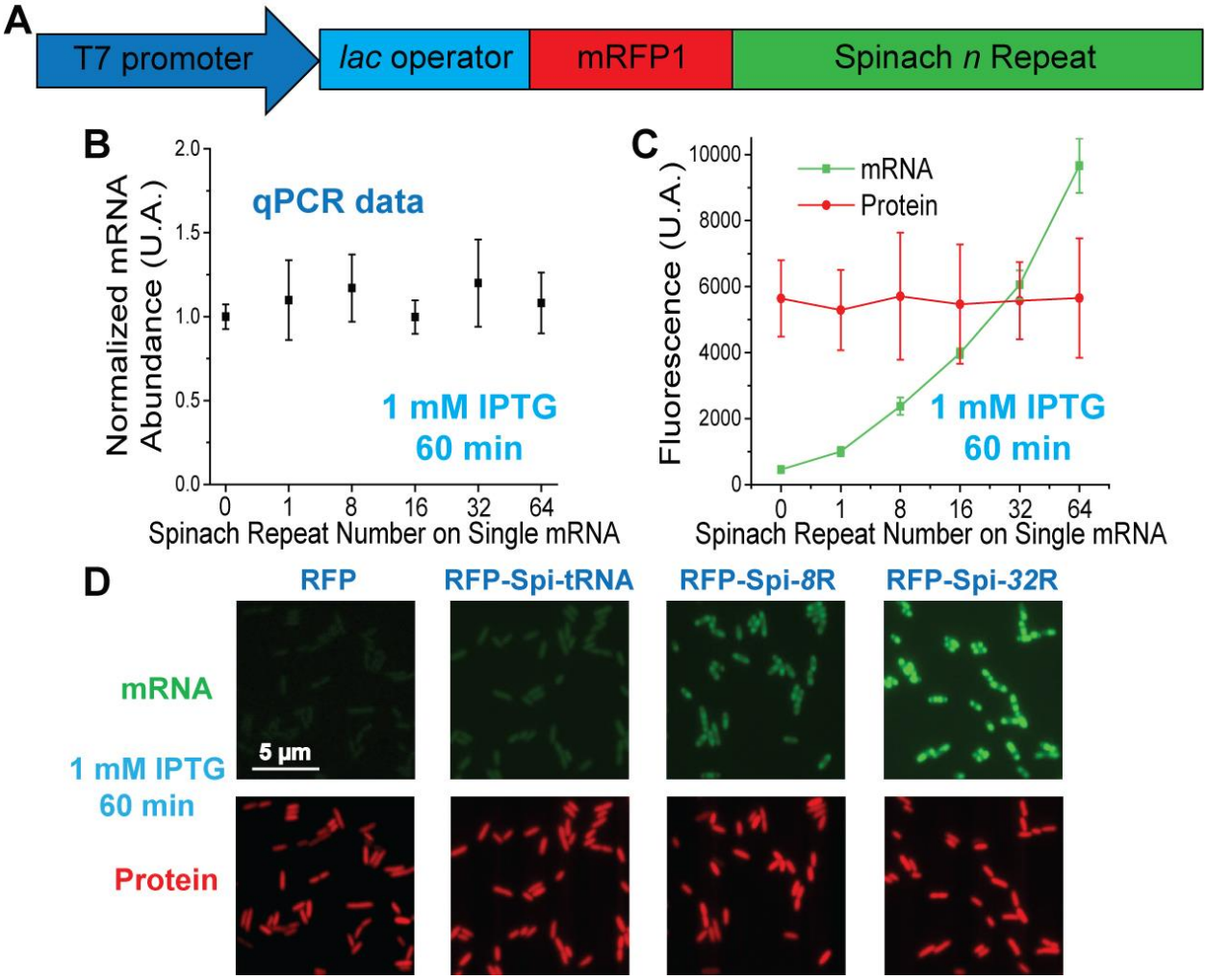

Figure 4.1. (A) Sketch of RFP-Spi-*n*R expression system in *E. coli*. (B) mRNA expression level of RFP or RFP-Spi-*n*R in *E. coli* after 60 min of 1 mM IPTG induction, measured by qPCR and normalized by the mRNA level of unmodified RFP in *E. coli* . (C) mRNA (Spinach) and protein (mRFP1) fluorescence in *E. coli* expressing RFP or RFP-Spi-*n*R upon induction, measured via epifluorescence imaging. (D) Representative fluorescence images of *E. coli* expressing RFP or RFP-Spi-*n*R upon induction.

Figure 4.2

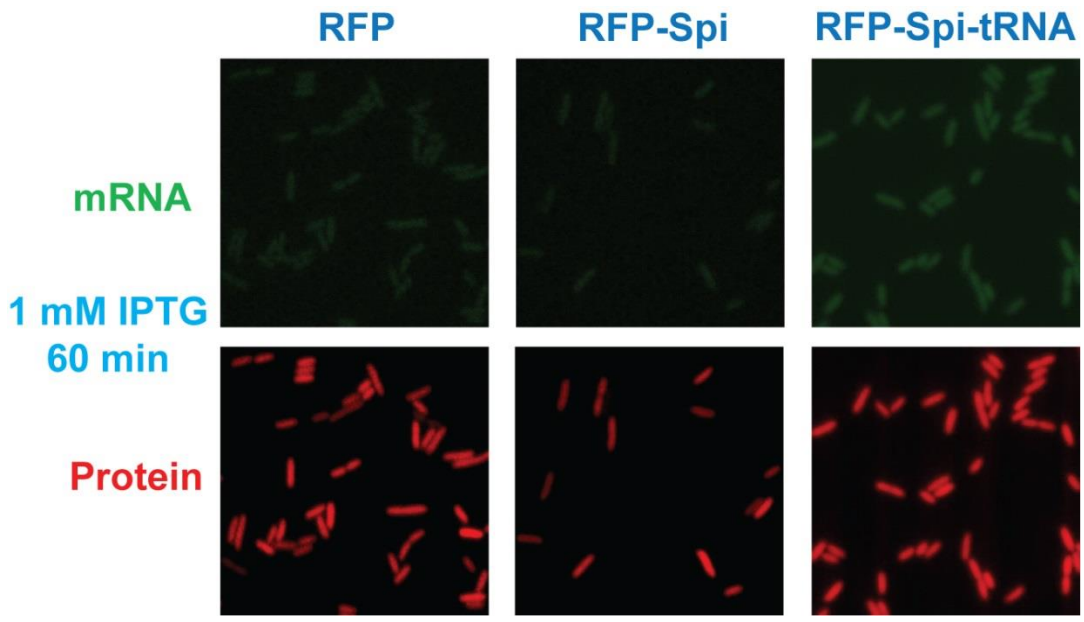

Figure 4.2. Representative fluorescence images of *E. coli* expressing RFP, RFP-Spi and RFP-Spi-tRNA

Figure 4.3

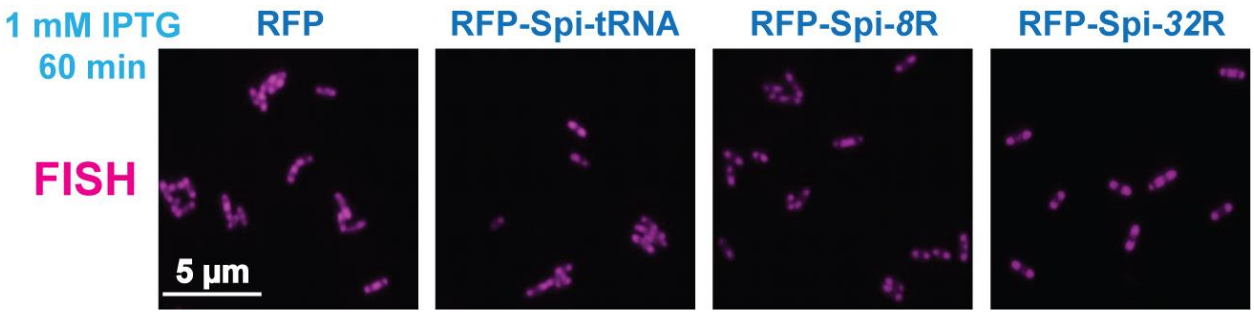

Figure 4.3. FISH validation of the localization of Spinach-tagged RFP mRNA, compared to that of unmodified RFP mRNA.

Figure 4.4

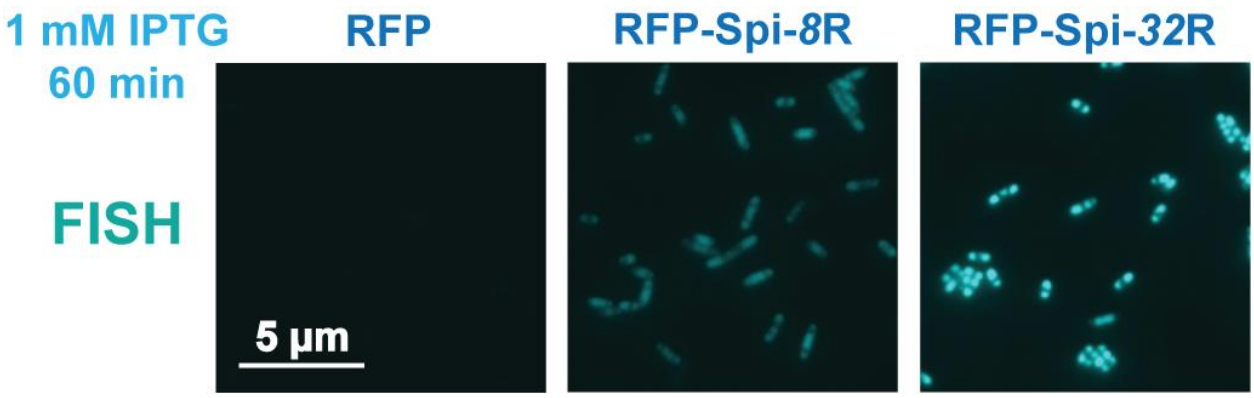

Figure 4.4. FISH validation of the localization of Spinach-tagged RFP mRNA, using FISH probes against the Spinach aptamer.

Figure 4.5

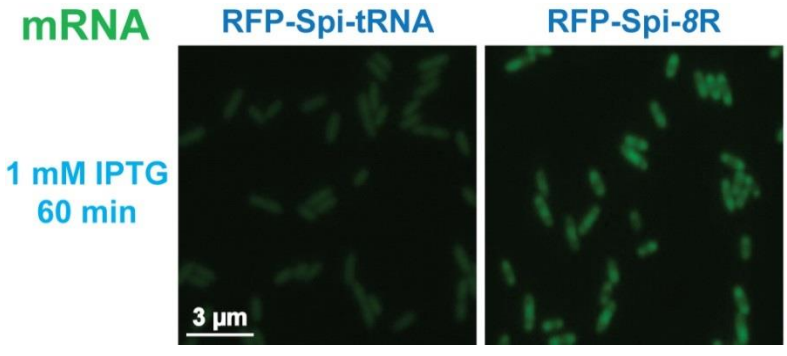

Figure 4.5. Representative fluorescence images of *E. coli* expressing mRNA tagged by Spi-tRNA and Spi-*8*R.

Figure 4.6

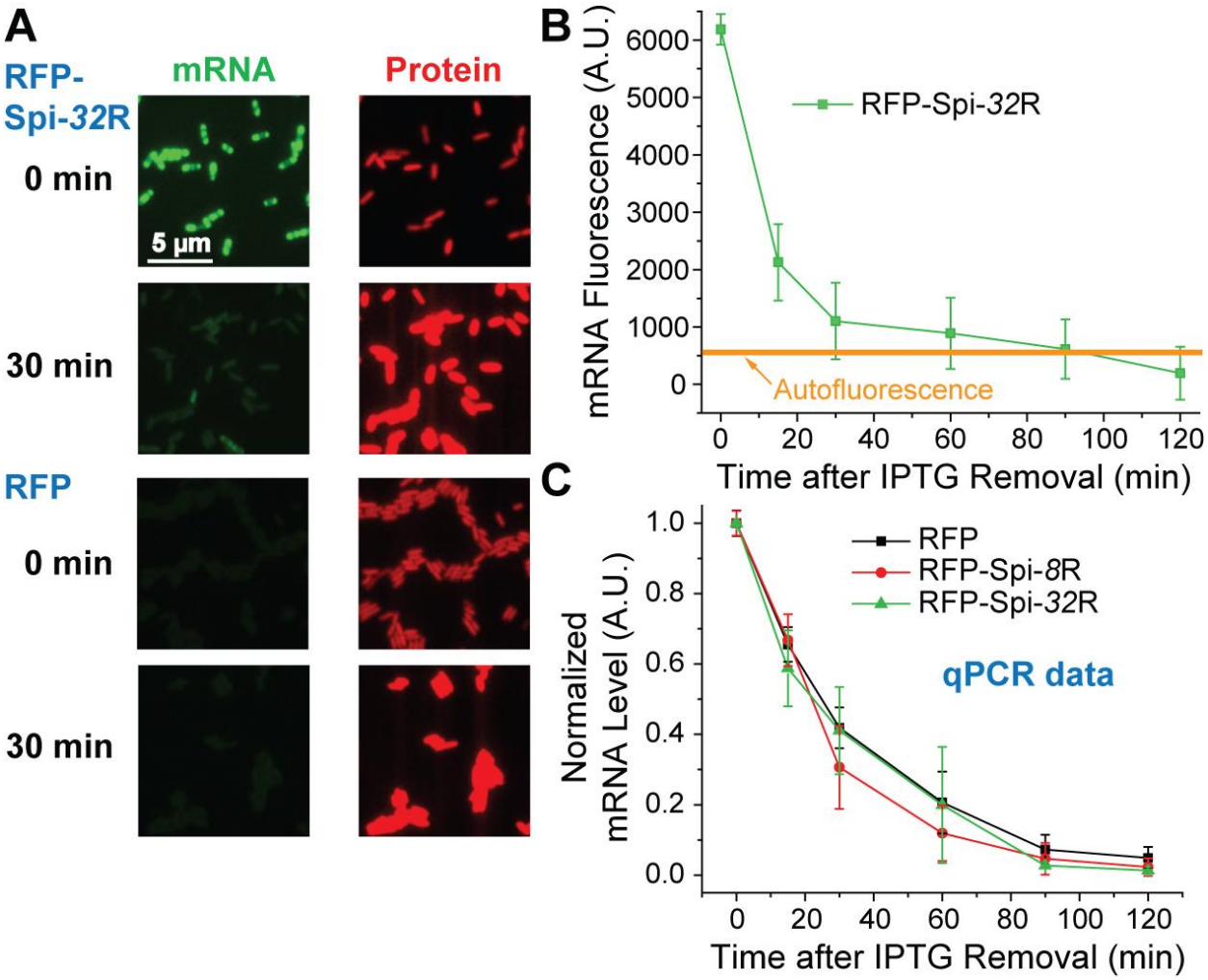

Figure 4.6. (A) mRNA (Spinach) and protein (mRFP1) fluorescence of *E. coli* expressing RFP-Spi-*32*R and RFP 0 min and 30 min after IPTG removal. (B) Spinach fluorescence as a function of time in the decay assay for RFP-Spi-*32*R, compared with the autofluorescence measured from *E. coli* expressing RFP. (C) RNA level in *E. coli* expressing RFP, RFP-Spi-*8*R and RFP-Spi-*32*R as a function of time in the decay assay, measured by qPCR.

