Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Tewoderos Ayele Date

Covalent Labeling of Protein and RNA Using Fluorescent Dyes

By

Tewoderos Ayele

Doctor of Philosophy

Chemistry

_____

Dr. Jennifer Heemstra

Advisor

_____

Dr. Dennis C. Liotta

Committee Member

_____

Dr. Khalid Salaita

Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.

Dean of the James T. Laney School of Graduate Studies

_____ Date

Covalent Labeling of Protein and RNA Using Fluorescent Dyes

By

Tewoderos Ayele

B.S., University of Colorado Denver, 2014

Advisor: Jennifer Heemstra, Ph.D.

An abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Chemistry

2020

Abstract

Covalent Labeling of Protein and RNA Using Fluorescent Dyes

By Tewoderos Ayele

Living systems are complex machines that are operated by tightly regulated interactions and organizations of biomolecules. Therefore, sensitive detection and tracking of these molecules are essential for understanding their complex biochemical properties. While there are various approaches for labeling and visualization of cellular biomolecules, the utility of the majority of these approaches is limited due to their non-covalent nature. To identify and image biomolecules in the complex environment, fluorescent probes must have several essential characteristics. In addition to being compatible with the cellular environment, these probes need to be cell-permeable, non-cytotoxic, selective, produce high signal-to-background ratio, and have minimal effect on the native property of the biomolecule. To date, no one technique fulfills all of these requirements, creating a unique demand for fluorescent probes that are adaptable for various experimental conditions. As a result, synergistic advances in organic chemistry, biology, physics, chemical biology, and spectroscopy are underway for improving these fluorescent probes. The work presented here aims to address this limitation by introducing a broadly applicable strategy for covalently and fluorescently labeling proteins and RNAs in complex environments. This strategy utilizes rationally designed fluorescent probes containing various reactive linkers.

Chapter 2 describes a photoaffinity protein labeling approach using a novel malachite green analog. We show that this technique selectively and covalently labels target proteins in live mammalian cells with temporal resolution and minimal background signal.

Chapter 3 describes the implementation of a similar photoaffinity approach to covalently label cellular RNA. In this work, we report the design of the first covalent light-up aptamer system for visualizing the spatiotemporal localization pattern of mRNA in live mammalian cells.

Chapter 4 describes an approach for covalent chemical labeling and affinity capture of inosine-containing RNAs using acrylomidofluorescein.

Taken together, the results demonstrated in this work highlight the advantages and utility of covalent and fluorescent labeling of biomolecules in complex environments.

Covalent Labeling of Protein and RNA Using Fluorescent Dyes

By

Tewoderos Ayele

B.S., University of Colorado Denver

Advisor: Jennifer Heemstra, Ph. D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2020

Acknowledgement

The journey towards my Ph.D. has been long and tiring. I would have never reached this milestone without the continued support of my family, friends, and mentors. Too many names come to mind when I write this acknowledgment, but I can only list a few of these individuals. First, I would like to express my sincere gratitude to my advisor Dr. Jennifer Heemstra for allowing me to join her lab. Throughout my Ph.D., she consistently provided me with abundant guidance and support. I feel very honored to earn my Ph.D. under her guidance. I would also like to express my sincere gratitude to all the past and current members of the Heemstra lab. Their friendship and support created a welcoming environment, I was eager to come to each day.

I would like to thank my committee members Dr. Khalid Salaita and Dr. Dennis Liotta, for their invaluable guidance and mentorship. I am sincerely thankful for both Dr. Liotta and Dr. Salaita for always making time to discuss anything I needed help with. Without their precious support and guidance, it would not be possible to achieve this milestone.

I would like to thank my excellent collaboration with Dr. Gary Bassell and his lab. In a relatively short time of collaboration, his counsel and expertise provided critical guidance for the completion of my main project and enabled the generation of several followup projects that I am proud to have started. His positive attitude and consistent encouragement also created a welcoming environment that made the collaboration one of the highlights of my Ph.D.

I would like to thank both the University of Utah and Emory University admission committees. I thank the University of Utah admission committee for accepting my

graduate school application and for allowing me to start my graduate career in the beautiful Salt Lake City. I thank Emory University for allowing me to transfer my Ph.D. and finish my studies without encountering any major hurdles and delay to my graduation timeline.

Next, special thanks to my undergraduate research advisor Dr. Marino Resendiz. A profound and sincere thanks for understanding, encouraging, supporting, and providing persistent guidance that I am grateful to receive throughout the years. If it was not for his encouragement and belief in my ability to do science, I would not have pursued science and discovered a career path I am now passionate about. His genuine kindness towards others and determination to help others is a quality I work to emulate in my personal and professional life. For these and many more other reasons, I am incredibly grateful to continue calling him my mentor.

Finally, I would like to give a heartfelt thank you to my family. Words cannot express my love and gratitude to the wonderful people I am privileged to call my family. Most Importantly, I would like to thank my mother and sister for always being by my side and providing never-ending love and support. I am deeply indebted to them for the sacrifices and support they tirelessly offer me. Making these amazing women proud has always been an impetus to me for success. Everything good I accomplish in my career and life, I attribute to their presence in my life. I would not have accomplished so much without their love, support, mentorship, and guidance!

Table of Contents

Chapter 1: Fluorescent labeling of RNA and proteins 1

1.1 Introduction 2

1.2 Protein labeling 2

1.2.1 FP-based protein labeling 2

1.2.2 Organic fluorescent dyes for protein labeling 4

1.3 RNA labeling 5

1.3.1 Single molecule fluorescent *in situ* hybridization for RNA labeling 5

1.3.2 Bacteriophage coat protein-derived RNA labeling 7

1.3.3 Light-up aptamer-based RNA labeling 9

1.3.4 Remaining challenges in RNA labeling 11

1.4 References 12

Chapter 2: Fluorogenic Photoaffinity Labeling of Proteins in Living Cells 16

2.1 Abstract 17

2.2 Introduction 17

2.3 Design and synthesis of fluorogenic photoaffinity labeling system 19

2.4 *In vitro* protein labeling 22

2.5 Protein labeling in live cells 25

2.6 Conclusion 27

2.7 Methods and supplemental information 28

2.7.1 Synthesis of MG-diazirine 28

2.7.2 Preparation of FAP fusion vectors 31

2.7.3 Preparation of mCer3-FAP plasmid vector 32

2.7.4 HeLa cell culture and maintenance 33

2.7.5 Plasmid Transfection 34

2.7.6 Imaging 35

2.7.7 Washout experiment 36

2.7.8 Kinetic experiment for protein concentration 37

2.7.9 MG-diazirine concentration dependent kinetic experiment 39

2.7.10 Cell viability experiment 40

2.8 References 41

Chapter 3: Imaging and tracking mRNA in live mammalian cells via fluorogenic photoaffinity labeling 48

3.1 Abstract 49

3.2 Introduction 50

3.3 Results 54

3.3.1 In vitro characterization 54

3.3.2 Fixed-cell imaging of RNA. 56

3.3.3 Comparison of fixed cell imaging of MGA/MGD2 with FISH. 59

3.3.4 Live cell imaging of mRNA. 61

3.4 Conclusion. 63

3.5 Methods and supplemental material 64

3.5.1 *In vitro* fluorescence enhancement 64

3.5.2 Cell culture and transfection 65

3.5.4 Immunofluorescence and MGA/MGD2 co-imaging 66

3.5.5 Confocal microscopy 67

3.5.6 Live cell imaging 68

3.5.7 MGA array plasmid construction 68

3.5.8 Synthesis of MGD2 70

3.5.9 Determination of MGD2 selectivity in cellular RNA. 71

3.6 References 78

Chapter 4: Chemical Labeling and Affinity Capture of Inosine-Containing RNAs Using Acrylamidofluorescein 82

4.1 Abstract 83

4.2 Introduction 83

4.3 Results and discussion 86

4.4 Supporting information 94

4.4.1 Synthesis of Acrylamidofluorescein 94

4.4.2 Ribonucleoside Labeling and HPLC Analysis 95

4.4.3 RNA Oligoribonucleotides 105

4.4.4 Oligoribonucleotide labelling and PAGE analysis 106

4.4.5 Oligoribonucleotide labelling and Immunoprecipitation pulldown..... 106

4.5 References 108

Chapter 5 : Conclusion and future direction..... 112

5.1 Conclusion and future direction..... 113

List of Figures

Figure 1-1 Methods for genetic tagging and fluorescent visualization of target proteins. 3

Figure 1-2 Methods for fluorescent labeling of Intracellular RNA. a, smFISH for RNA labeling 7

Figure 2-1 a, Temporally-controlled covalent labeling of a protein of interest (POI) ... 20

Figure 2-2 In vitro assessment of covalent photoaffinity labeling 23

Figure 2-3 Kinetics of labeling reaction 24

Figure 2-4 Testing the specificity of FAP labeling in cell lysate. 25

Figure 2-5 Fluorescence imaging of HeLa cells transfected with mCerulean3-FAP 27

Figure 2-6 Plasmid map of mCer3-FAP control expression vector highlighting important component for cloning and expression. 32

Figure 2-7 Representative fluorescence microscopy images of transfected HeLa cells incubated with MG-diazirine 36

Figure 2-8 Washout experiment 37

Figure 2-9 Protein concentration dependent kinetic experiment. 38

Figure 2-10 Kinetic experiment for MG-diazirine concentration 39

Figure 2-11 Cell viability 41

Figure 3-1 Characterization of MGA-functionalized mRNA in the presence of MGD2 52

Figure 3-2 RNA labeling in fixed cells. 57

Figure 3-3 Live cell tracking of RNA and protein granules 60

Figure 3-4 Selective labeling of MGA functionalized mRNA in the presence of cellular RNA extracted from HeLa cells 72

Figure 3-5 UV dependent labeling of RNA 76

Figure 4-1 Chemical labeling of inosine 85

Figure 4-2 Representative HPLC traces depicting the reaction between inosine and acrylamidofluorescein over 24 hours..... 86

Figure 4-3 Denaturing PAGE analysis of synthetic oligoribonucleotides labeled with acrylamidofluorescein 89

Figure 4-4 Workflow for quantifying pulldown efficiency with acrylamidofluorescein labeling and immunoprecipitation 91

Figure 4-5 Synthesis of acrylamidofluorescein 94

Figure 4-6 Representative HPLC traces of ribonucleoside reactivity with acrylonitrile 97

Figure 4-8 Percent conversion of ribonucleosides when reacted with acrylamidofluorescein 101

Figure 4-9 ESI-MS and MS/MS spectra of isolated product fraction for the reaction of inosine and acrylonitrile 102

Figure 4-10 ESI-MS and MS/MS spectra of isolated product fraction for the reaction of inosine and acrylamidofluorescein 103

Figure 4-11 ESI-MS and MS/MS spectra of isolated product fraction for the reaction of pseudouridine and acrylonitrile. 104

Figure 4-12 ESI-MS and MS/MS spectra of isolated product fraction for the reaction of pseudouridine and acrylamidofluorescein 105

List of Tables

Table 2-1 Tabular data for protein concentration dependent kinetic experiment.....38

Table 2-2 Tabular data for protein concentration dependent kinetic experiment 40

Table 2-3 Tabular data for cell viability experiment 41

Table 3-1 Tabular data fluorescence output of MGA array 73

Table 3-2 Tabular data for UV dependent fluorescence enhancement of 1xMGA-mRNA compared to MGD2 in 1xPBS..... 73

Table 3-3 Tabular data for UV dependent fluorescence enhancement of control mRNA compared to MGD2 in 1xPBS..... 73

Table 3-4 Raw data table for fluorescence intensity of RNA foci in untrasfected Neuro-2a cells or Neuro-2a cells expressing mCDK6 functionalized with 1xMGA or 6xMGA at the 5'UTR 74

Table 3-5 Tabular data for fluorescence of RNA foci 76

Table 4-1 Preparation of RNA-I-Cy5 and RNA-A-Cy3 mixture solutions 107

List of Frequently Used Abbreviations

μM	Micromolar
A	Adenine
ADAR	Adenosine deaminase acting on RNA
bp	Base pair
^{13}C	Carbon
C	Cytidine
Calcd	Calculated
CDK	Cyclin-dependent kinase
CMV	Cytomegalovirus
CP	Coat protein
d	Doublet
DCM	Dichloromethane
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EF-1a	Human elongation factor-1 alpha
ESI-TOF	Electrospray ionization time-of-flight
FAE^{1}I	N^1-fluoresceinacrylamidoethylinosine
FAP	Fluorogen activating Protein
FISH	Fluorescent *in situ* hybridization
FLAP	Fluorescent light-up proteins
FP	Fluorescent protein

G	Guanosine
^{1}H	Proton
HPLC	High performance liquid chromatography
HRMS	High-resolution mass spectrometry
I	Inosine
ICE	Inosine erasing
IF	Immunofluorescent
IP	Immunoprecipitation
K_d	Dissociation constant
kDA	Kilodalton
LB	Luria broth
LRMS	Low-resolution mass spectrometry
m	Multiplet
m/z	Mass-to-charge ratio
mCer3	mCerulean3
MG	Malachite green
mg	Milligram
MGA	Malachite green aptamer
MGD	Malachite green diazirine
MHz	Megahertz
ml	Milliliter
mM	Millimolar
mmol	Millimole
MS	Mass spectrometry

NLS	Nuclear localizing signal
NMR	Nuclear magnetic resonance
nt	Nucleotide
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PBST	Phosphate-buffered saline with tween
PCR	Polymerase chain reaction
Pd/C	Palladium on carbon
POI	Protein of interest
RBP	RNA binding proteins
RFU	Relative fluorescent unit
RNA	Ribonucleic acid
ROI	RNA of interest
s	Singlet
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SEM	Standard error of the mean
seq	Sequencing
smFISH	Single-molecule fluorescent *in situ* hybridization
TEAA	Triethylammonium acetate
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TLC	Thin-layer chromatography
UV	Ultraviolet

UV	Uridine
v/v	Percentage by volume
w/w	Percentage by weight
Ψ	Pseudouridine

Chapter 1: Fluorescent labeling of RNA and proteins

1.1 Introduction

Sensitive analytical labeling and detection of biomolecules are essential to uncover the complex biochemical processes in living systems. Chemical probes and intrinsically fluorescent proteins (FPs) are primarily employed for investigating the biological function and activity of DNA, RNA, and carbohydrates. Despite the availability of various labeling strategies, there remains a need for a broadly applicable and covalent functionalization approach that enables enrichment, analysis, and visualization of these cellular components. The work presented in this dissertation describes the development of various covalent labeling approaches of protein and RNA molecules that provide a variety of scientific utility enabling enrichment, analysis, and visualization of these cellular components.

1.2 Protein labeling

1.2.1 FP-based protein labeling

To enable imaging of proteins in live cells, an early strategy was developed in which a protein of interest (POI) is genetically fused in frame to a FP (Figure 1-1a). To obtain this gene fusion, the coding DNA sequence for the POI is amplified, isolated, and ligated to the gene of an appropriate FP, such as green fluorescent protein (GFP) or one of the many other color variants.[1] This fusion construct is then inserted into an appropriate mammalian expression plasmid vector and introduced into cells, which subsequently express the fluorescent fusion product to enable visualization. This technique has been widely adopted and is quite useful, as all components needed for visualization are genetically encodable and are produced inside of the cell. Further, molecular cloning

techniques are relatively cheap and straightforward, and various fusion protein designs and iterations can be quickly constructed and inserted into appropriate plasmid vectors using economical reagents.[2] These constructs can also be easily introduced into most cell types using standard transient transfection techniques or, when required, stably integrated into cellular genomes using lentiviral or CRISPR-based strategies.[3] However, the constitutive fluorescence of these FP fusions presents a large limitation in signal-to-noise resolution. To achieve sufficient fluorescent signal inside this cell, strong expression promoters are typically required in mammalian plasmid vectors, including cytomegalovirus and elongation factor-1a. While these promoters are often necessary to achieve sufficient protein levels for many POI fusions to enable robust visualization, this technique can also be problematic in that it results in the uncontrolled production of an unnaturally high copy number of the fusion protein. This can generate an overall diffuse signal throughout the cell and may not recapitulate natural expression levels or subcellular localization.[4, 5]

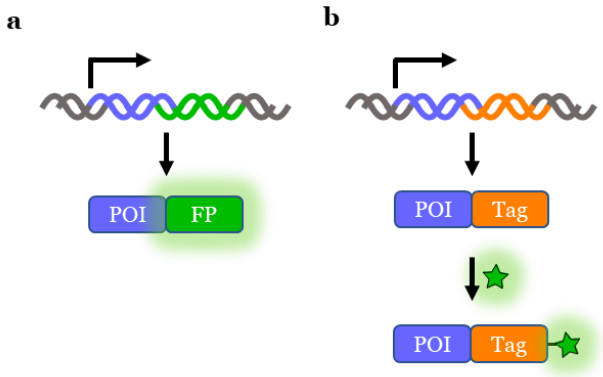


Figure 1-1 Methods for genetic tagging and fluorescent visualization of target proteins. (a) POI can be genetically fused to a FP, or (b) attached to a peptide or protein tag that is recognized by reactive organic fluorescent dye

1.2.2 Organic fluorescent dyes for protein labeling

Organic fluorescent dyes have desirable properties that overcome the limitations of FPs such as smaller size, better photochemical properties, lower background, and broader spectral range. As a result, several alternative protein labeling techniques have been developed that utilize chemical labeling strategy to functionalize POI with fluorescent organic dyes (Figure 1-1b). Most notably, self-labeling enzymes such as SNAP-tag,[6] CLIP-tag,[7] FIAsH, and ReAsH[8] allowed for the direct chemical labeling of the POI with organic fluorophores. These systems rely on the selective reactivity of the fluorophores with the cysteine residues found in the active site of the enzymes. The advantages of these methods include substantially improved control over the intensity and timing of fluorescent labeling and the ability to harness the diverse palette of small-molecule fluorophores. However, techniques of covalent protein labeling that target specific amino acids or short peptide sequences still lack the labeling selectivity in complex biological systems where other proteins or molecules are likely to have similar reactive groups.

Together, both FPs and fluorescent dye-based approaches have advanced the field of protein labeling and have been invaluable for the current understanding of protein structure and function. However, each of these methods have their limitations, and the ideal choice of protein labeling strategy is highly dependent on the biological question at hand. As a result, there is an active effort in the scientific community to add to the currently existing approaches by developing a more robust method that will enable unique experimental designs. Chapter 2 of this dissertation discusses in-depth about a new protein labeling strategy we developed to complement current approaches. By

appending a photocrosslinking group onto a fluorogenic dye, we showed a fast, selective, and covalent labeling of cellular POI. We anticipate this approach will elucidate novel cellular mechanism by enabling experimental designs that require protein labeling, imaging, isolation, and real-time pulse-chase analysis.

1.3 RNA labeling

In addition to *in situ* imaging and tracking of mature proteins, elucidating the upstream regulation mechanisms of protein synthesis is necessary to understand cellular biology. The process of protein synthesis starts from DNA molecules, which store the genetic information necessary for expression and regulation of proteins. Several key classes of RNA molecules then work in coordination to translate this genetic information into functional proteins. Therefore, the complex stages of RNA maturation, regulation, and spatiotemporal transport plays a key role in cellular development and physiology.

To enable better investigation of RNA biology, significant efforts have been directed towards the developments of sequence specific RNA imaging tools. However, unlike FPs, there are no intrinsically fluorescent nucleic acid sequences that can be used for visualizing intracellular RNA. Therefore, alternative approaches that utilize *in situ* hybridization, aptamers that bind to fluorescent dyes, or engineered RNA motifs that are recognized by FP-fused coat protein are used for imaging and studying RNA biology.

1.3.1 Single molecule fluorescent *in situ* hybridization for RNA labeling

While bulk analysis approaches such as RNA-sequencing, quantitative reverse transcription polymerase chain reaction (RT-qPCR), and northern blot can be used to quantify the abundance of transcripts in tissues or single cells, they do not inform the

spatial distribution of specific RNA in different subcellular compartments. Adding to these techniques, single-molecule fluorescence *in situ* hybridization (smFISH) has emerged as a powerful tool for investigating both quantity and spatial distribution of cellular transcripts.[9, 10] In this approach, short oligonucleotides that are complementary to different regions of a target cellular RNA are chemically synthesized and functionalized with fluorescent molecules. When fixed and permeabilized cells are then incubated with a solution containing these prefunctionalized sequences, Watson-Crick-Franklin base paring results in the hybridization of these probes to the target RNA in a 'tiling' fashion (Figure 1-2a). The use of multiple hybridization probes that recognize a single transcript generates bright signal that is distinguishable from the signal produced by excess probes. In recent years, smFISH has been used to elucidate various RNA biology, for example 1) cell-to-cell heterogeneity of gene expression,[11] 2) importance of *β-actin* mRNA localization for fibroblast motility,[12] and 3) multiplexed visualization of several RNA species in a single cell.[13]

Despite being the benchmark for analyzing both the localization and expression levels of transcripts, smFISH is limited to a static view of RNA. The requirement of cell fixation and permeabilization makes this technique not suitable for investigating the dynamic processes of RNA. Therefore, smFISH can only be used as an endpoint assay.

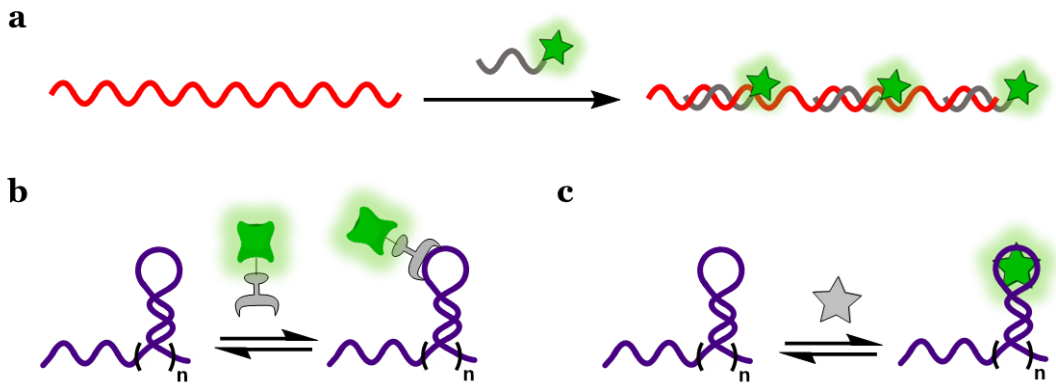


Figure 1-2 Methods for fluorescent labeling of Intracellular RNA. a, smFISH for RNA labeling. Fluorescently labeled synthetic oligos hybridize to their complimentary intracellular transcript. b, FP-ta2agged RBPs for RNA labeling. FP-tagged RBPs recognize and bind to hairpin sequence that is appended to the transcript of interest. c, FLAPs for RNA labeling. RNA aptamers appended to the target RNA bind and enhance the fluorescence of fluorogenic small molecules.

1.3.2 Bacteriophage coat protein-derived RNA labeling

To address the limitation of smFISH, Singer and coworkers reported the first use of RNA-binding proteins (RBPs) as RNA visualization tool.[10] These proteins recognize and bind to a distinct RNA sequence with high affinity. One such RBP that has been primarily used for RNA imaging is the MS2 bacteriophage coat protein (MCP). In addition to self-assembling to form the icosahedral capsid of the virus, MCP also binds to a specific hairpin structure and sequesters viral RNA inside the capsid. Taking advantage of the strong affinity and specificity of this protein-RNA pair, FP-fused MCP have been used to label RNA molecules that have tandem repeats of the hairpin structure at the 3' UTR (Figure 1-2b). This genetically encodable approach has since been a workhorse for

visualizing the localization and dynamics of single mRNA molecules in living yeast, plant, bacteria, and mammalian cells. [14-17]

The MS2 system also paved the path for the discovery of other RNA-protein pair systems such as PP7 and lambda N. Together, these orthogonal systems, allowed for visualization of the spatiotemporal behavior of multiple transcripts in a single cell. Although these RNA-protein based approaches are genetically encodable and can be used for concurrent labeling of multiple transcripts, several caveats have thwarted their robustness. The major drawback being the expression of excess FP-tagged coat proteins inside the cell, which results in substantial background signal. This limitation is marginally addressed by tagging unbound CP-FPs with nuclear localization signal (NLS) sequence, which sequesters the excess NLS-CP-FP proteins to the nucleus.[17] Incorporating a large number of coat-protein recognizing sequences into the transcript under study also increased the fluorescence intensity over background.

While these approaches somewhat improved the signal intensity, they introduce a new set of limitations to the system. The incorporation of NLS results in accumulation of the CP-FP to the nucleus hindering RNA analysis near the nucleus. This can be a major limitation for studying RNA biology as most RNA events such as nuclear export, transcription, splicing, editing, and translation occur near or inside the nucleus. Additionally, the incorporation of tandem repeats of the aptamer to the RNA of interest followed by CP-FP binding results in a significant increase in size to the transcript. This major alteration of the size and structure of the transcript could affect the localization pattern, secondary structure, and overall function of the RNA.[18]

1.3.3 Light-up aptamer-based RNA labeling

Aptamers are short oligonucleotides that have high specificity and affinity for a target ligand molecule by folding into a structured module. More specifically, fluorescent light-up aptamers (FLAPs) are progressively selected in vitro from a large library of randomized sequences to have a tight binding to a fluorescent molecule and thereby enhancing the fluorescence output. The fluorogenicity of these systems enables for minimization of background signal, which alleviates the need for washout of excess dye. Additionally, FLAPs can be either genetically encoded into the host DNA or exogenously introduced into the cell.

Tsien and coworkers reported the first FLAP by selecting for a short RNA sequence that binds and significantly enhances the fluorescence of the malachite green (MG) fluorogen.[19] Similarly, Paige and coworkers reported the Spinach aptamer that binds and activates the fluorescence of DFHBI, a small molecule mimic of the HBI chromophore found in GFP.[20] The spinach aptamer was also used for imaging cellular RNA. However, it was identified that the intracellular fluorescence enhancement of the Spinach was significantly reduced. This property is likely due to the complexity of the intracellular environment, which could compromise the small molecule and aptamer binding as well as the aptamer folding efficiency. Although multiple iterations of the Spinach aptamer (Spinach2,[21] Baby Spinach,[22] and iSpinach[23]) have been reported, the requirement for high ion concentration restricted their applicability to bacterial RNA imaging and *in vitro* assays.

The discovery of both the MG and Spinach aptamers spurred the development of other FLAPs such as Mango,[20] Corn,[24] and Broccoli.[25] These FLAPs aimed to address

some of the limitations of Spinach by providing improvements such as small size, higher brightness, reduced reliance on high ion concentration for folding, and stronger affinity to their respective chromophore. However, crystal structures of these aptamers show that the chromophore binding pockets all contain a G-quadruplex structure.[22, 26-30] This structural requirement of the FLAPs remains a significant constraint in their broad applicability as eukaryotic cells have robust machinery that globally unfolds RNA G-quadruplexes.[31, 32]

Recently, the Pepper FLAP was reported showing significant advancement of aptamer-based labeling approaches.[33] This aptamer has a strong binding affinity to its cognate chromophore with a dissociation constant of ~ 3.5 nM. Compared to Broccoli and Corn, the pepper aptamer produced ~ ten-fold greater fluorescence signal in mammalian cells. Further enhancing the pepper FLAP, Yang and coworkers demonstrated effective tuning of the fluorescence output by functionalizing the HBC chromophore with a different electron donor and acceptor groups.

Despite the constant advancements and promising potential of aptamer-based systems, their non-covalent nature limits their robustness. Even the Pepper and Mango aptamers that have high affinity to their respective ligands with K_d of 3 and 4 nM, were shown not to withstand washout conditions.[20, 30, 33] This inherent limitation prevents the use of aptamer-based systems for pulse-chase and other experiments where media exchange is required. Therefore, the development of FLAPs that address these limitations is highly desirable to widen their applicability in eukaryotic cells.

1.3.4 Remaining challenges in RNA labeling

The RNA imaging field is no longer in its infancy, and the development of a plethora of RNA labeling approaches described in this chapter substantiates this claim. These research tools have elucidated various properties of intracellular RNA and primed the curiosity for deeper understanding of transcript dynamics, birth, maturation, and death. Most notably, these approaches revealed that RNA localization is a very ubiquitous cellular phenomenon and cellular homeostasis is highly regulated by this asymmetric distribution of transcripts. A comprehensive understanding of RNA biology and its localization mechanisms, however, requires the development of new and improved tools that will enable visualization of the dynamic nature of RNA in its native environment.

There are numerous currently impossible experimental designs that will benefit from the advancement of robust RNA imaging approaches. These experiments include: pulse-chase labeling of RNA to track the localization pattern of transcripts at various stages of cell cycle or changing environmental conditions; live-cell tracking of RNA dynamics using tags that do not affect the native localization pattern of RNA; and sequence-specific pulldown of native RNA to better investigate the transcript-dependent role of RBPs. Chapter 3 and 4 of this dissertation describe novel RNA covalent labeling approaches that can be used to track RNA localization as well as identify epigenetic modifications. Adding to the RNA imaging toolbox, we anticipate these new RNA labeling strategies will open the door to new applications that require covalent and temporally controlled labeling.

1.4 References

1. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C., Green fluorescent protein as a marker for gene expression. *Science* **1994,** *263* (5148), 802-805.

2. Giepmans, B. N.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y., The fluorescent toolbox for assessing protein location and function. *science* **2006,** *312* (5771), 217-224.

