UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Control the Function of RNA by Site-Specific Bioconjugation

Permalink
https://escholarship.org/uc/item/6xk4b35h

Author
Zhang, Dongyang

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Control the Function of RNA by Site-Specific Bioconjugation

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Chemistry

by

Dongyang Zhang

Committee in charge:

       Professor Neal K. Devaraj, Chair
       Professor Joshua S. Figueroa
       Professor Xiang-Dong Fu
       Professor Tadeusz F. Molinski
       Professor Wei Wang

2020

The dissertation of Dongyang Zhang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Chair

University of California San Diego

2020

DEDICATION

I dedicate this work to my beloved family, sincere friends and mentors for their

unwavering love, support, encouragement and guidance. I would not have

accomplished my Ph.D. journey without them.

EPIGRAPH

The measure of intelligence

is the ability to change.

—Albert Einstein

TABLE OF CONTENTS

LIST OF FIGURES

2015	B. S. in Chemistry, Nanjing University, Jiangsu, China
2015-2020	Graduate Teaching Assistant, University of California, San Diego
2020	Ph. D. in Chemistry, University of California, San Diego

PUBLICATIONS

Zhang, D., Jin, S., Liu, L., Tota, E., & Devaraj, N. K. (2020). "Multiplexed photo-activation of CRISPR-Cas9 gene editing with single-cell resolution." *manuscript in preparation*.

Zhang, D., Jin, S., Piao, X., & Devaraj, N. K. (2020). "Multiplexed photo-activation of mRNA with single-cell resolution." *ACS Chemical Biology*, 15(7), 1773-1779.

Zhang, D., Zhou, C. Y., Busby, K. N., Alexander, S. C., & Devaraj, N. K. (2018). "Light-activated control of translation by enzymatic covalent mRNA Labeling." *Angewandte Chemie*, 130(11), 2872-2876.

Liu, L., Zhang, D., Mai, J., & Devaraj, N. K. (2020). "Live-cell compatible light triggered biorthogonal ligations enabled by photocaged-dihydrotetrazines." *manuscript submitted*.

Qian, H., Kang, X., Hu, J., Zhang, D., Liang, Z., Meng, F., Zhang, X., Xue, R., Maimon, S. D., Devaraj, N. K., Zhou, Z., Mobley, W. C., Cleveland, D. W., & Fu, Xiang-Dong. (2020). "Reversing a model of Parkinson's disease with in situ converted nigral neurons." *Nature*, 582(7813), 550-556.

Liu, L., Zou, Y., Bhattacharya, A., Zhang, D., Lang, S. Q., Houk, K. N., & Devaraj, N. K. (2020). "Enzyme-free synthesis of natural phospholipids in water." *Nature Chemistry*, 12, 1029–1034.

Busby, K. N., Fulzele, A., Zhang, D., Bennett, E. J., & Devaraj, N. K. (2020). "Enzymatic RNA Biotinylation for Affinity Purification and Identification of RNA-protein Interactions." *ACS Chemical Biology*, 15(8), 2247-2258.

Ehret, F., Zhou, C. Y., Alexander, S. C., Zhang, D., & Devaraj, N. K. (2017). "Site-specific covalent conjugation of modified mRNA by tRNA guanine transglycosylase." *Molecular Pharmaceutics*, 15(3), 737-742.

ABSTRACT OF THE DISSERTATION

Control the Function of RNA by Site-Specific Bioconjugation

by

Dongyang Zhang

Doctor of Philosophy in Chemistry

University of California San Diego, 2020

Professor Neal K. Devaraj, Chair

RNA is one of the most important biomacromolecules in the living systems, manipulating a highly complexed collection of functions which are critical to the regulation of numerous cellular processes. In the past few years, our lab has developed a versatile and powerful RNA modifying technique, named RNA-TAG. The RNA-TAG technique utilizes a bacterial tRNA guanine transglycosylase (TGT) to exchange a guanine nucleobase within a specific 17-nucleotide motif (Tag) for synthetic pre-queuosine1 (preQ1) derivatives. By inserting this Tag sequence into an RNA of interest, we can covalently and site-specifically conjugate functional small-molecules onto any RNA of interest. I applied our RNA-TAG technique to manipulate the functions of different types of cellular RNA, for example, mRNA for translation regulation and the singe guide

RNA (sgRNA) for CRISPR/Cas9 gene editing.

To interfere with the translation process of mature mRNA, I systematically conjugated three bulky visible-light photocleavable caging groups along the 5' untranslated region (5'-UTR) of an mRNA, severely reducing its translation activity. Upon visible-light photocleavage of these caging groups, mRNA translation was resorted. To expand our technology, we further designed two photo-sensitive caging groups which can be sequentially cleaved by two wavelength of lights (405 nm and 488 nm). In this manner, a sequential/multiplexed photo-activation of two mRNAs within the same cells was achieved.

Based on our RNA-TAG technique, I developed a novel technique named RNA-CLAMP, which can covalently and site-specifically crosslink ('clamp') two internal stem-loops within a RNA of interest. By designing small-molecule substrates which contain a cleavable linker and two preQ1 moieties, I achieved accurate 'clamping' and rapid release of the target RNA, which significantly alter its secondary/ternary structure, resulting in gain/loss of functions. I applied the RNA-CLAMP technique to the sgRNA of the CRISPR/Cas9 gene editing system. Upon 'clamping' of two internal loops (the tetra-loop and stem-loop2) of the sgRNA, the Cas9 gene editing system was completely deactivated. However, the cleavage of the crosslinker with an external stimulus, for example, light irradiation, released the sgRNA and fully activated gene editing in live mammalian cells. To the best of knowledge, this CRISPR/Cas9 gene editing photo-regulation technology offers the best dynamic range, flexibility, multiplexing capability and accessibility to date. We believe with further development, the RNA-CLAMP technique will have tremendous applications in manipulating other types of RNAs, such as ribozymes, microRNAs, and long non-coding RNAs. Such techniques will serve as powerful tools in studying complex cellular networks as well as developing novel therapeutics applications.

Chapter 1

Introduction

1.1   Overview of RNA-modifying techniques

RNA is one of the most important biomacromolecules in the living systems, manipulating a highly complex collection of functions which are critical to the regulation of numerous cellular pathways and processes.[1] Being the cornerstone of biology's central dogma, numerous approached has been developed to study and manipulate the functions of RNAs. However, compared to the study of proteins and DNAs/chromosomes, our understanding of RNA's cellular function is significantly lacking. This is partially because of the transient nature of RNA molecule. The half-life of RNA is significantly shorter than DNA and protein.[2,3] Besides, the detection of RNA suffers from low copy number as low as one copy per cell.[4] Many creative methodologies have been developed in the past few decades to address this challenging question: how to label and manipulate cellular RNAs. In this part of the introduction, I aim to briefly introduce some of the most recent and exciting techniques which enable the detection, visualization and manipulation of cellular RNAs.

One of the most well-established techniques to study cellular RNAs is the nucleic acid antisense in situ hybridization technique.[5,6] Antisense oligonucleotides labeled with different types

of fluorophores are directly conjugated to detect cellular RNA. Together with the development of high-resolution fluorescence microscopy, the RNA fluorescence in situ hybridization (RNA-FISH) method has become a more accessible technique. However, the RNA-FISH approaches have certain limitations. First, multiple RNA-FISH probes are usually required to target a single RNA molecule in order to image shorter and less abundant RNA targets.[7] Second, off-target binding and mismatches also limit the application of RNA-FISH approaches. Thus, a variety of modifications have been made to improve the sensitivity and specificity of the RNA-FISH method. For example, molecular beacons or FRET donor and acceptor hybridizing probes can be used to improve both the signal-to-noise ratio and specificity.[8,9]

Functional cellular RNAs are usually bound with one or more proteins, the RNA binding proteins (RBPs). Thus, a RBP fused to a fluorescence reporter, such as a fluorescence protein, can be used to detect and image RNA in live cells. The most successful development of this approach is the use of the RNA phage capsid protein MS2.[10] In this case, a short RNA hairpin structure is required to promote the binding of the RNA of interest with MS2. For example, an imaging tool was developed by the Singer lab back in 1998 which enables the imaging and tracking of RNAs in live yeast.[11] The issues with these approaches are: [1] the unbound MS2-GFP causes high imaging background and [2] the low abundance of some RNA targets severely limits the applications of these approaches. With many recent improvements and optimizations, the MS2 system is still a workhorse for live cell RNA imaging.[12–15]

Recently, fluorogenic RNA aptamers have also become a very powerful method for imaging RNAs of interest in live cells. In 2011, Jaffrey and his co-workers developed a live cell compatible 98-nucleotide RNA mimic of GFP, the Spinach aptamer, that binds to an analogue of GFP's chromophore, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) and turn on its fluorescence.[16] Some advancements were achieved after the initial development of the Spinach aptamer. For example, an optimized "superfolding" Spinach II was developed which resulted in brighter fluorescence signals.[17] Besides, a plug-and-play set of derivatives of the HBI

fluorogenic core enables different colors for imaging RNA of interests.[18] These successful RNA aptamer approaches opened up the potentials for live cell RNA imaging. However, they do have certain limitations. First, these aptamer approaches require a ¿50-nucleotide RNA structure to be encoded in the RNA of interest. The aptamer sequences must be properly folded in cellular conditions, limiting this approach to applications for RNAs that do not have rigid sequence length or structural requirements. Additionally, a pro-fluorescent small molecule must also be delivered together with the aptamers to achieve binding and fluorescence activation.

1.2 Site-Specific Bioconjugation of RNA by Enzymatic Transglycosylation at Guanosine (RNA-TAG)

Apart from all the approaches mentioned above, covalent RNA-modifying approaches have been challenging because of the difficulties in selectively modifying a single RNA of interest among the other RNAs in cellular conditions. Comparing to non-covalent interactions, covalent strategies provide an additional level of robustness in harsh cellular conditions. Due to the covalent linkage, the conjugated functional groups will not be disassociated from the RNA of interest in most conditions. Besides, the low-molecular weight of small-molecule (< 2 kDa) minimize the perturbation of normal RNA functions. While many covalent RNA-modifying approaches have been developed, few methods allow for the selective labeling of a single post-transcriptional RNA among the complex cellular RNA pool.[19]

To achieve covalent modification of RNA, researchers have designed modified nucleotides which can be then incorporated onto the RNA via transcription or solid-state synthesis. However, large functional groups can not be used in these approaches, limiting their applications. Additionally, solid-state-synthesis is not widely accessible in many cases. Thus, a method that enables post-transcriptional modification of RNA is high desired. Early work by Richardson and Gumport used T4 RNA ligase to conjugate functional small-molecules such as biotin, fluorophores onto

the 3' end of RNA transcripts.[20] However, this approach is unselective and the location of the modification is limited at the 3' end of the RNA. The labeling yield was also poor. As another example, Helm, etc., took advantage of the promiscuity of a t-RNA methyltransferase, Trm1, which then transfers a methyl group from S-adenosyl-L-methionine (AdoMet) to the N2 atom of Phe 71 guanosine 26 in the tRNA.[21] Furthermore, by synthesizing alternative AdoMet analogues bearing a small alkyne click handle, tRNA can be selective labeled for a subsequent click chemistry. In this manner, a bigger functional group can be conjugated onto the RNA through this two-step conjugation strategy. Many other approaches have been developed using this two-step labeling methods.[22–25]

tRNAs are well-known to undergo enzymatic post-transcriptional modifications, inspiring us to dive into this natural process and try to adapt these modification enzymes for covalent modification of a RNA of interest. Recently, tRNA-modifying enzymes have shown promise in RNA labeling and imaging.[25,26] Despite their success, limitations still exist.

[1] a large motif is required to promote enzymatic labeling;

[2] millimolar concentrations for both RNA and enzyme are required.

In 2015, our group has harnessed a bacterial tRNA guanine transglycosylase (TGT) to conjugate functional reporters such as fluorophores and affinity handles site-specifically to an RNA of interest in a single enzymatic step.[27] *Escherichia coli* has been extensively characterized and is known to natively incorporate the nucleobase preQ1 into the G34 wobble position of tRNAs that bear a UGU in their anticodon stem loop.[28,29] E. coli TGT is unique in that it does not require the entire tRNA structure for substrate recognition but instead can efficiently recognize the anticodon stem loop region of the tRNA, even when encoded into other RNA sequences. E. coli TGT is exclusive and distinctive from its eukaryotic counterpart, both in small molecule substrate recognition (queuine vs preQ1), as well as in RNA recognition requirements (whole tRNA vs hairpin alone). These key differences provide a unique condition that allows labeling by *E. coli* TGT to remain selective for a short encoded hairpin. While extensive selectivity studies

have yet to be performed, we would expect few off-target sites to be labeled to any significant extent because of both sequence and structural requirements as outlined in past work.[29] We demonstrate the site-specific incorporation of nucleobase derivatives bearing fluorophores or affinity labels into a short RNA stem loop recognition motif by exchange of a guanine residue. The RNA-TAG (RNA transglycosylation at guanosine) is carried out by a bacterial (*E. coli*) tRNA guanine transglycosylase (TGT), whose natural substrate is the nitrogenous base preQ1. Remarkably, we have successfully incorporated large functional groups including biotin, BODIPY, thiazole orange, and Cy7 through a polyethylene glycol linker attached to the exocyclic amine of preQ1. Larger RNAs, such as mRNA transcripts, can be site-specifically labeled if they possess the 17-nucleotide hairpin recognition motif. The RNA-TAG methodology could facilitate the detection and manipulation of RNA molecules by enabling the direct incorporation of functional artificial nucleobases using a simple hairpin recognition element (Figure 1.1).
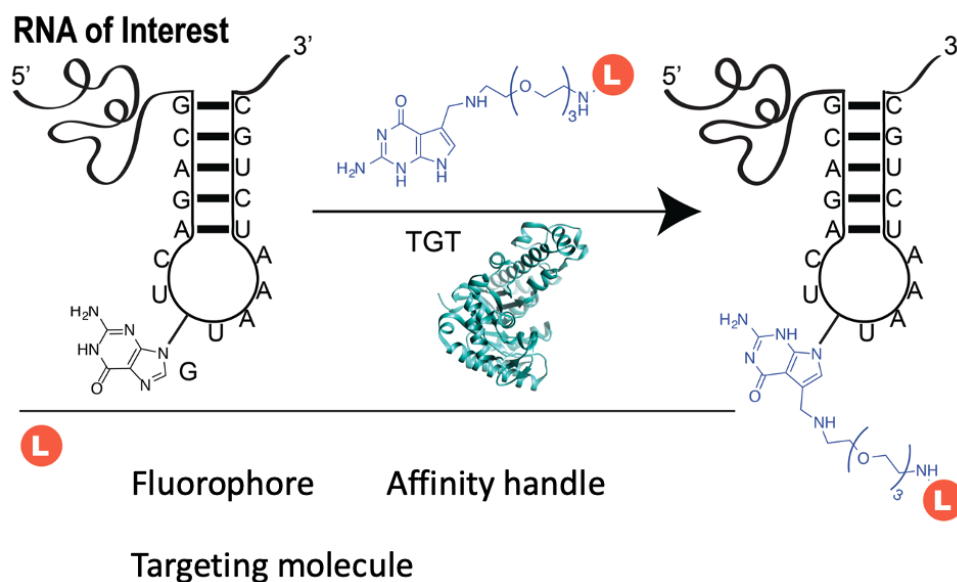
Figure 1.1: RNA Transglycosylation at Guanosine (RNA-TAG). TGT-catalyzed transglycosylation of an RNA of interest inserted with an ECY-A1 minihelix with preQ1 derivative nucleobases. Different functional small-molecules, for example, fluorophores, affinity handles and targeting molecules, can be covalently conjugated onto the RNA of interest via the one-step enzymatic labeling.

To extend the scope of the RNA-TAG methodology, we then developed an approach for

genetically encoded site- specific labeling of large mRNA transcripts, by employing mod-mRNA as substrate. As a proof of concept, we covalently attached a fluorescent probe to mCherry encoding mod-mRNA transcripts bearing 5-methylcytidine and/or pseudouridine substitutions with high labeling efficiencies. To provide a versatile labeling methodology with a wide range of possible applications, we employed a two-step strategy for functionalization of the mod-mRNA to highlight the therapeutic potential of this new methodology. We envision that this novel and facile labeling methodology of mod-RNA will have great potential in decorating both coding and noncoding therapeutic RNAs with a variety of diagnostic and functional moieties.[30]

Apart from RNA imaging cellular RNAs, we also demonstrated the ability to functionally label and isolate a target transcript by conjugating a biotin preQ1 derivative and selectively purifying the labeled transcript from total RNA extract with streptavidin-coated magnetic beads. Leveraging the ability to site-specifically and covalently label an RNA of interest using E. coli tRNA guanine transglycosylase and an unnatural nucleobase substrate, we establish the identification of RNA-protein interactions and the selective enrichment of cellular RNA in mammalian systems. We demonstrate the utility of this approach through the identification of known binding partners of 7SK snRNA via mass spectrometry. Through a minimal 4-nucleotide mutation of the long noncoding RNA HOTAIR, enzymatic biotinylation enables identification of putative HOTAIR binding partners in MCF7 breast cancer cells that suggest new potential pathways for oncogenic function. Furthermore, using RNA sequencing and qPCR, we establish that an engineered enzyme variant achieves high levels of labeling selectivity against the human transcriptome allowing for 145-fold enrichment of cellular RNA directly from mammalian cell lysates. The flexibility and breadth of this approach suggests that this system could be routinely applied to the functional characterization of RNA, greatly expanding the toolbox available for studying mammalian RNA biology.[31]

1.3 Photo-regulation of Cellular RNAs

Our RNA-TAG technique enables the incorporation of different types of functional small-molecules onto the RNA of interest. Have this versatile and powerful tool in hand, we were able to achieve precise and sophisticated manipulation of cellular RNA at a high spatiotemporal resolution. Here, I'll introduce two examples using our RNA-TAG technique, the photo-regulation of translation and the photo-regulation of the CRISPR/Cas9 genome editing.

1.3.1 Photo-regulation of Translation

Tools that allow researchers to manipulate gene expression are critical for basic research, biotechnology, and therapeutic applications.[32,33] There is an extensive body of research dedicated to controlling gene expression, at both the transcriptional and translational level, using approaches such as DNA/RNA methylation, histone modification and modulation of translation factors.[34–36] RNA biogenesis is a complex process, involving, for example, processing, transportation, and maturation, and this can lead to substantial lag time between stimulus and response in such transcriptional-based gene expression regulation systems.[37] On the other hand, directly controlling translation of mature mRNA involves much quicker response times. The cell controls various cellular functions and cell fate decisions in this manner, especially at early development stages, when rapid changes in protein concentration are needed.[38] More recently, IVT-mRNA has been widely used to selectively differentiate iPS cells, as well as to treat genetic disorders.[39] In comparison to DNA-based therapeutics, mRNA can be utilized by cells immediately upon entering the cytoplasm, leading to rapid onset of protein expression.[40] Furthermore, the use of mRNA minimizes the risk of causing insertional mutagenesis.

Regulating mRNA translation in eukaryotic cells typically focuses on the initiation stage, when initiation factors (IFs) assemble to form the 80S initiation complex. In 5' cap-dependent translation initiation, initiation factors elF4B and elF4F bind to the mRNA 5' cap, which then

recruit the 43S preinitiation complex.[41] This complex then scans through the 5' UTR until it encounters the start codon, which triggers the assembly of the ribosome and translation elongation. Accordingly, traditional strategies developed to tune the level of translation include the use of 5'- and 3'-UTR-binding factors and microRNAs (miRNAs), which non-covalently interact with target mRNA.[34,42] Alternative strategies using covalent RNA modifications have also been developed to regulate gene expression. For example, by modifying oligonucleotides such as antisense oligonucleotides, short interfering RNAs (siRNAs) or miRNAs with caging groups, optical control of translation can be achieved.[43–45] Apart from these nucleic acid-based approaches, modifying mRNA transcripts directly with small molecule effectors that are responsive to external stimuli can also lead to spatiotemporal regulation of gene expression. For example, Okamoto and his coworkers have developed strategies using diazo-coumarins to statistically cage mRNAs and photo-regulate gene expression.[46–49] To achieve straightforward mRNA labeling, an RNA modification strategy capable of incorporating a diverse array of functional groups onto RNA in a single step would be ideal. Furthermore, sequence specific modification can minimize over-modification, resulting in precise control of gene expression. This is where our RNA-TAG technique becomes handy.

Photo-regulation would offer rapid and non-invasive manipulation of target mRNA, and allow for the selective activation of protein expression with high spatiotemporal resolution. Furthermore, the ability to precisely control laser wavelength and illumination time allows the minimization of cell death.[50] Thus, a light-activated mRNA translation system would be an extremely useful method to regulate gene expression.

1.3.2 Photo-regulation of CRISPR/Cas9 Genome Editing

CRISPR/Cas9 genome editing has become an extremely active topic in both biological sciences and gene therapeutic applications. Tools that allow researchers to spatiotemporally control gene editing may promote study of gene functions with a high precision and reduce

off-targeting mutations by activating gene editing at a certain time and location.[51,52] We have seen a fast development of CRISPR/Cas9 gene editing techniques since 2012, when Martin Jinek, etc., discovered that Cas9 can act as a programmable RNA–guided DNA endonuclease.[53] Since then, many techniques have been developed to achieve chemical control of the CRISPR/Cas9 gene editing. For example, small-molecule inducers, such as doxycycline and rapamycin were used to induce gene editing in cells.[54,55] Also, conditional activation of CRISPR/Cas9 gene editing can also be achieved via transient delivery of purified Cas9:sgRNA complex.[56,57] These chemical methods have been used for generating conditional gene knockins/knockouts and reducing off-target gene editing. However, these approaches have limitations.

[1] some small-molecule inducers, for example rapamycin, have undesirable biological effects and toxicity;

[2] the slow diffusion rate of small-molecule inducers makes the regulation of gene editing less responsive and precise.

On the other hand, optical control of gene editing can overcome these challenges. First, a high spatiotemporal resolution can be achieved by the use of light. Second, by controlling the irradiation time period and light intensity, toxic effects can be minimized. Third, different wavelength of lights can be used to regulation different cellular process, enabling multiplexed manipulation of certain cellular processes. To date, there has been several approaches developed to achieve optical control of CRISPR/Cas9 gene editing. These techniques can be classified into two categories: the protein (Cas9) centric approaches and the guide RNA centric approaches.

Protein centric approaches aim to modify some key amino acid residues on the Cas9 in order to control the protein functions. Some alternative strategies use a split version of Cas9 that are fused to a pair of photo-dimerizing proteins, to achieve photo-control of CRISPR/Cas9. For example, James Hemphill, etc. developed a genetically encoded light-activated Cas9 by engineering a caged lysine amino acid within the Cas9 protein.[58] In this manner, several lysine residues were identified as possible caging sites. By irradiation with 365 nm UV light, the

activity of the caged Cas9 can be restored up to wild-type level. For another example, Yuta Nihongaki and the co-workers engineered a photo-activatable Cas9 which consists of split Cas9 fragments and photoinducible dimerization domains.[59] The blue light irradiation brings together the photoinducible dimerization domains, activating the Cas9. In contrast, the function of Cas9 was deactivated without light irradiation. Note that these approaches either uses toxic UV light or the dynamic range of activation/deactivation was not great.

Guide RNA centric approaches, in contrast, focus on modifying the single guide RNA or the trans-acting RNA/crispr RNA pair to control the function of gene editing. A typical strategy uses photocleavable DNA oligonucleotides that are complement to the target regions of the sgRNA to quench the activity of gene editing. Upon on photo-irradiation, the lower melting temperature of the shorter DNA oligos lose their ability to bind to the sgRNA, activating the gene editing (Figure 1.2).[60] More recent study engineered caged gRNAs by substituting four nucleobases evenly distributed throughout the 5'-protospacer region with caged nucleobases during solid-state RNA synthesis.[61] This approach has achieved better dynamic range. However, it still required solid-state synthesis and each sgRNA has to be individually designed for different DNA targets. Again, a toxic UV light has to be used for the activation of sgRNA.
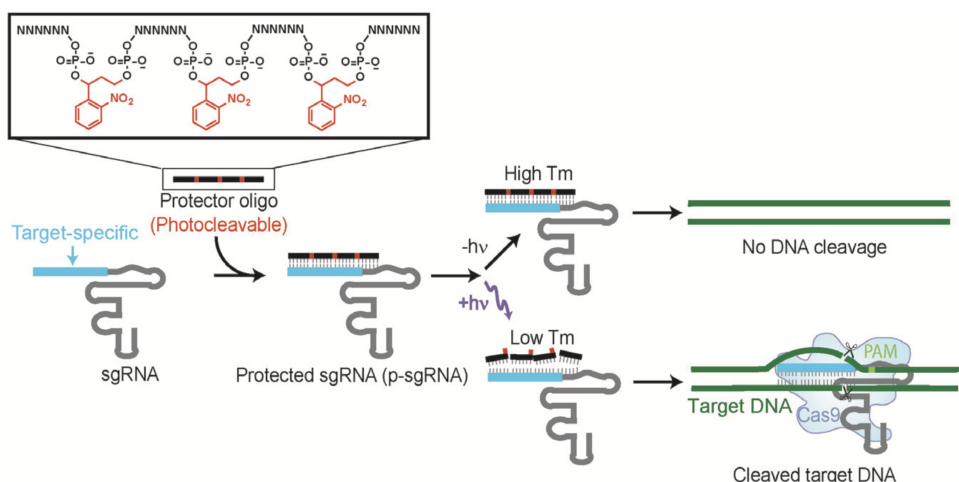
Figure 1.2: Light-Activated CRISPR Using Guide RNAs with Photocleavable Protectors.

1.4   Golden Standards

Having reviewed these many different strategies, we may ask what will be an ideal strategy to achieve photo-regulation of a cellular process? What are the golden standards? Here are my thought on some specifications that tool developers should keep in mind when trying to achieve light control of cellular processes.

[1] Toxic UV light should be avoided, if possible. As we know UV lights are toxic to cells. A longer wavelength of light should be considered for the activation/deactivation processes.

[2] The dynamic range much be high. There should be no background activity when the system is at the caged stage. When activated, the system should have high activity or as good as its wild-type level.

[3] The tool should be easy enough to use. It means that the majority of the field should be able to use such tool without complicated equipment and expensive/hard-to-obtain materials.

[4] The tool should be convenient to apply to different applications. There should be no specific protein or only minimal RNA processing involved to make the tool functional.

[5] The tool should accept different delivery platform. For many applications, especially therapeutics application, delivery is one of the most important process. A great tool should accept a variety of delivery methods, such as nucleic acid transfection, RNA-protein complex delivery, vector delivery, etc.

[6] Multiplexing capability is desirable. It means that different wavelength of light can be simultaneously or sequentially used to control different targets within the same cell.

[7] Reversible capability is desirable. The tool may be able to turn-on and turn-off under different stimuli.

Even though these golden standards may not be all check marked, we should aim to achieve as many of them as possible during the development processes. In the next three chapters, I'm going to share my journey where I tried to develop a good photo-regulation technique to

control different cellular processes, the translation and the CRISPR/Cas9 gene editing.

Chapter 2

Light-Activated Control of Translation by Enzymatic mRNA Labeling

2.1 Introduction

Much work has been done on engineering transcriptional regulators to control gene expression. A complementary approach is to modulate the translation of eukaryotic messenger RNA (mRNA) into protein. Here, we demonstrate the activation of cellular protein expression upon visible light photocleavage of small molecule cages tethered to the 5' untranslated region (5' UTR) of an mRNA. These photo-cleavable cages are conjugated to in vitro transcribed mRNA (IVT-mRNA) through RNA transglycosylation, an enzymatic process in which a bacterial tRNA guanine transglycosylase (TGT) exchanges a guanine nucleobase in a specific 17 nts motif (Tag) with synthetic pre-queuosine1 (preQ1) derivatives. The cages severely reduce mRNA translation efficiency when strategically placed in the 5' UTR. Using this method, we demonstrate the successful spatiotemporal photo-regulation of gene expression with single cell precision. Additionally, our method can be applied to therapeutically relevant chemically modified mRNA (mod-mRNA) transcripts. Photo-regulation of translation through covalent site-specific post-

transcriptional modification of mRNA provides a modular and efficient approach for developing synthetic gene regulatory circuits, biotechnological applications, and therapeutic discovery.
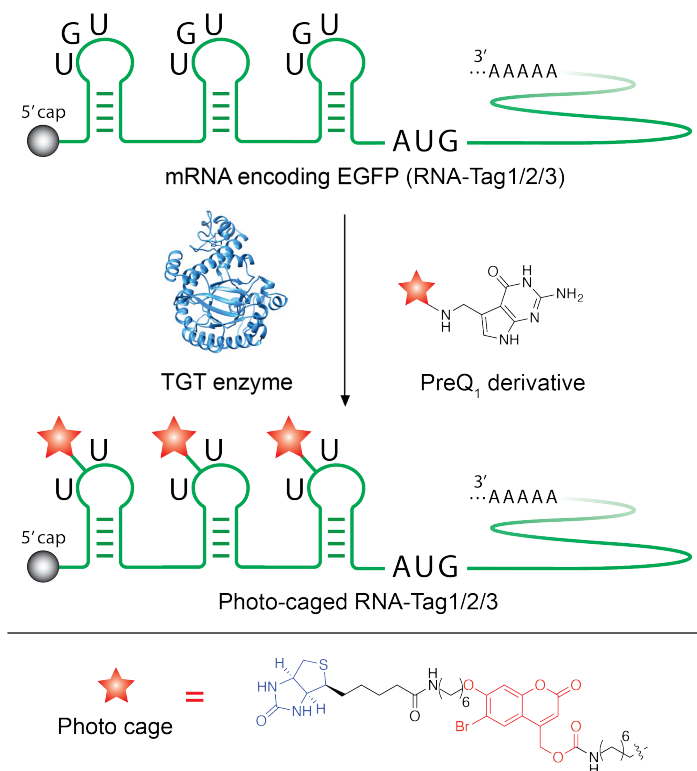
Figure 2.1: Covalent labeling of mRNA using RNA-TAG. RNA-TAG utilizes a bacterial tRNA guanine transglycosylase (TGT) to exchange a guanine nucleobase on a specific 17-nucleotide motif (Tag) for synthetic preQ1 derivatives. In this work, synthetic photocleavable caging groups were covalently incorporated onto the mRNA transcript.

To covalently modify mRNA in a site-specific manner, we applied our previously developed technology, RNA transglycosylation at guanosine (RNA-TAG), to an mRNA of interest.[27] RNA-TAG utilizes a bacterial (E. coli) tRNA guanine transglycosylase (TGT) to recognize a specific genetically encoded 17 nucleotides RNA sequence (Tag) and exchange the original guanine with a synthetic derivative of preQ1, the natural substrate of TGT, onto the mRNA(Figure 1.1). Previously, we demonstrated that a variety of preQ1 derivatives bearing fluorophores or affinity tags could be covalently conjugated to an mRNA in vitro with high yield.[27,30] We hypothesized that the incorporation of bulky preQ1 derivatives (cages) onto the 5'-UTR of an mRNA would

significantly hinder translation initiation, and that subsequent cleavage of the cages would activate translation (Figure 2.1).

Therefore, by designing cages that are cleaved upon light irradiation, photo-regulation of gene expression can be achieved through enzymatic mRNA labeling. As we also discussed in previous section photo-regulation would offer rapid and non-invasive manipulation of target mRNA, and allow for the selective activation of protein expression with high spatiotemporal resolution. Furthermore, the ability to precisely control laser wavelength and illumination time allows the minimization of cell death.[50] A light-activated mRNA translation system would be an extremely useful method to regulate gene expression.

2.2   Results and Discussions

2.2.1   Choice of The Photo-sensitive Cages

An ideal photo-cleavable cage would have high photolysis efficiency and be cleaved using visible light (> 400 nm) to limit cytotoxicity. Therefore, we chose 6-bromo-7-aminoethoxycoumarin-4-ylmethoxycarbonyl (Bac) as the photo-cleavable linker, which shows high cleavage efficiency at a wavelength of 405 nm in vivo.[62, 63] We synthesized the photo-cleavable preQ1 substrates (compound 7) using a strategy similar to previous reports (Figure 2.2).[27]
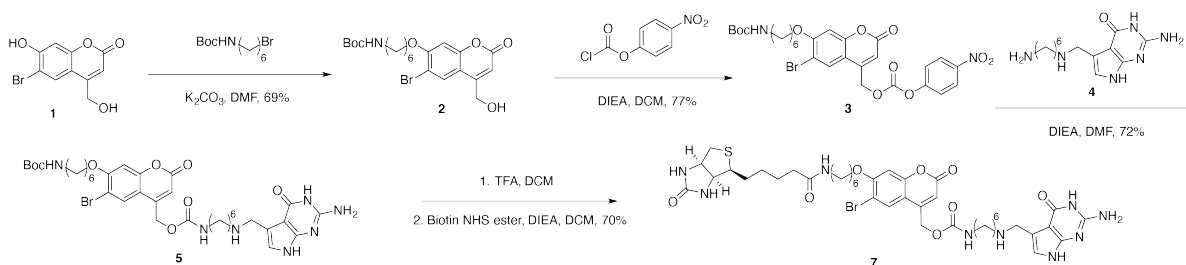
Figure 2.2: Route for synthesis of the TGT substrate biotin-Bac-preQ1 (7).   Boc = tert-butyloxycarbonyl protecting group, DMF = dimethylformamide, DIEA = N,N-Diisopropylethylamine, DCM = dichloromethane, TFA = trifluoroacetic acid.

Briefly, compound 4 was obtained through the alkylation of the exocyclic amine of preQ1. The generated preQ1 derivative was then coupled with the Bac linker building block (compound 3) to yield a Boc protected preQ1-Bac conjugate (compound 5).[64] To append an affinity biotin handle for purification of labeled mRNAs, the preQ1-Bac conjugate (compound 5) was deprotected and further coupled with biotin-NHS ester to obtain the final TGT substrate compound 7, which can be photo-cleaved by 405 nm laser light (Figure 2.3).[62,63]
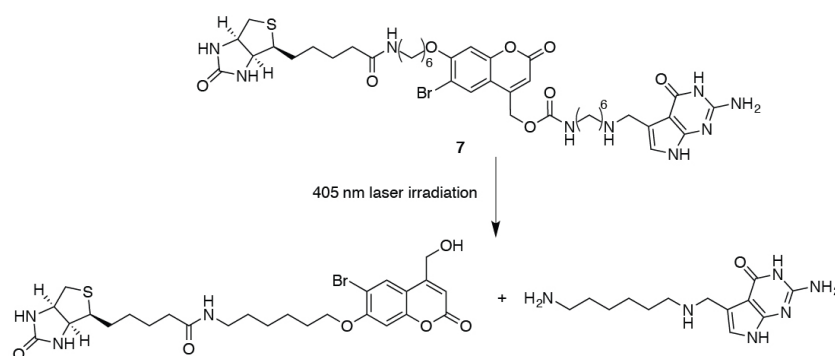
Figure 2.3: Photo-cleavage of compound 7 upon 405 nm laser irradiation.

2.2.2   mRNA Enzymatic Labeling and Activity Test

To test whether covalent labeling of mRNA with the bulky biotin-Bac groups affected the efficiency of translation, three Tag sequences were inserted at various locations within the 5' UTR of an IVT-mRNA encoding EGFP (Figure 2.1, RNA-Tag1/2/3). Tag-1 is located immediately adjacent to the 5' cap, beginning at the third base of the IVT-mRNA. Tag-2 is located in the middle of the 5' UTR. Tag-3 is located immediately before the start codon of the EGFP open reading frame. Following in vitro transcription, the EGFP mRNA transcript containing all three Tags (RNA-Tag1/2/3) was capped and polyadenylated to enable expression in mammalian cells. The RNA-Tag1/2/3 was then labeled with photo-cleavable biotin-Bac-preQ1 (compound 7) following standard TGT labeling conditions. TGT labeled RNA-Tag1/2/3 was isolated through streptavidin affinity purification. The caged RNA-Tag1/2/3 was then transfected into cultured HeLa cells.