Figure 4.7

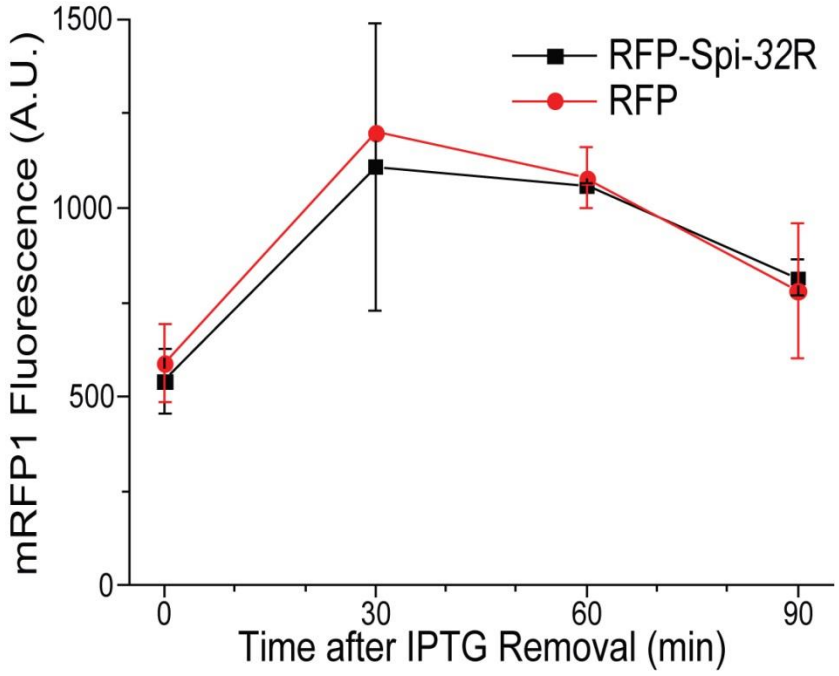

Figure 4.7. mRFP1 protein fluorescence level as a function of time in the decay assay, in *E. coli* expressing RFP-Spi-*32*R and unmodified RFP mRNA.

Figure 4.8

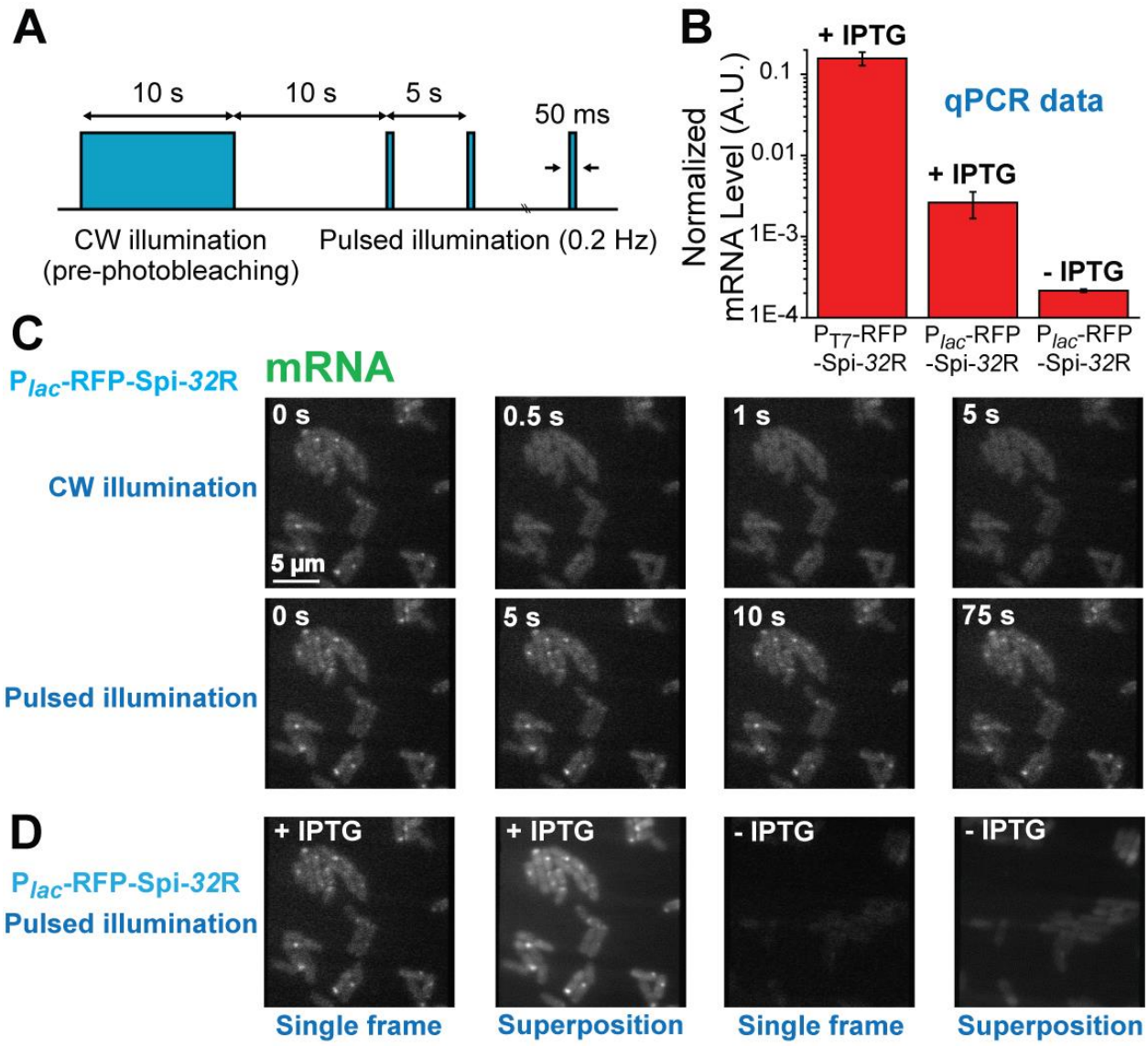

Figure 4.8. (A) The pulsed illumination strategy to observe Spinach fluorescence, in which a 10 s continuous-wave (CW) illumination was applied to pre-photobleach (pre-PB) cellular autofluorescence, with a 10 s wait period after illumination withdrawal, and then pulsed laser (power 0.2 mW, frequency 0.2 Hz, pulse duration 50 ms) was sent to illuminate the sample. (B) Expression level of RFP-Spi-*32*R mRNA under different promoters and induction conditions, measured by qPCR. (C) Representative fluorescence images of induced P_{lac}-RFP-Spi-*32*R cells under CW or pulsed illumination. (D) Fluorescence images of induced and uninduced P_{lac}-RFP-Spi-*32*R cells, shown in single frame (50 ms exposure time) or the superposition of 15 frames under pulsed illumination.

Figure 4.9

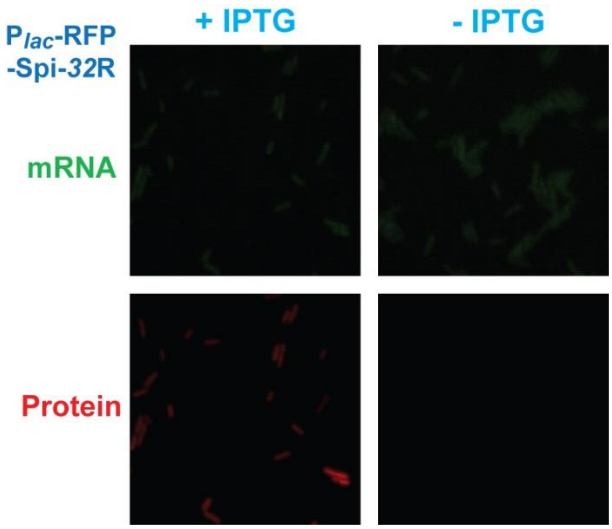

Figure 4.9. Representative epifluorescence images of E. coli expressing P_{lac}-RFP-Spi-*32*R, before and after IPTG induction.

Figure 4.10

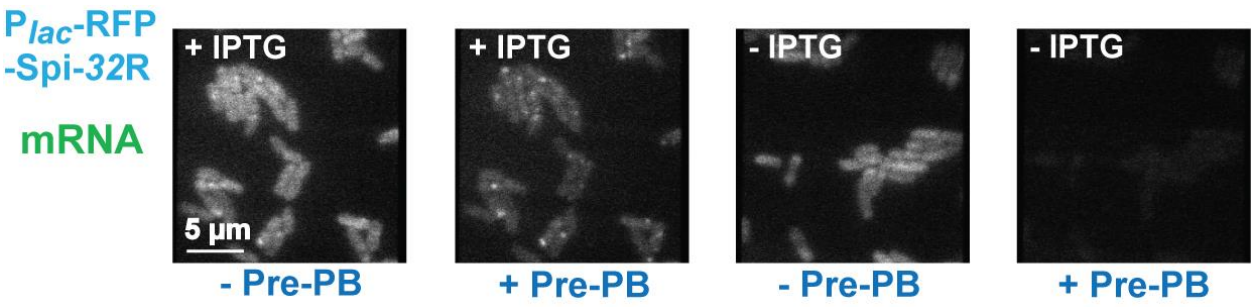

Figure 4.10. Fluorescence images of P_{lac}-RFP-Spi-*32*R cells before and after pre-photobleaching (pre-PB; CW illumination, 10 s) of autofluorescence.

Figure 4.11

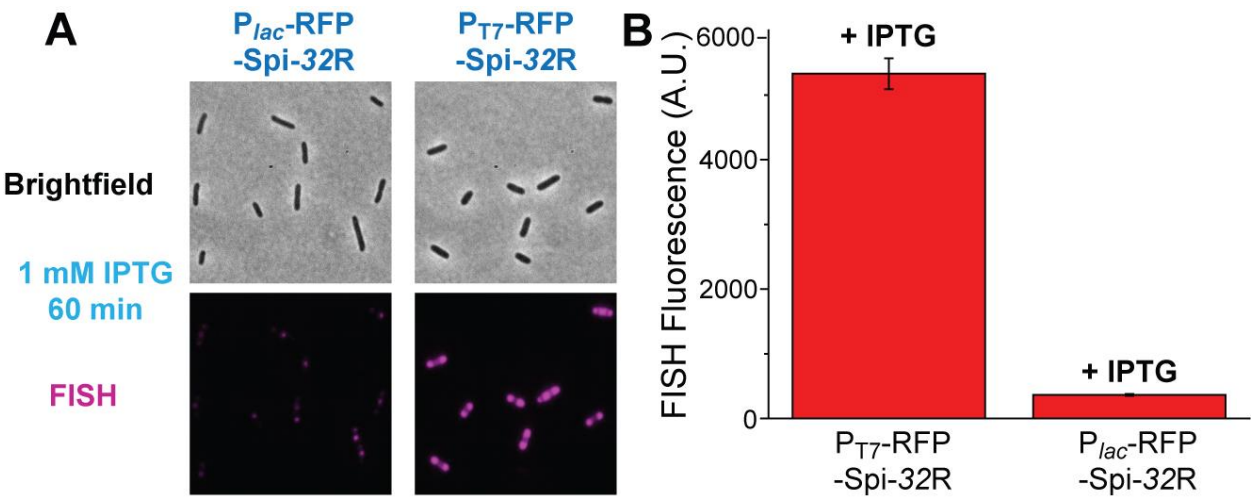

Figure 4.11. (A) FISH images indicating RFP-Spi-*32*R mRNA localization and expression level under the control of different promoters in *E. coli*. (B) Average FISH fluorescence intensity in *E. coli* expressing P_{lac}-RFP-Spi-*32*R and P_{T7}-RFP-Spi-*32*R.

