UNIVERSITY OF CALIFORNIA SAN DIEGO

# Enzymatic Tools for Enabling Nucleic Acid Driven Technologies

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

in

Chemistry

by

Ember Marai Tota

Committee in charge:

Professor Neal K. Devaraj, Chair
Professor Alexis C. Komor
Professor Yitzhak Tor
Professor Gene W. Yeo
Professor Brian M. Zid

2021

The Dissertation of Ember Marai Tota is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

# DEDICATION

I dedicate this dissertation to my family.

Thank you for believing in me and supporting me in my dreams,

in this work they have come true.

# EPIGRAPH

Life is not easy for any of us.

But what of that?

We must have perseverance and above all confidence in ourselves.

We must believe that we are gifted for something,

and that this thing, at whatever cost, must be attained.

Marie Curie

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGEMENTS

Getting a PhD has been the most challenging endeavor of my life thus far. My path to this point was far from linear and anything but traditional. With four associate's, a bachelor's, and a master's degree, this doctorate marks my seventh and final degree. Just like each one before, I could not be prouder. At each step I thought I had reached a culmination point and never in a thousand years did I imagine I would one day earn a PhD.

I would not have been able to get to this place without the support, mentorship, and encouragement of so many of my professors during my undergraduate studies that saw my potential and pushed me to push myself. I would like to thank my Anatomy and Physiology professor at Mesa College, Professor Anne Geller, who encouraged me to change my path to being a medical doctor. While that is not where I ended up, it is because of you that I transferred to UCSD to pursue my bachelor's degree and I would not be here without that. I would also like to thank the Chemistry Department at Grossmont College, particularly, Professor Tom Olmstead, Professor Judy George, and Professor Cary Willard. You all helped my love for chemistry flourish and gave me an incredible foundation to build on when I transferred to UCSD, without your support and excellent teaching I would not have been as successful as I have. Finally, I would like to thank Professor Haim Weizman, you have been one of the most influential people in my life and I would not be here without you. Thank you for pushing me, encouraging me to push myself, and always helping me aim higher. From asking how many A+'s there were on my transcript to introducing me to the idea of becoming a scientist, I am here in very large part because of you and your mentorship and friendship have been invaluable.

My entire academic career came to a culmination point in the lab of Professor Neal Devaraj, my PhD advisor. Neal, it is hard to express the gratitude I have for your mentorship, guidance, encouragement, and overall support. You helped me make my PhD exactly what I wanted it to be. Your never-ending support in helping me follow my path while encouraging and challenging me to question it and redirect has been instrumental in my academic career and future aspirations. The incredible group of scientists that you have brought into the lab has fostered an excellent environment that encourages scientific curiosity, mentorship, collaboration, and friendship. I would like to thank the amazing group of people that I have had the pleasure to call my colleagues: Eric Zhou, Xijun Piao, Andrés Seoane Fernandez, Roberto Brea, Henrike Neiderholtmeyer, Shuaijiang Jin, Mai Johnnson, Luping Liu, Jorge Jimenez, Youngjun Lee, Andrew Rudd, Alessandro Fracassi, Chih-Chin Chi, Mahta Moinpour, Ahanjit Bhattacharya, Kayla Busby, Dongyang Zhang, Kira Podolsky, Hetika Vora, Satyam Khanal, Judith Flores, Christy Cho, Jiyue Chen, Alexander Harjung, and Jacob Vance. In particular, I would like to thank Kayla, for your scientific guidance, immense friendship, and so, so much more. Thank you for all that you have done and continue to encourage me to do.

During my PhD I spent time in two labs, shifting after earning my master's degree. While this dissertation describes the work done in the lab of Neal Devaraj, I would be remiss not to mention my time in the lab of Kamil Godula. I would like to thank Kamil for his mentorship and support throughout my time in his lab and during my transition to a new lab. I would also like to acknowledge my postdoctoral mentor, Mia Huang, for her mentorship, support, and friendship, thank you for all you did to encourage and guide me.

Finally, I would like to thank members of the lab, Chris Fisher, Greg Trieger, Mitchell Christy, Dan Honigfort, Austen Michalak, Matt Naticchia, Taryn Lucas, Bryce Timm, Ariana Remmel, Mia Huang, Lisa Adamiak, and Steve Verespy. Beyond being colleagues, we became friends, and your support has helped me get to this point.

I would like to thank my undergraduate advisor, Professor Yitzhak Tor, for your support and mentorship, both then and now. I would also like to thank my graduate student mentor Patrycja Hopkins, thank you for your guidance and friendship.

A big thank you to my committee members, Alexis Komor, Yitzhak Tor, Gene Yeo, and Brian Zid for their encouragement and guidance over the last several years.

Finally, and very importantly, I would like to thank my family and all of my friends who have become a part of my family. To all of my friends, especially Paige, Alex, Lisa, Taryn, and Alyson, you have all guided me in some way, whether in life or in science, and I couldn't be more grateful. Mommy and Daddy, thank you so much for always encouraging me and supporting me in everything I have ever wanted to do. Without your faith in me I would not be here. Diora, Michael, and Mark, thank you for being an amazing part of my life and my bonus parents. To all of my siblings, thank you for dealing with my nerdiness and overachiever mindset and loving me through it all. Linda, AJ, Alicia, and Brandon, your love and support has meant more to me than you will ever know, thank you. I would like to thank my furry children, Phoebe and Stella, your unconditional love has brought me so much happiness and has helped me through the hardest of times. Finally, I would like to thank my partner and best friend for the last 14 years, Chris Tota, I would not be here without you, and I will forever be grateful for this time and the joy, support, and drive it has brought to my life.

## Notes About the Chapters

Chapter two, in part is currently being prepared for submission for publication. Tota, Ember M.; Devaraj, Neal K. The dissertation author was the primary researcher and author of this material.

Chapter three, in part is currently being prepared for submission for publication. Tota, Ember M.; Harjono, Vince; Devaraj, Neal K. I would like to thank Vince Harjono, who assisted in experiments involving oligonucleotide probe sets labeled with terminal deoxynucleotidyl transferase (TdT). The dissertation author was the primary researcher and author of this material.

Chapter four, in part is currently being prepared for submission for publication. Tota, Ember M.; Devaraj, Neal K. The dissertation author was the primary researcher and author of this material.

# VITA

## Education

**2013** Associate of Science, Chemistry, Physics, Mathematics, Cuyamaca College

**2015** Bachelor of Science, Biochemistry/Chemistry, UC San Diego

**2017** Master of Science, Biochemistry/Chemistry, UC San Diego

**2021** Doctor of Philosophy, Biochemistry/Chemistry, UC San Diego

## Publications

Zhang, D.; Jin, S.; Liu, L.; **Tota, E.M**.; Li, Z.; Piao, X.; Devaraj, N.K RNA-CLAMP Enables Photo-activated Control of CRISPR-Cas9 Gene Editing by Site-specific Intramolecular Cross-linking of the sgRNA. *preprint: bioRxiv.* **2021.**

Sandoval, D. R.; Gomez Toledo, A.; Painter, C. D.; **Tota, E.M**.; Sheikh, M. O.; West, A.; Frank, M. M.; Wells, L.; Xu, D.; Bicknell, R.; Corbett, K. D.; Esko, J. D. Proteomics-based screening of the endothelial heparan sulfate interactome reveals that C-type lectin 14a (CLEC14A) is a heparin-binding protein. *The Journal of biological chemistry*. **2020,** 2804–2821.

Huang, M.L.; **Tota, E.M.**; Lucas, T.M.; Godula, K. Influencing Early Stages of Neuromuscular Junction Formation through Glycocalyx Engineering. *ACS Chem Neurosci*. **2018,** 3086-3093.

Naticchia, M.R.; Laubach, L.K.; **Tota, E.M.**; Lucas, T.M.; Huang, M.L.; Godula, K. Embryonic Stem Cell Engineering with a Glycomimetic FGF2/BMP4 Co-Receptor Drives Mesodermal Differentiation in a Three-Dimensional Culture. *ACS Chem Biol*. **2018,** 2880-2887.

Huang, M.L.; **Tota, E.M.**; Verespy, S.; Godula, K. Glycocalyx Scaffolding to Control Cell Surface Glycan Displays. *Curr Protoc Chem Biol*. **2018**.

## Fellowships and Awards

**2016** National Science Foundation Graduate Research Fellowship

**2015** Katzin Prize Recipient

# ABSTRACT OF THE DISSERTATION

Enzymatic Tools for Enabling Nucleic Acid Driven Technologies

by

Ember Marai Tota

Doctor of Philosophy in Chemistry

University of California San Diego, 2021

Professor Neal K. Devaraj, Chair

The importance of nucleic acids has expended past their role in encoding life and into a space of scientific utility and discovery. Developing new, robust methods to study and manipulate them in their natural environments and harness their programmability for fundamental research tools and biomedical applications has become a critical goal for researchers. This work presents an enzymatic tool for functional modification of single stranded DNA oligonucleotides as well a strategy to assemble functional nucleic acid protein conjugates.

Due to its greater stability and ease of synthesis, DNA, rather than RNA, is the nucleic acid of choice for many functional applications. Often, these applications require

the DNA oligonucleotide to be modified in some way with an affinity tag or reporter. There are a limited number of methods to label DNA oligonucleotides, all of which have their shortcomings, including being cost prohibitive, inefficient, or a limited small molecule scope. In this dissertation I develop an enzymatic single stranded DNA modification method, expanding on the powerful RNA transglycosylation at guanine (RNA-TAG) modification tool developed in our lab. DNA-TAG offers researchers with an unprecedented tool to generate modified oligonucleotides in their lab. This work describes the rational and iterative design of DNA-TAG and demonstrates its utility through fluorescent northern blot and RNA fluorescent in situ hybridization (FISH).

Combining the programmability of nucleic acids with the functions of proteins has the potential to enable powerful assemblies with unique and critical functions, both natural and synthetic. Developing novel strategies to bring these classes of biomolecules together in a robust and specific manner continues to facilitate new applications. Here I present a fully enzymatic method to covalently assemble a functional RNA-Protein conjugate using a novel bifunctional small molecule probe and demonstrate he utility of the system by recruitment of an endonuclease to an RNA of interest to induce degradation.

With the herein presented advancements, we expand the application space of nucleic acid modification. Nucleic acid transglycosylation at guanine (NA-TAG) technologies provide researchers with a single tool capable of modifying nearly any single stranded nucleic acid substrate with a small molecule of their choice for downstream applications.

# 1 Introduction

## 1.1 Motivation

At the most fundamental level, nucleic acids serve as the blueprint for life. From tightly packed chromosomal DNA on the order of 100 million base pairs to siRNAs that are on the order of 10s of nucleotides, the heterogeneity of nucleic acid species necessitates a vast toolbox of methods to study them. Among these tools are a variety of analytical techniques including gel electrophoresis, sequencing, qPCR, imaging and detection techniques relying on protein-nucleic acid interactions, in situ hybridization (ISH), and covalent modification. Over the last century, scientists have not only worked diligently to better understand the native role of nucleic acids, but to develop creative ways to harness their unmatched versatility and programmability for biotechnological applications. These groundbreaking innovations include PCR, aptamers, next generation sequencing, DNA barcoding, data storage, genetic modification, and vaccine technologies, to name a few.

Many applications require the nucleic acid to be associated with a functional small molecule of some kind which is then used for visualization, immobilization, conjugation, and affinity interactions, among others. A variety of methods have been developed to detect and image nucleic acids transcripts including RNA binding fusion proteins, RNA aptamers, DNA and RNA targeting antisense probes, and repurposed RNA modification enzymes.[1–8] Of the techniques available, covalent modification is the most robust and non-perturbing. Modification strategies include the site-specific insertion of synthetic nucleotides bearing a functional group of interest, non-site-specific chemical modification of nucleobases or the sugar backbone, and a limited set of enzymatic modification techniques.[9–12] For many technical applications, precision is the key to success, and it is

necessary to have the means to carry out an efficient, site-specific modification of the nucleic acid substrate. Enzymatic modifications offer not only all of the benefits listed above but they can be carried out in complex environments under mild conditions and can be programmed to work in cellular systems.

Scientists have turned to tRNA modification enzymes, with at least 85 known posttranscriptional tRNA modifications, the space for inspiration and innovation is vast within this class of enzymes.[13] One modification strategy appends an entire tRNA to a transcript of interest and enzymatically installs a click handle for subsequent small molecule conjugation.[14] Our lab has turned to another tRNA modification enzyme, tRNA guanine transglycosylase (TGT), whose native function is to exchange a single guanine in the anticodon loop of its 4 cognate tRNAs for 7-deazaguanine derivatives. The promise of TGT as an RNA modification tool lies in its transglycosylation mechanism; by excising an entire nucleobase and replacing it with another it offers the potential to insert functionally modified analogs of its small molecule substrate.[15] Earlier work in the lab focused on the development of such a modification technique using TGT from *E.* coli. The developed technique is called RNA transglycosylation at guanine (RNA-TAG) and can efficiently and site selectively insert small molecule handles into an RNA of interest via recognition of a programmed minimal 17 nucleotide hairpin sequence.[16] Thus far our lab has worked to demonstrate the utility of this robust modification technique through the development of a variety of applications to study and manipulate RNAs.[17–23] However, until this work we had not yet explored the expanded potential of TGT to act on DNA substrates in addition to their RNA counterparts. While a great deal of innovation has been seen in developing tools to enzymatically modify RNA, there has been less progress

in the development of strategies to covalently modify single stranded DNA. In this work, I both expand the utility of RNA-TAG with a fully enzymatic strategy for functional nucleic acid – protein conjugation and extend the capabilities of the nucleic acid transglycosylation strategy, presenting an unprecedented tool to enzymatically functionalize single stranded DNA substrates (DNA-TAG). This set of complimentary techniques is referred to collectively as nucleic acid transglycosylation at guanine (NA-TAG).

## 1.2 tRNA Guanine Transglycosylases – Substrates and Catalytic Mechanism

tRNA guanine transglycosylases catalyze the exchange of guanine for 7-deazaguanine derivatives in cognate tRNAs. These enzymes are present across all three kingdoms of life, differing slightly in their small molecule or tRNA substrates.[24] Eubacterial and eukaryal TGTs insert PreQ1 and queuine respectively, into anticodon loops of the same target tRNAs: $tRNA^{Asn}$, $tRNA^{Asp}$, $tRNA^{Tyr}$, and $tRNA^{His}$, (Figure 1.1) by exchanging it for a guanine at the wobble position of the anticodon stem loop.[25] It has been well documented that TGT from Escherichia coli (E. coli) is able to modify a 17 nucleotide RNA hairpin derived from the anticodon loop of $tRNA^{Tyr}$ (ECY-A1), not requiring the full tRNA sequence for recognition, and that an extended stem 25 nucleotide hairpin (ECY-MH) can also be recognized and with higher efficiency than the 17 nucleotide hairpin.[26] Research from the same group showed that the actual minimal recognition unit for enzymatic activity is the UGU motif in the 33-35 positions of the cognate tRNAs and that the nucleobase at position 32 needed to be a pyrimidine, either C or U, and that their larger purine counterparts at that position, A and G, reduced enzyme activity (Figure 1.1.).[27]

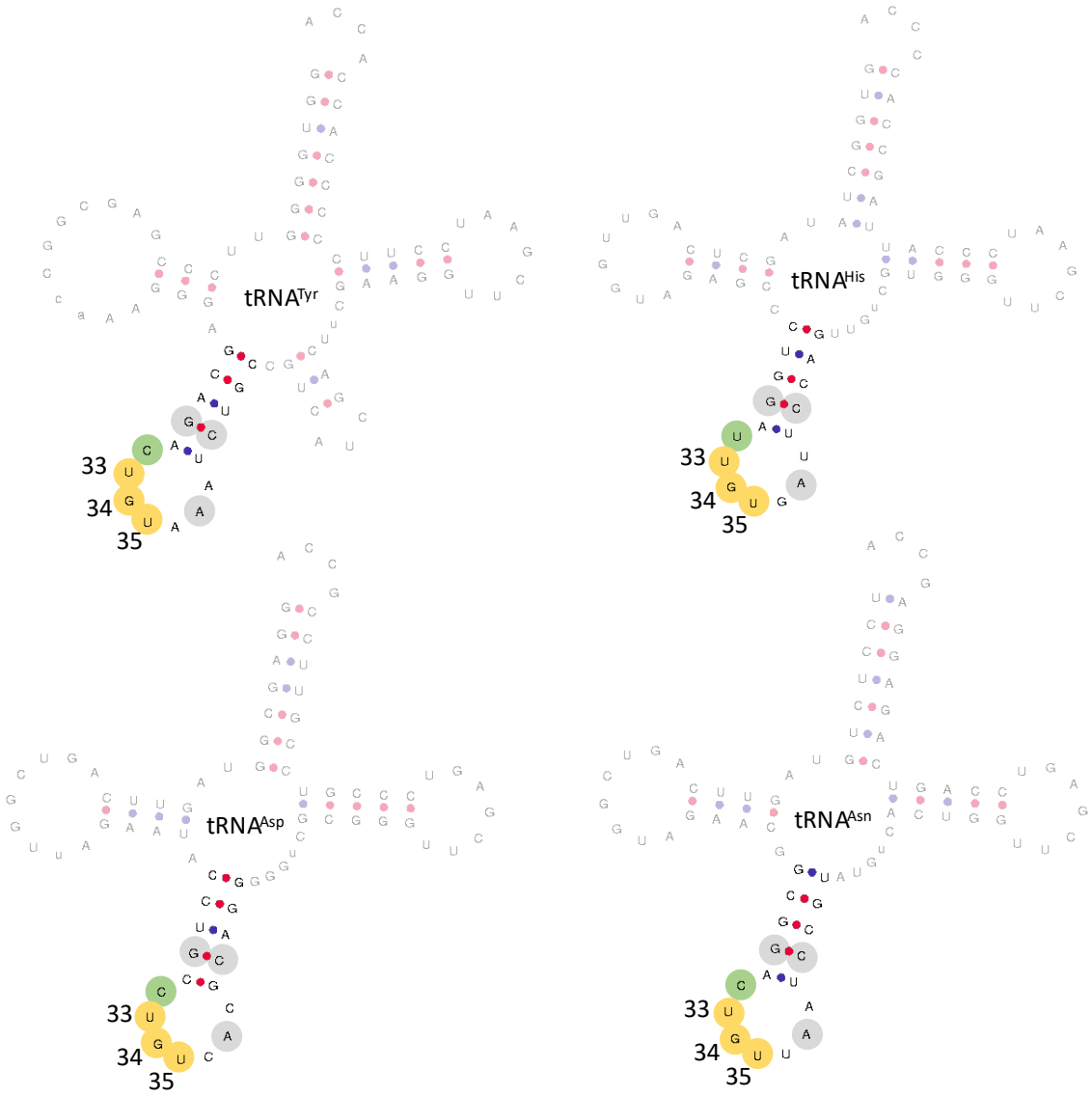



**Figure 1.1** Eukaryal and eubacterial TGT cognate tRNAs and anticodon loop numbering scheme based on $tRNA^{Tyr}$ numbering. Minimal recognition motif is indicated by the yellow circle, the required pyrimidine base by the green circle, and anticodon loop similarities between tRNAs by the grey circles.