16

Transfected cells were incubated overnight, and imaged by fluorescence confocal microscopy. A significant decrease in EGFP expression (90% decrease) was observed in cells transfected with photo-caged RNA-Tag1/2/3 compared to cells transfected with unlabeled RNA-Tag1/2/3 (Figure 2.4), suggesting that the introduction of the bulky biotin-Bac groups perturbs mRNA translation initiation.
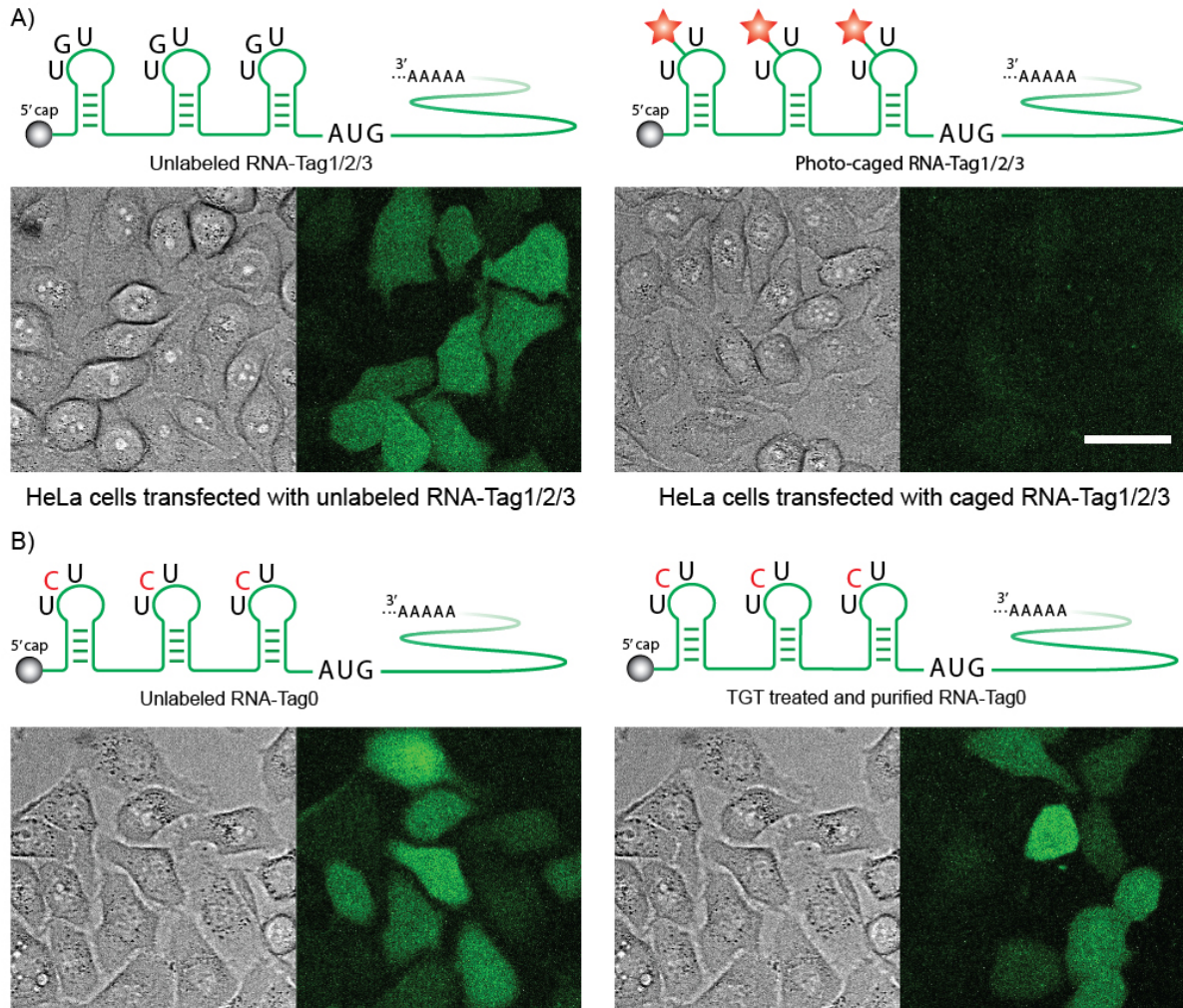
Figure 2.4: mRNA caging through RNA-TAG. TGT labeling is performed to incorporate photo-cleavable cages at the 5' UTR in order to hinder the translation activity of RNA-Tag1/2/3 or RNA-Tag0. Labeled or unlabeled RNA-Tag1/2/3 and RNA-Tag0 are transfected into HeLa cells. Cell images were taken after overnight incubation at 37 °C, 5% CO2. Scale bar = 40 μm.

2.2.3 Screening of mRNA Constructs

To investigate whether the site of covalent modification within the 5' UTR affects translation efficiency, mRNA transcripts with mutations at one or more Tag sites were assayed for activity. To minimize changes among these transcripts, a guanine to cytosine (G-to-C) mutation was introduced to the key U-G-U site of the Tag, rendering it unreactive for TGT-mediated labeling while maintaining the secondary structure of the RNA (Figure 2.5).[27] In addition to the RNA-Tag1/2/3 with three fully functional Tags, a single G-to-C mutation was introduced to the first, second or third Tag sequence of RNA-Tag1/2/3 to obtain RNA-Tag2/3, RNA-Tag1/3 and RNA-Tag1/2 transcripts, respectively. RNA-Tag3, RNA-Tag1 and RNA-Tag2 were obtained by introducing mutations at two of the three RNA-Tag1/2/3 Tags. RNA-Tag0 contains the G-to-C mutation at all three Tags and thus cannot be labeled with any photo-cleavable cage. All synthetic mRNA transcripts were treated with TGT enzyme and compound 7. The enzymatically labeled and purified mature mRNAs were transfected into cultured HeLa cells. EGFP expression of these cells was quantified using fluorescence microscopy (Figure 2.6).

A substantial decrease in EGFP expression was observed for the transcripts with only one functional Tag (RNA-Tag3, RNA-Tag1 and RNA-Tag2) compared to the unlabeled mRNA (RNA-Tag0). Notably, RNA-Tag3 demonstrated severely impaired translation activity, resulting in an EGFP expression level of 18.7% ($\pm$ 0.97% SEM) compared to the RNA-Tag0 (Figure 2.6, RNA-Tag3), indicating that the photocage closest to the start codon (Tag-3) has the strongest effect on reducing translation. Lower expression level (15% - 20%) was observed with transcripts containing two cages (RNA-Tag2/3, RNA-Tag1/3 and RNA-Tag1/2). Minimal EGFP expression was observed in cells transfected with labeled RNA-Tag1/2/3, with only 9.2% ($\pm$ 0.64% SEM) fluorescence intensity relative to the unlabeled RNA-Tag0. EGFP expression of RNA-Tag0 and labeled RNA-Tag1/2/3 was also examined with immunoblotting (Figure 2.7). Together, these data show a clear trend indicating that labeling mRNA with more photocages leads to lower translation activity, with modification at Tag-3 showing the greatest impact on translation.
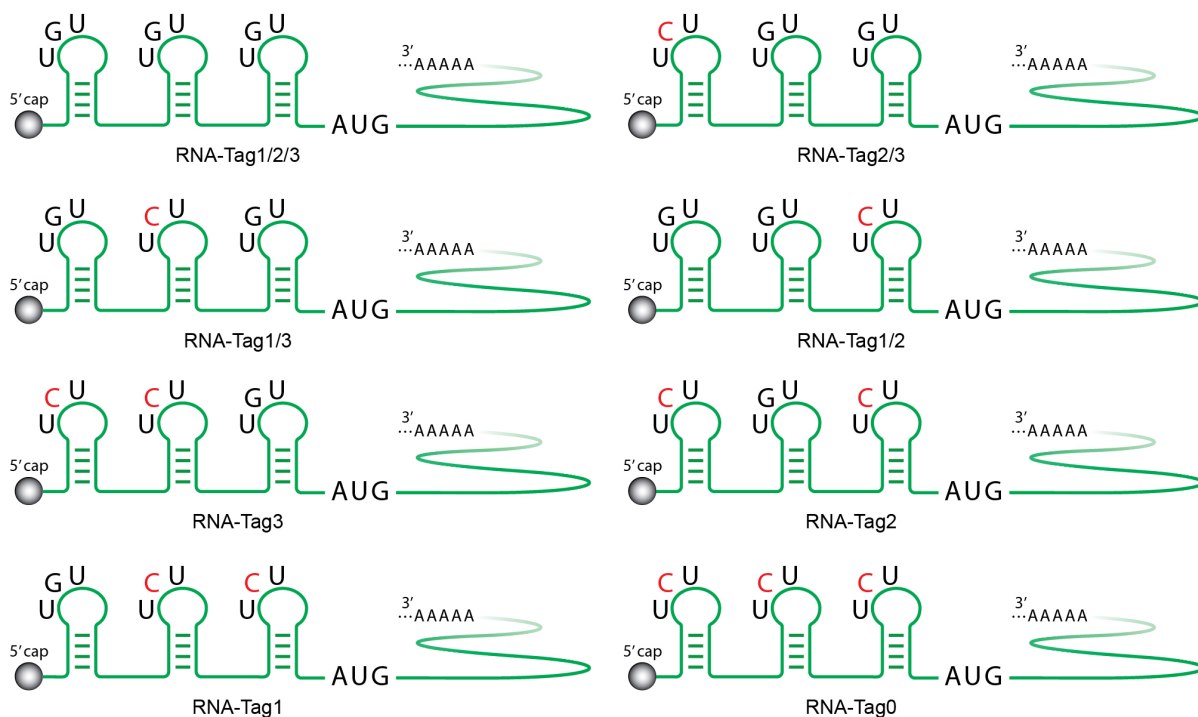
Figure 2.5: RNA constructs with different number of active TGT enzymatic labeling sites

2.2.4 Photo-activation of Translation

Having demonstrated that covalent labeling of three Tags on an mRNA transcript significantly hindered its translation efficiency, we next sought to explore whether protein expression could be restored by photo-uncaging. First, we demonstrated that upon 5000 μJ of in vitro UV irradiation, more than 90% of the labeled photo cages can be cleaved from the 17 nts RNA hairpin (Tag) (Figure 2.8, Figure 2.9). To demonstrate live cell photo uncaging of TGT labeled IVT-mRNA, cultured HeLa cells were transfected with labeled RNA-Tag1/2/3 or unlabeled RNA-Tag0. In order to activate protein expression while limit cytotoxicity, the labeled RNA-Tag1/2/3 transfected cells were irradiated with a 405 nm laser for 10 seconds to trigger photo-uncaging. To allow time for EGFP expression and maturation, cells were imaged 8 hours after the uncaging event (Figure 2.10). To our delight, recovered EGFP expression was observed in previously dark cells transfected with TGT labeled RNA-Tag1/2/3, suggesting a successful photo activation of gene
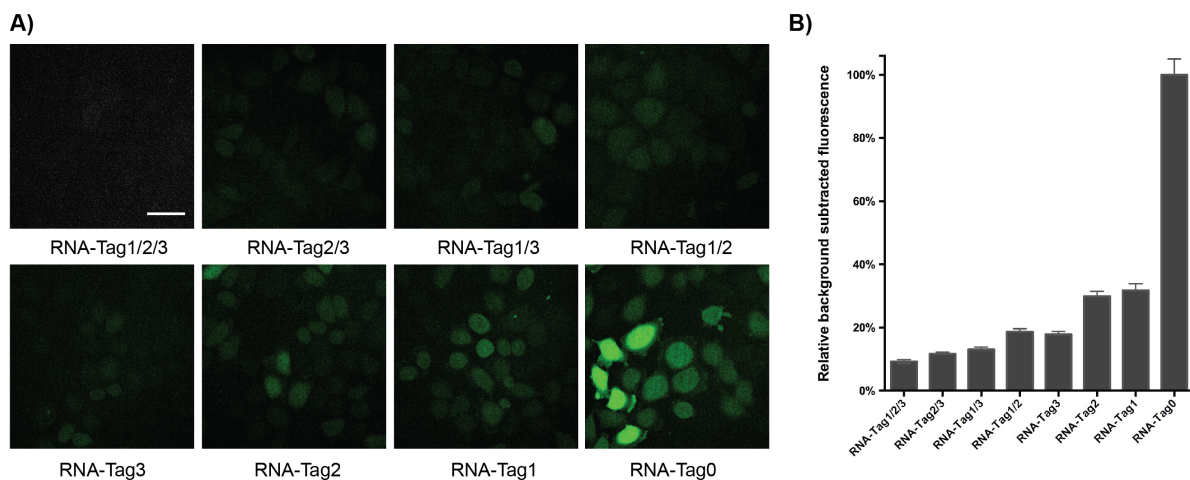
Figure 2.6: mRNA translation activity with different numbers of photocleavable caging groups incorporated. A) Cell images showing cells transfected with different mRNA constructs. Scale bar: 50 μm. B) Background-subtracted average cell fluorescence intensity, which indicates the protein expression level. Average fluorescence intensity values were measured from more than 80 cells. Error bars show the SEM.

expression (Figure 2.10). Moreover, the expressed EGFP was only found within the laser illuminated cells, but not in adjacent cells, demonstrating spatiotemporal regulation of gene expression with single cell precision. To examine whether the uncaged preQ1 nucleobase has effect on protein expression, HeLa cells were transfected with in vitro UV uncaged RNA-Tag1/2/3, which is expected to carry amine modifications at three Tags, followed by immunoblotting to evaluate EGFP expression. Similar EGFP expression levels were observed between cells transfected with TGT labeled, but in vitro UV uncaged, RNA-Tag1/2/3 and unlabeled RNA-Tag1/2/3, suggesting that the amine modifications at the 5'-UTR of an mRNA have little effect on translation.

Chemically modified mRNA (mod-mRNA), which contains modified nucleobases such as pseudouracil or 5-methylcytosine (5mC), has been shown to have increased serum stability and elicits a lower immune response compared to natural mRNA,[65, 66] making it a useful reagent for RNA gene therapy.[67, 68] Therefore, in addition to regulating protein expression through photo-uncaging of enzymatically labeled mRNA transcripts, the ability of TGT to label mod-mRNA would expand the scope of this technology and potentially enable spatiotemporal activation of

RNA-TAG0	+	−	−	−
RNA-TAG1/2/3	−	+	+	+
TGT labeling	−	−	+	+
RNA UV treated	−	−	−	+
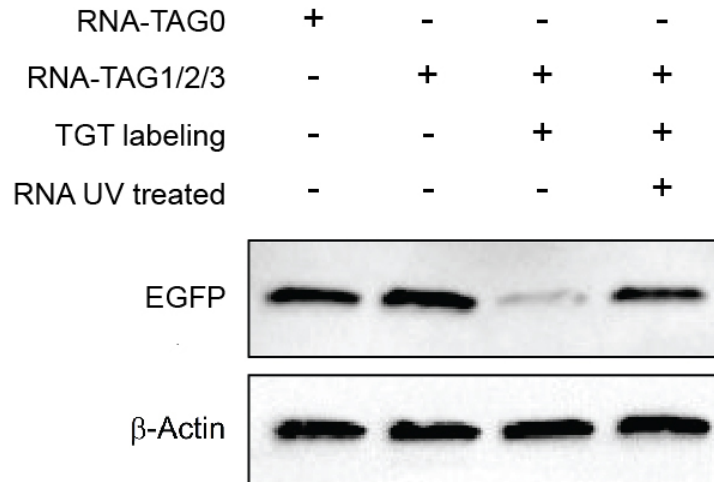
| EGFP | | | | |

| β-Actin | | | | |

Figure 2.7: Immunoblot analysis of EGFP expression. Cells were trans- fected with either unla-beled RNA-Tag0, unlabeled RNA-Tag1/2/3, labeled RNA-Tag1/2/3, or labeled RNA-Tag1/2/3 that was uncaged through UV irradiation prior to transfection.

therapeutically relevant RNAs.[30] Taking advantage of the versatility of the RNA-TAG technology, we applied our caging/uncaging strategy to a 5-methylcytidine (5mC) mod-mRNA transcript. IVT mod-mRNA encoding EGFP (mod-RNA-Tag1/2/3 and mod-RNA-Tag0) could be conveniently synthesized following a previously established protocol.[69] The IVT product was subsequently polyadenylated and labeled by TGT with compound 7. HeLa cells were transiently transfected with the photo-caged mod-mRNA and illuminated with a 405 nm laser as previously described. We observed successful activation of the photo-caged mod-mRNA and the switching on of protein expression (Figure 2.11). However, the observed background EGFP expression in cells transfected with caged mod-mRNA was somewhat higher than that observed with caged mRNA. This might be due to the decreased labeling efficiency of mod-mRNA by TGT, which has been previously described.[30] We believe that further development of our photo-uncaging technique for labeled mod-mRNA will allow precise control of the expression of mod-mRNA and may facilitate applications of mod-mRNA as an important therapeutic RNA in degenerative diseases, as well as in the treatment of genetic disorders.
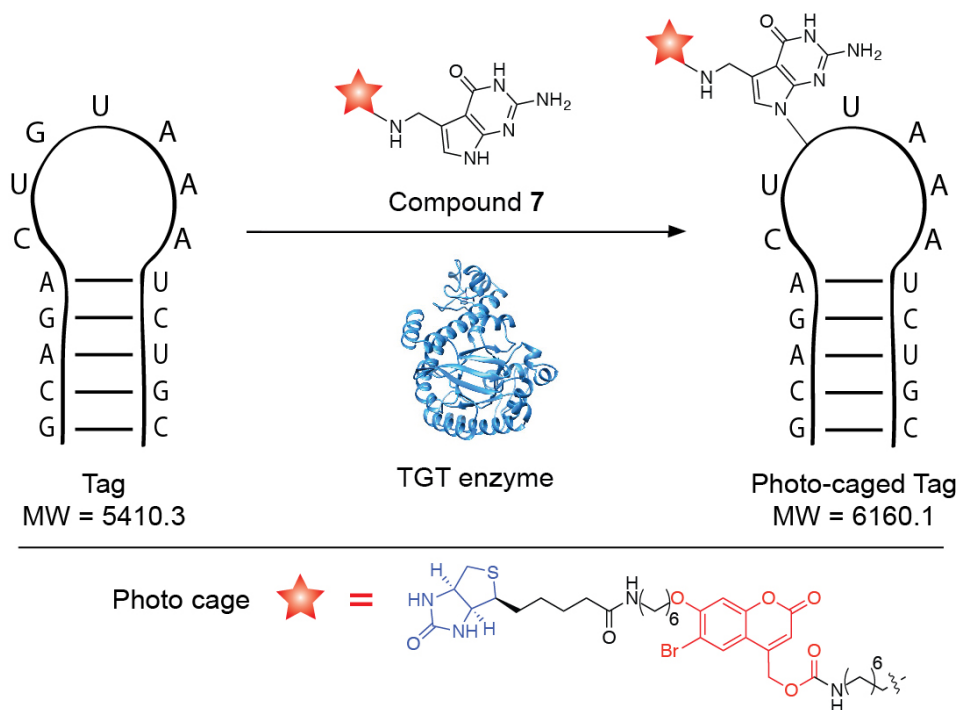
Figure 2.8: Schematic representation of TGT labeling of Tag RNA with Compound 7.

2.3 Conclusions

In conclusion, we have demonstrated a strategy to achieve light-activated control of mRNA translation in mammalian cells by the removal of photo-cleavable cages within the 5' UTR of an mRNA transcript. Using our RNA-TAG technology, we are able to incorporate photo-cleavable cages at site-specific positions within the 5' UTR of an mRNA transcript, significantly diminishing protein expression compared to the unlabeled mRNA. Translation was activated using a 405 nm laser with a single cell precision. Through careful design of orthogonal photo-cleavable probes, our approach should be amenable to the activation of multiple mRNAs using different wavelengths of light. We believe that the ability to spatiotemporally regulate gene expression through enzymatic, site-specific labeling of RNA will serve as a powerful and versatile tool with applications spanning from basic research in developmental biology and genetics to the development of novel mRNA-based therapeutics.
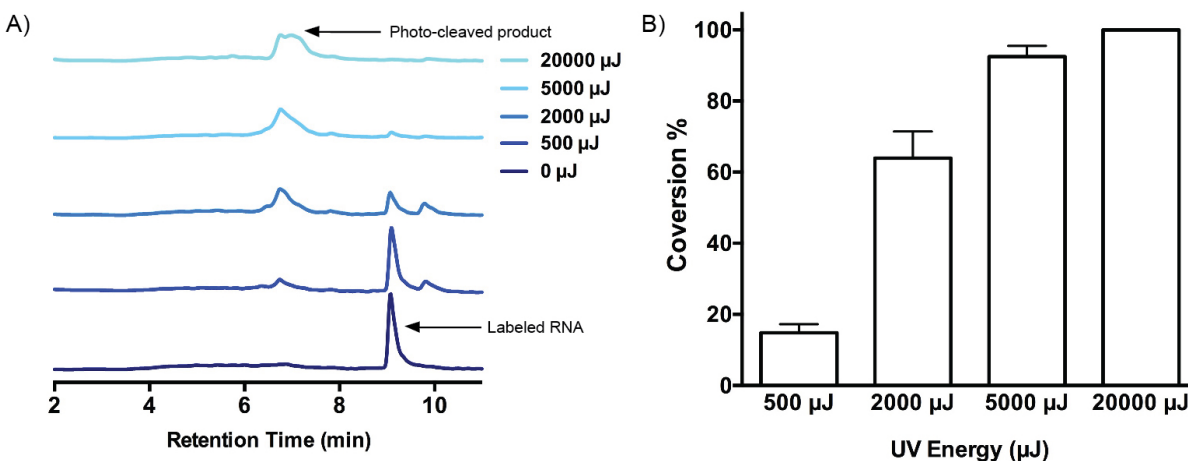
Figure 2.9: HPLC analysis of photo-cleavage of the caged 17-nt RNA TAG. A) HPLC traces (260 nm absorbance) represent that the labeled RNA undergoes photo-cleavage under different dosage of UV. B) The bar graph demonstrates the percentage conversion of the labeled Tag to photo-cleaved products. All measurements were measured in triplicates. Error bars denote the standard deviation of three measurements.

2.4 Methods and Materials

2.4.1 General Materials

Reagents and instruments

Chemical reagents used for synthesis of TGT substrate biotin-Bac-preQ1 were purchased from Sigma-Aldrich (St. Louis, MO) and Alfa Aesar (Haverhill, MA) without further purification. DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA) and Eton Bioscience (California, CA). Molecular biology reagents such as restriction digestion enzymes, Q5 DNA polymerase, T7 RNA polymerase, Vaccinia Capping System, *E. coli* Poly(A) Polymerase, nucleotide stains and competent bacterial were purchased from New England Biolabs (Ipswitch, MA), Promega (Madison, WI), or Life Technologies (Carlsbad, CA). DynabeadsTM M-280 Streptavidin was purchased from Thermo Fisher Scientific (Waltham, MA). H and C NMR spectra were recorded on a Varian VX 500 MHz NMR Spectrometer. High resolution mass spectroscopy was collected on an Agilent Infinity 1260 LC and tandem Agilent 6230
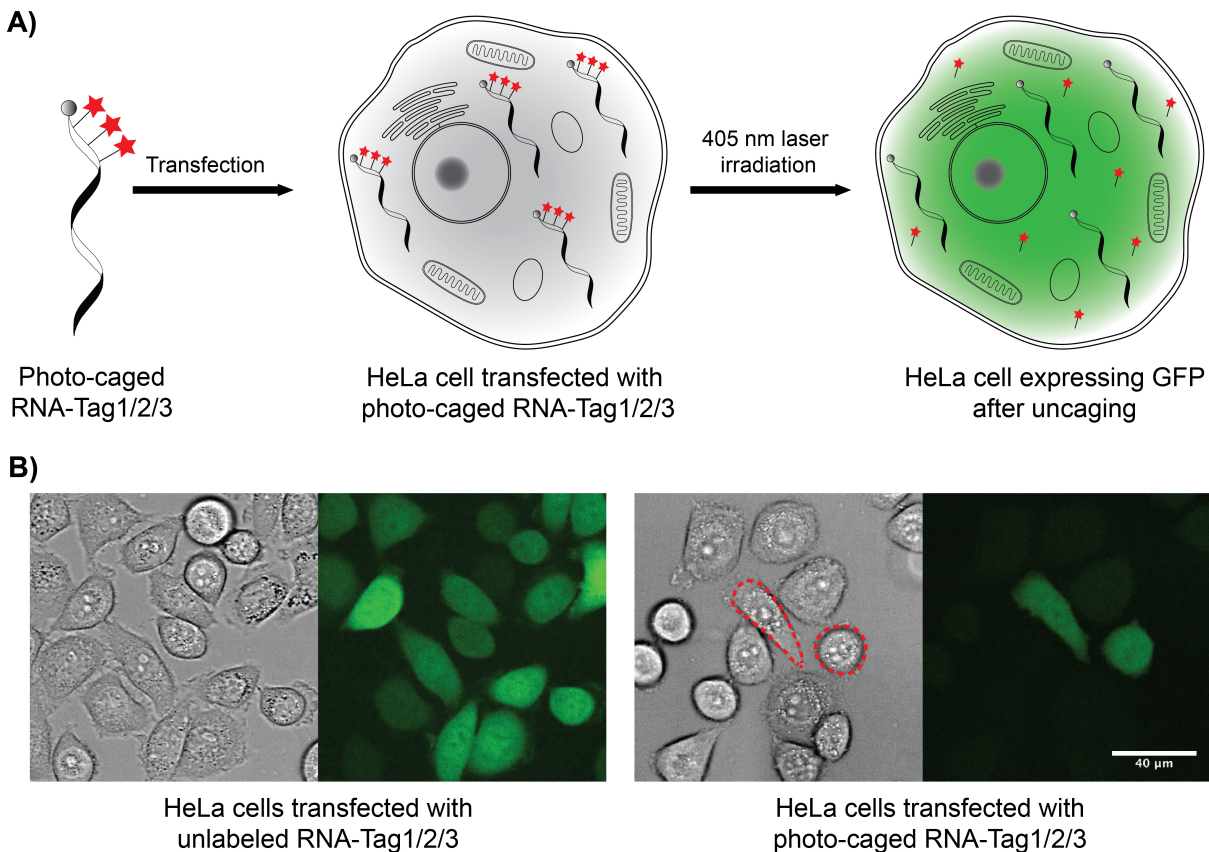
Figure 2.10: Live-cell uncaging of photocaged RNA-Tag1/2/3. A) Experi- mental procedure for live-cell uncaging. B) Live-cell uncaging using 405 nm laser light. Red circles indicate cells transfected with photo- caged RNA-Tag1/2/3 and irradiated with 405 nm laser light for 10 seconds. Scale bar: 40 μm.

high resolution time of flight (TOF) mass spectrometer managed by the UCSD Department of Chemistry and Biochemistry Molecular Mass Spectroscopy Facility. Reverse-phase HPLC purification and analysis were performed using an Agilent 1260 Infinity HPLC with Agilent 6120 Quadrapole mass spectrometer (Santa Clara, CA). preQ1 derivatives were prepared and analyzed with an Agilent Zorbax SB-C18 semi-prep column (ID 9.4 x 250 mm, 5 μm, 80 Å) and an Agilent Zorbax eclipse plus C8 column, using a water/methanol gradient containing 0.1% fomic acid. Fluorescence microscopy imaging was performed on an Axio Observer Z1 motorized inverted microscope (Carl Zeiss Microscopy Gmb, Germany) with Yokogawa CSU-X1 spinning disk confocal unit using a 20x, 1.42 NA objective to an Evolve 512x512 EMCCD camera
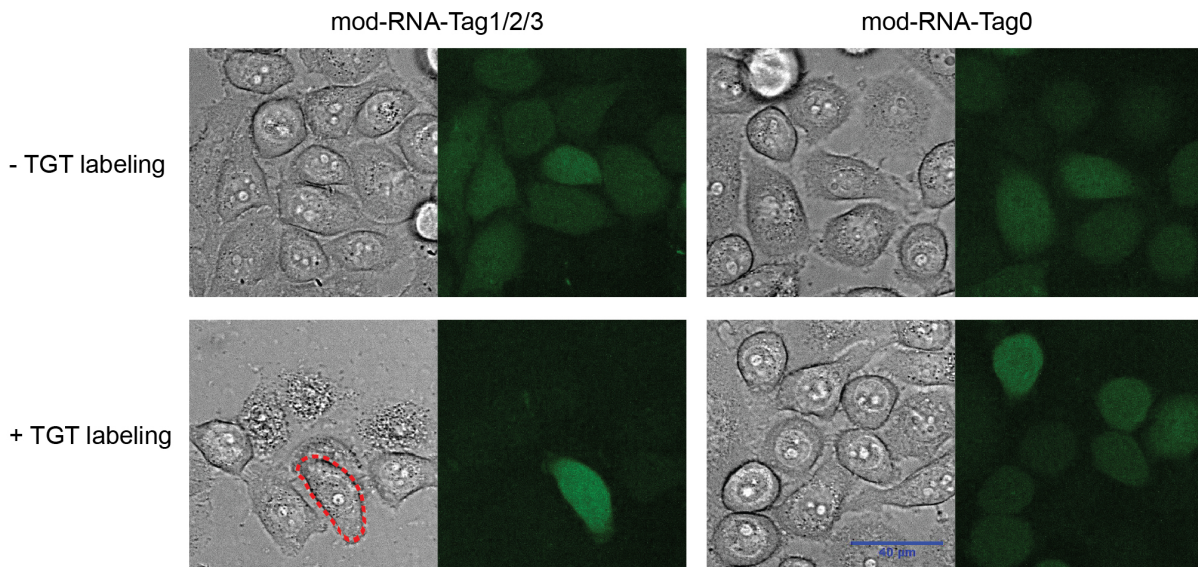
Figure 2.11: Mod-mRNA uncaging using 405 nm laser light. Red circles represent the selected HeLa cells irradiated with 405 nm laser light. HeLa cells were transfected with photo-caged mod-mRNA and illuminated with a 405 nm laser line at 100% laser power for 10 seconds. Cells were incubated at 37 °C in 5% CO2 for 8 hours and imaged using fluorescence microscopy. Successful uncaging was demonstrated by the recovery of protein expression only in selected cell treated with laser irradiation (Figure S9). Scale bar = 40 μm.

(Photometrics, Canada) using ZEN imaging software (Carl Zeiss Microscopy GmbH, Germany). Fluorophores were excited with laser diodes (405 nm; 20 mW, 488 nm; 30 mW). Photo-uncaging was performed using a DirectFRAP module built around the fluorescence microscope and the 405 nm laser (Carl Zeiss). Images were processed using Image J.

Reaction Buffers

TGT Storage Buffer: 25 mM HEPES, pH 7.3, 2 mM DTT, 1 mM EDTA, and 100 μM PMSF.

TGT Reaction Buffer: 100 mM HEPES, pH 7.3, and 20 mM $MgCl_2$.

T7 Reaction Buffer: 40 mM Tris pH 7.5, 5 mM DTT, 25 mM $MgCl_2$, 2 mM spermidine.

DynabeadsTM streptavidin binding and washing (B&W) buffer 2X: 10 mM Tris-HCl (pH 7.5), 1mM EDTA, 2 M NaCl.

DynabeadsTM streptavidin binding and washing buffer with Tween (BW&T) 2X: 10 mM

Tris-HCl (pH 7.5), 1mM EDTA, 2 M NaCl, 0.1% Tween 20.

DynabeadsTM streptavidin solution A: DEPC-treated 0.1 M NaOH, DEPC-treated 0.05 M NaCl. DynabeadsTM streptavidin solution B: DEPC-treated 0.1 M NaCl.

DynabeadsTM streptavidin elution buffer: 950 μL formamide, 20 μL 500 mM EDTA, 30 μL RNase free water.

2.4.2   Chemical Synthesis

Synthesis of compound 2
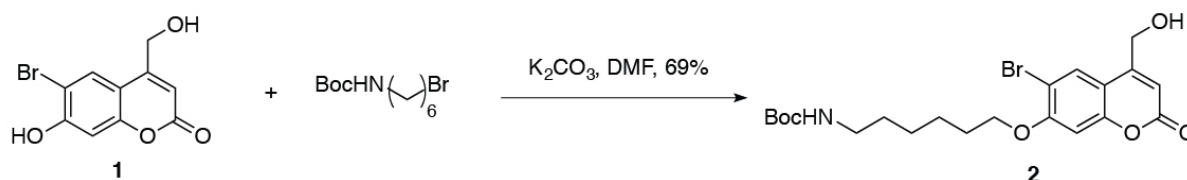
Figure 2.12: Synthesis of compound 2.

Compound 1 (1.00 g, 3.69 mmol) was suspended in 15 mL anhydrous DMF.[70] Potassium carbonate (1.02 g, 7.38 mmol) and 6-(Boc-amino)hexyl bromide (1.34 g, 4.80 mmol) were added to the mixture. The reaction mixture was stirred at room temperature for 40 hours. The reaction solution was concentrated, extracted with DCM, and washed with water for three times. The organic layer was dried and concentrated. Crude product was purified by column chromatography (hexane:ethyl acetate = 4:1) to yield 1.20 g (69%) of compound 2 as a white residue (Figure 2.12).

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.66 (s, 1H), 6.73 (s, 1H), 6.48 (s, 1H), 4.82 (d, J = 3.2 Hz, 2H), 4.60 (s, 1H), 4.03 (t, J = 6.2 Hz, 2H), 3.12 13 (m, 3H), 1.86 (p, J = 6.5 Hz, 2H), 1.56 – 1.49 (m, 4H), 1.44 – 1.42 (m, 11H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 161.36, 157.96, 156.21, 154.36, 153.96, 127.48, 111.73, 109.65, 108.15, 100.99, 79.35, 69.68, 60.74, 40.59, 30.08, 28.73, 28.54, 28.54, 28.54, 26.58, 25.78. HRMS [M]$^+$ m/z calcd. for [C$_{21}$H$_{28}$BrNO$_6$]$^+$ 469.1100, found 469.1098.

Synthesis of compound 3
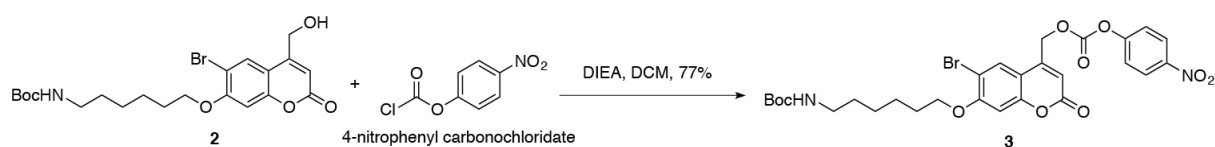
Figure 2.13: Synthesis of compound 3.

Compound 2 (500 mg, 1.06 mmol) was dissolved in 10 mL anhydrous DCM and the mixture was cooled to 0 °C. Diisopropylethylamine anhydrous (1.86 mL, 10.63 mmol) was added to the solution, followed by a slow addition of 4-nitrophenyl carbonochloridate (2.14 g, 10.63 mmol). The reaction mixture was stirred at room temperature for 30 minutes, and then warmed to room temperature and stirred for another 1 hour. Upon the completion of the reaction, the mixture was concentrated and purified by column chromatography (hexane:ethyl acetate = 4:1) to yield 520 mg (77%) of compound 3 as a white solid (Figure 2.13).

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.32 (d, J = 9.2 Hz, 2H), 7.68 (s, 1H), 7.43 (d, J = 9.1 Hz, 2H), 6.85 (s, 1H), 6.47 (s, 1H), 5.42 (s, 2H), 4.54 (s, 1H), 4.09 (t, J = 5.9 Hz, 2H), 3.15 – 3.12 (m, 2H), 1.89 (p, J = 6.3 Hz, 2H), 1.52 – 1.58 (m, 4H), 1.44 – 1.42 (m, 11H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 160.14, 158.55, 156.19, 155.17, 154.70, 152.09, 146.81, 145.78, 127.29, 125.68, 125.68, 121.73, 121.73, 111.33, 110.93, 108.63, 101.39, 79.37, 69.83, 65.38, 40.58, 30.11, 28.63, 28.47, 28.47, 28.47, 26.56, 25.78. HRMS [M$^+$Na]$^+$ m/z calcd. for [C$_{28}$H$_{31}$BrN$_2$O$_{10}$Na]$^+$ 657.1054, found 657.1052.

Synthesis of compound 5

Compound 4, prepared as previously described,[30] (20 mg, 72 μmol) was dissolved in 4 mL DMF in a 10 mL glass vial, followed by the addition of diisopropylethylamine anhydrous (38 μL, 216 μmole). A solution of compound 3 (32 mg, 50 μmol) in DMF was then slowly added to the mixture. The reaction proceeded at room temperature for 1 hour and was monitored by LCMS.

Figure 2.14: Synthesis of compound 5.

The reaction was concentrated and the residue was purified by HPLC. 40 mg of compound 5 was collected as a white solid in a 72% yield. HPLC gradient: 0 min – 2.5 min 50% Phase B in Phase A, 2.5 min – 18 min 50% Phase B in Phase A to 90% Phase B in Phase A, 18 min – 28 min 95% Phase B in Phase A, 28 min – 32 min 50% Phase B in Phase A (Phase A: H_2O with 0.1% formic acid, Phase B: MeOH with 0.1% formic acid).

^{1}H NMR (500 MHz, CD$_3$OD) δ 7.87 (s, 1H), 7.01 (s, 1H), 6.83 (s, 1H), 6.28 (s, 1H), 5.28 (s, 2H), 4.24 (s, 2H), 4.13 (s, 2H), 3.16 (s, 2H), 3.08 – 3.04 (m, 4H), 1.87 (s, 2H), 1.74 (s, 2H), 1.57 – 1.43 (m, 21H). ^{13}C NMR (126 MHz, CD$_3$OD) δ 162.50, 159.72, 158.57, 157.80, 155.82, 154.51, 153.93, 152.60, 129.07, 118.98, 112.64, 110.58, 110.33, 109.86, 109.23, 102.03, 99.67, 79.80, 70.87, 62.53, 47.59, 44.80, 41.59, 41.27, 30.91, 30.48, 29.82, 28.88, 28.70, 28.70, 28.70, 27.51, 27.24, 27.10, 26.80. HRMS [M]$^+$ m/z calcd. for [C$_{35}$H$_{49}$BrN$_7$O$_8$]$^+$ 774.2821, found 774.2818.

Synthesis of compound 7
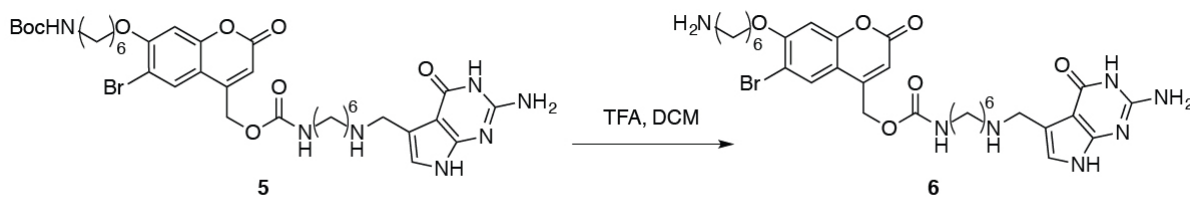
Figure 2.15: Deprotection of compound 5.