CHAPTER 5

SORT-SEQ STUDY ON *ptsG* REGULATION BY *sgrS*

5.1 Introduction

Bacterial small RNAs (sRNAs) play an important role in regulating gene expression at post-transcriptional level, especially in stress response[45,116,122]. Certain examples include ryhB in iron limitation stress response[123], dsrA in osmotic and acid stress response[53] and oxyS in oxidative stress response[54]. One of the sRNAs we have particular interest in is sgrS sRNA synthesized in response to sugar-phosphate stress[50]. Although sugar is the most important carbon and energy source for bacteria, the accumulation of sugar-phosphate, generated during sugar intake by specific sugar transporters, will lead to production of detrimental cellular metabolites, resulting in growth inhibition and cell damage. The sugar-phosphate accumulation can be relieved by sgrS regulation. As a 227-nt sRNA, sgrS primarily targets ptsG mRNA[124], coding for the EIICB domain of glucose phosphotransferase as the main membrane transporter for glucose. With the help of Hfq protein, sgrS anneals to ptsG mRNA, and thus blocks the ribosomal binding site (RBS), inhibiting translation initiation to synthesize new glucose transporter. At the same time, the sRNA/mRNA binding complex would be targeted and degraded by RNaseE[50], significantly decreasing the cellular level of ptsG mRNA. Negative regulation is also applied to manXYZ[50,125], coding for another type of sugar transporter. sgrS could simultaneously enhance the cellular level of a protein YigL which removes the phosphate group from sugar-phosphate[50,126], to relieve the stress. In addition to target mRNAs regulation, either positively or negatively, sgrS sequence even contains an open reading frame coding for SgrT protein[127], which functions to deactivate sugar transporter, slowing down further accumulation of sugar-phosphate.

In *Escherichia coli* (E. coli), the 227-nt sgrS sRNA can be functionally divided into 3 segments[124], the sgrT-coding sequence, the seed region, and 3’ Hfq-binding region, counted from 5’ to 3’ on the RNA sequence (Figure 5.1). The seed region is a stretch of sequence that has partial complementarity with diverse target RNAs and thus is responsible for annealing to target RNAs for further regulation. As an example, for ptsG and sgrS interaction in *E. coli*, sgrS

nucleotides 169 - 177 partially anneal with ptsG 5' UTR sequence, shielding the ribosomal binding site (RBS) and inhibiting translation initiation. Unlike the seed region that directly interacts with target mRNA, 3' Hfq-binding region contains a GC-rich sequence followed by a U-rich sequence; the GC-rich sequence is predicted to fold into multiple stem-loops, and along with the U-rich sequence, is recognized by Hfq protein in bacteria. Though biochemistry experiments showed that sgrS could solely bind target RNAs and trigger further regulation, in the cell Hfq is required for sgrS function[52]. It is thought that Hfq binds sgrS to increase its stability and to guide it to target RNAs, which were also shown to bind Hfq.

Our Lab has previously studied the *in vivo* target search kinetics for sgrS regulation on ptsG using super-resolution imaging combined with computational biology[128], and we briefly discussed the contributions of some parameters including Hfq binding, RNaseE participation and sequence complementarity. To further study the ptsG and sgrS interaction, we are interested in understanding how the function of sgrS is determined by its sequence. To answer the question, an interpretation of the sgrS segmentation and identification of key nucleotides within each segment is required. A usual way to achieve that is to generate site-directed or random mutations within the sgrS sequence, and by screening, interesting mutations that significantly change sgrS regulation on its target RNAs could be selected. Further characterization can be conducted to find out the mechanism through which the mutation alters sgrS function. As a pretty typical genetics method to study the relationship between RNA sequence and function, it is quite tedious to introduce and screen hundreds of mutations individually.

A high-throughput way to study the sRNA sequence-dependent function, Sort-Seq, was developed recently[129-131]. The basic principle was to construct a target-reporter system which fused the sRNA-regulated target sequence with a fluorescent protein-coding sequence (Figure 5.2), and the expression level of the system was characterized by the protein level, measured by cellular fluorescence; in addition, the expression system is under the control of certain inducers (Isopropyl β-D-1-thiogalactopyranoside, or IPTG, here); besides, the expression level could be well regulated by certain sRNAs, due to the interaction between target sequence and the sRNA. With the controllable target-reporter system, to study the sRNA sequence-dependent regulation, a library containing different sRNA sequence mutations generated by mutagenesis PCR was introduced (Figure 5.2); different mutations were introduced into a collection of cells expressing

the target-reporter, meanwhile the expression of the mutation library was also controlled by certain inducers (Anhyrotetracycline, or aTc, here). With the co-expression of the target-reporter and the sRNA mutation library, cells would show very diverse fluorescence signal, since different sRNA mutations would have arious regulation efficiencies on the target sequence, resulting in fluorescence discrepancy among cells. The great part of the Sort-Seq method is that the collection of cells with different fluorescence signal can be arranged into different groups using fluorescence-activated cell sorting (FACS). Afterwards by extracting the mutation sequence information of cells from different sorted groups by sample preparation and high-throughput sequencing, the sgrS mutant distribution in each sorted group would be measured and be correlated with regulation efficiency calculated based on average fluorescence signal. The single round of FACS-Sequencing experiment can provide the regulation efficiency information for hundreds of different mutations, avoiding the tedious construction and screening process.

Previous studies have applied Sort-Seq to understand the sequence-function relationship for sRNAs including dsrA[131], ryhB[131], Qrr1-5[129] *etc.*, and to identify important single nucleotide for sRNA function maintenance. Here we developed a Sort-Seq assay to study the regulation of sgrS on ptsG, and tried to find interesting sgrS mutations for further understanding of the regulation principles of sgrS.

5.2 Results

5.2.1 Construction of the target-reporter system for sgrS regulation on ptsG

To construct the reporter system which monitors target gene expression and regulation, we subcloned a fusion sequence containing partial ptsG sequence (105 nt 5' UTR plus the first 30 nt coding sequence of ptsG mRNA) and superfolder GFP-coding sequence (named ptsG-*sf*GFP), from pZEMB8 plasmid (a gift from Carin Vanderpool Lab), and further inserted the sequence into pAS05 (Amp^R) (a gift from Erel Levine Lab), a low-copy plasmid. The generated plasmid, named pAS06 (Figure 5.2), was transformed into *E. coli* strain MB1 (ΔptsG, ΔsgrS, LacIQ, TetR), for further imaging and cell sorting experiments. Another plasmid, pZAMB1 (Cm^R) (a gift from Carin Vanderpool Lab), consisting of sgrS-coding sequence, was co-transformed into MB1 strain to study sgrS regulation on the target-reporter system (Figure 5.2). The expression of

ptsG-*sf*GFP and sgrS, under the control of $P_{Llac-O1}$ and $P_{Ltet-O1}$, respectively, can be induced by Isopropyl β-D-1-thiogalactopyranoside (IPTG) and Anhyrotetracycline (aTc).

5.2.2 sgrS mutation library construction

The sgrS mutation library was generated by mutagenesis PCR followed by plasmid ligation, transformation, and bacterial culture-based amplification (Figure 5.2), as described in details in Experimental Procedure. The mutation library was transformed into MB1 strain carrying pAS06 plasmid (consisting of ptsG-*sf*GFP) and the transformed cells were aliquoted into frozen stocks for the following imaging, cell sorting and qPCR experiments.

5.2.3 Characterization of the target-reporter system and its regulation by sgrS

The target-reporter system was evaluated to confirm its applicability and usefulness for Sort-Seq experiments for ptsG and sgrS. The evaluation was conducted based on two simple criteria:

(1) Cells transformed with the target-reporter system should show significant fluorescence enhancement compared to empty cells under induction conditions, suggesting successful expression and function of the fluorescence reporter (superfolder GFP) under our experimental conditions. In some cases fluorescent proteins lose its function once fused to another protein due to misfolding. Although superfolder GFP has been optimized to maintain the correct folding in the context of protein fusion, a test is still needed to ensure that fluorescence protein reporter works well to indicate the expression of fused target.

(2) Cells co-transformed with the target-reporter system and the regulatory sRNA should show significant fluorescence decrease compared to those only transformed with the target-reporter system, suggesting observable sRNA regulation on the target-reporter system. Since the target-reporter system would be used in Sort-Seq experiments to qualify and quantify sRNA regulation on the target, the reporter fluorescence level should show considerable difference in the presence and absence of the sRNA. Usually when the expression level of the sRNA is too low and that of the reporter is too high, the target-reporter system would overwhelm the regulatory sRNA. The system (target-reporter and sRNA) can be optimized by adjusting the average copy-number of the plasmids carrying the target-reporter sequence and the sRNA sequence, and the induction

conditions under which the target-reporter and the sRNA are expressed. Since we constructed the regulation system (ptsG-sfGFP/sgrS) based on a previously reported one which has already been well optimized, we simply used the previously reported induction conditions to test the regulation system we constructed.

In terms of the target-reporter system we constructed, the expression of ptsG-*sf*GFP and its regulation can be characterized reporter protein translation. Traditionally we could use western blotting to quantify the protein level. Nevertheless, since we used a fluorescent protein as the reporter, we could directly measure the reporter fluorescence via fluorescence imaging or flow cytometry to quantify the protein level upon induction and sRNA regulation.

5.2.3.1 Cell imaging indicated that the target-reporter system was well optimized under our induction conditions

Three types of cells were imaged under fluorescence microscopy: cells expressing ptsG-*sf*GFP, with no induction; cells expressing ptsG-*sf*GFP, with IPTG induction; cells expressing ptsG-*sf*GFP and sgrS, with IPTG and aTc induction (Figure 5.3). The GFP fluorescence level of the listed three types of cells was useful for us to evaluate whether the target-reporter system was induced efficiently, and whether it could be efficiently repressed by the sRNA.

According to the fluorescent images, we found that upon IPTG induction of ptsG-*sf*GFP, cells showed bright GFP fluorescence (Figure 5.3), while with IPTG and aTc co-induction, cellular fluorescence signal was quite weak (Figure 5.3), indicating that sgrS induction repressed the expression of the reporter at translation level. Therefore it seems that the reporter system (ptsG-*sf*GFP) we constructed was useful for sgrS regulation study.

5.2.3.2 Flow cytometry analysis quantified reporter induction and regulation

Although we found that cells expressing ptsG-*sf*GFP could show significant fluorescence enhancement upon induction and were well regulated by sgrS, it is still important to check the cellular fluorescence using flow cytometry before we conduct Sort-Seq experiments, since the excitation light source used by the flow cytometry was not optimized for the fluorescent protein reporter, or the detection side of the flow cytometry instrument was not as sensitive as was the

fluorescent microscopy. Therefore we performed flow cytometry analysis on cells expressing ptsG-*sf*GFP and co-expressing ptsG-*sf*GFP and sgrS. We found that the reporter system we constructed showed great performance in the flow cytometry instruments (Figure 5.4), both for cells induced for ptsG-*sf*GFP expression, and for cells co-induced for ptsG-*sf*GFP and SgrS expression. The average cellular fluorescence level (center position of the fluorescence distribution histogram in Figure 5.4) decreased ~7-fold (from ~7×10^3 (black curve) to ~1×10^3 (red curve)) (Figure 5.4) with sgrS co-induction, compared to the condition where only ptsG-*sf*GFP was induced, showing that sgrS regulation could be efficiently detected using flow cytometry.

5.2.4 Cells transformed with sgrS mutation library showed various sRNA regulation efficiencies on the target-reporter system from cell to cell

We transformed plasmids carrying sgrS mutation library into MB1 strain (ptsG-*sf*GFP was co-transformed), and induced the expression of ptsG-*sf*GFP and sgrS mutation library. Cells were first imaged under fluorescence microscopy (Figure 5.3). We found that unlike cells co-expressing ptsG-*sf*GFP and wild-type (WT) sgrS, which showed low cellular fluorescence suggesting sgrS repression on ptsG-*sf*GFP expression, cells transformed with the mutation library of sgrS showed very different cellular fluorescence level (Figure 5.3): some showed similar low fluorescence to those expressing WT sgrS, while some showed high fluorescence. The difference of cellular fluorescence level suggested that cells transformed with different sgrS mutants showed very different repression extent on ptsG-*sf*GFP.

We further sent the cells to flow cytometry analysis and found that the cells showed very broad distribution of fluorescence (Figure 5.4), consistent with the cell imaging result. We could find from the flow cytometry analysis result that the average fluorescence intensity of cells transformed with sgrS mutation library, upon IPTG and aTc induction, spun from similar fluorescence level of cells expressing only ptsG-*sf*GFP, to that of cells expressing ptsG-*sf*GFP and WT sgrS. The broad distribution made it applicable for fluorescence-activated cell sorting (FACS), and therefore we further sent the cells for Sort-Seq experiments, where cells were distributed into different groups based on its average cellular fluorescence measured by flow cytometry.

5.2.5 Cell sorting and Next-Gen sequencing

Based on the flow cytometry analysis result (cell number as a function of average fluorescence), we set up 7 sorting bins with identical size (in log-scale) to sort the cells into 7 groups, named D1-D7 (Figure 5.5). Cells from group D1 showed the lowest average cellular fluorescence, suggesting sgrS mutants from this group have similar regulation efficiency on pts*G*-*sf*GFP compared to that of WT sgrS, while cells from group D7 showing the highest average cellular fluorescence indicated that sgrS mutants from those cells have its regulation function severely perturbed. Cells from each sorted group were sent for plasmid collection (pZAMB1-sgrS mutant). The mutated sgrS sequence within the plasmid were amplified and further adapted and barcoded for Next-Gen sequencing.