3. Lackner, D. H.; Carré, A.; Guzzardo, P. M.; Banning, C.; Mangena, R.; Henley, T.; Oberndorfer, S.; Gapp, B. V.; Nijman, S. M.; Brummelkamp, T. R., A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nature communications* **2015,** *6* (1), 1-7.

4. Deer, J. R.; Allison, D. S., High-level expression of proteins in mammalian cells using transcription regulatory sequences from the Chinese hamster EF-1α gene. *Biotechnology progress* **2004,** *20* (3), 880-889.

5. Qin, J. Y.; Zhang, L.; Clift, K. L.; Hulur, I.; Xiang, A. P.; Ren, B.-Z.; Lahn, B. T., Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PloS one* **2010,** *5* (5).

6. Keppler, A.; Pick, H.; Arrivoli, C.; Vogel, H.; Johnsson, K., Labeling of fusion proteins with synthetic fluorophores in live cells. *Proceedings of the National Academy of Sciences* **2004,** *101* (27), 9955-9959.

7. Gautier, A.; Juillerat, A.; Heinis, C.; Corrêa Jr, I. R.; Kindermann, M.; Beaufils, F.; Johnsson, K., An engineered protein tag for multiprotein labeling in living cells. *Chemistry & biology* **2008,** *15* (2), 128-136.

8. Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y., New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *Journal of the American Chemical Society* **2002,** *124* (21), 6063-6076.

9. Raj, A.; Van Den Bogaard, P.; Rifkin, S. A.; Van Oudenaarden, A.; Tyagi, S., Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods* **2008,** *5* (10), 877-879.

10. Femino, A. M.; Fay, F. S.; Fogarty, K.; Singer, R. H., Visualization of single RNA transcripts in situ. *Science* **1998,** *280* (5363), 585-590.

11. Larson, D. R.; Singer, R. H.; Zenklusen, D., A single molecule view of gene expression. *Trends in cell biology* **2009,** *19* (11), 630-637.

12. Jeffery, W. R.; Tomlinson, C. R.; Brodeur, R. D., Localization of actin messenger RNA during early ascidian development. *Developmental biology* **1983,** *99* (2), 408-417.

13. Chen, K. H.; Boettiger, A. N.; Moffitt, J. R.; Wang, S.; Zhuang, X., Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **2015,** *348* (6233), aaa6090.

14. Van Gijtenbeek, L. A.; Kok, J., Illuminating messengers: An update and outlook on RNA visualization in bacteria. *Frontiers in microbiology* **2017,** *8*, 1161.

15. Lionnet, T.; Czaplinski, K.; Darzacq, X.; Shav-Tal, Y.; Wells, A. L.; Chao, J. A.; Park, H. Y.; De Turris, V.; Lopez-Jones, M.; Singer, R. H., A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nature methods* **2011,** *8* (2), 165.

16. Luo, K.-R.; Huang, N.-C.; Yu, T.-S., Selective targeting of mobile mRNAs to plasmodesmata for cell-to-cell movement. *Plant physiology* **2018,** *177* (2), 604-614.

17. Bertrand, E.; Chartrand, P.; Schaefer, M.; Shenoy, S. M.; Singer, R. H.; Long, R. M., Localization of ASH1 mRNA particles in living yeast. *Molecular cell* **1998,** *2* (4), 437-445.

18. Heinrich, S.; Sidler, C. L.; Azzalin, C. M.; Weis, K., Stem–loop RNA labeling can affect nuclear and cytoplasmic mRNA processing. *Rna* **2017,** *23* (2), 134-141.

19. Babendure, J. R.; Adams, S. R.; Tsien, R. Y., Aptamers switch on fluorescence of triphenylmethane dyes. *Journal of the American Chemical Society* **2003,** *125* (48), 14716-14717.

20. Paige, J. S.; Wu, K. Y.; Jaffrey, S. R., RNA mimics of green fluorescent protein. *Science* **2011,** *333* (6042), 642-646.

21. Strack, R. L.; Disney, M. D.; Jaffrey, S. R., A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat–containing RNA. *Nature methods* **2013,** *10* (12), 1219.

22. Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferré-D'Amaré, A. R., Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature structural & molecular biology* **2014,** *21* (8), 658.

23. Autour, A.; Westhof, E.; Ryckelynck, M., iSpinach: a fluorogenic RNA aptamer optimized for in vitro applications. *Nucleic acids research* **2016,** *44* (6), 2491-2500.

24. Song, W.; Filonov, G. S.; Kim, H.; Hirsch, M.; Li, X.; Moon, J. D.; Jaffrey, S. R., Imaging RNA polymerase III transcription using a photostable RNA–fluorophore complex. *Nature chemical biology* **2017,** *13* (11), 1187.

25. Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R., Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society* **2014,** *136* (46), 16299-16308.

26. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S., Quadruplex DNA: sequence, topology and structure. *Nucleic acids research* **2006,** *34* (19), 5402-5415.

27. Millevoi, S.; Moine, H.; Vagner, S., G-quadruplexes in RNA biology. *Wiley Interdisciplinary Reviews: RNA* **2012,** *3* (4), 495-507.

28. Huang, H.; Suslov, N. B.; Li, N.-S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A., A G-quadruplex–containing RNA activates fluorescence in a GFP-like fluorophore. *Nature chemical biology* **2014,** *10* (8), 686.

29. Warner, K. D.; Sjekloća, L.; Song, W.; Filonov, G. S.; Jaffrey, S. R.; Ferré-D'Amaré, A. R., A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. *Nature chemical biology* **2017,** *13* (11), 1195.

30. Trachman III, R. J.; Demeshkina, N. A.; Lau, M. W.; Panchapakesan, S. S. S.; Jeng, S. C.; Unrau, P. J.; Ferré-D'Amaré, A. R., Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Nature chemical biology* **2017,** *13* (7), 807.

31. Shelke, S. A.; Shao, Y.; Laski, A.; Koirala, D.; Weissman, B. P.; Fuller, J. R.; Tan, X.; Constantin, T. P.; Waggoner, A. S.; Bruchez, M. P., Structural basis for activation of fluorogenic dyes by an RNA aptamer lacking a G-quadruplex motif. *Nature communications* **2018,** *9* (1), 1-10.

32. Guo, J. U.; Bartel, D. P., RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **2016,** *353* (6306), aaf5371.

33. Chen, X.; Zhang, D.; Su, N.; Bao, B.; Xie, X.; Zuo, F.; Yang, L.; Wang, H.; Jiang, L.; Lin, Q., Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nature biotechnology* **2019,** *37* (11), 1287-1293.

Chapter 2: Fluorogenic Photoaffinity Labeling of Proteins in Living Cells

This chapter was derived from published manuscript: *Fluorogenic Photoaffinity Labeling of Proteins in Living Cells*

Tewoderos M. Ayele, Steve D. Knutson, Satheesh Ellipilli, Hyun Hwang, and Jennifer M. Heemstra

Published on April 12, 2019, available online:

https://doi.org/10.1021/acs.bioconjchem.9b00203

This chapter is also derived from a published book chapter:

Chapter Sixteen - Covalent live-cell labeling of proteins using a photoreactive fluorogen

Tewoderos M. Ayele, Steve D. Knutson, and Jennifer M. Heemstra

Published on April 28, 2020, available online:

https://doi.org/10.1016/bs.mie.2020.04.019

2.1 Abstract

Genetically encoded fluorescent proteins or small-molecule probes that recognize specific protein binding partners can be used to label proteins to study their localization and function with fluorescence microscopy. However, these approaches are limited in signal-to-background resolution and the ability to temporally control labeling. Herein, we describe a covalent protein labeling technique using a fluorogenic malachite green probe functionalized with a photoreactive crosslinker. This enables a controlled covalent attachment to a genetically encodable fluorogen activating protein (FAP) with low background signal. We demonstrate covalent labeling of a protein *in vitro* as well as in live mammalian cells. This method is straightforward, displays high labeling specificity, and results in improved signal-to-background ratios in photoaffinity labeling of target proteins. Additionally, this probe provides temporal control over reactivity, enabling future applications in real-time monitoring of cellular events.

2.2 Introduction

Fluorescent labeling of proteins in living systems has provided an invaluable window into cellular function.[1, 2] Protein expression,[3] folding,[4] post-translational modification,[5] and localization[6, 7] can all now be monitored with fluorescent labels and microscopic analysis. The most common protein labeling technique employs genetic fusion of a POI with an intrinsically fluorescent protein (FP) such as green fluorescent protein or its variants.[8] This process is typically accomplished through transfection of cells with a vector containing a promoter that induces the expression of the transgene. In most cases, overexpression of FP fusion proteins does not have an observable effect on the native properties and localization of the target protein.[9, 10] However, the intrinsic

fluorescence of these FP labels often results in diffuse signals throughout the cell, making it difficult to monitor protein dynamics with high resolution. Additionally, promoters such as CMV and EF-1a, commonly used for expression of FP fusions, result in the production of large protein copy numbers, which in turn can compromise signal-to-noise ratios.[11, 12]

Additional protein labeling techniques have been developed using non-covalent or covalent attachment of a fluorescent ligand to a peptide or protein sequence. One common strategy is the tetracysteine-biarsenical system to covalently attach a green or red fluorescent small molecule to a protein of interest.[13, 14] Despite the small size and high quantum yield of these systems, the arsenic-containing small-molecule dyes are shown to have elevated photobleaching properties, diminishing the clarity of subcellular tracking of proteins. Systems such as SNAP/Clip-tag,[15, 16] LAP,[17, 18] Halo-tag,[15, 19] and coiled-coil tag[20] also employ a system where a small organic molecule covalently attaches to the POI. While these techniques are simple and enable covalent modification of target proteins, they still rely on constitutively active fluorophores, and hence background fluorescence is quite high in unreacted probes, requiring additional wash-out steps for effective visualization, rendering it difficult to use these systems for real time monitoring of cellular behavior.

Techniques that use fluorogenic molecules[21-23] have significantly improved this resolution issue, overcoming the background signal generated from unbound molecules. However, these techniques lack either covalent attachment or temporal-control of target protein labeling, and hence their use is limited to non-covalent associations with short, diffusion-limited visualization lifetimes. The ability to temporally control protein labeling would enable pulse-chase experiments for studying time-dependent processes such as

response to a therapeutic agent or flux in metabolic reactions.[24-26] Temporally-controlled covalent protein labeling would also prove useful in pull-down experiments, as the protein of interest (POI) could be selectively labeled and isolated from a complex system.[27, 28] Thus, there remains a need for the development of a protein labeling strategy that can provide robust, stable, and temporally-controlled labeling with low background signal.

2.3 Design and synthesis of fluorogenic photoaffinity labeling system

Herein, we address these needs using a photoaffinity labeling strategy, as light-induced reactions offer covalent attachment with precise temporal control and can also offer spatial control of labeling if desired. Specifically, we use a fluorogen activating protein (FAP) that has been developed by Waggoner and coworkers and elaborated by Bruchez and coworkers. This FAP is composed of a single-chain variable fragment antibody that binds to and enhances the fluorescence of malachite green (MG).[29] Non-covalent labeling using this protein-fluorophore pair has been used to image proteins *in vitro* and in live cells,[30, 31] however, this strategy has yet to be elaborated to enable covalent attachment of the fluorogenic label. Here we extend the functionality of this labeling system by appending a reactive diazirine group to the MG fluorogen, such that irradiation with light triggers covalent attachment to the FAP (Figure 2-1a). The diazirine reactive group is small in size and has been used as a photoactive cross-linker for proteomic profiling and fluorescent labeling of proteins,[32-34] and the high photostability of MG[35, 36] minimizes challenges associated with photobleaching. The ability to form a covalent interaction will open the door to additional experimental possibilities such as immunoprecipitation and enhanced imaging contrast after wash-out.

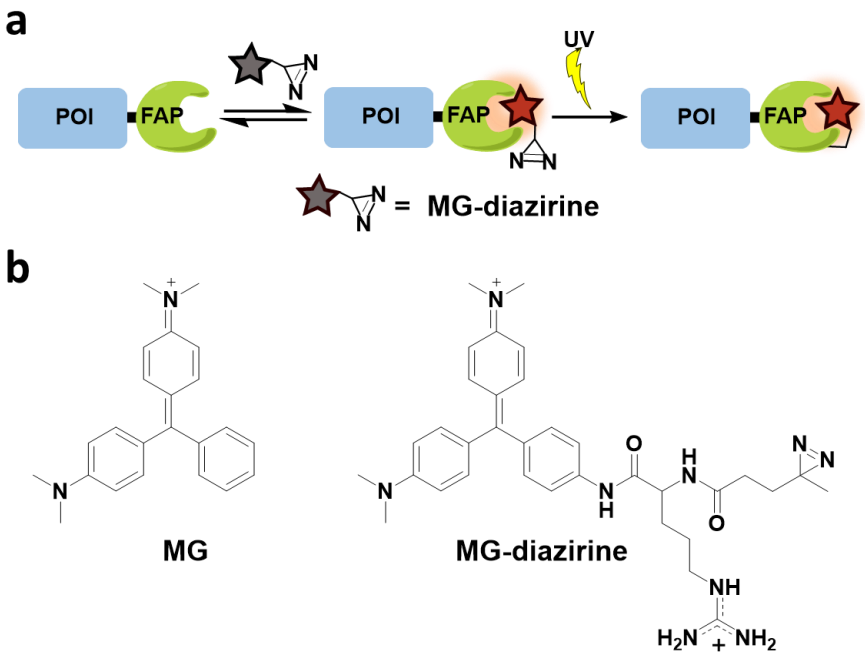


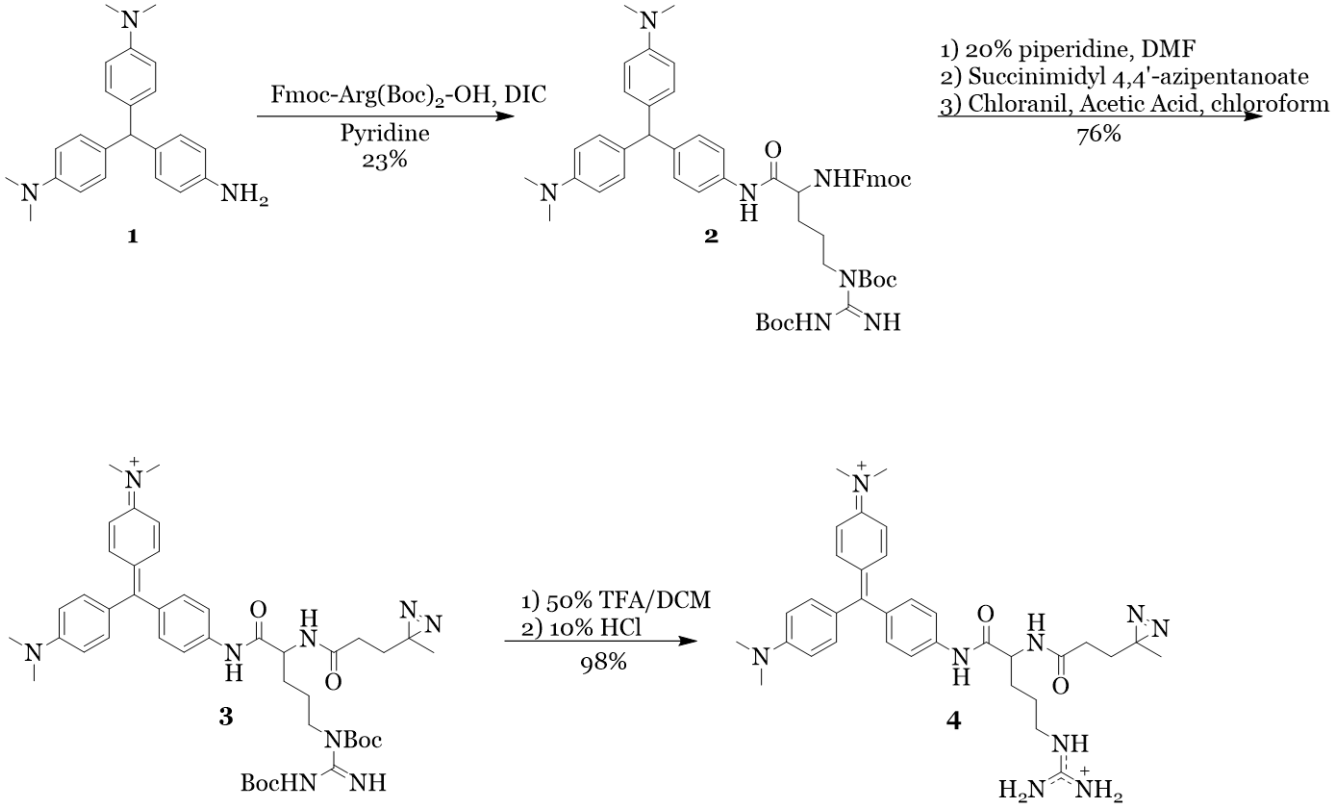
Figure 2-1 a, Temporally-controlled covalent labeling of a protein of interest (POI). b, Chemical structure of MG and MG-diazirine.

To generate our photoaffinity probe, we designed a malachite green (MG) fluorophore functionalized with a diazirine reactive group (MG-diazirine) (Figure 2-1b). Malachite green, a triphenylmethane dye, has very low fluorescence in solution, but can display up to a 20,000-fold increase in brightness when bound to a receptor such as the single-chain antibody-derived FAP. This property of MG provides low background signal for unbound and unreacted probe molecules, as well as for probes that may non-specifically react with biomolecules other than the FAP receptor. Another desirable feature of the MG-FAP system is its fluorescence emission maximum at 680 nm. This far red-shifted fluorescence emission further minimizes background signal by enabling imaging outside of the window where cellular autofluorescence is problematic. For the photoreactive group, we chose diazirine, which forms a highly reactive carbene species upon short bursts of UV irradiation at 355 nm, and is commonly used for photo-

crosslinking of proteins in cells.[32, 37] Finally, we incorporated an arginine moiety in the linker connecting the reactive diazirine group and the MG, as this was found to improve solubility and membrane permeability of the fluorogenic probe.

The synthesis of MG-diazirine is outlined in Scheme 1. Amine-modified malachite green was generated using the method reported by Deng and coworkers and was coupled with Fmoc-Arg(Boc)$_2$-OH.[38] After deprotection of the Fmoc group using 20% piperidine in CH_2Cl_2, the resulting amine was coupled with succinimidyl 4,4'-azipentanoate. Oxidation of the triphenylmethane using chloranil, followed by deprotection of the Boc groups, yielded MG-diazirine.

Scheme 2-1 Synthesis of MG-diazirine



2.4 *In vitro* protein labeling

With the photoreactive fluorogen molecule in hand, we turned to testing the feasibility of our labeling approach *in vitro*. FAP protein was expressed and purified from bacteria, and this was incubated with 50 μM MG-diazirine for 10 min to allow for binding, then irradiated with 365 nm UV light for 5 min to initiate the diazirine group activation. The solution was then mixed 1:1 (v/v) with SDS denaturing loading dye and was heated to 90 °C for 10 min. Importantly, this denaturation step disrupts any non-covalent interactions of the fluorogen with the FAP, allowing for visualization of only the covalently bound FAP-MG-diazirine complexes. After cooling to room temperature, the samples were separated using SDS-PAGE followed by analysis using both Coomassie staining and

fluorescence imaging. MG has very minimal fluorescence after denaturation of the FAP binding partner, which introduces a challenge for detection in SDS-PAGE. However, we have found that freezing the gel by incubation on dry ice provides a similar rotational confinement, producing sufficient fluorescence enhancement to enable visualization of the MG signal. As shown in Figure 2-2, labeling is only observed for the sample having both FAP protein and MG-diazirine, along with exposure to UV irradiation. No labeling is observed when the FAP is not expressed, no UV irradiation is used, or when MG-diazirine is replaced with MG lacking the photoreactive group. These control experiments demonstrate that covalent labeling of the FAP protein is attributable to binding of MG and subsequent light-induced reaction of the diazirine group.

To quantify the rate of covalent labeling, we next conducted a time-course experiment. Solutions containing FAP protein and 50 μM MG-diazirine in 1x PBS were prepared and incubated at 37 °C to allow for noncovalent equilibration of the MG-diazirine with the FAP protein. After the initial incubation time, the samples were subjected to UV irradiation for varying quantities of time, and SDS-PAGE analysis was carried out using similar techniques as described above to determine the extent of covalent labeling (Figure 2-3a-b). Under these conditions, 50% labeling was observed within 10 min of UV activation (Figure 2-3b), which is comparable to previously reported photoaffinity labeling strategies.[39, 40]

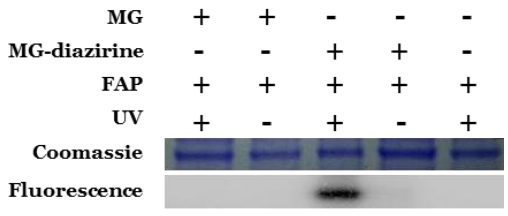


Figure 2-2 In vitro assessment of covalent photoaffinity labeling

While the MG probe only becomes fluorescent after binding to the FAP, we were still curious to know whether the diazirine could react non-specifically with other proteins. To investigate this question, we carried out the labeling reaction in a whole-cell lysate both with and without addition of FAP (Figure 2-4). The solutions were incubated for 30 min, then UV irradiated for 5 min and analyzed by SDS-PAGE. Encouragingly, the only bands visible in the fluorescence channel are those corresponding to the molecular weight of the FAP (major) and uncleaved FAP-His fusion (minor), and these bands are not observed in the no-FAP control. The absence of any other visible bands indicates that labeling is specific to the FAP and that the diazirine does not react with other proteins in a complex biological solution.

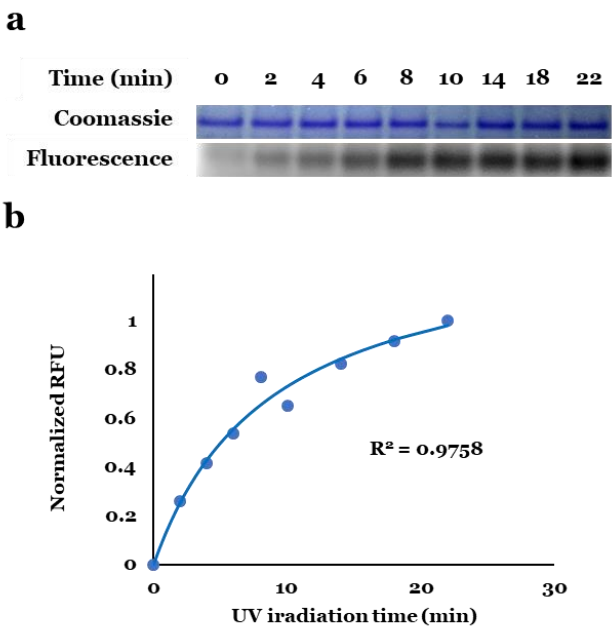


Figure 2-3 Kinetics of labeling reaction. a, SDS-PAGE analysis of labeling as a function of time. b, Normalized fluorescence as a function of time.

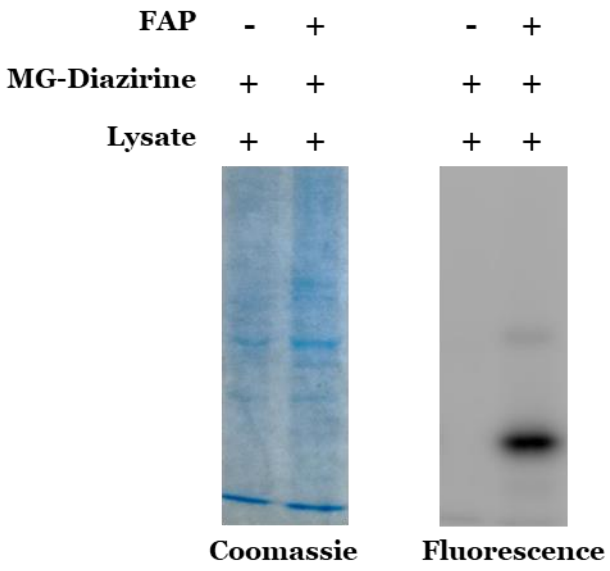


Figure 2-4 Testing the specificity of FAP labeling in cell lysate.

2.5 Protein labeling in live cells

After validating the ability of MG-diazirine to covalently label the FAP *in vitro*, we turned toward our ultimate goal of demonstrating protein labeling in live cells. HeLa cells were transfected with a plasmid encoding a FAP-mCerulean3 fusion protein. In this design, the blue fluorescent mCerulean3 protein serves as both a transfection reporter and as a model POI to evaluate the use of our FAP-MG-diazirine system. Importantly, the excitation and emission wavelengths of the mCerulean3 and MG-diazirine do not overlap, and thus there is no interference from fluorescence crosstalk or FRET signal. The cells were grown to 90% confluency and incubated with 50 μM MG-diazirine for 15 min. UV irradiation was then applied for 5 min, which we previously found to be sufficient for covalent labeling *in vitro*, and we verified that this irradiation procedure did not affect cell viability (Section 2-7-10 and Figure 2-11). UV photolabeling of cells was followed by exchange of the media and fluorescence imaging. The media was subsequently exchanged every 5 min and wash-out of unreacted probe followed by fluorescence imaging. While

not necessary for using this imaging method in future applications, we performed these wash-out experiments to assess the stability of photolabeled versus non-covalent fluorogenic signal over time. Initially, the cells showed similar fluorescence intensity to control cells incubated with MG-diazirine but not exposed to UV irradiation. As expected, cells that were not transfected did not exhibit detectable MG-diazirine fluorescence (Figure 2-5a). This result indicates that the MG-diazirine fluorescence signal results from binding to the FAP and that UV activation of MG-diazirine in the presence of other cellular components does not result in unwanted background fluorescence. After each washing cycle, the transfected cells that were not UV irradiated showed a steady decrease in fluorescence, while the transfected cells that were irradiated with UV light maintained a consistent fluorescence intensity over 40 min of wash-out time (Figure 2-5b&c and Figure 2-8). This indicates that the covalent labeling reaction is also robust in the context of living cells, and that covalent labeling produced robust signal duration and stability. While the fluorescence of non-irradiated cells does not fully reach background levels, we hypothesize that this is due to the high binding affinity of the FAP for MG, making it difficult to fully deplete the non-covalently FAP-bound MG fraction in this control condition. However, the dramatic fluorescence decrease that we do observe suggests that unbound MG-diazirine is increasingly depleted from the cytoplasm, and thus our system would still meet this key requirement for pulse-chase labeling applications.

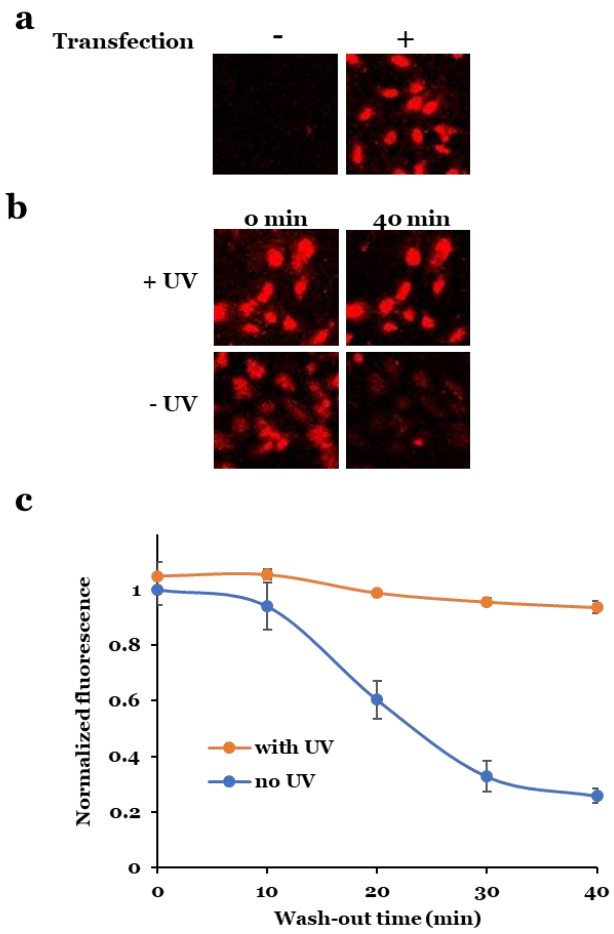


Figure 2-5 Fluorescence imaging of HeLa cells transfected with mCerulean3-FAP. a, Transfected cells and non-transfection control. b, Wash-out of MG-diazirine from non-irradiated cells demonstrates robustness of covalent labeling approach. b, Fluorescence signal as a function of washout time for cells irradiated and not irradiated with UV light.

2.6 Conclusion

In conclusion, we report a new approach for covalent fluorescent labeling of proteins of interest *in vitro* as well as in living cells. To accomplish this, we have synthesized a fluorogenic MG-diazirine probe that has an arginine unit to provide cell permeability and a diazirine reactive group for photoaffinity labeling. The MG-diazirine exhibits very low background signal and has red-shifted fluorescence emission that is well

outside of the cellular auto-fluorescence region. We show that this probe is capable of covalently labeling its FAP binding partner. The MG-diazirine probe is facile to synthesize, and plasmids encoding the FAP protein are widely available, making this approach very user friendly and broadly applicable. Unlike other covalent protein labeling approaches, the use of UV light to initiate probe attachment to the protein of interest provides temporal control over the labeling reaction to enable future applications such as pulse-chase imaging and studies of cellular behavior with enhanced signal stability and reduced background. Complimenting currently existing protein labeling strategies, this research provides a facile approach to covalent protein labeling in a format that is anticipated to open the door to new applications requiring temporal control.