Compound 5 (5.1 mg, 6.45 μmole) was suspended in 1 mL DCM in a 4 mL glass vial. 50 μL of TFA was slowly added to the stirring mixture at room temperature for 30 minutes. The
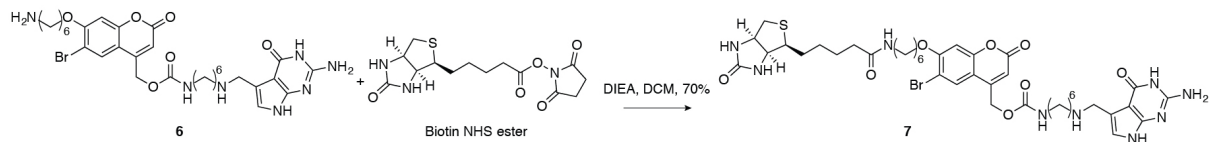
Figure 2.16: Synthesis of compound 7.

solvent was removed in vacuo and compound 6 was obtained. The residue was used for the next step without further purification.

Compound 6 was dissolved in 0.6 mL of DCM. DIEA anhydrous (3.88 μL, 22 μmole) was added to the mixture. Biotin NHS ester (Broadpharm, San Diego, CA) (2.02 mg, 5.93 μmole) was added to the stirring mixture and the reaction proceeded at room temperature for 30 minutes. The reaction mixture was concentrated and purified by HPLC. 4.7 mg of compound 7 was collected as a white solid in a 70% yield. HPLC gradient was as reported for the purification of compound 5.

^{1}H NMR (500 MHz, CD$_3$OD) δ 7.90 (s, 1H), 7.05 (s, 1H), 6.84 (s, 1H), 6.29 (s, 1H), 5.29 (s, 2H), 4.59 (s, 1H), 4.48 (dd, J = 7.9, 4.9 Hz, 1H), 4.29 (dd, J = 7.9, 4.4 Hz, 1H), 4.24 (s, 2H), 4.16 (t, J = 6.2 Hz, 2H), 3.23 – 3.13 (m, 4H), 3.06 (t, J = 7.5 Hz, 2H), 2.91 (d, J = 12.8, 4.9 Hz, 1H), 2.69 (d, J = 12.8 Hz, 1H), 2.19 (t, J = 7.3 Hz, 2H), 1.88 (p, J = 6.6 Hz, 2H), 1.79 – 1.36 (m, 18H), 1.29 (s, 2H). ^{13}C NMR (126 MHz, CD$_3$OD) δ 174.58, 161.24, 161.13, 158.36, 156.43, 154.48, 153.07, 152.57, 151.25, 127.81, 127.67, 117.88, 111.32, 109.30, 109.02, 108.45, 107.88, 101.17, 69.48, 62.09, 61.86, 61.16, 60.10, 55.45, 46.23, 43.44, 40.17, 38.84, 35.42, 29.08, 28.92, 28.36, 28.11, 26.16, 26.01 25.83, 25.66, 25.58, 25.55, 25.36. HRMS [M]$^+$ m/z calcd. for [C$_{40}$H$_{55}$BrN$_9$O$_8$S]$^+$ 900.3072, found 900.3075.

Synthesis of compound 8

preQ1-C6-NH$_2$ compound 4 (2.26 mg, 8.1 μmole) was dissolved in 200 μL DMF in a 4 mL glass vial, followed by the addition of DIEA anhydrous (4.1 μL, 24.3 μmole). A solution of Cy5 NHS ester (Broadpharm, San Diego, CA) in 200 μL DMF was then slowly added to the mixture. The reaction was allowed at room temperature for 1 h and monitored by LCMS. The
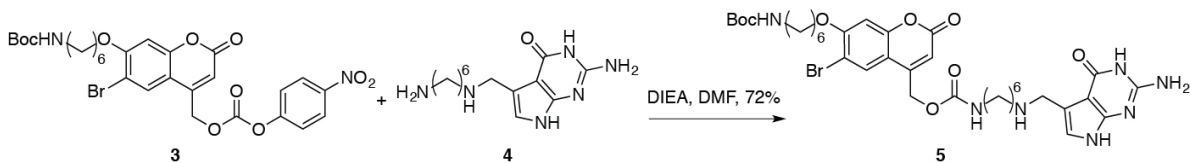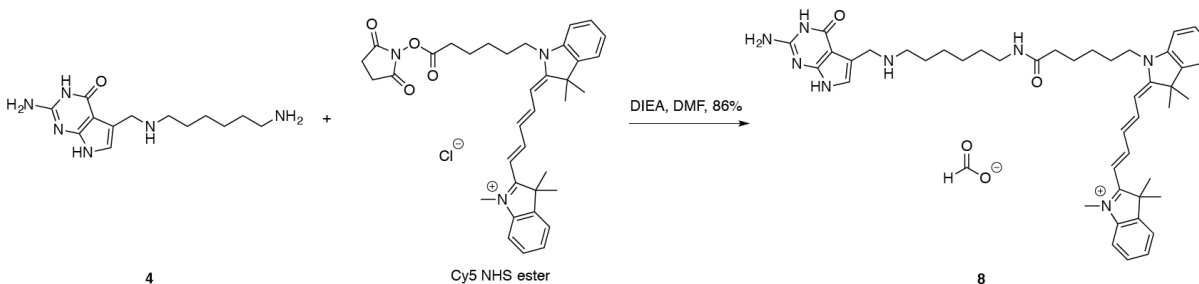
Figure 2.17: Synthesis of compound 8.

reaction was concentrated and the residue was purified by HPLC. 5.5 mg of compound 8 was collected as a blue solid in an 86% yield. HPLC gradient was as reported for the purification of compound 5.

^{1}H NMR (500 MHz, CD$_3$OD) δ 8.25 (t, J = 12.9 Hz, 2H), 7.49 (d, J = 7.4 Hz, 2H), 7.41 (q, J = 7.6 Hz, 2H), 7.33 – 7.23 (m, 4H), 6.83 (s, 1H), 6.62 (t, J = 12.3 Hz, 1H), 6.27 (dd, J = 13.6, 4.4 Hz, 2H), 4.24 (s, 2H), 4.10 (t, J = 7.3 Hz, 2H), 3.62 (s, 3H), 3.13 (t, J = 7.1 Hz, 2H), 3.04 (t, J = 7.3 Hz, 2H), 2.20 (t, J = 7.3 Hz, 2H), 1.87 – 1.78 (m, 2H), 1.72 – 1.67 (m, 14H), 1.52 – 1.40 (m, 6H), 1.37 – 1.32 (m, 2H), 1.30 (d, J = 9.6 Hz, 2H). HRMS [M]$^+$ m/z calcd. for [C$_{45}$H$_{59}$N$_8$O$_2$]$^+$ 743.4755, found 743.4758.

2.4.3   In vitro mRNA Transcription

Transcription Templates

The EGFP expression vector pcDNA3-EGFP was purchased from Addgene (plasmid #13031) and as a gift from Doug Golenbock (delivered throught Add Gene). Ultramer DNA oligo containing three active tags was purchased from Integrated DNA Technologies (San Diego, CA). EGFP-multiTag vector was obtained by inserting the ultramer DNA oligo into 5'-UTR of the EGFP coding region through Gibson DNA assembly technique (NEB, Ipswitch, MA). DH5a competent cells (Life Technologies, Carlsbad, CA) were transformed with the ligation product and screened against ampicillin on agar plates overnight. Colonies were selected and overgrown

for 16 hours. The overgrowth was subjected to DNA extraction with a QIAGEN Plasmid Maxi Kit (QIAGEN, Venlo, Limburg Netherlands). Sequencing was performed to verify the inserted sequence.

EGFP-multiTag sequence from T7 promoter to 3' XbaI restriction enzyme cut site (three 17 nts Tag sequences were underlined):

```
TAATACGACTCACTATAGGGGCAGACTGTAAATCTGCAGACCCAAGCTTGGTAGGTCAGTTGCAGTTAC
CGAGCTCGGATCCACTAGTAACGGCCGCGCAGACTGTAAATCTGCCAGTGTGCTAGTCAGACAGATGGA
ATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCAGACTGTAAATCTGCGATGGTGAGCAAGGGC
GAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTC
AGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACC
GGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGC
TACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGC
ACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG
GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAG
TACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTC
AAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCC
AACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAATCTAGA
```

In vitro Transcription

EGFP-multiTag vector was linearized using XbaI restriction digestion enzyme (NEB, Ipswitch, MA) to ensure uniform length transcription. For each IVT reaction, 10 μg of DNA plasmid was dissolved in cutsmart buffer (NEB, Ipswitch, MA) to a final concentration of 300 ng/μL. 40 units of XbaI restriction digestion enzyme was added into the reaction mixture. The reaction was carried out at 37 °C for 4 hours to ensure complete digestion. The reaction solution was allowed to cool down to room temperature. The DNA product was extracted with an equal volume of molecular biology grade phenol / chloroform / isoamyl alcohol (25:24:1) and vortexed for 2 minutes followed by a 5-minute centrifugation at 10,000 RCF. The top aqueous layer was transferred to a fresh tube and an equal volume of chloroform was added. The mixture was vortexed for 2 minutes followed by a 5 minute centrifugation at 10,000 RCF. The aqueous layer

was transferred into a fresh tube for DNA precipitation. The linearized DNA was precipitated by the addition of 0.1x volumes of 3M sodium acetate (pH = 5.2) and 2.5x volumes of 95% ethanol. The sample was chilled to -20 °C overnight, centrifuged at 16,100 RCF for 20 minutes at 4 °C, followed by gently removing ethanol. The DNA pellet was air dried and resuspended in 200 μL of RNase free water. The obtained DNA solution was used directly for IVT reaction. Each IVT reaction was set up with 50 ng/μL of linearized DNA template, 5 mM of each NTPs (ATP, CTP, UTP), 9 mM of GTP (NEB, Ipswitch, MA), 0.004 unit/μL of thermostable inorganic pyrophosphatase (NEB, Ipswitch, MA), 0.25 μg/μL T7 RNA polymerase, 0.05% Triton X-100 (Sigma, St. Louis, MO) and 1 unit/μL RNase Inhibitors, Murine (NEB, Ipswitch, MA). The IVT reaction was carried out at 37 °C for 4 hours to allow for sufficient RNA synthesis. To remove DNA template, 2 μL of 100 mM CaCl2 and 20 units of Turbo DNase (Life Technologies, Carlsbad, CA) were added to the mixture and incubated at 37 °C for 1 hour. The mixture was then centrifuged at 10,000 RCF for 5 minutes at room temperature to pellet any remaining magnesium pyrophosphate. The supernatant was resuspended in 200 μL RNase free water. To precipitate the in vitro transcribed mRNA (IVT-mRNA) product, the solution was added with 100 μL of 8 M LiCl and chilled to -20 °C for 4 hours, followed by 20 minutes of centrifugation at 16,000 RCF at 4 °C. The supernatant was removed gently. The remaining RNA pellet was resuspended in 200 μL RNase free water and quantified at 260 nm. The IVT-mRNA was confirmed as a single observable UV shadowing band by 4% denaturing PAGE (4% polyacrylamide in TBE with 8M urea) and kept frozen at -20 °C until used

Maturation of in vitro transcribed mRNA

Vaccinia capping system (NEB, Ipswitch, MA) was used to add a 7-methylguanylate cap structure (Cap 0) to the 5' end of IVT-mRNA to allow 5'-cap dependent translation initiation. IVT-mRNA was diluted in 29.2 μL RNase free water with a final concentration of 0.5 μg/μL. The RNA solution was heated to 65 °C and held for 5 minutes to denature RNA. The mixture

was then chilled on ice and held for another 5 minutes, followed by the addition of 4 μL of 10x capping buffer, 2 μL of 10 mM GTP, 2 μL of 4 mM S-adenosylmethionine (SAM), 32 units of RNase inhibitor and 2 μL of Vaccinia Capping Enzyme. The reaction mixture was incubated at 37 °C for 45 minutes. The capped IVT-mRNA product was purified by LiCl precipitation and directly used for polyadenylation. The overall yield for capping reaction and LiCl precipitation was around 90%. E. coli Poly(A) Polymerase (NEB, Ipswitch, MA) was used to polyadenylate capped IVT-mRNA. 30 mL of reaction mixture containing 15 mg of capped IVT-mRNA, 1 mM ATP, 30 units of RNase inhibitor was prepared in 1x E. coli Poly(A) polymerase reaction buffer and incubated at 37 °C for 1 hour. The maturated IVT-mRNA was purified by LiCl precipitation and directly used for in vitro TGT labeling.

In vitro Transcription of mod-mRNA

Appropriate IVT reaction mixtures were assembled following a previously described protocol.[71] For a 5mC substituted mod-mRNA IVT reaction, a mixture of 7.5 mM ATP, 1.9 mM GTP, 7.5 mM UTP and 7.5 mM 5mCTP (Trilink, San Diego, CA), 2.5 mM anti-reverse cap analog (ARCA) (Trilink, San Diego, CA), 1 unit RNase inhibitor, 2 μg DNA template and 1x MegaScript T7 RNA polymerase master mix were assembled in 1x MegaScript T7 RNAP reaction buffer to give a total volume of 40 mL. The reaction mixture was carried out at 37 °C for 4 hours. To remove the DNA template, 2 μL of 100 mM CaCl2 and 20 units of Turbo DNase (Life Technologies, Carlsbad, CA) was added into the reaction mixture and incubated for 1 hour at 37 °C. The mixture was then centrifuged at 10,000 RCF for 5 minutes at room temperature to pellet any remaining magnesium pyrophosphate. The supernatant was resuspended in 200 μL RNase free water. To precipitate the mod-mRNA product, the solution was added with 100 μL of 8 M LiCl and chilled to -20 °C for 4 hours, followed by 20 minutes of centrifugation at 16,000 RCF at 4 °C. The supernatant was removed gently. The remaining RNA pellet was resuspended in 200 μL RNase free water and quantified at 260 nm.

2.4.4 In vitro TGT Labeling and Purification

TGT Labeling of IVT-mRNA

TGT labeling conditions were adapted from our previous study.[27] Synthesized TGT substrate compound 7 was dissolved in RNase free water to obtain a concentration of 650 μM. For in vitro TGT reaction, 1 μM of IVT-mRNA, 50 μM of compound 7, 2 unit/μL RNase inhibitor and 1.5 μM of TGT enzyme was assembled in 1x TGT reaction buffer. The reaction mixture was incubated at 37 °C for 4 hours. The crude labeled RNA transcript was purified by LiCl precipitation.

Purification of TGT-labeled mRNA Transcript

The crude TGT labeling product was further purified using Dynabeads[TM] M-280 Strepta-vidin. Dynabeads[TM] M-280 Streptavidin beads stock solution (Thermo Fisher Scientific, Waltham, MA) was vortexed for 30 seconds. 150 μL of the well-mixed stock solution was transferred to a 2 mL Eppendorf tube. Following manufacture's protocal, the magnetic beads were washed with 300 μL of 1x BW&T buffer for 3 times, 300 μL of solution A for 2 times, 300 μL of solution B for 2 times and 300 μL of 1x BW&T buffer for 2 times After washing, the beads were incubated with labeled IVT-mRNA (use amount if available) in 1x BW&T buffer for 25 minutes to allow for the binding of IVT-mRNA to the magnetic beads. The solution was then rotated at room temperature for 25 minutes to allow for the binding of IVT-mRNA and the magnetic beads. The RNA bound beads were washed with 300 μL of 1x BW&T buffer 2 times and 300 μL of 1x B&W buffer 1 time. The beads were then resuspended in 100 μL elution buffer and incubated at 65 °C for 3 minutes, followed by centrifugation at 7,000 RCF for 1 minute. The resulting supernatant was then subjected to ethanol precipitation with the addition of 0.1x volumes of 3M sodium acetate and 2.5x volumes of ethanol. The mixture was held at - 80 °C overnight to precipitate RNA, followed by centrifugation at 7,000 RCF for 3 minutes and 17,000 RCF for 17

minutes. The pellet was resuspended in 15 uL RNAse free water. The Dynabeads purified labeled IVT-mRNA was stored in - 20 °C until used. The overall purification yields for RNA constructs with three tags, two tags and one tag were around 57%, 35% and 24%. It should be noted that the Dynabeads purification yields didn't mean the TGT labeling yield. The low purification yield might because of insufficient binding of the labeled IVT-mRNAs with the streptavidin beads and RNA degradation during washing and elution processes. The TGT labeling efficiency was discussed in the degree of labeling session.

Degree of Labeling of Compound 7 on IVT-mRNA Transcripts

To determine the number of photo cages (compound 7) that are actually labeled on each mRNA transcript, a fluorescence biotin quantitation assay (Thermo Scientific, Waltham, MA) was performed. This commercially available microplate-based assay measures the fluorescence of the Thermo Scientific DyLight Reporter (fluorescent avidin and HABA premix) upon binding with a biotinylated sample, in this case, the TGT labeled biotinylated RNA. The avidin fluoresces when the weakly interacting HABA (4'-hydroxyazobenzene-2-carboxylic acid) is displaced by the biotinylated RNA. The amount of biotin for each sample can be determined by comparing the fluorescence intensity of the sample with a standard curve. To obtain the biotin standard curve, 0 pmol/μL, 0.5 pmol/μL, 1 pmol/μL, 2 pmol/μL and 4 pmol/μL biocytin standard solutions was prepared. For this fluorescence biotin quantitation assay, all TGT labeled IVT-mRNA transcripts were purified using Dynabeads streptavidin pull-down to remove unlabeled RNA transcripts. The Thermo Scientific DyLight Reporter has an excitation maximum at 494 nm and an emission maximum at 520 nm. To measure the fluorescence, 2 μL of standard biocytin solution or RNA sample was mixed with 18 uL of 1x DyLight Reporter. The mixture was incubated for 5 minutes at room temperature and measured using fluorescence plate reader. The standard curve and all measurements were performed in triplicates. By comparing the biotin and RNA concentration for each sample, the average biotin labels per mRNA transcript was obtained (Figure 2.18). For

IVT-mRNA transcript with one, two or three active Tags, there is an average number of 0.94, 1.60 or 2.20 biotins per mRNA construct, respectively.
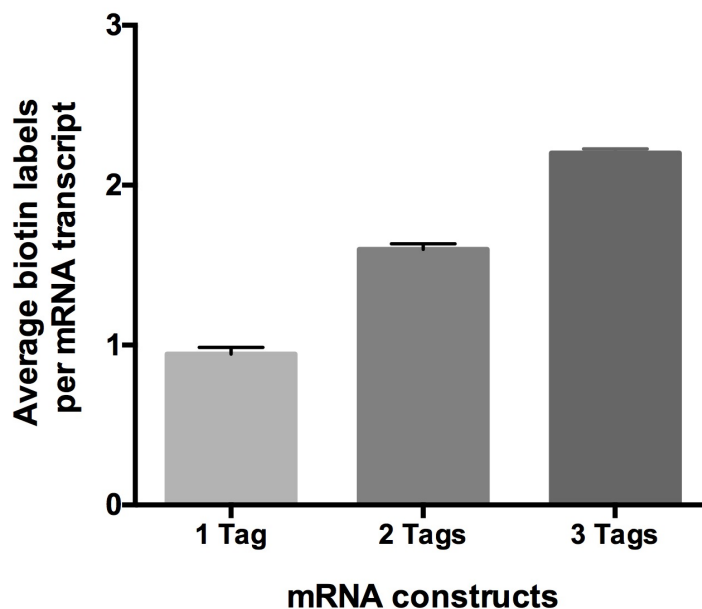
Figure 2.18: The average number of TGT labeled biotins per IVT-mRNA construct. An average of 0.94, 1.60 and 2.20 biotins were conjugated to each mRNA, respectively. All measurements were performed in triplicates. The error bars denote the standard deviation of three measurements.

Determination of TGT Labeling Efficiency

In addition to using the biotin quantification assay to determine the efficiency of TGT labeling, we have also enzymatically labeled the mRNA constructs containing one, two and three functional Tags with compound 8, a Cy5 fluorophore conjugated preQ1 derivative that has a high extinction coefficient ($\varepsilon = 250,000$ cm^{-1}M^{-1}) allowing for a direct absorption measurement. The concentration of the mRNA and the Cy5 were determined by measuring absorptions at 260 nm and 650 nm using nanodrop, respectively. The reason that the TGT labeling efficiency determined using this method (Figure 2.19) is lower than that in Figure 2.18, is presumably because that the mRNA constructs used in this experiment were not purified using affinity columns to remove

unlabeled constructs, suggesting that an affinity purification is important to enhance mRNA purity.
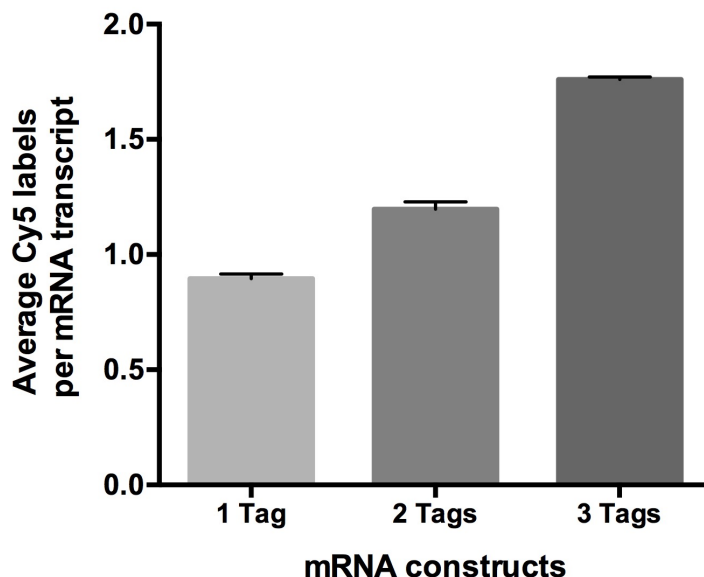
Figure 2.19: Bar graph demonstrates a quantitative degree of labeling of the mRNA. Absorbance of enzymatic labeled RNA at 650 and 260 nm were measured to determine the concentration of Cy5 and mRNA, respectively. The degree of labeling was calculated as the ratio of the two concentrations. An average of 0.90, 1.20 and 1.76 Cy5 probes were conjugated to each mRNA construct. All measurements were measured in triplicates. Error bars denote the standard deviation of three measurements.

2.4.5   In vitro Uncaging of Photo-caged RNA

To investigate whether the photo cages (compound 7) can be released from RNA upon light irradiation, in vitro UV uncaging assay of the photo-caged 17 nts RNA hairpin (photo-caged Tag) was performed.

Synthesis of Photo-caged Tag

In a 1.7 mL Eppendorf tube, a 1.2 mL TGT labeling reaction was assembled with the following components in 1x TGT reaction buffer: 10 μM of Tag (17 nts RNA hairpin), 5 μM of

TGT enzyme, 30 μM of compound 7, 5 mM DTT. The reaction was incubated at 37 °C for 3 hours. The reaction mixture was then centrifuged at 16,000g for 10 min, and the supernatant was subject to ethanol precipitation, followed by centrifugation at 16,000 for 15 min. The resulting pellet was resuspended in 100 μL RNase free water. The biotin labeled RNA was purified by HPLC. HPLC purification method: 0 min – 1 min: 5% Phase D in Phase C, 1 min – 10 min: 5% Phase D in Phase C to 45% Phase D in Phase C, 10 min – 12 min: 5% Phase D in Phase C (Phase C: 20% HFIP in H_2O with 0.1% TEA, Phase D: MeOH). Retention time: 9.2 min.

ESI-TOF MS of biotin labeled Tag. The original ESI-TOF MS spectrum and deconvoluted mass are shown in Figure 2.20 and Figure 2.21, respectively.
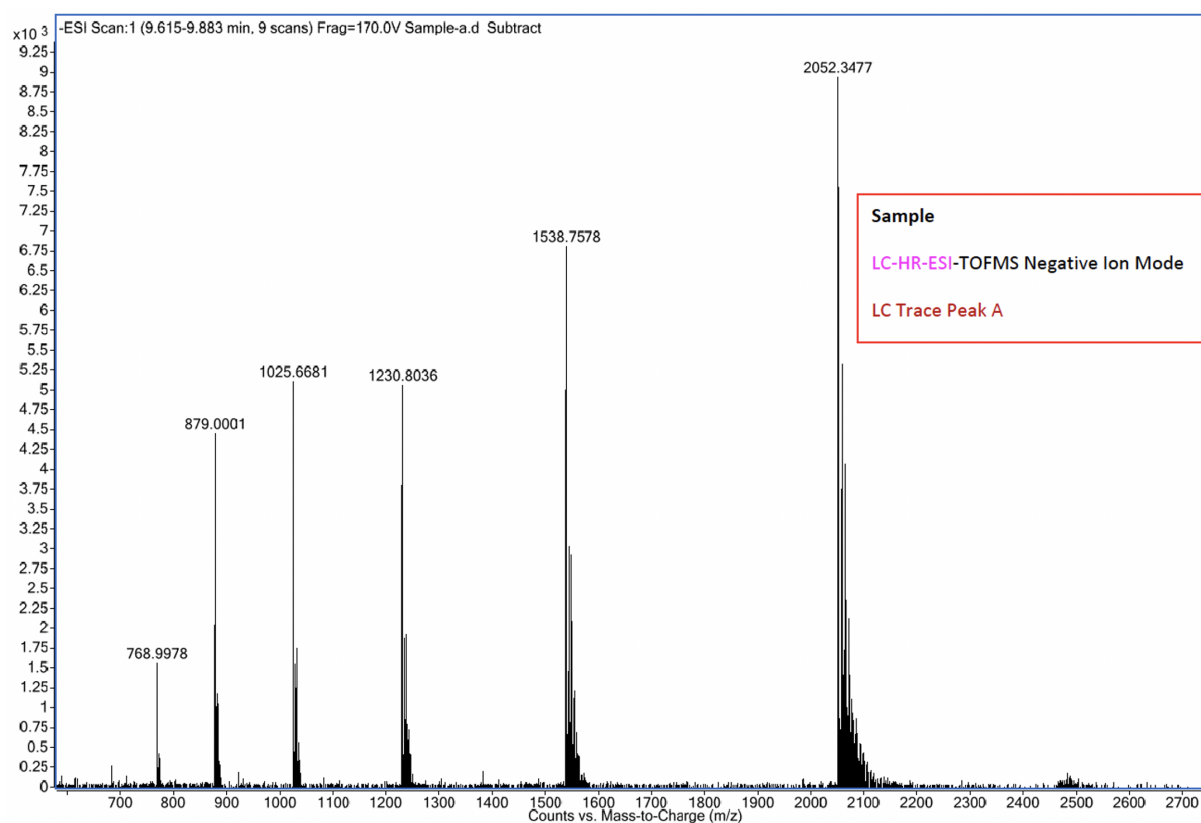
Figure 2.20: ESI-TOF MS spectrum of photo-caged Tag.

6160.1

Deconvoluted mass spectrum

Sample

LC-HR-ESI-TOFMS Negative Ion Mode

LC Trace Peak A
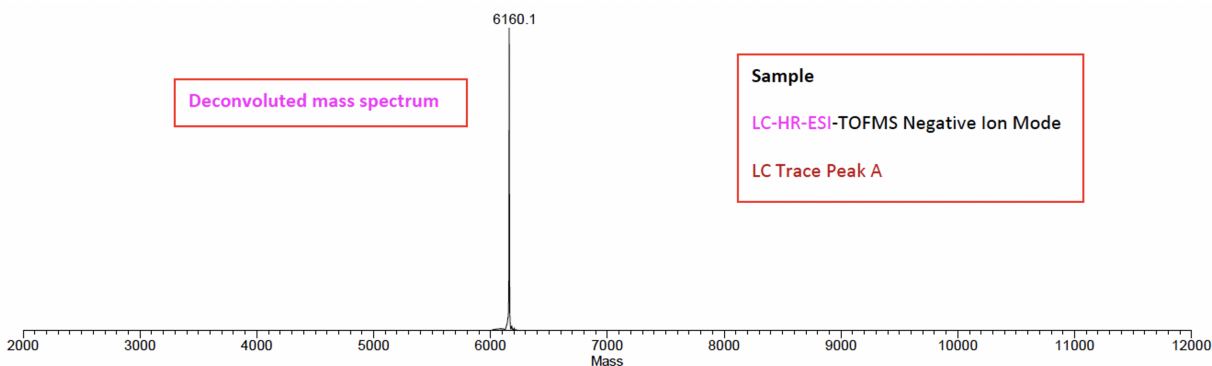
Figure 2.21: deconvoluted mass of photo-caged Tag.

UV cleavage of Photo-caged RNA Tag

To set up the in vitro UV uncaging experiments, a Nunc Lab-Tek 8-well chamber slide containing 100 μL 25 μM photo-caged Tag was placed in a Stratalinker UV 1800 (Stratagene, La Jolla, CA). UV energy ranging from 500 μJ to 20,000 μJ was applied to the photo-caged Tag, and the UV irradiated samples were subsequently analyzed by HPLC. Normalized representative HPLC traces for photo-caged Tag treated with different UV dosage demonstrate that the caged RNA could be fully uncaged (Figure 2.9). The HPLC method was the same as described in previous section. The peaks were integrated to determine the percentage conversion (Figure 2.9).

2.4.6   Live-cell Imaging and Photo-uncaging

IVT-mRNA Transfection

HeLa cells (ATCC, Manassas, VA) were cultured in DMEM media (Life Technologies, Carlsbad, CA) with 10% FBS and pen/strep. HeLa cells were plated at an initial density of around 40,000 cells per well in a Nunc Lab-Tek 8 well chamber slide (Thermo Scientific, Waltham, MA). Cells were allowed to adhere overnight, washed with Opti-MEM media (Life Technologies, Carlsbad, CA) and subsequently transfected in Opti-MEM media with an addition of 400 ng of IVT-mRNA and 1.0 μL of Lipofectamine RNAiMAX (Life Technologies, Carlsbad, CA). Cells

were transfected for 4 hours before washed with DMEM to remove any transfection reagents. Cells were allowed to grow in DMEM at 37 °C overnight to allow for sufficient protein expression.

Fluorescence Microscopy Imaging

All images, unless otherwise indicated, were acquired on a Yokagawa spinning disk system (Yokagawa, Japan) built around an Axio Observer Z1 motorized inverted microscope (Carl Zeiss Microscopy GmbH, Germany) with a 20x 1.42 NA objective to an Evolve 512x512 EMCCD camera (Photometrics, Canada) using ZEN imaging software (Carl Zeiss Microscopy GmbH, Germany). EGFP was excited with a 488 nm, 100 mW OPSL laser, green. EGFP expression from the IVT-mRNA was quantified by measuring average fluorescence level among more than 80 HeLa cells. Cells images were processed using ImageJ software.

Activities of Labeled RNA-Tag1/2/3 And Unlabeled RNA-Tag1/2/3

We used fluorescence microscopy to determine the translation efficiency of photo-caged RNA-Tag1/2/3 and unlabeled RNA-Tag1/2/3. A significant decrease in EGFP expression (90% decrease) was observed in cells transfected with photo-caged RNA-Tag1/2/3 compared to cells transfected with unlabeled RNA-Tag1/2/3 (Figure 2.4).

405 nm Laser Toxicity

To test the toxicity of the 405 nm laser used for live cell uncaging experiment, we did a 405 nm toxicity assay. HeLa cells were transfected with RNA-Tag0 following the protocol described above in the "IVT-mRNA" transfection session. After transfection, cells were washed with HBSS twice and incubated in DMEM. Selected cells (indicated with a blue arrow) were irradiated using a DirectFrap module with 405 nm laser line at 100% laser power for 0 second, 5 seconds, 10 seconds, 15 seconds and 25 seconds. Cell images were taken after 1 hour, 2 hours, 3.5 hours and 5 hours after the laser irradiation event (Figure 2.22). We observed that only 25

seconds of 405 nm laser irradiation killed the HeLa cells.

Live-cell Photo-uncaging

HeLa cells were transfected with photo-caged IVT-mRNA. After transfection, cells were washed with HBSS twice and incubated in DMEM. Selected cells were irradiated using a DirectFrap module with 405 nm laser line at 100% laser power for 10 seconds to allow for the removal of labeled photo-cages on IVT-mRNA. Upon photo-uncaging, cells were incubated at 37 °C in 5% CO2 for 8 hours to allow for the recovery of EGFP expression. For mod-mRNA live cell photo-uncaging, HeLa cells were transfected with photo-caged mod-mRNA and illuminated with a 405 nm laser line at 100% laser power for 10 seconds. Cells were incubated at 37 °C in 5% CO2 for 8 hours and imaged using fluorescence microscopy. Successful uncaging was demonstrated by the recovery of protein expression only in selected cell treated with laser irradiation (Figure 2.23). Scale bar = 40 μm.

2.4.7 NMR Spectra

NMR spectrums of compounds in Charpter 2 are shown in Figure 2.24 to Figure 2.32.

2.5 Acknowledgements
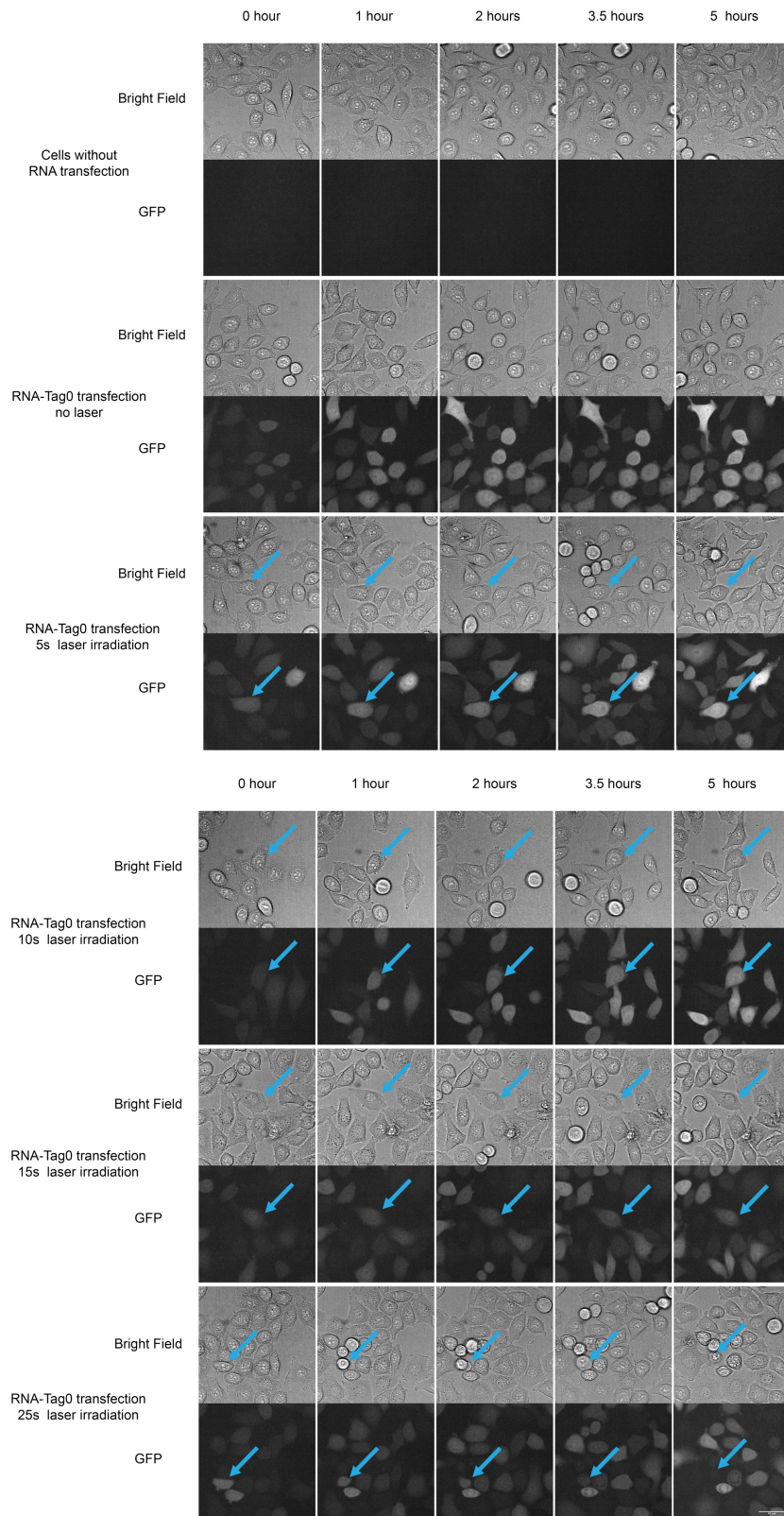
Figure 2.22: 405 nm laser cell toxicity. Blue arrow selected cells were irradiated with 405 nm laser for 0 second, 5 seconds, 10 seconds, 15 seconds and 25 seconds. Scale bar = 40 μm.

Figure 2.23: Mod-mRNA uncaging using 405 nm laser light. Red circles represent the selected HeLa cells irradiated with 405 nm laser light.