5.2.6 Sequencing data analysis

We used a Matlab code to process the FASTQ file (the raw sequencing data) and extracted the sequencing reads. Individual reads were compared with the standard sequence, where the WT sgrS sequence plus the adaptor sequence were used, to generate the mutation information of the sgrS mutant.

5.2.6.1 Sequencing data quality evaluation

In the sequencing experiment, we used Illumina MiSeq Reagent Kit v3 (600-cycle) (Illumina), which allowed 300-nt reading length from both ends of the sequencing sample, for total length of 600 nt. Since 300-nt reading length from one end allowed the full coverage of the sgrS sequence we applied (254 nt) plus the adaptor length, initially we planned to use the 300-nt paired reading to sequence our sample from forward and reverse directions. In this case we could select the data with identical reading results by forward and reverse sequencing, to decrease the occurrence possibility of false reading. To test the applicability of paired reading, we first evaluate the data qualify for forward and reverse sequencing result, respectively.

The first step was to filter out the sequencing data that had less than 300-nt reading length, i.e. the sequencing process failed to complete the full-length reading. Surprisingly, we found that although forward reading data showed high percentage of qualified reads (~100%), reverse

reading data showed much less successful rate (<50%) to achieve 300-nt reading length (Figure 5.6).

Since reverse reading was initiated after the completion of the forward reading, it was possible that at the late stage of the sequencing experiment, where the reverse reading was conducted, the sample quality was much lower compared to that at the initial stage, leading to lower completion percentage for reverse sequencing.

Furthermore, for those sequencing reads that achieved 300-nt reading length, we analyzed the average Q-score for each nucleotide position within the total sequence, and compared the average quality of forward and reverse sequencing data, at single-nucleotide resolution.

We found that for forward sequencing data, within the 300-nt reading length, the average Q-score remained quite high (Q-score $\geq$ 27) (Figure 5.7). Nucleotide positions 6 - 258 showed stable high average Q-score (Q-score $\geq$ 32); nucleotide positions 1 - 5 showed a little lower average Q-score, but the value was still higher than 32; the average Q-score became lower starting from nucleotide position 259, but generally the value remained higher than 27. According to the nucleotide-specific Q-score plot, we concluded that the forward sequencing data were quite reliable. In a big contrast, for reverse sequencing data, from nucleotide position 144, the average Q-score started decreasing (Figure 5.7); the value decreased dramatically along the sequencing direction and at nucleotide position 209, the average Q-score value was significantly smaller than 27[132]; at nucleotide 300, which was the final sequencing read, the Q-score was just barely larger than 10. In this case we could not confirm the accuracy of reverse sequencing data.

According to the analysis, we found that for our paired reading sequencing data, the quality of forward sequencing data was quite fine, while reverse sequencing data were not quite reliable. Since forward sequencing data were long enough to cover the total sgrS mutation sequence, we discarded the reverse sequencing data and only analyzed the forward sequencing data for our Sort-Seq experiments.

5.2.6.2 Percentage of multi-mutation sequence increases from D1 to D7 sorted group

We did simple statistics on the sequencing data to find out the occurrence rate of non-mutation, mutation and single-mutation sequence for the sgrS sequence, in different sorted groups (Figure 5.8). We expected that sequences from D7 group, which showed the highest fluorescence and thus suggested the lowest regulation efficiencies of sgrS mutants on ptsG-*sf*GFP, should have the lowest percentage of non-mutation sequence and the highest percentage of multiple-mutation sequence, because generally the more mutations within the sequence, the higher possibility for that specific sRNA mutation to show lower binding affinity to target mRNAs or to suffer misfolding problems. The analysis result accorded with our expectation. From D1 to D7, the occurrence rate of non-mutation sequence (WT sgrS) decreased, while that of multiple-mutation sequence (more than one mutated nucleotide within the sgrS sequence) increased; the occurrence rate of single mutation sequence was generally unchanged, perhaps due to a compensation effect.

5.2.6.3 Mutual information analysis revealed important nucleotides for sgrS regulation on ptsG

We conducted mutual information analysis on sequencing data[129,130] (Figure 5.9). We only analyzed those sequencing readings showing single nucleotide mutation within the sgrS sequence, as it helped us identify the nucleotides too critical to be altered. Mutual information analysis gave each nucleotide position a mutual information value, and the larger the value is, the stronger the sgrS regulation efficiency is perturbed upon single mutation taking place at this nucleotide position.

The mutual information analysis suggested several important nucleotides: G176, G178, G183, G184, C199, C200, U222, U223, U224, *etc*. Next we would generate a heat map showing the mutual information value for each nucleotide within the sgrS sequence and discuss the effect of each single mutation in the context of the secondary structure and functional segmentation of sgrS.

5.2.6.4 Mapping of single nucleotides important for SgrS regulatory ability

To show the positions of important nucleotides within the sequence and to better understand how individual nucleotides affect sgrS function structurally, we constructed a heat map where we showed the predicted sgrS secondary structure and colored each nucleotide according to the

calculated mutual information value (Figure 5.10 and Figure 5.11). The redder the color was, the stronger the regulation efficiency was perturbed once the nucleotide was mutated (Figure 5.10 and Figure 5.11).

The fact that G176 and G178 are very important for sgrS regulation on ptsG, is consistent with previous study showing that mutating either G to C totally abolished sgrS regulation over ptsG. Some other discoveries in the Sort-Seq experiment are: the 3' U-rich sequence of sgrS, containing 8 U nucleotides, is very important for sgrS function, and 5 out of the 8 U nucleotides showed strong effect; the 3' GC-rich sequence, forming 2 stem-loop structures, is also very important, as for the 24 nucleotides forming base-pairs, 19 of them showed strong perturbation upon mutation (Figure 5.10 and Figure 5.11). Since both the U-rich and GC-rich sequence were suggested to play a role in Hfq binding, it confirmed that Hfq interaction is extremely important to maintain the sgrS function in the cell. It is not clear whether the significance of those individual nucleotides is due to its direct interaction with Hfq, or its possible role in maintaining the stem-loop structure.

5.3 Discussion

In this chapter, we constructed a Sort-Seq assay to study sequence-dependent sgrS regulation on target mRNAs, more specifically, ptsG. Unlike commonly used genetics methods where people would generate random mutations on sgrS sequence, and conducted various biochemistry or genetics assays to screen various mutations and identify interesting mutations, which is time-consuming and tedious, Sort-Seq allows simultaneously quantification of hundreds of mutations using two high-throughput technologies, fluorescence-activated cell sorting and next-generation sequencing. By co-introduction of the target-reporter system, which is the fusion of ptsG 5' UTR and superfolder GFP-coding sequence, and the mutation library of the sgrS, we acquired a collection of cells carrying different sgrS mutations, and thus lead to regulation efficiency discrepancy. The power of flow cytometry enables the separation of cells carrying weak mutations and strong mutations by sorting cells with different fluorescence levels into multiple groups. Further application of the Next-Gen Sequencing enables extraction of the mutation distribution among different groups, thus informing us how the hundreds of mutations within the library affect sgrS regulation efficiency individually.

With our Sort-Seq experiment result, we identified several nucleotides that are very important for sgrS regulation on ptsG (Figure 5.12). G176 and G178 belong to sgrS/ptsG basepairing region. Although during sgrS/ptsG annealing, 15 Watson-Crick pairs will form spanning 19 bp, only G176 and G178 have been reported to be extremely important for the annealing process as mutation of either nucleotide would deprive sgrS regulation on ptsG[124,128]. Our Sort-Seq result is consistent with the report. Interestingly, other identified significant nucleotides all belong to the 3' region responsible for Hfq interaction, including 5 nucleotides from the first short stem-loop, 14 nucleotides from the second long stem-loop, and 5 U nucleotides from the poly-U sequence. The single mutations of those nucleotides within either stem-loop might affect the formation of the secondary structure, as all of those nucleotides are predicted to basepair with another nucleotide, and the structure breakage might affect or even abolish Hfq binding. The poly-U sequence, though important for Hfq binding also, might be recognized by Hfq via the U-rich sequence instead of the structure, as poly-U usually does not participate in secondary structure formation. It would be helpful to introduce a second mutation to regenerate the basepairing with the first mutation within the stem-loop region. If the sgrS function is recovered, the effect of the original single mutation might function structurally instead of sequence-wise.

In addition to plan more experiments to confirm the Sort-Seq result, we can also select interesting mutations to conduct *in vivo* kinetics measurement, as we did before for WT sgrS and ptsG. Since kinetics measurement is able to extract binding and dissociation coefficients for all level of interactions[128], including ptsG/sgrS binding/dissociation and sgrS/Hfq binding/dissociation, using sgrS mutations to conduct *in vivo* kinetics measurement will help us understand how different components involved in sgrS regulation play their roles and influence the overall regulation process.

5.4 Experimental procedure

5.4.1 Construction of plasmids

The reporter system, ptsG-*sf*GFP sequence, including 105 nt 5' UTR and 30 nt coding sequence (coding first 10 amino acids of PtsG protein) of PtsG mRNA was fused by 42 nt linker sequence and superfolder GFP coding sequence, was subcloned from the plasmid pZEMB8 (a gift from

Carin Vanderpool Lab). It was inserted into the low-copy plasmid pAS05 (a gift from Erel Levine Lab) between XhoI and XbaI restriction sites to make the plasmid pAS06. The expression of ptsG-*sf*GFP was under the control of $P_{Llac\text{-}O1}$.

A medium-copy plasmid pZAMB1 (a gift from Carin Vanderpool Lab) was applied for the expression of the sRNA, sgrS, whose sequence was inserted between NdeI and BamHI restriction sites. The expression was under the control of $P_{Ltet\text{-}O1}$. The plasmid pZAMB1 was also utilized as the template of mutagenesis PCR to construct sgrS mutation library, and as the vector to insert the sgrS mutation sequence.

5.4.2 Cell growth and induction condition

The *E. coli* MB1 strain (ΔPtsG, ΔSgrS, LacIq, TetR) was used for fluorescence imaging, flow cytometry and sequencing sample preparation experiments. *E. coli* cells were grown at 37 °C in Lysogeny Broth (LB) (LB Broth Miller, EMD Millipore) liquid media with antibiotics as plasmid selection markers. 100 μg/mL ampicillin (Gold Biotechnology, Inc) was applied for pZEMB8, pAS05 and pAS06 plasmid, and 30 μg/mL Chloramphenicol (Sigma-Aldrich) was applied for pZAMB1 plasmid and derivative SgrS mutation library. Cell density was characterized using optical density (OD) at 600 nm using a plastic cuvette in a Fisher Scientific Educational Spectrophotometer (Fisher Science Education).

To induce the expression of the reporter system (ptsG-*sf*GFP) and/or the regulatory sRNA (sgrS), *E. coli* MB1 cells transformed with certain plasmids (pAS06 for ptsG-*sf*GFP and pZAMB1 for sgrS or its mutation library) were grown at 37 °C in LB medium overnight with specific antibiotics from a single colony or frozen liquid stock (for sgrS mutation library expression). The liquid culture was diluted 200-fold into fresh LB medium and kept growing until at OD_{600} = 0.1-0.2. The liquid culture was diluted to OD_{600} = 0.001 and was supplemented with 1 mM IPTG (Sigma-Aldrich) and/or 50 ng/mL aTc to induce the expression of PtsG-sfGFP and/or SgrS or SgrS mutation library. When OD_{600} = 0.1-0.2, the *E. coli* cells were collected and further treated for fluorescence imaging, flow cytometry analysis and cell sorting, or RNA extraction and quantitative PCR (qPCR) analysis.

5.4.3 sgrS sRNA mutagenesis

Mutagenesis PCR was conducted on SgrS using Agilent Genemorph II Random Mutagenesis Kit (Agilent Technologies). The protocol was adapted from Erel Levine paper[131]. Mutagenesis PCR was performed with 1.8 ng pZAMB1 plasmid (sgrS sequence included) for 30 cycles. The product was purified by gel electrophoresis and 5 ng was amplified using standard Taq PCR (SureStart Taq DNA Polymerase, Agilent Technologies) to increase the yield of individual mutants. PCR products were digested with NdeI and BamHI followed by gel purification. Along with the pZAMB1 vector generated by NdeI and BamHI digestion and gel purification, 5 ligation reactions were set up for pZAMB1 vector and the prepared PCR insert (sgrS mutation sequence library) using T4 Ligase (New England BioLabs).

The ligation products from the 5 reactions were combined and purified using P-6 micro bio-spin column (Bio-Rad) into water. 4 μL ligation products were transformed into 100 μL 5-alpha Electrocompetent *E. coli* (New England BioLabs), and the transformed cells were recovered for 1 hr in 2 mL Super Optimal Broth (SOB) medium (New England BioLabs). 1 mL of the culture (representing 50% of the transformed cells) were diluted into LB medium supplemented with 30 μg/mL Chloramphenicol to select cells successfully carrying plasmid ligation. Miniprep was conducted to collect the plasmids, which made up of the sgrS mutation library, from the overnight culture.

To generate the cells with sgrS mutation library for Sort-Seq experiments, 10 ng collected plasmids (sgrS mutation library) was transformed into MB1 strain pre-transformed with pAS06 plasmid and expressing ptsG-sfGFP. The transformed cells were recovered and were diluted into LB medium supplemented with 100 μg/mL Ampicillin (pAS06 selection) and 30 μg/mL Chloramphenicol (pZAMB1-sgrS mutant selection) for overnight culture. The liquid culture were spun down and aliquoted as frozen stocks for the following imaging and flow cytometry experiments.