Upon recognition of its RNA substrate, the enzyme excises a guanine and becomes covalently bound to the sugar backbone through the carboxyl group of Asp143.[28] Subsequently, a PreQ1 substrate enters the active site and covalently inserts at the anomeric sugar carbon, releasing the enzyme (Figure 1.2.).[29–33] This covalent modification is efficient, irreversible, and is able to proceed at a mild 37˚C. This and past

work from our own lab has demonstrated the versatility of TGT in its ability to accept a variety of small molecule substrates as well as its flexibility in recognizing its nucleic acid substrate in different contexts. Namely, TGT can recognize its 17-nucleotide hairpin when it is inserted in a nucleic acid substrate.[16]

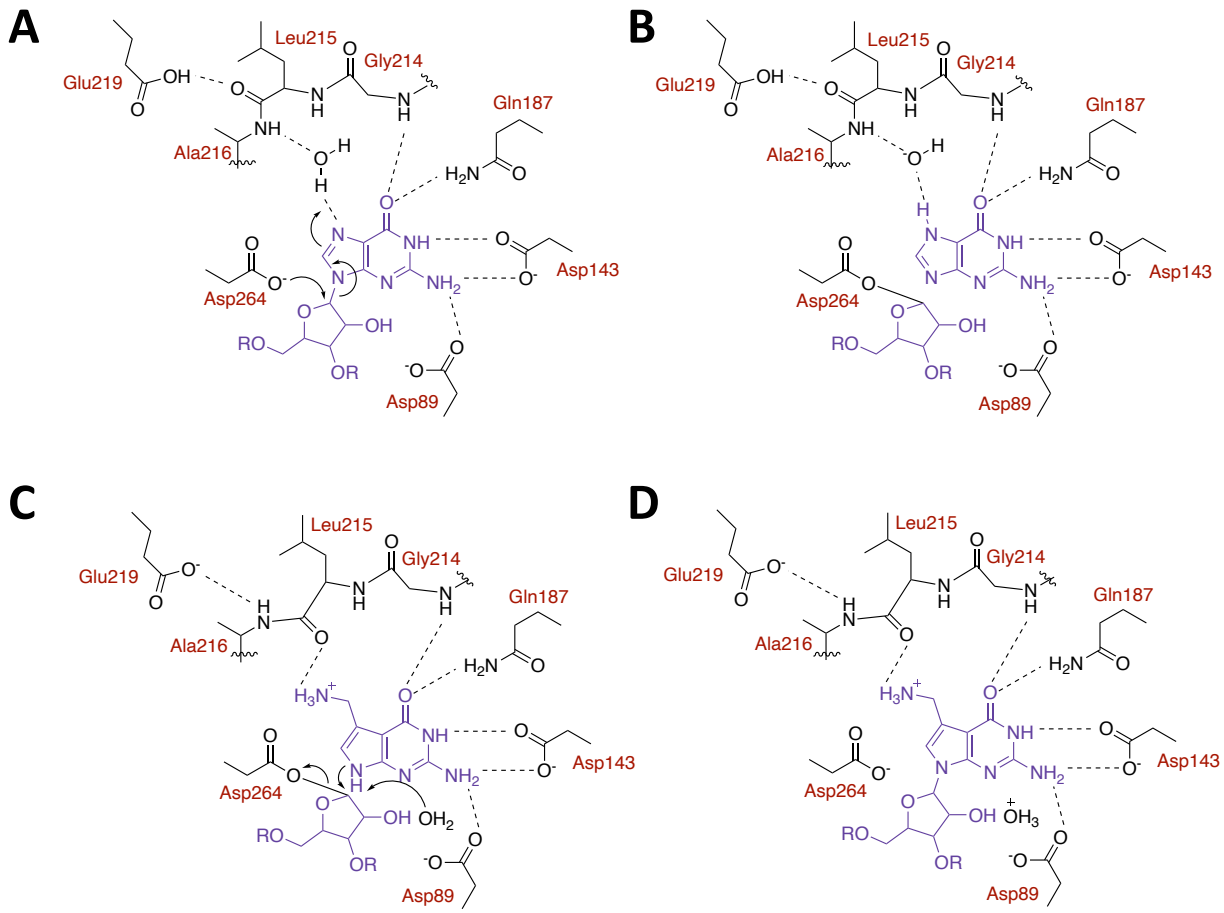



**Figure 1.2** Mechanism of RNA modification by E. coli TGT. (A) G34 binds the active site of TGT and Asp264 releases guanine by nucleophilic attack of the ribose anomeric carbon. (B) The covalent RNA-enzyme adduct. (C) PreQ1 replaces guanine in the active site and N9 releases the enzyme via nucleophilic attack at the anomeric carbon of the ribose at the 34 position. (D) The free enzyme and PreQ1 modified RNA substrate.

## 1.3 DNA modification by *E. coli* TGT

It has been documented in the literature that TGT is unable to modify the analogous DNA version of the 17 nucleotide ECY-A1 or extended ECY-MH hairpin substrates.[34] According to the published catalytic mechanism for TGTs and published findings, the deoxyribose backbone is not the culprit for the inability of TGT to turn over

DNA substrates (Figure 1.2.). For instance, it has been shown that a DNA substrate where all of the thymidine residues are swapped for 2-deoxyuridine is in fact a suitable substrate for *E. coli* TGT.[34] It has been hypothesized that the inability of the enzyme to act on DNA oligonucleotides containing thymidines is due to either specific recognition of the uracil bases or conformational differences induced by the different nucleobase. As such, it had not been shown that *E. coli* TGT can, in fact, recognize and modify DNA substrates in their native form containing only A, T, G, and C nucleobases.

A crystal structure of TGT from *E. coli* has yet to be determined, however, *Zymomonas mobilis* TGT has been crystalized many times in the presence of its RNA substrate and various small molecules.[35–42] A refined model of *E. coli* TGT based on a structure of *Z. mobilis* co-crystalized with a PreQ1 modified hairpin was generated using Schrodinger Prime modeling software and used to examine the potential limiting interactions between the substrate and enzymatic residues.[41] Based on these observations, and previously published data, I hypothesized that the limits of *E. coli* TGT toward modifying DNA substrates were in fact steric. Based on this hypothesis, this dissertation describes the rationally designed, iterative development of an unmodified DNA substrate that can be recognized and efficiently labeled by *E. coli* TGT. The versatility and efficiency of this labeling strategy holds great promise for its utility in cost effectively generating labeled oligonucleotides and its use in generating programmable macromolecular assemblies.

The third chapter of this work demonstrates the utility of my developed DNA modification system by employing it to generate fluorescent oligonucleotide probes for use with two widely used analytical techniques: northern blot analysis and fluorescence

in situ hybridization (FISH). Both analytical tools utilize antisense probes modified with a reporter and the current state-of-the-art probes for both approaches have their drawbacks. Namely, the cost of purchasing fluorescently labeled oligonucleotides and safety issues surrounding the use of radioactive probes. The versatility and efficiency of DNA-TAG offers a cost effective, easily executed, self-generated alternative for scientists to label oligonucleotides with their probe of choice and use them in their research.

### 1.4 Further Expanding the TGT Toolbox

Macromolecular assemblies are at the core of how we define and understand life. Combinations of peptides, nucleotides, glycans, and lipids, the four macromolecular building blocks, are the foundation of biology, making up all vital cellular components. These assemblies range in size and function, from the ribosome to the cell membrane, and are critical to every aspect of life. The utility of synthetic assemblies has become increasingly apparent as scientists uncover the complexity and function of their native counterparts. CRISPR/Cas systems are perhaps the most well-known molecular machines.[43–47] Scientists have hijacked this class of systems, native to bacteria and archaea, and developed a technology that employs the encodability of nucleic acids to program the functionality of enzymes. Beyond CRISPR/Cas systems, combining the programmability of base pairing and the functional importance of nucleic acids with the specificity and activity of polypeptides, lipids, and glycans has begun to transform the way we approach biological problem solving and therapeutic intervention.[48–50]es Naturally occurring macromolecular assemblies maintain order and structure through highly evolved intermolecular interactions; synthetic assemblies, on the other hand, must program in the desired order and structure, sometimes through the use of known affinity interactions but often requiring covalent linkages to join the different biomolecular classes.

Most current technologies which allow covalent conjugation of DNA and RNA to other biomolecules rely on chemical modifications, limiting their availability and utility with synthetic challenges and a high price tag.

The fourth chapter of this dissertation focuses on the development of a fully enzymatic protein – nucleic acid modification strategy. Self-labeling proteins such as SNAP-tag, CLIP-tag, and HaloTag are useful tools for labeling or functionalizing a protein of interest.[51–53] Utilizing a bifunctional small molecule probe bearing PreQ1 on one end and benzylguanine on the other, I successfully conjugate a functional enzyme SNAP-tag fusion to a target RNA. This simple conjugation strategy can be combined with DNA-TAG or nucleic acid guided strategies to expand its applications and utility.

## 1.5 Terminology, definitions, and assumptions

Throughout this dissertation I use several terms interchangeably and have several standard practices. Terminology, definitions, and standard practice assumptions are discussed here.

The literature on *E. coli* TGT substrate acceptance generally refers to the minimal and extended hairpin substrates as ECY-A1 and ECY-MH, respectively.[54] I will reference them as such in addition to the terminology used for RNA TAG: ECY-A1 = ECYA1 = TAG2 = TAG2.0, and ECY-MH = ECYMH = TAG3 = TAG3.0.[16] DNA versions of these substrates are indicated with a lowercase “d” in front of the term; for instance, dTAG3. When referring to mutant hairpins as TAG2 or TAG3 analogs or something similar I am referring to the length of the oligo not the sequence. The literature has also set a precedent for the numbering scheme of the ECY-A1 hairpin that is derived from the original $tRNA^{Tyr}$ sequence; namely, the UGU at the 7-8-9 positions of the hairpin are referenced in the literature as the 33-34-35 positions.[54] I use this literature precedent

when referring to a residue position, despite the length or origin of the hairpin sequence. This numbering scheme can be seen above (Figure 1.2.)

In general, when referring to tRNA guanine transglycosylases (TGT) it should be assumed that I am referencing TGT from *E. coli*, unless otherwise noted. RNA-TAG is the RNA modification technology previously developed by members of the lab and stands for RNA transglycosylation at guanine. The analogous term DNA-TAG is the name for the modification technology developed in this work and stands for DNA transglycosylation at guanine. For all DNA-TAG development experiments the small molecule probe should be assumed to be PreQ1-biotin unless otherwise noted (Figure 2.18.). Additionally, gel shift analysis of the hairpin labeling experiments is generally run on 20% urea PAGE in TBE unless otherwise noted. In general, gel shift assays are used to assess an increase in molecular weight from the insertion of the PreQ1 probe.

In designing the hairpin mutants, I used two labeling schemes: a contiguous number-based scheme that includes all mutants, as well as a naming scheme that indicated the intention of that mutant. A full list of numbers and names can be found in Table 2.1. Notably, the key DNA-TAG hairpin is referred to interchangeably as HP 176 and HIS M10, it is also the same as HP 130 although it is not referred to as such.

# 2 Developing DNA-TAG

## 2.1 Introduction

While harnessing the programmable power of nucleic acids is no new revelation for science, novel innovative applications that realize this power have been crucial to scientific advancements of late. These innovative strategies often rely heavily on nucleic acid modifications used for visualization, immobilization, conjugation, and affinity interactions, among others. Modification strategies include the site-specific insertion of synthetic nucleotides bearing a functional group of interest, non-site-specific chemical modification of nucleobases or the sugar backbone, and a limited set of enzymatic modification techniques. For many technical applications, precision is the key to its success, and it is necessary to have the means to carry out an efficient, site-specific modification of the nucleic acid substrate. While a variety of site-specific enzymatic RNA modification strategies have been well established, the same is not true for single stranded DNA (ssDNA) modification. Currently enzymatic modification of ssDNA is limited to 3′ insertion of modified nucleosides and the 5′ insertion of modified phosphate groups.[55,56] Previous work from our group has established RNA-TAG, an RNA modification strategy using tRNA guanine transglycosylase (TGT) (Figure 2.1).[16,57] TGT catalyzes the exchange of guanine for 7-deazaguanine derivatives on its native substrate, the anticodon loop of queuine-cognate tRNAs, however, it has been shown that the minimum recognition sequence for TGT is a short 17 nt hairpin with a uridine flanked target guanine in its loop.[27] In RNA-TAG, *E. coli* TGT recognizes this minimal hairpin loop incorporated into an RNA of interest and catalyzes the exchange of the target guanine for synthetically modified PreQ1 derivatives bearing a reporter, thus allowing the downstream application or investigation of the RNA of interest.

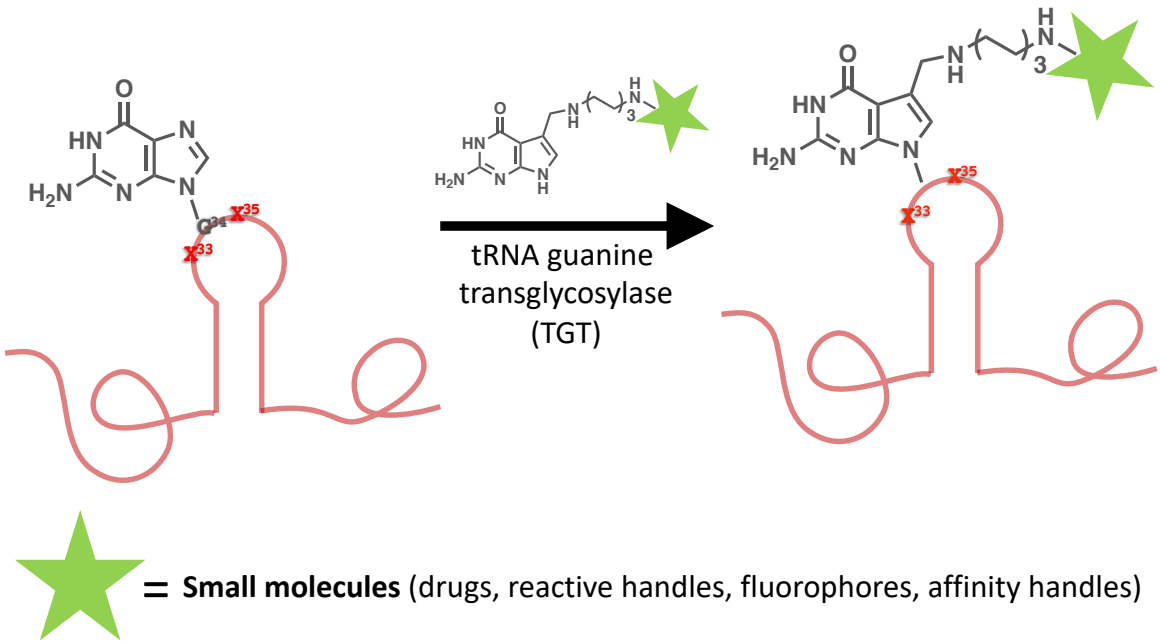



**Figure 2.1** RNA/DNA-TAG modification scheme. Inserting a 17-nucleotide hairpin into a ssDNA or RNA substrate allows for recognition and modification by *E. coli* TGT. A variety of reactive handles attached to PreQ1 via a 6-carbon linker on the exocyclic amine are successfully inserted at the 34 position, replacing the target guanine.

Previously, TGT was deemed incapable of modifying native DNA substrates, showing that nucleobase modifications were required for recognition and efficient recognition of the nucleic acid substrate.[34] Preliminary experiments using carefully designed DNA and RNA mutant oligonucleotide substrates led us to hypothesize that the inability of TGT to turn over the analogous DNA hairpin was likely due to steric limitations rather than specific base recognition. Through iterative testing of rationally designed DNA hairpins, we uncovered a series of several 17 nucleotide DNA hairpins that are efficiently recognized by the enzyme. These 17 nucleotide recognition sequences can either be inserted internally or appended to either end of ssDNA of interest allowing for installation of a functional handle. Additionally, the versatility of TGT and its tolerance toward substrate modification allows for the one step incorporation of all reporters tested, varying in size and overall charge.

Alluding to the sister technology, I have named this technique DNA transglycosylation at guanine (DNA-TAG). This chapter discusses the iterative design and development of DNA-TAG. As a three-component modification system, there were two avenues to follow to generate a system capable of inserting our PreQ1 derivatives into a DNA of interest. First, I explored *E. coli* TGT mutants designed based on a structural model estimated from a *Z. mobilis* TGT crystal structure.[41] Ultimately, this approach didn't generate any enzymes with higher activity than the wild type and I instead turned to finding a new DNA substrate with inspiration from the other TGT cognate tRNA anticodon loops, namely, $tRNA^{Asn}$, $tRNA^{Asp}$, and $tRNA^{His}$ (Figure 1.1.).