2.7 Methods and supplemental information

2.7.1 Synthesis of MG-diazirine

p-nitro-leucomalachite green

This molecule was synthesized using the same protocol described by Deng and coworkers.[38] First *p*-nitro-malachite green was synthesized by reacting *p*-nitrobenzaldehyde (250 mg, 1.6543 mmol) and *N,N*-dimethylaniline (0.503 ml, 3.9704 mmol) in the presence of zinc chloride (451 mg, 3.3086 mmol) at 100 ^{0}C for 5 h. The reaction was then cooled down to room temperature. The solution was filtered and concentrated under reduced pressure. The product was then purified with column chromatography by gradually changing the polarity of the mobile phase from 100% Hexane to 40% Hexane/ ethyl acetate. ^{1}H NMR (400 MHz, $CDCl_3$) δ *2.93* (s, 12H), 5.45 (s, 1H), 6.66 (d, 4H, J = *8.3* Hz), 6.94 (d, 4H, J = *8.8* Hz), 7.28 (m, 2H, J = *8.3* Hz), 8.101

(d, 2H, *J* = *8.3* Hz). LRMS (ESI-TOF) *m/z* Calcd for $C_{23}H_{26}N_3O_2$ $[M + H]^+$ 376.20195; Found 376.19890.

_p_-amino-leucomalachite green (1)

p-nitro-leucomalachite green (100 mg, 0.2665 mmol) was then reduced using Pd/C (5% w/w) and 9 ml of methanol and THF solution (1/2 v/v) under hydrogen for 3h. The solution was then filtered through celite and concentrated under r educed pressure to yield a light blue solid. Yield = 91.0 mg, 98.9%. ^{1}H NMR (400 MHz, $CDCl_3$) δ 2.90 (s, 12H), 3.53 (s, 2H), 5.27 (s, 1H), 6.58-6.66 (m, 6H), 6.89-6.98 (m, 6H). LRMS (ESI-TOF) *m/z* Calcd for $C_{23}H_{27}N_3$ $[M + H]^+$ 346.22832; Found 346.22868

(9*H*-fluoren-9-yl)methyl (1-((4-(bis(4-(dimethylamino)phenyl)methyl)phenyl)amino)-5-N,N-diboc-guanidino-1-oxopentan-2-yl)carbamate (2)

To a solution of compound **1** (50.00mg, 0.1483 mmol) in pyridine (500.0 µL), *N,N′*-Diisopropylcarbodiimide (70.0 µL, 0.4449 mmol) and Fmoc-Arg(Boc)$_2$-OH (106.0 mg, 0.1780 mmol) were added at room temperature and the reaction mixture stirred for 12 h. After completion of the reaction, the solvent was removed *in vacuo* and the product was extracted with dichloromethane. The organic layer was dried over anhydrous Na_2SO_4 and purified using column chromatography (50% EtOAc/hexane) to obtain compound **2**. Yield = 30.0 mg, 23%. ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.36 (s, 9H), 1.49 (s, 9H), 1.65-1.72 (m, 2H), 1.78-1.91 (m, 2H), 2.89 (s, 12H), 3.41-3.55 (m, 2H), 4.20 (t, 1H, *J* = *8.0 Hz*), 4.36-4.47 (m, 2H), 5.33 (s, 1H), 6.30 (d, 1H, *J* = *8.0 Hz*), 6.63 (d, 4H, *J* = *8.0 Hz*), 6.94 (d, 4H, *J* = *8.0 Hz*), 7.06 (d, 2H, *J* = *8.0 Hz*), 7.26-7.39 (m, 5H), 7.57-7.62 (m, 2H), 7.74 (d, 2H, *J* = *8.0* Hz); 8.41 (s, 1H), 8.48 (t, 1H, *J* = *8.0 Hz*), 11.44 (s, 1H); LRMS (ESI-TOF) *m/z* Calcd. for $C_{56}H_{67}N_5O_7$ $[M + H]^+$ 924.5024; Found 924.5016.

(*Z*)-*N*-(4-((4-(5-(2,3-bis(*tert*-butoxycarbonyl)guanidino)-2-(3-(3-methyl-3*H*-diazirin-3-yl)propanamido)pentanamido)phenyl)(4-(dimethylamino)phenyl)methylene)cyclohexa-2,5-dien-1-ylidene)-*N*-methylmethanaminium (3)

To a solution of compound **2** (30.00mg, 0.0325 mmol) in DMF (500.0 μL), 0.5 mL of 20% piperidine in DMF was added and stirred room temperature for 1h. The free amine was purified by preparative TLC using 2% MeOH/DCM, then coupled with NHS-diazirine (22.0 mg, 0.0975 mmol) in DMF. The solvent was evaporated in vacuo and the coupled product was subjected to oxidation using chloranil (16.0 mg, 0.0650 mmol) and acetic acid (30.0 μL) in chloroform. The reaction mixture was refluxed for 4h. After completion of the reaction, the solvent was removed *in vacuo*, and the compound **3** was purified by preparative TLC using 10% MeOH/DCM. Yield = 20.0 mg, 76%. ^{1}H NMR (400 MHz, CD_3OD) δ 1.00 (d, 3H, *J = 4.0 Hz*), 1.42-1.51 (m, 18H), 1.66-1.70 (m, 4H), 2.14-2.22 (m, 2H), 2.89 (s, 6H), 3.31 (s, 6H), 3.34-3.44 (m, 2H), 4.44-4.61 (m, 3H), 6.68 (d, 2H, *J* = 8.0 Hz), 7.03 (d, 2H, *J* = 8.0 Hz), 7.17 (d, 2H, *J* = 8.0 Hz), 7.33 (m, 2H), 7.42(d, 2H, *J* = 8.0 Hz), 7.86 (d, 2H, *J* = 8.0 Hz); LRMS (ESI-TOF) *m/z* Calcd for $C_{44}H_{60}N_9O_6^+$ $[M]^+$ 810.4661; Found 810.4664.

***N*-(4-((4-(dimethylamino)phenyl)(4-(5-guanidino-2-(3-(3-methyl-3*H*-diazirin-3-yl)propanamido)pentanamido)phenyl)methylene)cyclohexa-2,5-dien-1-ylidene)-*N*-methylmethanaminium (5)**

Compound **3** (20.00mg, 0.0246 mmol) was subjected to Boc-deprotection using 2.0 mL of 50% TFA/DCM for 1 h to obtain compound **4**. The solvent was removed *in vacuo* and the product was treated with 10% *aq*. HCl in order to remove excess TFA to obtain pure compound **5**. Yield = 14.8 mg, 98%. ^{1}H NMR (400 MHz, CD_3OD) δ 1.01 (s,

3H), 1.67-1.72 (m, 2H), 1.77-1.98 (m, 4H), 2.20 (t, 2H, *J* = 8.0 Hz), 3.22-3.26 (m, 2H), 3.32 (s, 12H), 4.52-4.57 (m, 1H), 7.04 (d, 2H, *J* = 8.0 Hz), 7.34 (d, 2H, *J* = 8.0 Hz), 7.42 (d, 2H, *J* = 8.0 Hz), 7.87 (d, 2H, *J* = 8.0 Hz); LRMS (ESI-TOF) *m/z* Calcd for $C_{34}H_{44}N_9O_2^+$ $[M]^+$ 610.3612; Found 610.3589.

2.7.2 Preparation of FAP fusion vectors

To prepare a plasmid for transfection into HeLa cells for subsequent live cell imaging with MG-diazirine probe, we utilized a commercially available pcDNA vector from Addgene (Figure 7), which contains the FAP protein fused to a fluorescent mCerulean3 (mCer3) reporter protein.[23] The fluorescence reporter protein served as a transfection control in the initial validation of this approach, but use of a fluorescent protein is not required for imaging applications. The vector also contains both bacterial replication features and appropriate promoters for mammalian expression. These vectors are typically supplied as agar stabs containing pre-transformed *E. coli* cells harboring the plasmid of interest. We outline how to replicate and extract this plasmid from bacteria for eventual transfection into HeLa cells for expression. This vector also contains numerous restriction enzyme sites surrounding the mCer3 gene (Figure 2-6), allowing users to clone and insert virtually any POI into this backbone and fuse to FAP.

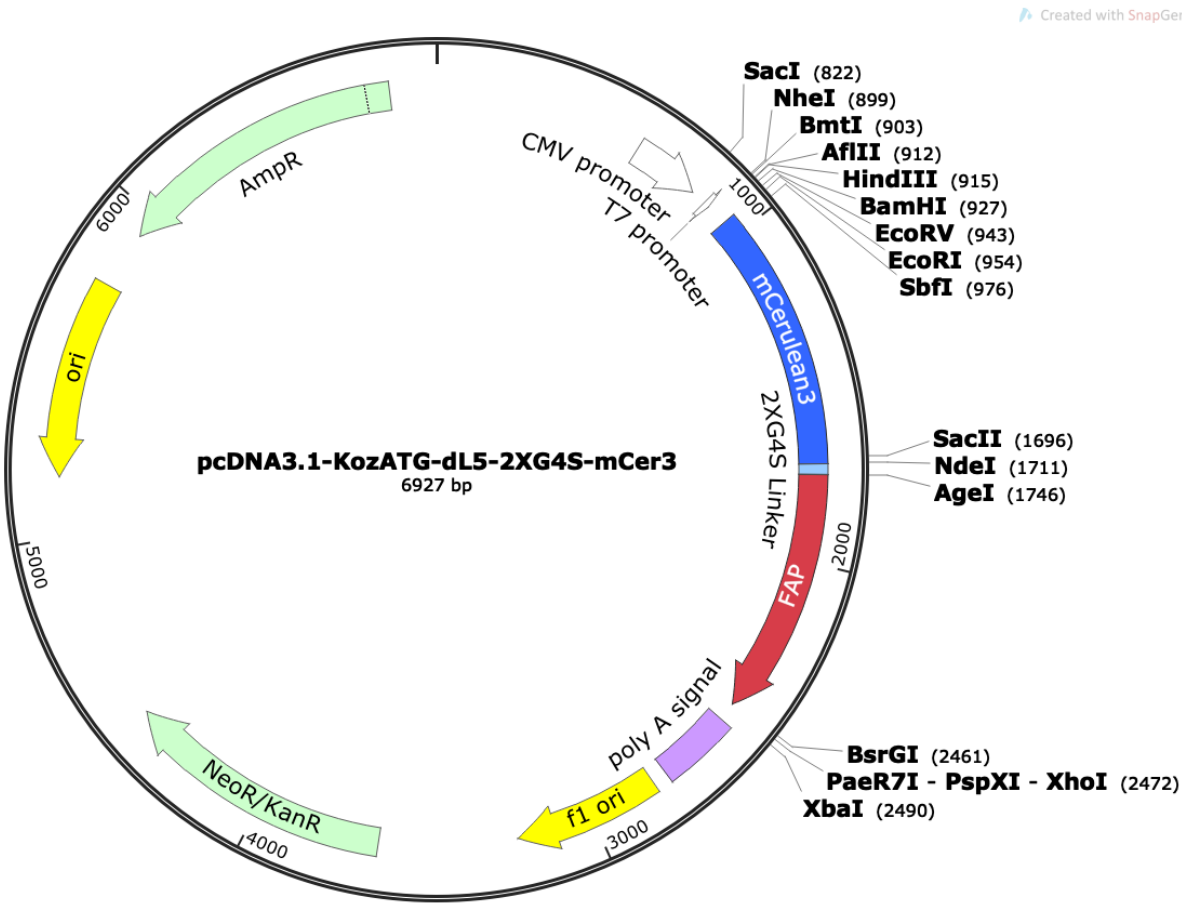


Figure 2-6 Plasmid map of mCer3-FAP control expression vector highlighting important component for cloning and expression.

2.7.3 Preparation of mCer3-FAP plasmid vector

First, luria broth (LB) aga (Miller) plates were supplemented with ampicillin by combining 37 g of the powder media mixture for every liter of ultrapure water. This solution was dissolved completely and sterilized in a suitable container by autoclaving. Separately, LB broth without ampicillin was prepared by combining 25 g of the powder media mixture for every liter of ultrapure water followed by autoclave sterilization. The hot LB broth and agar mixtures were placed in water bath equilibrated to 50 ^{0}C. When

LB broth and agar are cool to touch, a concentrated solution of ampicillin (amp) was added to give a final concentration of 100 μg/ml. While LB agar is still a molten liquid, ~20-25 mL was poured into 100 mm x 15 mm sterile petri dishes and allowed to solidified at room temperature.

Addgene vectors typically arrive as agar stabs which contain *E. coli* cells transformed with the plasmid of interest. To grow more bacteria through replication, a sterile pipet tip was used to streak the bacteria onto a LB + amp agar plate. The plates were then incubated upside down at 37 ^{0}C overnight. The next day, the streak plates showed robust bacterial growth and many distinct colonies was observed. Individual colonies were then picked using sterile pipet tip and dropped in a culture tube with broth and cap. This starter culture was placed in a 37 ^{0}C shaker overnight. The next day, the culture tubes were noticeably turbid indicating successful bacterial growth. The plasmid was then extracted using QIAprep Spin Miniprep kit according to the manufacturer's instructions.

2.7.4 HeLa cell culture and maintenance

Complete EMEM media was prepared by combining 450 mL of EMEM base media and 50 mL FBS. Starter culture was prepared by fully thawing cryovials in 37 ^{0}C and transferring the contents fo the vial to a 15 ml conical tube containing 10 ml of EMEM complete media prewarmed to 37 ^{0}C. The cells were then pelleted by spinning at 500 x g for 5 minutes. The supernatant was then carefully removed, and the cell pellet was resuspended in 10 ml of fresh EMEM complete media. This solution was transferred to a T25 flask and incubated at 37 ^{0}C, 5% CO_2 for 12 hours. The next day, using brightfield microscope adherence of cells and absence of floating or dead cells was confirmed. The

cells were then placed back into the incubator and allowed to grow to ~80% confluency while replacing media as necessary. When cells reached the desired confluency, the media was removed and cells were rinsed with 2-3 mL of -.25% Trypsin/EDTA and this solution was discarded. A fresh 2 mL of 0.25% Trypsin/EDTA was added and the cells were placed in the incubator for 6 min, or until the cells detached from the T25 surface. After full detachment, the cell suspension in 0.25% Trypsin/EDTA was transferred to 9.5 ml of fresh EMEM complete media. The cells were split 1:2 into a large T75 flask by combining 5 mL of the resuspended mixture with 20 mL of fresh media. The cells were split 2-3 times before they are suitable for tranfection.

2.7.5 Plasmid Transfection

After a suitable number of passages, HeLa cells were ready for transfection at confluencies between 50 – 75% in 96-well plates. The media from the wells was removed and replaced with 100 µL pre-warmed Opti-MEM reduced serum medium. Transfection solution A was then prepared by combining 5 µL Opti-MEME reduce serum medium and 0.15 µL Lipofectamine 3000. Transfection solution B was prepared by combining 10 µL Opti-MEM reduced serum medium, 200 ng pcDNA FAP Plasmid Vector, and 0.4 µL P3000 Reagent. Transfection solutions A and B were combined and incubated at room temperature for at least 30 minutes. 10 µL of this combined A/B mixture was added into the desired well of the 96-well plate. The cells were then incubated at 37 °C, 5% CO_2 for 12 hours. After this period, the Opti-MEM from each well was removed and replaced with 100 µL fresh EMEM complete media and the cells were placed back into the incubator for additional 6 hours.

50 mM MG-diazirine stock solution (~30.5 mg/mL) in DMSO was prepared. A working solution of 50 μM MG-diazirine was prepared by dilution the stock solution 1:1000 in EMEM complete media. The media from each of the wells was removed and 100 μL of this working solution was added. The cells were incubated for 15 minutes at 37 37 °C, 5% CO_2 then the MG-diazirine solution from each well was removed. The wells were briefly washed three times with 100 μL fresh EMEM complete media. The cells were then UV-irradiated for 5 minutes before imaging.

2.7.6 Imaging

Fluorescent images were obtained using Leica DMi8 confocal laser scanning microscope equipped with HC Plan Fluotar x10/0.15 air objective and two HYD GaAsP detectors. For mCer3-FAP plasmid, excitation of 433 nm and emission of 475 nm were used. For imaging of the FAP/MG-diazirine, excitation of 640 nm and emission of 669 nm were used. Images for each appropriate channel were obtained using the microscope where clear distinction were shown when comparing between cells that have been transfected with the reporter mCer3-FAP plasmid versus non-transfected controls. Representative images are shown in Figure 2-7, illustrating signal from both mCer3 and FAP-MG diazirine. This experiment thus confirms successful synthesis of MG-diazirine and functional introduction into cells. Additionally, the images obtained from this experiment also provide a good overall measure of plasmid transfection efficiency in HeLa cells.

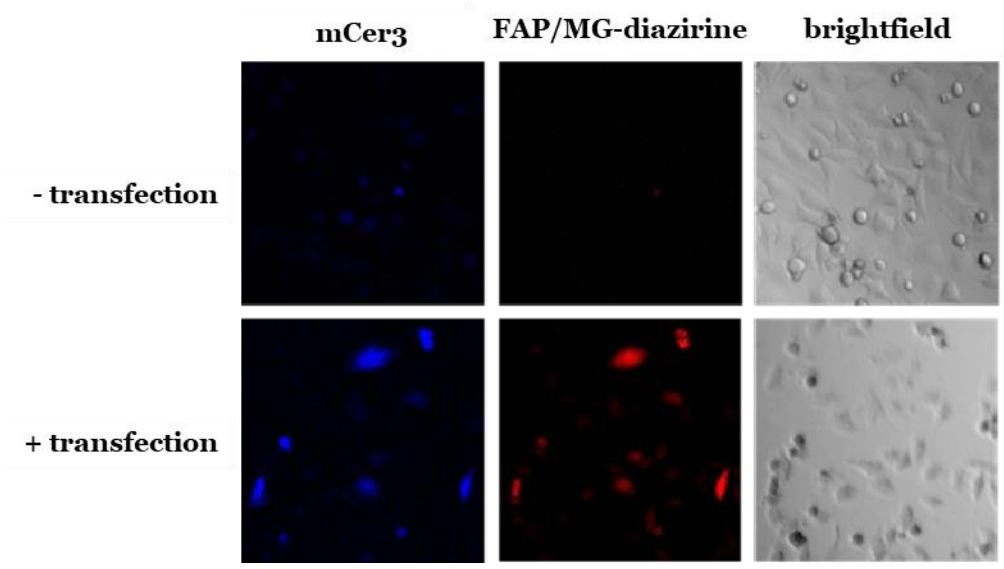


Figure 2-7 Representative fluorescence microscopy images of transfected HeLa cells incubated with MG-diazirine. Transfection with the reporter mCer3-FAP plasmid results in both mCer3 and MG-diazirine fluorescent signal confirming successful transfection and expression of the construct, as well as functional MG-diazirine binding and fluorogenic activity. Cells not exposed to the vector produce no detectable background fluorescence.

2.7.7 Washout experiment

After the cells were transfected, incubated with MG-diazirine solution, and UV-crosslinked, images of the same cells were taken after replacing the media every 10 minutes. Figure 2-8 shows that UV-treated cells retained signal after 40 minutes of washout. However, cells that were not UV-treated showed a decrease in fluorescence brightness over time.

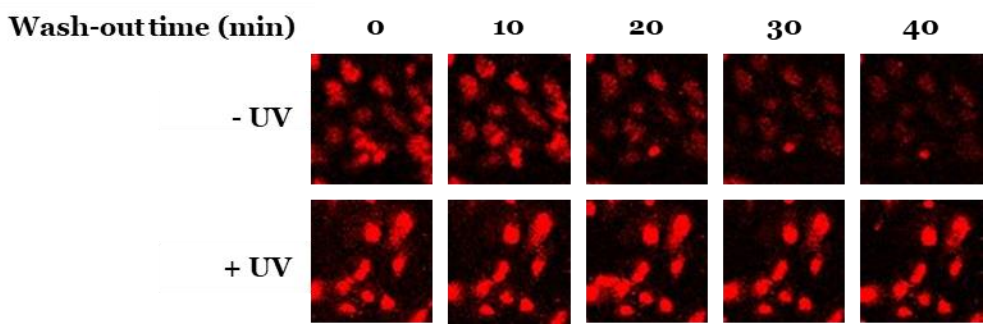


Figure 2-8 Washout experiment. Fluorescence microscopy images in transfected cells with or without UV irradiation. Increasing wash time produced a steady loss of fluorescence in cells without UV treatment, while maintenance of signal is indicative of successful photo-crosslinking of MG-diazirine with FAP.

2.7.8 Kinetic experiment for protein concentration

50 μM MG-diazirine was incubated at 37 ^{0}C in a solution containing different FAP concentrations for 30 min. After the incubation time, the solutions were then mixed 1:1 (v/v) with SDS denaturing loading dye and were heated to 90 ^{0}C for 10 min and allowed to cool down to room temperature. These samples were loaded on SDS-PAGE and the fluorescence of the bands were quantified using the same technique described in the manuscript.

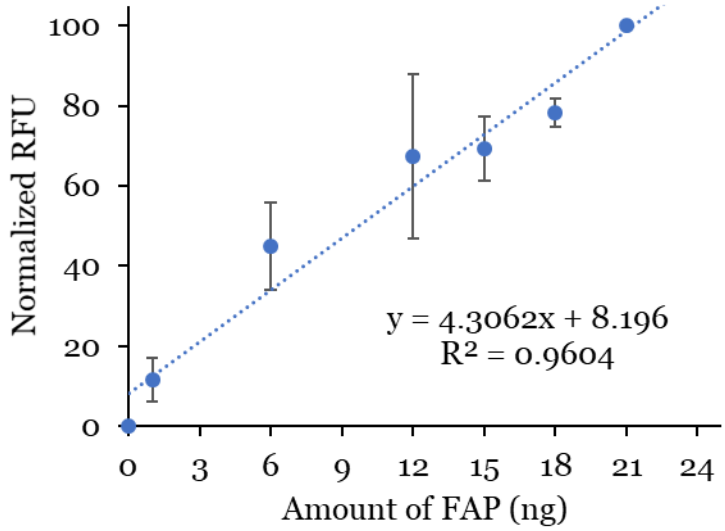


Figure 2-9 Protein concentration dependent kinetic experiment.

Table 2-1 Tabular data for protein concentration dependent kinetic experiment

FAP concentration (ng)	Average Normalized RFU	Standard deviation
0	0.000	0.000
1	11.621	5.588
6	45.025	10.903
12	67.384	20.615
15	69.342	8.058
18	78.356	3.528
21	100.000	0.000

2.7.9 MG-diazirine concentration dependent kinetic experiment

20 µg of FAP was dissolved in 1XPBS. MG-diazirine was added to this solution to give different final concentration of MG-diazirine. The solution was then incubated at 37 ^{0}C for 30 min. After the initial incubation time the protein was denature by missing 1:1 (v/v) with SDS denaturing loading dye and were incubated at 90 ^{0}C for 10 min and allowed to cool down to room temperature. These samples were loaded on SDS-PAGE and the fluorescence of the bands were quantified using the same technique describe in the manuscript.

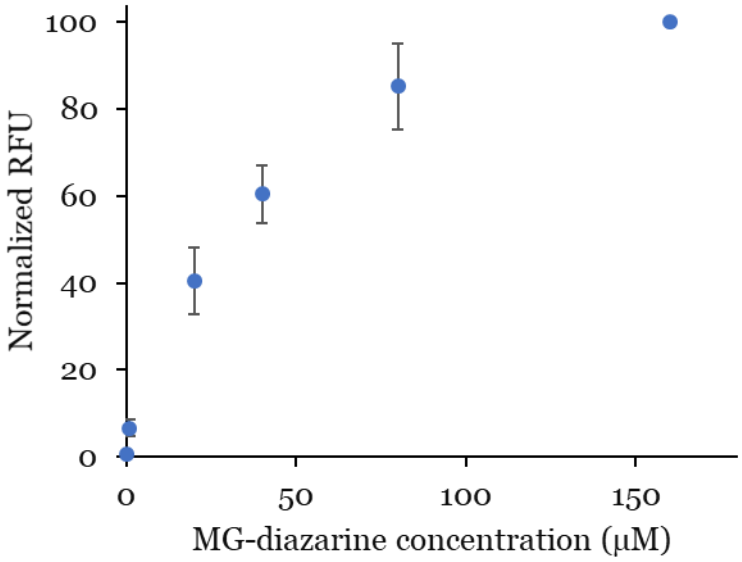


Figure 2-10 Kinetic experiment for MG-diazirine concentration.

Table 2-2 Tabular data for protein concentration dependent kinetic experiment

MG-diazirine concertation (μM)	Average RFU	Standard deviation
0	0.609	0.542
1	6.653	2.057
20	40.474	7.762
40	60.360	6.617
80	85.145	9.752
160	100.000	0.000

2.7.10 Cell viability experiment

The effect of UV-irradiation on cell viability was tested using HeLa cells. The cells were grown on 96-well plates containing 100 μl media until they reached 90% confluency and were exposed to UV-irradiation for different amount to time. A well without cells containing 100 μl media was used as a negative control. After the UV-irradiation, the 96-well plate was placed back in 37 ^{0}C cell incubator for 12 h. Then 10 μl of WST-1 (4-[3-(4-lodophenyl)-2-(4-nitrophenyl)-2*H*-5-tetrazolio]-1,3-benzene disulfonate) (Thermo Fisher) was added to each well and allowed to incubate at 37 ^{0}C cell incubator for 4 h. After the 4h incubation, the plate was shaken thoroughly for 1 min on a shaker and 440 nm absorbance of each of the wells was measured.

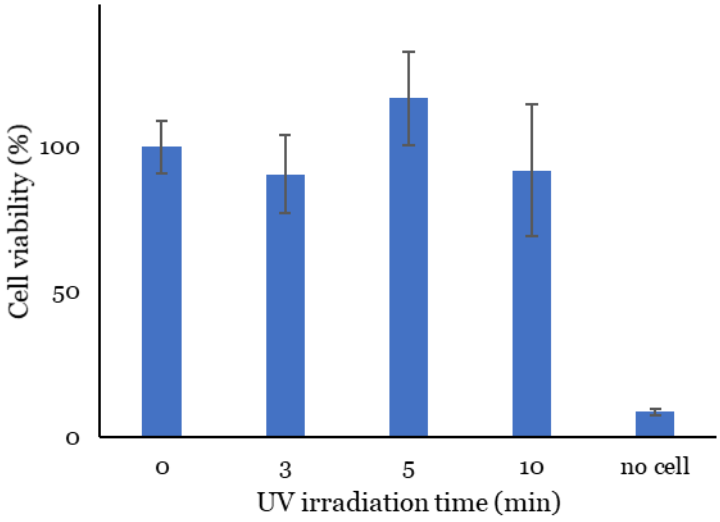


Figure 2-11 Cell viability

Table 2-3 Tabular data for cell viability experiment

UV-irradiation Time (min)	0	3	5	10	no cell
Normalized cell viability	100.000	90.570	116.700	91.886	8.755
Standard deviation	9.111	13.490	15.992	22.554	0.966

2.8 References

1. Toseland, C. P., Fluorescent labeling and modification of proteins. *J Chem Biol* **2013,** *6* (3), 85-95.

2. Crivat, G.; Taraska, J. W., Imaging proteins inside cells with fluorescent tags. *Trends Biotechnol* **2012,** *30* (1), 8-16.

3. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C., Green Fluorescent Protein as a Marker for Gene-Expression. *Science* **1994,** *263* (5148), 802-805.

4. Waldo, G. S.; Standish, B. M.; Berendzen, J.; Terwilliger, T. C., Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology* **1999,** *17* (7), 691-695.

5. Hertel, F.; Zhang, J., Monitoring of post-translational modification dynamics with genetically encoded fluorescent reporters. *Biopolymers* **2014,** *101* (2), 180-187.

6. Lippincott-Schwartz, J.; Snapp, E. L.; Phair, R. D., The Development and Enhancement of FRAP as a Key Tool for Investigating Protein Dynamics. *Biophysical Journal* **2018,** *115* (7), 1146-1155.

7. Dangol, S.; Singh, R.; Chen, Y. F.; Jwa, N. S., Visualization of Multicolored in vivo Organelle Markers for Co-Localization Studies in Oryza sativa. *Molecules and Cells* **2017,** *40* (11), 828-836.

8. Tsien, R. Y., The green fluorescent protein. *Annual Review of Biochemistry* **1998,** *67*, 509-544.

9. Das, S. C.; Panda, D.; Nayak, D.; Pattnaik, A. K., Biarsenical Labeling of Vesicular Stomatitis Virus Encoding Tetracysteine-Tagged M Protein Allows Dynamic Imaging of M Protein and Virus Uncoating in Infected Cells. *Journal of Virology* **2009,** *83* (6), 2611-2622.

10. Doyon, J. B.; Zeitler, B.; Cheng, J.; Cheng, A. T.; Cherone, J. M.; Santiago, Y.; Lee, A. H.; Vo, T. D.; Doyon, Y.; Miller, J. C.; Paschon, D. E.; Zhang, L.; Rebar, E. J.; Gregory, P. D.; Urnov, F. D.; Drubin, D. G., Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nature Cell Biology* **2011,** *13* (3), 331-U327.