5.4.4 Fluorescence microscopy imaging

The protocol of epifluorescence imaging was adapted from that described in Chapter 4, Methods and Materials Section. 1 mL *E. coli* liquid culture (OD_{600} = 0.1-0.2) were chilled on ice followed

by centrifugation (6,000 g, 4 °C, 1 min) to form a cell pellet. The cells were washed by ice-cold 1X PBS twice, and were then resuspended in 100 µL 1X PBS.

To image the sample, a few µL of cell suspension was sandwiched between a glass coverslip (No. 1.5) and a thin slab of 1.5 % (w/v) agarose gel, by dissolving the agarose in 1X PBS.

All epifluorescence images were taken with a Zeiss Axiovert 200M with the Apotome Structured Illumination Optical Sectioning System. An oil immersion objective (1.46 NA 100x) was used for DIC and fluorescence imaging. A lamp light source (X-Cite 120Q, Excelitas Technologies) was applied to illuminate the sample and an emCCD camera (iXon3 DV887, Andor Technology) to acquire the fluorescence image. The filter sets applied were: DIC (no filter, autoexposure), sfGFP (Ex 480-500 nm, Em 509-547 nm, exposure time 200 ms). The images were acquired and processed by the software ZEN provided by Zeiss.

5.4.5 Cell sorting

Flow cytometry analysis and cell sorting was conducted by Sony SH 800 Cell Sorter (Sony Biotechnology). To prepare the flow cytometry or cell sorting sample, cells were cultured overnight in LB medium with antibiotics; the liquid culture were dilute 200-fold and were kept culturing until OD_{600} reached ~0.1 - 0.2; the cells were further diluted to OD_{600} = 0.001 in LB medium with antibiotics, and 1mM IPTG and/or 50 ng/mL aTc were added according to the *E. coli* strain and its carrying plasmids; cells kept growing until OD_{600} reached ~0.1 - 0.2, and the cells were washed twice with ice-cold 1x PBS, and were kept on ice before flow cytometry analysis or cell sorting experiments.

5.4.6 Sequencing sample preparation

Sorted cells were grown in LB medium supplemented with 30 µg/mL Chloramphenicol to saturation, and Miniprep was conducted to collect the plasmids. Three rounds of PCR amplification were conducted before we obtained the prepared sequencing library: in the first round, we used primers annealing to the flanking region of the sgrS sequence and amplified the sgrS mutation sequence from the extracted plasmids; in the second and third round, adapters specific for Illumina MiSeq sequencing were added onto the originally amplified sequence. The

primer design and experimental protocol could be found in the description of 16S Metagenomic Sequencing Library Preparation provided by Illumina (http://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf).

5.4.7 Illumina Next-Gen sequencing

The sequencing samples prepared from different sorted groups were barcoded with dual-index adapters (Illumina Nextera Index Kit), and were mixed and treated with Illumina MiSeq Reagent Kit v3 (600-cycle) according to the protocol. Since the samples had very low sequence diversity, we spiked in 40% PhiX DNA (PhiX Control v3, Illumina). The samples were sent to Illumina MiSeq Sequencer for sequencing, and sequencing data were distributed into different sets suggesting samples prepared from different sorted groups, according to the adaptors added to the prepared samples.

5.4.8 Mutual information analysis

Mutual information analysis was conducted using the protocol described in the literature introducing Sort-Seq method[129,130].

5.5 Figures

Figure 5.1

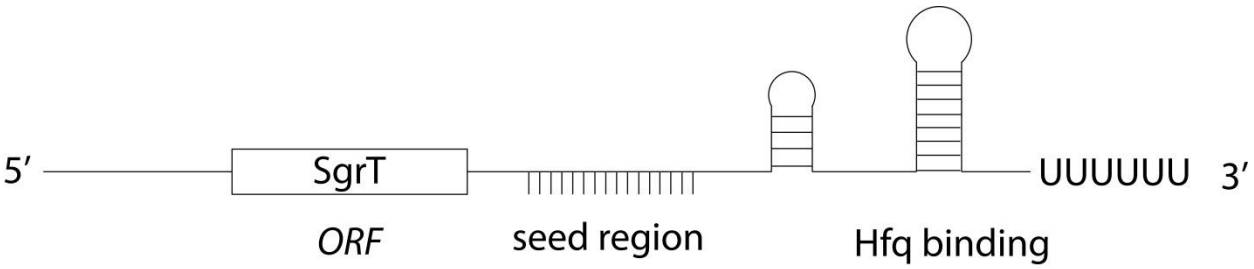

Figure 5.1. The scheme of sgrS functional segments.

Figure 5.2

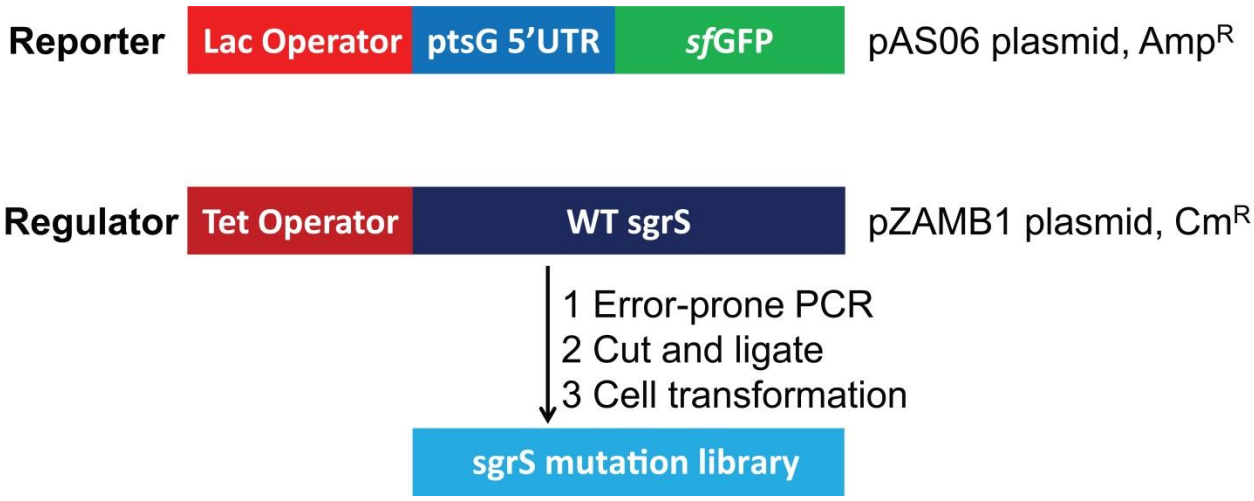

Figure 5.2. The scheme of target-reporter system, WT sgrS expression system, and sgrS mutation library construction.

Figure 5.3

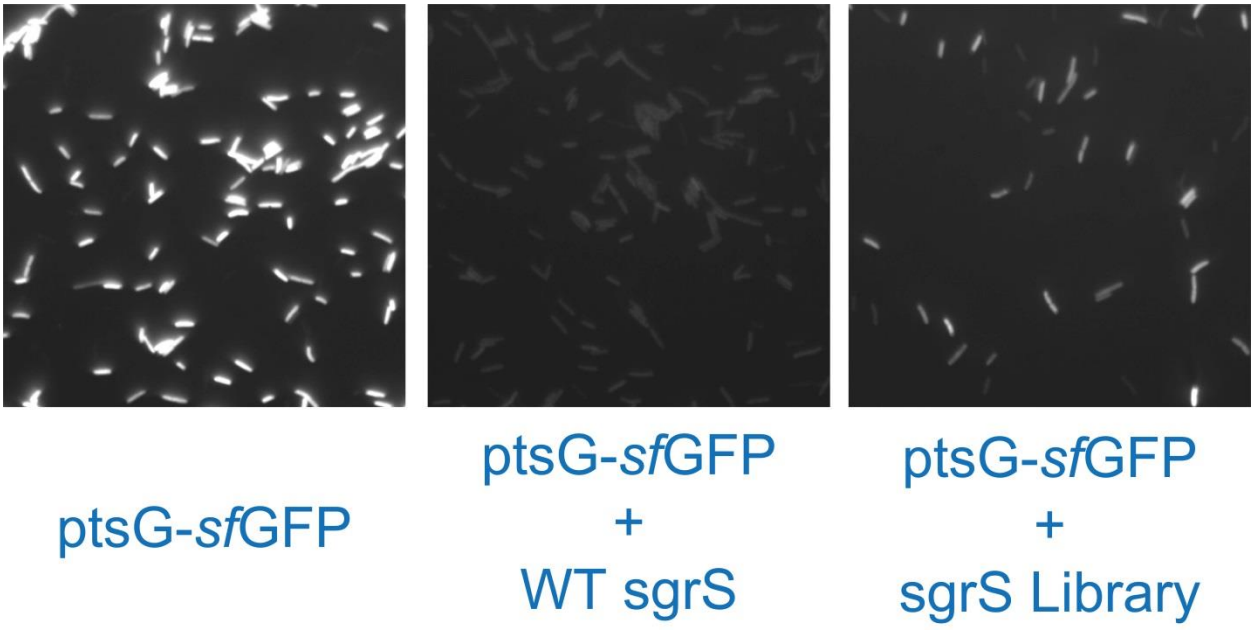

Figure 5.3. Epifluorescence imaging of cells expressing only ptsG-*sf*GFP, ptsG-*sf*GFP and WT sgrS, and ptsG-*sf*GFP and sgrS mutation library.

Figure 5.4

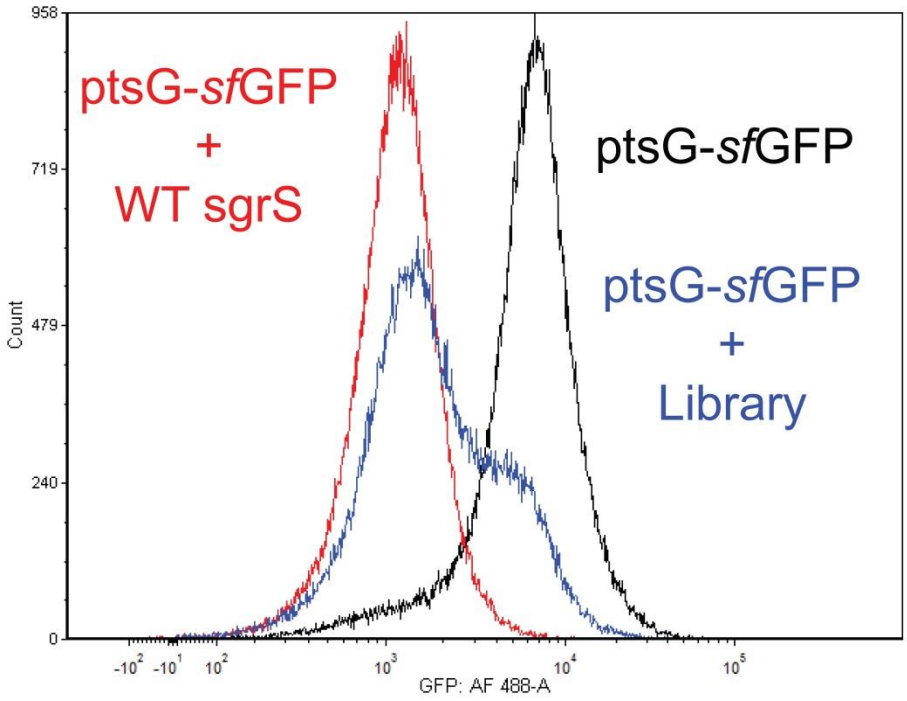

Figure 5.4. Flow cytometry analysis on three types of cell collection, which showed the cell number distribution as a function of cellular fluorescence (log-scale on cellular fluorescence intensity).

Figure 5.5

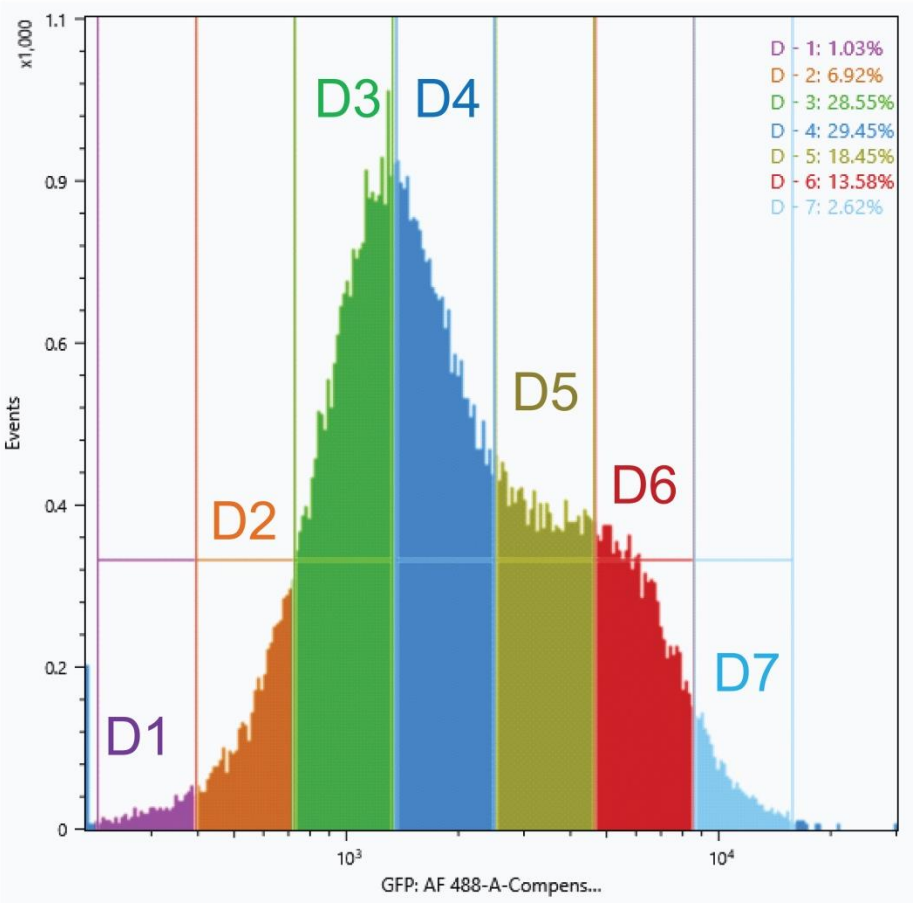

Figure 5.5. The cells were sorted into 7 groups (D1 - D7). The occupancy percentages of each sorted group in the total cell collection are: 1.03% for D1, 6.92% for D2, 28.55% for D3, 29.45% for D4, 18.45% for D5, 13.58% for D6 and 2.62% for D7.