Figure 2.24: ^{1}H NMR of compound 2 in Chapter 2

44
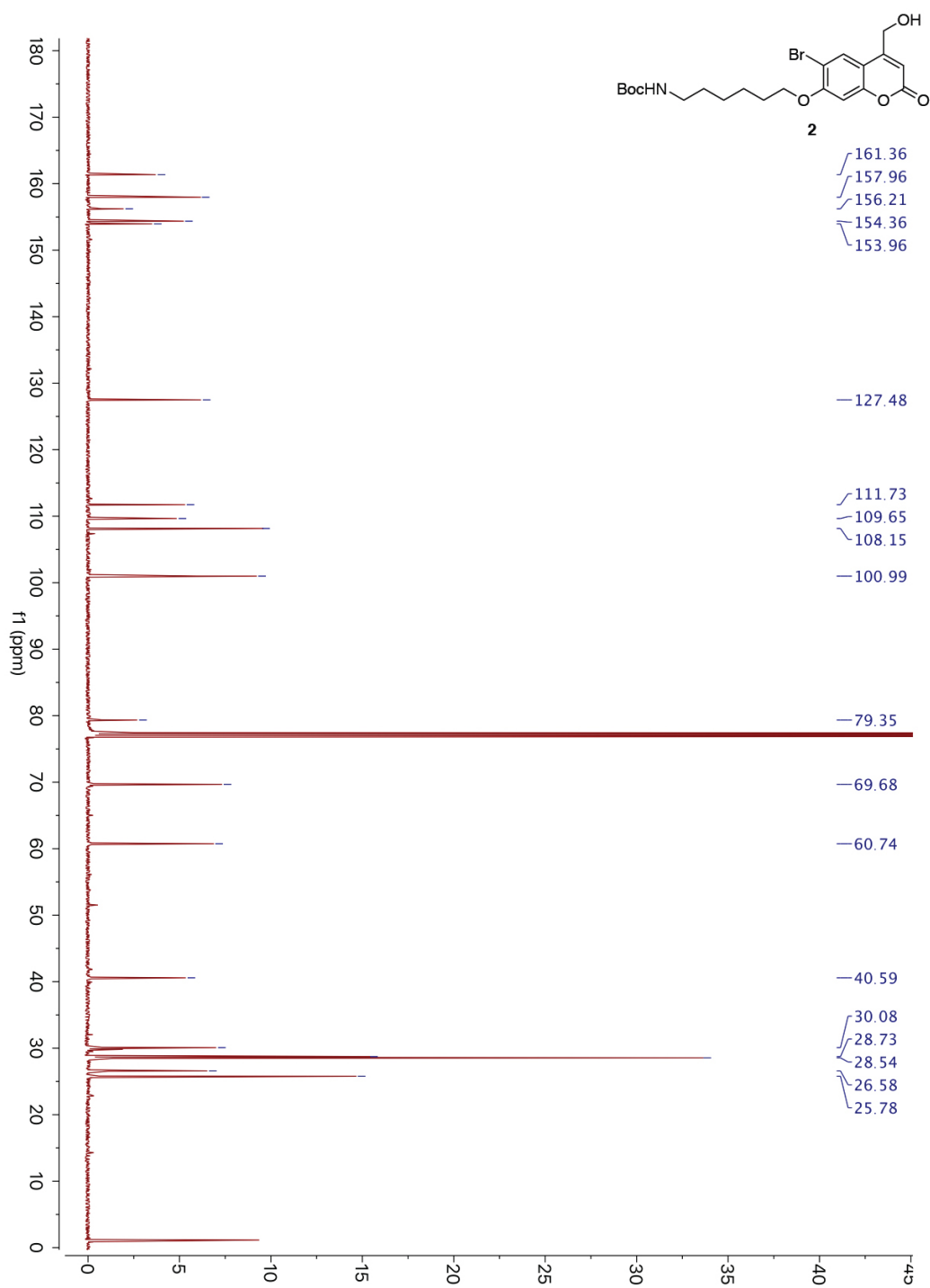
Figure 2.25: ^{13}C NMR of compound 2 in Chapter 2

45
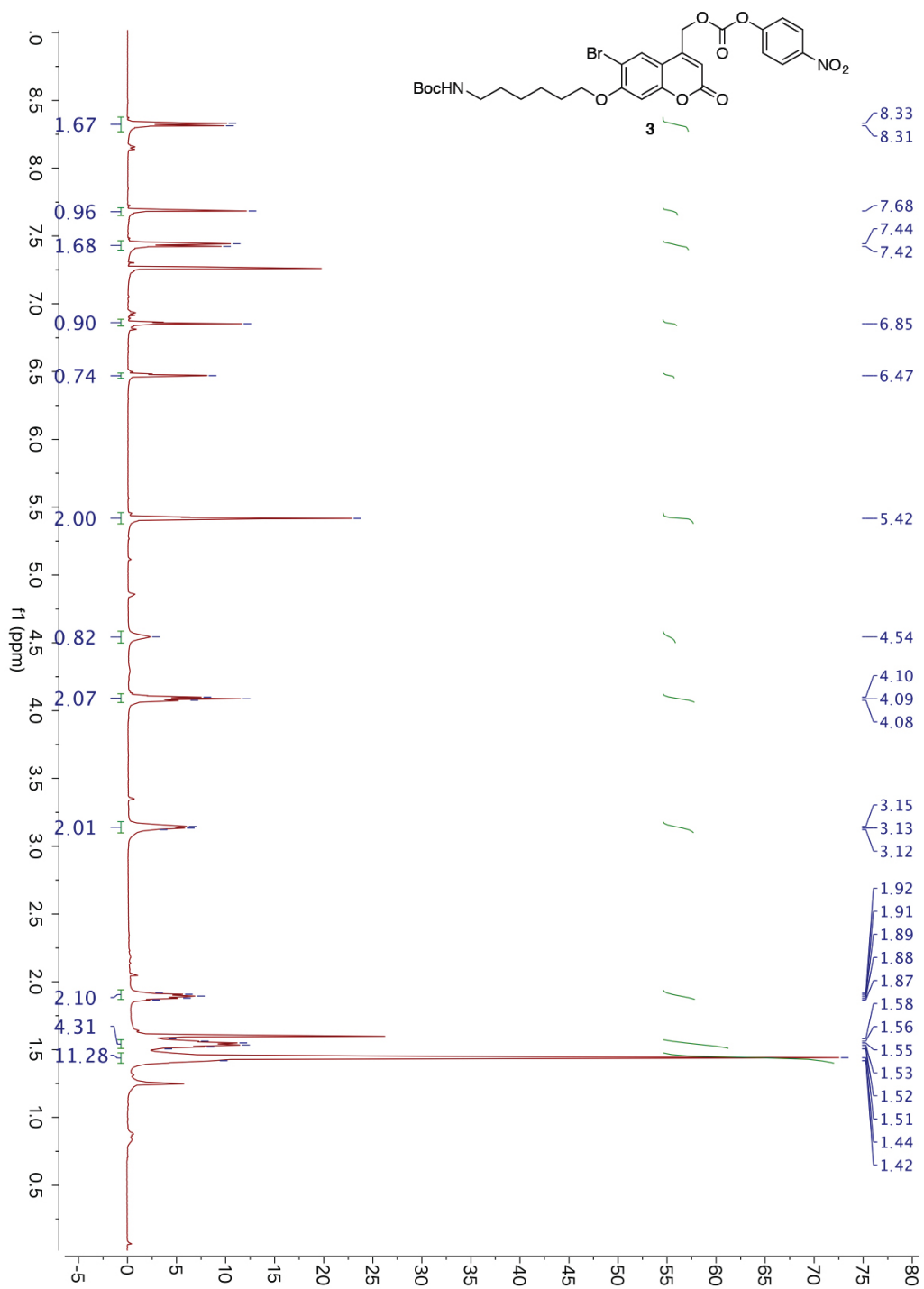
Figure 2.26: ^{1}H NMR of compound 3 in Chapter 2

46
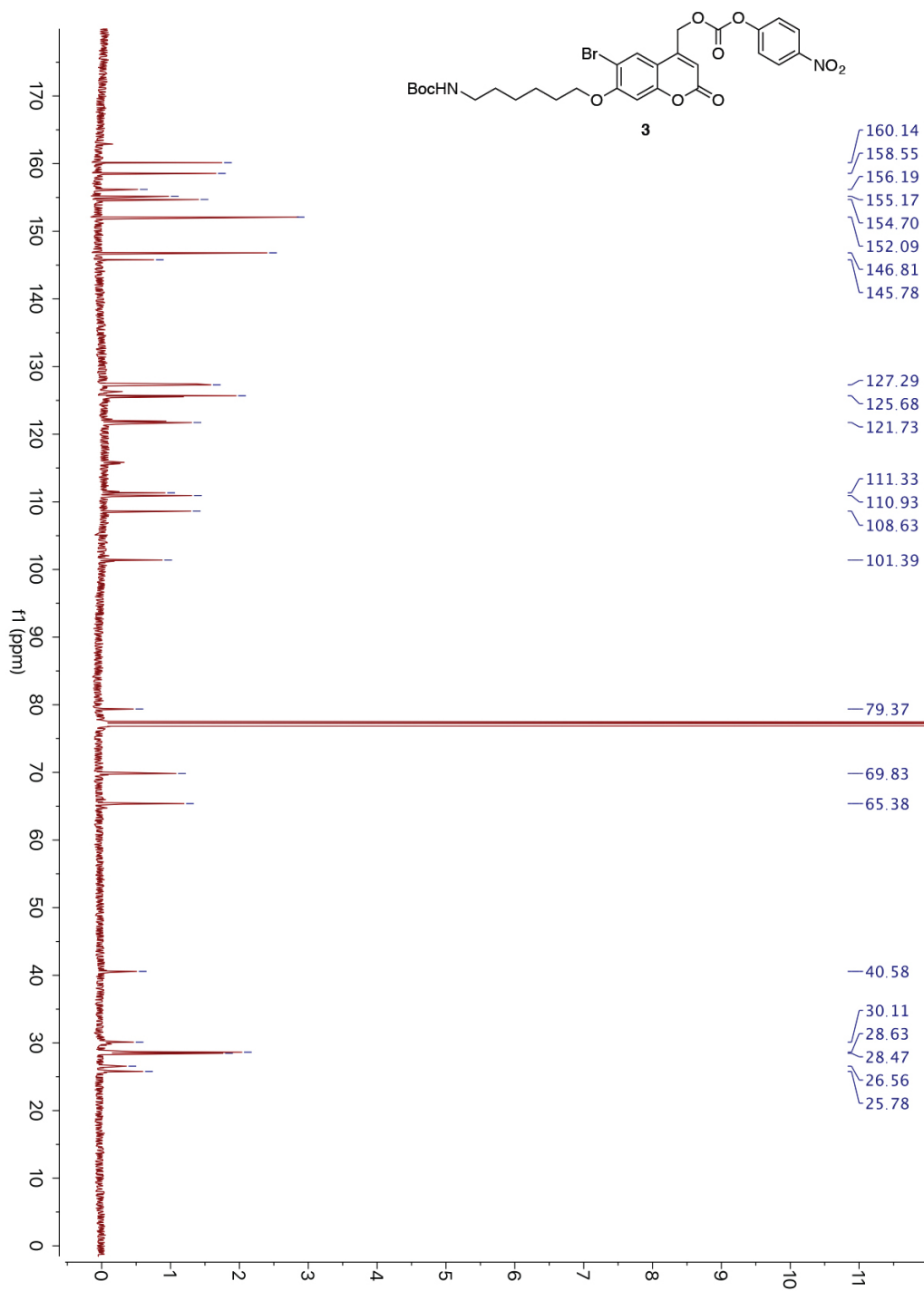
Figure 2.27: ^{13}C NMR of compound 3 in Chapter 2

47
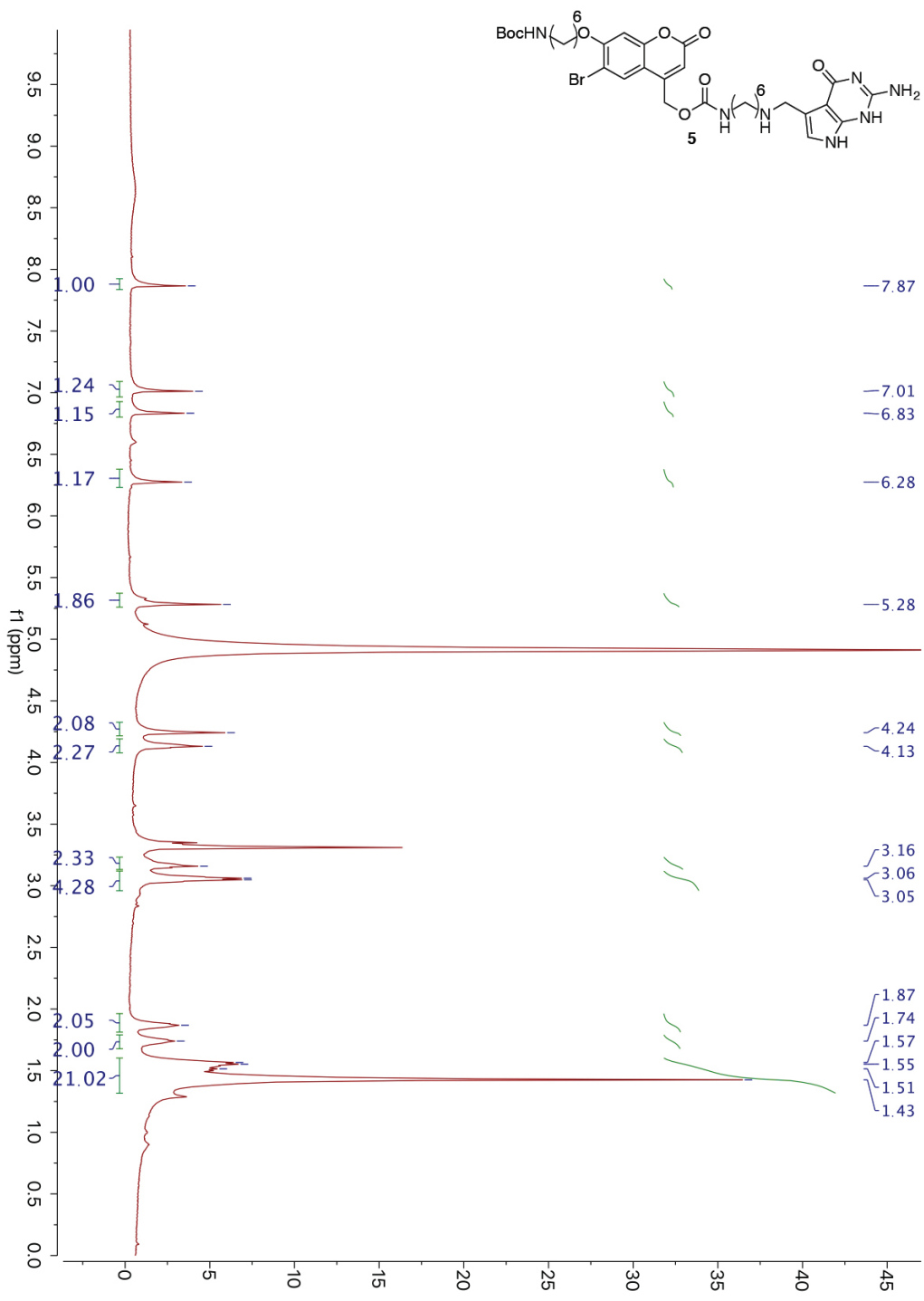
Figure 2.28: [^1]H NMR of compound 5 in Chapter 2

48
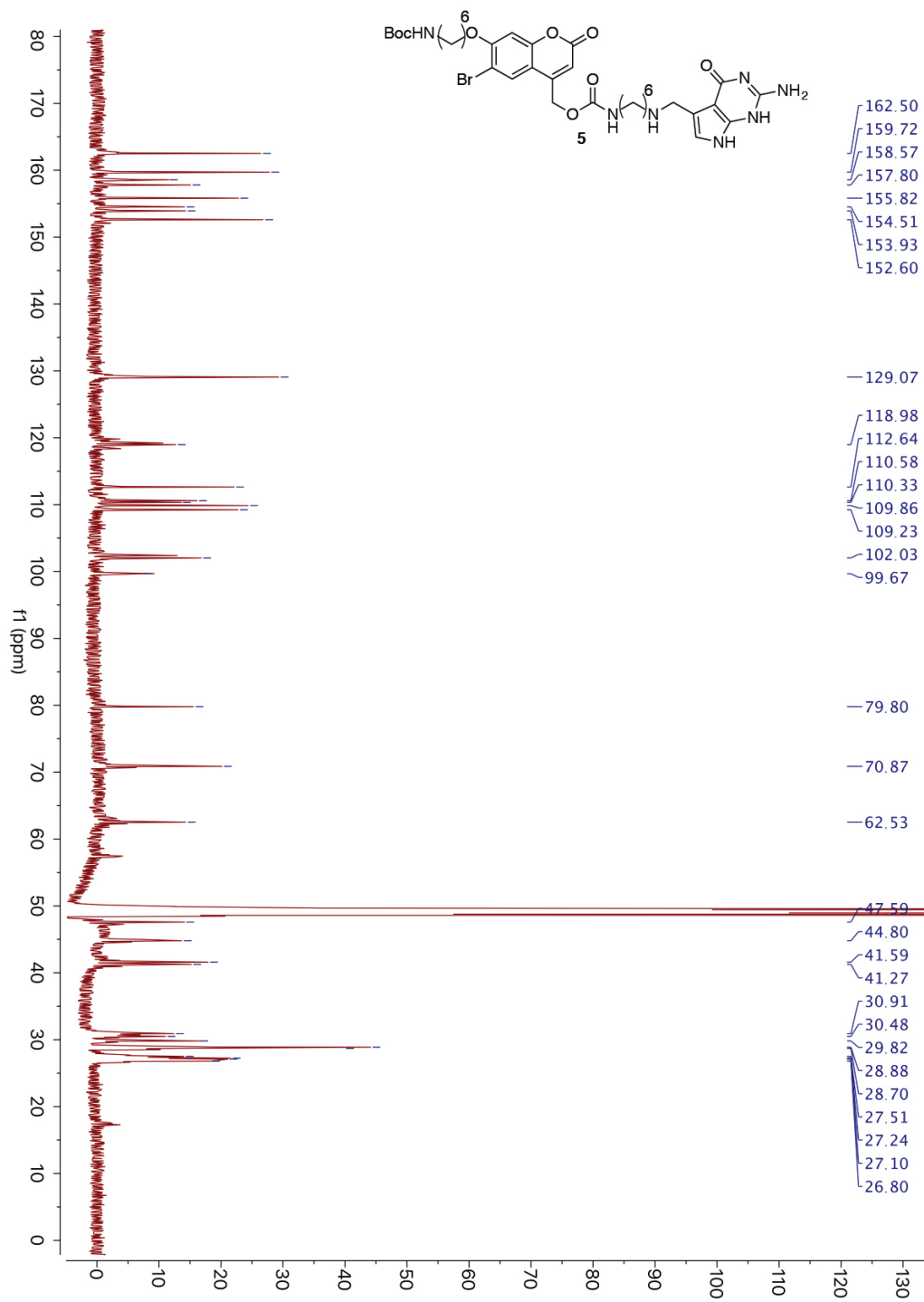
Figure 2.29: ^{13}C NMR of compound 5 in Chapter 2

49

Figure 2.30: ^{1}H NMR of compound 7 in Chapter 2
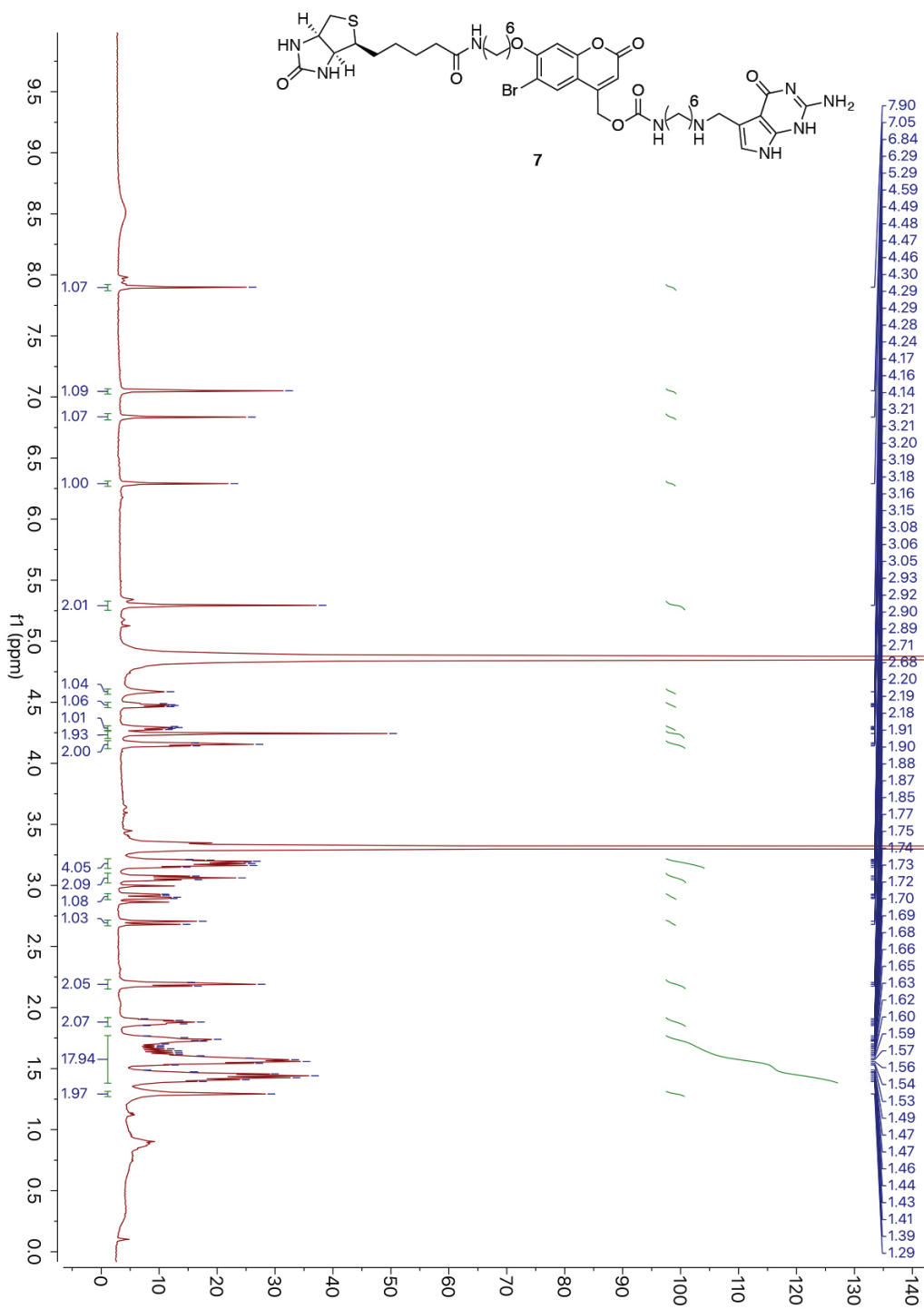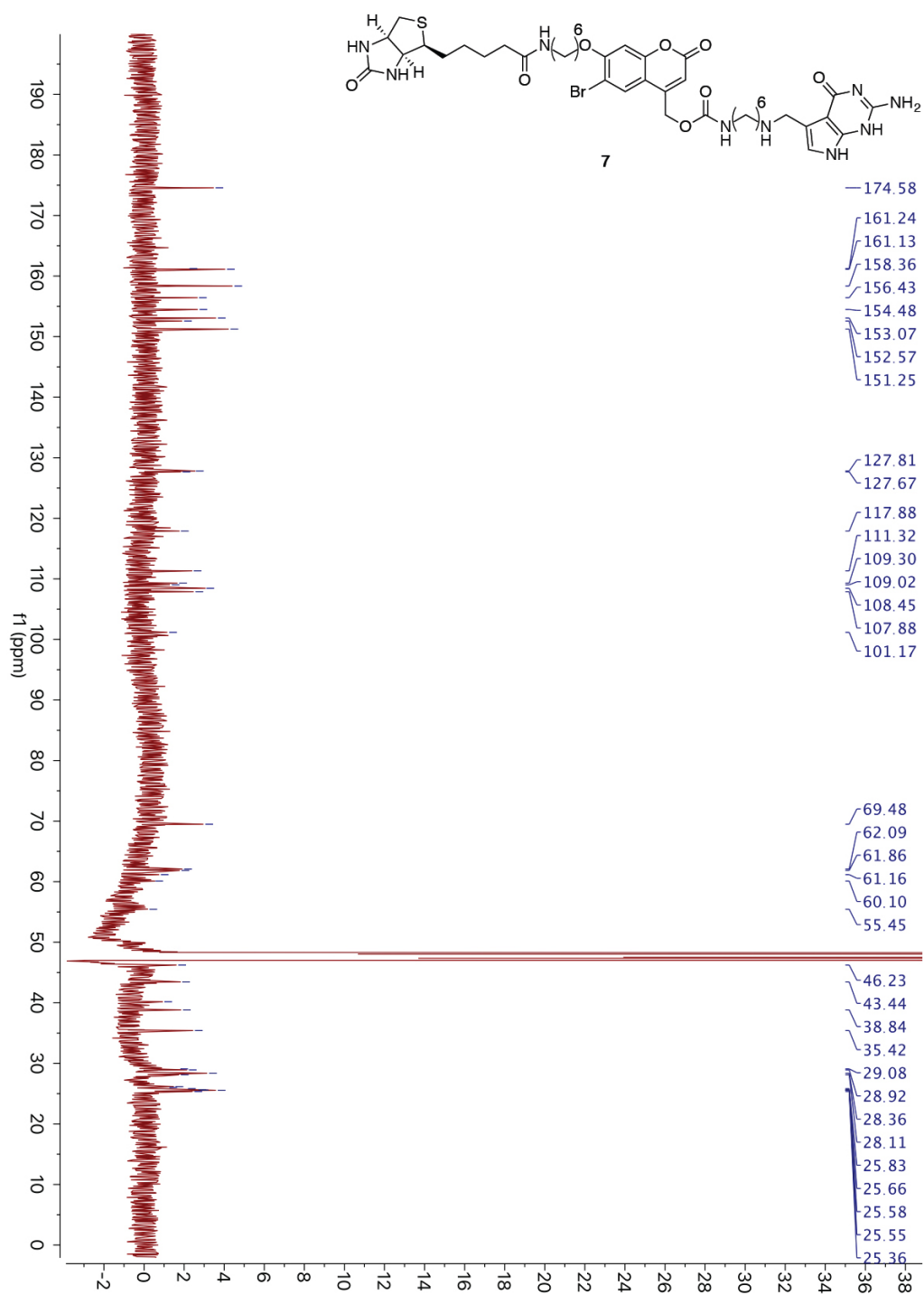
Figure 2.31: ^{13}C NMR of compound 7 in Chapter 2

51
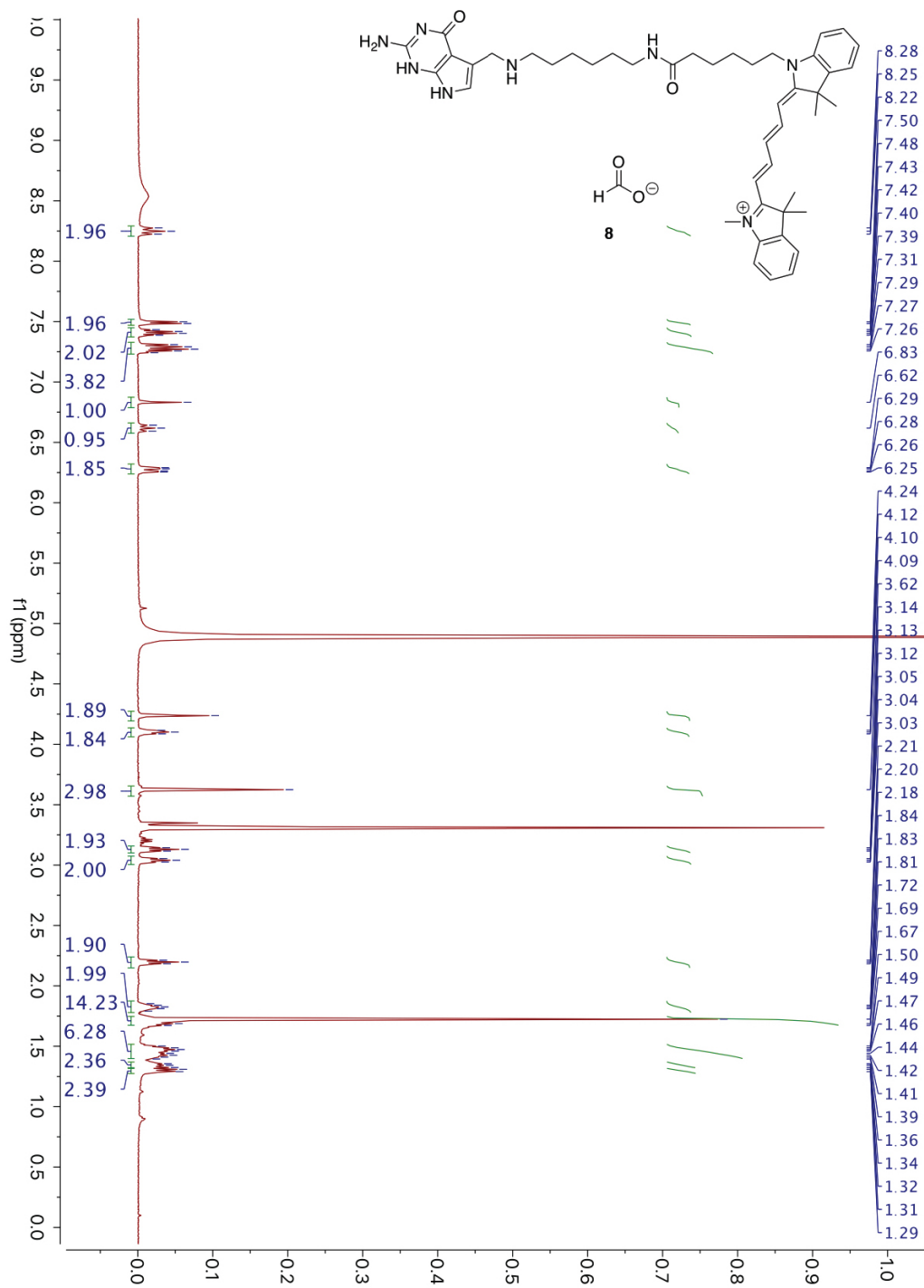
Figure 2.32: ^{1}H NMR of compound 8 in Chapter 2

Chapter 3

Multiplexed Light-activation of mRNA Translation With Single-cell Resolution

3.1 Introduction

We demonstrate sequential optical activation of two types of mRNAs in the same mammalian cell through the sequential photocleavage of small molecule caging groups ('photo-cages') tethered to the 5' untranslated region (5'-UTR) of an mRNA. Synthetic 'photo-cages' were conjugated onto target mRNA using RNA-TAG, an enzymatic site-specific RNA modification technique. Translation of mRNA was severely reduced upon conjugation of the 'photo-cages' onto the 5'-UTR. However, subsequent photo-release of the 'cages' from the mRNA transcript triggered activation of translation with single-cell spatiotemporal resolution. To achieve sequential photo-activation of two mRNAs in the same cell, we synthesized a pair of 'photo-cages' which can be selectively cleaved from mRNA upon photo-irradiation with different wavelengths of light. Sequential photo-activation of two mRNAs enabled precise optical control of translation of two unique transcripts. We believe that this modular approach to precisely and rapidly control gene expression will serve as a powerful tool in future biological studies that require controlling

translation of multiple transcripts with high spatiotemporal resolution.

Previously, we demonstrated a technique which enabled precise optical control of mRNA translation through the conjugation of light-sensitive 'cages' onto IVT-mRNA.[72] Laser irradiation (405 nm) on live cells removed the 'photo-cages' from cytoplasmic mRNA, subsequently activating translation with single cell precision. We speculated we could significantly expand this tool by enabling sequential photo-activation of two mRNAs within the same cell. Here, leveraging the multiplex capability of our mRNA caging/uncaging platform, we describe a sequentially light-activated translation regulatory system that utilizes two 'photo-cages' to cage two types of mRNAs. Irradiation with longer wavelength light (456-488 nm) activates one mRNA, while subsequent irradiation with shorter wavelength light (365-405 nm) activates the other mRNA.3.1 This multiplexed gene expression regulatory system provides a high degree of flexibility and shows potential for enabling the study of multiple regulatory genes with high spatial-temporal resolution.
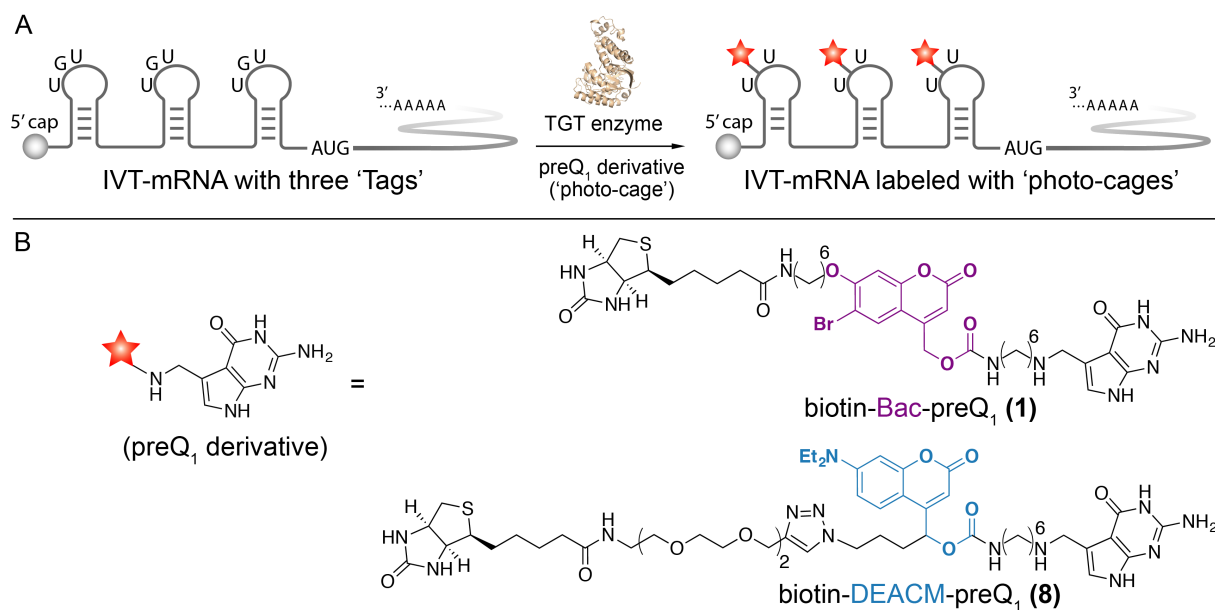
Figure 3.1: mRNA photo-caging using two 'photo-cages'. A) To facilitate TGT enzymatic labeling, three enzyme recognition sequences, 'Tags', are genetically inserted along the 5'-UTR of an IVT-mRNA. Subsequent conjugation of the 'photo-cages' severely reduces mRNA translation activity. B) Chemical structures of two sequentially activable preQ1 derivatives ('photo-cages'), biotin-Bac-preQ1 and biotin-DEACM-preQ1.

3.2 Results and Discussions

3.2.1 Choice of The Second Photo-sensitive Cage

Previously, we chose the photo-cleavable linker 6-bromo-7-aminoethoxycoumarin-4-ylmethoxycarbonyl (Bac) to synthesize the 'photo-cage', biotin-Bac-preQ1 (1).[72] To allow for sequential photo-activation of mRNA translation, an additional photo-sensitive linker that is responsive to a longer wavelength of light was desired. Inspired by previously reported work,[73] we chose to explore [7-(diethylamino)coumarin-4-yl]-methyl (DEACM) as an additional photo-activable linker that can be sequentially activated. The DEACM linker has a wide absorbance spectrum and was previously reported to be cleaved in cellular conditions by irradiation with 470 nm light.[44] Thus, the DEACM linker should form a sequentially photo-activable linker pair with our previously reported Bac linker, which is uncaged by irradiation with 365-405 nm light. To synthesize the new 'photo-cage' biotin-DEACM-preQ1 (8) (Figure 3.2), the DEACM-based building block (2) was subjected to sequential mesylation and azidation to generate azide compound (3). Next, the DEACM NHS-ester (4) was obtained through deprotection of the silica protecting group and N-succinimidyl carbonate (DSC) treatment. Subsequently, the preQ1 derivative (5) was coupled to the DEACM NHS-ester (4) to yield the preQ1-DEACM conjugate (6). To introduce a biotin affinity handle to facilitate purification of the labeled mRNAs, the preQ1-DEACM conjugate (6) was further coupled with commercially available biotin-PEG4-alkyne (7) via click chemistry to obtain the final TGT enzymatic substrate biotin-DEACM-preQ1 (8).