11. Qin, J. Y.; Zhang, L.; Clift, K. L.; Hulur, I.; Xiang, A. P.; Ren, B. Z.; Lahn, B. T., Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *Plos One* **2010,** *5* (5).

12. Deer, J. R.; Allison, D. S., High-level expression of proteins in mammalian cells using transcription regulatory sequences from the Chinese hamster EF-1 alpha gene. *Biotechnology Progress* **2004,** *20* (3), 880-889.

13. Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y., New biarsenical Ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications. *Journal of the American Chemical Society* **2002,** *124* (21), 6063-6076.

14. Griffin, B. A.; Adams, S. R.; Tsien, R. Y., Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998,** *281* (5374), 269-272.

15. Stagge, F.; Mitronova, G. Y.; Belov, V. N.; Wurm, C. A.; Jakobs, S., Snap-, CLIP- and Halo-Tag Labelling of Budding Yeast Cells. *Plos One* **2013,** *8* (10).

16. Gautier, A.; Juillerat, A.; Heinis, C.; Correa, I. R.; Kindermann, M.; Beaufils, F.; Johnsson, K., An engineered protein tag for multiprotein labeling in living cells. *Chemistry & Biology* **2008,** *15* (2), 128-136.

17. Yao, J. Z.; Uttamapinant, C.; Poloukhtine, A.; Baskin, J. M.; Codelli, J. A.; Sletten, E. M.; Bertozzi, C. R.; Popik, V. V.; Ting, A. Y., Fluorophore Targeting to Cellular Proteins via Enzyme-Mediated Azide Ligation and Strain-Promoted Cycloaddition. *Journal of the American Chemical Society* **2012,** *134* (8), 3720-3728.

18. Fernandez-Suarez, M.; Baruah, H.; Martinez-Hernandez, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y., Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nature Biotechnology* **2007,** *25* (12), 1483-1487.

19. Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohane, R. F.; Urh, M.; Simpson, D.; Mendez, J.; Zimmerman, K.; Otto, P.; Vidugiris, G.; Zhu, J.; Darzins, A.; Klaubert, D. H.; Bulleit, R. F.; Wood, K. V., HatoTag: A novel protein labeling technology for cell imaging and protein analysis. *Acs Chemical Biology* **2008,** *3* (6), 373-382.

20. Reinhardt, U.; Lotze, J.; Morl, K.; Beck-Sickinger, A. G.; Seitz, O., Rapid Covalent Fluorescence Labeling of Membrane Proteins on Live Cells via Coiled-Coil Templated Acyl Transfer. *Bioconjugate Chemistry* **2015,** *26* (10), 2106-2117.

21. Bruchez, M. P., Dark dyes-bright complexes: fluorogenic protein labeling. *Current Opinion in Chemical Biology* **2015,** *27*, 18-23.

22. Chen, Y. C.; Clouthier, C. M.; Tsao, K.; Strmiskova, M.; Lachance, H.; Keillor, J. W., Coumarin-Based Fluorogenic Probes for No-Wash Protein Labeling. *Angewandte Chemie-International Edition* **2014,** *53* (50), 13785-13788.

23. Telmer, C. A.; Verma, R.; Teng, H. B.; Andreko, S.; Law, L.; Bruchez, M. P., Rapid, Specific, No-wash, Far-red Fluorogen Activation in Subcellular Compartments by Targeted Fluorogen Activating Proteins. *Acs Chemical Biology* **2015,** *10* (5), 1239-1246.

24. Peraro, L.; Deprey, K. L.; Moser, M. K.; Zou, Z.; Ball, H. L.; Levine, B.; Kritzer, J. A., Cell Penetration Profiling Using the Chloroalkane Penetration Assay. *Journal of the American Chemical Society* **2018,** *140* (36), 11360-11369.

25. Darabedian, N.; Gao, J. X.; Chuh, K. N.; Woo, C. M.; Pratt, M. R., The Metabolic Chemical Reporter 6-Azido-6-deoxy-glucose Further Reveals the Substrate Promiscuity of O-GlcNAc Transferase and Catalyzes the Discovery of Intracellular Protein Modification by O-Glucose. *Journal of the American Chemical Society* **2018,** *140* (23), 7092-7100.

26. Schneider, N.; Gabelein, C.; Wiener, J.; Georgiev, T.; Gobet, N.; Weber, W.; Meier, M., Genetic Code Expansion Method for Temporal Labeling of Endogenously Expressed Proteins. *Acs Chemical Biology* **2018,** *13* (11), 3049-3053.

27. Titeca, K.; Lemmens, I.; Tavernier, J.; Eyckerman, S., Discovering cellular protein-protein interactions: Technological strategies and opportunities. *Mass Spectrom Rev* **2019,** *38* (1), 79-111.

28. Jain, A.; Liu, R. J.; Ramani, B.; Arauz, E.; Ishitsuka, Y.; Ragunathan, K.; Park, J.; Chen, J.; Xiang, Y. K.; Ha, T., Probing cellular protein complexes using single-molecule pull-down. *Nature* **2011,** *473* (7348), 484-U322.

29. Szent-Gyorgyi, C.; Schmidt, B. A.; Creeger, Y.; Fisher, G. W.; Zakel, K. L.; Adler, S.; Fitzpatrick, J. A. J.; Woolford, C. A.; Yan, Q.; Vasilev, K. V.; Berget, P. B.; Bruchez, M. P.; Jarvik, J. W.; Waggoner, A., Fluorogen-activating single-chain antibodies for imaging cell surface proteins. *Nature Biotechnology* **2008,** *26* (2), 235-240.

30. Yan, Q.; Schmidt, B. F.; Perkins, L. A.; Naganbabu, M.; Saurabh, S.; Andreko, S. K.; Bruchez, M. P., Near-instant surface-selective fluorogenic protein quantification using sulfonated triarylmethane dyes and fluorogen activating proteins. *Organic & Biomolecular Chemistry* **2015,** *13* (7), 2078-2086.

31. Zeng, G.; Wang, Y.; Bruchez, M. P.; Liang, F.-S., Self-Reporting Chemically Induced Protein Proximity System Based on a Malachite Green Derivative and the L5** Fluorogen Activating Protein. *Bioconjugate Chemistry* **2018,** *29* (9), 3010-3015.

32. Chan, E. W.; Chattopadhaya, S.; Panicker, R. C.; Huang, X.; Yao, S. Q. J. J. o. t. A. C. S., Developing photoactive affinity probes for proteomic profiling: hydroxamate-based probes for metalloproteases. **2004,** *126* (44), 14435-14446.

33. Tomohiro, T.; Kato, K.; Masuda, S.; Kishi, H.; Hatanaka, Y. J. B. c., Photochemical construction of coumarin fluorophore on affinity-anchored protein. **2011,** *22* (3), 315-318.

34. Bringmann, G.; Gampe, C. M.; Reichert, Y.; Bruhn, T.; Faber, J. H.; Mikyna, M.; Reichert, M.; Leippe, M.; Brun, R.; Gelhaus, C. J. J. o. m. c., Synthesis and pharmacological evaluation of fluorescent and photoactivatable analogues of antiplasmodial naphthylisoquinolines. **2007,** *50* (24), 6104-6115.

35. Saurabh, S.; Zhang, M.; Mann, V. R.; Costello, A. M.; Bruchez, M. P., Kinetically Tunable Photostability of Fluorogen-Activating Peptide–Fluorogen Complexes. *ChemPhysChem* **2015,** *16* (14), 2974-2980.

36. Saurabh, S.; Perez, A. M.; Comerci, C. J.; Shapiro, L.; Moerner, W. E., Super-resolution Imaging of Live Bacteria Cells Using a Genetically Directed, Highly Photostable Fluoromodule. *Journal of the American Chemical Society* **2016,** *138* (33), 10398-10401.

37. Hoffmann, J. E.; Dziuba, D.; Stein, F.; Schultz, C., A Bifunctional Noncanonical Amino Acid: Synthesis, Expression, and Residue-Specific Proteome-wide Incorporation. *Biochemistry* **2018,** *57* (31), 4747-4752.

38. Xing, W. W.; He, L.; Yang, H.; Sun, C. J.; Li, D. W.; Yang, X. M.; Li, Y.; Deng, A. P., Development of a sensitive and group-specific polyclonal antibody-based enzyme-linked immunosorbent assay (ELISA) for detection of malachite green and leucomalachite green in water and fish samples. *Journal of the Science of Food and Agriculture* **2009,** *89* (13), 2165-2173.

39. Uchinomiya, S.; Nonaka, H.; Wakayama, S.; Ojida, A.; Hamachi, I., In-cell covalent labeling of reactive His-tag fused proteins. *Chemical Communications* **2013,** *49* (44), 5022-5024.

40. Chen, Q.; Hu, X.; Zhang, D. D.; Chen, X. W.; Wang, J. H., Selective Isolation of Myosin Subfragment-1 with a DNA-Polyoxovanadate Bioconjugate. *Bioconjugate Chemistry* **2017,** *28* (12), 2976-2984.

Chapter 3: Imaging and tracking mRNA in live mammalian cells via fluorogenic photoaffinity labeling

This chapter was derived from a bioRxiv paper: *Imaging and tracking mRNA in live mammalian cells via fluorogenic photoaffinity labeling*

Tewoderos M. Ayele, Travis Loya, Arielle N. Valdez-Sinon, Gary J. Bassell, and Jennifer M. Heemstra

Published on February 11, 2020, available online:

https://doi.org/10.1101/2020.02.10.942482

3.1 Abstract

Cellular RNA labeling using light-up aptamers that bind to and activate fluorogenic molecules has gained interest in recent years as an alternative to protein-based RNA labeling approaches. Aptamer-based systems are genetically encodable and cover the entire visible spectrum. However, the relatively weak nature of the non-covalent aptamer-fluorogen interaction limits the utility of these systems in that multiple copies of the aptamer are often required, and in most cases the aptamer must be expressed on a second scaffold such as a transfer RNA. We propose that these limitations can be averted through covalent RNA labeling, and here we describe a photoaffinity approach in which the aptamer ligand is functionalized with a photoactivatable reactive group such that irradiation with UV light results in covalent attachment to the RNA of interest. In addition to the robustness of the covalent linkage, this approach benefits from the ability to temporally control RNA labeling. To demonstrate this method, we incorporated a photoaffinity linker onto malachite green and fused the malachite green aptamer to a specific mRNA reporter of interest. We observed markedly improved sensitivity for fixed cell imaging of mRNA using this approach compared to *in situ* hybridization. Additionally, we demonstrate visualization of RNA dynamics in live cells using an mRNA having only a single copy of the aptamer, minimizing perturbation of the structure and localization. Our initial biological application utilizes the photoaffinity labeling approach to monitor RNA stress granule dynamics and we envision future application of this method for a wide range of investigations into the cellular localization, dynamics, and protein binding properties of cellular RNAs.

3.2 Introduction

Trafficking of messenger RNA (mRNA) to subcellular compartments plays an essential role in RNA homeostasis and cellular function. This spatiotemporal control of mRNA localization is a common characteristic for a significant fraction of transcripts,[1-3] and in recent years, fluorescent microscopy has dramatically increased our understanding of the heterogeneity of transcript regulation and the complex subcellular interactions of RNAs and proteins. However, this relies on the ability to fluorescently label cellular RNAs without significantly perturbing their structure or localization. The earliest approaches to fluorescently tagging cellular RNAs utilized probes capable of binding to the RNA of interest (ROI) through Watson-Crick-Franklin base pairing, including fluorescent *in situ* hybridization (FISH) and molecular beacons. While these methods yielded much of the current day knowledge on RNA localization, they generally require cell fixation, and thus cannot provide insight into trafficking and dynamics of cellular RNAs.[4-6] Currently, the most ubiquitous approach for visualizing mRNA uses GFP-fused RNA binding proteins such as MS2, λ_N, PCP, or Cas proteins.[7-10] These fluorescently-tagged proteins recognize a specific sequence that is incorporated multiple times onto ROI. While this does enable visualization of RNAs in living cells, these methodologies suffer from the fact that the unbound fluorescent protein creates significant background signal. This necessitates functionalization of the ROI with multiple copies of the target RNA sequence, and the size of that sequence as well as the heavy load of the associated proteins (>1300 kDa) can alter the native localization and functional properties of the RNA.[11, 12]

In 2003, Tsien and coworkers proposed that the small-molecule recognition capabilities of RNA could potentially be used for RNA labeling, and they reported an RNA

aptamer that binds to the fluorogenic dye malachite green (MG). Over the past decade, a number of other RNA aptamers binding to fluorogenic dyes have been reported, including the Spinach[13, 14], Broccoli[15], and Mango[16] aptamers. These aptamers have been fused to RNAs and expressed in cells to enable RNA visualization. However, these approaches remain predominantly used in bacterial cells and typically require fusion of multiple copies of the aptamer to enable fluorescence imaging. Recently, Yang and coworkers reported the Peppers aptamer system, which has a high signal-to-background ratio and was used to image genomic loci in mammalian cells.[17] While the stability and cellular brightness of the Peppers aptamer was a significant improvement compared to the previously reported aptamer-based systems, this approach still suffers from inherent limitations as a result of its non-covalent nature. For example, the reversibility of the fluorescent probe interaction with the ROI makes Peppers and other non-covalent systems unusable for applications that require media exchange, as labeling does not withstand washout steps. This inherent limitation of non-covalent approaches also makes them unsuitable for time-resolved investigations such as pulse-chase experiments.

We hypothesized that the limitations of the current RNA labeling approaches could be overcome through covalent attachment of the fluorescent molecule to the target RNA. Compared to all of the existing methods, which rely on non-covalent binding, covalent attachment would provide increased robustness to maximize signal-to-background and would allow the labeling to withstand media exchange or washing steps. Additionally, we envisioned that using a photoactivatable reactive group would provide temporal control over the labeling process, which is not possible using any of the existing methods. To achieve this goal, we utilized the malachite green aptamer (MGA) first reported by Tsien and coworkers.[18] Similar to the Peppers aptamer, MGA binds to the MG fluorogen and

induces significant red-shifted fluorescence enhancement.[19] The excitation and emission maxima for the MG fluorogen are also located in the far-red region of the UV spectrum, averting the inherent challenges associated with cellular auto-fluorescence and making this aptamer-ligand pair exceptionally well-suited for live cell imaging.

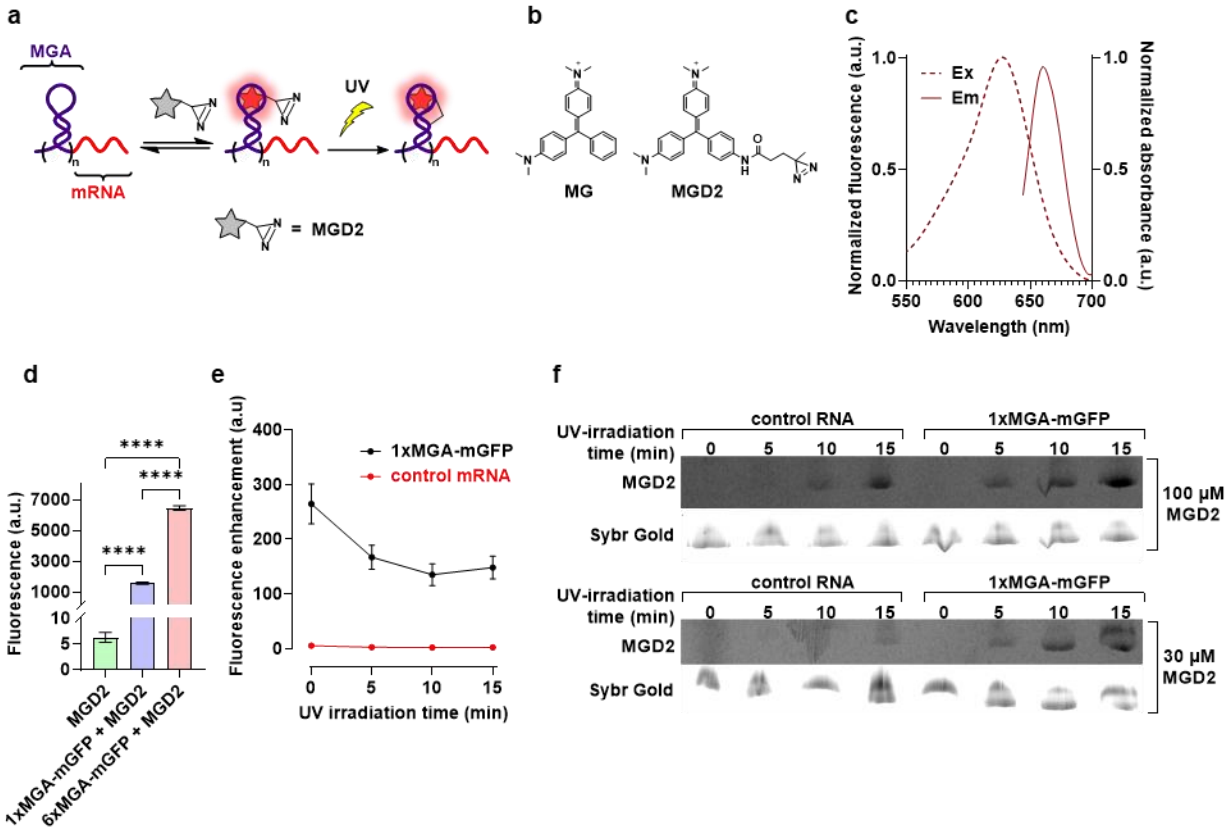


Figure 3-1 Characterization of MGA-functionalized mRNA in the presence of MGD2 a, Schematic representation of fluorogenic photoaffinity labeling of MGA-functionalized mRNA. The MGA-functionalized mRNA binds to the fluorogenic dye and induces fluorescence enhancement. UV-irradiation results in covalent attachment of the dye to the ROI. b, Structures of MG and MGD2. The canonical MG molecule was functionalized with a diazirine linker to enable photoaffinity labeling of MGA. c, Emission (solid) and excitation (dashed) spectra of MGD2 bound to 1xMGA-mGFP. d, Fluorescence of MGD2, 1xMGA-mGFP, and 6xMGA-mGFP in 1xPBS. Statistical comparison was performed using

one-way ANOVA. (****P<0.0001). Bars and error bars represent the mean and standard deviation from n = 4 independent samples. e, Time-dependent fluorescence enhancement of 1xMGA-mRNA and control mRNA upon irradiation at 360 nm in 1x PBS. Points and error bars represent the mean and standard deviation from n = 3 independent samples. f, Denaturing PAGE gel analysis of UV-dependent covalent labeling of 1x MGA-mGFP and control GFP mRNAs with MGD2.

To covalently label the target RNA, we envisioned that the aptamer could be fused to the ROI, expressed in cells, and the cells incubated with the MG ligand having a photoactivatable handle. UV-irradiation would convert the non-covalent binding interaction into a covalent linkage, resulting in robust and temporally-controlled labeling of the ROI (Fig. 1a). To create the photo-reactive fluorogen, we appended a diazirine linker to the dye (Fig. 1b) and termed this new MG derivative malachite green diazirine-2 (MGD2). We previously used a similar approach to design a derivative of MG termed MGD for photoaffinity labeling of proteins in live cells;[20] however, this strategy has yet to be applied to RNA labeling. Upon UV-A irradiation at 365 nm, the diazirine linker is activated and produces a carbene moiety that readily reacts with nearby C-H and heteroatom-H bonds. UV-C (254 nm) irradiation has been used for cross-linking RNA-protein interactions by taking advantage of the photo-responsiveness of natural amino acids such as Cys, Lys, Phe, Trp, and Tyr.[21, 22] However, the longer wavelength of 365 nm used to activate MGD2 ensures that our design does not result in unwanted cross-linkage of RNA with cellular proteins.

Using our photoaffinity approach, we demonstrate that an mRNA of interest can be labeled and imaged in both fixed and live cells using a single 57 nt fusion. This is significantly smaller than the fusions required in other aptamer-based methods,

minimizing perturbation of RNA structure and localization. We demonstrate that covalent labeling enables RNA visualization under conditions where the non-covalent system fails, and we utilize our approach to monitor the dynamics of RNAs in stress granules. Together, this research introduces the first covalent method for cellular RNA labeling and provides an effective and easy-to-use tool for the RNA community. The added robustness and temporal control achieved using this approach is anticipated to significantly advance RNA imaging capabilities, providing important new insights into the role of RNA trafficking in biological processes such as development and disease.

3.3 Results

3.3.1 In vitro characterization

We first utilized *in vitro* studies to investigate the reactivity and selectivity of the aptamer when functionalized to an mRNA. This was accomplished by transcribing the acGFP mRNA appended with one or six copies of the aptamer sequence at the 5' UTR, which will be referred to as 1xMGA-mGFP and 6xMGA-mGFP, respectively. Both MGA-functionalized mRNAs displayed a well-defined absorbance and fluorescence profile with an excitation maximum at 625 nm and an emission maximum 660 nm in the presence of MGD2 (Fig. 1c). Prior to UV-irradiation, fluorescence measurements revealed a 251-fold fluorescence enhancement for MGD2 bound to 1xMGA-mGFP and >1000-fold fluorescence enhancement for MGD2 bound to 6xMGA-mGFP (Fig. 1d). We next investigated the effect of UV-irradiation on the fluorescence enhancement. For this experiment, we used 1xMGA-mGFP and a control mGFP mRNA that does not contain the aptamer sequence (Fig. 1e). We observed that up to 15 minutes of UV-irradiation did not

result in any detectable fluorescence enhancement of MGD2 in the presence of the control mRNA. However, the fluorescence enhancement of 1xMGA-mGFP declined slightly over time and stabilized at approximately 140-fold after 15 minutes of UV irradiation. This decrease in enhancement was not entirely unexpected, as covalent attachment may limit the fluorogen to binding modes that have slightly less rotational restriction, but nonetheless we remained encouraged by these data as they indicated that a single copy of MGA could produce a detectable and stable signal-to-background ratio for cellular imaging experiments.

We also investigated the specificity of labeling using denaturing PAGE analysis. Both the control GFP mRNA lacking MGA (4 μM) and 1xMGA-GFP mRNA (4μM) were incubated with MGD2 and UV-irradiated for different lengths of time (Fig. 1f). We observed non-specific labeling of the control mRNA after 15 minutes of irradiation in the presence of 30 μM MGD2 and after 10 minutes of irradiation in the presence of 100 μM MGD2. These data were somewhat surprising given the lack of fluorescence enhancement observed for the control RNA in the previous experiment. However, this observation can be explained by considering that in the case of non-specific labeling, the energy from the absorbed light is dissipated through nonradiative rotational relaxation of the phenyl groups of MGD2.[18, 23] However, the tight target-specific binding of the aptamer restricts the rotational relaxation of the dye and results in the enhanced fluorescence output. Thus, while our approach does result in some non-specific RNA labeling, we anticipated that this would not create problematic background signal during imaging experiments. We were further encouraged to this notion upon testing the selective reaction of MGD2 (30 μM) with 6xMGA-mGFP in the presence of cellular RNA (Supplementary Fig. 1). After 10 min of UV irradiation, we observed bands in the MG channel corresponding to the target

RNA and slight impurities, but no bands from labeling of other cellular RNAs. The collective observations from these experiments served as guidelines for irradiation time and dye concentrations used in subsequent cellular labeling experiments.

3.3.2 Fixed-cell imaging of RNA.

Having demonstrated our photoaffinity labeling approach *in vitro*, we next turned to fixed cell experiments, as this would enable us to directly compare our method to FISH and validate labeling of the target RNA in cells. As a biological context for testing our labeling method, we chose stress granules. In response to stress conditions, cells form non-membrane-bound cytosolic and nuclear RNA-protein assemblies to stall the translation of mRNA until the cells are no longer under stress. While most mRNAs can be concentrated to stress granules, different mRNAs have vastly distinct localization efficiencies.[5, 24, 25] Using FISH, Parker and coworkers showed that the CDK6 mRNA is highly enriched in stress granules of mammalian cells.[5] Inspired by this finding, we used mCDK6 as a model system to fluorescently label and image the unique distribution pattern of the mRNA. Both the 6x and 1x MGA sequences were inserted at the 5' UTR of the transcript, and the construct was placed under the control of the cytomegalovirus (CMV) promoter (Fig. 2a). Mouse neuroblastoma Neuro-2a cells were then transfected with these plasmids and incubated with 30 μM MGD2 for 20 min. The cells were then irradiated using 365 nm UV light to allow for covalent labeling of the aptamer. The media was replaced to washout excess dye, and cell stress conditions were induced by 45 min of arsenite exposure, a well characterized paradigm to induce formation of stress granules. Following fixation and immunofluorescence labeling of one stress granule marker protein (G3BP1), the cells were imaged using confocal microscopy.

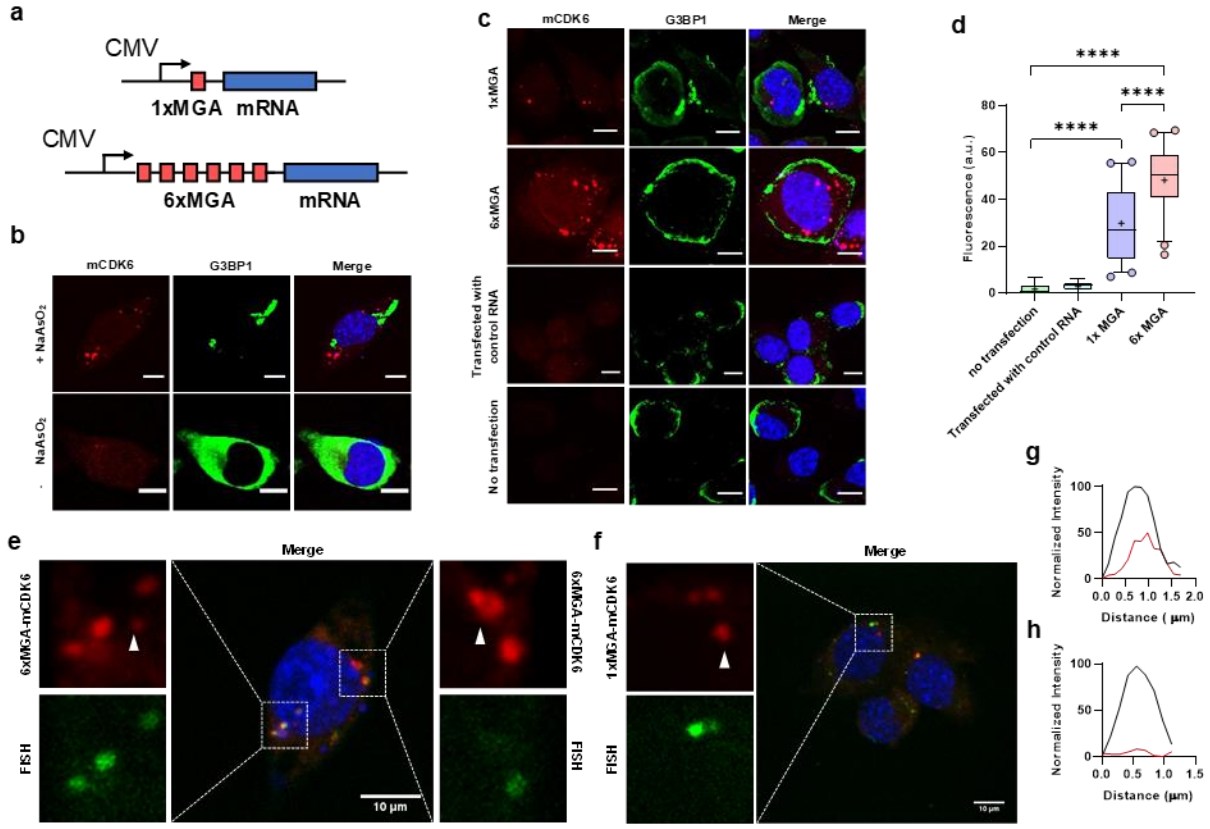


Figure 3-2 RNA labeling in fixed cells a, Schematic representation of MGA-functionalized mRNA constructs. b, Representative confocal images of cellular granule formation in Neuro-2a cells with and without $NaAsO_2$ stress. For this experiment, 6xMGA was used to label mCDK6 RNA. c, Representative confocal imaging of Neuro-2a cells transfected with 1xMGA-mCDK6, 6xMGA-mCDK6, control mCDK6, and cells that were not transfected. For figures 2b and c, G3BP1 protein immunolabeling was used to see the formation of granules. d, Fluorescence intensity of RNA foci in untransfected Neuro-2a cells or Neuro-2a cells expressing mCDK6 functionalized with 1xMGA or 6xMGA at the 5'UTR. (n = 6 foci for no transfection and transfection with control RNA, n = 50 foci for 1xMGA and 6xMGA). Statistical comparison was performed using one-way ANOVA. (****P<0.0001). Box plots show median, upper and lower quartiles, whiskers extending

to 5th and 95th percentile, and mean represented by a cross sign. e, Confocal mCDK6 RNA labeling in Neuro-2a cells using 6xMGA and FISH. f, Confocal mCDK6 RNA labeling in Neuro-2a cells using 1xMGA and FISH. White arrow represents RNA granules detected by 6xMGA but not by FISH. Scale bars represent 10 μm. g, Representative normalized line-scan of colocalized FISH (red line) and MGA-mCDK6 (black line) labeling with MGD2. h, Representative normalized line-scan of RNA granules detected by MGA-functionalized mCDK6 (black line) but not with FISH (red line).