Figure 5.6

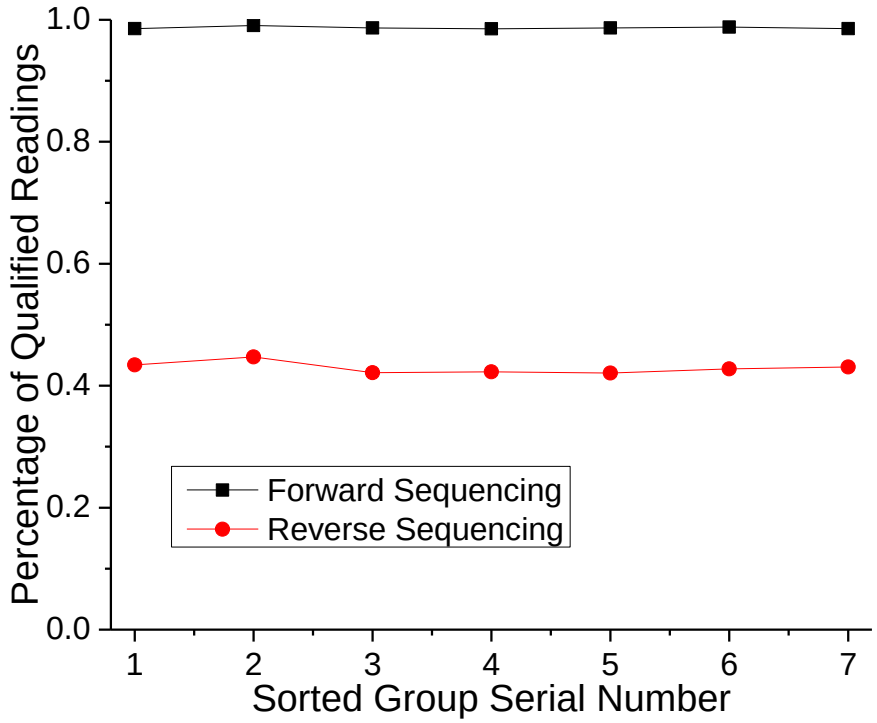

Figure 5.6. Percentage of sequences having 300-nt length, from Illumina sequencing data. The value for sequencing data from sorted group D1 - D7, either achieved by forward or reverse reading, is shown in the figure.

Figure 5.7

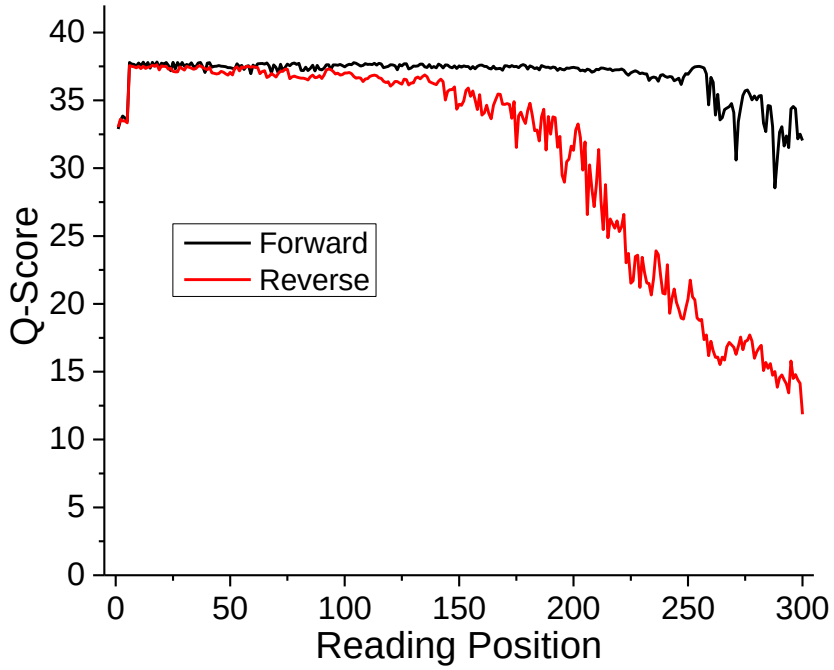

Figure 5.7. The average Q-score for the sequencing results from 7 sorted groups (D1 - D7) for each reading position for forward and reverse reading, respectively.

Figure 5.8

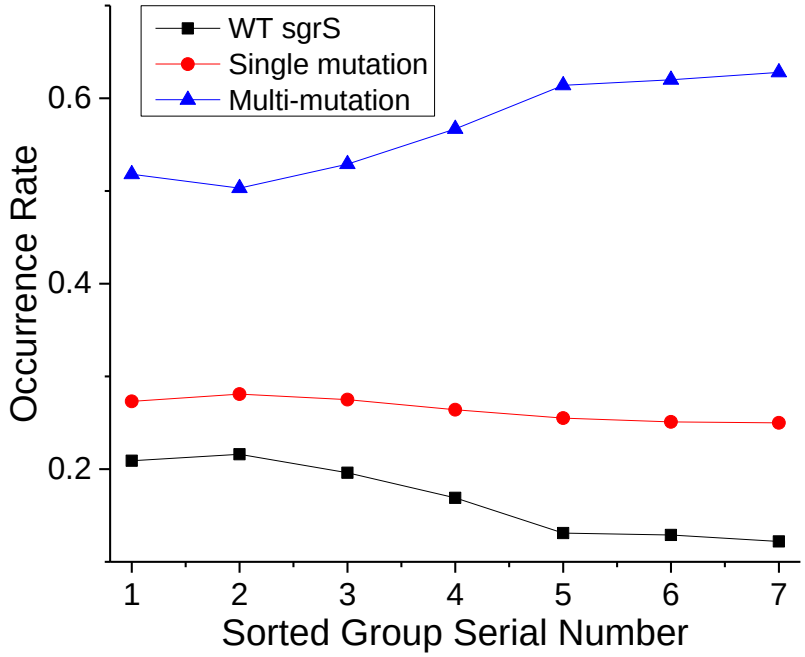

Figure 5.8. Percentage of WT sgrS, single-mutation sgrS and mult-mutation sgrS, respectively, from sorted group D1 - D7, achieved by calculating forward sequencing data.

Figure 5.9

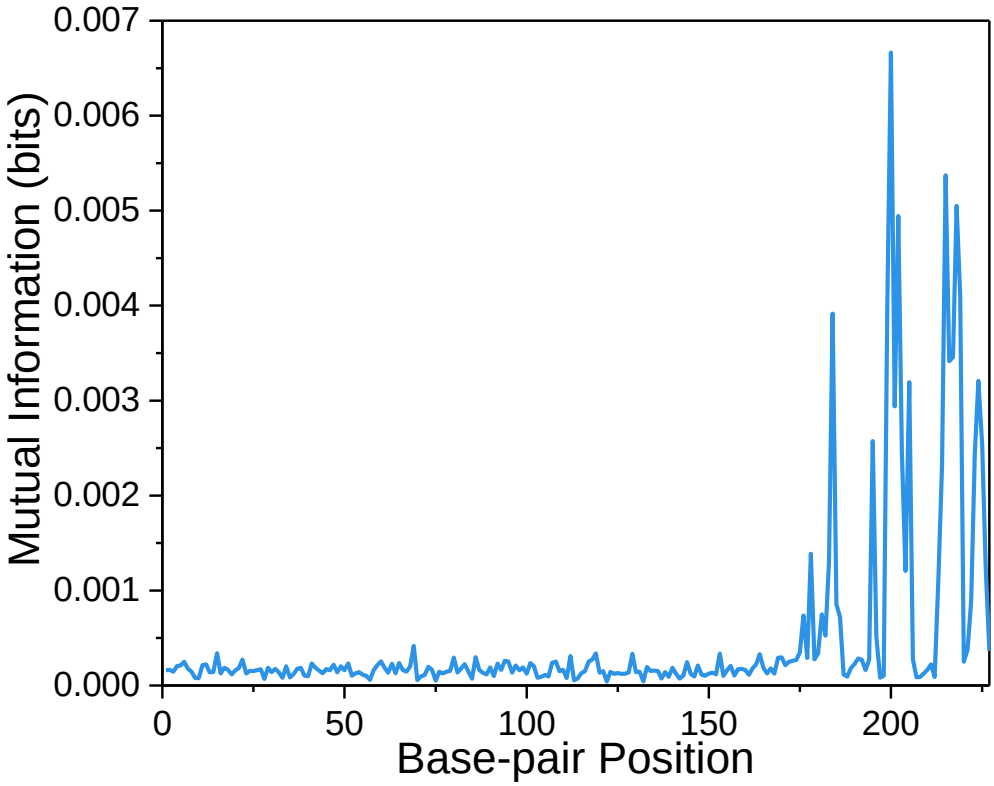

Figure 5.9. Mutual information value for nucleotide 1 - 227 for the applied sgrS calculated from Sort-Seq data.

Figure 5.10

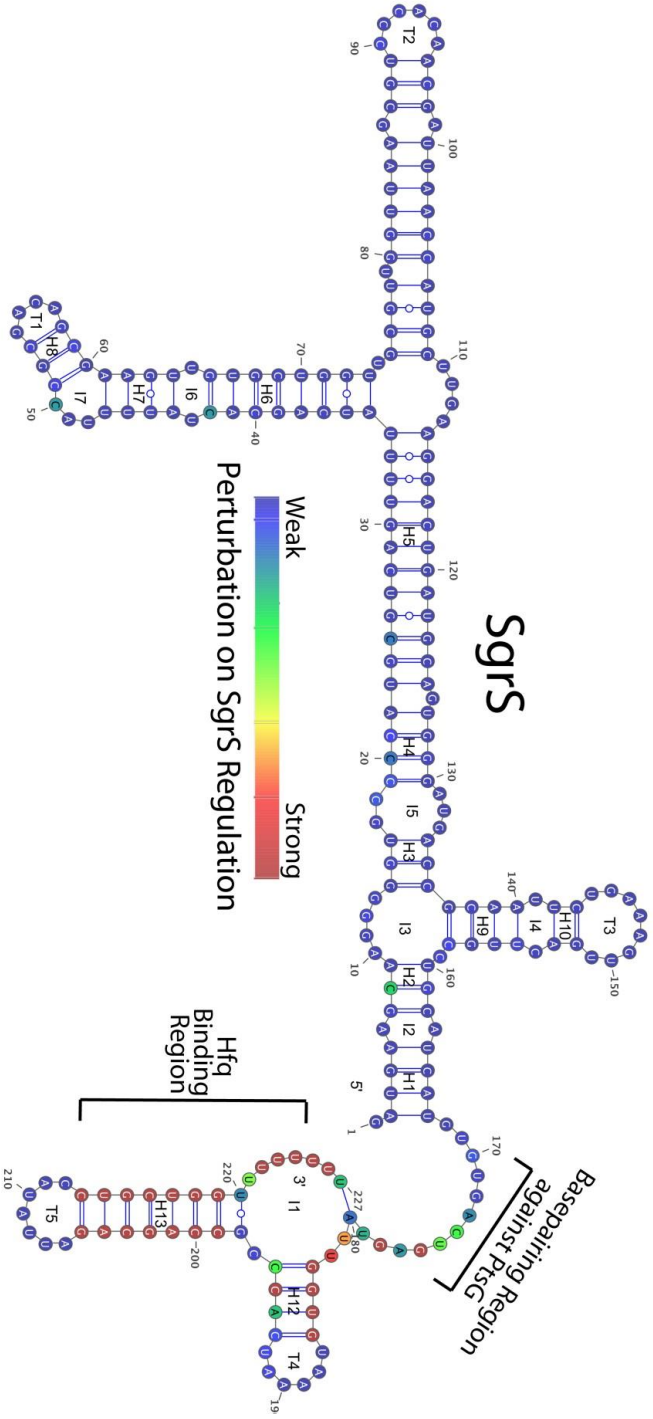

Figure 5.10. A heat map showing how significant individual nucleotides are in maintaining sgrS regulation function on ptsG in the context of predicted sgrS secondary structure. Individual nucleotides are colored and the redder the color is, the more perturbation is brought about when mutations take place on the specific nucleotide.