3.2.2 Sequential Release of 'photo-cages' from RNA Oligos

To demonstrate that biotin-Bac-preQ1 (1) and biotin-DEACM-preQ1 (8) can be sequentially released from RNA upon photo-irradiation with two wavelengths of light, in-vitro uncaging of labeled RNA oligos was performed using either 365 nm or 456 nm light followed by 18%
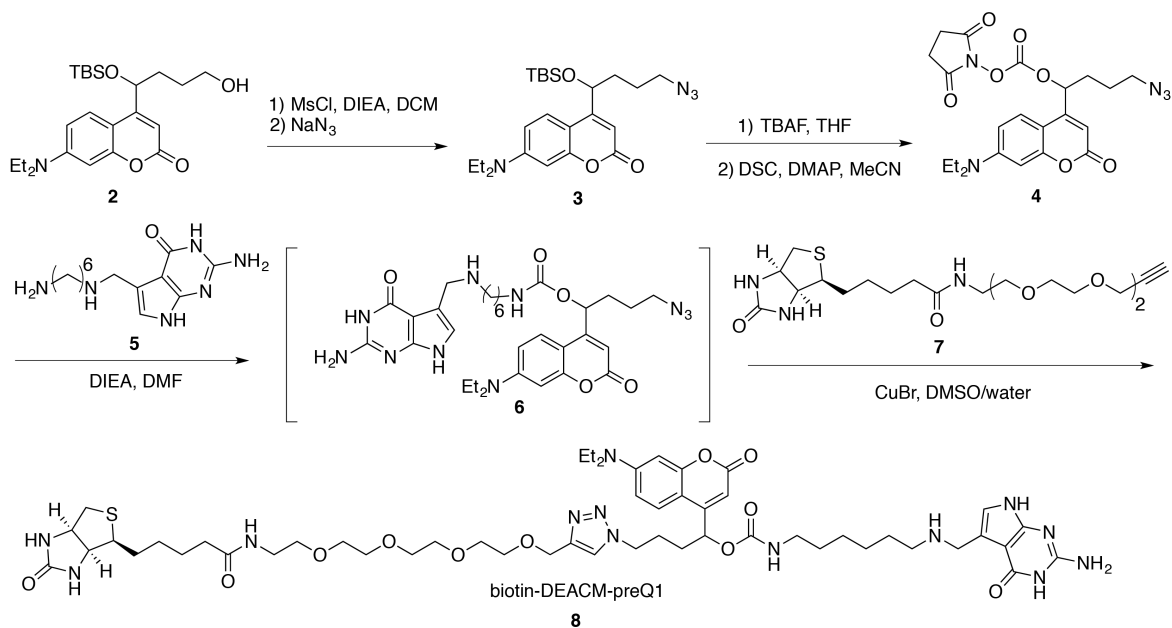
Figure 3.2: Synthesis of the biotin-DEACM-preQ1 compound 8. Boc = tert-butyloxycarbonyl protecting group, MsCl = methanesulfonyl chloride, DCM = dichloromethane, TBS = tert-butyldimethylsilyl protective group, TBAF = tetra-n-butylammonium fluoride, THF = tetrahydrofuran, DSC = N-succinimidyl carbonate, DMAP = 4-dimethylaminopyridine, DIEA = N,N-diisopropylethylamine, DMF = dimethylformamide, DMSO = dimethyl sulfoxide.

denaturing polyacrylamide gel electrophoresis (PAGE) analysis. The 'Tag' oligo was used as RNA substrate for TGT labeling (Figure 3.3A). To covalently conjugate the photo-cage onto the 'Tag' oligo, TGT labeling using either biotin-Bac-preQ1 or biotin-DEACM-preQ1 as small molecule substrate was carried out in a dark room with minimal red ambient light to prevent undesired degradation of the light-sensitive preQ1 derivatives. Labeled 'Tag' oligo was further purified by streptavidin-biotin pull-down followed by denaturing-PAGE analysis. Covalent conjugation of the 'photo-cage' increased the molecular weight of the 'Tag' oligo, resulting in significant RNA band shifts shown in denaturing-PAGE (Figure 3.3B).[27,72] To trigger the cleavage of the 'photo-cage', the 'Tag' oligo labeled with biotin-DEACM-preQ1 was irradiated with a 456 nm lamp, while the 'Tag' oligo labeled with biotin-Bac-preQ1 was irradiated with a 365 nm lamp. We observed that upon irradiation with 456 nm light, only biotin-DEACM-preQ1 was released from the 'Tag' oligo, not the biotin-Bac-preQ1. Biotin-Bac-preQ1 was released from the 'Tag' oligo upon irradiation

with 365 nm light. Thus, by using 456 nm and 365 nm light sources, sequential release of the 'photo-cages' from RNA was achieved.
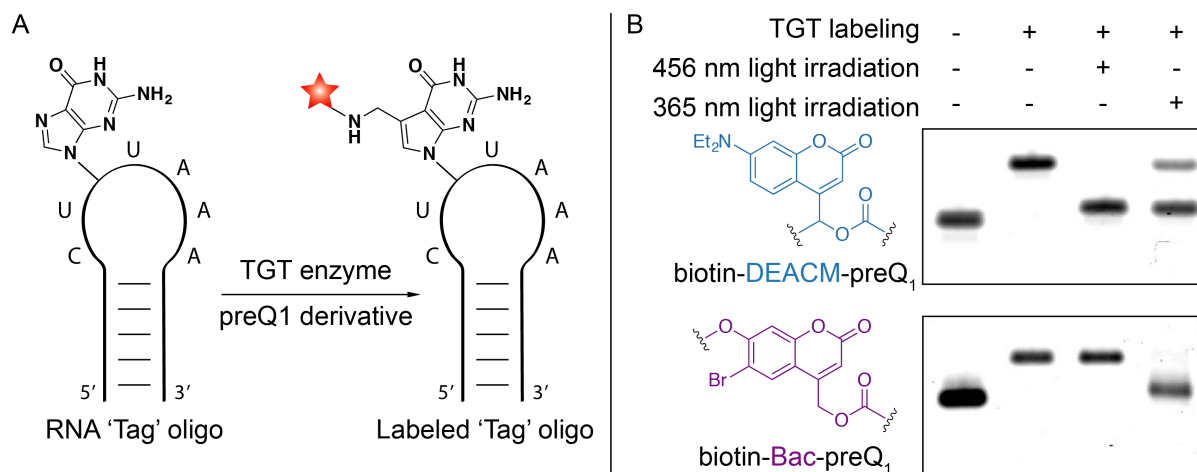
Figure 3.3: Sequential Release of 'photo-cages' from RNA oligos. A) In-vitro conjugation of RNA 'Tag' oligo with preQ1 derivatives through RNA-TAG. B) Denaturing 18% PAGE analysis of enzymatic labeling and subsequent photo-uncaging of the RNA 'Tag' oligos. Compared to the unlabeled RNA oligo, an upper gel shift in the second column demonstrates covalent attachment of the 'photo-cage' onto the RNA oligo. Lower gel shift in the third or fourth column demonstrates photo-release of the 'photo-cage' from the RNA oligo. After photo-cleavage, the resulting RNA band was only slightly up-shifted compared to the starting material, supporting that only a minimal residue was left on the RNA after photo-cleavage.

3.2.3 Sequential Live-cell Photo-activation of mRNA Translation

Having demonstrated that biotin-DEACM-preQ1 and biotin-Bac-preQ1 can be sequentially released from RNA oligos in-vitro, we examined whether these 'photo-cages' can be sequentially photo-released from mRNA in live mammalian cells. Mature IVT-mRNAs coding for either the green fluorescent protein (GFP) or the red fluorescent protein (RFP) mCherry, with three TGT labeling sites located along the 5'-UTR, were synthesized following a previously reported mRNA transcription and maturation protocol.[72] Using TGT enzymatic labeling, the mature IVT-mRNA coding for GFP was conjugated with biotin-DEACM-preQ1 while the IVT-mRNA coding for RFP was conjugated with biotin-Bac-preQ1. To get rid of unlabeled

IVT-mRNAs, biotin-streptavidin affinity purification was performed (60% recovery). Equal amounts of photocaged IVT-mRNAs coding for GFP or RFP were co-transfected into HEK-293 cells using lipofectamine reagent (Figure 3.4A). mRNA translation activity was quantified by fluorescence imaging. Minimal translation activity was observed for the photocaged mRNAs, shown as dark cells (Figure 3.4B, Figure S7). Two hours post transfection, selected cells were irradiated with either 488 nm for 30 seconds or 405 nm wavelength of laser for 10 seconds to trigger the release of the 'photo-cages' from mRNA. [32] To allow time for sufficient mRNA translation and maturation, cells were imaged 8 hours after laser irradiation. As expected from our in-vitro studies, the 488 nm irradiation only triggered the release of the biotin-DEACM-preQ1 from mRNA. As a result, recovered protein expression of GFP was observed only in cells that were irradiated with the 488 nm laser (Figure 3.4B). Importantly, expression of RFP was not observed in these cells, which was expected because the biotin-Bac-preQ1 used to cage the RFP-mRNA is not responsive to 488 nm irradiation. In contrast, expression of both GFP and RFP was observed in cells that were irradiated with 405 nm wavelength of laser, because both biotin-Bac-preQ1 and biotin-DEACM-preQ1 have absorbance at 405 nm wavelength. These observations were consistent with our in-vitro uncaging experiments, that the longer wavelength of light only triggered the release of biotin-DEACM-preQ1, whereas the shorter wavelength of light triggered the release of both 'photocages' from mRNA. Moreover, expression of fluorescent protein was only observed in laser irradiated cells, not in adjacent cells, demonstrating photo-activation of gene expression with high cellular resolution.

Next, we demonstrated sequential photo-activation of two mRNAs in the same living cell. Caged IVT-mRNAs coding for GFP and RFP were co-transfected into HEK-293 cells. Two hours post transfection, the selected cell was first irradiated with 488 nm laser light to trigger the release of biotin-DEACM-preQ1 from GFP-mRNA. Five hours post transfection, the same cell was then irradiated with 405 nm laser light to trigger the release of the biotin-Bac-preQ1 from RFP-mRNA. Cells were continuously imaged to quantify protein expression level (Figure 3.5A, Figure 3.6).
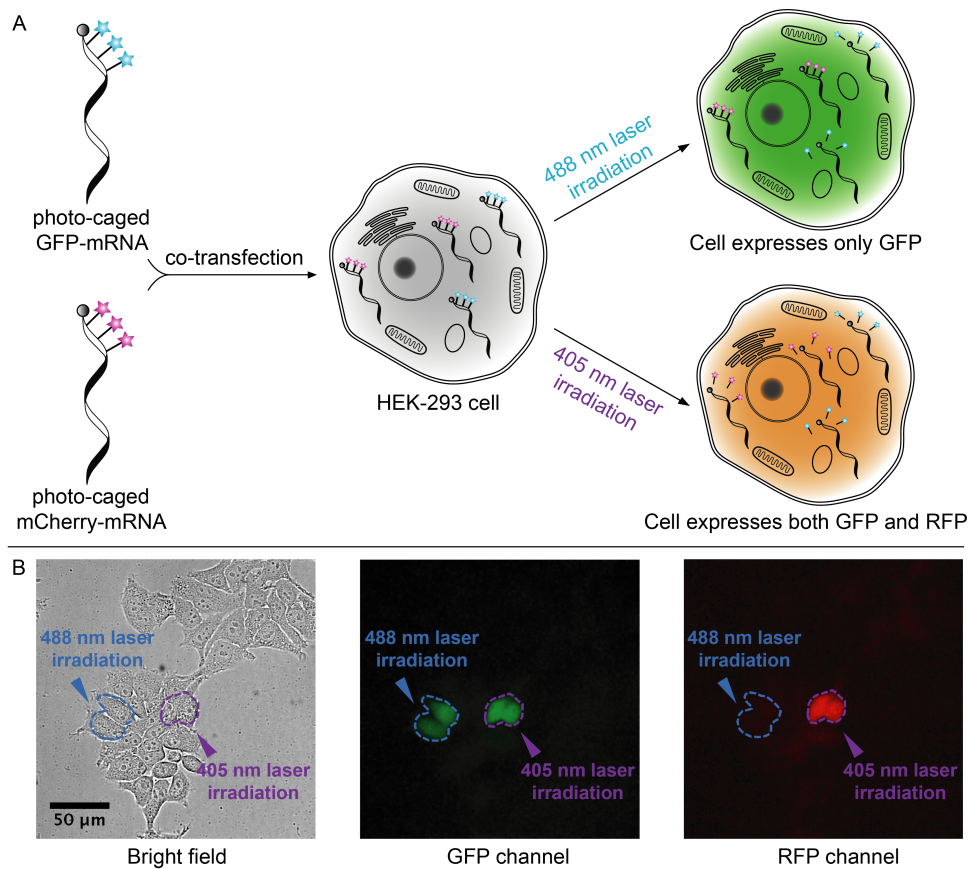
Figure 3.4: Live cell photo-activation of mRNA translation. A) HEK-293 cells are co-transfected with caged GFP-mRNA and caged mCherry-mRNA, followed by photo-uncaging with either 488 nm (30 seconds) or 405 nm laser (10 seconds) irradiation. B) Live cell fluorescence imaging. Selected cells (circled in blue) irradiated with 488 nm laser only express GFP, shown as green cells in the GFP channel and dark cells in the RFP channel. Selected cells (circled in purple) irradiated with 405 nm light express both GFP and RFP. Scale bar = 50 μm.

As shown in the fluorescence images, the first irradiation with 488 nm laser light triggered the expression of GFP, while the expression of RFP remained silenced. As expected, the subsequent irradiation with 405 nm laser light activated the expression of RFP. GFP and RFP expression levels in laser irradiated cells were quantified by measuring relative fluorescence intensity (Figure 3.5B). Using this model, we demonstrated sequential photo-activation of two mRNAs in the same cell by using two wavelengths of lights.
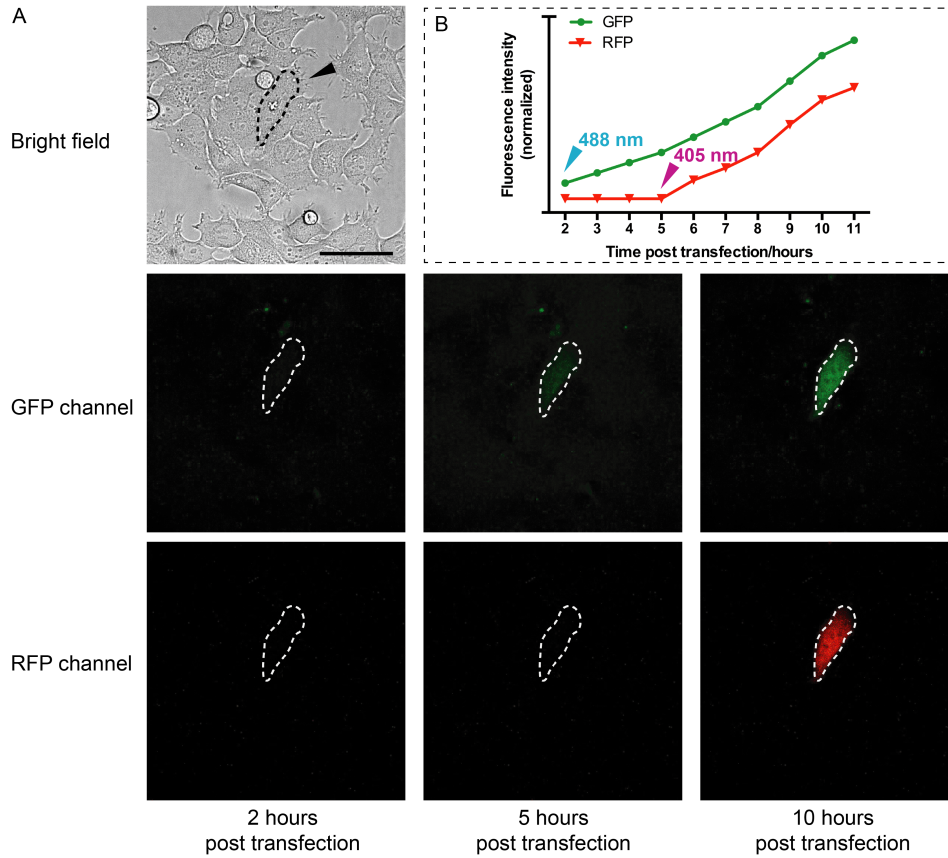
Figure 3.5: Live cell sequential photo-activation of two mRNAs within the same cell. A) 2 hours post transfection, the selected cell (circled in dash line) was first irradiated with 488 nm laser light to activate GFP-mRNA. 5 hours post transfection, the same cell was irradiated with 405 nm laser light to activate RFP-mRNA (Scale bar = 100 μm). 3 more replicates were shown in Figure S9. B) Relative fluorescence unit (RFU) of GFP and RFP from the selected cells was measured using Fiji software as mean gray value. Average fluorescence intensities from 3 replicates were plotted against time. Error bars represent standard deviation calculated from 3 replicates.
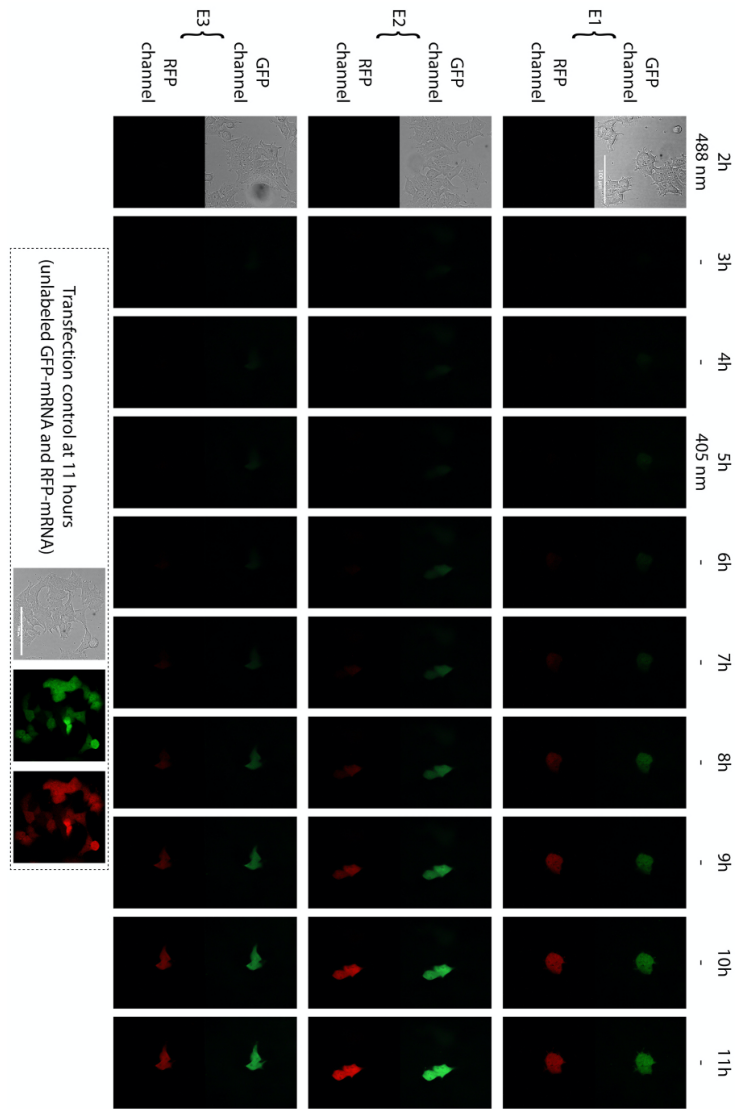
Figure 3.6: Caged IVT-mRNAs coding for GFP and RFP were co-transfected into HEK-293 cells. Two hours post transfection, the selected cell was first irradiated with 488 nm laser light to trigger the release of biotin-DEACM-preQ1 from GFP-mRNA. Five hours post transfection, the same cell was then irradiated with 405 nm laser light to trigger the release of the biotin-Bac-preQ1 from RFP-mRNA. Cells were continuously imaged to quantify protein expression level. As control, cells transfected with unlabeled mRNA coding for GFP and RFP were images at the end time point (11 hours post transfection). Fluorescence intensity was measured and averaged. The average relative fluorescence intensity of GFP from laser irradiated cells was 78.2% comparing to cells transfected with unlabeled GFP-mRNA. The average relative fluorescence intensity of RFP from laser irradiated cells was 82.5% comparing to cells transfected with unlabeled RFP-mRNA.

61

3.3 Conclusions

In conclusion, we have developed a technique that allows sequential photo-activation of two mRNAs with single-cell resolution. We demonstrated the synthesis and photochemical properties of two 'photo-cages', biotin-Bac-preQ1 (1) and biotin-DEACM-preQ1 (8). These 'photocages' were covalently and site-specifically conjugated onto the 5'-UTR of mRNA through TGT enzymatic labeling. As a result, translation efficiency of the labeled mRNA was severely diminished compared to the unlabeled mRNA. This pair of 'photo-cages' can be released from mRNA transcripts sequentially upon irradiation with 365 nm/405 nm light (lamp excitation) or 405 nm/488 nm light (laser excitation), leading to translational activation of the corresponding mRNA. Irradiation with a longer wavelength of light only cleaves one 'photo-cage' (biotin-DEACM-preQ1) from RNA, while a shorter wavelength of light cleaves both 'photo-cages'. By using the appropriate order of photo-irradiation, sequential photo-activation of two mRNAs within the same cell was demonstrated by live cell fluorescence imaging. We believe that the ability to sequentially photo-activate two genes with high spatial-temporal resolution provides a powerful and versatile optogenetic tool to build robust, complex, and scalable synthetic gene networks. Such a tool may improve capabilities to precisely manipulate biological networks, which can aid studies of gene regulatory mechanisms, promote the engineering of artificial biological systems, and facilitate the development of novel therapeutic applications.

3.4 Methods and Materials

Some of the methods, such as in vitro transcription, TGT labeling and purification, etc. have been covered in Chapter 2 - Methods and Materials section. In this section, new methods used in Chapter 3 will be covered.

3.4.1 Chemical Synthesis

Synthesis of compound (3)

Figure 3.7: Synthesis of compound (3) 4-(4-azido-1-((tert-butyldimethylsilyl)oxy)butyl)-7-(diethylamino)-2H-chromen-2-one

A solution of the previously reported 4-(1-((tert-butyldimethylsilyl)oxy)-4-hydroxybutyl)-7-(diethylamino)-2H-chromen-2-one (2) (50.0 mg, 0.12 mmol) in DCM (5 mL) was treated with DIEA (29.8 mg, 0.24 mmol) and MsCl (20.6 mg, 0.18 mmol).[44,73] Then the reaction mixture was stirred for 12 hours at room temperature. Afterwards, the reaction solution was quenched with water in ice bath and extracted with 5 mL DCM for 3 times. The combined organic solution was washed with brine and dried with Na_2SO_4. All organic solvent was removed in vacuo. The resulting product was dissolved in DMF (3 mL). Next, NaN_3 (15 mg, 0.24 mmol) was added to the solution. The reaction solution was heated at 80°C for 5 hours. After cooling the reaction to room temperature, water (10 mL) was added to quench the reaction, followed by EtOAc (5mL*6) extraction. Then the organic solution was dried with Na_2SO_4 and removed in vacuo. The resulting crude material was then purified by flash chromatography (0-10% EtOAc in hexanes) to afford compound (3) as a yellow solid (43.2 mg, 81%) (Figure 3.7). ^{1}H NMR (500 MHz, $CDCl_3$, δ): 7.45 (d, J = 9.0 Hz, 1H), 6.57 (d, J = 9.1 Hz, 1H), 6.52 (d, J = 2.5 Hz, 1H), 6.18 (s, 1H), 4.93-4.88 (m, 1H), 3.45-3.37 (m, 4H), 3.37-3.23 (m, 2H), 1.92-1.74 (m, 2H), 1.74-1.66 (m, 2H), 1.21 (t, J = 7.1 Hz, 6H), 0.92 (s, 9H), 0.09 (s, 3H), –0.03 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$, δ): 162.47, 158.10, 156.56, 150.27, 125.04, 108.35, 105.87, 105.87, 97.90, 70.50, 51.27, 44.68, 44.68, 35.17, 25.77, 25.77, 25.77, 24.69, 18.15, 12.44, 12.44, – 4.70, –5.15. HRMS (M[+]H[+]) calcd for $[C_{23}H_{37}N_4O_3Si]^+$ 445.2629, found 445.2633.
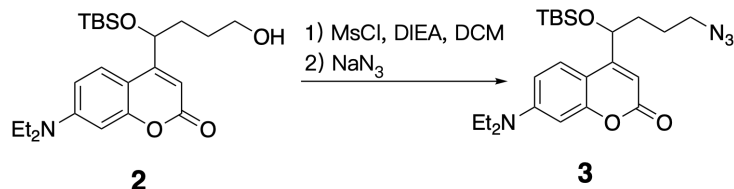
Synthesis of compound (4)
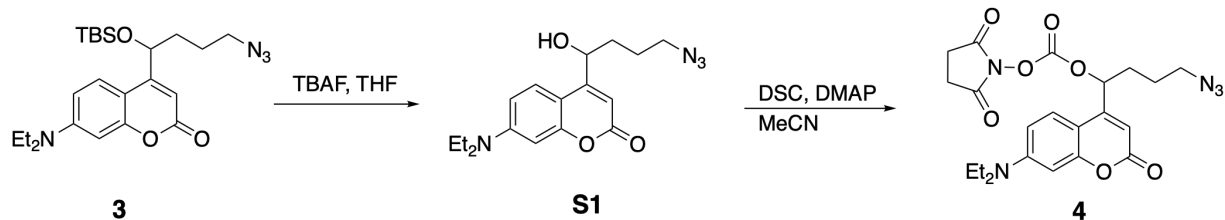
Figure 3.8: Synthesis of compound (4).

4-(1-((tert-butyldimethylsilyl)oxy)-4-hydroxybutyl)-7-(diethylamino)-2H-chromen-2-one (S1)

A solution of compound (3) (20 mg, 0.045 mmol) in THF (5 mL) was treated with TBAF solution (1 M, 1.35 mmol). The resulting solution was stirred at room temperature for 6 hours. After the deprotection, the solution was directly subjected to semipreparative HPLC purification, using a C18 column [gradient of H_2O with 0.1% formic acid and MeOH with 0.1% formic acid 95:5 (0 min) to 5:95 (10 min to 18min)]. Note: We tried to purify the compound by silica flash chromatography, which leads to complete decomposition of compound (S1). The fractions containing product compound (S1) from semipreparative HPLC purification were dried in vacuo at room temperature on rotary evaporator. The resulting product was further dried in vacuo with mechanical pump for 1 hour at room temperature. Compound (S1) was obtained as yellow oil (18.1 mg, 96%). ^{1}H NMR (500 MHz, CDCl$_3$, δ): 7.34 (d, J = 8.9 Hz, 1H), 6.54 (d, J = 8.8 Hz, 1H), 6.43 (s, 1H), 6.20 (s, 1H), 4.97-4.91 (m, 1H), 3.41 (s, 1H), 3.36-3.27 (m, 4H), 1.88 (m, 1H), 1.82-1.64 (m, 3H), 1.13 (t, J = 7.1 Hz, 6H). ^{13}C NMR (126 MHz, CDCl$_3$, δ): 162.88, 158.64, 156.39, 150.26, 125.25, 109.17, 105.91, 105.26, 98.20, 77.35, 69.33, 51.24, 44.95, 33.84, 25.11, 12.51, 12.36. HRMS (M$^+$Na$^+$) calcd for [C$_{17}$H$_{22}$N$_4$O$_3$Na]$^+$ 353.1384, found 353.1586.

4-azido-1-(7-(diethylamino)-2-oxo-2H-chromen-4-yl)butyl (2,5-dioxopyrrolidin-1-yl) carbonate (4)

Compound (S1) (18.1 mg, 0.043 mmol) was dissolved in MeCN, followed by addition of

DSC (22.0 mg, 0.086 mmol) and DMAP (15.9 mg, 0.13 mmol). The reaction solution was stirred for 12 hours at room temperature, followed by semipreparative HPLC purification, using a C18 column [gradient of H_2O with 0.1% formic acid and MeOH with 0.1% formic acid 95:5 (0 min) to 5:95 (10 min to 18 min)]. The fractions containing product (4) from semipreparative HPLC purification were dried in vacuo on rotary evaporator without heat. The resulting product (9.6 mg, 50%) was further dried in vacuo with mechanical pump for 1 hour at room temperature. Note: After dried in in vacuo at room temperature, there were small amount of unknown by-product generated. Thus, the NMR spectrums of compound (4) were not obtained (Figure 3.8). However, compound (4) is confirmed by low resolution mass spectrometry, using fresh fraction from semipreparative HPLC. LRMS (M^+H^+) calcd for $[C_{22}H_{26}N_5O_7]^+$ 472.2, found 472.1.

Synthesis of compound (8)
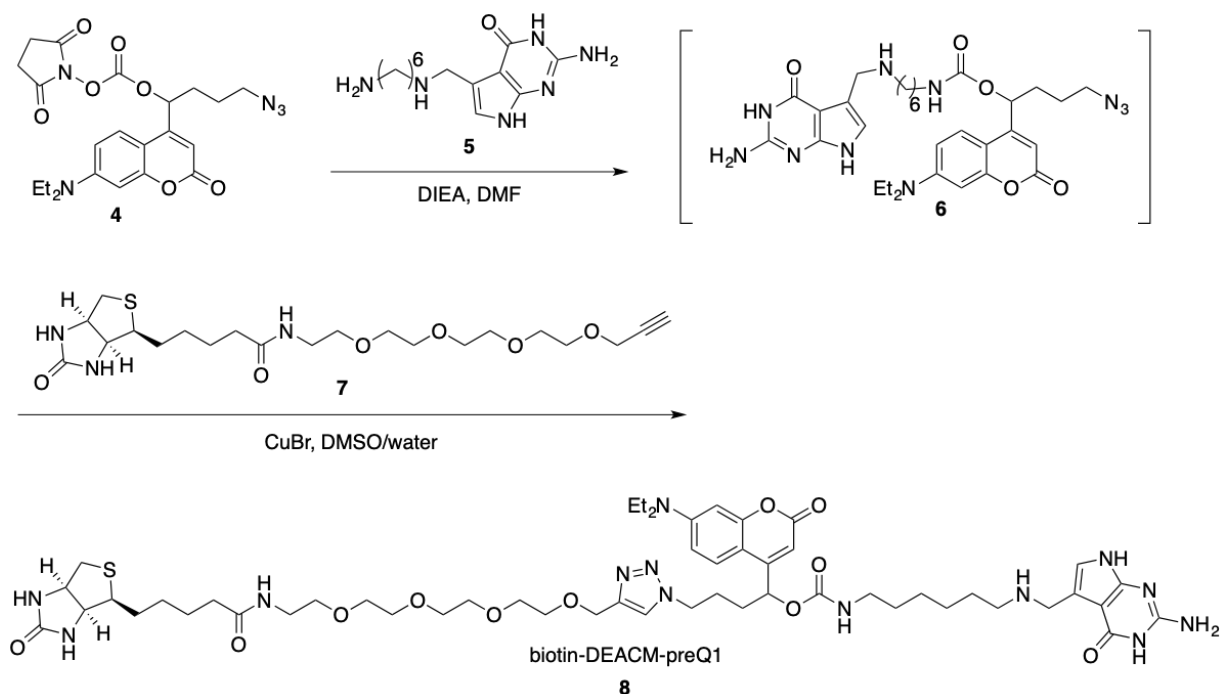
Figure 3.9: Synthesis of compound (8).

Compound (4) (5.0 mg, 0.011 mmol) was dissolved in DMF (0.5 mL) followed by slow addition of the previously reported compound (5) (8.8 mg, 0.032 mmol in 0.5 mL DMF) and DIEA (7.1 mg, 0.055 mmol).[72] The reaction solution was stirred for 1 hour at room temperature. The crude substitution product was directly subjected to semipreparative HPLC purification, using a C18 column [gradient of H_2O with 0.1% formic acid and MeOH with 0.1% formic acid 95:5 (0 min) to 5:95 (10 min to 18min)]. Shielded from light, the semipreparative HPLC fractions containing product compound (6) (confirmed by low resolution mass spectrometry, LRMS (M^+H^+) calcd for $[C_{31}H_{43}N_{10}O_5]^+$ 635.3, found 635.3) were dried in vacuo with ice bath. The resulting product was directly used for the next step of synthesis. Protected from light, CuBr (0.8 mg, 0.0055 mmol in DMSO:H_2O=0.3 ml:0.3 mL solution) was added to compound (6) followed by the addition of commercial available compound (7) (5.0 mg, 0.011 mmol in 0.4 mL DMSO; commercially available at BroadPharm, San Diego, USA; CAS# 1458576-00-5). The reaction mixture was stirred for 2 hours at room temperature, followed by semipreparative HPLC purification, using a C18 column [gradient of H_2O with 0.1% formic acid and MeOH with 0.1% formic acid 95:5 (0 min) to 5:95 (10 min to 18min)]. Protected from light, the fractions containing compound (8) from semipreparative HPLC were dried in vacuo at room temperature. After further vacuum drying in ice bath, there was a trace amount of water left in the vial due to the low temperature. To completely dry the final product compound (8), a solvent mixture of H_2O:MeCN=0.3 mL:0.3 mL was added to the vial containing compound (8), followed by lyophilization. Compound (8) was obtained as white residue (0.6 mg, 5% for 2 steps). HRMS (M^+H^+) calcd for $[C_{52}H_{78}N_{13}O_{11}S]^+$ 1092.5659, found 1092.5664.

TGT enzymatic labeling	-	+	+
biotin-Bac-preQ1	-	+	-
biotin-DEACM-preQ1	-	-	+
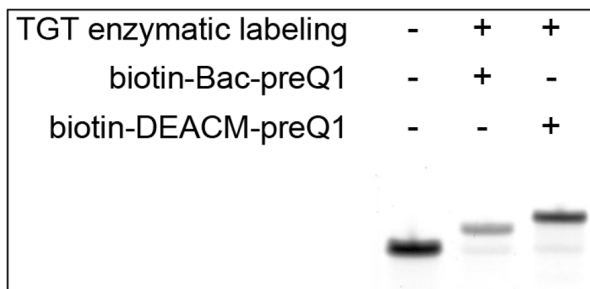
Figure 3.10: 18% denaturing PAGE analysis of in vitro TGT labeling of the 17-nt RNA oligo without affinity purification. Gel was stained with 1X GelRed.

3.4.2   In vitro TGT Labeling and Purification

TGT labeling of 17-nt RNA oligos

TGT labeling reaction was assembled with the following components in 1x TGT reaction buffer: 5 μM of 17-nt RNA oligo, 5 μM of TGT enzyme, 50 μM of small-molecule substrate (biotin-Bac-preQ1 or biotin-DEACM-preQ1), and 5 mM DTT. The reaction mixture was incubated at 37 °C for 4 hours. 1 μL of proteinase K was added into reaction mixture and incubated at 37°C for 30 minutes to terminate the labeling reaction. The crude labeled RNA was purified by EtOH precipitation. Crude labeling product was analyzed on 18% denaturing PAGE (Figure 3.2). By measuring RNA band intensity, the labeling efficiency of RNA oligo using biotin-Bac-preQ1 was about 93.6% and the labeling efficiency of RNA oligo using biotin-DEACM-preQ1 was about 96.7%. After biotin-streptavidin affinity purification, the unlabeled RNA oligo can be completely get rid of (Figure 3.3). Note that sense biotin-Bac-preQ1 and biotin-DEACM-perQ1 are extremely photo-sensitive, all reaction and analysis show be performed in a dark room with minimal red lamp as ambient light. All materials should be handled carefully to protect from photo-degradation.

Electrophoresis of Biotin-labeled mRNA

An mRNA with one active Tag sequence located at the 5'-UTR of the mRNA was labeled with biotin through TGT enzymatic labeling. The labeled mRNA was further purified by biotin-streptavidin affinity purification. 100 ng biotin-labeled and purified mRNA or control mRNA without TGT labeling was incubated with 1 μM streptavidin in 10 μL solution at room temperature for 15 minutes. The samples were then analyzed on 1% GelRed pre-stained agarose gel under 90 V for 45 minutes (Figure 3.11).
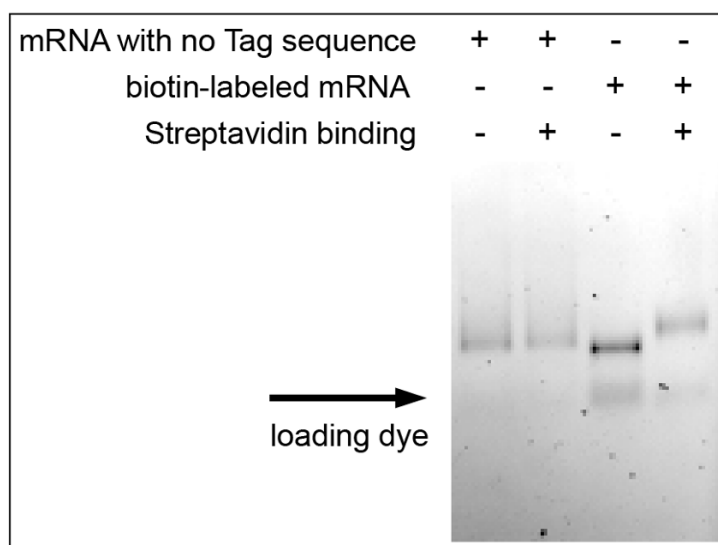
Figure 3.11: Agarose gel analysis of biotin-labeled mRNA. Biotinylation of mRNA by TGT labeling. 1% agarose Gel-shift analysis was performed to demonstrate successful incorporation of biotin at the 5'-UTR of the mRNA (as shown in line 4).

Degree of Labeling of biotin-DEACM-preQ1 on IVT-mRNA Transcript

In our previous paper,[72] we performed mRNA labeling efficiency test using the biotin-Bac-preQ1 analog. We performed the following experiment to determine the labeling efficiency of mRNA using biotin-DEACM-preQ1 analog. To quantify the labeling efficiency of biotin-DEACM-preQ1 onto mRNA, we performed a fluorescence biotin quantitation assay (Thermo Scientific, Waltham, MA). This commercially available microplate-based assay measures the

fluorescence of the Thermo Scientific DyLight Reporter (fluorescent avidin and HABA premix) upon binding with a biotinylated sample, in this case, the TGT labeled biotinylated RNA. The avidin fluoresces when the weakly interacting HABA (4'-hydroxyazobenzene-2-carboxylic acid) is displaced by the biotinylated RNA. The amount of biotin for each sample can be determined by comparing the fluorescence intensity of the sample with a purified and signally biotinylated mRNA reference (one labeled biotin per mRNA molecule, Figure 3.11). As shown in Figure 3.12, there is an average number of 2.12 biotins per mRNA transcript for the labeled but unpurified mRNA with 3 Tags inserted at the 5'-UTR. After Dynabeads™ streptavidin purification, there is an average number of 2.32 biotins per mRNA transcript. After photo-irradiation with 456 nm light for 30 seconds, there is an average number of 0.04 biotins per mRNA transcript which is similar to unmodified mRNA control (0.037), demonstrating complete photo-cleavage of the biotin-DEACM-preQ1 analog from mRNA transcript using 456 nm LED light.
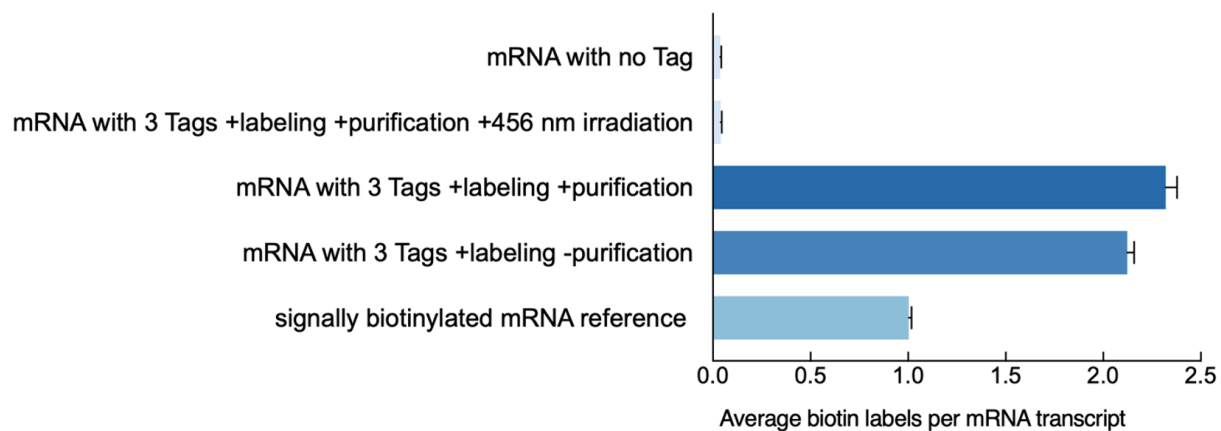
Figure 3.12: Degree of labeling of biotin-DEACM-preQ1 on IVT-mRNA transcript. The average number of TGT labeled biotins per mRNA transcript obtained from the fluorescence biotin quantitation assay. The signally biotinylated mRNA was used as normalization reference.