## 2.2 TGT activity toward DNA analogs of RNA substrates and their derivatives

The efficiency and reliability with which TGT inserts functional PreQ1 probes into an RNA of interest coupled with the published mechanism of TGT catalysis alludes to the possibility that the enzyme could similarly recognize DNA substrates (Figure 1.2.). While it was previously reported that TGT does not recognize the wild type DNA substrate analogous to the minimal 17 nt RNA hairpin substrate, when all "dT" nucleotides were replaced with "dU" the enzyme was able to recognize the DNA substrate.[34] With this information I designed a series of experiments to assess the limitations of TGT toward DNA substrates, limiting the number of dT→U mutations to only those that are part of the determined minimal recognition element in RNA, "UGU". For all of the preliminary experiments, the DNA oligos tested used the extended stem DNA analog of the ECY-A1 sequence, ECY-MH (GGGAGCAGAC**XGX**AAATCTGCTCCC) where the loop sequence is underlined, and the bolded residues are considered the "minimum recognition element" (known for RNA substrates to be UGU) (Figure 2.2.).

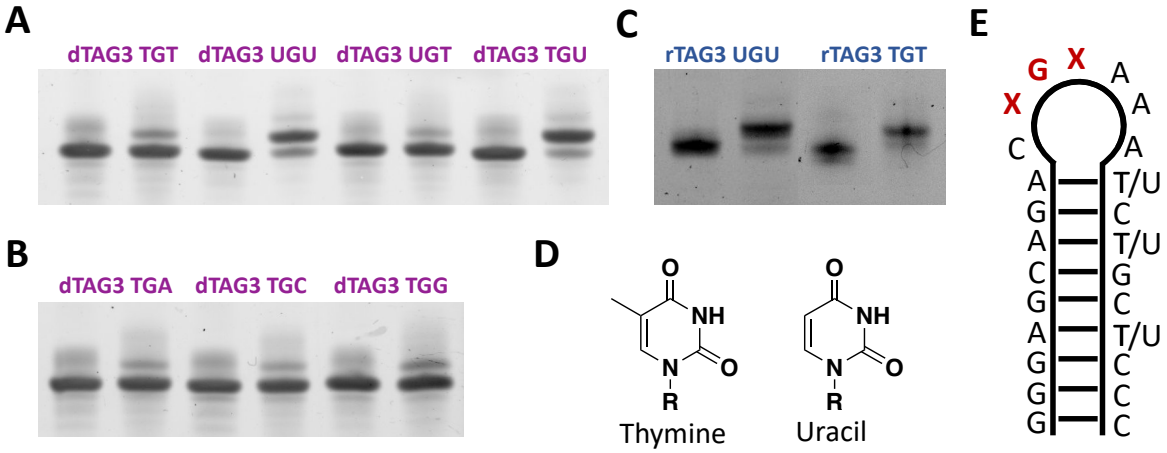



**Figure 2.2** PreQ1-biotin modifications of dTAG3.0 derived substrates. The inability of TGT to modify dECYMH is due to sterics/conformation of the substrate, not the presence of T rather than U. (A) Replacing both T residues or T35 only in the minimal recognition motif with dU allows *E. coli* TGT to recognize and modify DNA substrates with high efficiency. (B) Position 35 must have a dU residue to be modified by *E. coli* TGT; A, G, and C residues generate similar labeling efficiency to original T35 containing substrate. (C) RNA TAG3.0 labeling and TAG3.0 transcribed with 5-methyluridine instead of uridine are both fully labeled by TGT. (D) Nucleobase structures. (E) TAG3.0 hairpin constructs.

DNA oligonucleotides containing the following minimal recognition element mutations were tested for TGT activity using a PreQ1-biotin probe: TGT, dUGdU, dUGT, TGdU. While TGT and dUGT showed no appreciable turn over by the enzyme, both dUGdU and TGdU showed comparable turn over to the wildtype RNA substrate. Additional mutations of T35 in the loop sequence also proved not to be substrates for *E. coli* TGT. The single mutation of the minimal recognition element to TGdU allows for recognition and efficient turnover of a DNA substrate by TGT (Figure 2.2). This expands on previous knowledge which only assessed the turnover of a substrate where all dT nucleotides were replaced with dU. This promising information led to the next question: can TGT turn over an RNA substrate where all of the U nucleotides are replaced with 5-meU ("rT")? This substrate had a similar turnover to that of the wildtype RNA hairpin (Figure 2.2.). Ultimately, these experiments confirmed that *E. coli* TGT can accept substrates containing "T" bases and that having dU in position 35 is the minimal necessary modification for *E. coli* TGT to recognize and modify dECYA1.

## 2.3 Zymomonas mobilis

The benefit of DNA-TAG would be maximized with a TGT-DNA substrate pair that requires no chemical modification. At this point I was convinced that TGTs inability to turnover dECYA1 was merely steric and could therefore be overcome by mutating bulky residues in the active site of the enzyme. TGTs are conserved across all three kingdoms of life along with their mechanism and key residues.[24] The availability of a crystal structure led me to test the DNA labeling ability of another TGT from the bacteria *Zymomonas mobilis* (*Z. mobilis*).[41] *Z. mobilis* TGT was less amenable to changes, labeling only the wild type RNA hairpin (Figure 2.3.). While the *E. coli* and *Z. mobilis* TGTs have different primary sequences, their active site residues are conserved. I designed and expressed mutant *E. coli* TGTs using a model based on the *Z. mobilis* crystal structure and tested their activity toward d/rTAG substrates.

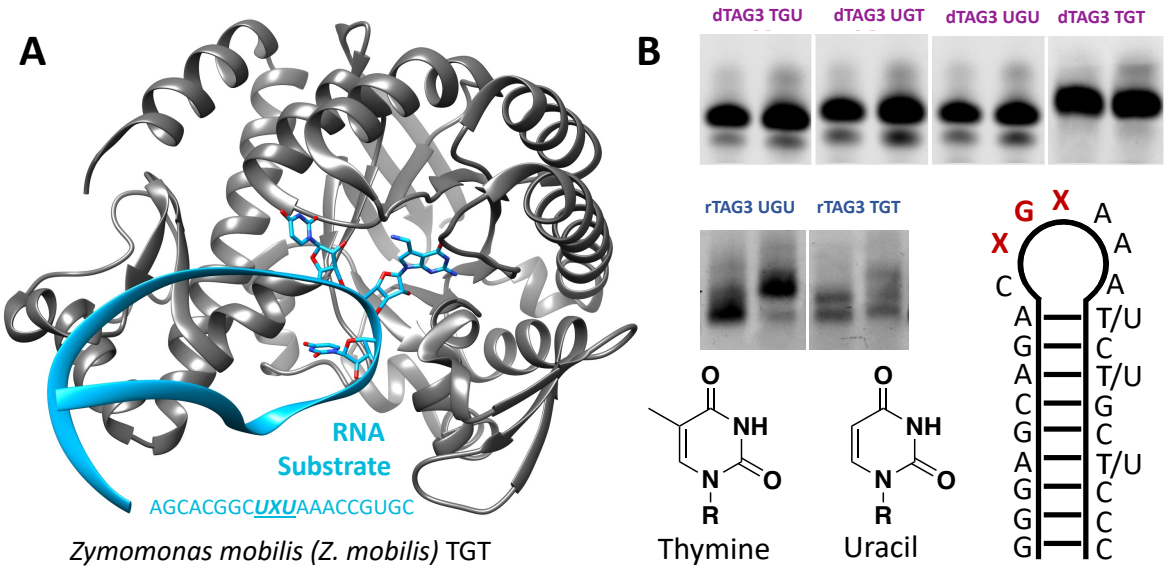



**Figure 2.3** TAG3.0 modification with PreQ1-biotin by Z. mobilis TGT. (A) Crystal structure of *Z. mobilis* bound to a PreQ1 modified RNA substrate. (B) *Z. mobilis* TGT modification of DNA and RNA TAG3.0 based constructs. *Z. mobilis* TGT fidelity is much higher than that of *E. coli* TGT, only efficiently modifying the original ECYA1 hairpin.

## 2.4 TGT mutants

Uracil and thymine are very similar nucleobases, the only difference being that thymine has a methyl group on the 5 position of its heterocycle while uracil has only a hydrogen (Figure 2.3). Since substrates with deoxyuridine at the 35 position are efficiently labeled by *E. coli* TGT while those with thymidine are not, I reasoned that this bulky methyl group was likely sterically hindering the oligo from sitting properly in the active site. A refined model of *E. coli* TGT based on a structure of *Z. mobilis* co-crystalized with a PreQ1 modified hairpin was generated by another lab member using Schrodinger Prime modeling software.[41] Residues in the binding pocket that interact with the U35 in the RNA substrate were identified and assumed to be those that were likely clashing with the T35 of the DNA hairpin. These residues include: Val32, Thr34, Val266, Thr269, and Arg273.

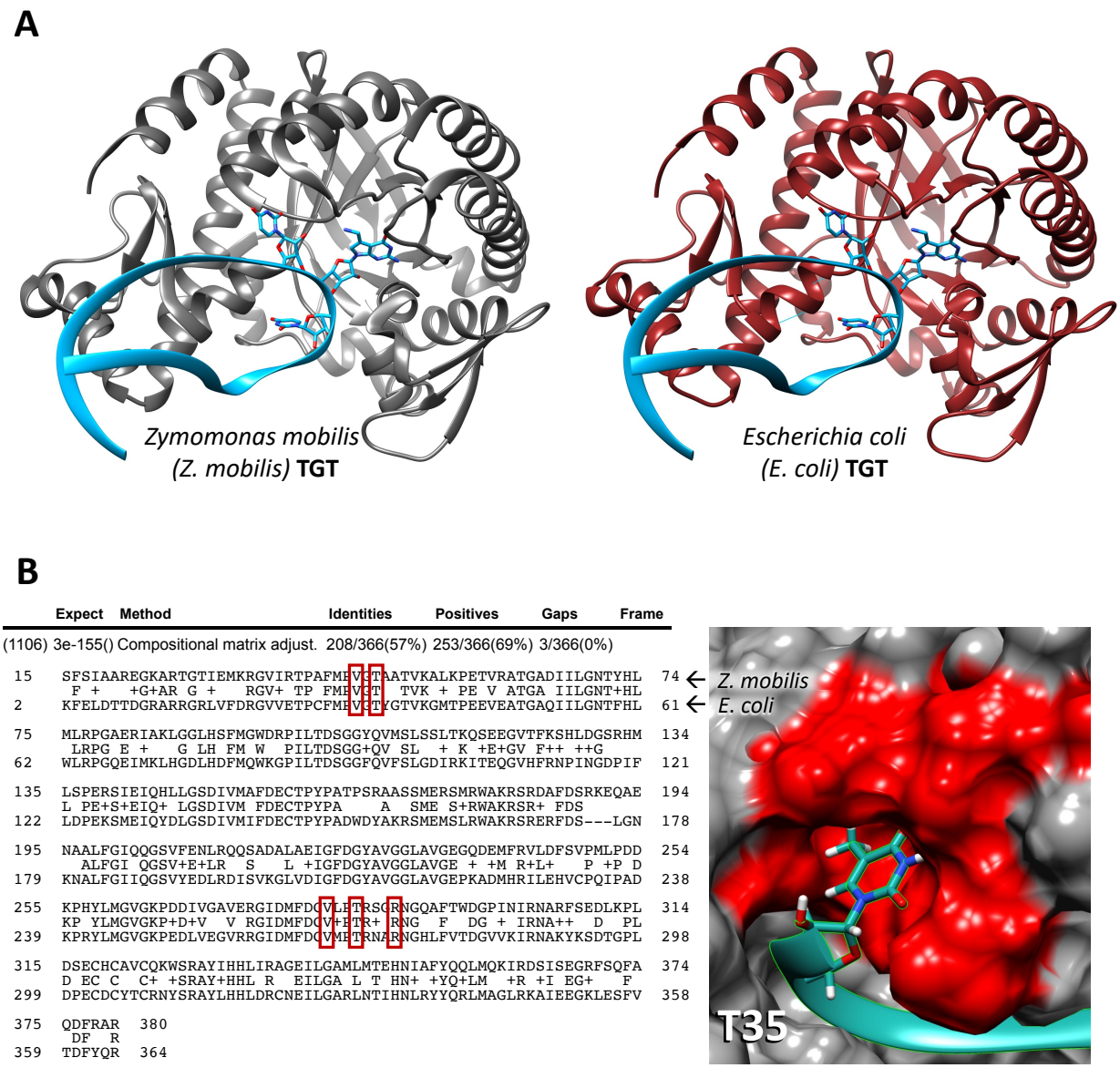



**Figure 2.4** Z. mobilis and E. coli TGT sequence alignment and modeling. (A) *Z. mobilis* TGT crystal structure and *E. coli* TGT fitted structure model. (B) *E. coli* and *Z. mobilis* sequence alignment and residues thought to interact with the residue at position 35 of the nucleic acid substrate. The highlighted residues were selected for mutation experiments.

A series of mutants targeting these residues were designed, cloned, expressed, and assessed for their activity toward dTAG3, the original RNA substrate, and a variety of *in vitro* transcribed RNA constructs. While all mutants maintained at least some activity toward the canonical RNA substrate, none of them showed improved activity toward the desired DNA substrate (Figure 2.5). Most mutations led to a more specific recognition of

the canonical RNA substrate however they were not further explored as alternative enzymes for the previously developed RNA-TAG technology. Interestingly and somewhat unsurprisingly, AGU, CGU, and GGU RNA substrates were modified fairly efficiently, challenging the previously published requirement of UGU as the minimal recognition motif for TGT activity.[27] Ultimately, enzymatic manipulation to extend TGT labeling to DNA substrates was not proving to be an efficient or productive direction. The next rational step was to try different DNA substrates.

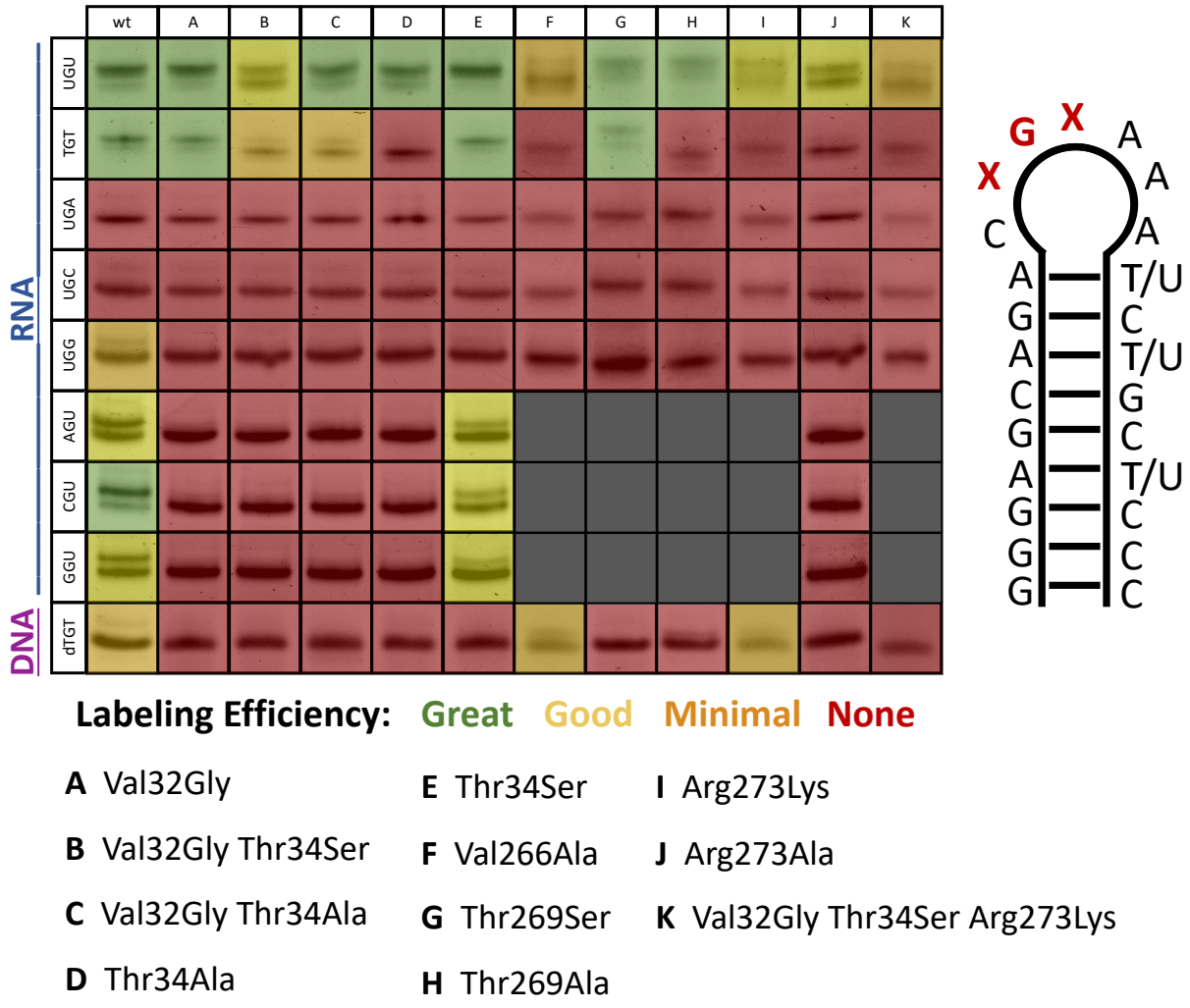



**Figure 2.5** TAG3.0 construct PreQ1-biotin modification by mutant E. coli TGTs. All mutants retained at least some activity toward the original RNA TAG3.0 substrate. None of the mutants improved TGT modification of DNA substrates.

## 2.5 DNA labeling of Hairpins based on cognate tRNAs

While there is likely a TGT mutant with the capability of labeling dECYA1, the process of designing and expressing TGT mutants proved time consuming and unproductive. As such, I decided to turn to the identity of the DNA substrate. We reasoned that by altering bases in both the stem and loop of the dECYA1 sequence, the conformation would alter just enough to sterically allow TGT to act on it. The clearest place to start was with DNA hairpins derived from other cognate tRNAs. The ECYA1 hairpin is derived from the anticodon loop of $tRNA^{Tyr}$. However, TGT has 3 other tRNA substrates: $tRNA^{His}$, $tRNA^{Asp}$, and $tRNA^{Asn}$ (Figure 1.1.). While they all have the same "UGU" minimal recognition element in their anticodon loops, the rest of the hairpin sequence varies. The 17-nucleotide hairpin oligos were ordered from IDT and tested with the wildtype TGT enzyme. To our surprise both the $tRNA^{Asp}$ and $tRNA^{His}$ derived hairpins were labeled with far greater efficiency, approximately 40%, as compared to that of the $tRNA^{Tyr}$ derived hairpin, approximately 5% (Figure 2.6). The $tRNA^{Asn}$ derived hairpin showed a strange degradation band that will be briefly discussed in a subsequent section. The extended 25 nucleotide hairpins (ECYMH) were also tested, and a similar labeling pattern was observed, although less efficiently.