Figure 5.11

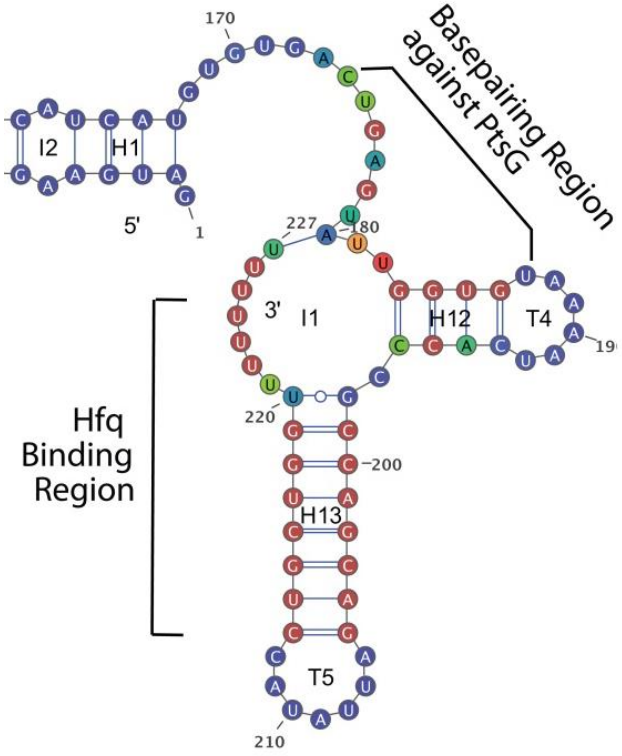

Figure 5.11. A zoom-in figure of the seed region and the Hfq-binding region of sgrS which underlines the identified significant individual nucleotides for ptsG regulation.

Figure 5.12

Figure 5.12. The scheme of ptsG and SgrS interaction and annealing.

REFERENCES

1. Clegg RM: **Fluorescence resonance energy transfer and nucleic acids**. *Methods Enzymol* 1992, **211**:353-388.

2. Roy R, Hohng S, Ha T: **A practical guide to single-molecule FRET**. *Nat Methods* 2008, **5**:507-516.

3. Ha T, Enderle T, Ogletree DF, Chemla DS, Selvin PR, Weiss S: **Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor**. *Proc Natl Acad Sci U S A* 1996, **93**:6264-6268.

4. Lohman TM, Ferrari ME: **Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities**. *Annu Rev Biochem* 1994, **63**:527-570.

5. Wold MS: **Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism**. *Annu Rev Biochem* 1997, **66**:61-92.

6. Nam EA, Cortez D: **SOSS1/2: Sensors of Single-Stranded DNA at a Break**. *Molecular Cell* 2009, **35**:258-259.

7. Richard DJ, Bolderson E, Khanna KK: **Multiple human single-stranded DNA binding proteins function in genome maintenance: structural, biochemical and functional analysis**. *Crit Rev Biochem Mol Biol* 2009, **44**:98-116.

8. Shereda RD, Kozlov AG, Lohman TM, Cox MM, Keck JL: **SSB as an Organizer/Mobilizer of Genome Maintenance Complexes**. *Critical Reviews in Biochemistry and Molecular Biology* 2008, **43**:289-318.

9. Theobald DL, Mitton-Fry RM, Wuttke DS: **Nucleic acid recognition by OB-fold proteins**. *Annu Rev Biophys Biomol Struct* 2003, **32**:115-133.

10. Murzin AG: **OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences**. *EMBO J* 1993, **12**:861-867.

11. Flynn RL, Zou L: **Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians**. *Crit Rev Biochem Mol Biol* 2010, **45**:266-275.

12. Alberts B, Johnson A, Lewis J, Raff MC, Roberts K, Walter P, Wilson J, Hunt T: *Molecular biology of the cell* edn fifth. New York: Garland Science; 2008.

13. Agrawal N, Dasaradhi PV, Mohmmed A, Malhotra P, Bhatnagar RK, Mukherjee SK: **RNA interference: biology, mechanism, and applications**. *Microbiol Mol Biol Rev* 2003, **67**:657-685.

14. Martin KC, Ephrussi A: **mRNA localization: gene expression in the spatial dimension**. *Cell* 2009, **136**:719-730.

15. Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, Hughes TR, Tomancak P, Krause HM: **Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function**. *Cell* 2007, **131**:174-187.

16. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB: **Gene regulation at the single-cell level**. *Science* 2005, **307**:1962-1965.

17. Tang F, Lao K, Surani MA: **Development and applications of single-cell transcriptome analysis**. *Nat Methods* 2011, **8**:S6-11.

18. Li GW, Xie XS: **Central dogma at the single-molecule level in living cells**. *Nature* 2011, **475**:308-315.

19. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S: **Imaging individual mRNA molecules using multiple singly labeled probes**. *Nat Methods* 2008, **5**:877-879.

20. Femino AM, Fay FS, Fogarty K, Singer RH: **Visualization of single RNA transcripts in situ**. *Science* 1998, **280**:585-590.

21. Hovelmann F, Gaspar I, Loibl S, Ermilov EA, Roder B, Wengel J, Ephrussi A, Seitz O: **Brightness through local constraint--LNA-enhanced FIT hybridization probes for in vivo ribonucleotide particle tracking**. *Angew Chem Int Ed Engl* 2014, **53**:11370-11375.

22. Tyagi S: **Imaging intracellular RNA distribution and dynamics in living cells**. *Nat Methods* 2009, **6**:331-338.
23. Bao G, Rhee WJ, Tsourkas A: **Fluorescent probes for live-cell RNA detection**. *Annu Rev Biomed Eng* 2009, **11**:25-47.

24. Wadia JS, Dowdy SF: **Protein transduction technology**. *Curr Opin Biotechnol* 2002, **13**:52-56.

25. Giles RV, Spiller DG, Grzybowski J, Clark RE, Nicklin P, Tidd DM: **Selecting optimal oligonucleotide composition for maximal antisense effect following streptolysin O-mediated delivery into human leukaemia cells**. *Nucleic Acids Res* 1998, **26**:1567-1575.

26. Santangelo PJ, Lifland AW, Curt P, Sasaki Y, Bassell GJ, Lindquist ME, Crowe JE, Jr.: **Single molecule-sensitive probes for imaging RNA in live cells**. *Nat Methods* 2009, **6**:347-349.

27. Belaya K, St Johnston D: **Using the mRNA-MS2/MS2CP-FP system to study mRNA transport during Drosophila oogenesis**. *Methods Mol Biol* 2011, **714**:265-283.

28. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM: **Localization of ASH1 mRNA particles in living yeast**. *Mol Cell* 1998, **2**:437-445.

29. Lionnet T, Czaplinski K, Darzacq X, Shav-Tal Y, Wells AL, Chao JA, Park HY, de Turris V, Lopez-Jones M, Singer RH: **A transgenic mouse for in vivo detection of endogenous labeled mRNA**. *Nat Methods* 2011, **8**:165-170.

30. Larson DR, Zenklusen D, Wu B, Chao JA, Singer RH: **Real-time observation of transcription initiation and elongation on an endogenous yeast gene**. *Science* 2011, **332**:475-478.

31. Lim F, Downey TP, Peabody DS: **Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7**. *J Biol Chem* 2001, **276**:22507-22513.

32. Cilley CD, Williamson JR: **Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE)**. *RNA* 1997, **3**:57-67.

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

34. Ozawa T, Natori Y, Sato M, Umezawa Y: **Imaging dynamics of endogenous mitochondrial RNA in single living cells**. *Nat Methods* 2007, **4**:413-419.

35. Rath AK, Rentmeister A: **Genetically encoded tools for RNA imaging in living cells**. *Curr Opin Biotechnol* 2014, **31C**:42-49.

36. Wu B, Chao JA, Singer RH: **Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells**. *Biophys J* 2012, **102**:2936-2944.

37. Golding I, Cox EC: **RNA dynamics in live Escherichia coli cells**. *Proc Natl Acad Sci U S A* 2004, **101**:11310-11315.

38. Golding I, Paulsson J, Zawilski SM, Cox EC: **Real-time kinetics of gene activity in individual bacteria**. *Cell* 2005, **123**:1025-1036.

39. Craggs TD: **Green fluorescent protein: structure, folding and chromophore maturation**. *Chem Soc Rev* 2009, **38**:2865-2875.

40. Dean KM, Palmer AE: **Advances in fluorescence labeling strategies for dynamic cellular imaging**. *Nat Chem Biol* 2014, **10**:512-523.

41. Shin I, Ray J, Gupta V, Ilgu M, Beasley J, Bendickson L, Mehanovic S, Kraus GA, Nilsen-Hamilton M: **Live-cell imaging of Pol II promoter activity to monitor gene expression with RNA IMAGEtag reporters**. *Nucleic Acids Res* 2014, **42**:e90.

42. Babendure JR, Adams SR, Tsien RY: **Aptamers switch on fluorescence of triphenylmethane dyes**. *J Am Chem Soc* 2003, **125**:14716-14717.

43. Constantin TP, Silva GL, Robertson KL, Hamilton TP, Fague K, Waggoner AS, Armitage BA: **Synthesis of new fluorogenic cyanine dyes and incorporation into RNA fluoromodules**. *Org Lett* 2008, **10**:1561-1564.
44. Lee J, Lee KH, Jeon J, Dragulescu-Andrasi A, Xiao F, Rao J: **Combining SELEX screening and rational design to develop light-up fluorophore-RNA aptamer pairs for RNA tagging**. *ACS Chem Biol* 2010, **5**:1065-1074.

45. Richards GR, Vanderpool CK: **Molecular call and response: the physiology of bacterial small RNAs**. *Biochim Biophys Acta* 2011, **1809**:525-531.

46. Gottesman S, McCullen CA, Guillier M, Vanderpool CK, Majdalani N, Benhammou J, Thompson KM, FitzGerald PC, Sowa NA, FitzGerald DJ: **Small RNA regulators and the bacterial response to stress**. *Cold Spring Harb Symp Quant Biol* 2006, **71**:1-11.

47. Schweizer HP: **Understanding efflux in Gram-negative bacteria: opportunities for drug discovery**. *Expert Opin Drug Discov* 2012, **7**:633-642.

48. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL: **The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae**. *Cell* 2004, **118**:69-82.

49. Li J, Attila C, Wang L, Wood TK, Valdes JJ, Bentley WE: **Quorum sensing in Escherichia coli is signaled by AI-2/LsrR: effects on small RNA and biofilm architecture**. *J Bacteriol* 2007, **189**:6011-6020.

50. Bobrovskyy M, Vanderpool CK: **The small RNA SgrS: roles in metabolism and pathogenesis of enteric bacteria**. *Front Cell Infect Microbiol* 2014, **4**:61.

51. Gottesman S, Storz G, Rosenow C, Majdalani N, Repoila F, Wassarman KM: **Small RNA regulators of translation: mechanisms of action and approaches for identifying new small RNAs**. *Cold Spring Harb Symp Quant Biol* 2001, **66**:353-362.

52. De Lay N, Schu DJ, Gottesman S: **Bacterial small RNA-based negative regulation: Hfq and its accomplices**. *J Biol Chem* 2013, **288**:7996-8003.

53. Lease RA, Smith D, McDonough K, Belfort M: **The small noncoding DsrA RNA is an acid resistance regulator in Escherichia coli**. *J Bacteriol* 2004, **186**:6179-6185.

54. Domenech P, Honore N, Heym B, Cole ST: **Role of OxyS of Mycobacterium tuberculosis in oxidative stress: overexpression confers increased sensitivity to organic hydroperoxides**. *Microbes Infect* 2001, **3**:713-721.

55. Raghunathan S, Ricard CS, Lohman TM, Waksman G: **Crystal structure of the homo-tetrameric DNA binding domain of Escherichia coli single-stranded DNA-binding protein determined by multiwavelength x-**

ray diffraction on the selenomethionyl protein at 2.9-A resolution. *Proc Natl Acad Sci U S A* 1997, **94**:6652-6657.

56. Raghunathan S, Kozlov AG, Lohman TM, Waksman G: **Structure of the DNA binding domain of E. coli SSB bound to ssDNA**. *Nat Struct Biol* 2000, **7**:648-652.

57. Yang C, Curth U, Urbanke C, Kang CH: **Crystal structure of human mitochondrial single stranded DNA binding protein at 2.4 angstrom resolution**. *Nature Structural Biology* 1997, **4**:153-157.

58. Fanning E, Klimovich V, Nager AR: **A dynamic model for replication protein A (RPA) function in DNA processing pathways**. *Nucleic Acids Res* 2006, **34**:4126-4137.

59. Dabrowski S, Olszewski M, Piatek R, Kur J: **Novel thermostable ssDNA-binding proteins from Thermus thermophilus and T. aquaticus-expression and purification**. *Protein Expr Purif* 2002, **26**:131-138.

60. Dabrowski S, Olszewski M, Piatek R, Brillowska-Dabrowska A, Konopa G, Kur J: **Identification and characterization of single-stranded-DNA-binding proteins from Thermus thermophilus and Thermus aquaticus - new arrangement of binding domains**. *Microbiology* 2002, **148**:3307-3315.

61. Bernstein DA, Eggington JM, Killoran MP, Misic AM, Cox MM, Keck JL: **Crystal structure of the Deinococcus radiodurans single-stranded DNA-binding protein suggests a mechanism for coping with DNA damage**. *Proc Natl Acad Sci U S A* 2004, **101**:8575-8580.

62. Eggington JM, Haruta N, Wood EA, Cox MM: **The single-stranded DNA-binding protein of Deinococcus radiodurans**. *BMC Microbiol* 2004, **4**:2.

63. Jedrzejczak R, Dauter M, Dauter Z, Olszewski M, Dlugolecka A, Kur J: **Structure of the single-stranded DNA-binding protein SSB from Thermus aquaticus**. *Acta Crystallographica Section D-Biological Crystallography* 2006, **62**:1407-1412.