In Neuro-2a cells exposed to arsenite stress, we observed the formation of distinct cytoplasmic RNA granules with the 6xMGA-functionalized mCDK6 (6xMGA-mCDK6). In contrast, the signal from cells that were not exposed to arsenite was diffused uniformly throughout the cytoplasm, and no detectable stress granule enriched mCDK6 was observed (Fig. 2b). Encouraged by the ability to visualize RNA granules having six copies of the aptamer appended to the mRNA, we attempted to image cells that were transfected with 1xMGA-functionalized mCDK6 (1xMGA-mCDK6). After similar arsenite treatment, we observed that RNA granules could be detected with a single copy of the aptamer fused to the mRNA. Moreover, when arsenite-treated cells were not UV-irradiated, no detectable foci formation was observed, indicating that RNA visualization is dependent on covalent attachment of the probe to the aptamer. We were especially excited by this observation, as it validated our hypothesis that covalent RNA labeling would provide a more robust imaging method than the existing non-covalent approaches. To further validate the specificity of this system, we transfected cells with CDK6 lacking the MGA sequence and then incubated the cells with MGD2 and performed UV irradiation. These control cells did not show any labeling, indicating the absence of non-specific labeling of other cellular components (Fig. 2c). When comparing the stress granules detected with

1xMGA versus 6xMGA-functionalized mRNA, we did observe reduced fluorescence intensity in cytoplasmic mRNA granules (Fig 2c and d). However, the difference was only ~2-fold compared to the 6-fold smaller fusion of the 1xMGA construct, and the ability to detect RNA localization using a single aptamer fusion allows for minimal alteration of the target mRNA. Together, these results demonstrate that we are able to fluorescently label cellular mRNAs in a sequence-specific manner and observe their localization to stress granules.

3.3.3 Comparison of fixed cell imaging of MGA/MGD2 with FISH.

To validate that the fluorescence signal observed was arising from labeling of the target CDK6 mRNA, we simultaneously incubated the Neuro-2a cells with FISH probes complementary to the CDK6 sequence but bearing a spectrally orthogonal fluorophore. This experiment also enabled us to directly compare the sensitivity of the MGA/MGD2 system to the commonly used FISH technique for RNA labeling in fixed cells. Following arsenite stress, MGD2 labeling, and cell fixation, we incubated the cells with our custom FISH probes. Merged image analysis of the 1xMGA- and 6xMGA-labeled mRNA with the FISH signal showed an overlap of the fluorescence generated from these two approaches (Fig. 2e-g). Interestingly, we observed some RNAs by our MGA/MGD2 system that were not identified by FISH (Fig. 2e, f, and h). This enhanced sensitivity was observed for both 1xMGA- and 6xMGA-functionalized mRNAs, indicating that our approach has a higher sensitivity for RNA detection than FISH. We reason this is because during FISH labeling, the fluorescent probe is hybridized after the ROI is localized to granules. While this approach can identify ROIs that are spatially accessible to the FISH probes, other proteins and RNAs found within the granules compromise the ability of the probes to hybridize

with the ROI. In contrast to FISH, our approach ensures that the MGA-functionalized mRNA is labeled with the fluorescent reporter before it is localized to the granules. This important distinction in the timing of RNA labeling allows for the detection of mRNA that otherwise would be inaccessible for FISH labeling.

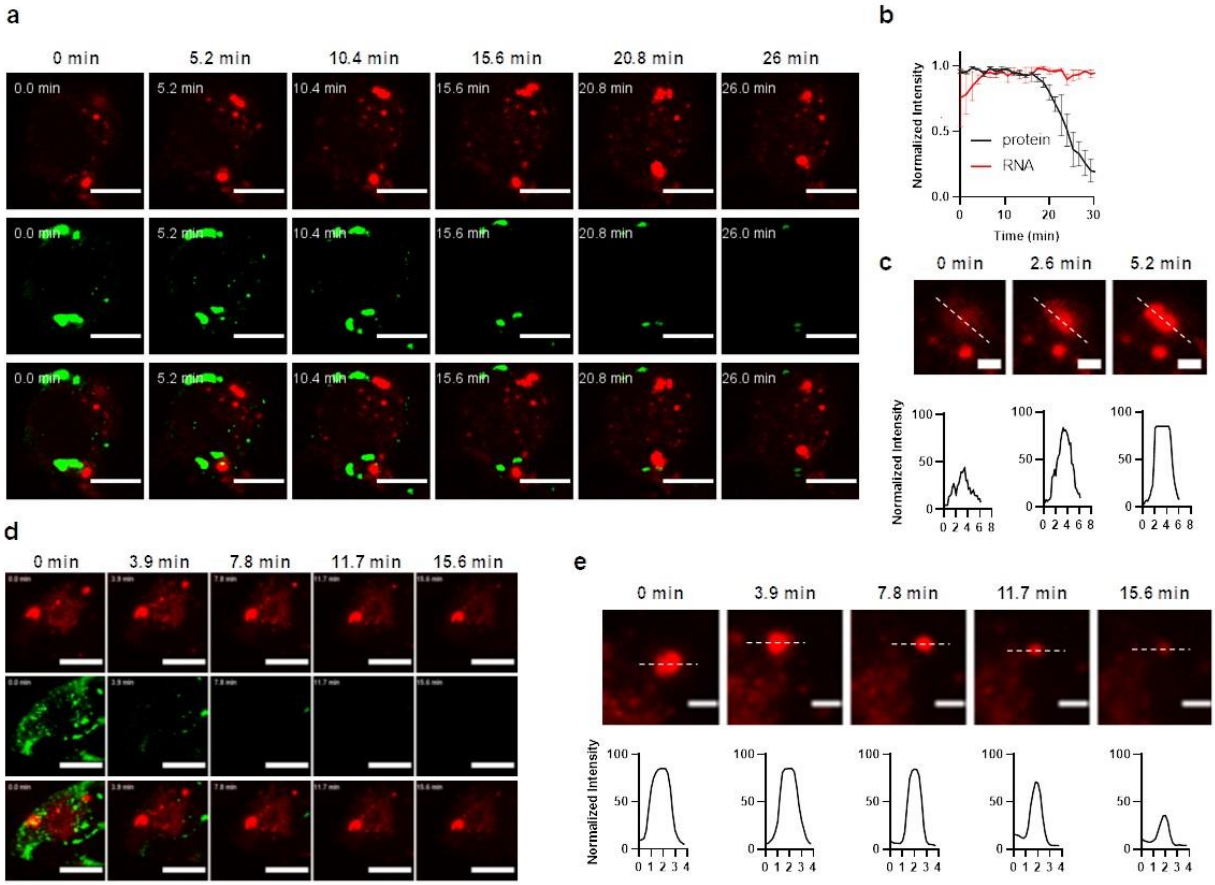


Figure 3-3 Live cell tracking of RNA and protein granules a, Real-time confocal microscopy images showing dynamic localization of mCDK6 (top panel), G3BP1 protein (middle panel), and merged (bottom panel) in Neuro-2a cells with 5.2 min intervals. Scale bars represent 10 μm. b, Fluorescence signal comparison of MGD2-labeled mRNA vs GFP-labeled G3BP1 protein. Line graph represents the mean and error bars represent SEM from n = 3 granules. c, Image showing phase separation and RNA granule

maturation. Top panel shows image slices taken every 2.8 min, and bottom panel is line-scan showing relative intensity of the dotted lines. Scale bar represents 2 μm d, Real-time confocal microscopy images showing granule deliquescence of mCDK6 (top panel), G3BP1 protein (middle panel), and merged (bottom panel) upon the addition of 5% 1,6-Hexanediol. Time slices were taken every 3.9 min. Scale bar represents 10 μm. e, Disappearance of RNA granule with 5% 1,6-Hexandiol treatment. Top panel shows confocal microcopy image of RNA granule and bottom panel shows relative intensity line-scan of the dotted lines. Scale bars represent 2 μm

3.3.4 Live cell imaging of mRNA.

Having established the applicability and robustness of our approach for both *in vitro* and fixed cell RNA imaging, we next investigated whether the MGD2/MGA system would be suitable for live cell imaging of endogenous mRNA localization and dynamics. Moreover, we sought to determine whether we could simultaneously track the real-time localization properties of both RNA and proteins within a living cell. For this assay, we co-transfected Neuro-2a cells with expression plasmids for 1xMGA-mCDK6 and dual GFP-tagged G3BP1 protein. The expressed mRNA was labeled with MGD2 and photo-crosslinked. The cells were then imaged to record the spatiotemporal features of both mRNA and protein in stress granules. After arsenite exposure, real-time confocal microscopy imaging of the granules revealed the dynamic nature of both the mRNA and the protein (Fig 3a and Supplemental video 1). Monitoring the signal intensity, we observed the fluorescent signal generated from the MGD2 labeled granules remained consistent over >25 min of imaging. However, the GFP signal showed a sharp signal decrease after 17 min of imaging (Fig 3a and b). Time-lapse imaging of the RNA granules

also showed a gradual phase separation and maturation of RNA granules (Fig 3c). These data highlight the ability of our approach to enable live-cell monitoring of RNA dynamics using only a single copy of the aptamer fusion and demonstrate the high photostability of the MG-aptamer pair.

After observing the motility and maturation of the granules, we questioned whether we could also observe their dissolution. In mammalian cells, stress granules disassemble in the presence of small aliphatic molecules that disrupt weak hydrophobic interactions.[26] Therefore, we added 1,6-hexanediol, an aliphatic alcohol commonly used for disassembly of stress granules.[26, 27] After arsenite-induced stress granule formation, we incubated the cells in 5% 1,6-hexanediol solution and observed the fluorescence signal of the labeled 1xMGA-mCDK6. In this experiment, signal from the protein granules disappeared after 7 min of incubation with 5% hexanediol. In contrast, most of the RNA granules exhibited a more sustained structural integrity, indicating that granules having high RNA content may be more resistant to dissolution (Fig 3d). In some RNA granules, however, hexanediol triggered an observable dissolution of these phase-separated compartments (Fig 3e and Supplemental video 2). This observation indicates that the strength of intermolecular forces of the RNA granule components is heterogeneous across different granules. Although we did not further investigate this property of RNA granules, the heterogeneity of the intermolecular forces of RNA granules is reported to be highly dependent on the composition of RNA, well-folded domains of proteins, and proteins having intrinsically disordered regions.[28] Therefore, these data indicate that our RNA labeling approach is able to validate previously held assumptions of RNA properties as well as uncover lesser-known physical and biological characteristics of these biomolecules.

3.4 Conclusion.

Fluorescent labeling and imaging of RNA is key to understanding its roles in cellular function and disease processes. While a number of methods have been reported for labeling RNA, these all suffer from limitations related to either the requirement for cell fixation, the need for a large fusion added to the RNA, or lack of robustness due to the weak nature of non-covalent interactions. We sought to develop a broadly applicable strategy that can be implemented for both fixed cell and live cell imaging and that would enable robust labeling with only a single small RNA fusion and fluorophore tag. To achieve these goals, we recognized that all of the current methods rely on non-covalent binding interactions, and that covalent tethering of the fluorescent probe and the RNA of interest could provide increased signal-to-background ratio with a smaller RNA fusion. Specifically, we envisioned a photoaffinity labeling approach, as this would also provide temporal control over the RNA labeling process. We modified the malachite green fluorogen to incorporate a photo-reactive diazirine linker, which allowed for covalent labeling of its cognate aptamer upon irradiation with UV light. By placing this aptamer at the 5' UTR of the mRNA, we showed target-specific fluorescence enhancement and labeling of the ROI. Fixed cell imaging of aptamer-functionalized mRNA showed formation of RNA stress granules in response to arsenite exposure. Compared to hybridization-based RNA labeling, we obtained enhanced sensitivity and lower background signal with our MGD2/MGA system. Furthermore, we showed that the dynamics of RNA granules containing a single aptamer-functionalized ROI can be tracked in live cells upon covalent attachment of the fluorogenic probe.

This novel strategy provides several advantages for RNA imaging applications. First, the far red-shifted fluorescence emission wavelength and the fluorogenic nature of

the MGD2 dye allows for minimal background signal generated from cellular autofluorescence and unbound small molecules, respectively. Second, the temporally controlled covalent attachment of the fluorogen to its cognate aptamer enables the labeling to withstand washout steps and allows tracking of RNAs labeled during a specific time window, a feature which is necessary for pulse-chase studies and other experiments that require media exchange. Third, a single aptamer fusion of 57 nt was sufficient to image RNA granules in both live and fixed cells. This is significantly smaller than the fusions required in other RNA labeling approaches, which typically append numerous copies of the respective tag and fluorescent probe or proteins, producing a fusion that can add thousands of kDa. Finally, this strategy is anticipated to be highly generalizable to enable the development of additional aptamer-photoaffinity probe combinations for multiplexed and multicolor imaging. Together, the research reported here significantly advances RNA labeling technology and introduces a robust and reliable tool for use by the RNA community to study basic mechanisms that underlie localization and dynamics of diverse types of RNA granules and how these mechanism go awry in disease models and other important biological contexts.

3.5 Methods and supplemental material

3.5.1 *In vitro* fluorescence enhancement

A solution of 30 μM of MGD2 was mixed with 4 μM of the corresponding mRNA in 1x PBS (1.54 mM Potassium monobasic, 155.17 mM Sodium Chloride, and 2.71 mM Sodium Phosphate dibasic with pH = 7.2) (ThermoFisher) and incubated at room temperature for 20 min. For comparison, a solution of MGD2 without any RNA was also

prepared in 1x PBS. The fluorescence of these solutions was measured using a BioTek Cytation 5 plate reader with ex = 620 ± 20 nm and em = 680 ± 20 nm. The average fluorescence value from replicate experiments was used to calculate fluorescence enhancement ($F_{enhancement}$) using Eq. 1, where $F_{MGA\text{-}mRNA}$ is the fluorescence of the solution containing both MGA-fused mRNA and MGD2 and F_{MGD2} is the fluorescence of the MGD2 solution.

$$F_{enhncement} = \frac{F_{MGA\text{-}mRNA}}{F_{MGD2}} \quad \text{(Eq. 1)}$$

3.5.2 Cell culture and transfection

Neuro-2a cells were used for all cellular experiments. These cells were cultured in DMEM (high glucose, Gibco) supplemented with 10% FBS at 37 ^{0}C in a humidified atmosphere of 5% CO_2 and 95% air. The cells were split every two days or once they reached > 85% confluency. All cellular imaging was done on Cellview cell culture slides (Greiner Bio-One cat. No. 543079). All transfections were done using the Lipofectamine 3000 (Invitrogen) transfection reagent following the manufacturer recommended protocol with minor modifications: a solution of 5 µl of Opti-MEM media (Gibco) and 0.3 µl of Lipofectamine 3000 was mixed with a solution of 5 µL of Opti-MEM, 0.3 µg of DNA, and 0.8 µL of P3000. This solution was incubated at room temperature for 20 min. The media from the cell culture wells was removed and the DNA lipofectamine mixture was added directly to each chamber containing 60-80% confluent cells. Immediately after, 90 µl of 37 ^{0}C warmed media was added to each well for a total of 100 µL of solution. The cells were then placed back into the cell culture incubator for 12 h before conducting further experimentation.

3.5.3 RNA FISH and MGA/MGD2 co-imaging

RNA FISH probes were designed against mCDK6 by using the Stellaris® RNA FISH Probe Designer (Biosearch Technologies, Inc) available online at www.biosearchtech.com/stellarisdesigner (Version 4.2). The synthesis of these probes was done in-house using a solid-phase oligonucleotide synthesizer (MerMade 12). The probes were then labeled with Fluorescein using the Label IT® nucleic acid labeling reagent (Mirus) using the manufacturer recommended protocol. Cells were incubated with 20 μM MGD2 in 37 ^{0}C prewarmed media for 20 min and irradiated with UV light at 365 nm for 10 min. The cells were then incubated in 0.5 mM sodium arsenite for 45 min at 37 ^{0}C. Cells wells were fixed with 4% paraformaldehyde (Biotium) for 15 min at room temperature and permeabilized with 0.1 % Triton X-100 (Sigma) for 1h. FISH probes were hybridized to the target mRNA for 12 h at 37 ^{0}C using Stellaris® hybridization buffer containing 10% formamide. Cells were washed with 200 μL Wash Buffer A (Stellaris®) for 30 min in the dark followed by a wash with 200 μL Wash Buffer B for 5 min. Cells were then imaged in Vectashield® antifade mounting media with DAPI (Vector Laboratories).

3.5.4 Immunofluorescence and MGA/MGD2 co-imaging

Immunofluorescence and MGA/MGD2 imaging on fixed Neuro-2α cells were performed following transfection, MGA/MGD2 labeling, and cell fixation protocol described in the **RNA FISH and MGA/MGD2 co-imaging** methods section above. After fixing the MGD2 labeled cells with 4% formaldehyde, the cells were permeabilized for 1 h using blocking buffer containing: 5% rabbit serum (Millipore), 0.1% bovine serum albumin (Millipore), and 0.1% Triton (Sigma) in 1x PBS. The blocking buffer was then

removed and replaced with 1:200 (v/v) diluted primary rabbit anti-G3BP1 antibody (Cell Signaling Technology® # 17798). The primary antibody was diluted in dilution buffer containing: 1% bovine serum albumin (Millipore), and 0.1 % Triton X-100 (Sigma) in 1x PBS. After 12 h incubation with the primary antibody buffer solution, the buffer was removed and cells were washed for three times, 5 min each, with 1x PBS. The cells were then incubated in a solution of secondary goat anti-rabbit IgG H&L (Alexa Fluor® 488) antibody (abcam #150077) for 1 h. The secondary antibody solution was prepared by diluting the secondary antibody to 1:200 (v/v) in the same dilution buffer as above. The cells were then washed three times for 5 min each with 1x PBS and imaged in Vectashield® antifade mounting media with DAPI (Vector Laboratories).

3.5.5 Confocal microscopy

Live and fixed cell images were taken on Cellview cell culture slides (Greiner Bio-One cat. No. 543079) using a Leica SP8 confocal laser scanning microscope equipped with an HC Plan Fluotar x10/0.15 air objective, an HC PL APO CS2 x20/0.75 air objective, an HC PL APO CS2 x63/1.4 oil objective, and two HYD GaAsP detectors. 405 nm Argon laser excitation was used to image DAPI; 488 nm Argon laser excitation was used to image Alexa488 labeled secondary antibody and FAM labeled FISH probes; 633 nm Helium-Neon laser was used to image MGD2.

Cellular images were obtained by taking Z-stack images with the instrument optimized step size and enough steps to cover the depth of each cell. Gain for each channel was optimized to minimize oversaturation while obtaining a clear fluorescent foci signal above background.

3.5.6 Live cell imaging

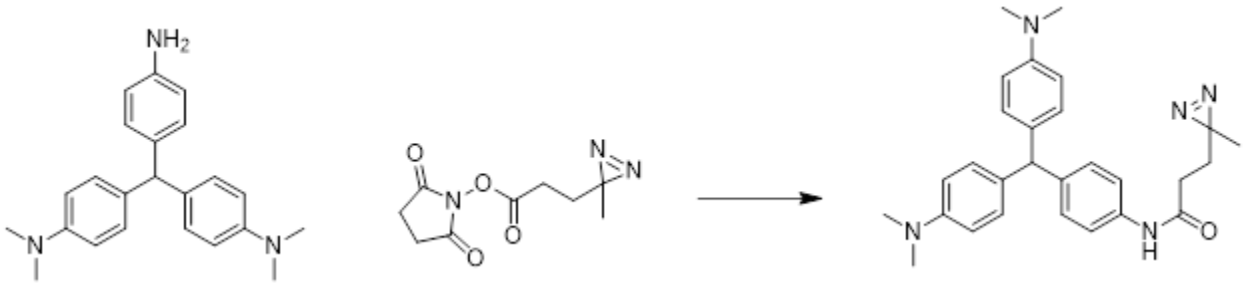
Cells were transfected with Phage Ubic G3BP1-GFP-GFP which was gifted from Jeffery Chao (Addgene plasmid # 119950; http://n2t.net/addgene:119950 ; RRID:Addgene_119950). Twelve hours after transfection, the cells were incubated in 30 μM MGD2 in media for 20 min followed by 10 min UV-irradiation. The media was then removed and replaced with 37 ^{0}C prewarmed media. The live cell images were acquired using the same microscope and settings outlined above with the following modifications: the 6-well cell culture slides were placed in an environmental chamber to control humidity and temperature during imaging, and images were taken using a 60x oil-immersion objective with instrument minimized framerate. The MTrackJ ImageJ plugin was used for granule tracking in live cells.

3.5.7 MGA array plasmid construction

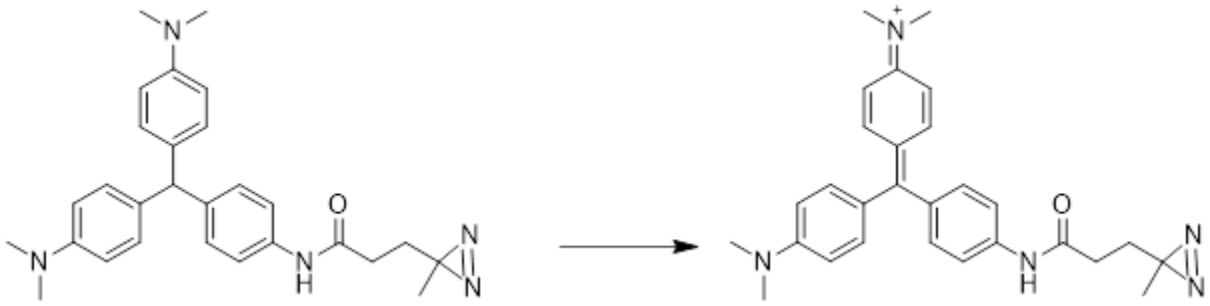
Single copy malachite green aptamer-containing plasmids were derived from pcDNA3.1(+) (Thermo), digested with *BamHI* and *NotI* and similarly cut acGFP from pAcGFP-N1-SialT[29] (pAcGFP1-N1-SialT was a gift from Lei Lu (Addgene plasmid #87324 ; http://n2t.net/addgene:87324 ; RRID:Addgene_87324) was inserted to create pcDNA3.1-acGFP. pcDNA3.1-acGPF was digested with *AgeI* and *XbaI* and inserted a similarly digested PCR product of mCDK6 from pcDNA3.1 mouse cdk6 wt (pcDNA3.1-mouse cdk6 wt was a gift from Martine Roussel (Addgene plasmid #75170; http://n2t.net/addgene:75170; RRID: Addgene_75170). Amplification was achieved using 5'- ATATATACCGGTACCATGGAGAAGGACAGCCT-3' and 5'-ATATATTCTAGAATCAGGCTGTGTTCAGCTCC-3', resulting in pcDNA3.1-mouse *Cdk6*

wt. The single copy malachite green aptamer[19] was inserted by digesting pcDNA3.1acGFP with *NheI* and *BamHI* and inserting an annealed oligo pair, 5'–CTAGCGGATCCCGACTGGCGAGAGCCAGGTAACGAATGGATCC-3' and 5'-GATCCGGATCCATTCGTTACCTGGCTCTCGCCAGTCGGGATCC-3' with compatible overhangs. The resulting vector, pcDNA3.1-1xMGA-acGFP was digested with *AgeI* and *XbaI* and inserted a similarly digested PCR product of mCDK6 from pcDNA3.1 mouse cdk6 wt. Amplification was achieved using 5'-ATATATACCGGTACCATGGAGAAGGACAGCCT-3' and 5'-ATATATTCTAGAATCAGGCTGTGTTCAGCTCC-3', resulting in pcDNA3.1-1xMGA-mouse *Cdk6 wt*. pcDNA3.1-6xMGA-acGFP was made by digesting pcDNA3.1acGFP with *NheI* and *AflII* and inserting similarly cut 6xMGA[30] PCR amplified from pUC57-6xMGA (Genscript) with 5'-ATATATGCTAGCTAGATGGTGTTTTGGTTTGG-3' and 5'-ATATATCTTAAGCGAATTCGGATCCGCG-3'. pcDNA3.1-6xMGA-mouse Cdk6 was made by digesting pcDNA3.1-6xMGA-acGFP with *AgeI* and *XbaI* and inserted a similarly digested PCR product of mCDK6 from pcDNA3.1 mouse cdk6 wt. Amplification was achieved using 5'- ATATATACCGGTACCATGGAGAAGGACAGCCT-3' and 5'-ATATATTCTAGAATCAGGCTGTGTTCAGCTCC-3', resulting in pcDNA3.1-6xMGA-mouse *Cdk6*.

3.5.8 Synthesis of MGD2



Leucomalachite green diazirine. The synthesis of *p*-amino-leucomalachite green was performed using the protocol described by Deng and coworkers.[31] In an oven dried 5 mL flask, p-amino-leucomalachite green (25.0 mg, 72.7 μmol) was dissolved in 1 mL of dry pyridine under inert gas. To this solution, 1.2 eq. of NHS-diazirine (19.6 mg, 86.83 μmol) was added and the solution was stirred at room temperature overnight. The reaction was then concentrated under reduced pressure to obtain dark green oil. The crude oil was dissolved in minimal amount of methanol and loaded on a preparative TLC with 30% EtOAc/MeOH as a mobile phase. The product band was scraped off from the preparative TLC and the product was filtered from the silica using MeOH and dried under reduced pressure. Yield = 11.6 mg, 35.31%. 1H NMR (500 MHz, DMSO) δ 0.99 (s, 3H), 1.64 (t, 2H, J = 7.57 Hz), 2.20 (t, 2H, J = 7.3 Hz), 3.04 (s, 12H), 5.63 (s, 1H), 7.03 (d, 2H, J = 8.3 Hz), 7.24 (d, 4H, J = 7.3), 7.52-7.69 (m, 6H), 10.15 (s, 1H). 13C NMR (500 MHz, DMSO) δ 19.36, 25.76, 29.53, 30.69, 44.52, 53.93, 119.21, 129.06, 130.14, 137.68, 169.75. LRMS (ESI-TOP) m/z Calcd for C28H33N5O [M + H]+ 456.2758; Found 456.2746



Malachite green diazirine (MGD2). 11.64 grams (25.55 μmol) of leucomalachite green diazirine was dissolved in 20% MeOH/EtOAc. To this solution, 1.2 eq. of chloranil (7.54 mg, 30.66 μmol) was added and the solution was stirred at room temperature for 3 h. The dark green solution was concentrated under reduced pressure and flash column purified using EtOAc to remove excess chloranil. Then mobile phase was then switched to 50% MeOH/DCM to isolate the crude product. The collected crude product was concentrated under reduced pressure and further purified by preparative TLC using 20% MeOH/DCM as mobile phase. Yield = 10.30 mg, 88.68%. 1H NMR (500 MHz, DMSO) δ 1.03 (s, 3H), 1.69 (t, 2H, J = 7.1 Hz), 2.37 (t, 2H, J = 7.3 Hz), 3.14 (s, 6H), 3.27 (s, 6H), 7.07 (d, 4H, J = 8.8 Hz), 7.31 (t, 6H, J = 9.3 Hz), 7.94 (d, 2H, J = 8.3 Hz). 13C NMR (500 MHz, DMSO) δ 19.35, 25.70, 28.93, 29.37, 40.39, 48.50, 113.63, 118.52, 126.24, 133.20, 136.25, 140.06, 156.23. LRMS (ESI-TOP) m/z Calcd. for C28H32N5O+ [M]+ 454.2601; Found 454.2595

3.5.9 Determination of MGD2 selectivity in cellular RNA.

To determine the selectivity of MGD2 labeling, we spiked in 6MGA-mGFP (200 ng) into cellular RNA extracted from HeLa cells (400 ng) in 1x PBS. To this solution, MGD2 (50 μM final concentration) was added and the solution was incubated for 20 min

at room temperature. After 10 min of UV irradiation, the sample was mixed with equal volume of 50% glycerol solution and heated to 70 ^{0}C to denature the RNA. This sample was loaded on to 1% agarose gel containing 1% v/v Clorox® bleach and Sybr Gold (Thermofisher). After running the gel for 1 h at a constant 100 V, the gel was frozen with dry ice for 10 min and imaged using GE Amersham Typhoon (Supplementary Fig 1). This freezing step enhanced the fluorescence output of the malachite green molecule. Analysis of this gel indicates that the MGD2 selectively labeled the aptamer functionalized mRNA and did not have any detectable labeling of cellular RNA.

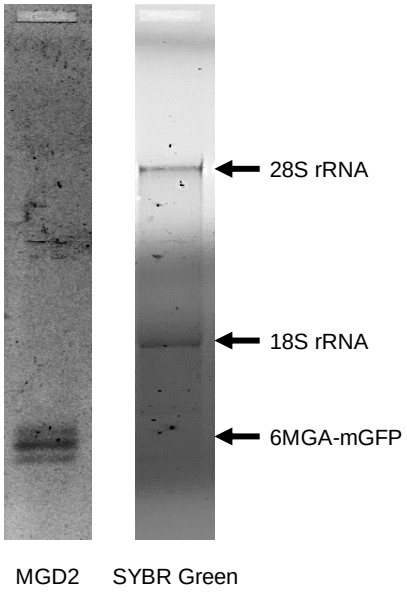


Figure 3-4 Selective labeling of MGA functionalized mRNA in the presence of cellular RNA extracted from HeLa cells.