69

3.4.3 Live-cell Imaging and Photo-uncaging

mRNA Stability of the Unlabeled mRNAs and the Caged mRNAs

Same amount of unlabeled GFP-mRNA, biotin-Bac-preQ1 caged GFP-mRNA and biotin-DEACM-preQ1 caged GFP-mRNA were transfected into HEK-293 cells. 2 hours post transfection, transfection medium was removed and change to full growth medium (DMEM, 10% FBS). Total cellular mRNA was extracted 4 hours and 24 hours after transfection, following by reverse transcription and qPCR analysis of the GFP-mRNA (primers were shown below). Zymo Direct-zol RNA Purification Kits was used to extra total RNA from cell samples. Oligo-dT(18)-VN (purchased from IDT) was used as primer for reverse transcription (Maxima Reverse Transcriptase, EP0742, Thermo Scientific, TM). NEB-luna qPCR 2X master mix was used for qPCR analysis (Figure 3.14). qPCR primers were listed below. GFP-mRNA level was plotted as relative mRNA levels normalized to actin-mRNA. As a result, we didn't observe significant mRNA stability difference among these 3 mRNAs at 4 hours post transfection. After 24 hours, we observed slightly lower (about 20%) relative mRNA level of the caged mRNA compared to unlabeled mRNA. This difference might due to the fact that the caged mRNA went through extra in vitro labeling and purification steps, which may cause mRNA degradation, not necessarily due to the incorporation of the photo-cages. The degradation of mRNA after in vitro TGT labeling and biotin-streptavidin purification can be visualized on 4% denaturing PAGE (Figure 3.13).

GFP-mRNA forward primer: ACGTAAACGGCCACAAGTTC

GFP-mRNA reverse primer: AAGTCGTGCTGCTTCATGTG

Actin-mRNA forward primer: agagctacgagctgcctgac

Actin-mRNA reverse primer: ctccatgcccaggaaggaagg

IVT-mRNA with 3 Tags	+	+	+
TGT labeling with biotin-DEACM-preQ1	-	+	+
biotin-streptavidin affinity purification	-	-	+
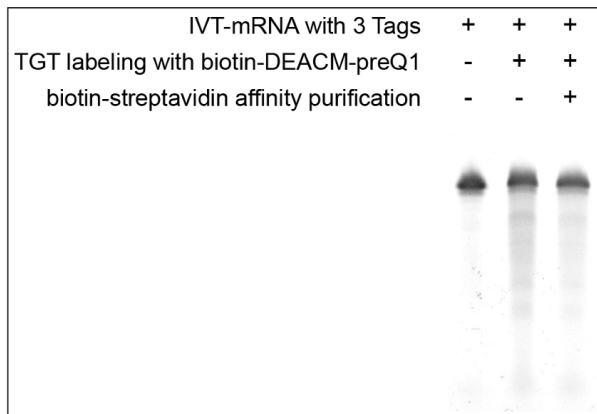
Figure 3.13: 4% denaturing PAGE analysis of IVT-mRNA labeling with biotin-DEACM-preQ1. Unlike labeling short RNA oligos, the molecular weight difference of labeled mRNA and unlabeled mRNA is too small to be visualized on 4% denaturing PAGE. Thus, in order to determine the labeling efficiency of biotin-DEACM-preQ1 onto mRNA transcript, we performed fluorescence biotin quantitation assay (Figure 3.11, Figure 3.12).

Translation Activity of the biotin-DEACM-preQ1 Caged GFP-mRNA

In order to estimate mRNA translation activity of biotin-DEACM-preQ1 caged GFP-mRNA and how much protein expression could be restored after light irradiation, we performed the following cell imaging and photo-uncaging experiments. Following mRNA transfection protocol described in previous section, we transfected the same amount (500 ng mRNA, 0.9 μL Lipofectamine RNAiMAX) of unlabeled mature GFP-mRNA, biotin-DEACM-preQ1 caged GFP-mRNA, biotin-DEACM-preQ1 caged and then in vitro 456 nm LED uncaged GFP-mRNA into cultured HEK-293 cells. Two hours post transfection, transfection medium was removed from cultured cells and the plate of cells that were transfected with biotin-DEACM-preQ1 caged mRNA were irradiated with 456 nm LED light for 30 seconds. Cells were imaged 8 hours after the light irradiation event (Figure 3.15). For fluorescent protein expression quantification, average fluorescence level (relative fluorescence units, RFU) was measured from more than 80 cells (Figure 3.15). As shown in Figure 3.15, the installment of biotin-DEACM-preQ1 reduced mRNA translation activity to about 12% relevant to the unlabeled mRNA. The in vitro photo-uncaged mRNA showed about 96% translation activity relevant to the unlabeled mRNA. The in vivo
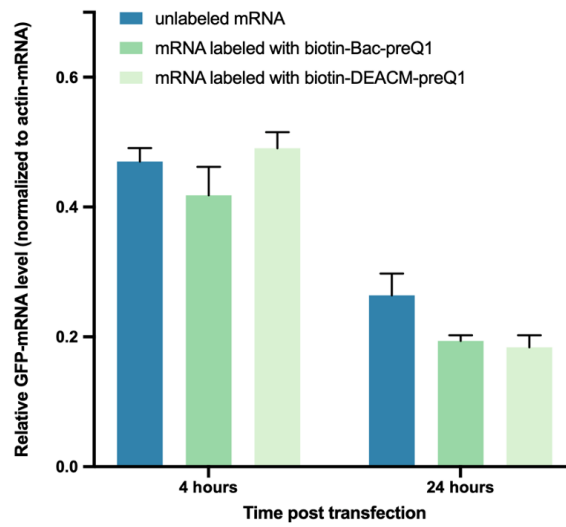
Figure 3.14: qPCR analysis of mRNA stability of unlabeled and caged mRNAs. GFP-mRNA level was normalized to actin-mRNA shown as relative GFP-mRNA level.

photo-uncaged mRNA showed about 76% translation activity relevant to the unlabeled mRNA.
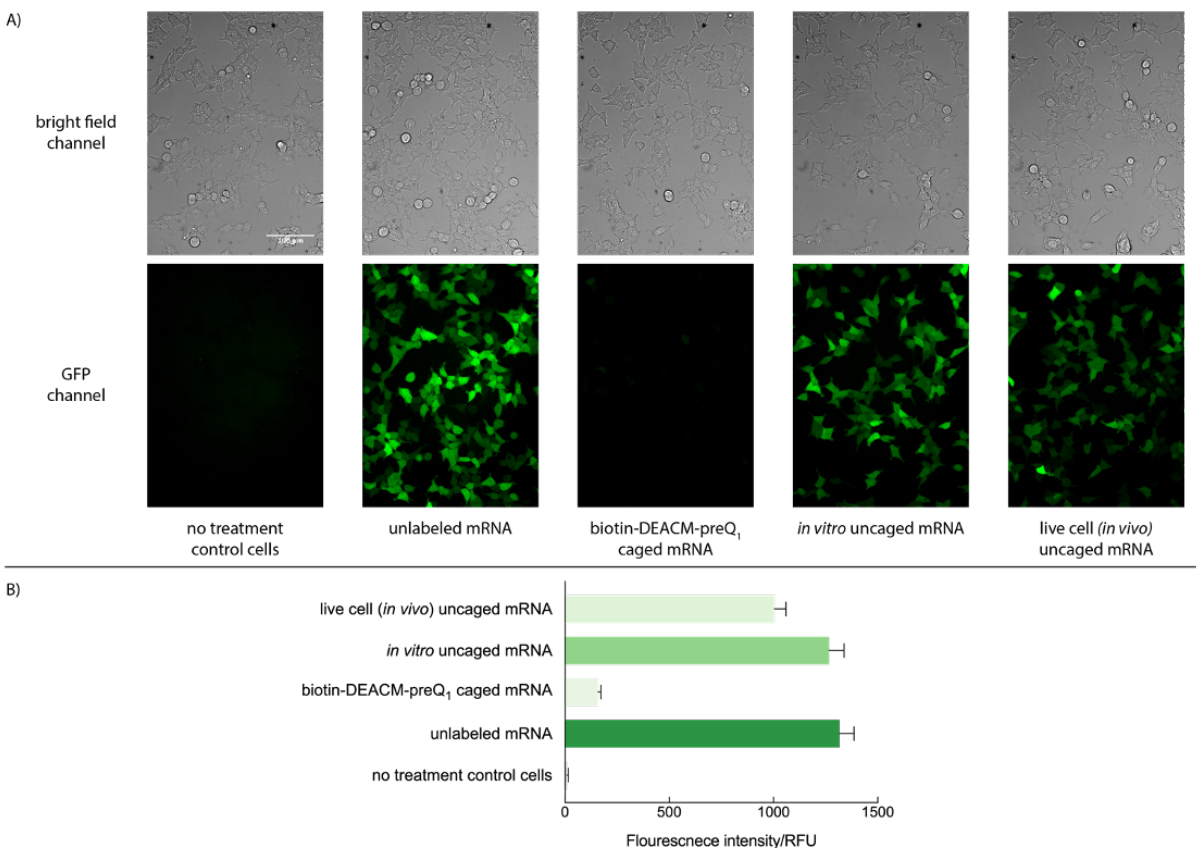
Figure 3.15: Cell imaging and protein expression quantification. A) Images of HEK-293 cells with different treatments. GFP fluorescence channels of all images were window leveled the same. Fluorescence intensity was relevant to GFP protein expression level. Scale bar = 100 μm. B) To obtain a quantitative estimation of GFP expression among differently treated cells, fluorescence intensity of individual cell in each cell image was measured using Fiji software. Fluorescence intensity of more than 80 cells were measured and averaged. The average relative fluorescence units (RFU) value with error bar (SEM) from each sample was plotted.

3.4.4   NMR Spectra and HRMS

3.5   Acknowledgements

Chapter 3, in full, is a reprint of the material as it appears in the publication: Zhang, D., Jin, S., Piao, X., & Devaraj, N. K. (2020). "Multiplexed photo-activation of mRNA with single-cell resolution." *ACS Chemical Biology*, 15(7), 1773-1779. The dissertation author was the co-primary investigator and author of this paper. Dr Shuaijiang Jin synthesized the preQ1-

DEACM-biotin enzymatic probe and developed the synthetic method to access the compounds presented in this chapter. I would like to thank Professor Neal Devaraj for directing this research and in the preparation of the manuscript.
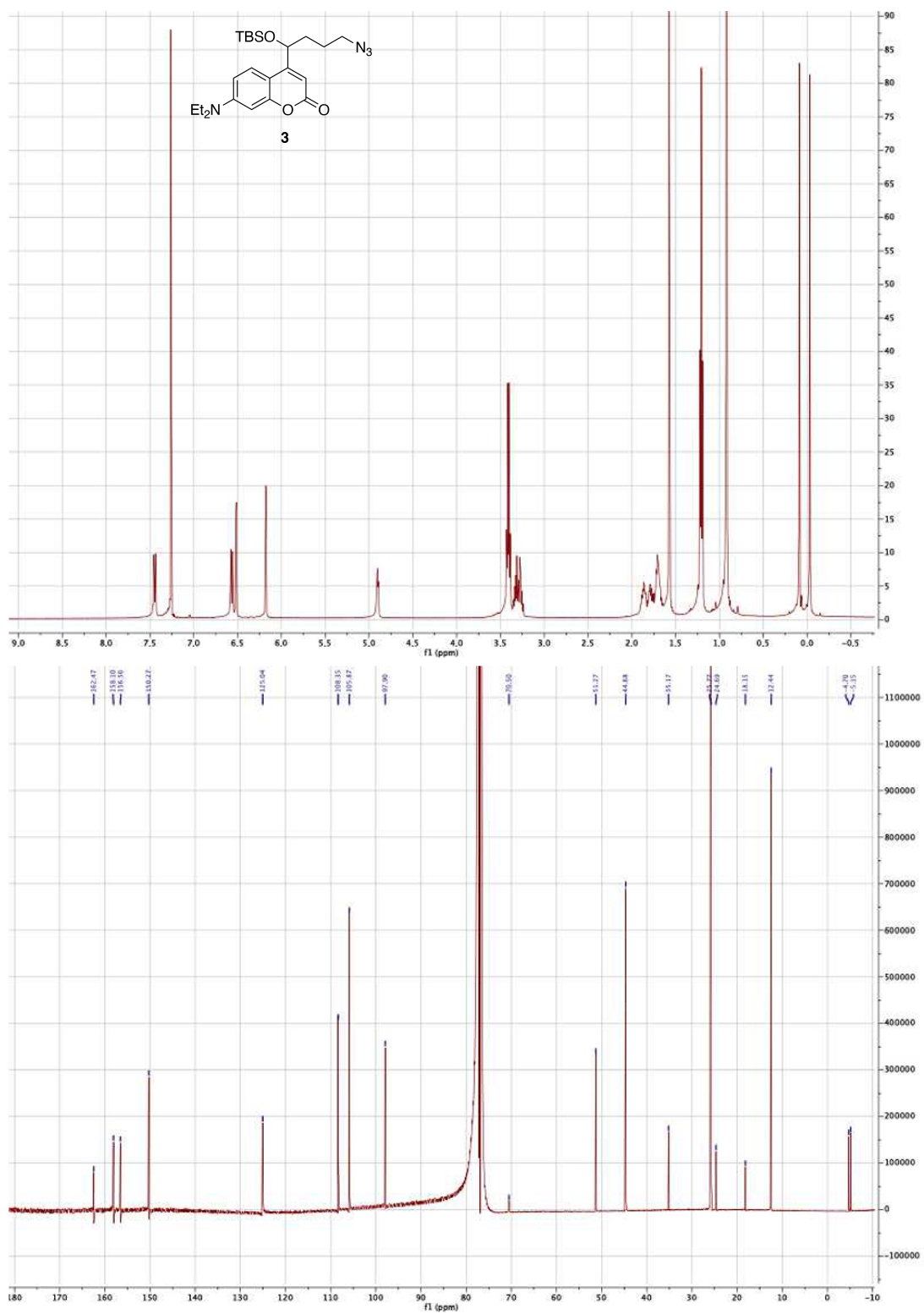
Figure 3.16: Chapter3 compound 3 NMRs.
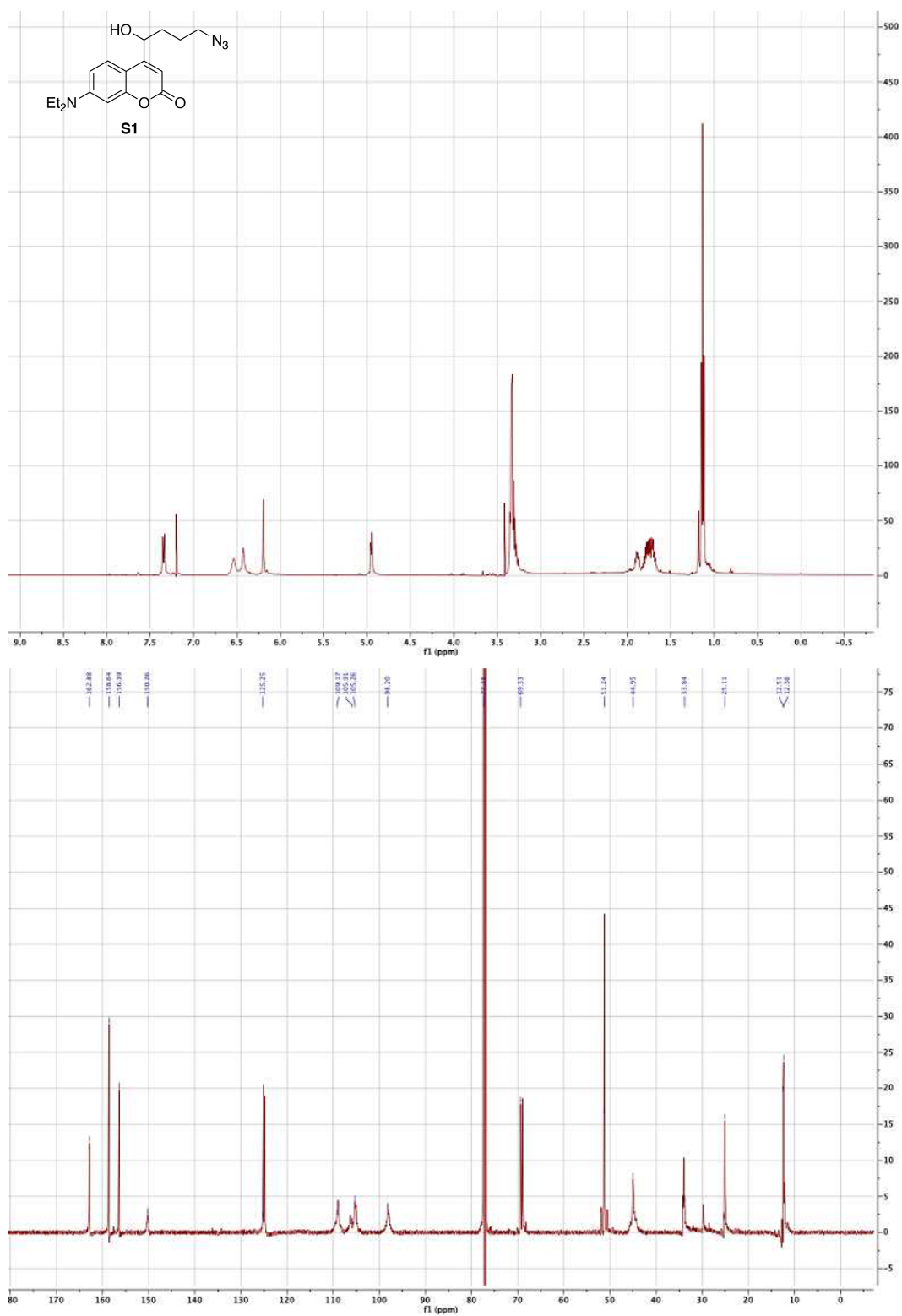
Figure 3.17: Chapter 3 compound S1 NMRs.
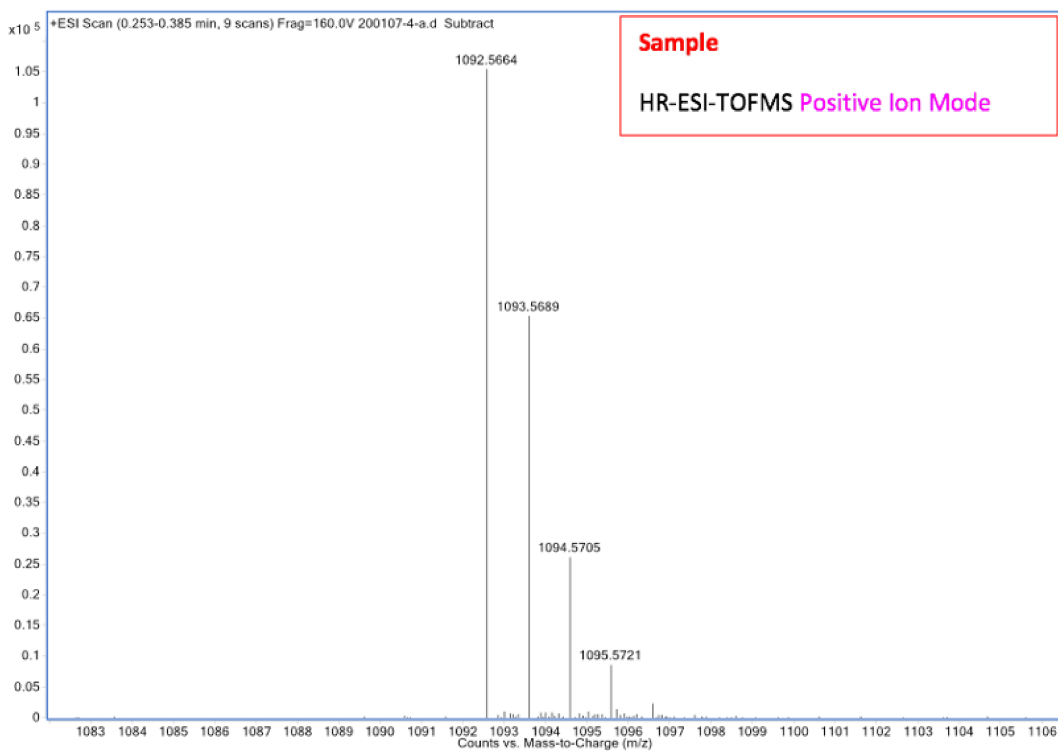
Mass Measured	Theo. Mass	Delta (ppm)	Composition
1092.5664	1092.5659	-0.5	$[C_{52} H_{78} N_{13} O_{11} S]^+$
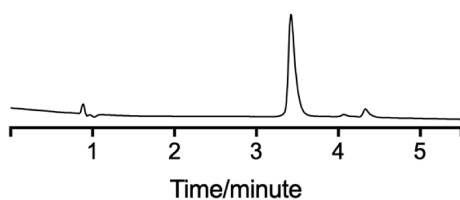
Figure 3.18: Chapter 3 compound 8 HRMS and HPLC trace. HPLC (405 nm absorbance) trace of compound (8). Analytical HPLC: tR = 3.48 min (Analytical HPLC condition 20 to 95% Phase B over 5 min, then 95% Phase B for 1 min, Eclipse Plus C8 analytical column. Phase A: H_2O with 0.1% formic acid; Phase B: MeOH with 0.1% formic acid.)

77

Chapter 4

Control the CRISPR/Cas9 Gene Editing by 'Clamping' the sgRNA

4.1 Introduction

Photo-regulation offers rapid and non-invasive manipulation of cellular processes. As we discussed in Chapter 1.3.2, there have been many approaches developed for the photo-regulation of CRISPR-Cas9 gene editing. Also, as we discussed in Chapter 1.4, most of the current methods suffer from major drop backs. Therefore, a method that enables photo-regulation of gene editing using visible light with high dynamic range will be highly desired. To achieve this goal, we need to have a completely different mind set regarding the strategies. Most current methods use bulky groups to hinder the binding between the Cas9 with sgRNA or the binding between sgRNA with target DNA, in order to block the gene editing process.[58,59,61] Because of the lack of robustness of such non-covalent interaction, it is challenging to achieve a complete inhibition of gene editing. Besides, different blocking molecules, such as anti-sense blocking oligo, have to by synthesized for each sgRNA. This process is expensive and time-consuming. The inhibition efficacy could also vary between targets, making such approaches less robust.

Fortunately, the crystal structure of dCas9 in complex with sgRNA and the target DNA has been well-studies, providing us tremendous insights into how this complex gene editing system works in the living system.[74]

The sgRNA is a chimeric version of the wild-type trans-acting RNA and crRNA hybrid (Figure 4.1). By inserting an artificial loop, the tetra-loop between the trans-acting RNA and the crRNA, the RNA hybrid can be stabilized, resulting easy-of-use and greater editing efficiency. In total, there are four stem-loops within the sgRNA, including the tetra-loop. As shown in the crystal structure (Figure 4.2), Cas9 consists of two lobes: a recognition (REC) lobe (shown in gray and light gray) and a nuclease (NUC) lobe (shown in pink) (Figure 4.2). The REC lobe can be divided into three regions, a long a helix referred to as the bridge helix (residues 60–93), the REC1 (residues 94–179 and 308–713) domain, and the REC2 (residues 180–307) domain. This lobe is primarily interacting with the repeat:anti-Repeat duplex of the sgRNA. The NUC lobe consists of the RuvC (residues 1–59, 718–769, and 909–1098), HNH (residues 775–908), and PAM-interacting (PI) (residues 1099–1368) domains. This lobe is responsible for cleaving the target DNA under proper conformation. And the negatively charged sgRNA:target DNA heteroduplex is placed in the positively charged groove at the interface between the REC and NUC lobes.

We can also divide the sgRNA into two parts, the REC-interacting sequences and the RuvC interacting sequences. These two parts are connected flexible linker (UUAUC), which is shown in purple in Figure 4.1. The guide:target and repeat:anti-repeat duplexes are deeply buried in a positively charged groove at the interface of the two lobes, whereas the rest of the sgRNA extensively interacts with the positively charged surface on the back side of the protein. It is also found that the sgRNA:DNA Complex adopts a T-Shaped architecture, which is critical for the Cas9 to be in a functional conformation to perform the DNA cleavage. Based on the crystal structure, the researchers also hypothesized that the flexible of Cas9 and sgRNA is likely to play a role in the assembly of the Cas9-sgRNA-DNA ternary complex. In summary the CRISPR/Cas9
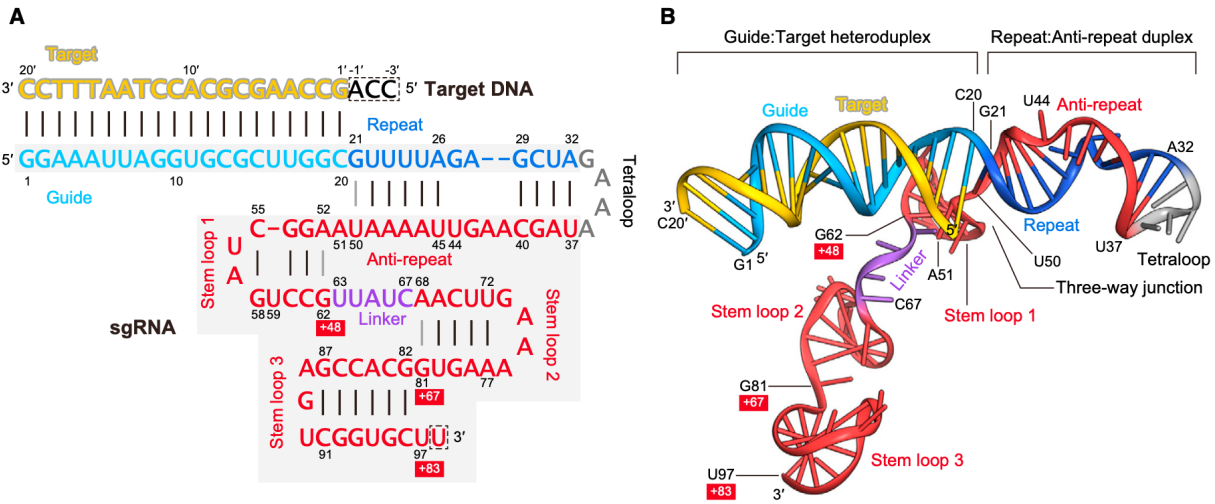
Figure 4.1: Structure of the sgRNA in CRISPR/Cas9 system.[74] The crystal structure of dCas9 in complex with sgRNA and target DNA has been solved.

system may cleave the target DNA following three steps (Figure 4.3).

[1] Cas9 alone is complete inactive and unable to cleave DNA. The REC and RuvC lobes are not in a functional conformation. The additional of the sgRNA stabilize the Cas9 protein by binding to the Bridge Helix. The stem-loop of the sgRNA also interact with different domains of the Cas9, reinforcing the interactions.

[2] The sgRNA then guided the Cas9-sgRNA complex to the genome target, where the PAM interacting domain of the Cas9 interacts the the PAM sequencing, forcing the Cas9 into a functional conformation.

[3] After the conformational activation, the DNA cleavage occurs at both the complementary and non-complementary strands.

Since the flexibility of the Cas9-sgRNA complex is critical to the function of the DNA cleavage, we may be able to structurally hinder this process by destroying this flexibility. RNA-TAG is a versatile and powerful RNA labeling tool. And sgRNA contains multiple stem-loop, which can be potentially labeled by TGT enzyme with sequence modifications. Thus, I decided to take the challenge to use the RNA-TAG technique to modify these stem-loop within the sgRNA sequence in order to hinder the Cas9 DNA cleavage activity. Since we were able to incorporate a
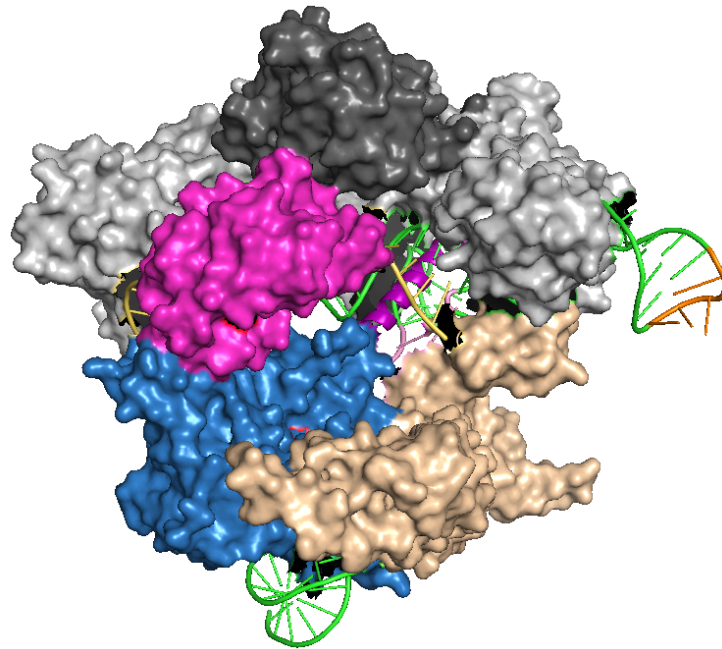
Figure 4.2: The crystal structure of dCas9 in complex with sgRNA and target DNA.[74]
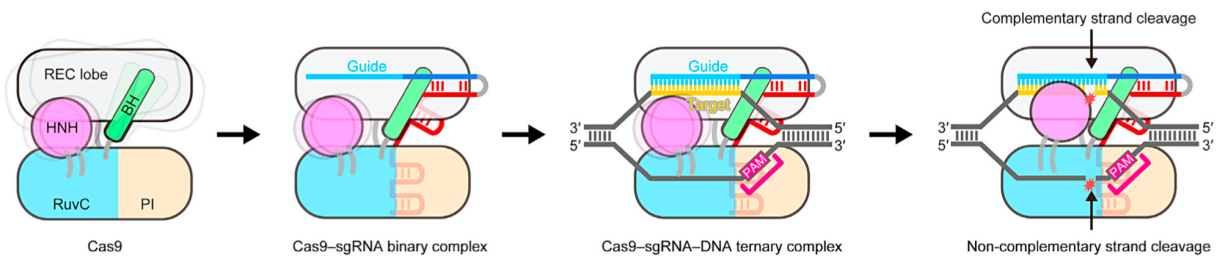
Figure 4.3: Proposed Model of RNA-guided DNA cleavage by Cas9.

photo-sensitive cage onto the RNA of interest, in this case, the sgRNA, we could use external light to release the 'photo-cage', therefore activating CRISPR/Cas9 gene editing upon light irradiation.

4.2 Results and Discussions

4.2.1 RNA-CLAMP: a Powerful Technique to Manipulate RNA

RNA Dimerization by RNA-TAG

Previously, we have demonstrated that our RNA-TAG technique was able to incorporate small-molecule functional groups onto the RNA of interest. However, we have never examined whether such approach can achieve RNA dimerization or cyclization. The dimerization/cyclization process is different than traditional RNA labeling using the RNA-TAG technique. For instance, we can think the RNA dimerization process as a two-step process (Figure 4.4). The RNA enzyme substrates are the 17-nt RNA oligos. The small-molecule substrate is the synthesized preQ1-PEG10-preQ1, bearing two preQ1 moieties to facilitate dual-modification. Thus, the first step of the dimerization process is the traditional TGT labeling process, which involves the incorporation of the preQ1-PEG-preQ1 on the RNA oligo. The product is a singly labeled RNA. For the second labeling, however, the traditional small-molecule substrate is now a conjugated RNA oligo, which is much bulkier than a small-molecule substrate. Fortunately, I discovered that the TGT enzyme is still able to accept this conjugated RNA as its substrates (Figure 4.5). In the 18% denaturing TBE-PAGE analysis, the bottom RNA bands represent the starting material, which is the 17-nt RNA oligo. The middle bands represent the singly labeled RNA oligo. The top bands represents the RNA dimer. By altering the RNA:preQ1-PEG10-preQ1 ratio, we also observed different conversions for the singly labeled product and dimer product. The results demonstrated that the TGT enzyme can accept a preQ1 moiety bearing a large biomacromolecule, in this case a RNA oligo, and incorporated it onto an RNA of interest.

RNA Cyclization by RNA-TAG

Having demonstrated that RNA oligos can be dimerized by TGT labeling, we next examined whether the RNA-TAG technique can be used for RNA cyclization. To facilitate
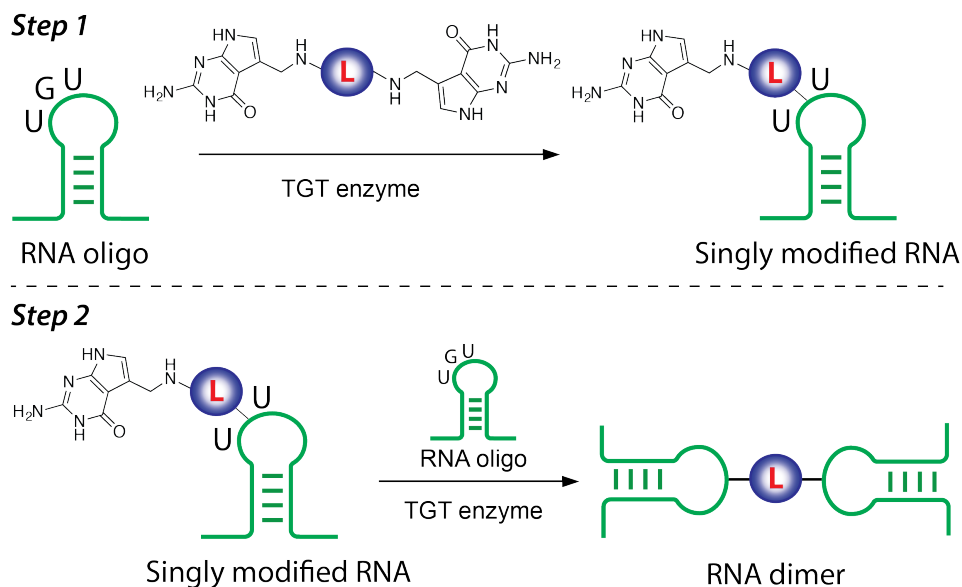
Figure 4.4: Two-step RNA oligo dimerization.

intramolecular cyclization, a RNA substrate needs to have two TGT enzymatic recognition sites. By using the small-molecule substrate, preQ1-PEG10-preQ1, a proposed RNA intramolecular cyclization reaction is shown in Figure 4.6. In this proposed reaction, two TGT enzymes bind to each RNA hairpin within the longer RNA. Next, the small-molecule substrate, preQ1-PEG10-preQ1 comes and connect the two RNA hairpins together, forming a open-ended circular RNA structure. It should be noted that unlike regular circular RNA, this 'clamped' RNA still have two free ends. Additionally, the sequence between these two hairpins should be flexible enough to enable the RNA 'clamping' reaction. As a proof of concept, we designed a 100-nt RNA molecule with two Tags located at the ends. To ensure the flexibility of this RNA molecule, we designed the RNA sequence using m-fold to avoid strong secondary structures between these to hairpins.[75–77] I performed the TGT labeling reaction using the preQ1-PEG10-preQ1 linker to test whether an internal 'clamping' reaction can occur. In Figure 4.7, the left band represents the RNA starting material, the right bands represent the labeling products. Unfortunately, the reaction was quite messy initially. I observed multiple products. Thus, it is critical to identify all the labeling product and purify the 'clamped' RNA.
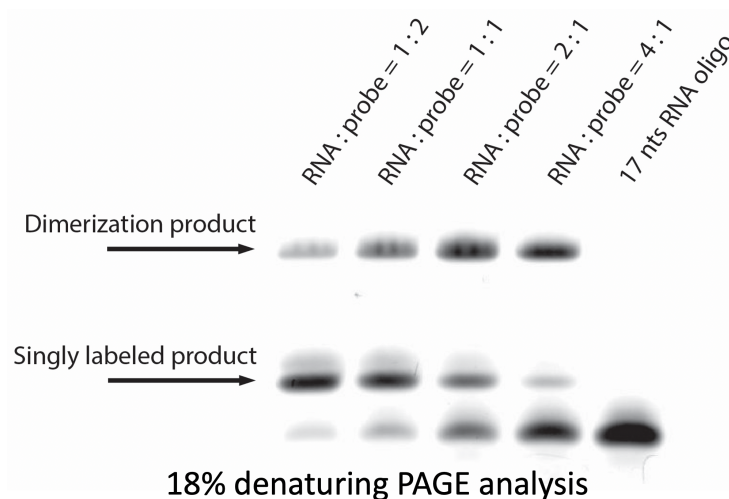
Figure 4.5: 10% denaturing PAGE analysis of RNA dimerization
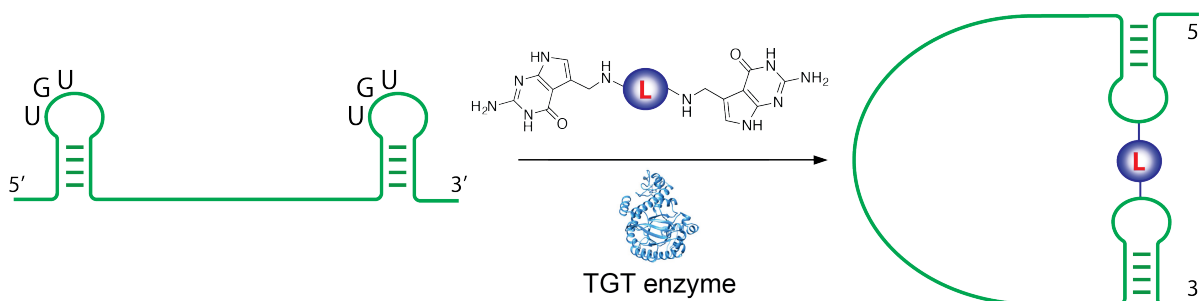
Figure 4.6: Proposed RNA 'clamping' reaction by using the RNA-TAG technique.