64. Fedorov R, Witte G, Urbanke C, Manstein DJ, Curth U: **3D structure of Thermus aquaticus single-stranded DNA-binding protein gives insight into the functioning of SSB proteins**. *Nucleic Acids Research* 2006, **34**:6708-6717.

65. Roy R, Kozlov AG, Lohman TM, Ha T: **SSB protein diffusion on single-stranded DNA stimulates RecA filament formation**. *Nature* 2009, **461**:1092-1097.

66. Zhou R, Kozlov AG, Roy R, Zhang J, Korolev S, Lohman TM, Ha T: **SSB functions as a sliding platform that migrates on DNA via reptation**. *Cell* 2011, **146**:222-232.

67. Kozlov AG, Eggington JM, Cox MM, Lohman TM: **Binding of the dimeric Deinococcus radiodurans single-stranded DNA binding protein to single-stranded DNA**. *Biochemistry* 2010, **49**:8266-8275.

68. Kumaran S, Kozlov AG, Lohman TM: **Saccharomyces cerevisiae replication protein A binds to single-stranded DNA in multiple salt-dependent modes**. *Biochemistry* 2006, **45**:11958-11973.

69. Roy R, Kozlov AG, Lohman TM, Ha T: **Dynamic structural rearrangements between DNA binding modes of E. coli SSB protein**. *Journal of Molecular Biology* 2007, **369**:1244-1257.

70. Filipkowski P, Koziatek M, Kur J: **A highly thermostable, homodimeric single-stranded DNA-binding protein from Deinococcus radiopugnans**. *Extremophiles* 2006, **10**:607-614.

71. Filipkowski P, Duraj-Thatte A, Kur J: **Novel thermostable single-stranded DNA-binding protein (SSB) from Deinococcus geothermalis**. *Arch Microbiol* 2006, **186**:129-137.

72. Filipkowski P, Duraj-Thatte A, Kur J: **Identification, cloning, expression, and characterization of a highly thermostable single-stranded-DNA-binding protein (SSB) from Deinococcus murrayi**. *Protein Expr Purif* 2007, **53**:201-208.

73. Filipkowski P, Kur J: **Identification and properties of the Deinococcus grandis and Deinococcus proteolyticus single-stranded DNA binding proteins (SSB)**. *Acta Biochim Pol* 2007, **54**:79-87.

74. Witte G, Fedorov R, Curth U: **Biophysical analysis of Thermus aquaticus single-stranded DNA binding protein**. *Biophys J* 2008, **94**:2269-2279.

75. Witte G, Urbanke C, Curth U: **Single-stranded DNA-binding protein of Deinococcus radiodurans: a biophysical characterization**. *Nucleic Acids Res* 2005, **33**:1662-1670.

76. Eggington JM, Kozlov AG, Cox MM, Lohman TM: **Polar destabilization of DNA duplexes with single-stranded overhangs by the Deinococcus radiodurans SSB protein**. *Biochemistry* 2006, **45**:14490-14502.

77. Ha T, Enderle T, Ogletree DF, Chemla DS, Selvin PR, Weiss S: **Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**:6264-6268.

78. Murphy MC, Rasnik I, Cheng W, Lohman TM, Ha T: **Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy**. *Biophys J* 2004, **86**:2530-2537.

79. Antony E, Weiland EA, Korolev S, Lohman TM: **Plasmodium falciparum SSB tetramer wraps single-stranded DNA with similar topology but opposite polarity to E. coli SSB**. *J Mol Biol* 2012, **420**:269-283.
80. Zhou R, Ha T: **Single-molecule analysis of SSB dynamics on single-stranded DNA**. *Methods Mol Biol* 2012, **922**:85-100.

81. Kim HD, Nienhaus GU, Ha T, Orr JW, Williamson JR, Chu S: **Mg2+-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules**. *Proc Natl Acad Sci U S A* 2002, **99**:4284-4289.

82. Karymov M, Daniel D, Sankey OF, Lyubchenko YL: **Holliday junction dynamics and branch migration: single-molecule analysis**. *Proc Natl Acad Sci U S A* 2005, **102**:8186-8191.

83. Joo C, McKinney SA, Lilley DM, Ha T: **Exploring rare conformational species and ionic effects in DNA Holliday junctions using single-molecule spectroscopy**. *J Mol Biol* 2004, **341**:739-751.

84. Antony E, Kozlov AG, Nguyen B, Lohman TM: **Plasmodium falciparum SSB Tetramer Binds Single-Stranded DNA Only in a Fully Wrapped Mode**. *Journal of Molecular Biology* 2012, **420**:284-295.

85. McKinney SA, Joo C, Ha T: **Analysis of single-molecule FRET trajectories using hidden Markov modeling**. *Biophysical Journal* 2006, **91**:1941-1951.

86. Perales C, Cava F, Meijer WJJ, Berenguer J: **Enhancement of DNA, cDNA synthesis and fidelity at high temperatures by a dimeric single-stranded DNA-binding protein**. *Nucleic Acids Research* 2003, **31**:6473-6480.

87. Kur J, Olszewski M, Dlugolecka A, Filipkowski P: **Single-stranded DNA-binding proteins (SSBs) - sources and applications in molecular biology**. *Acta Biochimica Polonica* 2005, **52**:569-574.

88. Inoue J, Shigemori Y, Mikawa T: **Improvements of rolling circle amplification (RCA) efficiency and accuracy using Thermus thermophilus SSB mutant protein**. *Nucleic Acids Research* 2006, **34**.

89. Inoue J, Honda M, Ikawa S, Shibata T, Mikawa T: **The process of displacing the single-stranded DNA-binding protein from single-stranded DNA by RecO and RecR proteins**. *Nucleic Acids Res* 2008, **36**:94-109.

90. Zhou R, Kozlov AG, Roy R, Zhang JC, Korolev S, Lohman TM, Ha T: **SSB Functions as a Sliding Platform that Migrates on DNA via Reptation (vol 146, pg 222, 2011)**. *Cell* 2011, **146**:485-485.

91. Paige JS, Wu KY, Jaffrey SR: **RNA mimics of green fluorescent protein**. *Science* 2011, **333**:642-646.

92. Strack RL, Disney MD, Jaffrey SR: **A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA**. *Nat Methods* 2013, **10**:1219-1224.

93. Dolgosheina EV, Jeng SC, Panchapakesan SS, Cojocaru R, Chen PS, Wilson PD, Hawkins N, Wiggins PA, Unrau PJ: **RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking**. *ACS Chem Biol* 2014, **9**:2412-2420.

94. Filonov GS, Moon JD, Svensen N, Jaffrey SR: **Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution**. *J Am Chem Soc* 2014, **136**:16299-16308.

95. Pothoulakis G, Ceroni F, Reeve B, Ellis T: **The spinach RNA aptamer as a characterization tool for synthetic biology**. *ACS Synth Biol* 2014, **3**:182-187.

96. Huang H, Suslov NB, Li NS, Shelke SA, Evans ME, Koldobskaya Y, Rice PA, Piccirilli JA: **A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore**. *Nat Chem Biol* 2014, **10**:686-691.

97. Warner KD, Chen MC, Song W, Strack RL, Thorn A, Jaffrey SR, Ferre-D'Amare AR: **Structural basis for activity of highly efficient RNA mimics of green fluorescent protein**. *Nat Struct Mol Biol* 2014, **21**:658-663.

98. You M, Jaffrey SR: **Structure and Mechanism of RNA Mimics of Green Fluorescent Protein**. *Annu Rev Biophys* 2015, **44**:187-206.

99. Paige JS, Nguyen-Duc T, Song W, Jaffrey SR: **Fluorescence imaging of cellular metabolites with RNA**. *Science* 2012, **335**:1194.

100. Song W, Strack RL, Jaffrey SR: **Imaging bacterial protein expression using genetically encoded RNA sensors**. *Nat Methods* 2013, **10**:873-875.

101. Hocine S, Raymond P, Zenklusen D, Chao JA, Singer RH: **Single-molecule analysis of gene expression using two-color RNA labeling in live yeast**. *Nat Methods* 2013, **10**:119-121.

102. Ha T: **Single-molecule approaches embrace molecular cohorts**. *Cell* 2013, **154**:723-726.

103. Han KY, Leslie BJ, Fei J, Zhang J, Ha T: **Understanding the photophysics of the spinach-DFHBI RNA aptamer-fluorogen complex to improve live-cell RNA imaging**. *J Am Chem Soc* 2013, **135**:19033-19038.

104. Ponchon L, Dardel F: **Recombinant RNA technology: the tRNA scaffold**. *Nat Methods* 2007, **4**:571-576.
105. Nevo-Dinur K, Nussbaum-Shochat A, Ben-Yehuda S, Amster-Choder O: **Translation-independent localization of mRNA in E. coli**. *Science* 2011, **331**:1081-1084.

106. Nath K, Koch AL: **Protein degradation in Escherichia coli. II. Strain differences in the degradation of protein and nucleic acid resulting from starvation**. *J Biol Chem* 1971, **246**:6956-6967.

107. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN: **Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays**. *Proc Natl Acad Sci U S A* 2002, **99**:9697-9702.

108. Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C: **Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation**. *Genome Res* 2003, **13**:216-223.

109. Jach G, Pesch M, Richter K, Frings S, Uhrig JF: **An improved mRFP1 adds red to bimolecular fluorescence complementation**. *Nat Methods* 2006, **3**:597-600.

110. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY: **A monomeric red fluorescent protein**. *Proc Natl Acad Sci U S A* 2002, **99**:7877-7882.

111. Gourse RL, Gaal T, Bartlett MS, Appleman JA, Ross W: **rRNA transcription and growth rate-dependent regulation of ribosome synthesis in Escherichia coli**. *Annu Rev Microbiol* 1996, **50**:645-677.

112. Miroux B, Walker JE: **Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels**. *J Mol Biol* 1996, **260**:289-298.

113. Terpe K: **Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems**. *Appl Microbiol Biotechnol* 2006, **72**:211-222.

114. Robin FB, McFadden WM, Yao B, Munro EM: **Single-molecule analysis of cell surface dynamics in Caenorhabditis elegans embryos**. *Nat Methods* 2014, **11**:677-682.

115. Curtis EA, Liu DR: **A naturally occurring, noncanonical GTP aptamer made of simple tandem repeats**. *RNA Biol* 2014, **11**:682-692.

116. Desnoyers G, Bouchard MP, Masse E: **New insights into small RNA-dependent translational regulation in prokaryotes**. *Trends Genet* 2013, **29**:92-98.

117. Borgia MB, Borgia A, Best RB, Steward A, Nettels D, Wunderlich B, Schuler B, Clarke J: **Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins**. *Nature* 2011, **474**:662-665.

118. Stigler J, Ziegler F, Gieseke A, Gebhardt JC, Rief M: **The complex folding network of single calmodulin molecules**. *Science* 2011, **334**:512-516.

119. Schmittgen TD, Livak KJ: **Analyzing real-time PCR data by the comparative C(T) method**. *Nat Protoc* 2008, **3**:1101-1108.

120. Neidhardt FC, Ingraham JL, Schaechter M: *Physiology of the Bacterial cell: A Molecular Approach*. Sunderland, MA: Sinauer associates; 1990.

121. So LH, Ghosh A, Zong C, Sepulveda LA, Segev R, Golding I: **General properties of transcriptional time series in Escherichia coli**. *Nat Genet* 2011, **43**:554-560.

122. Storz G, Vogel J, Wassarman KM: **Regulation by small RNAs in bacteria: expanding frontiers**. *Mol Cell* 2011, **43**:880-891.

123. Masse E, Vanderpool CK, Gottesman S: **Effect of RyhB small RNA on global iron use in Escherichia coli**. *J Bacteriol* 2005, **187**:6962-6971.

124. Rice JB, Vanderpool CK: **The small RNA SgrS controls sugar-phosphate accumulation by regulating multiple PTS genes**. *Nucleic Acids Res* 2011, **39**:3806-3819.

125. Rice JB, Balasubramanian D, Vanderpool CK: **Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mRNA**. *Proc Natl Acad Sci U S A* 2012, **109**:E2691-2698.

126. Papenfort K, Sun Y, Miyakoshi M, Vanderpool CK, Vogel J: **Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis**. *Cell* 2013, **153**:426-437.

127. Wadler CS, Vanderpool CK: **A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide**. *Proc Natl Acad Sci U S A* 2007, **104**:20454-20459.

128. Fei J, Singh D, Zhang Q, Park S, Balasubramanian D, Golding I, Vanderpool CK, Ha T: **RNA biochemistry. Determination of in vivo target search kinetics of regulatory noncoding RNA**. *Science* 2015, **347**:1371-1374.

129. Rutherford ST, Valastyan JS, Taillefumier T, Wingreen NS, Bassler BL: **Comprehensive analysis reveals how single nucleotides contribute to noncoding RNA function in bacterial quorum sensing**. *Proc Natl Acad Sci U S A* 2015, **112**:E6038-6047.

130. Kinney JB, Murugan A, Callan CG, Jr., Cox EC: **Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence**. *Proc Natl Acad Sci U S A* 2010, **107**:9158-9163.

131. Peterman N, Lavi-Itzkovitz A, Levine E: **Large-scale mapping of sequence-function relations in small regulatory RNAs reveals plasticity and modularity**. *Nucleic Acids Res* 2014, **42**:12177-12188.

132. Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C: **Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform**. *Nucleic Acids Res* 2015, **43**:e37.