Table 3-1 Tabular data fluorescence output of MGA array

	MGD2	1xMGA-mGFP + MGD2	6xMGA-mGFP + MGD2
Sample 1	5	1621	6594
Sample 2	7	1507	6342
Sample 3	7	1646	6568
Sample 4	6	1507	6312
mean	6.25	1570	6454
stdev	0.96	73	150

Table 3-2 Tabular data for UV dependent fluorescence enhancement of 1xMGA-mRNA compared to MGD2 in 1xPBS

1xMGA-mGFP

UV irradiation time (min)	Sample 1	Sample 2	Sample 3	mean	stdev
0	337.25	223.83	232.33	264	63
5	210.75	142.83	147.83	167	38
10	175	114	116.17	135	35
15	189.25	127.83	127.33	148	36

Table 3-3 Tabular data for UV dependent fluorescence enhancement of control mRNA compared to MGD2 in 1xPBS

control mRNA

UV irradiation time (min)	Sample 1	Sample 2	Sample 3	mean	stedv
0	7.5	4.1667	4.5	5.4	1.8
5	3.5	2	2.5	2.67	0.76
10	2.5	2.1667	2	2.22	0.25
15	3.5	2	1.6667	2.39	0.98

Table 3-4 Raw data table for fluorescence intensity of RNA foci in untrasfected Neuro-2a cells or Neuro-2a cells expressing mCDK6 functionalized with 1xMGA or 6xMGA at the 5'UTR

no transfection	Transfected with control RNA	1x MGA	6x MGA
6.7234	3.647	55.4355	50.7431
-0.0716	-2.631	30.9835	27.3111
2.0024	6.093	55.4635	42.1601
1.4524	2.766	55.4105	25.8611
0.0904	3.403	45.8785	61.7121
-0.2626	3.162	21.8995	65.4781
		35.7165	68.4491
		36.8405	48.6931
		48.8965	40.0551
		47.6815	24.4571
		27.4985	27.0291
		34.3785	36.3681
		43.5045	50.4931
		54.8165	45.1181
		48.3165	65.2311
		21.9635	48.2431
		19.3935	48.0551
		22.5045	54.8381
		21.8785	59.6751
		34.3165	52.9041
		42.9705	63.7431
		53.0815	69.4491
		9.0855	54.6871

24.7335	45.1491
55.9545	60.2431
38.8005	46.1751
34.9555	40.7911
42.0895	30.2591
26.3785	27.7431
17.2335	54.9141
11.8455	56.9251
41.0265	66.8281
35.8545	64.7431
38.6915	56.0101
48.0305	63.9621
14.6915	68.2431
11.7225	50.0081
8.6915	52.9251
6.9355	43.3881
14.4755	51.8051
8.9105	58.5821
14.0615	49.1811
14.4415	45.7431
23.6495	23.0291
17.8665	37.5051
15.8545	51.4101
13.7335	20.4571
15.3935	45.1001
11.3165	51.3971
11.5385	16.4021

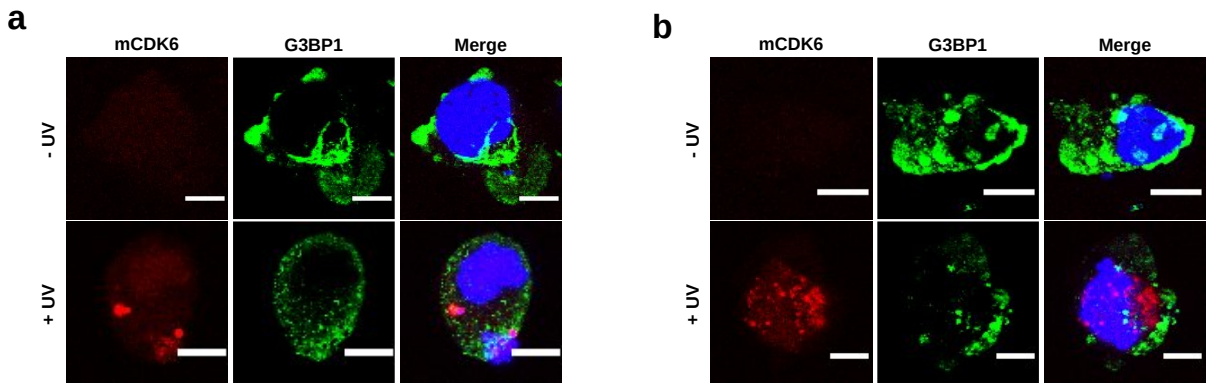


Figure 3-5 UV dependent labeling of RNA. a, Representative confocal images of RNA labeling with and without UV irradiation using cells transfected with 6xMGA. b, Representative confocal images of RNA labeling with and without UV irradiation using cells transfected with 1xMGA. G3BP1 protein immunolabeling was used to see the formation of granules. Scale bars represent 10 μm.

Table 3-5. Tabular data for fluorescence of RNA foci

	no transfection	Transfected with control RNA	1x MGA	6x MGA
Lower quartile	-0.119	1.42	14.6	40.6
Median	0.771	3.28	26.9	50.3
Upper quartile	3.18	4.26	43.1	58.9
5% Percentile	-0.263	-2.63	8.81	21.9
95% Percentile	6.72	6.09	55.5	68.3
Mean	1.66	2.74	29.7	48.2

Table 3-6 Tabular data for fluorescence signal comparison of GFP-tagged cellular protein and MGD2 labeled RNA granules

	Protein		RNA	
Time (min)	Mean	SEM	Mean	SEM
0	0.955	0.0116	0.756	0.220
1.334	0.952	0.0106	0.783	0.154
2.668	0.985	0.00747	0.846	0.112
4.002	0.974	0.00603	0.895	0.0371
5.337	0.940	0.0187	0.939	0.0231
6.671	0.982	0.0114	0.950	0.0394
8.005	0.957	0.0197	0.939	0.0284
9.339	0.972	0.0187	0.959	0.0151
10.673	0.974	0.00936	0.924	0.0492
12.007	0.944	0.0119	0.955	0.0322
13.342	0.944	0.0175	0.935	0.0344
14.676	0.924	0.0112	0.928	0.0603
16.01	0.942	0.0073	0.934	0.0540
17.344	0.908	0.0284	0.980	0.00924
18.678	0.873	0.0361	0.980	0.0133
20.012	0.799	0.0577	0.959	0.0226
21.347	0.708	0.0556	0.960	0.0098
22.681	0.622	0.155	0.970	0.0116
24.015	0.509	0.122	0.897	0.0419
25.349	0.362	0.129	0.928	0.0411
26.683	0.330	0.0933	0.933	0.0347
28.017	0.259	0.0964	0.960	0.00347
29.352	0.199	0.0863	0.938	0.0314

30.686	0.188	0.0906	0.948	0.0164

3.6 References

1. Martin, K. C.; Ephrussi, A., mRNA localization: gene expression in the spatial dimension. *Cell* **2009,** *136* (4), 719-730.

2. Lécuyer, E.; Yoshida, H.; Parthasarathy, N.; Alm, C.; Babak, T.; Cerovina, T.; Hughes, T. R.; Tomancak, P.; Krause, H. M., Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **2007,** *131* (1), 174-187.

3. Jansen, R.-P., mRNA localization: message on the move. *Nature reviews Molecular cell biology* **2001,** *2* (4), 247-247.

4. Hobro, A. J.; Smith, N. I., An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging. *Vibrational Spectroscopy* **2017,** *91*, 31-45.

5. Khong, A.; Matheny, T.; Jain, S.; Mitchell, S. F.; Wheeler, J. R.; Parker, R., The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Molecular cell* **2017,** *68* (4), 808-820. e5.

6. Lawrence, J. B.; Singer, R. H., Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell* **1986,** *45* (3), 407-415.

7. Rackham, O.; Brown, C. M., Visualization of RNA–protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *The EMBO journal* **2004,** *23* (16), 3346-3355.

8. Daigle, N.; Ellenberg, J., λ N-GFP: an RNA reporter system for live-cell imaging. *Nature methods* **2007,** *4* (8), 633.

9. Wu, B.; Chen, J.; Singer, R. H., Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Scientific reports* **2014,** *4*, 3615.

10. Yang, L.-Z.; Wang, Y.; Li, S.-Q.; Yao, R.-W.; Luan, P.-F.; Wu, H.; Carmichael, G. G.; Chen, L.-L., Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 Systems. *Molecular cell* **2019,** *76* (6), 981-997. e7.

11. Heinrich, S.; Sidler, C. L.; Azzalin, C. M.; Weis, K., Stem–loop RNA labeling can affect nuclear and cytoplasmic mRNA processing. *Rna* **2017,** *23* (2), 134-141.

12. Tutucci, E.; Vera, M.; Biswas, J.; Garcia, J.; Parker, R.; Singer, R. H., An improved MS2 system for accurate reporting of the mRNA life cycle. *nAture methods* **2018,** *15* (1), 81.

13. Huang, H.; Suslov, N. B.; Li, N.-S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A., A G-quadruplex–containing RNA activates fluorescence in a GFP-like fluorophore. *Nature chemical biology* **2014,** *10* (8), 686.

14. Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferré-D'Amaré, A. R., Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature structural & molecular biology* **2014,** *21* (8), 658.

15. Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R., Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society* **2014,** *136* (46), 16299-16308.

16. Trachman III, R. J.; Demeshkina, N. A.; Lau, M. W.; Panchapakesan, S. S. S.; Jeng, S. C.; Unrau, P. J.; Ferré-D'Amaré, A. R., Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Nature chemical biology* **2017,** *13* (7), 807.

17. Chen, X.; Zhang, D.; Su, N.; Bao, B.; Xie, X.; Zuo, F.; Yang, L.; Wang, H.; Jiang, L.; Lin, Q., Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nature biotechnology* **2019,** *37* (11), 1287-1293.

18. Babendure, J. R.; Adams, S. R.; Tsien, R. Y., Aptamers switch on fluorescence of triphenylmethane dyes. *Journal of the American Chemical Society* **2003,** *125* (48), 14716-14717.

19. Grate, D.; Wilson, C., Laser-mediated, site-specific inactivation of RNA transcripts. *Proceedings of the National Academy of Sciences* **1999,** *96* (11), 6131-6136.

20. Ayele, T. M.; Knutson, S. D.; Ellipilli, S.; Hwang, H.; Heemstra, J. M., Fluorogenic Photoaffinity Labeling of Proteins in Living Cells. *Bioconjugate chemistry* **2019,** *30* (5), 1309-1313.

21. Ule, J.; Jensen, K.; Mele, A.; Darnell, R. B., CLIP: a method for identifying protein–RNA interaction sites in living cells. *Methods* **2005,** *37* (4), 376-386.

22. Ule, J.; Jensen, K. B.; Ruggiu, M.; Mele, A.; Ule, A.; Darnell, R. B., CLIP identifies Nova-regulated RNA networks in the brain. *Science* **2003,** *302* (5648), 1212-1215.

23. Abedin, K. M.; Ye, J. Y.; Inouye, H.; Hattori, T.; Nakatsuka, H., Local dynamics in solid matrices investigated by malachite green optical microprobes. *Journal of luminescence* **1995,** *64* (1-6), 135-140.

24. Lavut, A.; Raveh, D., Sequestration of highly expressed mRNAs in cytoplasmic granules, P-bodies, and stress granules enhances cell viability. *PLoS genetics* **2012,** *8* (2), e1002527.

25. Thomas, M. G.; Loschi, M.; Desbats, M. A.; Boccaccio, G. L., RNA granules: the good, the bad and the ugly. *Cellular signalling* **2011,** *23* (2), 324-334.

26. Kroschwald, S.; Maharana, S.; Mateju, D.; Malinovska, L.; Nüske, E.; Poser, I.; Richter, D.; Alberti, S., Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. *elife* **2015,** *4*, e06807.

27. Wheeler, J. R.; Matheny, T.; Jain, S.; Abrisch, R.; Parker, R., Distinct stages in stress granule assembly and disassembly. *Elife* **2016,** *5*, e18413.

28. Van Treeck, B.; Parker, R., Emerging roles for intermolecular RNA-RNA interactions in RNP assemblies. *Cell* **2018,** *174* (4), 791-802.

29. Tie, H. C.; Mahajan, D.; Chen, B.; Cheng, L.; VanDongen, A. M.; Lu, L., A novel imaging method for quantitative Golgi localization reveals differential intra-Golgi trafficking of secretory cargoes. *Molecular biology of the cell* **2016,** *27* (5), 848-861.

30. Yerramilli, V. S.; Kim, K. H., Labeling RNAs in Live Cells Using Malachite Green Aptamer Scaffolds as Fluorescent Probes. *ACS Synthetic Biology* **2018,** *7* (3), 758-766.

31. Xing, W. W.; He, L.; Yang, H.; Sun, C. J.; Li, D. W.; Yang, X. M.; Li, Y.; Deng, A. P., Development of a sensitive and group-specific polyclonal antibody-based enzyme-linked immunosorbent assay (ELISA) for detection of malachite green and leucomalachite green in water and fish samples. *Journal of the Science of Food and Agriculture* **2009,** *89* (13), 2165-2173.

Chapter 4: Chemical Labeling and Affinity Capture of Inosine-Containing RNAs Using Acrylamidofluorescein

Tewoderos ayele contributed to the design and synthesis of Acrylamidofluorescein

This chapter was derived from a manuscript: *Chemical Labeling and Affinity Capture of Inosine-Containing RNAs Using Acrylamidofluorescein*

Steve D. Knutson, Tewoderos M. Ayele, and Jennifer M. Heemstra

Published on August 27, 2018, available online:

https://doi.org/10.1021/acs.bioconjchem.8b00541

4.1 Abstract

Adenosine-to-inosine (A-to-I) RNA editing is a wide-spread and conserved post-transcriptional modification, producing significant changes in cellular function and behavior. Accurately identifying, detecting, and quanti-fying these sites in the transcriptome is necessary to improve our understanding of editing dynamics, its broad-er biological roles, and connections with diseases. Chemical labeling of edited bases coupled with affinity enrichment has enabled improved characterization of several forms of RNA editing. However, there are no approaches currently available for pull-down of in-osines. To address this need, we explore acrylamide as a labeling motif and report here an acrylamidofluorescein reagent that reacts with inosine and enables enrichment of inosine-containing RNA transcripts. This method provides improved sensitivity in the detection and iden-tification of inosines towards a more comprehensive transcriptome-wide analysis of A-to-I editing. Acryla-mide derivatization is also highly generalizable, provid-ing potential for the labeling of inosine with a wide va-riety of probes and affinity handles.

4.2 Introduction

RNA is extensively edited after transcription. Adenosine-to-inosine (A-to-I) conversion is of one of the most common and impactful forms of editing and is catalyzed by adenosine deaminases acting on RNA (ADARs).[1] Resulting inosines base pair with cytidine and are effectively decoded as guanosine by cellular machinery. A-to-I editing occurs in both coding and non-coding RNA transcripts, eliciting dramatic changes in overall cellular function and behavior. Editing of mRNA can alter protein sequence through direct modification of codons or by altering splice sites and regulatory elements

in untranslated regions. A-to-I editing events are also extensive in non-coding RNAs, including microRNA and small-interfering RNA precursors, significantly altering their biosynthesis, trafficking, specificity, and gene regulation properties.[2-4] Accurately identifying A-to-I RNA editing sites in the transcriptome is necessary to improve our understanding of these modifications and their biological functions. A recently developed method to map A-to-I editing locations employs chemical modification of inosines with acrylonitrile to form *N*[1]-cyanoethylinosine (Figure 1a).[5, 6] Termed inosine chemical erasing sequencing (ICE-seq), this technique leverages the observation that inosine cyanoethylation inhibits Watson-Crick base pairing and effectively arrests reverse transcription at A-to-I editing sites. Resulting truncated cDNAs fail to undergo PCR amplification and are "erased" from RNA sequencing chromatograms, allowing bioinformatic detection of editing sites. Although ICE-seq has improved the accuracy and scalability of mapping and discovering A-to-I RNA editing sites, this method is also limited in sensitivity, as labeled inosine-containing transcripts cannot be enriched.

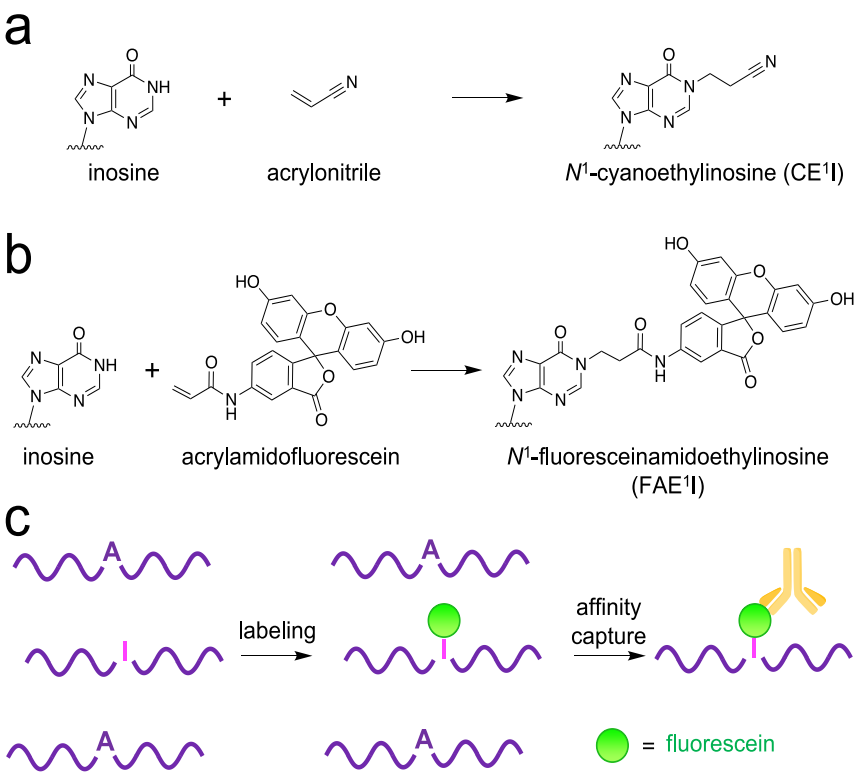


Figure 4-1 Chemical labeling of inosine. a, Acrylonitrile and b, acrylamidofluorescein produce N1 addition products. c, Acrylamidofluorescein labeling enables affinity capture of transcripts containing inosine.

Additionally, while millions of A-to-I sites have been identified across the human transcriptome, actual editing rates at these sites are highly variable and dependent on cellular and environmental cues, rendering them difficult to detect, characterize, and measure with these techniques. This is particularly true in coding RNAs, where I/A ratios can range anywhere from <0.001-5% depending on tissue type or external stimuli.[7-9] Together, these challenges mask the overall prevalence and true landscape of A-to-I RNA editing across the transcriptome.

The ability to enrich A-to-I edited transcripts from more complex total RNA samples would largely address this limitation and allow for deeper interrogation and

characterization of the epitranscriptome. Approaches using chemical labeling and/or antibody immunoprecipitation to capture edited transcripts have enabled significant advances in identifying and cataloging a number of other RNA modifications, including N^1- and N^6-methyladenosine, 5-methylcytidine, 5-hydroxymethylcytidine, and pseudouridine (Ψ).[10-18] While a previous study reported the production of antibodies targeting inosine for the enrichment of tRNAs, this method also displayed adsorptivity to other nucleobases and has not been further demonstrated in any other contexts.[19] Thus, no generally applicable methods currently exist for the derivatization and/or enrichment of inosines in RNA, significantly limiting both depth and sensitivity in identifying and studying A-to-I RNA editing dynamics across the transcriptome.

4.3 Results and discussion

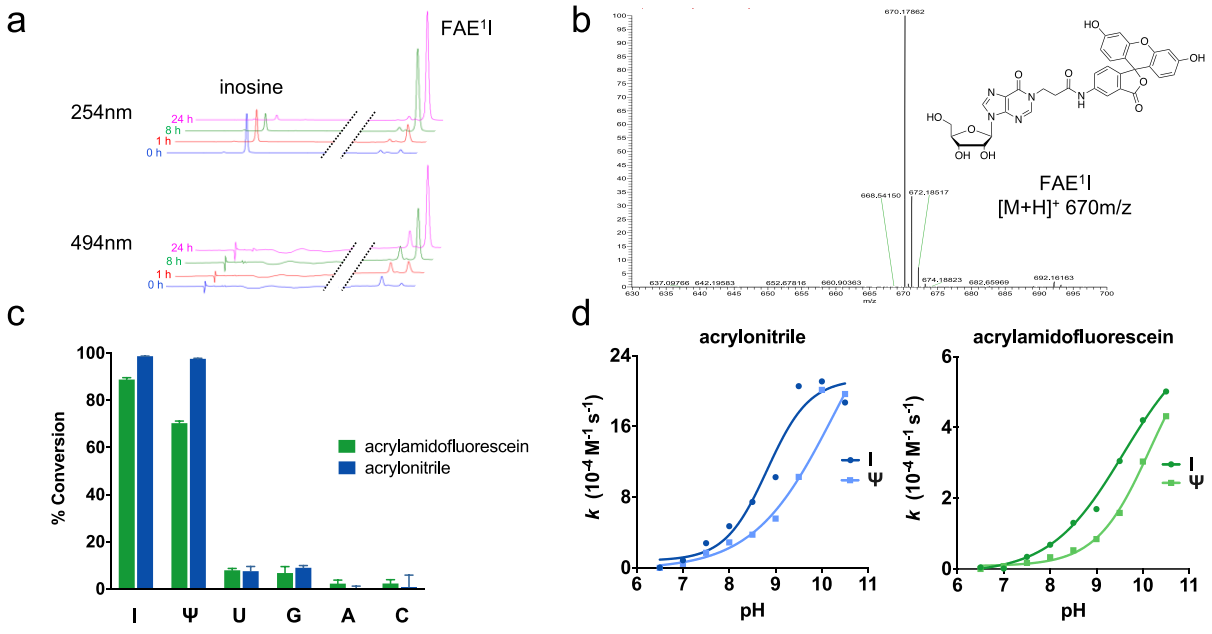


Figure 4-2 a, Representative HPLC traces depicting the reaction between inosine and acrylamidofluorescein over 24 hours. Disappearance of inosine (I) correlates with the appearance of a new putative N^1-fluoresceinacrylamidoethylinosine (FAE¹I) product

peak. b, ESI-MS analysis confirming mass identity of FAE^{1}I product. c, Reactivity panel of acrylonitrile and acrylamidofluorescein with ribonucleosides after 24 hours. d, Dependence of reaction rate constants on pH for the major reacting nucleosides inosine (I) and pseudouridine (Ψ).

In the design of a reagent for affinity capture of inosine-containing RNAs, we hypothesized that an acrylamide electrophile would provide similar reactivity towards inosine as acrylonitrile, while offering the structural flexibility to install an affinity handle for enrichment. To test this hypothesis, we carried out a facile synthesis to generate acrylamidofluorescein (Figure S1), as this reagent would provide both fluorescent labeling of inosines and the ability to perform affinity capture of A-to-I edited RNA transcripts using a commercially available anti-fluorescein antibody (Figure 1b,c).

After designing and synthesizing the acrylamidofluorescein reagent, we assessed initial labeling performance by reacting acrylamidofluorescein and acrylonitrile with each of the major ribonucleosides: inosine (I), pseudouridine (Ψ), uridine (U), guanosine (G), adenosine (A), and cytidine (C). Closely mimicking the ICE reaction conditions, a mixture comprising 50 mM ribonucleoside and 250 mM of either acrylonitrile or acrylamidofluorescein was prepared in 50:50 triethylammonium acetate:ethanol at pH 8.6. The solutions were incubated at 70 °C and the reaction was monitored by HPLC over 24 hours. As illustrated in Figure 2a, disappearance of inosine peaks is clearly shown along with the formation of a new product peak in both 254 nm and 494 nm chromatograms. This product peak was isolated and analyzed using ESI-MS and MS/MS analysis, confirming the identity of the predicted *N*^{1}-fluoresceinamidoethylinosine (FAE^{1}I) product (Figure 2b, S9). Using ribonucleoside peak areas in the chromatograms,

we determined the ratio of reacted vs unreacted ribonucleoside to calculate average conversion percentages for each base at various time points over 24 hours (Figures 2c, S7).

While acrylamidofluorescein and acrylonitrile exhibit similar reactivity trends, it is clear from the data that acrylonitrile has higher reaction efficiency (Figures 2d, S7). This is likely due to the difference in electron withdrawing properties between the two reagents, which contributes significantly to the kinetics of addition reactions.[20, 21] Given that the amide group is less withdrawing than the nitrile moiety, these results are then unsurprising. Regardless, acrylamidofluorescein and acrylonitrile display similar overall labeling selectivity, exhibiting major product formation with I and Ψ, minimal reactivity with U and G, and virtually no reactivity with A and C throughout extended reaction times. While both reagents display reactivity with Ψ, these observations are consistent with previous studies using acrylonitrile and serve to demonstrate the similar reactivity profiles of both acrylonitrile and acrylamidofluorescein. Indeed, the first reports of acrylonitrile-nucleoside labeling demonstrated its robust reactivity with N^1 on both inosine and Ψ.[20, 21]

To further validate addition of acrylamidofluorescein at N^1 of inosine, we assessed the effect of pH on reaction rates. Early characterizations of acrylonitrile reactivity with inosine showed that cyanoethylation is strongly pH dependent, suggesting N^1 deprotonation is required for reactivity. Similarly, the data in Figure 2d illustrate the direct correlation between reaction rate and pH and highlight the preferred reactivity with inosine at ~pH 8.5-8.6, consistent with the known p$\underline{K}_a$ values of N^1 for inosine (8.7)[22] and pseudouridine (9.5).[23] Taken together with the MS spectra, these results strongly support

the predicted N^1 addition to inosine and further suggest a similar labeling mechanism of acrylamidofluorescein compared with the well characterized chemistry of acrylonitrile.

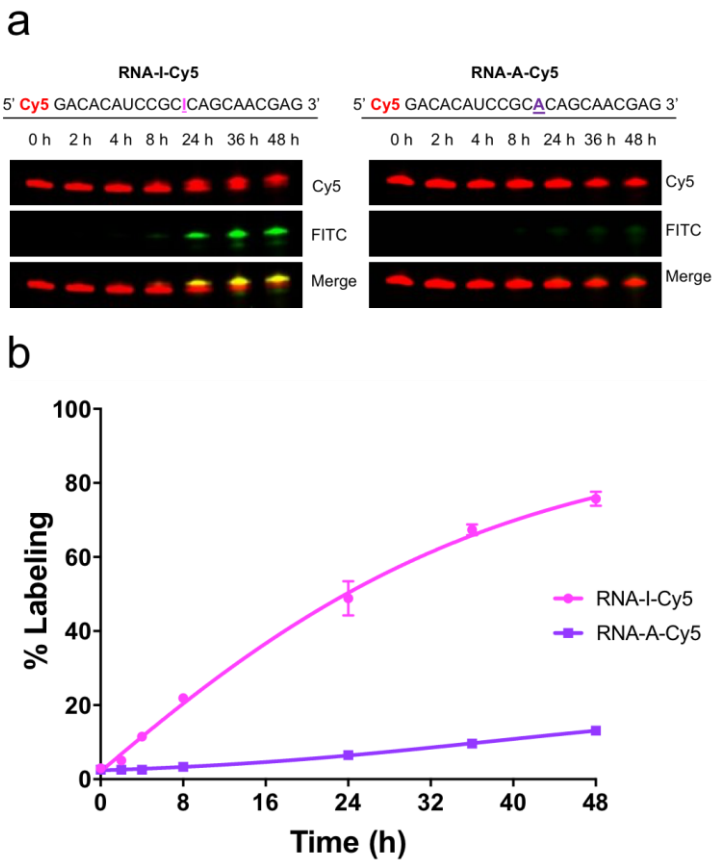


Figure 4-3 a, Denaturing PAGE analysis of synthetic oligoribonucleotides labeled with acrylamidofluorescein. b, Densitometric quantification of oligoribonucleotide labeling.

Given the promising results of our reagent with ribonucleosides, we next sought to demonstrate acrylamidofluorescein labeling of inosine in RNA oligoribonucleotides. As a test system for these studies, we chemically synthesized two short RNAs containing a 5' Cy5 fluorescent label and an adenosine (RNA-A-Cy5) or inosine (RNA-I-Cy5) at a defined position. We subjected each of these RNAs to acrylamidofluorescein labeling and denaturing PAGE analysis. As shown in Figure 3a, fluorescein labeling is clearly observed in RNA-I-Cy5 with increasing reaction times, and the labeled product exhibits a slight

decrease in migration rate. In comparison, only a faint signal is observed for RNA-A-Cy5, even after a 48 hour reaction time. Given that the presence of inosine is the only molecular difference between these two RNA strands, these data are indicative of selective fluorescein addition at this nucleotide position. Densitometric analysis was performed on the labeled RNA bands in the gels and normalized to standard amounts of fluorescein and Cy5-labeled control oligo nucleotides. These data were then used to calculate labeling yield as a function of reaction time (Figure 3b), which illustrates good selectivity for labeling of RNA-I-Cy5 compared to RNA-A-Cy5. This experiment also highlights the importance of reaction time in maximizing inosine labeling efficiency while maintaining selectivity, as we observe optimal RNA-I:RNA-A labeling ratios at approximately 24 hours. While longer RNA transcripts can undergo hydrolysis in mild alkaline conditions at elevated temperatures, these data demonstrate the stability of shorter RNA segments under our reaction conditions. We envision the use of this labeling method with high-throughput RNA-seq workflows, which require fragmentation of longer RNAs prior to library preparation and amplification. This fragmentation step is employed upstream of chemical labeling and pulldown in the analogous strategies described above for mapping other RNA modifications,[10, 11, 15-17, 24] and thus our results indicate compatibility with these platforms.