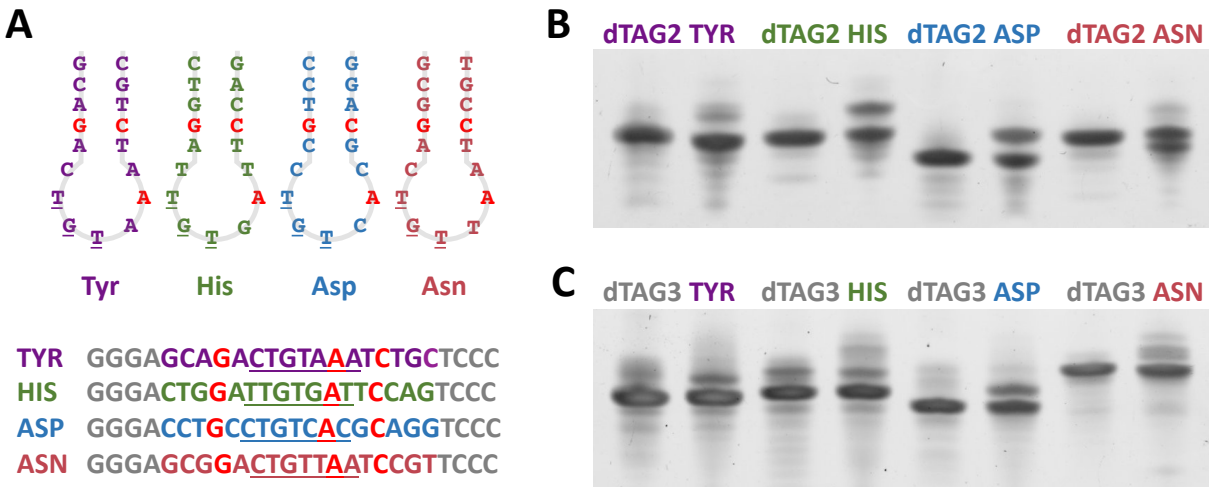



**Figure 2.6** TAG hairpins derived from TGT cognate tRNA substrates. (A) DNA hairpins and sequences of the four cognate tRNAs. The extended TAG3.0 version of the hairpins is indicated in grey. (B) *E. coli* TGT exhibits increased activity toward HIS and ASP based hairpins compared to the TYR based TAG hairpin. (C) Extended TAG3.0 analogs are accepted as well, although to a lesser extent than TAG2.0 analogs.

## 2.6 TGT Labeling of cognate tRNA chimera hairpins

To determine the impact of the identity of the stem and loop sequences on the ability of TGT to modify the DNA substrate I designed chimera hairpins, swapping the stems and loops of the 4 different sequences. These 16 chimera hairpins showed that, indeed, the stem and loop sequences do play different important roles in the ability of the enzyme to modify the substrate. Namely, the increased labeling of the substrate containing the tyrosine stem and aspartate loop sequence. Overall, the ASP-HIS and HIS-ASP mutants labeled most efficiently. Interestingly, all sequences containing the asparagine stem display the same degradation band, both in the labeled and unlabeled species. The next section outlines the iterative process used to find a DNA substrate that is efficiently labeled by *E. coli* TGT.

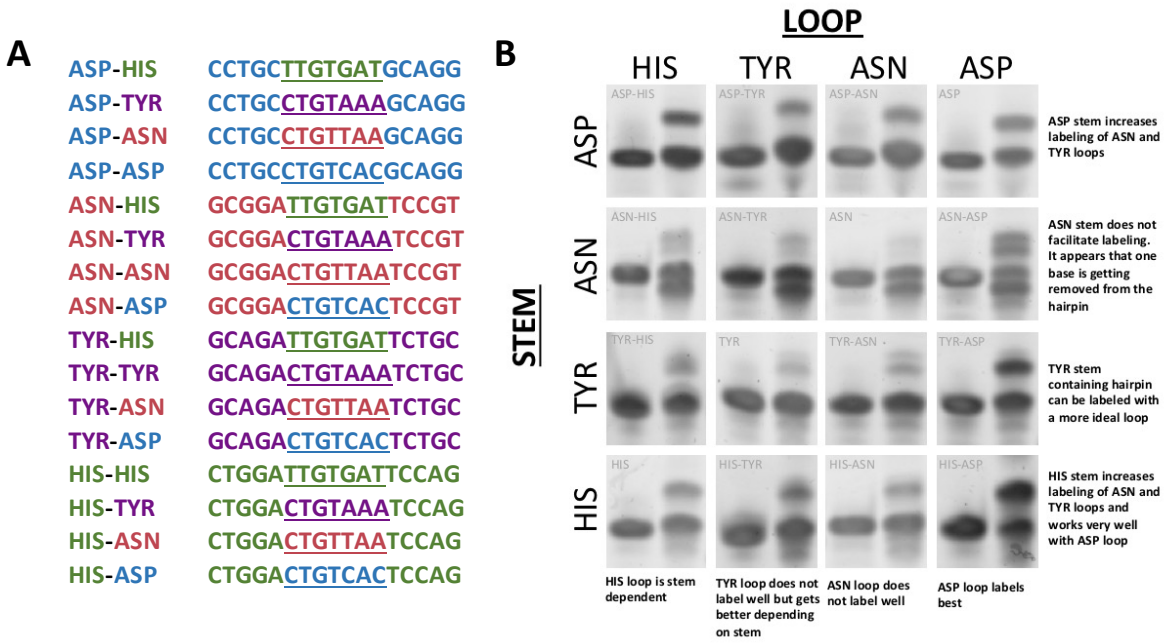



**Figure 2.7** E. coli TGT modification of stem-loop hybrid hairpins. (A) Hybrid hairpin sequences. (B) TGT modification via gel shift analysis of all hybrid hairpins. The ASN stem has a very regular degradation pattern. Overall, the loop seems to be more critical than the stem, as evidenced by ASP loop containing hairpins all retaining at least 40% labeling.

## 2.7 Rational Design and Testing of a Series of DNA Hairpin Substrates

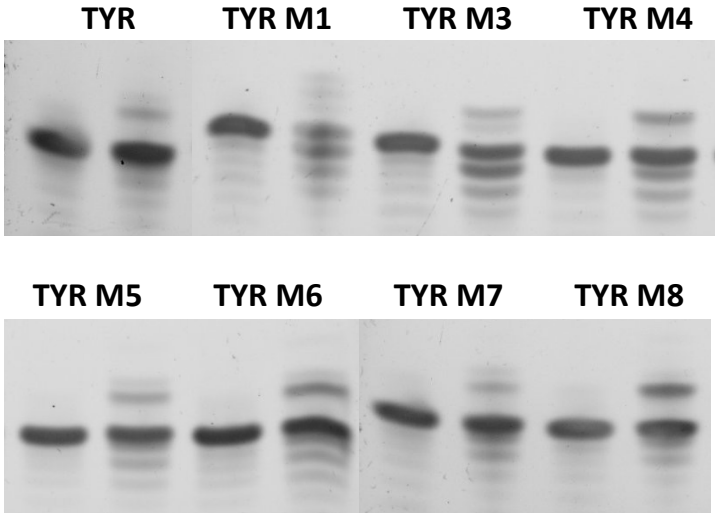
The sequences of the anticodon stem loops of the four cognate tRNAs vary greatly with the exception of the UGU motif, an A in the 37 position, and a G-C pair at the 30-40 position in the anticodon stem (Figure 2.7). This hinted that the sequence itself has very little if not no effect on the ability of the enzyme to recognize its substrate. I designed a variety of mutants of each tRNA analog, varying bases in both the stem and loop. While changes in the stem had some effect on the labeling, it was clear that modifying residues in the loop had a greater impact. I systematically and iteratively designed mutants, ultimately finding a histidine loop mutant that was able to be fully labeled.

I subsequently inserted this loop into a variety of stem sequences, revealing that while the stem did indeed play some role in the ability of TGT to label the substrate, the loop sequence is far more important. A variety of unique stems were able to be labeled

with approximately 100% efficiency and I moved forward with HIS M10 as the primary DNA-TAG substrate.

### *2.7.1 TYR Hairpin Mutants*

Mutations of the original $tRNA^{Tyr}$ based hairpin were designed to probe the site of degradation in the asparagine hairpin as well as the importance of the identity of base 36 or 37 for TGT recognition and turn over. When A's are replaced with C's in these positions I observed a slight increase in labeling efficiency, however nothing greater than the HIS-ASP or ASP-HIS chimera hairpins (Figure 2.8).




TYR GCAGA**CTGTAAA**TCTGC

TYR M1 GCAGA**CTGTAAA**TCTGT

~~TYR M2 GCAAA**CTGTAAA**TTGC~~

TYR M3 ACAGA**CTGTAAA**TCTGT

TYR M4 CCAGA**CTGTAAA**TCTGG

TYR M5 GCAGA**CTGTAAT**TCTGC

TYR M6 GCAGA**CTGTAAC**TCTGC

TYR M7 GCAGA**CTGTTAA**TCTGC

TYR M8 GCAGA**CTGTCAA**TCTGC

**Figure 2.8** TAG2.0-TYR mutant hairpin labeling. The only mutation that increased labeling of the TYR mutants was a A→C in the 36 position. Mutating the last residue to T caused a degradation similar to that seen in the ASN stem-based hairpins.

### *2.7.2 ASN Hairpin Mutants*

Hairpins containing stem sequences based on $tRNA^{Asn}$ were consistently showing a degradation product after the TGT reaction. The degradation was not random but rather seemed to be a single base judging from the gel shift. Being that TGT is a transglycosylase I hypothesized that the enzyme was acting on a base in addition to its target guanine, removing it, then being kicked off by water rather than a PreQ1 derivative.

The hairpin contains a G-T mismatch base pair at the base of the stem that I thought might be the likely culprit. Through testing a series of ASN mutants I showed that mutating it to an A-T pair did not eliminate the degradation while changing it to a G-C did, indicating that it was likely the “T” causing the degradation but not because it was a part of the mismatch (Figure 3.9). I used LCMS to determine the identity of the degradation product. Indeed, the mass of the degradation product was that of the hairpin minus a single T nucleobase (Figure 2.11). At this point it was unclear whether the degradation was from TGT, or a contaminant enzyme left over during purification.

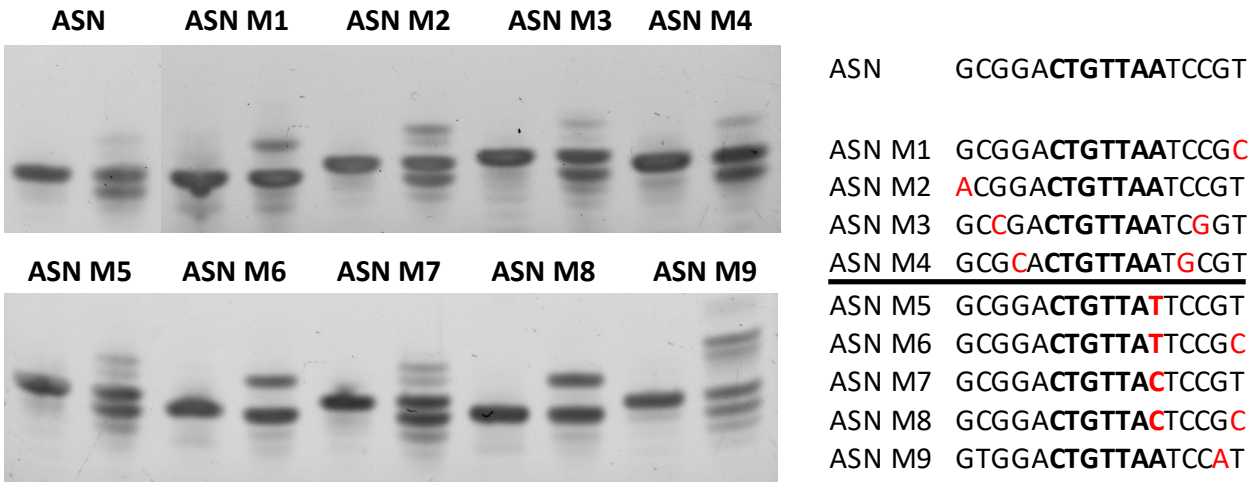



**Figure 2.9** TAG2.0-ASN mutant hairpin labeling. Mutating residue 38 to a smaller base (C or T) generated a significant increase modification by TGT. Changing the 3′ T residue to C eliminated degradation products.

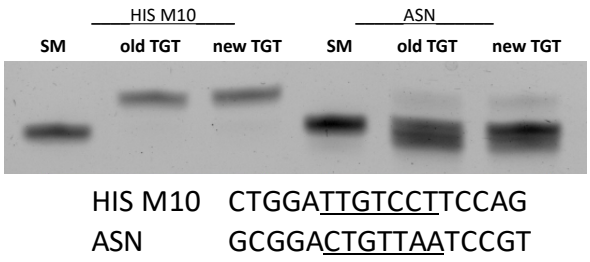
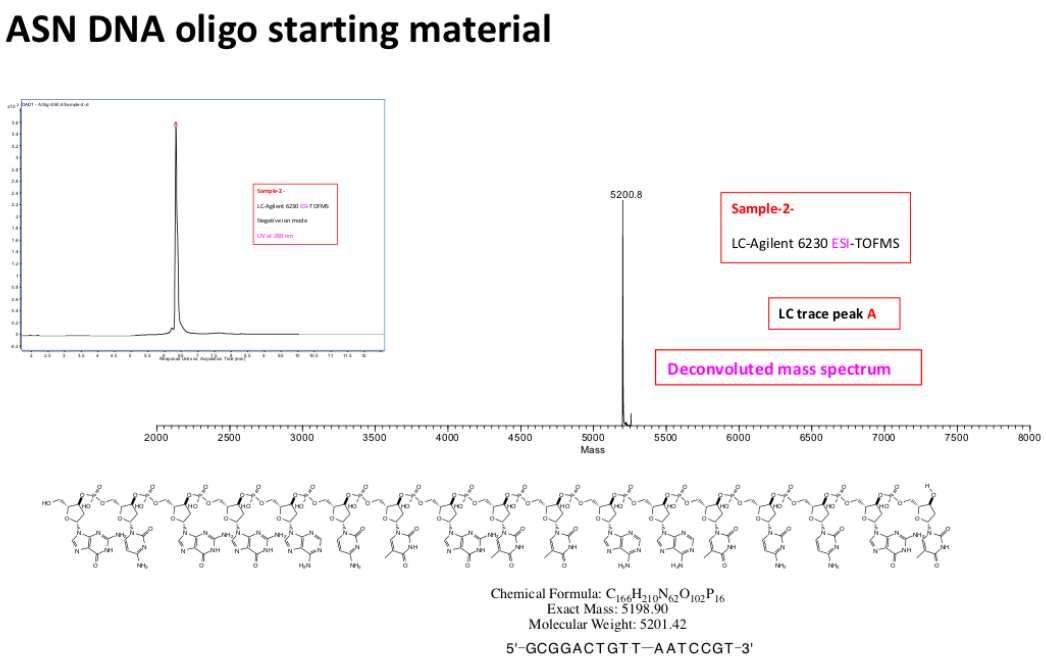



**Figure 2.10** Extra pure TGT prep labeling of ASN hairpin. Even with the additional purification handle the ASN based hairpin is still experiencing an excision of the 3′ T nucleobase. More experiments should be done to verify this is indeed being done by TGT and not some other contaminant.

At that point the enzyme prep involved two back-to-back His TAG purifications, which was sufficient to remove any RNAses but was perhaps not enough to remove an exonuclease with the ability to act on the 3′ end of this hairpin. Via protein gel the enzyme seemed to be pure, with no abundant contaminants. Because of this we were not sure that FPLC would give an answer to the problem and instead decided to clone in another, more specific affinity tag. The new construct has an N-terminal streptactin tag and a C-terminal His tag. The prep was first His purified and subsequently strep tag purified. The resulting enzyme prep still causes degradation (Figure 2.10). Since this phenomenon is only observed for oligonucleotides with 3′ Ts we did not look into the phenomenon further, although it would be an interesting endeavor for future interrogations.