I hypothesized that there will be at lease three major types of products out of the RNA 'clamping' reaction (Figure 4.8).

1. Both of the two hairpins within the RNA molecule get singly labeled by the TGT enzyme, generating a linear RNA molecule.

2. As desired, the preQ1-PEG10-preQ1 connects two hairpins, forming a 'clamped' RNA molecule. It should be noted that this RNA molecule still have two free ends. Thus, this is not a typical circular RNA molecule.

3. Because each RNA substrate has two potential modification sites, the reaction can generate a polymer containing different numbers of RNA molecules (the starting material).

To test these hypotheses, I performed a Ribonuclease H (RNAse-H) digestion assay.
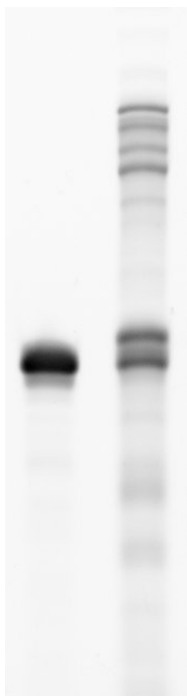
Figure 4.7: 7% denaturing PAGE analysis of the crude products out of RNA 'clamping' reaction by RNA-TAG.

RNAse-H is know to catalyze the cleavage of RNA in an RNA/DNA substrate via a hydrolytic mechanism.[78] By designing a DNA template which is complementary to the RNA sequence along the RNA molecule, we can use RNAse-H to selectively cleave the RNA molecule. Through the DNA-templated RNAse-H digestion assay, the linear RNA product (hypothesis 1) should be cleaved and generates two RNA fragments. The 'clamped' RNA products can also be cleaved. However, only one RNA fragment should be detected after the cleavage because the two ends of the 'clamped' RNA are covalently connected by a PEG10 linker. The cleavage of the polymer RNA construct should generate multiple RNA fragments as digestion products. This RNAse-H digestion model is shown in Figure 4.8). The denaturing PAGE analysis of the RNAse-H assay proves our hypothesis.

As shown in Figure 4.9, line 2 to line 3 are isolated products out of the crude RNA 'clamping' reaction. The linear RNA product (hypothesis 1) results in two RNA fragments out of the RNAse-H digestion assay as shown in line 7 and line 8. The polymer RNA products result
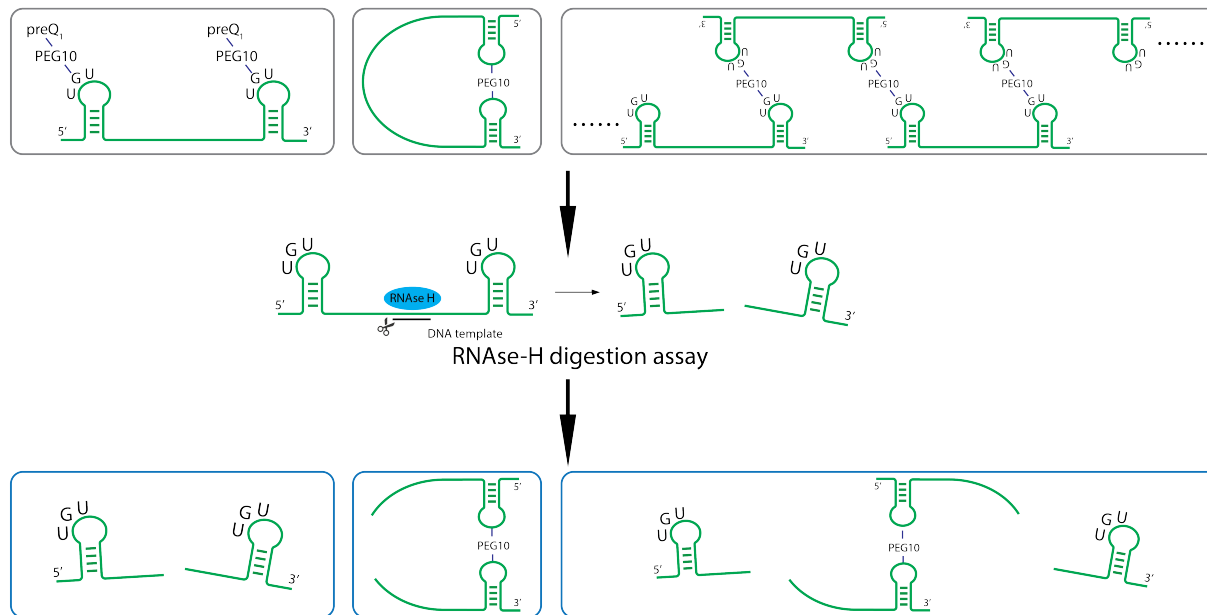
Figure 4.8: Ribonuclease H digestion assay to identify products out of the RNA 'clamping' reaction.

in multiple RNA fragments as shown in line 9 and line 10. Interesting, one RNA product only generates a single major band after the RNAse-H digestion assay (line 5 and line 6), indicating that the two ends of this RNA product are covalently cross-linked. Thus, we determined that this band represented the successfully 'clamped' RNA.

4.2.2 Screening of sgRNA Constructs for RNA-CLAMP

Having demonstrated that we can successfully dimerize or 'clamp' of a RNA of interest. We next thought how do we apply our technique to the sgRNA. As we discussed in previous section (Figure 4.1), the sgRNA contains four stem-loops and two free ends. All these positions can be potentially inserted with the 17-nt RNA Tag sequence. Previously, it has been shown that many of these positions can be altered without affecting the sgRNA activity.[74] Thus, I made a series of sgRNA with one of these positions mutated (Figure 4.10). As shown in Figure 4.10, one out of the six positions of the sgRNA is modified to allow for the insertion of the 17-nt Tag sequence in order to promote TGT labeling. The modified sequence is shown in orange
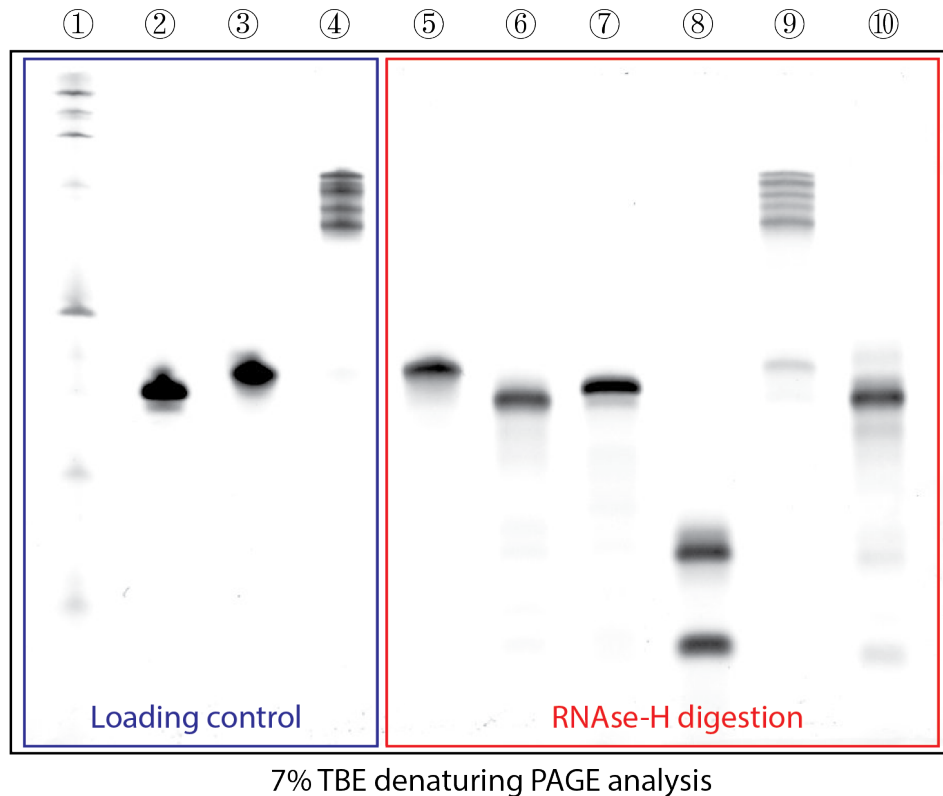
Figure 4.9: 7% denaturing PAGE analysis of the RNAse-H digestion assay.

color. First, we tested all the activities of the sequence modified sgRNAs (sgRNA1 to sgRNA6), compared to the wt-sgRNA. The genome locus DYRK1A is chosen as the sgRNA targeting site as it has been shown previously used to quantify gene editing efficiency.[79]

To obtain optimal gene editing efficiency, we examined different Cas9 delivery approaches, including transfection or electroporation of Cas9 vector or mRNA, electroporation of Cas9-sgRNA RNP complex and sequential transfection of Cas9-mRNA and sgRNA. The gene editing efficiency varies from 20% INDEL to 80% INDEL with different delivery strategies. In order to accurately and consistently quantify the activity of sgRNA, less experimental variables is preferred. Thus, we constructed an engineered mammalian single-colony HEK-293 cell line that stably expresses Cas9 enzyme using the lentiviral approach. To quantify the activity of sgRNA, we delivered sgRNA1 to sgRNA6 into the HEK-293-Cas9 cells using lipofectamine RNAiMAX. One day after transfection, culturing medium was exchanged with fresh full-growth medium (DMEM with

10% FBS). Three days after transfection, genomic DNA was extracted from the cells, following by PCR amplification of the sgRNA targeted region. The PCR amplified DNA fragments were further analyzed by either the T7E1 nuclease digestion assay or by sanger sequencing to quantify gene knockouts efficiency of the targeted region.[79, 80]
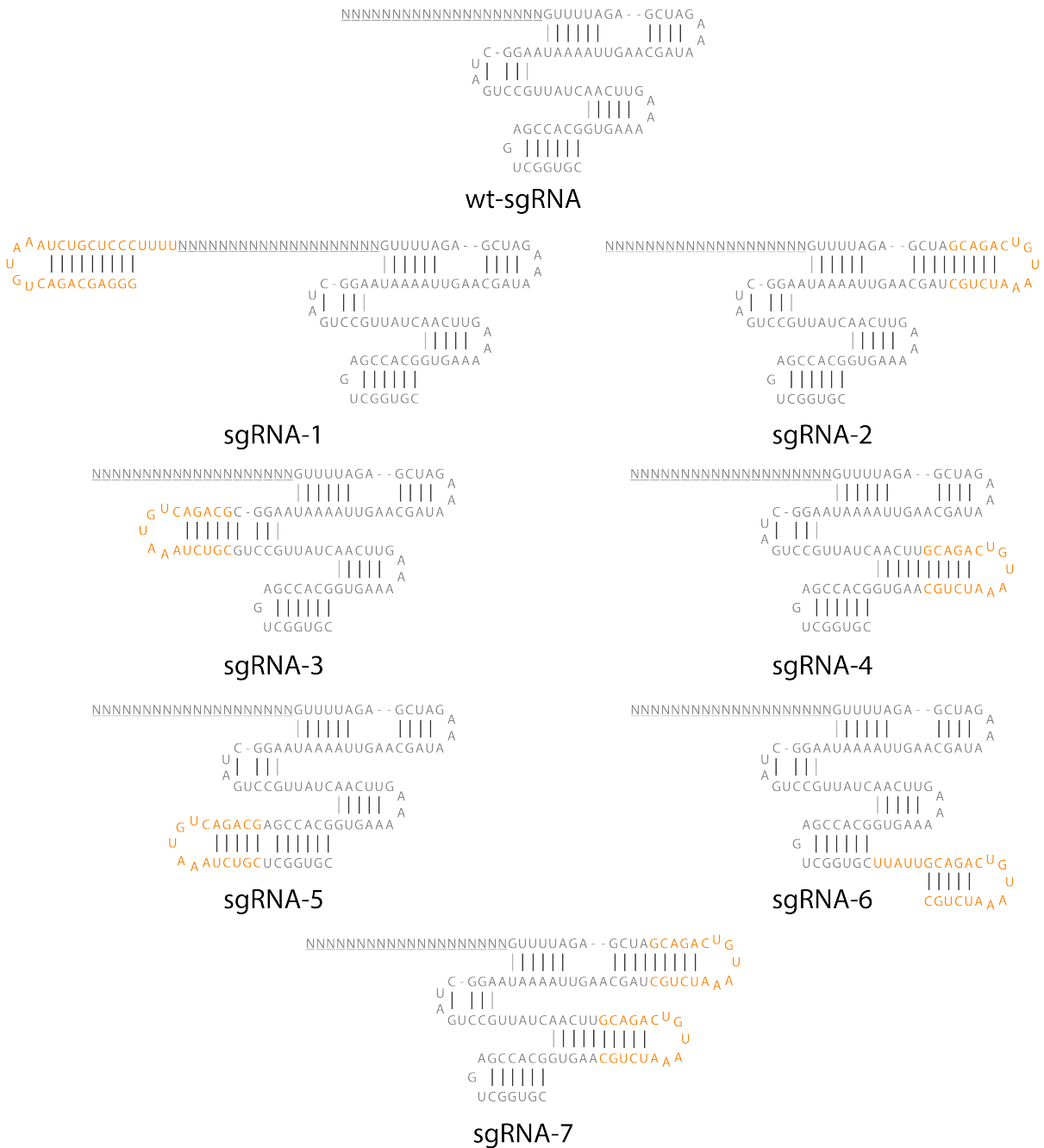
Figure 4.10: Sequences of sgRNA1 to sgRNA7.

As shown in Figure 4.11, sgRNA2 and sgRNA4 have wild-type level of gene knockout efficiency, which is in line with previous reports.[79] For sgRNA2, the tetra-loop of the sgRNA is modified. In the crystal structure, this tetra-loop is reaching out of the Cas9-sgRNA-DNA ternary
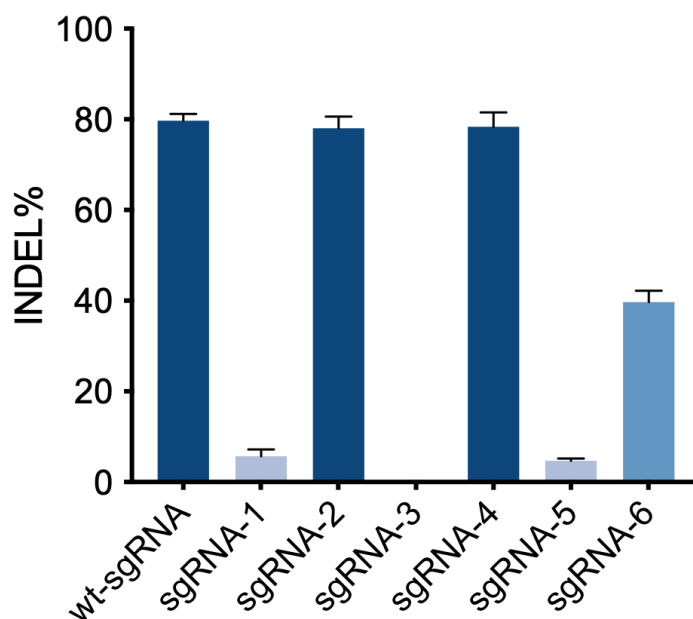
Figure 4.11: Activities of sgRNA1 to sgRNA6.

complex, indicating that modification at this position may have little or no effect on the activity of the Cas9 binding and DNA cleavage activity (Figure 4.1, Figure 4.2, Figure 4.12). Additionally, the sgRNA is a chimeric version of the trans-acting RNA:crRNA hybrid, suggesting that this tetra-loop has no structural function. Thus, we reasoned that this tetra-loop is acceptable to any sequence modifications. We also observed wild-type level activity of sgRNA4. The stem-loop 2 of the sgRNA is solvent accessible according to the crystal structure of the dCas9-sgRNA-DNA ternary complex (Figure 4.12). Although the stem part of the hairpin structure interacts with the Cas9 protein, the extension of this stem preserves these interactions. By swapping the original loop sequence of the stem-loop 2 (GAAA) with a 17-nt Tag sequence, the sgRNA4 still binds to the Cas9 and cleaves the target DNA. The stem-loop 1 of the sgRNA is critical to the formation of the three-way-junction, which is essential for the binding of sgRNA and Cas9 protein. Sequence modification at this position (sgRNA3) completely diminished Cas9 activity. While the 5' extension of the sgRNA may do not have affect on the Cas9-sgRNA binding, it may hinder the strand invasion of the sgRNA towards the DNA double helix structure of genome,
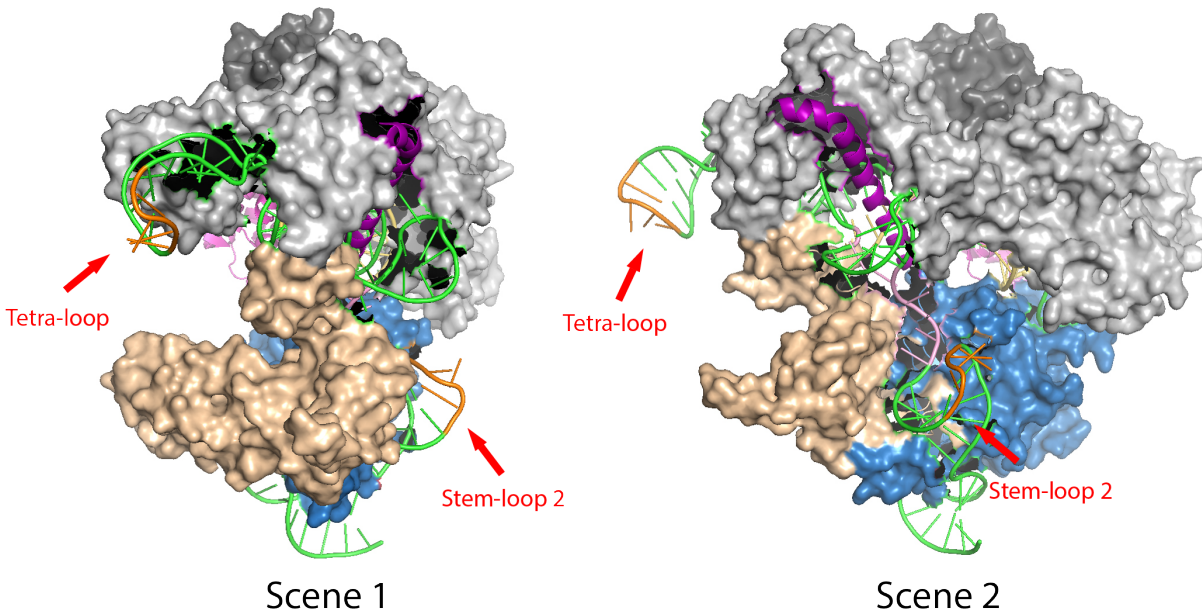
Figure 4.12: Tretra-loop and Stem-loop 2 of the Cas9. Side scene of the dCas9-sgRNA-DNA ternary complex with the tetra-loop and stem-loop 2 highlighted. Both of the two loops are solvent accessible.

causing significant decrease (95% decrease) of the Cas9 gene editing efficiency in vivo (Figure 4.11). Next, we determined whether a dual modification of the sgRNA at both the tetra-loop and the stem-loop 2 will affect its gene editing activity. Therefore, I constructed sgRNA7, which contains sequence modifications at two stem-loops. As expected, sgRNA7 also showed wild-type gene editing activity level.

In summary, by screening all the six potential TGT labeling sites along the sequence of sgRNA, we identified two locations, the tetra-loop and the stem-loop 2, which tolerant sequence modifications. As a result, we are able to replace the original loop sequences with a 17-nt Tag sequence to promote TGT enzymatic labeling.

4.2.3 CRISPR-CLAMP: a sgRNA Caging Strategies

The ability to dimerize or 'clamp' the RNA provides potential to cage the sgRNA by altering its ternary structure. The first strategy is to dimerize the sgRNA by covalently linking

the tetra-loop or the stem-loop 2 between two sgRNAs (Figure 4.13). I has been previously discovered that dimerization of the Cas9-sgRNA complex may lead to reduction in activity.[81] In this study, the researcher showed that a small bacteriophage-encoded anti-CRISPR protein blocks activity of a single Cas9 ortholog and induces Cas9 dimerization while preventing binding to the target DNA. Thus, it is possible that we can also inhibit the Cas9 activity by dimerizing the Cas9-sgRNA complexing through our TGT enzymatic dimerization of the sgRNAs (Figure 4.13). Interestingly, we found that instead of inhibiting the Cas9 activity, the dimerization of the sgRNA only reduced the Cas9 activity by about 50

Dimerization via the tetra-loop   Dimerization via the stem-loop 2
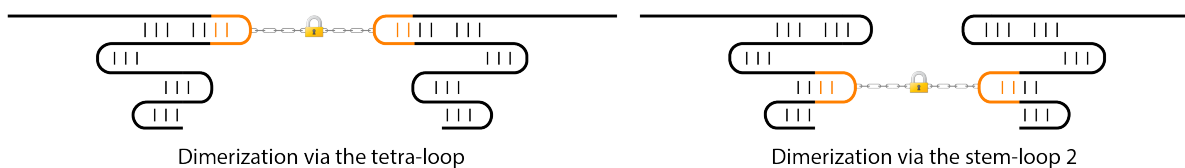
Figure 4.13: Caging of sgRNA by dimierization.
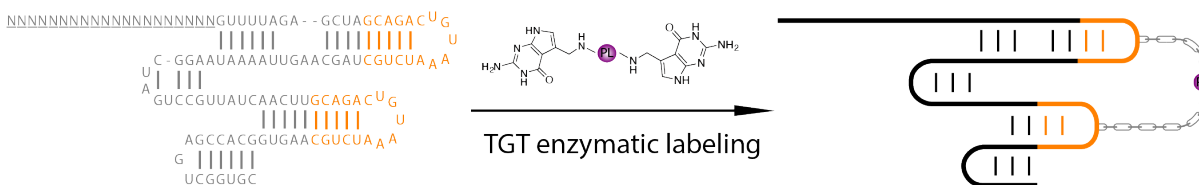
Figure 4.14: Caging of sgRNA by internal 'clamping'. By clamping two internal loops of the sgRNA, the flexibility of the Cas9-sgRNA is significantly reduced, casuing inhibition of the CRISPR/Cas9 gene editing system.

Next, we examined the effect of RNA 'clamping' on the Cas9 activity. Using our established RNA 'clamping' technique, we successfully 'clamped' the tetraloop and the stem-loop 2 of the sgRNA7 (Figure 4.14). Gel purification (denaturing TBE-PAGE) was performed to get rid of any unmodified sgRNAs. Then, we tested the activity of the 'clamped' sgRNA7 in HEK-293 cells. As a result, the gene editing activity dropped to 15% compared to the linear sgRNA7. Based on this promising result, we decided to further improve this inhibition effect by altering our RNA 'clamping' technique.

92

As we discussed in Chapter 4.1 (Introduction), the flexibility of the Cas9-sgRNA complex is critical to the function of the DNA cleavage. By 'clamping' the internal stem-loops of the sgRNA, we reduced such flexibility, causing inhibition of the DNA cleavage. We reasoned that if we further increase the force of such 'clamping', in another word, shorten the distance between the 'clamping' sites, better inhibition effect can be achieved. Thus, I constructed three derivatives of sgRNA7, which are sgRNA8, sgRNA9, sgRNA10 (Figure 4.15).
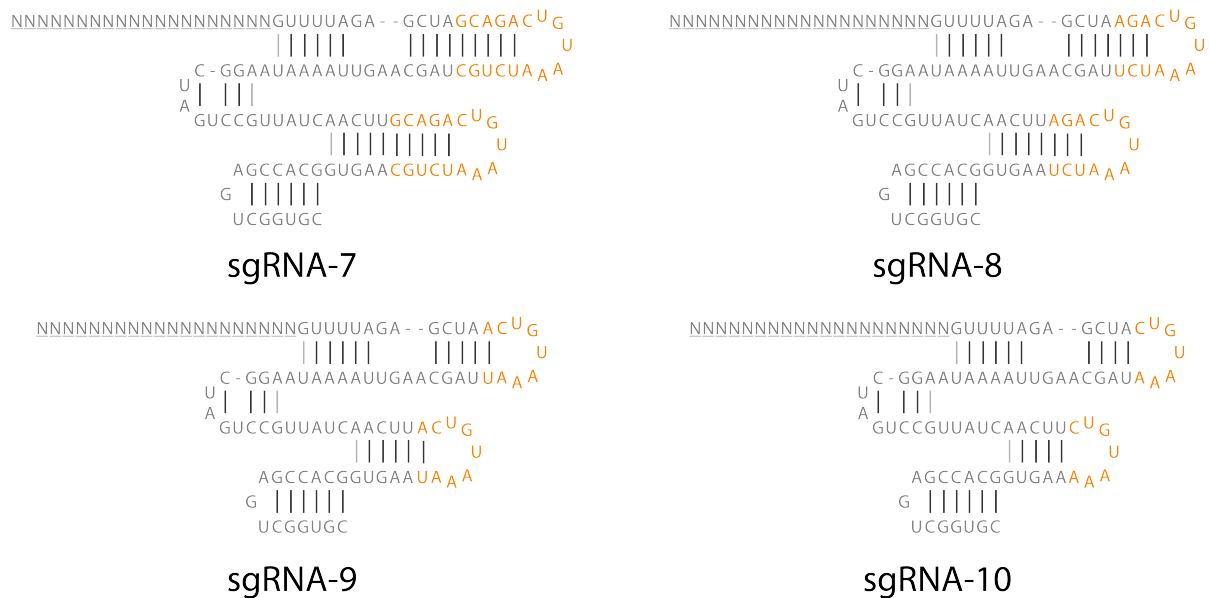
Figure 4.15: Sequences of sgRNA7 to sgRNA10.

sgRNA8, sgRNA9 and sgRNA10 were formed by truncating 2 bp, 4 bp and 5 bp of the stem part of the 17-nt RNA Tag sequence, accordingly. sgRNA10 was designed by exchanging the GAAA sequence of the loop part of the tetra-loop and stem-loop 2 with the new sequence CUGUAAA. Thus, only six nucleotides were added into the wt-sgRNA sequence. Based on previous research,[27] TGT accepts a variety of stem sequences and label the guanine located at the third position of the loop (CUGUAAA). Thus, we reasoned that the TGT enzyme should still 'clamp' sgRNA8, sgRNA9 and sgRNA10 with different stem structure. This was proven by denaturing PAGE analysis. Next, we tested the gene editing activities of 'clamped' sgRNA8 to sgRNA10. In Figure 4.16, we can see that sgRNA8 to sgRNA10 have similar gene editing

activity compared to sgRNA-7 in their linear form. As we previously discussed, the 'clamped' sgRNA7 still has about 15% activity compared to its linear form. We observed similar effect for sgRNA8. Interestingly, however, after clamping sgRNA9 and sgRNA10, their activities completely diminished. This proved the hypothesis that the distance between the 'clamping' sites (the guanine residue within the CUGUAAA loop) has significant effect on the inhibition of gene editing. The shorter the distance, the tighter the 'clamp', resulting in a more rigid sgRNA ternary structure. When the rigidity reached to a certain point, the Cas9-sgRNA complex can no longer be functional as a DNA nuclease.
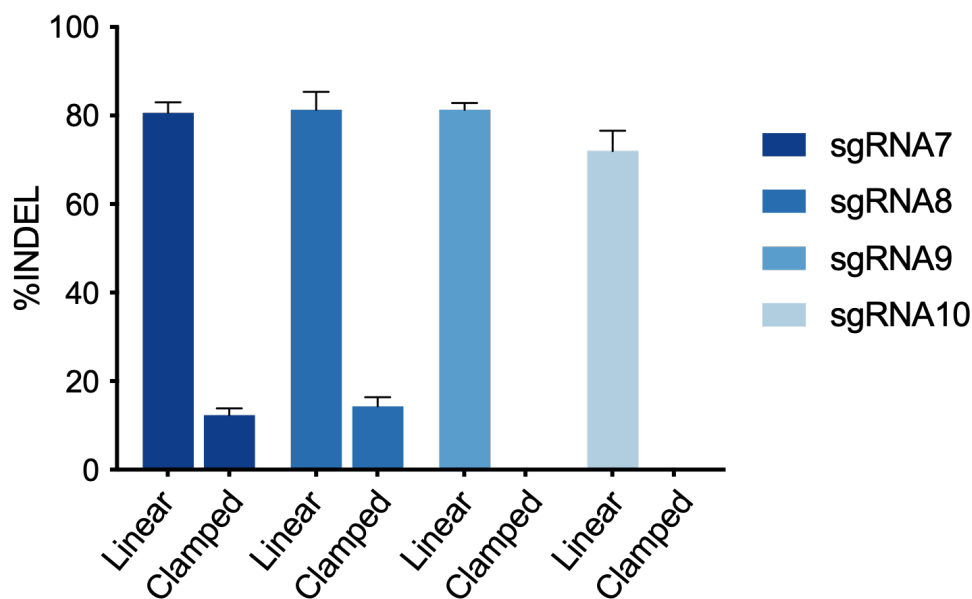
Figure 4.16: Gene editing activity of 'clamped' sgRNA8 to sgRNA10.

4.2.4 Photo-activation of Gene Editing

To achieve photo-activation of CRISPR/Cas9 gene editing, we incorporated a photo-cleavable linker within the TGT small-molecule substrate. As we discussed in Chapter 1.4, an ideal photo-regulation approach should avoid the use of toxic UV light. Therefore, we chose a previous reported [7-(diethylamino)coumarin-4-yl]-methyl (DEACM) linker as the photo-cleavable linker, which can be cleaved by 405 nm to 488 nm wavelength of light.[44,73,82] The

chemical structures of this preQ1-DEACM-preQ1 probe, together with the non-cleavable preQ1-PEG10-preQ1 probe were shown in Figure 4.17. It should be noted that the head-to-head distance of the preQ1-PEG10-preQ1 is around 71 Å, while the head-to-head distance of preQ1-DEACM-preQ1 is only about 43 Å. We purposely designed a shorter probe to reduce the distance between the two guanine residues after the internal 'clamping' reaction, which should result in a greater inhibition effect. We observed that the new preQ1-DEACM-preQ1 substrate was able to inhibit all sgRNAs with different stem length (sgRNA7-10), upon TGT enzymatic 'clamping'.

preQ1-PEG10-preQ1
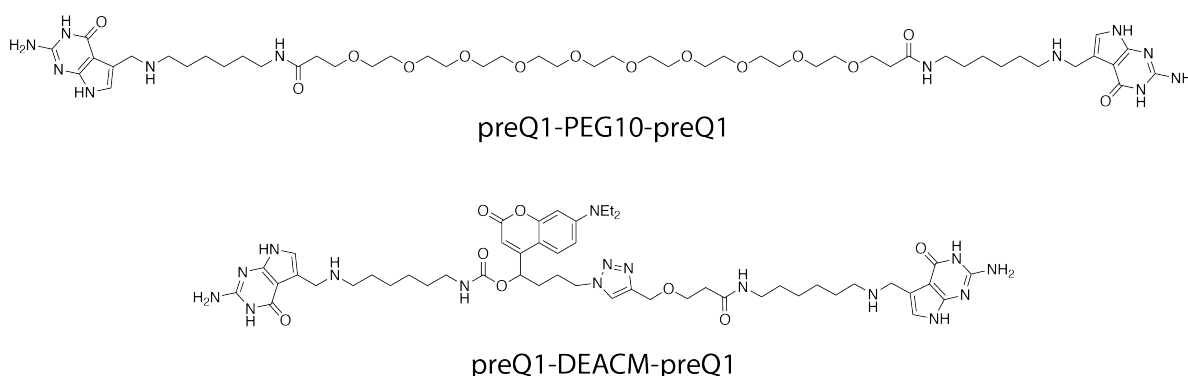
preQ1-DEACM-preQ1

Figure 4.17: Structures of preQ1-PEG10-preQ1 and preQ1-DEACM-preQ1.

| preQ1-DEACM-preQ1 'clamping' | - | + | - | + |
| 456 nm LED irradiation | - | - | + | + |
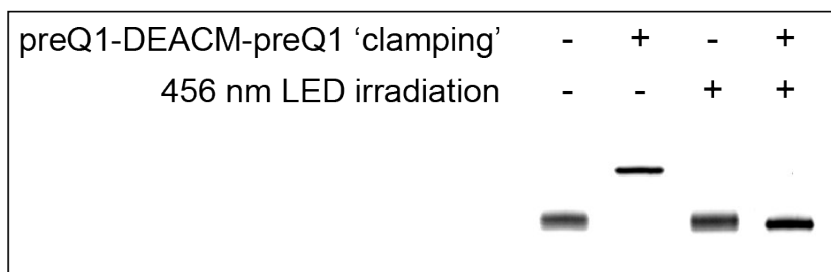
Figure 4.18: 12% denaturing PAGE analysis of TGT labeling reaction using the preQ1-DEACM-preQ1 probe and sgRNA7, and the subsequent in vitro photo-release using 456 nm LED light source.

We then tested the reactivity of the newly synthesized preQ1-DEACM-preQ1 probe using sgRNA7 under a standard TGT labeling condition. The 'clamped' sgRNA7 was further gel purified to get rid of undesired RNA products. To test the photo-cleavage activity of the DEACM linker, a 456 nm wavelength of LED light was used to irradiate the 'clamped' sgRNA (water

solution) for 3 minutes, following by denaturing PAGE analysis (Figure 4.18). As shown in the PAGE analysis, irradiation with 456 nm wavelength of light was harmless to the integrity of the unlabeled sgRNA (1st and 3rd lines). However, three minutes of irradiation completely cleaved the DEACM linker, resulting in a significant gel shift of the photo-linearized sgRNA (4th line) compared to the 'clamped' sgRNA (2nd line).
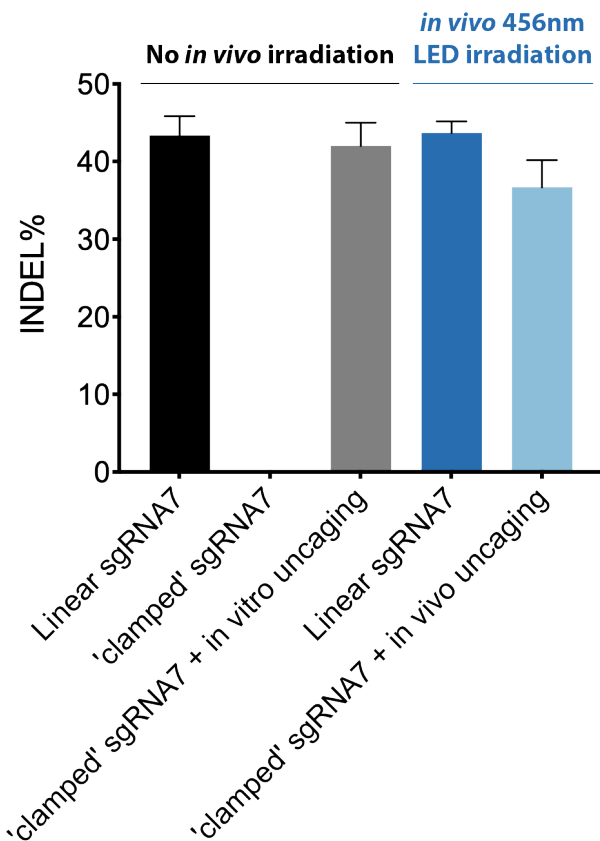
Figure 4.19: In vivo photo-activation of Cas9 gene editing by 456 nm LED.

Next, I tested the gene editing efficiency of the in vitro photo-cleaved sgRNA7. Specifically, I delivered the linear sgRNA7, 'clamped' sgRNA7 and 'clamped' but in vitro photo-cleaved sgRNA into the HEK-293 cells to promote gene editing. Gene editing efficiency (INDEL rate) was quantified by sanger sequencing. To our delight, we observed wild-type gene editing efficiency of the 'clamped' but in vitro photo-cleaved sgRNA. This demonstrated that the photo-cleavage of the DEACM-based cross-linker released the sgRNA and completely restored the CRISPR/Cas9 gene

editing activity. Furthermore, we demonstrated that we can photo-activate the 'clamped' sgRNA in live cells. To achieve live cell photo-activation of CRISPR/Cas9 gene editing, we first transfected the 'clamped' sgRNA into cultured HEK-293 cells. 3 hours after transfection, cells were washed with fresh full-growth medium (DMEM, 10% FBS) to get rid of the lipofectamine-RNA complex. Then, the cells were irradiated with a 456 nm wavelength of LED light for 30 seconds to cleave the DEACM linker and activate the 'clamped' sgRNA. Gene editing efficiency (INDEL%) was quantified by sanger sequencing. As we can see in Figure 4.19, 456 nm wavelength of irradiation (LED) light source successfully activated gene editing in live cells. The INDEL mutation rate for the in vivo activation was 36.7%, which is about 84% compared to the linear sgRNA (43.7%). Thus, we achieved a high dynamic range of deactivation (0% activity) and in vivo photo-activation (84% compared to the wild-type level).