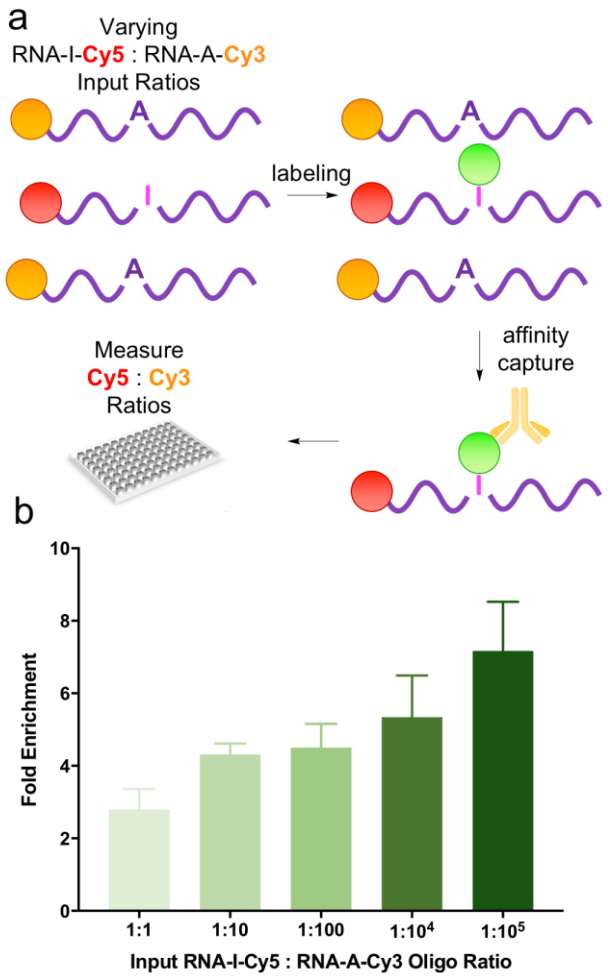


Figure 4-4 a, Workflow for quantifying pulldown efficiency with acrylamidofluorescein labeling and immunoprecipitation. b, Fold enrichment of inosine-containing oligoribonucleotides from varying mixtures.

Encouraged by these results, we sought to establish feasibility for our ultimate goal of enriching inosine-containing transcripts via immunoprecipitation (IP) of labeled oligonucleotides. To test this approach, we utilized the same RNA sequences from the previous experiment but labeled the inosine and adenosine variants with Cy5 and Cy3, respectively, to allow for simultaneous fluorescence-based quantification of each species. RNA-I-Cy5 and RNA-A-Cy3 were combined in varying ratios, subjected to

acrylamidofluorescein labeling, and then affinity captured using an anti-fluorescein monoclonal antibody and protein A/G magnetic beads. After extensive washing, bound oligoribonucleotides were eluted and quantified using a fluorescence plate reader (Figure 4a). Final concentrations of RNA-A-Cy3 and RNA-I-Cy5 after pull-down were compared to input ratios to calculate fold-enrichment. As shown in Figure 4b, acrylamidofluorescein labeling coupled with IP enables upwards of 7-fold enrichment of inosine-containing oligoribonucleotides, with the highest enrichment factors achieved for samples containing the lowest ratios of the inosine-containing RNA.

Chemical modification strategies coupled with affinity capture have significantly improved the sensitivity and accuracy in sequencing, mapping, and characterizing several modified RNA bases.[10-18] However, there are no extant methods for enriching A-to-I edited transcripts, greatly limiting our ability to understand the true scale and impact of A-to-I modifications on cell and tissue function. Here we address this challenge through the synthesis of a novel acrylamidofluorescein reagent that chemically labels inosine and enables the enrichment of A-to-I edited transcripts.

While the observed reactivity between acrylamidofluorescein and Ψ may seem problematic for the effective isolation and enrichment of inosine-containing transcripts from biological samples, Ψ is found predominantly in ribosomal RNAs and tRNAs, and thus effective fractionation of total RNA samples can remove significant quantities of this modified base. In coding RNAs, I also vastly outnumbers Ψ, with current estimates of ~500:1 I:Ψ.[7, 17, 25] Additionally, methods have now been developed to selectively label and/or deplete Ψ from total RNA pools using biotinylated carbodiimide reagents.[17] We envision that acrylamidofluorescein could be coupled with carbodiimide labeling to

achieve simultaneous selective modification and separate enrichment of transcripts containing I and Ψ, respectively. We also recognize the potential to improve enrichment by reducing reactivity with the natural ribonucleosides U and G, and efforts are underway to explore alternative acrylamide structures toward this goal. Regardless, given the present lack of methods for isolating inosine-containing RNAs, the research presented here represents a critical first step toward integrating chemical labeling and enrichment methods for this important application.

A-to-I RNA editing is among the most widespread epitranscriptomic modifications and is integral to a variety of cellular processes. Additionally, direct links to malfunctions in A-to-I RNA editing are being rapidly discovered for a growing number of diseases. Robust identification and characterization of these RNA modifications is vital to understanding their biological function and dynamics. The research reported here is anticipated to advance the study of A-to-I RNA editing by enabling a more comprehensive and deeper detection of inosines in the transcriptome through pre-enrichment of edited transcripts from complex RNA mixtures. While our initial investigation utilized acrylamidofluorescein, the acrylamide scaffold offers considerable flexibility for the attachment of other affinity handles and functional probes. Thus, we envision that our labeling and affinity capture approach can be expanded into a rich toolbox for elucidating the true scale and dynamics of A-to-I editing.

4.4 Supporting information

4.4.1 Synthesis of Acrylamidofluorescein

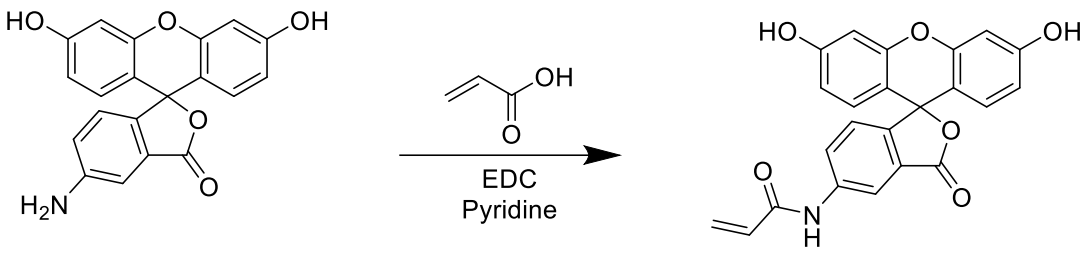


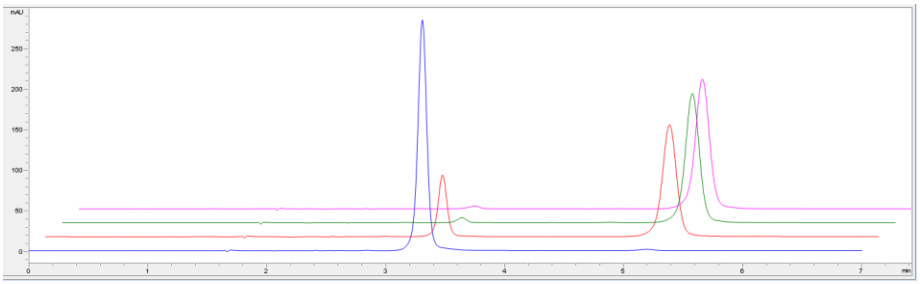
Figure 4-5 Synthesis of acrylamidofluorescein

To a solution of 5-aminofluorescein (1.00 g, 2.88 mmol) in pyridine (8.00 ml, 98.9 mmol), *N*-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (828 mg, 4.32 mmol) and acrylic acid (0.390 ml, 5.76 mmol) were added and left to stir at room temperature overnight. Once the 5-aminofluorescein was consumed, as determined by TLC, the reaction was dried under reduced pressure to form a crude oil. The crude oil was added into 20 mL of 10% sodium hydroxide and extracted using dichloromethane. The organic layer was collected and acidified by adding concentrated hydrochloric acid until orange precipitates formed. The product was then vacuum filtered and dried to yield 0.930 g (80.7%) of orange powder. 1H NMR (400 MHz, DMSO-d6) δ 10.93 (s, 1H), 8.47 (s, 1H), 7.96 (d, *J* = 8.6, 1H), 7.22 (d, *J* = 8.2, 1H), 6.77 (s, 2H), 6.53-6.68 (m, 5H), 6.35 (d, *J* = 16, 1H), 5.84 (d, *J* = 11.3, 1H). 13C NMR (400 MHz, DMSO-d6) δ 168.9, 164.2, 160.9, 152.8, 141.2, 132.0, 130.0, 128.2, 127.7, 126.7, 125.4, 114.8, 113.7, 110.6, 102.7. HRMS m/z (ESI) calcd for C23H15NO6 (M+H)+ 402.09776, found 402.09658.

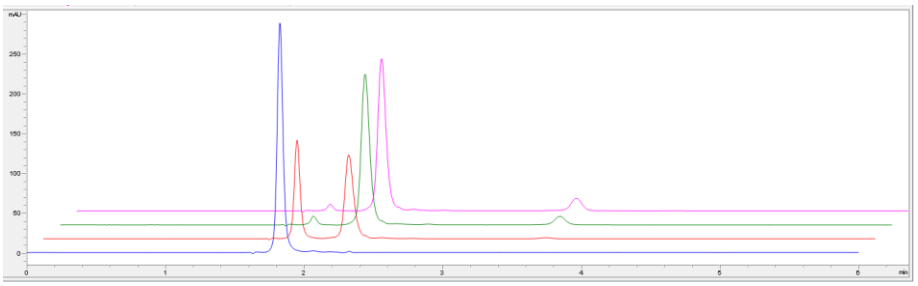
4.4.2 Ribonucleoside Labeling and HPLC Analysis

Ribonucleosides inosine, guanosine, adenosine, cytidine and uridine were purchased from Sigma Aldrich Corporation (St. Louis, MO). Pseudouridine was obtained from MP Biomedicals LLC (Santa Ana, CA). Labeling reaction mixtures were comprised of 50 mM ribonucleoside and 250 mM reagent (acrylonitrile or acrylamidofluorescein) in 50:50 EtOH:reaction buffer. Phosphate buffered saline (PBS) was used for reactions from pH 6.5-7.5 and 1M triethylammonium acetate (TEAA) for pH 8.0-10.5. Reactions were incubated at 70 ^{0}C for the time periods indicated. Reversed-phase HPLC analysis was performed on an Agilent 1260 Infinity II system using a 4µm, 150 x 4.6 mm Phenomenex Synergi Fusion-RP 80A C18 column. Samples were prepared in a stationary phase solution of 5% acetonitrile in PBS. Acrylonitrile reactions were analyzed using an isocratic mobile phase of 5:95 acetonitrile:water. Acrylamidofluorescein reactions were analyzed using a linear mobile phase gradient from 5% to 45% acetonitrile in water over 25 minutes. All mobile phases contained 0.1% trifluoracetic acid.

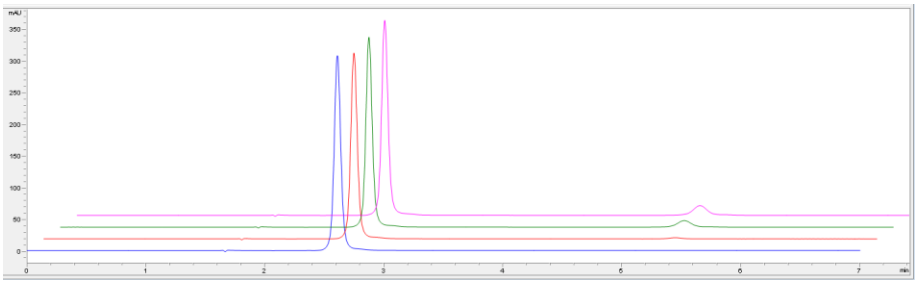
Inosine



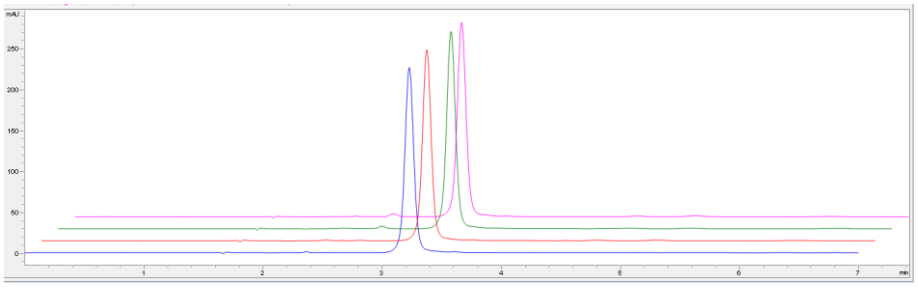
Pseudouridine



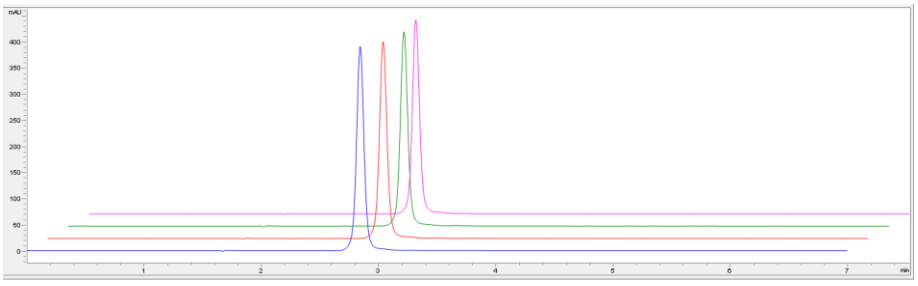
Uridine



Guanosine



Adenosine



Cytidine

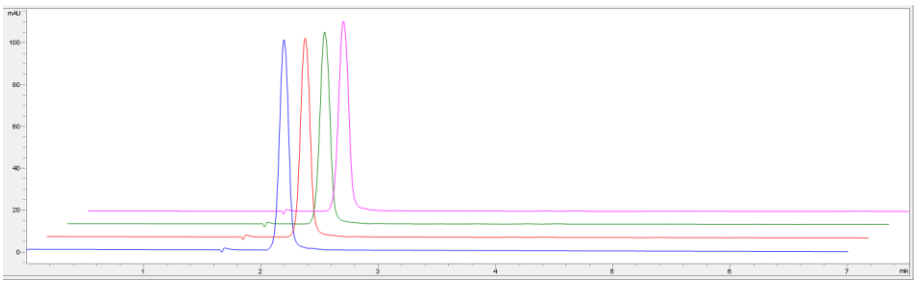
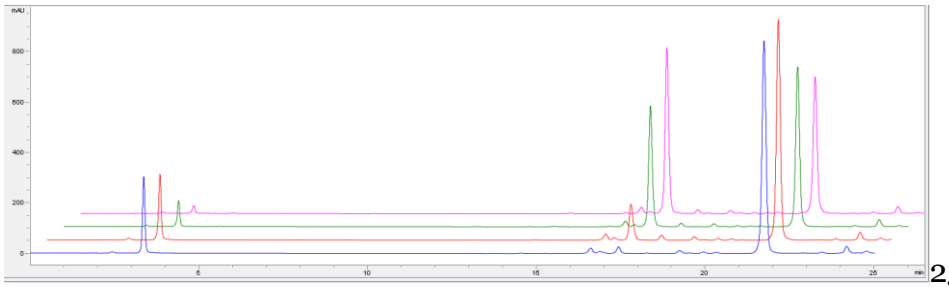
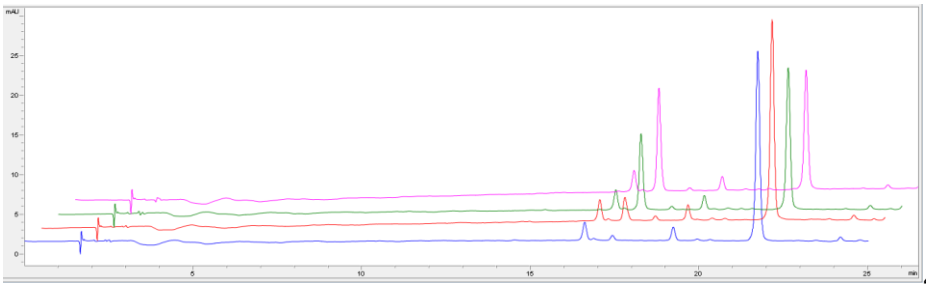


Figure 4-6 Representative HPLC traces of ribonucleoside reactivity with acrylonitrile. All chromatograms were monitored at 254nm. Blue = 0 hours, red = 1 hour, green = 8 hours, pink = 24 hours.

Inosine

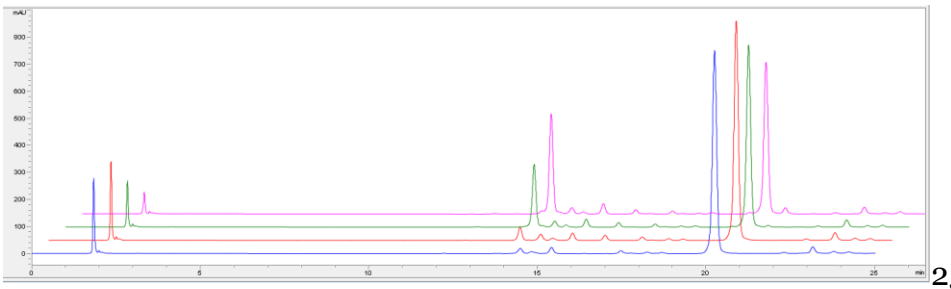


254 nm

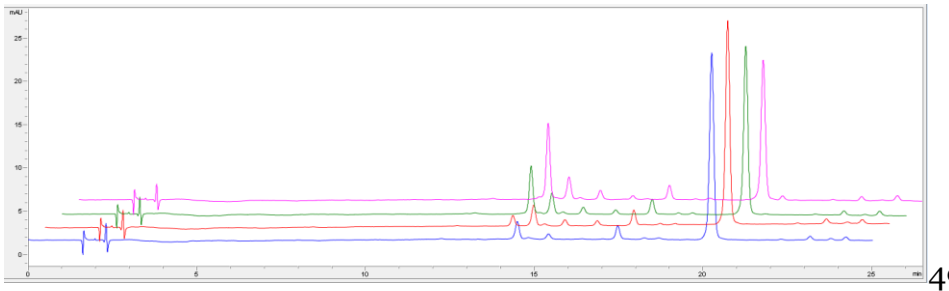


494 nm

Pseudouridine

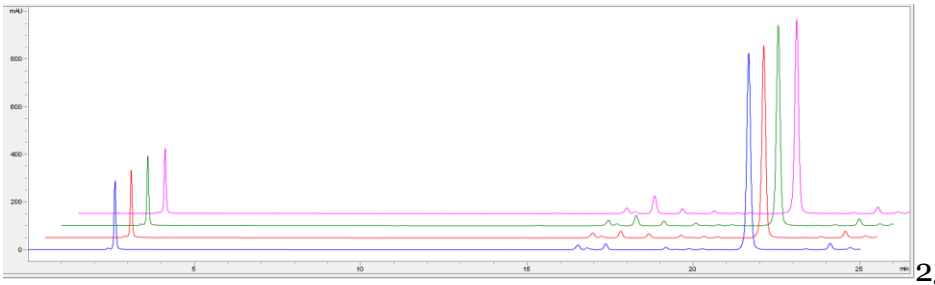


254 nm

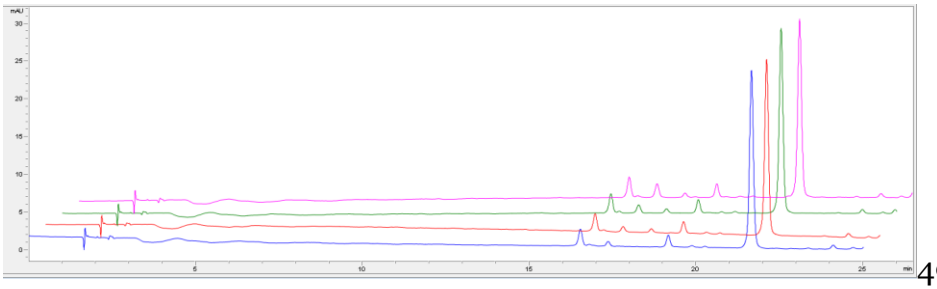


494 nm

Uridine

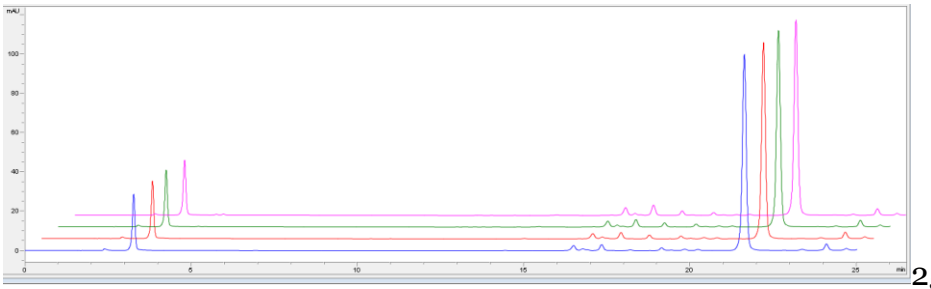


254 nm

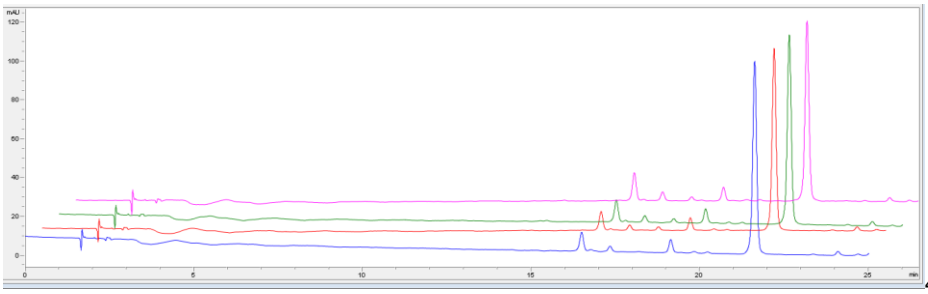


494 nm

Guanosine

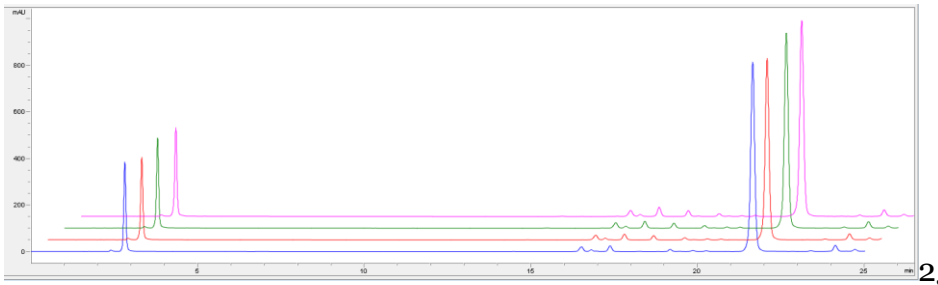


254 nm

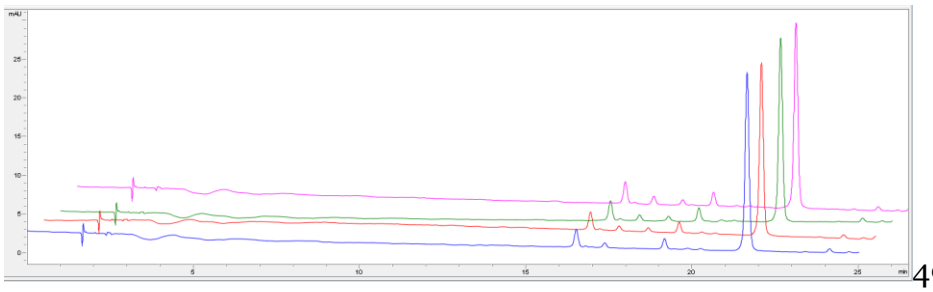


494 nm

Adenosine

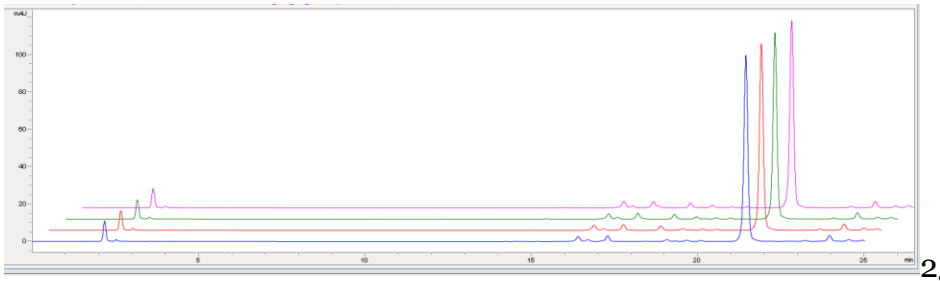


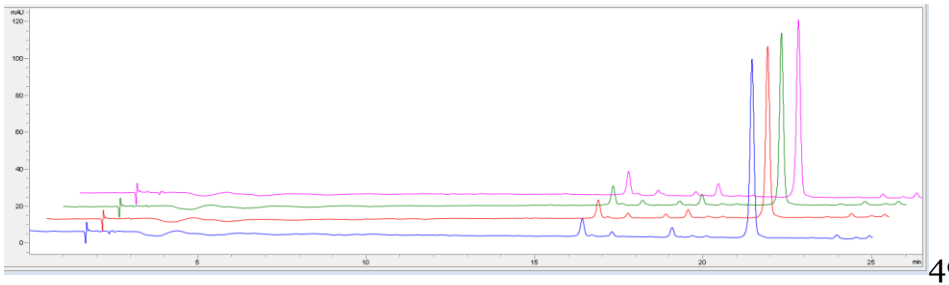
254 nm



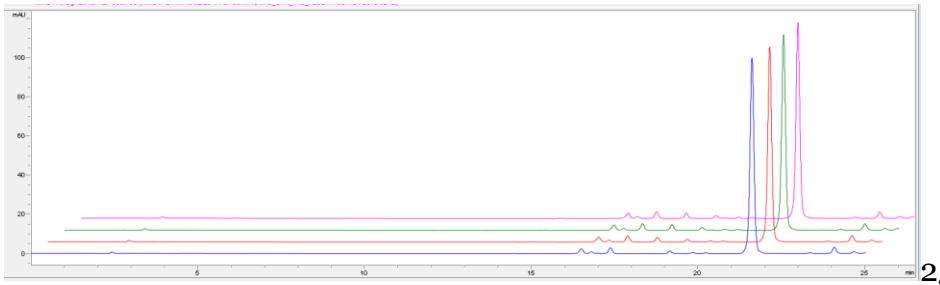
494 nm

Cytidine





Acrylamidofluorescein alone (no ribonucleoside)



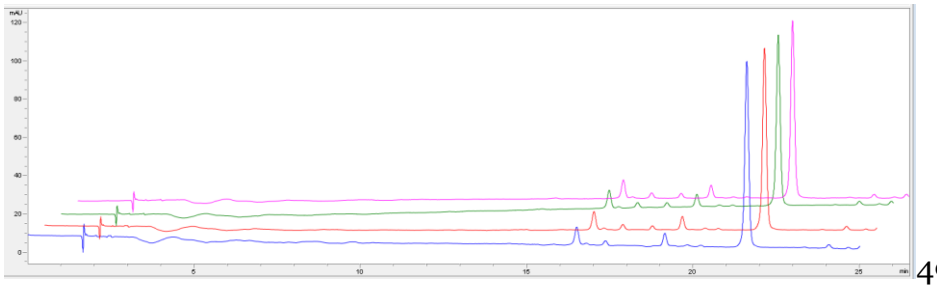


Figure 4-7 Representative HPLC traces of ribonucleoside reactivity with acrylamidofluorescein. Chromatograms were monitored at 254 and 494 nm, as indicated. Blue = 0 hours, red = 1 hour, green = 8 hours, pink = 24 hours.