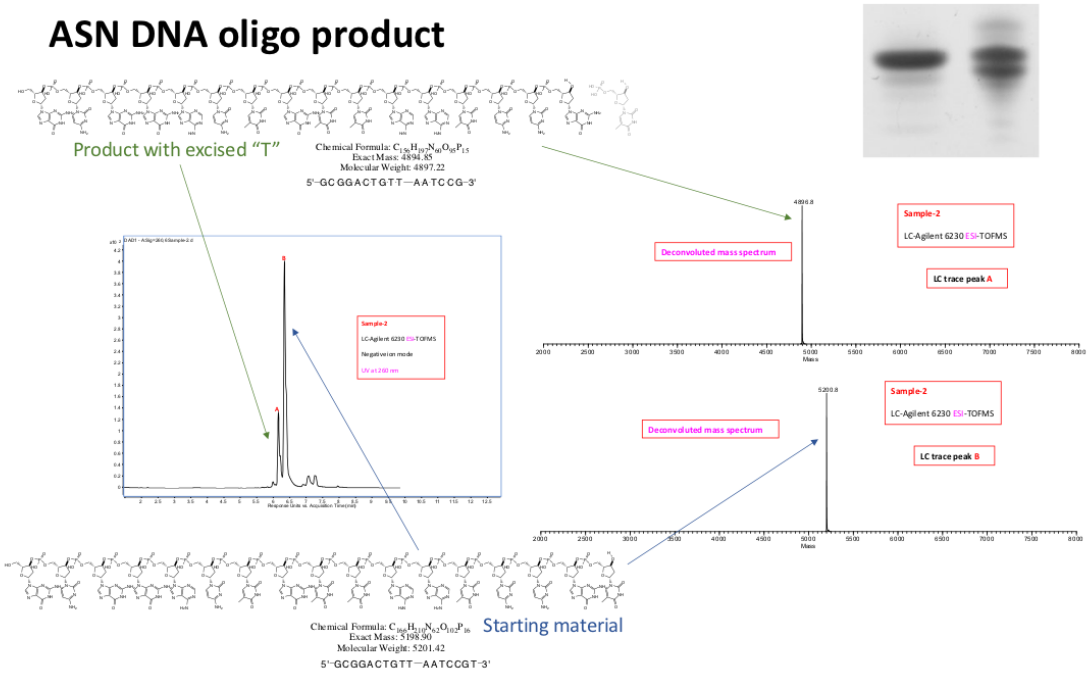
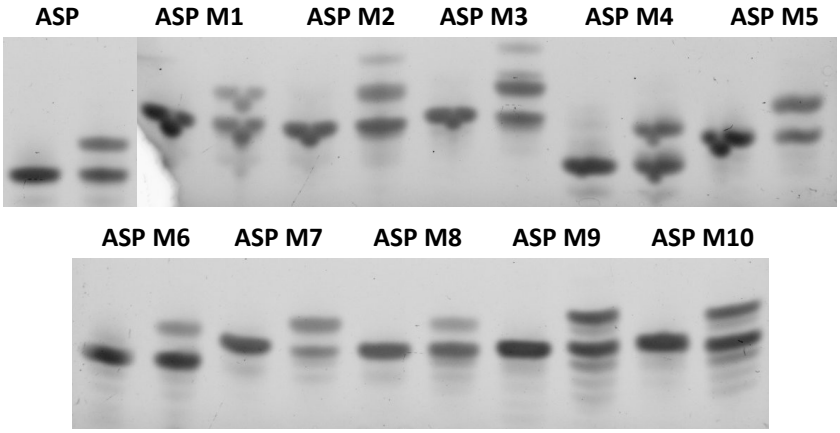




**Figure 2.11** LCMS analysis of ASN degradation product. The starting material and product LC traces have one and two predominant peaks, respectively, as expected. The difference in mass of the degradation product to the starting material is exactly that of thymine, which is in agreement with mutation experiments.

### *2.7.3 ASP Hairpin Mutants*

As one of the more promising starting sequences, the $tRNA^{Asp}$ mutants were designed to explicitly report on the importance of both the stem identity and the loop sequence. Notably, just like with other mutants, changes to the loop seemed to have a greater impact to TGT labeling than those made in the stem sequence. For instance, using a mirrored stem sequence (ASP M2) did not seem to have a huge effect on the extent of PreQ1 labeling by TGT while changing A37 to C (ASP M5 and M7) visually bumped labeling efficiency over 50% (Figure 2.12).



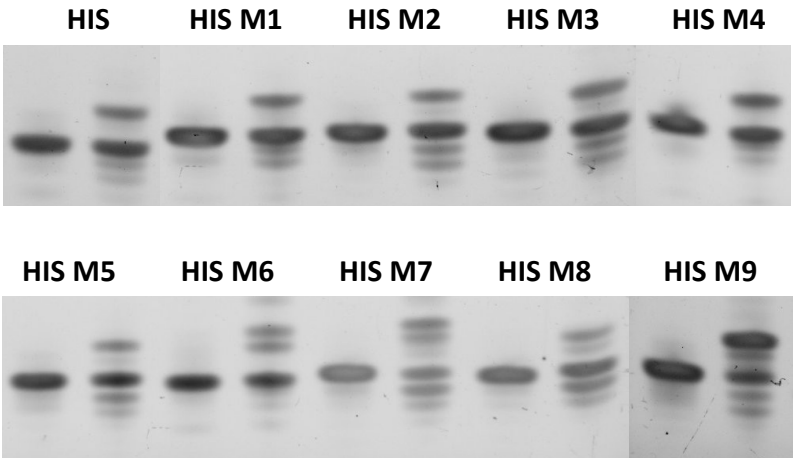

| | |
|---|---|
| ASP | CCTGC**CTGTCAC**GCAGG |
| ASP M1 | CCTCG**CTGTCAC**CGAGG |
| ASP M2 | GGACG**CTGTCAC**CGTCC |
| ASP M3 | GGAGC**CTGTCAC**GCTCC |
| ASP M4 | CCGCG**CTGTCAC**CGCGG |
| ASP M5 | CCTGC**CTGTCCC**GCAGG |
| ASP M6 | CCTGC**CTGTCTC**GCAGG |
| ASP M7 | CCTCG**CTGTCCC**CGAGG |
| ASP M8 | CCTCG**CTGTCTC**CGAGG |
| ASP M9 | CCTGA**CTGTCAC**TCAGG |
| ASP M10 | CCTCA**CTGTCAC**TGAGG |

**Figure 2.12** TAG2.0-ASP mutant hairpin labeling. This round of mutants confirms a greater importance of loop sequence over stem sequence as evidenced by M2, which has a reversed stem sequence but exhibits a similar degree of labeling. Replacing the A37 residue with a C seems to have the greatest positive effect, as seen in M5 and M7. Interestingly, changinf that residue to the other smaller pyrimidine base, T, had a negative impact on labeling efficiency.

### *2.7.4 HIS Hairpin Mutants*

The first round of $tRNA^{His}$ mutants provided insight supporting the hypothesis that the ASN degradation product is likely due to the terminal 3′ T residue as well as confirming once again that loop residues play a greater role in determining TGT recognition than those of the stem. The most promising candidate proved to be the HIS M9 mutant. With a single mutation of base 36 from A→C (HIS M9) the labeling seemed to increase to ~75% by eye (Figure 2.13). The first round of mutants provided a lot of insight into the

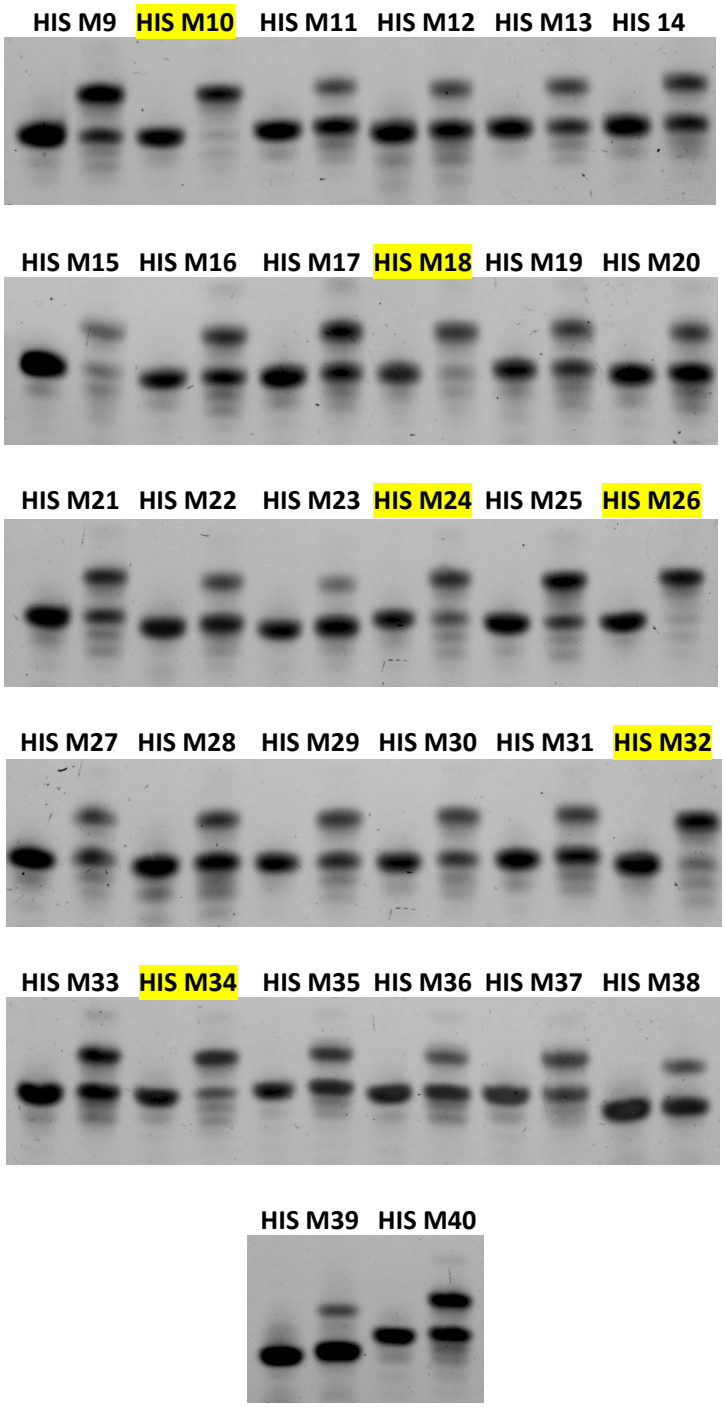
allowances TGT could tolerate. With this information I designed an additional 30 $tRNA^{His}$ derived mutants focusing on changing the bases in the loop.




| | |
|---|---|
| HIS | CTGGA**TTGTGAT**TCCAG |
| HIS M1 | CTCGA**TTGTGAT**TCGAG |
| HIS M2 | CTAGA**TTGTGAT**TCTAG |
| HIS M3 | CTGGA**TTGTTAT**TCCAG |
| HIS M4 | CTGCA**TTGTGAT**TGCAG |
| HIS M5 | CTGAA**TTGTGAT**TTCAG |
| HIS M6 | GTGGA**TTGTGAT**TCCAC |
| HIS M7 | GTGGA**TTGTGAT**TCCAT |
| HIS M8 | ATGGA**TTGTGAT**TCCAT |
| HIS M9 | CTGGA**TTGTCAT**TCCAG |

**Figure 2.13** Round one of TAG2.0-HIS mutant hairpin labeling. Just as with the TYR and ASP mutants, changing the last reside to a T induced degradation as seen with the ASN stem containing hairpins. Interestingly, mutating to a 5′ G seems to cause double labeling. Overall, HIS M9 showed the greatest improvement in labeling efficiency and prompted a second set of HIS mutants where I focused on the loop.

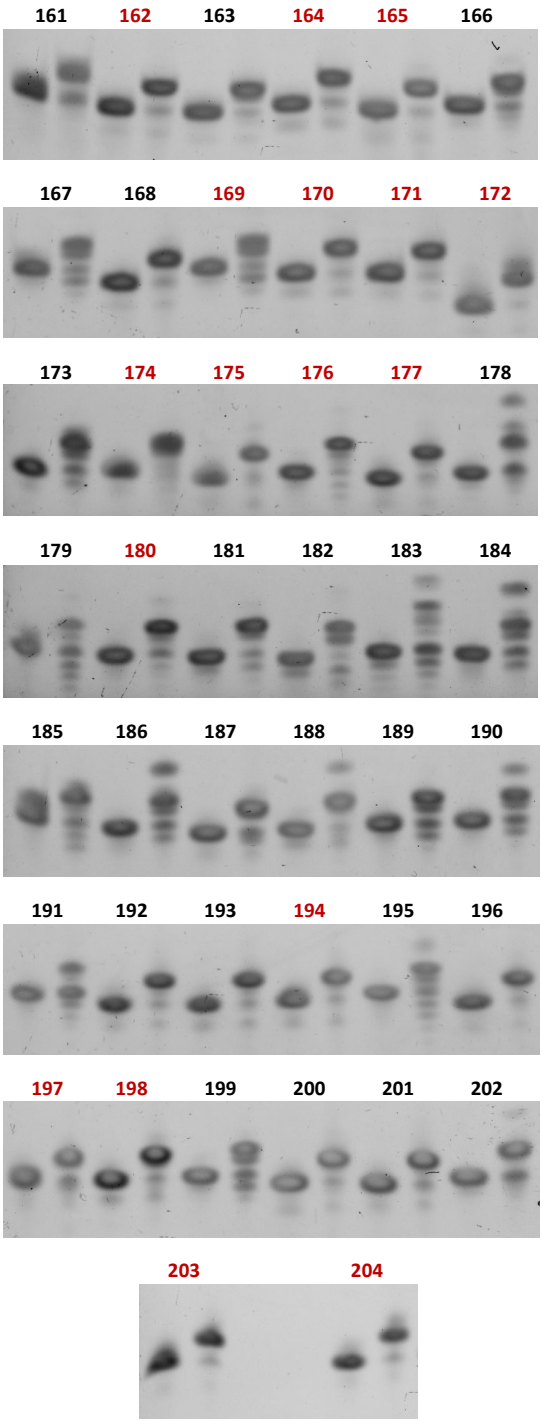
This set of mutants provided insight into the minimal requirements for TGT labeling of DNA hairpins with PreQ1 derivatives. Namely, a minimal substrate loop requirement of either XTGTCCX or XTGTXCC where X contains smaller pyrimidine bases (C or T) (Figure 2.14). in general, having a purine base other than the target guanine in the loop decreases labeling efficiency, confirming the major role of sterics in determining TGT activity toward DNA substrates. More work might be done in this area to test the limits and utility of TGT for DNA labeling. For instance, published work has shown that TGT is able to recognize RNA hairpins with a shifted UGU motif. With an ideal loop sequence in hand, I decided to move forward with HIS M10: CTGGATTGTCCTTCCAG to test the effect of the stem sequence.




| | |
|---|---|
| HIS M9 | CTGGA**TTGTCA**TTCCAG |
| HIS M10 | CTGGA**TTGTCC**TTCCAG |
| HIS M11 | CTGGA**TTGTTT**TCCAG |
| HIS M12 | CTGGA**TTGTTA**TTCCAG |
| HIS M13 | CTGGA**TTGTCT**TCCAG |
| HIS M14 | CTGGA**TTGTCG**TTCCAG |
| HIS M15 | CTGGA**TTGTTG**TTCCAG |
| HIS M16 | CTGGA**TTGTTC**TTCCAG |
| HIS M17 | CTGGA**CTGTCA**CTCCAG |
| HIS M18 | CTGGA**CTGTCCC**TCCAG |
| HIS M19 | CTGGA**CTGTTTC**TCCAG |
| HIS M20 | CTGGA**CTGTTA**CTCCAG |
| HIS M21 | CTGGA**CTGTCTC**TCCAG |
| HIS M22 | CTGGA**CTGTCGC**TCCAG |
| HIS M23 | CTGGA**CTGTTGC**TCCAG |
| HIS M24 | CTGGA**CTGTTCC**TCCAG |
| HIS M25 | CTGGA**TTGTCA**CTCCAG |
| HIS M26 | CTGGA**TTGTCCC**TCCAG |
| HIS M27 | CTGGA**TTGTTTC**TCCAG |
| HIS M28 | CTGGA**TTGTTA**CTCCAG |
| HIS M29 | CTGGA**TTGTCTC**TCCAG |
| HIS M30 | CTGGA**TTGTCGC**TCCAG |
| HIS M31 | CTGGA**TTGTTGC**TCCAG |
| HIS M32 | CTGGA**TTGTTCC**TCCAG |
| HIS M33 | CTGGA**CTGTCA**TTCCAG |
| HIS M34 | CTGGA**CTGTCC**TTCCAG |
| HIS M35 | CTGGA**CTGTTT**TCCAG |
| HIS M36 | CTGGA**CTGTTA**TTCCAG |
| HIS M37 | CTGGA**CTGTCT**TCCAG |
| HIS M38 | CTGGA**CTGTCG**TTCCAG |
| HIS M39 | CTGGA**CTGTTG**TTCCAG |
| HIS M40 | CTGGA**CTGTTC**TTCCAG |

**Figure 2.14** Round two of TAG2.0-HIS mutant hairpin labeling. This round of mutations was the most informative providing the two key aspects of the loop sequence that allow efficient DNA labeling: C or T in position 32 and CCX or XCC (X=C or T) in positions 36-38. HIS M10 was used as the primary DNA-TAG hairpin for all subsequent experiments.