4.2.5 Single-cell Photo-activation of Gene Editing

Light-activation offers excellent spatiotemporal resolution. Previously, we demonstrated single-cell light-activation of mRNA translation.[72,82] To demonstrate the single-cell gene editing activation, we constructed a fluorescence protein reporter cell line. Inspired by previous work,,[83] we inserted three fluorescence reporter genes (1 RFP and 2 GFPs) into the genome of the cells. The RFP is constitutively expressed by the EF-1α promoter, whereas the two GFP genes are not expressed without gene editing activity because the GFP sequences are out of frame and there is a stop codon before the first GFP gene (Figure 4.20). Once a double-strand brake is generated by Cas9 nuclease, error-prone nonhomologous end joining will occur which often results in INDEL mutations. This indel formation may cause frame shifts, making either of the downstream GFP genes in frame and expresses. Theoretically, if a frameshift is randomly generated, the expression activation rate of the downstream GFP is 66.6%. To deliver this surrogate fluorescence reporter genes into the genome of the HEK-293 cells, we applied a 'sleeping beauty' (SB) transposon system.[84,85] First, the three fluorescence protein genes were cloned into a SB vector

containing blasticidin resistance genes. A Cas9 targeting site (SURR) was inserted between the RFP and the first GFP genes. Then, the constructed SB vector and the SB100X transposase vector were co-delivered into HEK-293 cells by lipofectamine transfection. Three days after transfection, blasticidin selection (10 μg/mL) was performed. During the 14-day selection, cells were frequently propagated to increase blasticidin selection efficacy. To test the function of the surrogate reporter, engineered HEK-293 reporter cells were transfected with the 'clamped' sgRNA7. Three hours after transfection, cells were irradiated with a 456 nm wavelength of LED light for 30 seconds to cleave the DEACM linker and activate the 'clamped' sgRNA. Three days after transfection, cell images were taken to quantify RFP and GFP expression levels (Figure 4.21). As expected, most cells expressed RFP. However, the cells transfected with 'clamped' sgRNA did not express GFP, meaning that the 'clamped' sgRNA has no activity, which was in line with our previous observation (Figure 4.19). Excitingly, strong GFP expression was observed among 456 nm wavelength LED irradiated cells, revealing that Cas9 gene editing was activated in these surrogate fluorescence reporter cells.
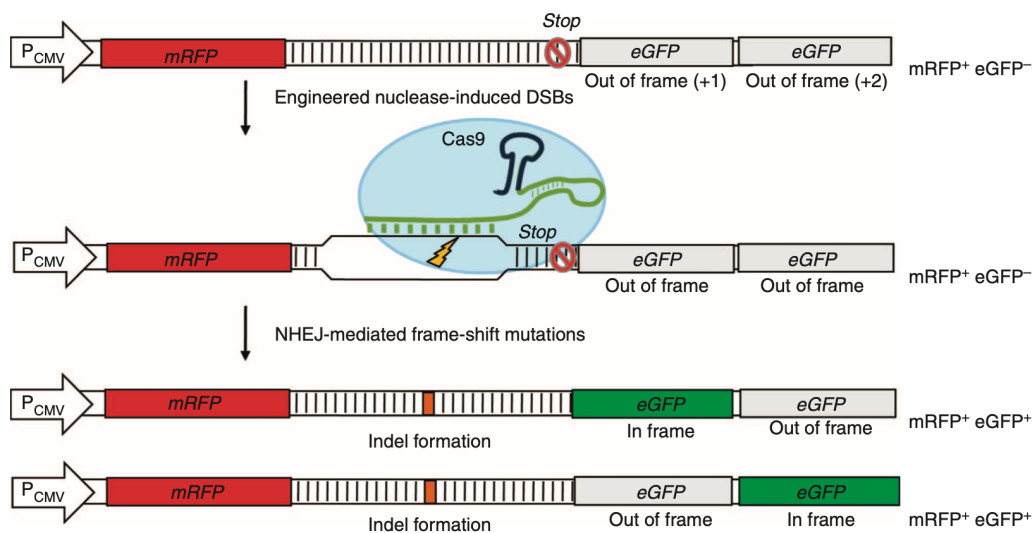
Figure 4.20: Demonstration of the mechanism of the surrogate GFP reporter.

Furthermore, after transfection the 'clamped' sgRNA7 into the HEK-293 surrogate reporter cells, a single cell was selected and irradiated with a 405 nm wavelength of laser using a
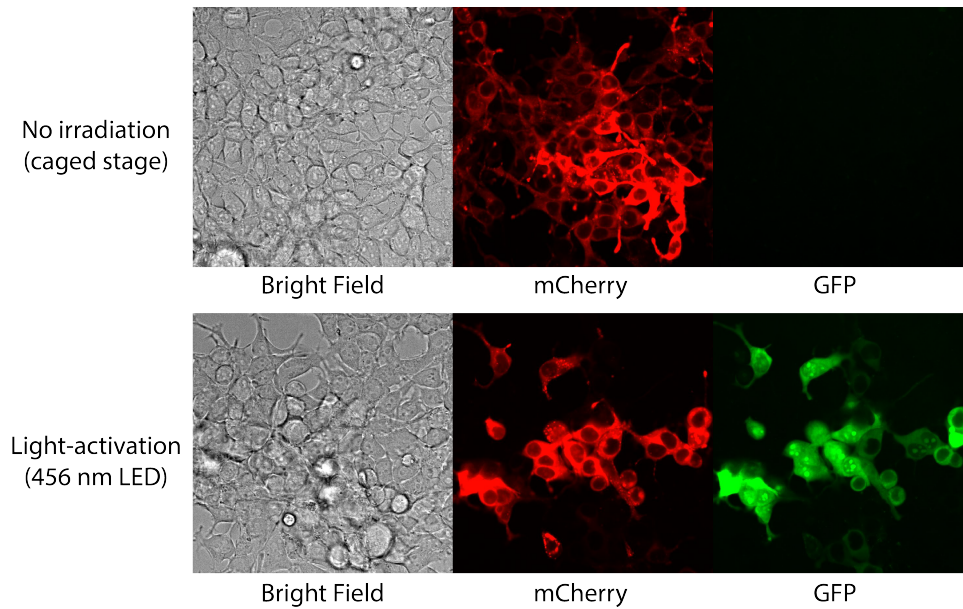
Figure 4.21: Photo-activation of gene editing using surrogate fluorescence reporter cells.

confocal microscope. Cells were continuously images to observe cell growth and the expression of gene editing reporter genes (Figure 4.22). The two images places at the first row (00:00 hour) were the first image taken right after the laser irradiation event. The circled cell was irradiated by a 405 nm wavelength of laser for 10 seconds to cleave the DEACM linker and activation gene editing. Since the sgRNA was at the caged stage, most cells expressed RFP, but no cell expressed GFP. 5 hours after the photo-irradiation event, the laser irradiated cells splitted into two sister cells. 15 hours after irradiation, one of the sister cell started expressing GFP shown as slight GFP fluorescence signal. 21.5 hours after irradiation, this cell showed stronger GFP expression level, while its sister cell also started express GFP. 4 hours later, both of the cells showed strong GFP fluorescence signal as a result of continuous GFP expression. 28.5 hours after irradiation, the first sister cell splitted for the second time. 2 hours later, it sister cell also splitted for the second time. At the end of the time-series imaging, we observed four cells expressing GFP. This clean lineage tracking of gene editing was benefited from the zero background of the 'clamped' sgRNA7. This time-series imaging of the single-cell gene editing photo-activation experiment clearly demonstrated the precision of our sgRNA photo-regulatory technique.
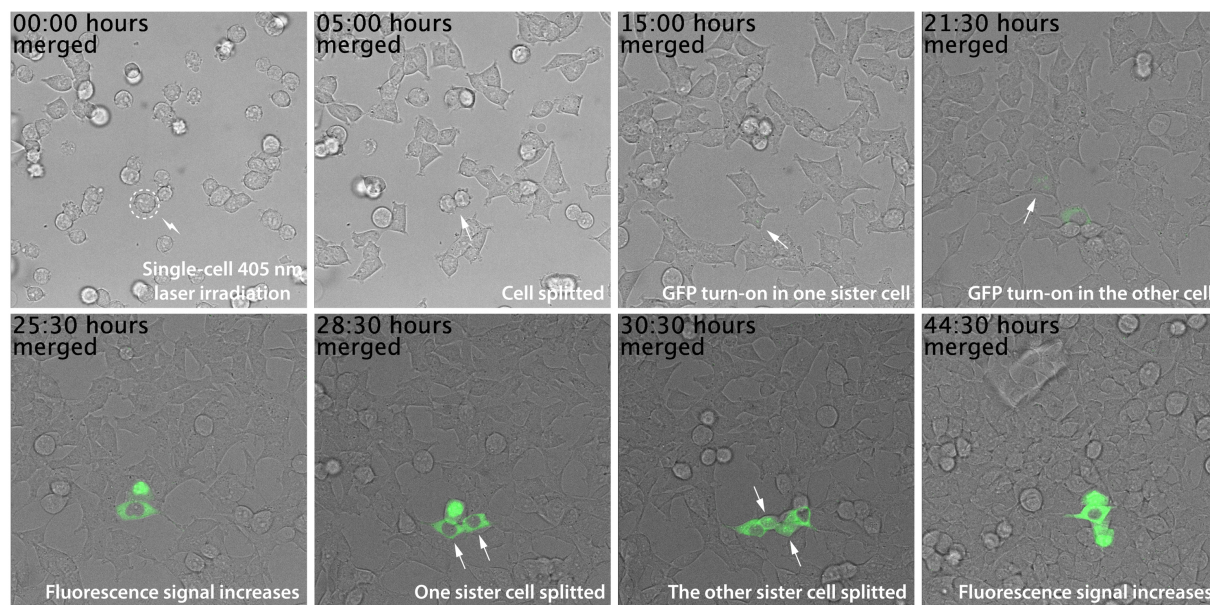
Figure 4.22: Single-cell photo-activation of gene editing. Images are merged images of the bright field and the GFP fluorescence channel.

4.2.6 Multiplex Photo-activation of Gene Editing

RNA-TAG is a powerful RNA-modifying tool because of its flexibility to accept basically any small-molecule substrates baring a preQ1 moiety. This versatility offers potentials to develop a multiplexed gene editing photo-activation platform. To achieve this, we synthesized another photo-cleavable small-molecule TGT substrate, preQ1-NB-preQ1 (Figure 4.23). The NB linker is responsive to 390 nm to 405 nm wavelength of lights. Thus, together with the DEACM linker which is responsive to 405 nm to 456 nm wavelength of light, we were able to build a multiplexed photo-activated gene editing system. First, we select the other genome loci GRIN2B (GGAGAACAGCACTCCGCTCT) as the second targeting site.[79] We chose the sgRNA9 sequence for sgRNA-GRIN2B because this sgRNA sequence has wild-type level of gene editing efficiency. Most importantly, the stem-loops of the 'clamping' sites were significantly shortened to increase inhibition efficacy upon 'clamping'. The sgRNA-GRIN2B sgRNA was 'clamped' by the preQ1-NB-preQ1 probe while the sgRNA-DYRK1A sgRNA was 'clamped' by the preQ1-DEACM-preQ1 probe. Both caged sgRNAs were transfected into HEK-293-Cas9 cells to induce gene

editing. 3 hours after transfection, cells were washed with fresh DMEM to get rid of excess of lipo-RNA complex. Next, cells were irradiated with either 456 nm or 405 nm wavelength of LED lights for 30 seconds. After 48 hours of incubation, gene editing efficiency was quantified by sanger sequencing for both the DYRK1A and GRIN2B sites (Figure 4.24).

preQ1-NB-preQ1

Figure 4.23: Structure of the small-molecule substrate preQ1-NB-preQ1.

Figure 4.24: Multiplexed photo-activation of gene editing.

Irradiation with a 456 nm wavelength of light only cleaved the DEACM linker, releasing the sgRNA-DYRK1A. As a result, we detected an INDEL efficiency of 18.3%. We didn't detect any INDEL mutation at the GRIN2B locus, demonstrating that the 456 nm wavelength of light
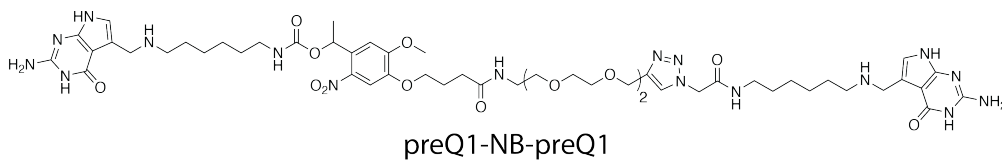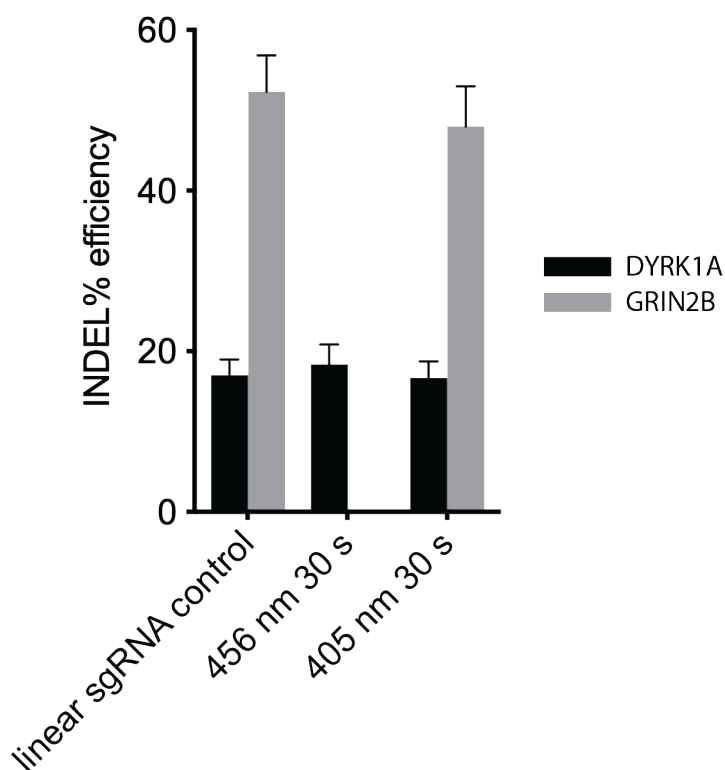
only activated gene editing at the DYRK1A site. In contrast, 405 nm wavelength of light cleaved both the DEACM linker and the NB linker, resulting in gene editing at both genome loci. We detected a INDEL efficiency of 16.7% at the DYRK1A site, and 48.0% at the GRIN2B site. The gene editing activation rate of the DYRK1A site was around 100%, while the activation rate of the GRIN2B site was around 92%. We attribute this discrepancy of activation rate is due to the fact that DEACM linker has higher photo-cleavage efficiency.

4.3 Conclusion

In conclusion, we demonstrated a versatile, highly efficient and multiplexed RNA caging and photo-decaging strategy. We applied this strategy to the sgRNA of the CRISPR/Cas9 gene editing system. Specifically, by internally cross-linking two internal stem-loops within the sgRNA sequence using our RNA-CLAMP technique, Cas9 completely lost its DNA cleavage activity. This caging effect was resulted from rigidification of the sgRNA ternary structure upon enzymatic 'clamping'. Notably, no gene editing was detected when the sgRNA was at the 'clamped' stage. Light irradiation efficiently cleaved the photo-sensitive linker and subsequently activated gene editing in live cells. By using photo-cleavable linker that are responsive to longer wavelength ($>=$ 456 nm) of lights, we avoided the use of toxic UV light for the activation. Furthermore, by using two photo-cages that are responsive to two different wavelength of lights, we achieved a sequential/multiplexed photo-activation of gene editing at two genome loci. Additionally, the 'clamped' sgRNAs were obtained by in vitro transcription, following by post-transcriptional enzymatic modification. No solid-state RNA synthesizer and special caged nucleotides are required. This offer great accessibility and flexibility for many labs. To the best of our knowledge, CRISPR-CLAMP provides the most activation wavelength flexibility, lowest background (no background activity at the caged stage) and multiplexing capability among all the CRISPR-based gene editing photo-activation technique available so far. CRISPR-CLAMP will service a

extremely powerful and versatile technology in the study of complexed gene networks and lineage tracing. A photo-activated CRISPR/Cas9 gene editing platform can also be potentially useful in many therapeutic application to reduce off-targeting rate and improve spatiotemporal gene editing resolution. Besides, our novel RNA-CLAMP technique can be used for many other types of RNA molecules, for example, ribozymes, to regulate their functions. We believed the RNA-CLAMP and CRISPR-CLAMP techniques we presented here provide exciting new potential to regulate the function of RNAs and achieve the photo-activation of gene editing at a whole new level never possible before.

4.4   Acknowledgements

Bibliography

[1] Thomas R Cech and Joan A Steitz. The noncoding rna revolution—trashing old rules to forge new ones. *Cell*, 157(1):77–94, 2014.

[2] Chyi-Ying A Chen, Nader Ezzeddine, and Ann-Bin Shyu. Messenger rna half-life measurements in mammalian cells. *Methods in enzymology*, 448:335–357, 2008.

[3] Caroline C Friedel, Lars Dölken, Zsolt Ruzsics, Ulrich H Koszinowski, and Ralf Zimmer. Conserved principles of mammalian transcriptional regulation revealed by rna half-life. *Nucleic acids research*, 37(17):e115–e115, 2009.

[4] Björn Schwanhäusser, Dorothea Busse, Na Li, Gunnar Dittmar, Johannes Schuchhardt, Jana Wolf, Wei Chen, and Matthias Selbach. Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342, 2011.

[5] Chenghua Cui, Wei Shu, and Peining Li. Fluorescence in situ hybridization: cell-based genetic diagnostic and research applications. *Frontiers in cell and developmental biology*, 4:89, 2016.

[6] Andrea M Femino, Fredric S Fay, Kevin Fogarty, and Robert H Singer. Visualization of single rna transcripts in situ. *Science*, 280(5363):585–590, 1998.

[7] Arjun Raj, Patrick Van Den Bogaard, Scott A Rifkin, Alexander Van Oudenaarden, and Sanjay Tyagi. Imaging individual mrna molecules using multiple singly labeled probes. *Nature methods*, 5(10):877–879, 2008.

[8] Sanjay Tyagi and Fred Russell Kramer. Molecular beacons: probes that fluoresce upon hybridization. *Nature biotechnology*, 14(3):303–308, 1996.

[9] Philip J Santangelo, Brent Nix, Andrew Tsourkas, and Gang Bao. Dual fret molecular beacons for mrna detection in living cells. *Nucleic acids research*, 32(6):e57–e57, 2004.

[10] Vivian J Bardwell and Marvin Wickens. Purification of rna and rna-protein complexes by an r17 coat protein affinity method. *Nucleic acids research*, 18(22):6587–6594, 1990.

[11] Edouard Bertrand, Pascal Chartrand, Matthias Schaefer, Shailesh M Shenoy, Robert H Singer, and Roy M Long. Localization of ash1 mrna particles in living yeast. *Molecular cell*, 2(4):437–445, 1998.

[12] Philip D Campbell, Jeffrey A Chao, Robert H Singer, and Florence L Marlow. Dynamic visualization of transcription and rna subcellular localization in zebrafish. *Development*, 142(7):1368–1374, 2015.

[13] Charles P Lai, Edward Y Kim, Christian E Badr, Ralph Weissleder, Thorsten R Mempel, Bakhos A Tannous, and Xandra O Breakefield. Visualization and tracking of tumour extracellular vesicle delivery and rna translation using multiplexed reporters. *Nature communications*, 6(1):1–12, 2015.

[14] Mireia Ferrer, Caroline Clerté, Célia Chamontin, Eugenia Basyuk, Sébastien Lainé, Jérome Hottin, Edouard Bertrand, Emmanuel Margeat, and Marylene Mougel. Imaging hiv-1 rna dimerization in cells by multicolor super-resolution and fluctuation microscopies. *Nucleic acids research*, 44(16):7922–7934, 2016.

[15] Hans E Johansson, Dagmar Dertinger, Karen A LeCuyer, Linda S Behlen, Charles H Greef, and Olke C Uhlenbeck. A thermodynamic analysis of the sequence-specific binding of rna by bacteriophage ms2 coat protein. *Proceedings of the National Academy of Sciences*, 95(16):9244–9249, 1998.

[16] Jeremy S Paige, Karen Y Wu, and Samie R Jaffrey. Rna mimics of green fluorescent protein. *Science*, 333(6042):642–646, 2011.

[17] Rita L Strack, Matthew D Disney, and Samie R Jaffrey. A superfolding spinach2 reveals the dynamic nature of trinucleotide repeat–containing rna. *Nature methods*, 10(12):1219, 2013.

[18] Grigory S Filonov, Jared D Moon, Nina Svensen, and Samie R Jaffrey. Broccoli: rapid selection of an rna mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society*, 136(46):16299–16308, 2014.

[19] Josephin M Holstein and Andrea Rentmeister. Current covalent modification methods for detecting rna in fixed and living cells. *Methods*, 98:18–25, 2016.

[20] Ross W Richardson and Richard I Gumport. Biotin and fluorescent labeling of rna using t4 rna ligase. *Nucleic acids research*, 11(18):6167–6184, 1983.

[21] Yuri Motorin, Jürgen Burhenne, Roman Teimer, Kaloian Koynov, Sophie Willnow, Elmar Weinhold, and Mark Helm. Expanding the chemical scope of rna: methyltransferases to site-specific alkynylation of rna for click labeling. *Nucleic acids research*, 39(5):1943–1952, 2011.

[22] Daniela Schulz, Josephin Marie Holstein, and Andrea Rentmeister. A chemo-enzymatic approach for site-specific modification of the rna cap. *Angewandte Chemie International Edition*, 52(30):7874–7878, 2013.

[23] Josephin Marie Holstein, Daniela Stummer, and Andrea Rentmeister. Engineering giardia lamblia trimethylguanosine synthase (glatgs2) to transfer non-natural modifications to the rna 5'-cap. *Protein Engineering, Design and Selection*, 28(6):179–186, 2015.

24 Josephin Marie Holstein, Daniela Stummer, and Andrea Rentmeister. Enzymatic modification of 5-capped rna with a 4-vinylbenzyl group provides a platform for photoclick and inverse electron-demand diels–alder reaction. *Chemical science*, 6(2):1362–1369, 2015.

25 Josephin M Holstein, Fabian Muttach, Stephan HH Schiefelbein, and Andrea Rentmeister. Dual 5 cap labeling based on regioselective rna methyltransferases and bioorthogonal reactions. *Chemistry–A European Journal*, 23(25):6165–6173, 2017.

26 Fahui Li, Jianshu Dong, Xiaosong Hu, Weimin Gong, Jiasong Li, Jing Shen, Huifang Tian, and Jiangyun Wang. A covalent approach for site-specific rna labeling in mammalian cells. *Angewandte Chemie*, 127(15):4680–4685, 2015.

27 Seth C Alexander, Kayla N Busby, Christian M Cole, Cun Yu Zhou, and Neal K Devaraj. Site-specific covalent labeling of rna by enzymatic transglycosylation. *Journal of the American Chemical Society*, 137(40):12756–12759, 2015.

28 Geoffrey C Hoops, Leroy B Townsend, and George A Garcia. trna-guanine transglycosylase from escherichia coli: structure-activity studies investigating the role of the aminomethyl substituent of the heterocyclic substrate preq1. *Biochemistry*, 34(46):15381–15387, 1995.

29 Alan W Curnow, Fan Lu Kung, Keith A Koch, and George A Garcia. trna-guanine transglycosylase from escherichia coli: gross trna structural requirements for recognition. *Biochemistry*, 32(19):5239–5246, 1993.

30 Fabian Ehret, Cun Yu Zhou, Seth C Alexander, Dongyang Zhang, and Neal K Devaraj. Site-specific covalent conjugation of modified mrna by trna guanine transglycosylase. *Molecular pharmaceutics*, 15(3):737–742, 2017.

31 Kayla Nicole Busby, Amitkumar Fulzele, Dongyang Zhang, Eric J Bennett, and Neal K Devaraj. Enzymatic rna biotinylation for affinity purification and identification of rna-protein interactions. *bioRxiv*, 2020.

32 Christian Lanctôt, Thierry Cheutin, Marion Cremer, Giacomo Cavalli, and Thomas Cremer. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature Reviews Genetics*, 8(2):104–115, 2007.

33 Rudolf Jaenisch and Adrian Bird. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics*, 33(3):245–254, 2003.

34 Gavin S Wilkie, Kirsten S Dickson, and Nicola K Gray. Regulation of mrna translation by 5-and 3-utr-binding factors. *Trends in biochemical sciences*, 28(4):182–188, 2003.

35 Ye Fu, Dan Dominissini, Gideon Rechavi, and Chuan He. Gene expression regulation mediated through reversible m 6 a rna methylation. *Nature Reviews Genetics*, 15(5):293, 2014.

36 En Li. Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics*, 3(9):662–673, 2002.

[37] Nahum Sonenberg and Alan G Hinnebusch. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, 136(4):731–745, 2009.

[38] Daniel Curtis, Ruth Lehmann, and Phillip D Zamore. Translational regulation in development. *Cell*, 81(2):171–178, 1995.

[39] Alessandro Rosa and Ali H Brivanlou. Synthetic mrnas: powerful tools for reprogramming and differentiation of human cells. *Cell Stem Cell*, 7(5):549–550, 2010.

[40] Ugur Sahin, Katalin Karikó, and Özlem Türeci. mrna-based therapeutics—developing a new class of drugs. *Nature reviews Drug discovery*, 13(10):759–780, 2014.

[41] Richard J Jackson, Christopher UT Hellen, and Tatyana V Pestova. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature reviews Molecular cell biology*, 11(2):113–127, 2010.

[42] Marc Robert Fabian, Nahum Sonenberg, and Witold Filipowicz. Regulation of mrna translation and stability by micrornas. *Annual review of biochemistry*, 79:351–379, 2010.

[43] Jeane M Govan, Rajendra Uprety, Meryl Thomas, Hrvoje Lusic, Mark O Lively, and Alexander Deiters. Cellular delivery and photochemical activation of antisense agents through a nucleobase caging strategy. *ACS chemical biology*, 8(10):2272–2282, 2013.

[44] Sayumi Yamazoe, Qingyang Liu, Lindsey E McQuade, Alexander Deiters, and James K Chen. Sequential gene silencing using wavelength-selective caged morpholino oligonucleotides. *Angewandte Chemie International Edition*, 53(38):10114–10118, 2014.

[45] Jeane M Govan, Douglas D Young, Hrvoje Lusic, Qingyang Liu, Mark O Lively, and Alexander Deiters. Optochemical control of rna interference in mammalian cells. *Nucleic acids research*, 41(22):10518–10528, 2013.

[46] Hideki Ando, Toshiaki Furuta, Roger Y Tsien, and Hitoshi Okamoto. Photo-mediated gene activation using caged rna/dna in zebrafish embryos. *Nature genetics*, 28(4):317–325, 2001.

[47] Shinzi Ogasawara. Control of cellular function by reversible photoregulation of translation. *ChemBioChem*, 15(18):2652–2655, 2014.

[48] Josephin M Holstein, Lea Anhäuser, and Andrea Rentmeister. Modifying the 5-cap for click reactions of eukaryotic mrna and to tune translation efficiency in living cells. *Angewandte Chemie International Edition*, 55(36):10899–10903, 2016.

[49] Hideki Ando and Hitoshi Okamoto. Practical procedures for ectopic induction of gene expression in zebrafish embryos using bhc-diazo-caged mrna. *Methods in cell science*, 25(1-2):25–31, 2003.

[50] Xue Wang, Xianjun Chen, and Yi Yang. Spatiotemporal control of gene expression by a light-switchable transgene system. *Nature methods*, 9(3):266, 2012.

[51] Yanfang Fu, Jennifer A Foden, Cyd Khayter, Morgan L Maeder, Deepak Reyon, J Keith Joung, and Jeffry D Sander. High-frequency off-target mutagenesis induced by crispr-cas nucleases in human cells. *Nature biotechnology*, 31(9):822–826, 2013.

[52] Yanni Lin, Thomas J Cradick, Matthew T Brown, Harshavardhan Deshmukh, Piyush Ranjan, Neha Sarode, Brian M Wile, Paula M Vertino, Frank J Stewart, and Gang Bao. Crispr/cas9 systems have off-target activity with insertions or deletions between target dna and guide rna sequences. *Nucleic acids research*, 42(11):7473–7485, 2014.

[53] Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A Doudna, and Emmanuelle Charpentier. A programmable dual-rna–guided dna endonuclease in adaptive bacterial immunity. *science*, 337(6096):816–821, 2012.

[54] Lukas E Dow, Jonathan Fisher, Kevin P O'rourke, Ashlesha Muley, Edward R Kastenhuber, Geulah Livshits, Darjus F Tschaharganeh, Nicholas D Socci, and Scott W Lowe. Inducible in vivo genome editing with crispr-cas9. *Nature biotechnology*, 33(4):390–394, 2015.

[55] Federico González, Zengrong Zhu, Zhong-Dong Shi, Katherine Lelli, Nipun Verma, Qing V Li, and Danwei Huangfu. An icrispr platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell stem cell*, 15(2):215–226, 2014.

[56] Sojung Kim, Daesik Kim, Seung Woo Cho, Jungeun Kim, and Jin-Soo Kim. Highly efficient rna-guided genome editing in human cells via delivery of purified cas9 ribonucleoproteins. *Genome research*, 24(6):1012–1019, 2014.

[57] Suresh Ramakrishna, Abu-Bonsrah Kwaku Dad, Jagadish Beloor, Ramu Gopalappa, Sang-Kyung Lee, and Hyongbum Kim. Gene disruption by cell-penetrating peptide-mediated delivery of cas9 protein and guide rna. *Genome research*, 24(6):1020–1027, 2014.

[58] James Hemphill, Erin K Borchardt, Kalyn Brown, Aravind Asokan, and Alexander Deiters. Optical control of crispr/cas9 gene editing. *Journal of the American Chemical Society*, 137(17):5642–5645, 2015.

[59] Yuta Nihongaki, Fuun Kawano, Takahiro Nakajima, and Moritoshi Sato. Photoactivatable crispr-cas9 for optogenetic genome editing. *Nature biotechnology*, 33(7):755–760, 2015.

[60] Piyush K Jain, Vyas Ramanan, Arnout G Schepers, Nisha S Dalvie, Apekshya Panda, Heather E Fleming, and Sangeeta N Bhatia. Development of light-activated crispr using guide rnas with photocleavable protectors. *Angewandte Chemie International Edition*, 55(40):12440–12444, 2016.

[61] Wenyuan Zhou, Wes Brown, Anirban Bardhan, Michael Delaney, Amber S Ilk, Randy R Rauen, Shoeb I Kahn, Michael Tsang, and Alexander Deiters. Spatiotemporal control of crispr/cas9 function in cells and zebrafish using light-activated guide rna. *Angewandte Chemie*, 132(23):9083–9088, 2020.

[62] Ji Luo, Rajendra Uprety, Yuta Naro, Chungjung Chou, Duy P Nguyen, Jason W Chin, and Alexander Deiters. Genetically encoded optochemical probes for simultaneous fluorescence reporting and light activation of protein function with two-photon excitation. *Journal of the American Chemical Society*, 136(44):15551–15558, 2014.

[63] Toshiaki Furuta and Kousei Noguchi. Controlling cellular systems with bhc-caged compounds. *TrAC Trends in Analytical Chemistry*, 23(7):511–519, 2004.

[64] Kentaro Katayama, Shinya Tsukiji, Toshiaki Furuta, and Teruyuki Nagamune. A bromocoumarin-based linker for synthesis of photocleavable peptidoconjugates with high photosensitivity. *Chemical communications*, (42):5399–5401, 2008.

[65] Satoshi Uchida, Kazunori Kataoka, and Keiji Itaka. Screening of mrna chemical modification to maximize protein expression with reduced immunogenicity. *Pharmaceutics*, 7(3):137–151, 2015.

[66] Katalin Karikó, Hiromi Muramatsu, Frank A Welsh, János Ludwig, Hiroki Kato, Shizuo Akira, and Drew Weissman. Incorporation of pseudouridine into mrna yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Molecular therapy*, 16(11):1833–1840, 2008.

[67] Lior Zangi, Kathy O Lui, Alexander Von Gise, Qing Ma, Wataru Ebina, Leon M Ptaszek, Daniela Später, Huansheng Xu, Mohammadsharif Tabebordbar, Rostic Gorbatov, et al. Modified mrna directs the fate of heart progenitor cells and induces vascular regeneration after myocardial infarction. *Nature biotechnology*, 31(10):898–907, 2013.

[68] Satheesh Elangovan, Behnoush Khorsand, Anh-Vu Do, Liu Hong, Alexander Dewerth, Michael Kormann, Ryan D Ross, D Rick Sumner, Chantal Allamargot, and Aliasger K Salem. Chemically modified rna activated matrices enhance bone regeneration. *Journal of Controlled Release*, 218:22–28, 2015.

[69] Meltem Avci-Adali, Andreas Behring, Heidrun Steinle, Timea Keller, Stefanie Krajeweski, Christian Schlensak, and Hans P Wendel. In vitro synthesis of modified mrna for induction of protein expression in human cells. *JoVE (Journal of Visualized Experiments)*, (93):e51943, 2014.

[70] Toshiaki Furuta, Samuel S-H Wang, Jami L Dantzker, Timothy M Dore, Wendy J Bybee, Edward M Callaway, Winfried Denk, and Roger Y Tsien. Brominated 7-hydroxycoumarin-4-ylmethyls: photolabile protecting groups with biologically useful cross-sections for two photon photolysis. *Proceedings of the National Academy of Sciences*, 96(4):1193–1200, 1999.

[71] Tatjana Michel, Daniel Luft, Meike-Kristin Abraham, Sabrina Reinhardt, Martha L Salinas Medina, Julia Kurz, Martin Schaller, Meltem Avci-Adali, Christian Schlensak, Karlheinz Peter, et al. Cationic nanoliposomes meet mrna: Efficient delivery of modified mrna using hemocompatible and stable vectors for therapeutic applications. *Molecular Therapy-Nucleic Acids*, 8:459–468, 2017.

[72] Dongyang Zhang, Cun Yu Zhou, Kayla N Busby, Seth C Alexander, and Neal K Devaraj. Light-activated control of translation by enzymatic covalent mrna labeling. *Angewandte Chemie*, 130(11):2872–2876, 2018.

[73] Brenda N Goguen, Andreas Aemissegger, and Barbara Imperiali. Sequential activation and deactivation of protein function using spectrally differentiated caged phosphoamino acids. *Journal of the American Chemical Society*, 133(29):11038–11041, 2011.

[74] Hiroshi Nishimasu, F Ann Ran, Patrick D Hsu, Silvana Konermann, Soraya I Shehata, Naoshi Dohmae, Ryuichiro Ishitani, Feng Zhang, and Osamu Nureki. Crystal structure of cas9 in complex with guide rna and target dna. *Cell*, 156(5):935–949, 2014.

[75] Michael Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, 31(13):3406–3415, 2003.

[76] Allison Waugh, Patrick Gendron, Russ Altman, James W Brown, David Case, Daniel Gautheret, Stephen C Harvey, Neocles Leontis, John Westbrook, Eric Westhof, et al. Rnaml: a standard syntax for exchanging rna information. *Rna*, 8(6):707–717, 2002.

[77] MICHAEL Zuker and AB Jacobson. Using reliability information to annotate rna secondary structures. *Rna*, 4(6):669–679, 1998.

[78] Susana M Cerritelli and Robert J Crouch. Ribonuclease h: the enzymes in eukaryotes. *The FEBS journal*, 276(6):1494–1505, 2009.

[79] F Ann Ran, Patrick D Hsu, Jason Wright, Vineeta Agarwala, David A Scott, and Feng Zhang. Genome engineering using the crispr-cas9 system. *Nature protocols*, 8(11):2281–2308, 2013.

[80] Tim Hsiau, Travis Maures, Kelsey Waite, Joyce Yang, Reed Kelso, Kevin Holden, and Rich Stoner. Inference of crispr edits from sanger trace data. *BioRxiv*, page 251082, 2018.

[81] Lucas B Harrington, Kevin W Doxzen, Enbo Ma, Jun-Jie Liu, Gavin J Knott, Alireza Edraki, Bianca Garcia, Nadia Amrani, Janice S Chen, Joshua C Cofsky, et al. A broad-spectrum inhibitor of crispr-cas9. *Cell*, 170(6):1224–1233, 2017.

[82] Dongyang Zhang, Shuaijiang Jin, Xijun Piao, and Neal K Devaraj. Multiplexed photo-activation of mrna with single-cell resolution. *ACS Chemical Biology*, 2020.

[83] Suresh Ramakrishna, Seung Woo Cho, Sojung Kim, Myungjae Song, Ramu Gopalappa, Jin-Soo Kim, and Hyongbum Kim. Surrogate reporter-based enrichment of cells containing rna-guided cas9 nuclease-induced mutations. *Nature communications*, 5(1):1–10, 2014.

[84] Adam J Dupuy, Keiko Akagi, David A Largaespada, Neal G Copeland, and Nancy A Jenkins. Mammalian mutagenesis using a highly mobile somatic sleeping beauty transposon system. *Nature*, 436(7048):221–226, 2005.

[85] Eric Kowarz, Denise Löscher, and Rolf Marschalek. Optimized sleeping beauty transposons rapidly generate stable transgenic cell lines. *Biotechnology journal*, 10(4):647–653, 2015.