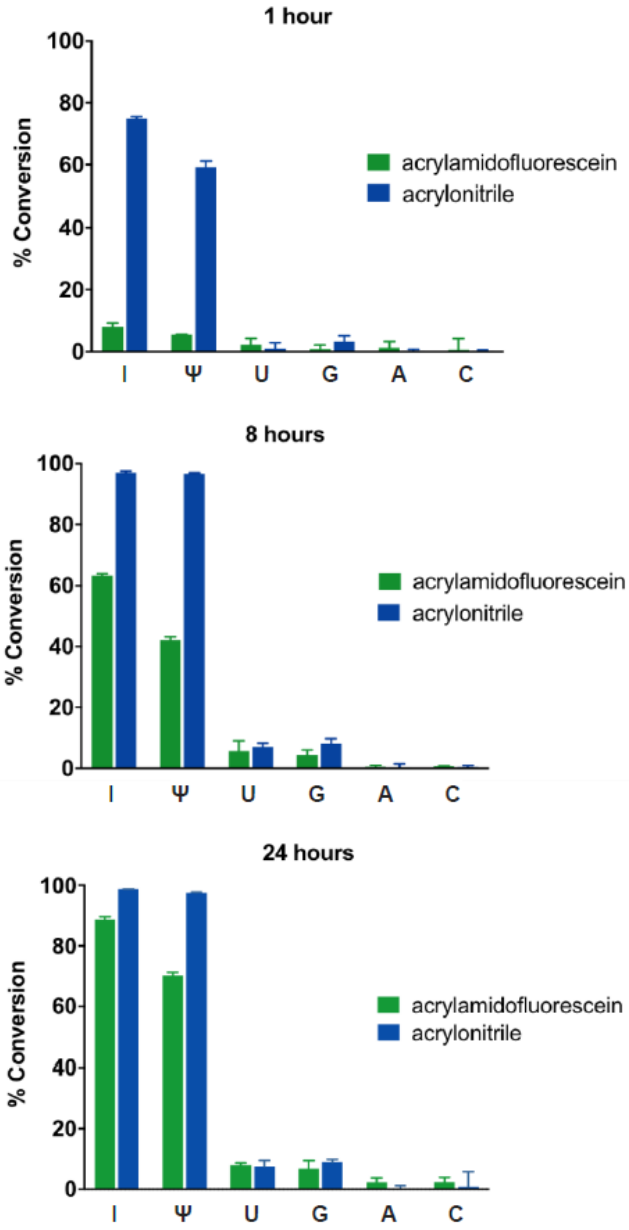


Figure 4-8 Percent conversion of ribonucleosides when reacted with acrylamidofluorescein (green bars) or acrylonitrile (blue bars) after 1, 8 and 24 hours at 70 ^{0}C, pH 8.6

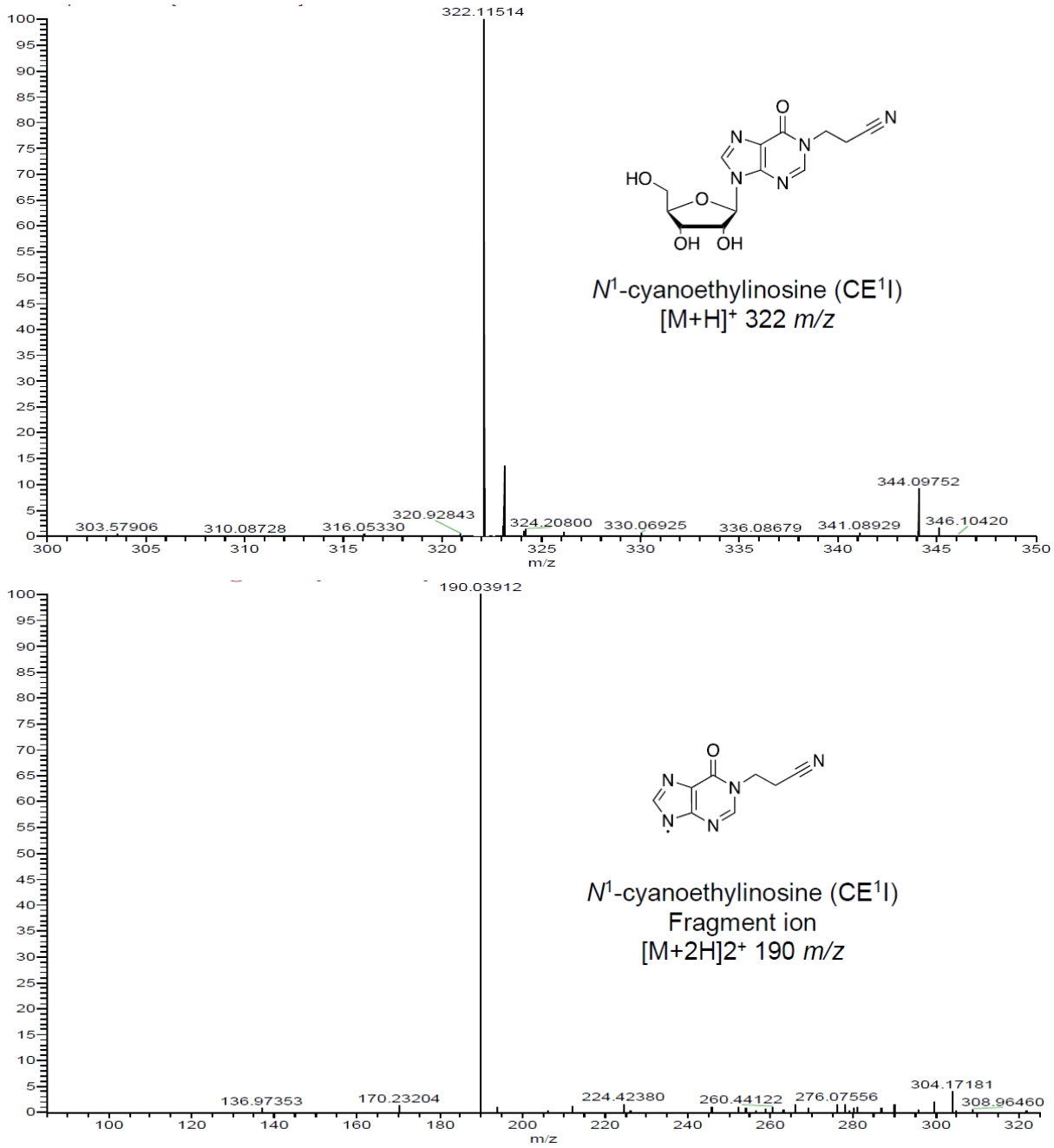


Figure 4-9 ESI-MS and MS/MS spectra of isolated product fraction for the reaction of inosine and acrylonitrile

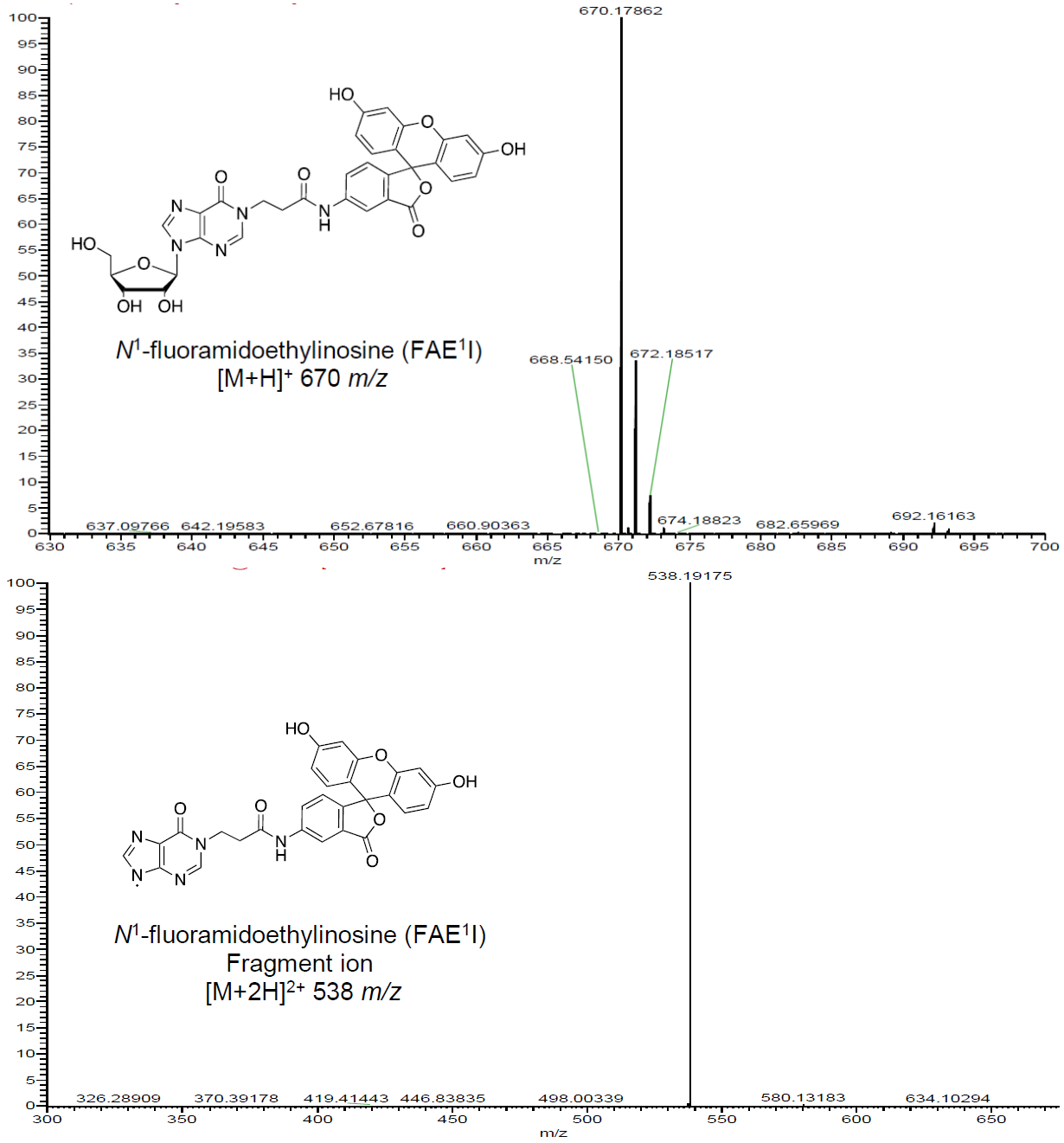


Figure 4-10 ESI-MS and MS/MS spectra of isolated product fraction for the reaction of inosine and acrylamidofluorescein

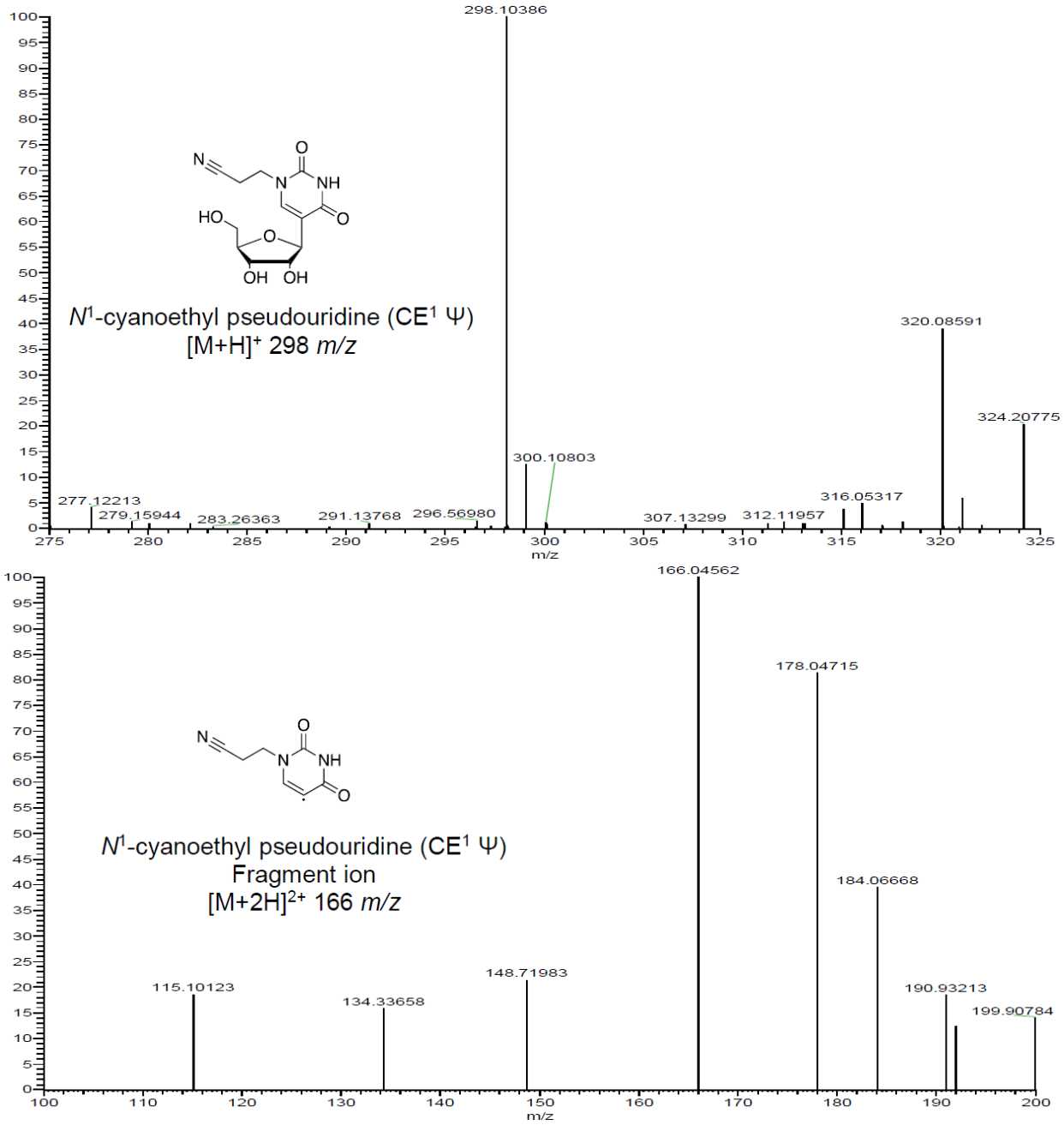


Figure 4-11 ESI-MS and MS/MS spectra of isolated product fraction for the reaction of pseudouridine and acrylonitrile.

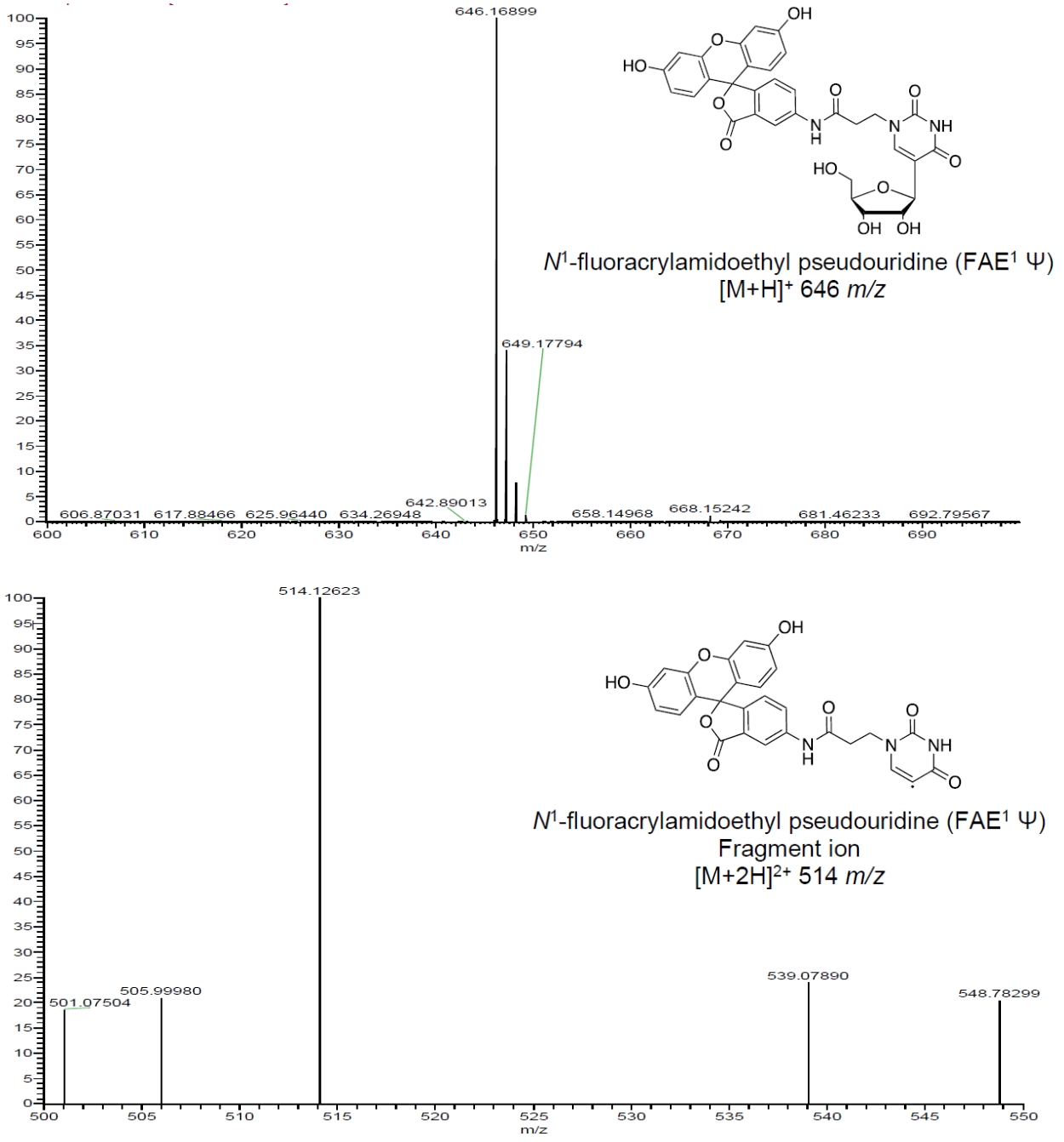


Figure 4-12 ESI-MS and MS/MS spectra of isolated product fraction for the reaction of pseudouridine and acrylamidofluorescein

4.4.3 RNA Oligoribonucleotides

RNA oligoribonucleotides were custom designed and synthesized from

Integrated DNA Technologies (Skokie, IL). Edited and non-edited controls were synthesized with either Cy5 or Cy3 at the 5' terminus as shown below.

RNA-I-Cy5 5' Cy5/GACACAUCCG**C**I**C**AGCAACGAG 3'

RNA-A-Cy3 5' Cy3/GACACAUCCGC**A**CAGCAACGAG 3'

RNA-A-Cy5 5' Cy5/GACACAUCCGC**A**CAGCAACGAG 3'

4.4.4 Oligoribonucleotide labelling and PAGE analysis

In triplicate, 1000 pmol of either RNA-A-Cy5 or RNA-I-Cy5 was added to a 0.1 mL solution of 250 mM acrylamidofluorescein in 50:50 EtOH:TEAA buffer and adjusted to pH 8.6. Reactions were incubated at 70°C. At indicated time points, crude reaction mixture was diluted 1:200 in tris-EDTA pH 7.5 buffer and ethanol precipitated. Samples were resuspended in Tris-EDTA buffer and quantified via Cy5 fluorescence. 1 pmol of each purified sample was loaded into each well, resolved on a 10% denaturing polyacrylamide gel, and imaged with a GE Amersham Typhoon. Densitometric quantification of bands was performed using ImageJ software. Each sample was normalized by comparing intensity of purified reaction bands to known amounts of RNA-I-Cy5 and a fluorescein labeled DNA oligonucleotide. Percent conversion was defined as the molar ratio of fluorescein to Cy5 for each well. All reactions were analyzed in triplicate.

4.4.5 Oligoribonucleotide labelling and Immunoprecipitation pulldown

In triplicate, varying mixtures of RNA-I-Cy5 and RNA-A-Cy3 were prepared in a 0.1 mL solution of 250 mM acrylamidofluorescein in 50:50 EtOH:TEAA buffer, adjusted to pH 8.6, and incubated at 70 °C for 24 hours. Mixtures were defined as follows:

Table 4-1 Preparation of RNA-I-Cy5 and RNA-A-Cy3 mixture solutions

Input Ratio (RNA-I-Cy5:RNA-A-Cy3)	pmol RNA-I-Cy5	pmol RNA-A-Cy3
1:1	500	500
1:10	100	1000
1:100	10	1000
$1:10^4$	1	1000
$1:10^5$	0.1	1000

After incubation, crude reaction mixtures were diluted 1:10 in tris-EDTA pH 7.5 buffer and ethanol precipitated. Samples were resuspended in 0.5 mL PBS + 0.05% tween 20 (PBST). An excess of monoclonal mouse anti-fluorescein antibody (MIF2901, Thermo Fisher Scientific, Rockford, IL) was added to each tube and incubated with end over end rotation for 2 hours at 4 °C. 0.02 mLof Protein A/G magnetic agarose beads (Thermo Fisher Scientific, Rockford, IL) was then added to each tube and incubated with end over end rotation for an additional 2 hours at 4 °C. Beads were then washed extensively with PBST, and bound oligoribonucleotides were eluted by heating to 95 °C for 20 minutes. Eluates were analyzed on a BioTek Cytation 5 spectrophotometer, and Cy5 and Cy3 concentrations were determined by correlating to a standard curve of RNA-I-Cy5 and RNA-Cy3. Fold-enrichment was defined as $\frac{[RNA\text{-}I\text{-}Cy5]_{final}/[RNA\text{-}A\text{-}Cy3]_{final}}{[RNA\text{-}I\text{-}Cy5]_{initial}/[RNA\text{-}A\text{-}Cy3]_{initial}}$

4.5 References

1. Bass, B. L., RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* **2002,** *71*, 817-46.

2. Valente, L.; Nishikura, K., ADAR gene family and A-to-I RNA editing: diverse roles in posttranscriptional gene regulation. *Progress in nucleic acid research and molecular biology* **2005,** *79*, 299-338.

3. Kawahara, Y. Z., Boris ; Sethupathy, Praveen ; Iizasa, Hisashi ; Hatzigeorgiou, Artemis G ; Nishikura, Kazuko, Redirection of Silencing Targets by Adenosine-to-Inosine Editing of miRNAs. *Science Vol. 315* (5815), 1137-1140.

4. Nishikura, K., A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* **2016,** *17* (2), 83-96.

5. Sakurai, M.; Yano, T.; Kawabata, H.; Ueda, H.; Suzuki, T., Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. *Nat Chem Biol* **2010,** *6* (10), 733-40.

6. Sakurai, M.; Suzuki, T., Biochemical identification of A-to-I RNA editing sites by the inosine chemical erasing (ICE) method. In *RNA and DNA Editing*, Springer: 2011; pp 89-99.

7. Tan, M. H.; Li, Q.; Shanmugam, R.; Piskol, R.; Kohler, J.; Young, A. N.; Liu, K. I.; Zhang, R.; Ramaswami, G.; Ariyoshi, K., Dynamic landscape and regulation of RNA editing in mammals. *Nature* **2017,** *550* (7675), 249.

8. Paul, M. S.; Bass, B. L., Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *The EMBO journal* **1998,** *17* (4), 1120-1127.

9. Yang, J. H.; Luo, X.; Nie, Y.; Su, Y.; Zhao, Q.; Kabir, K.; Zhang, D.; Rabinovici, R., Widespread inosine-containing mRNA in lymphocytes regulated by ADAR1 in response to inflammation. *Immunology* **2003,** *109* (1), 15-23.

10. Li, X.; Xiong, X.; Wang, K.; Wang, L.; Shu, X.; Ma, S.; Yi, C., Transcriptome-wide mapping reveals reversible and dynamic N 1-methyladenosine methylome. *Nature chemical biology* **2016,** *12* (5), 311.

11. Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M., Topology of the human and mouse m 6 A RNA methylomes revealed by m 6 A-seq. *Nature* **2012,** *485* (7397), 201.

12. Edelheit, S.; Schwartz, S.; Mumbach, M. R.; Wurtzel, O.; Sorek, R., Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS genetics* **2013,** *9* (6), e1003602.

13. Delatte, B.; Wang, F.; Ngoc, L. V.; Collignon, E.; Bonvin, E.; Deplus, R.; Calonne, E.; Hassabi, B.; Putmans, P.; Awe, S., Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **2016,** *351* (6270), 282-285.

14. Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; León-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.; Satija, R.; Lander, E. S., Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **2014,** *159* (1), 148-162.

15. Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. V., Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **2014,** *515* (7525), 143.

16. Lovejoy, A. F.; Riordan, D. P.; Brown, P. O., Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PLoS One* **2014,** *9* (10), e110799.

17. Li, X.; Zhu, P.; Ma, S.; Song, J.; Bai, J.; Sun, F.; Yi, C., Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nature chemical biology* **2015,** *11* (8), 592.

18. Wu, Q.; Amrutkar, S. M.; Shao, F., Sulfinate Based Selective Labeling of 5-Hydroxymethylcytosine: Application to Biotin Pull Down Assay. *Bioconjugate chemistry* **2018,** *29* (2), 245-249.

19. Inouye, H.; Fuchs, S.; Sela, M.; Littauer, U. Z., Detection of inosine-containing transfer ribonucleic acid species by affinity chromatography on columns of anti-inosine antibodies. *Journal of Biological Chemistry* **1973,** *248* (23), 8125-8129.

20. YOSHIDA, M.; UKITA, T., Selective Modifications of Inosine and Ø-Uridine with Acrylonitrile out of the Other Ribonucleosides. *The Journal of Biochemistry* **1965,** *57* (6), 818-821.

21. Ofengand, J., The Function of Pseudouridylic Acid in Transfer Ribonucleic Acid I. the specific cyanoethylation of pseudouridine, inosine, and 4-thiouridine by acrylonitrile. *Journal of Biological Chemistry* **1967,** *242* (21), 5034-5045.

22. Fox, J. J.; Wempen, I.; Hampton, A.; Doerr, I. L., Thiation of Nucleosides. I. Synthesis of 2-Amino-6-mercapto-9-β-D-ribofuranosylpurine ("Thioguanosine") and Related Purine Nucleosides1. *Journal of the American Chemical Society* **1958,** *80* (7), 1669-1675.

23. Cohn, W. E., Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *Journal of Biological Chemistry* **1960,** *235* (5), 1488-1498.

24. Suzuki, T.; Ueda, H.; Okada, S.; Sakurai, M., Transcriptome-wide identification of adenosine-to-inosine editing using the ICE-seq method. *Nature protocols* **2015,** *10* (5), 715.

25. Addepalli, B.; Limbach, P. A., Mass spectrometry-based quantification of pseudouridine in RNA. *Journal of the American Society for Mass Spectrometry* **2011,** *22* (8), 1363-1372.

Chapter 5 : Conclusion and future direction

5.1 Conclusion and future direction

The intricate and dynamic nature of cellular components governs their organization and specialized function. Therefore, a broadly applicable fluorescent labeling approach is essential for studying the complex properties of biomolecules in their native environments. One of the biggest challenges yet to be addressed in the pursuit of fluorescent labeling biomolecules is the scarcity of approaches that are robust, target-specific, and produce reduced background signal. To overcome these limitations, here, we described the utilization of small molecules with various reactive linkers for covalent and fluorescent labeling of RNAs and proteins.

Chapter 2: Genetically encoded fluorescent proteins or small-molecule probes are widely used to label proteins for fluorescence microscopy. These tools have dramatically advanced our understanding of protein localization, expression, and cellular function. Despite the availability of multiple protein labeling techniques, there remains a need for a robust approach that provides temporal control over labeling. Here we report a photoaffinity labeling approach using a novel malachite green analogue with an existing fluorogen activating protein receptor. We show that this technique selectively and covalently labels target proteins in live mammalian cells with temporal resolution and minimal background signal. We envision that this approach will facilitate more in-depth interrogation of protein dynamics and improve our understanding of protein activity. Like the RNA labeling approach discussed in chapter 2, we also anticipate the future direction of this project involves the implementation of a similar approach for covalently labeling POIs using other FAP/fluorogen pairs. In addition to enabling multiplex labeling

of proteins in a single system, future studies will also focus on combining both the photoaffinity labeling strategy of cellular RNAs and proteins in a single system.

Chapter 3: Fluorescence imaging of cellular RNA has yielded significant insight into the role of transcript localization and dynamics in biological processes. While there are several reported RNA labeling methods, they all suffer from limitations related to either the requirement for cell fixation, the need for large tags that can disrupt localization, and the lack of robustness due to the weak nature of the non-covalent interactions upon which they rely. To overcome these shortcomings, we developed a broadly applicable strategy that can be implemented for imaging any transcript in either fixed or live cell experiments. Specifically, we envisioned a photoaffinity labeling approach, as this would also provide temporal control over the RNA labeling process. We synthesized and modified the malachite green fluorogen to incorporate a photoaffinity linker, which allowed for covalent labeling of its cognate aptamer upon irradiation with UV light. By placing this aptamer at the 5' UTR of the mRNA, we reported the first covalent light-up aptamer system for visualizing the spatiotemporal localization pattern of mRNA in live mammalian cells. This RNA-labeling method requires only a short 57 nt fusion, which is significantly smaller than the fusion needed for any of the other RNA labeling systems reported to date. Using the CDK6 mRNA as a model system, we showed the gradual phase separation and maturation of RNA granules in mammalian cells. Taking advantage of the covalent labeling approach, we were able to perform media exchange to also visualize the dissolution of these granules in real-time. In summary, this novel strategy considerably advances the RNA imaging technology and provides a robust and reliable tool for the RNA community. The future direction of this approach involves

the application of strategy to investigate the localization patterns of different transcripts as it relates to various disease models. We also anticipate the implementation of a similar approach to improve other FLAP system and enable multiplexed covalent labeling of transcripts in a single cell.

Chapter 4: Inosine has an essential role in the proper folding and translation of RNA, however, there are no known strategies to selectively label and isolate inosine containing RNA molecules. Here, we report the implementation of acrylamide reactive group as selective covalent labeling motif. In addition to covalent labeling, by appending this reactive linker onto a fluorescein molecule, we also reported the quantification and enrichment of Inosine-containing RNA molecules. This strategy lays the groundwork for future derivatization of acrylamide linkers with different molecules that will enable alkylation of inosine with a variety of probes and affinity handles.

In conclusion, by harnessing the molecular recognition capabilities of biomolecules and temporally controlled reactivity of different linkers, we described the development and implementation of a highly generalizable and robust approach for visualizing, tracking, quantifying, and isolating specific biomolecules in a complex biological environment.