### 2.7.5 *His M10 Loop Stem Mutant Hairpins*

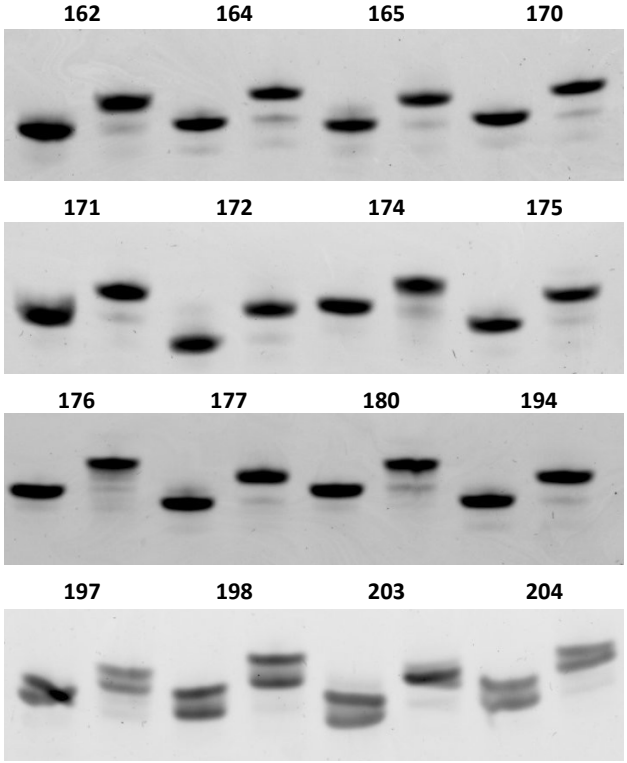



161_Ala1_H10 CTT GCT TGT CCT GCA AG
**162_Ala2_H10 CCT GCT TGT CCT GCA GG**
163_arg1_H10 AGC CCT TGT CCT GGG CT
**164_arg2_H10 CTC GGT TGT CCT CCG AG**
**165_arg3_H10 CTG CCT TGT CCT GGC AG**
166_arg4_H10 ACG ACT TGT CCT GTC GT
167_asn1_H10 GCG GAT TGT CCT TCC GT
168_asp1_H10 CCT GCT TGT CCT GCA GG
169_cys1_H10 GCG GAT TGT CCT TCC GT
**170_gln1_H10 CCG GAT TGT CCT TCC GG**
**171_gln2_H10 CCG GTT TGT CCT ACC GG**
**172_glu1_H10 CCG CCT TGT CCT GGC GG**
173_gly1_H10 AGA GCT TGT CCT GCT CT
**174_gly2_H10 TCA GCT TGT CCT GCT GA**
**175_gly3_H10 CGA CCT TGT CCT GGT CG**
**176_his1_H10 CTG GAT TGT CCT TCC AG**
**177_ile1_H10 CAC CCT TGT CCT GGG TG**
178_leu1_H10 GTT GAT TGT CCT TCA AC
179_leu2_H10 CTA GCT TGT CCT GTT AG
**180_leu3_H10 CTA CCT TGT CCT GGT AG**
181_leu4_H10 CCA GAT TGT CCT TCT GG
182_leu5_H10 AGG GAT TGT CCT TCC CT
183_lys1_H10 GTT GAT TGT CCT TCA AT
184_met1_H10 GGC GAT TGT CCT TCG CT
185_met2_H10 CAT CAT TGT CCT TGA TG
186_met3_H10 GGC GAT TGT CCT TCG CT
187_met4_H10 TCG GGT TGT CCT CCC GA
188_phe1_H10 GGG GAT TGT CCT TCC CC
189_pro1_H10 ACT GGT TGT CCT CCA GT
190_pro2_H10 CCG TCT TGT CCT GTC GG
191_pro3_H10 CTT CGT TGT CCT CGA GG
192_ser1_H10 CCG GTT TGT CCT ACC GG
193_ser2_H10 CCG GTT TGT CCT ACC GG
**194_ser3_H10 CTC CCT TGT CCT GGG AG**
195_ser4_H10 CAC GCT TGT CCT GTG TG
196_thr1_H10 CAC CCT TGT CCT GGG TG
**197_thr2_H10 TCG CAT TGT CCT TGC GA**
**198_thr3_H10 CGC ATT TGT CCT ATG CG**
199_thr4_H10 ACT GAT TGT CCT TCA GT
200_thr5_H10 CAC CCT TGT CCT GGG TG
201_trp1_H10 CCG GTT TGT CCT ACC GG
202_tyr1_H10 GCA GAT TGT CCT TCT GC
**203_val1_H10 CCA CCT TGT CCT GGT GG**
**204_val2_H10 CCT CCT TGT CCT GGA GG**

**Figure 2.15** TAG2.0-HIS M10 stem mutant hairpin labeling. The stems of all *E. coli* tRNA anticodon stem loops were appended with the HIS M10 loop sequence. All stem mutants were labeled although some suffered from 3′ T excision and others were labeled multiple times. 16 candidate sequences were taken as the optimal hairpins to be used in the DNA-TAG labeling system (labeled in red). HIS M10 (176 here) was still taken forward for all subsequent experiments unless otherwise noted.

With a loop that I know is efficiently labeled I set to assess the tolerance of TGT toward different stem sequences. I pulled the anticodon loop stem sequences from human tRNAs and inserted the HIS M10 loop into each one. Overall, the data show that TGT is not selective toward stem sequence as long as the loop sequence is accepted. Except for some multi-labeling events and degradation similar to the $tRNA^{Asn}$ stem hairpin analogs, every sequence was labeled and most with high efficiency (Figure 2.15).

From this small set of possible hairpins, I identified 16 unique hairpins that are effectively quantitatively modified by TGT, arbitrarily denoted as: 162, 164, 165, 170, 171, 172, 174, 175, 176, 177, 180, 194, 197, 198, 203, 204. I chose to move forward with the HIS M10 mutant, which is hairpin 176 in this list, for all further experiments unless otherwise noted (Figure 2.16).

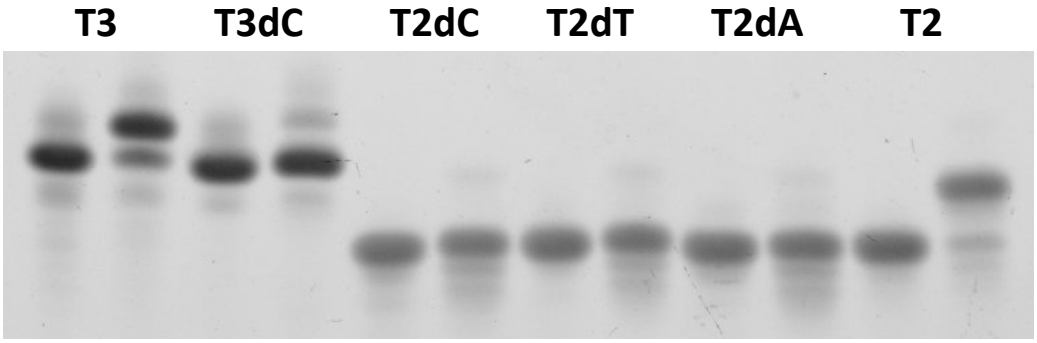



| | |
|---|---|
| 162_Ala2_H10 | CCT GCT TGT CCT GCA GG |
| 164_arg2_H10 | CTC GGT TGT CCT CCG AG |
| 165_arg3_H10 | CTG CCT TGT CCT GGC AG |
| 170_gln1_H10 | CCG GAT TGT CCT TCC GG |
| 171_gln2_H10 | CCG GTT TGT CCT ACC GG |
| 172_glu1_H10 | CCG CCT TGT CCT GGC GG |
| 174_gly2_H10 | TCA GCT TGT CCT GCT GA |
| 175_gly3_H10 | CGA CCT TGT CCT GGT CG |
| 176_his1_H10 | CTG GAT TGT CCT TCC AG |
| 177_ile1_H10 | CAC CCT TGT CCT GGG TG |
| 180_leu3_H10 | CTA CCT TGT CCT GGT AG |
| 194_ser3_H10 | CTC CCT TGT CCT GGG AG |
| 197_thr2_H10 | TCG CAT TGT CCT TGC GA |
| 198_thr3_H10 | CGC ATT TGT CCT ATG CG |
| 203_val1_H10 | CCA CCT TGT CCT GGT GG |
| 204_val2_H10 | CCT CCT TGT CCT GGA GG |

**Figure 2.16** Stem mutant top hits labeling. The top hits from the stem mutant set were tested once more, confirming their efficient labeling by *E. coli* TGT. The dual bands the gel containing HP 197-204 are an artifact and additional experiments confirmed that these hairpins were indeed fully labeled.

## 2.8 HIS M10 Final DNA-TAG Substrate Characterization

Finally, I fully characterized the HIS M10 mutant, otherwise noted as hairpin 130 or 176, as it is the primary hairpin used for all DNA-TAG experiments and applications going forward. It has been shown in the literature that an extended 25 nucleotide version of the RNA hairpin can be more efficiently and reliably labeled as it has a higher melting temperature.[57] The analogous extended HIS M10 DNA hairpin was assessed for labeling and was labeled with similar efficiency to its 17-nucleotide counterpart. Additionally, the ΔC versions of the 25-nucleotide hairpin and the ΔC, ΔT, and ΔA versions of the 17-nucleotide hairpin were not labeled by TGT, confirming that the G34 is in fact what is being modified by the enzyme (Figure 2.17).




| | |
|---|---|
| His M10 T3 | GGGACTGGA**TTGTCCT**TCCAGTCC |
| His M10 T3 dC | GGGACTGGA**TTCTCCT**TCCAGTCC |
| His M10 T2 dC | CTGGA**TTCTCCT**TCCAG |
| His M10 T2 dT | CTGGA**TTTTCCT**TCCAG |
| His M10 T2 dA | CTGGA**TTATCCT**TCCAG |
| His M10 T2 | CTGGA**TTGTCCT**TCCAG |

**Figure 2.17** HIS M10 TAG2.0 and TAG3.0 negative controls. The extended TAG3.0 version of HIS M10 was labelled efficiently. TAG2.0 mutants with ΔC, ΔT, and ΔA mutations at position 34 were unable to be labeled, as expected as well as TAG3.0 ΔC. This data confirms that G34 is what is being replaced by *E. coli* TGT.

A valuable feature of nucleic acid labeling by TGT is the versatility of the probes which are able to be accepted by the enzyme. Previous work has shown that TGT is tolerant of a variety of probes for RNA modification, including neutral, positive, and negatively charged reporters.[16,18–21,57–59] The 17 nucleotide HIS M10 DNA substrate was efficiently labeled with all probes tested. The probes include: PreQ1-biotin, -tetramethylrhodamine, -Cy5, -AlexaFluor647, -AlexaFluor488, -benzyl guanine, and -SiR (Figures 2.18, 3.4).

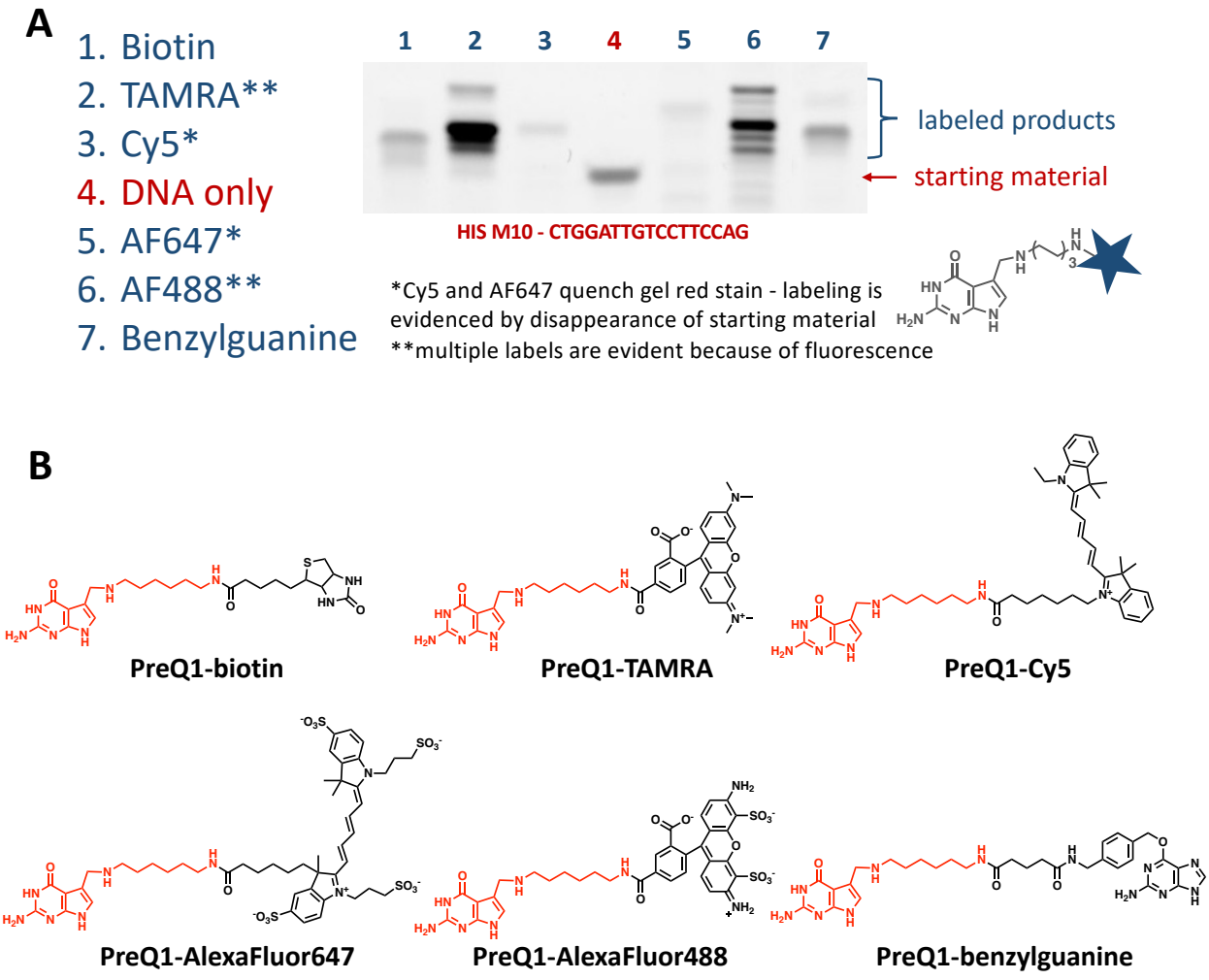
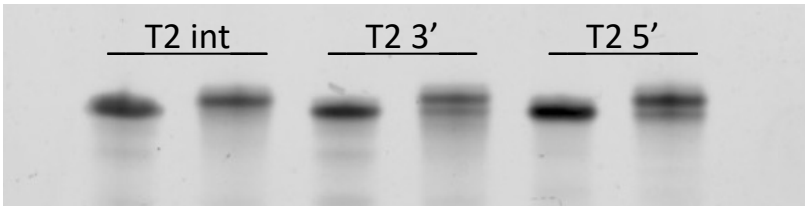



**Figure 2.18** HIS M10 TGT labeling with various probes. (A) TGT is able to efficiently modify the HIS M10 DNA substrate with neutral, positive, and negatively charged probes as evidenced by the disappearance of starting material and/or the appearance of a product band. Multiple labels are evident in highly fluorescent products although they are in small amounts (likely <5%). (B) PreQ1 probe structures. All probes are synthesized via NHS chemistry, purified by HPLC, and verified by LCMS.

## 2.9 Further Development of DNA-TAG Modification Technology

In order for DNA-TAG labeling technology to be a useful tool it is imperative that the 17-nucleotide hairpin can be recognized in the context of a DNA species of interest. The hairpin was inserted internally and into the 5′ or 3′ ends of a 59 mer oligonucleotide and subjected to TGT labeling. All constructs were efficiently labeled by TGT, confirming its potential utility for a number of applications (Figure 2.19).




His M10 T2 int ACCATCATTTTCATATCCTCCACTGGA**TTGTCCT**TCCAGACCACCATCATTTGCAATGA

His M10 T2 3' ACCATCATTTTCATATCCTCCAACCACCATCATTTGCAATGACTGGA**TTGTCCT**TCCAG

His M10 T2 5' CTGGA**TTGTCCT**TCCAGACCATCATTTTCATATCCTCCAACCACCATCATTTGCAATGA

**Figure 2.19** HIS M10 construct placement. The HIS M10 construct can be efficiently labeled when it is at either end of a ssDNA construct or internal. This data is important for the utility of the technology in interesting applications.

Next, I assessed the ability of TGT to insert multiple probes into a single oligonucleotide. For these experiments I either used two of the same or two different TAG hairpins (HP 176 and HP 177), placed them both on the 5′ or 3′ or opposite ends, and also tested single ΔC mutants to see the difference between single and double labeling. Most constructs were fully labeled as expected, however when HP 176 was at the 3′ end of the oligo incomplete labeling was observed. This could be due to improper annealing or conformation of the oligos themselves. This was not further investigated as all other constructs were labeled as expected. Notably, having tandem and split identical hairpins did not affect modification by TGT (Figure 2.20).

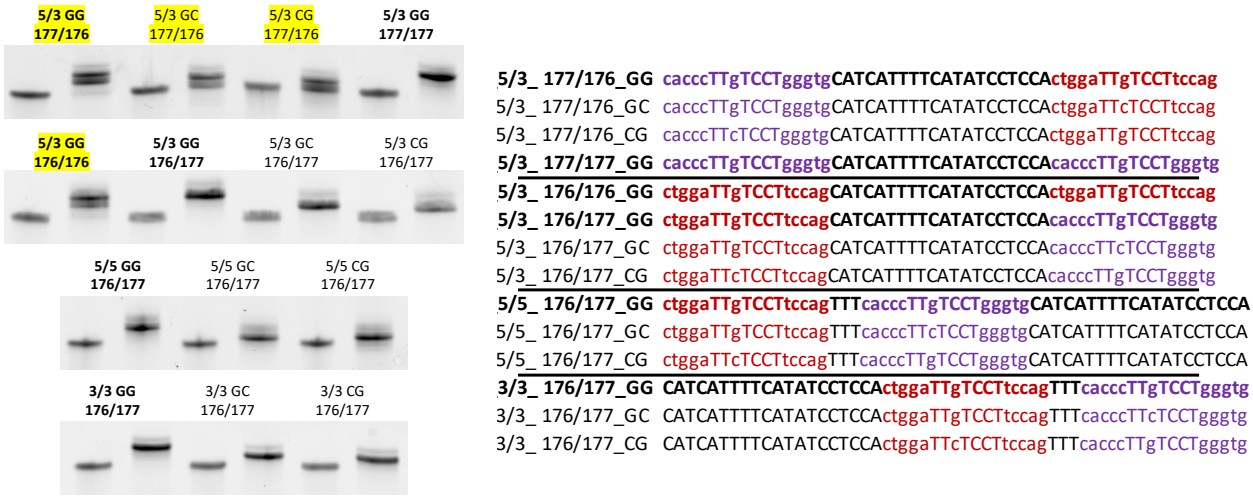
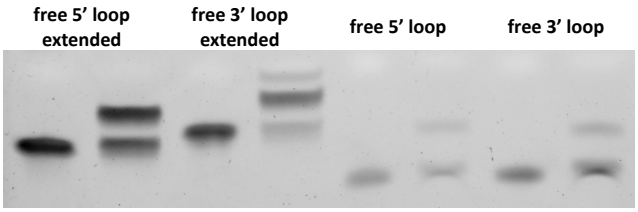



**Figure 2.20** Inserting multiple labels into a construct. For applications that might require increased sensitivity, inserting multiple probes into the same construct can be beneficial. TGT is able to insert multiple labels in tandem as well as at either end into identical hairpins or two unique TAG sequences. Multiple labels can be confirmed by the increased gel shift as compared to ΔC mutations. Notably, constructs with hairpin 176 at the 3′ end seemed to be labeled less efficiently.

The flexibility of TGT toward different nucleic acid substrates led me to ask the question of whether a hairpin was needed at all or if the appropriate sequence found its way into the active site of the enzyme it would still be able to be labeled. I designed two sets of linear constructs, either 17 or 12 nucleotides long with the free loop sequence either at the 5′ or 3′ end. While all constructs were labeled, the longer oligonucleotides were better accepted by the enzyme, labeling approximately 55% and 75% for the 5′ and 3′ free loops, respectively (Figure 2.21). This I likely due to an increased association with the enzyme induced by a longer oligonucleotide sequence. With further optimization this labeling strategy could be promising for applications that may be sensitive to the addition of a structured 17 nucleotide hairpin if a short 7 nucleotide extension can be added to the end of the sequence.




| | |
|---|---|
| free 5' loop extended | TTGTCCTtccagcaccc |
| free 3' loop extended | gggtgctggaTTGTCCT |
| free 5' loop | TTGTCCTtccag |
| free 3' loop | ctggaTTGTCCT |

**Figure 2.21** Free loop sequence labeling by TGT. TGT can modify a substrate with the appropriate loop sequence when it is at either end of the construct. Notably, the 3′ end of longer oligos is preferred.

There are nucleases that are able to recognize and cut at specific modified nucleobases, such as deoxyuridine and inosine. I hypothesized that by incorporating these modified bases into the hairpin constructs I could design modified oligos with a minimal number of extra nucleotides and therefor minimize any perturbation inserted into the construct of interest. Specifically, I thought to use two products from NEB, one containing an uracil DNA glycosylase (UDG) called USER and the other utilizing Endonuclease V which cuts after inosine. I designed hairpin constructs which incorporate dU or dI in such a way as to minimize the perturbation of the oligo of interest. Initial experiments testing this hypothesis are promising but in need of optimization. Both enzymes were able to degrade the modified oligo, but the initial labeling reactions still require some optimization, leading to inconclusive product lanes via urea page (Figure 2.22).

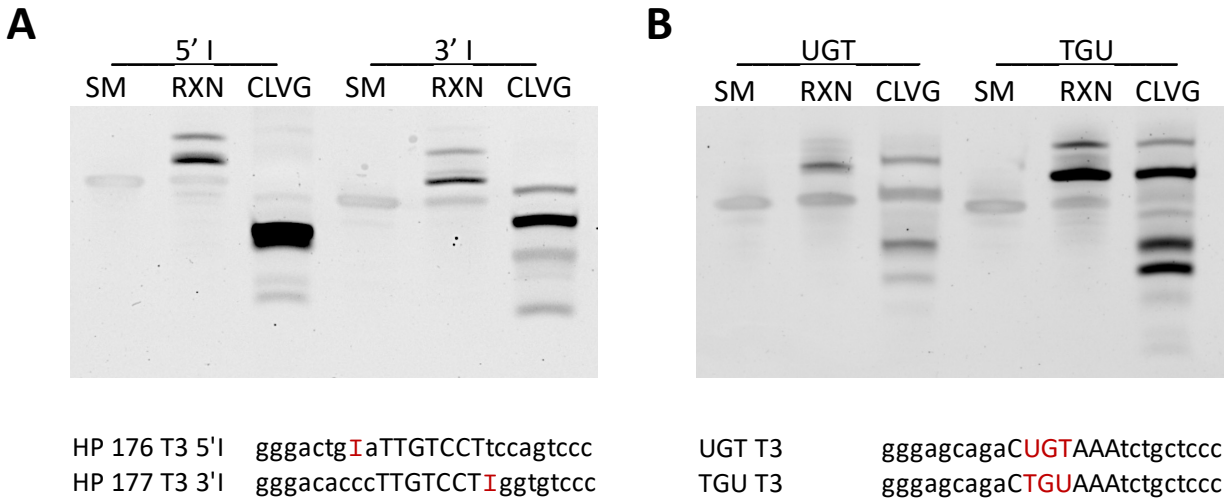
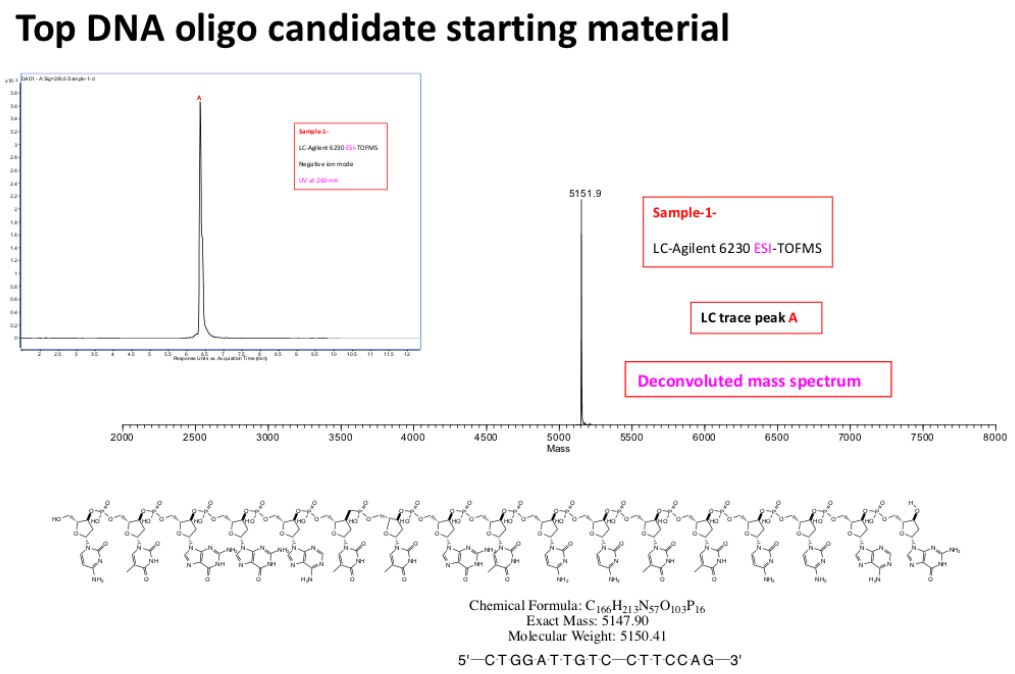



**Figure 2.22** TAMRA labeled hairpin construct cleavage by endonuclease enzymes. (A) Endonuclease V cleavage of labeled TAG3.0 hairpins containing inosine. (B) The USER enzyme mix cleaves constructs containing dU.

While gel shift assays are a well-established and reliable method of characterizing and qualifying TGT labeling efficiency in our lab, I corroborated this data more precisely with LCMS. The mass increase and peak area indicate that labeling is quantitative, although there are some small impurities or degradation products that could likely be avoided with a purer TGT prep (Figure 2.23).



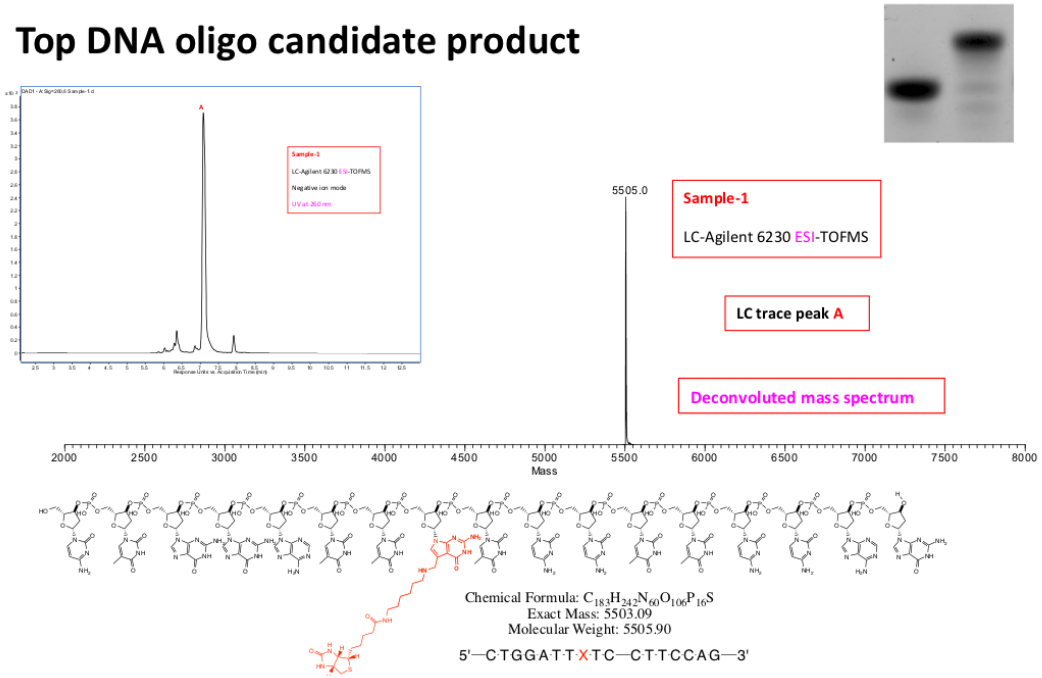




**Figure 2.23** LCMS verification of HIS M10–biotin adduct. The starting material and product LC traces have one predominant peak that matches the expected species, confirming the nearly quantitative labeling of the hairpin by TGT.

## 2.10 Conclusions

Beyond being central to life, nucleic acids have proven to be an invaluable tool for various applications ranging from critical therapeutics to fundamental scientific research. Harnessing the unparalleled programmability and versatility afforded by these biological macromolecules requires tools that install functional handles or facilitate detection or function. Increasing the versatility and diversity of these tools is critical to the further advancement of relevant fields. Here we introduce a technology that allows for the enzymatic insertion of a variety of functional small molecules into ssDNA substrates of interest for downstream applications. This system is compatible with both internal and end installation of DNA-TAG hairpins and leaves both the 5′ and 3′ ends free, making it compatible with other DNA modification techniques. Additionally, it is tolerant of a variety of small molecule substrates. Additionally, this technology allows an inexpensive and straightforward method by which researchers can quickly label several DNA oligos, either simultaneously or in parallel, in a single step and with a short spin column purification. The applications discussed in the next chapter show promise for the breadth of capabilities that this technology will unlock.

## 2.11 Materials and Methods

### *2.11.1 Buffers*

**In Vitro Transcription**: 40 mM Tris pH 7.5, 5 mM DTT, 25 mM MgCl2, 2 mM spermidine

**TGT reaction buffer 1X**: 100 mM HEPES, pH 7.3, 5 mM DTT, and 20 mM MgCl2

**His lysis and binding buffer (Zymo)**: 50 mM sodium phosphate buffer pH 7.7, 300 mM sodium chloride, 50 mM imidazole, 0.03 % Triton X-100.

**His elution buffer (Zymo)**: 50 mM sodium phosphate buffer pH 7.7, 300 mM sodium chloride, 250 mM imidazole.

**TGT storage buffer**: 25 mM HEPES pH 7.5, 100 mM NaCl.

*2.11.2 In vitro transcription of RNA hairpins*

RNA oligonucleotides are very expensive. In order to test all of the constructs that I wanted I decided to *in vitro* transcribe the hairpins in the lab. The transcriptions were carried out with the following conditions using T7 RNA polymerase previously prepared by another lab member: 1X T7 RNAP Buffer, 5mM MeUTP, 5mM ATP, 5mM CTP, 9mM GTP, 5mM Annealed template, 0.004 U/μL PPTase, 0.05% Triton X, and 0.15 ug/μL T7 RNAP and incubated at 4˚C for 4 hours. Subsequently, 1.2 μL 100 mM CaCl2 and 1 μL Turbo DNAse were added to a 100uL IVT reaction, incubated at 37˚C for 20 minutes. The RNA product was then purified by ethanol precipitation and the transcript analyzed by urea PAGE.

*2.11.3 Synthesis of PreQ1-small molecule probes*

PreQ1-biotin was the key small molecule used in this chapter for the development of DNA-TAG. Additional probes used in this chapter include PreQ1-tetramethylrhodamine, -Cy5, -AlexaFluor647, -AlexaFluor488, -SiR, and -benzyl guanine. All probes were synthesized as previously described with the caveat of the PreQ1-NHBoc precursor being ordered and characterized by a commercial source.[57] As such, LCMS characterization was sufficient for confirmation of the small molecule identity. Briefly, the Boc protected starting material was deprotected in 10% TFA/DCM for one hour. The TFA was neutralized 3 times with 10% TEA in DCM, rotoevaporating the solution to near dryness between each wash. The crude oil product was taken to the next step where it was dissolved in dry DMF followed by the dropwise addition of TEA which causes the free amine to enter into solution. Each NHS probe was dissolved in dry DMF and added to the

reaction dropwise with stirring. The reaction was allowed to proceed for 2 hours at room temperature, the DMF evaporated, and purified by HPLC. The final product was confirmed by LMCS.

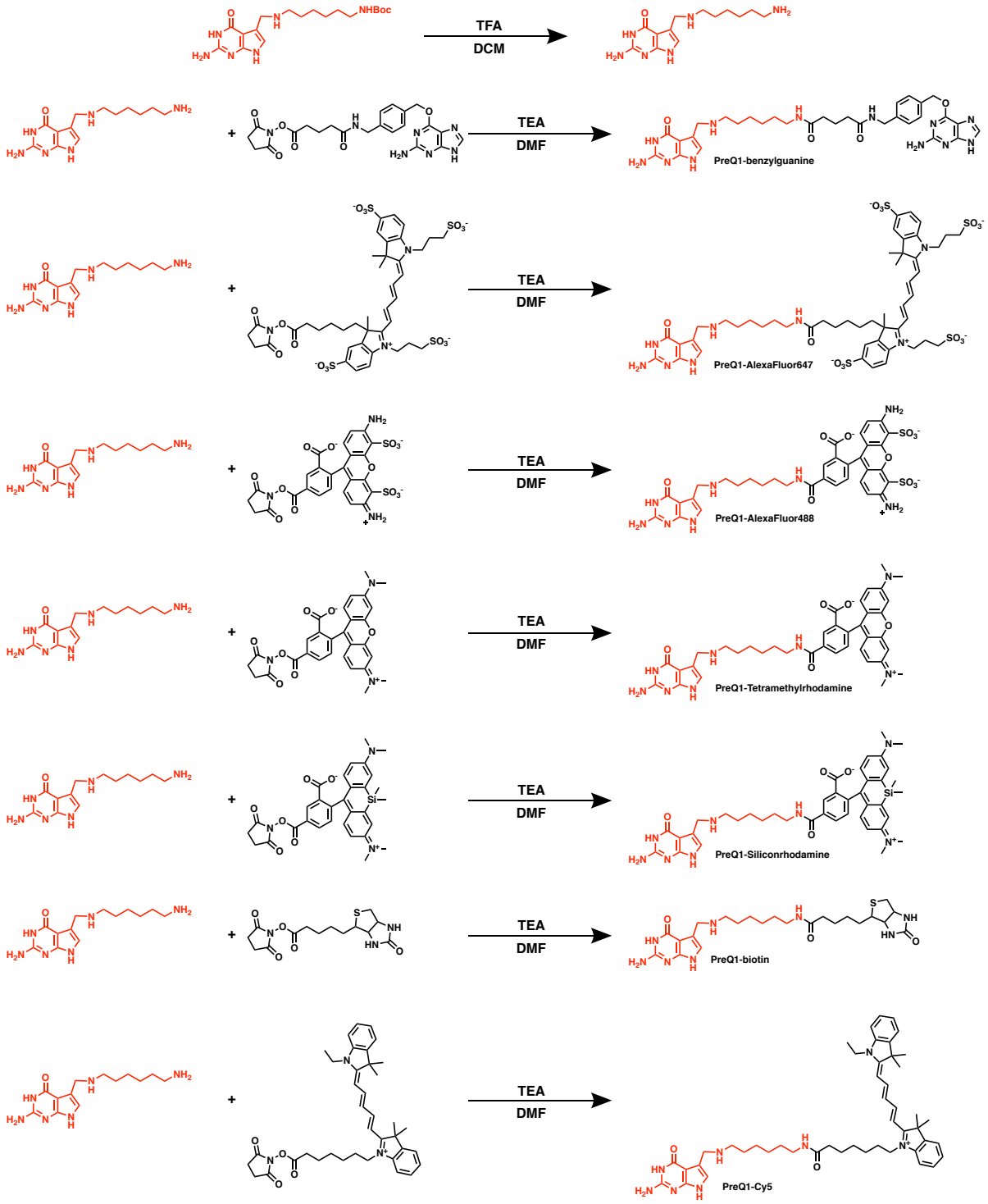



**Figure 2.24** synthesis and structures of PreQ1 probes used in this chapter.

*2.11.4 Expression of E. coli TGT*

*E. coli* TGT was expressed as previously described from plasmid #138201 in the Addgene database. It should be noted that the second construct with an added C terminal Strep tag was cloned in the same plasmid and sequenced by Eton Biosciences. Briefly, the plasmid carrying TGT was transformed into BL21(DE3) Competent *E. coli* cells from New England Biosciences using the protocol described on the Addgene website. A starter culture was grown overnight in the presence of kanamycin for selection. Approximately 5mL of starter culture was transferred into 200 mL of kanamycin containing LB and allowed to grow to an optical density of ~0.6. At this point expression was induced with by adding IPTG to a final concentration of 1mM and shaken at 37˚C for 4 hours. The bacteria were pelleted 6,000xg for 20 minutes at 4˚C. Cell pellets can be resuspended and purified immediately or stored at -20˚C if desired.

*2.11.5 Purification of TGT*

In general, expression is done on a 200 mL scale and His purification carried out using HisPur™ Ni-NTA Resin spin columns from Thermofisher. The cell pellet is resuspended in 2 mL total of lysis buffer and sonicated on ice using the following settings: 4 cycles (30 sec on / 2 min off); output control = 4; duty cycle% = 30-40. The lysate is centrifuged at 10,000g for 30 min at 4 C to remove cellular debris. In tandem, the resin from the spin column is equilibrated 2X with 2 resin bed volumes of lysis buffer. The supernatant of the centrifuged cell lysate is transferred to the capped spin column and incubated for at least one hour at 4˚C with end over end mixing. The resin is then washed, and the protein eluted according to the HisPur manual. The lysis and wash buffer are the same as it was noted that a higher imidazole concentration during binding reduced

nonspecific carry over into the final elutions. Buffers were ordered from Zymo and supplemented with 400uM PMSF: **His-Wash buffer**: 50 mM sodium phosphate buffer pH 7.7, 300 mM sodium chloride, 50 mM imidazole, 0.03 % Triton X-100 and **His-Elution buffer**: 50 mM sodium phosphate buffer pH 7.7, 300 mM sodium chloride, 250 mM imidazole. The final elutions were dialyzed overnight using a 30 kDa dialysis cassette into storage buffer: 25 mM HEPES, 2 mM DTT, 1 mM EDTA, 0.5 mM NaCl, 100 µM PMSF, pH 7.3. The same process was repeated to ensure that the final solution was RNAse free. For the second construct with a dual purification tag, dialysis was skipped, and the solution taken directly to the streptactin purification step after elution from the nickel NTA column and dialyzed in the same buffer to remove biotin. An approximate concentration was determined by absorbance on the nanodrop and the final protein aliquoted in 20uL fractions and stored at -80˚C.

*2.11.6 Cloning*

Mutant enzymes were all prepared by mutating the Addgene plasmid referenced above. the online NEBaseChanger tool was used to design quick change primers. PCR amplification and product ligation were carried out using the Q5® Site-Directed Mutagenesis Kit. After ligation the product was transformed into NEB® 5-alpha Competent *E. coli* cells and spread on a kanamycin agar plate. The plate was incubated over night at 37˚C. the next day, several colonies were selected and 12 mL cultures frown in kanamycin containing LB overnight. The new plasmids were extracted with the Biomiga mini prep kit and submitted to Eton Biosciences for sanger sequencing.

*2.11.7 TGT labeling reaction*

TGT modification of DNA and RNA substrates was carried out at 37˚C for 4 hours in a thermocycler with a heated lid. Briefly, the reaction conditions are: 10 μM enzyme, 5 μM nucleic acid substrate, and 50 μM PreQ1-biotin (or other small molecule probe), 1X TGT reaction buffer, and 1-unit RNAsin from Promega for RNA substrates. Components are mixed in no particular order, gently vortexed, briefly spun down, and incubated on a thermocycler. In the case that the reaction was set up at the end of the day, the thermocycler is set to go to 4˚C indefinitely after the 4-hour reaction time. For labeling reactions of constructs containing 2 hairpins the small molecule substrate concentration was doubled. For analysis purposes reactions were not purified before PAGE which did not affect the gel shift.

*2.11.8 Labeled oligo purification*

For the experiments where it was necessary to purify the labeled construct, the oligo clean and concentrator kit from Zymo was used and the product eluted in pure water.

*2.11.9 Urea PAGE*

Urea PAGE was the primary method of analysis to determine TGT activity toward specific substrates. For all gels, I used the 19:1 Acrylamide:Bisacrylamide SequaGel® UreaGel System and the Mini-PROTEAN Tetra Vertical Electrophoresis Cell from Bio-Rad. Briefly, 1.5mm plates assembled on the Bio-Rad gel casting stand, 10 mL of the desired percentage gel was freshly prepared in a 15 mL centrifuge tube, 14 μL of TEMED was added and gently mixed by inverting the tube, followed by 80 μL of 10% ammonium per sulfate. After briefly mixing the solution is poured between the plates, a comb inserted, and allowed to solidify (~20-30 min). All oligo samples were prepared for gel

electrophoresis using 2X RNA loading dye from New England Biosciences. For all constructs, 150-200 ng of the oligonucleotide was loaded onto the gel for visualization. The solutions were heated to 98˚C for 5 minutes on a thermocycler. The electrophoresis cassette was assembled, and the chamber filled with 1X TBE. After thorough washing of the wells to remove leached urea, the samples were loaded into the wells and the gel was run for 90 min at 200V. Afterward, the gel was stained with gel red in TBE (1:1000 dilution) for ~5 min and imaged on a Bio-Rad ChemiDoc gel imager system.

*2.11.10 Gel Shift Analysis*

Images were exported for analysis form the ChemiDoc software and analyzed qualitatively by eye. Successful labeling is indicated by an upward shift in the product lane due to an increase in molecular weight of the labeled species. In the cases where a quantitative measurement was desired, the bands were analyzed via densitometry using FIGI. Briefly, the percent signal of each band compared to the total in the gel was calculated.

*2.11.11 Restriction enzyme reactions*

For the endonuclease V and USER reactions, the TGT reactions were purified using the zymo clean and concentrator kit and the product treated to the enzyme. For the initial experiments 1 unit of enzyme was used for the 12 µL 5uM reaction. These conditions need to be optimized for future experiments.

## 2.12 Acknowledgments

Chapter two, in part is currently being prepared for submission for publication. Tota, Ember M.; Devaraj, Neal K. The dissertation author was the primary researcher and author of this material.

# 3 Applications of DNA-TAG

## 3.1 Introduction

Oligonucleotide modification and functionalization is becoming increasingly important for a variety of applications in field such as biology, biochemistry, materials science, bio and chemical engineering, and chemistry, among others. As such, finding inexpensive, reliable strategies to functionalize the nucleic acid substrate with a functional handle of interest is imperative to related advancements in these fields. Currently, there are limited methods for these modifications, including chemical insertion during or after solid phase synthesis and enzymatic strategies. Chemical synthesis is low throughput and costly, requiring at least one round of HPLC purification during the process. Current enzymatic strategies are limited to 3′ insertion of modified nucleosides and the 5′ insertion of modified phosphate groups and by their small molecule substrate scope. For instance, terminal deoxynucleotidyl transferase (TdT) catalyzes the transfer of nucleotides to the 3′ end of DNA constructs. This enzyme has been used to install reporters appended to deoxy UTP, however, the scope of tolerated small molecules is slim and it is often necessary to install a small molecule handle and preform a secondary bioorthogonal modification step to insert the actual reporter. DNA-TAG provides an unprecedented tool for the modification of DNA oligonucleotides, able to efficiently insert a variety of small molecule reporters into a small 17 nucleotide recognition sequence in a single step with a very simple spin column purification.

To demonstrate my technology, I turned to two common uses for DNA oligonucleotide reporters: RNA detection via northern blot (Figure 3.1), and RNA FISH (Figure 3.2). Both techniques utilize antisense probes modified with a reporter. Historically, northern blot detection has relied on radioactive residues inserted into the

antisense probe.[60] Radioactive probes are non-ideal due to safety concerns and the resulting restrictions and challenges to minimize them. However, it has been difficult to find an accessible detection method with matched sensitivity. Fluorescent probes are a promising alternative.[61] Although they are generally less sensitive than radioactive detection methods, they offer the potential for multiplexing with different dyes, have a longer shelf life, and are far safer to work with.

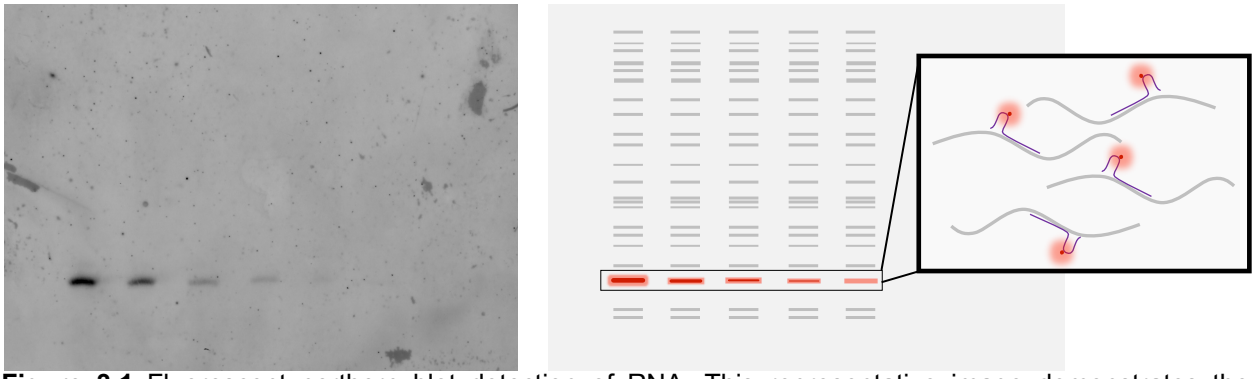


**Figure 3.1** Fluorescent northern blot detection of RNA. This representative image demonstrates the concept of fluorescent northern detection of a transcript of choice DNA-TAG generated probes.

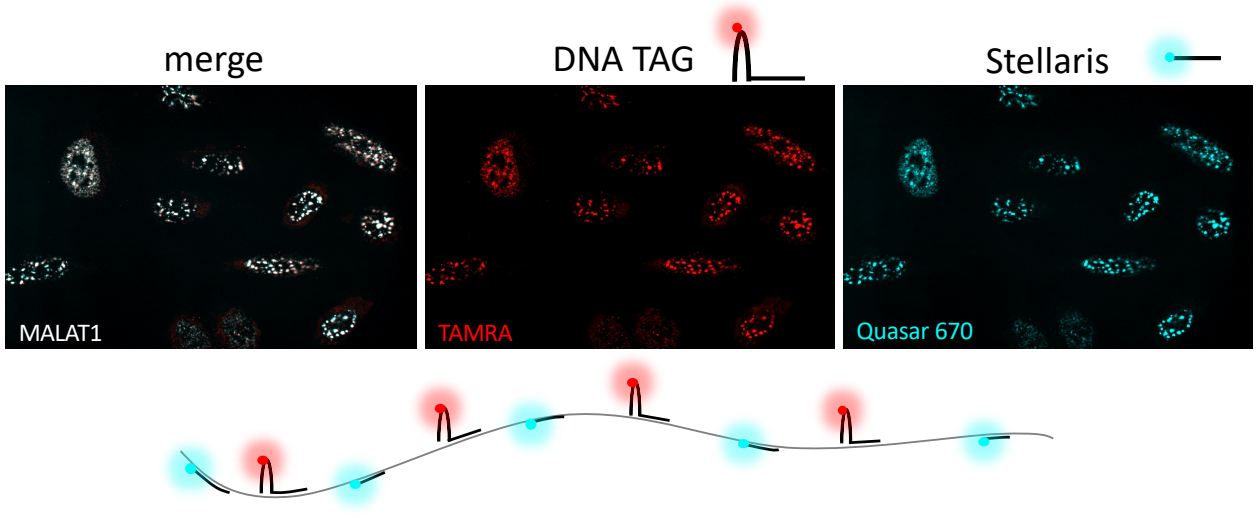



**Figure 3.2** Single molecule FISH. This representative image demonstrates the concept of FISH and the ability of DNA-TAG generated probes to report on a transcript of interest.

In situ hybridization (ISH) detection of RNA and DNA in cells and tissue provides insight into the localization, function, and prevalence of specific NA species. ISH and Fluorescent ISH (FISH) techniques are a critical tool for these types of experiments and therefore, developing inexpensive, sensitive, reliable hybridization tools are important to the field of RNA biology and biochemistry.[62–73] Researchers have developed a number of strategies to increase sensitivity and specificity including signal amplification by exchange reaction (SABER) FISH, multiplexed error-robust fluorescence in situ hybridization (MERFISH), and RNAscope, among others. Single molecule FISH is a useful tool to detect a single RNA species via localized fluorescence of a set of labeled antisense NA probes.[74–77] This method is highly specific and reliable; however, its use is limited because it can be very expensive. The single molecule FISH technique employed here is generally performed with probes sets of 30-48 unique fluorescently modified oligonucleotides. As mentioned earlier, synthesizing modified oligonucleotides can be low yielding and costly. Custom ready to go probe sets can be ordered from Stellaris but they are expensive, come in small amounts, and are limited in fluorophore identity and diversity. Enzymatic generation of labeled probe sets promises a simple, inexpensive alternative which would allow labeling with a variety of fluorophores, immediate generation of probe sets, and flexibility in developing probe sets for difficult to detect transcripts. Challenges for enzymatic labeling strategies include uniformity in number of labels and labeling efficiency. Both aspects are key to producing probe sets that can reliably and sensitively report on transcript abundance and location. Currently, an enzymatic labeling strategy employing terminal deoxynucleotidyl transferase (TdT) is used along with dideoxy nucleoside probes to install a single functional modification at

the 3′ end of probe sets.[78] This method only allows the direct installation of a select few fluorophores with most requiring a two-step modification process, first installing a reactive handle and secondarily clicking on a fluorophore of interest.

The efficiency of DNA-TAG labeling and the flexibility of TGT to accept a variety of small molecule substrates make it a great candidate for the production of fluorescently labeled DNA oligonucleotide probes. As a proof of concept for the utility of the technology I employed DNA-TAG to generate a fluorescent northern blot probe as well as FISH probe sets. This labeling strategy was high yielding and produced reliable results.

## 3.2 Northern blot

Fluorescent northern blot strategies are a fairly recent development.[61] DNA-TAG offers a simple, efficient, means to generate fluorescent probes to reliably detect RNAs of interest. Two key advantages of this labeling strategy would be the potential to introduce multiple labels into a single oligonucleotide to increase sensitivity and decrease detection limits as well as the inexpensive generation of a probe set using several oligos to target a single transcript, also resulting in increased sensitivity and specificity. Preliminary experiments discussed here use a single labeled probe to detect U6 RNA both in cellular extracts and in pure samples *in vitro* transcribed in the lab.

U6 RNA is a small, approximately 100 nucleotide, snRNA that is part of the spliceosome. To synthesize the U6 RNA for my experiments, I extracted DNA from U2OS cells and amplified the U6 sequence with a primer set that also includes the T7 promoter sequence. This amplified product was purified and used as a template for *in vitro* transcription. The IVT RNA was then purified by ethanol precipitation and analyzed via urea PAGE. Simultaneously, total RNA was extracted from U2OS cells. The U6 PCR

amplified template, IVT U6 RNA, and total extracted U2OS RNA were analyzed via gel electrophoresis (Figure 3.3.).

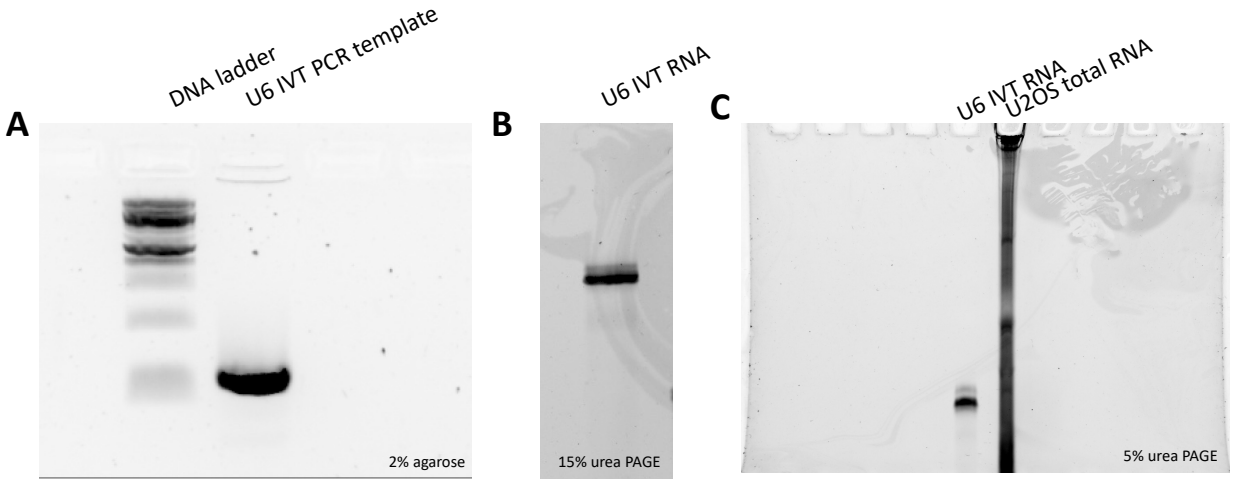
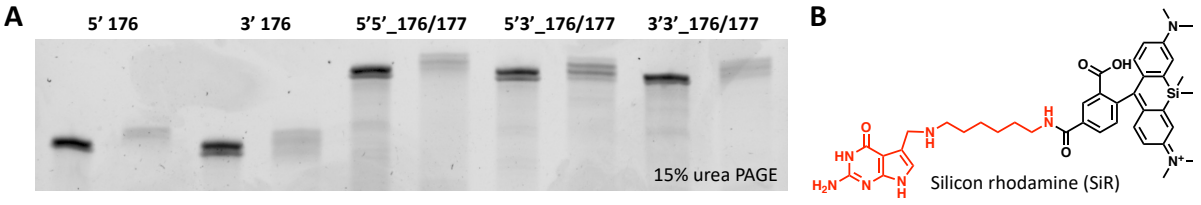



**Figure 3.3** U6 RNA preparation and U2OS RNA extract. (A) Amplification of U6 T7 IVT template from U2OS genomic DNA. (B) Ethanol precipitated IVT U6 RNA. (C) IVT U6 RNA and U2OS extracted total RNA.

Several antisense probe constructs were designed to test the limits of DNA-TAG for use in RNA northern blot detection. One or two TAG hairpins were placed at either end of the antisense sequence. The probes were modified with PreQ1-siliconrhodamine with the DNA-TAG technology (Figure 3.4.). All probes were efficiently labeled, as evidenced by the disappearance of the starting material band on the gel. Each construct was tested for its ability to detect U6 RNA via northern blot. Indeed, all probes were able to hybridize to the target (Figure 3.5.). The workflow was optimized to increase signal in future experiments.




| | |
|---|---|
| 5' U6 probe 176 | ctggaTTGTCCTtccagTTTGCAGGGGCCATGCTAATCTTCTCTGTATCG |
| 3' U6 probe 176 | GCAGGGGCCATGCTAATCTTCTCTGTATCGTTTctggaTTGTCCTtccag |
| 5' 5' U6 probe 176 177 | ctggaTTGTCCTtccagTTTcacccTTGTCCTgggtgTTTGCAGGGGCCATGCTAATCTTCTCTGTATCG |
| 5' 3' U6 probe 176 177 | ctggaTTGTCCTtccagTTTGCAGGGGCCATGCTAATCTTCTCTGTATCGTTTcacccTTGTCCTgggtg |
| 3' 3' U6 probe 176 177 | GCAGGGGCCATGCTAATCTTCTCTGTATCGTTTctggaTTGTCCTtccagTTTcacccTTGTCCTgggtg |

**Figure 3.4** U6 antisense probe constructs and labeling. (A) All probe constructs are efficiently labeled by TGT. SiR quenches gel red which is why the product band seems dim. (B) Structure of PreQ1-siliconrhodamine.

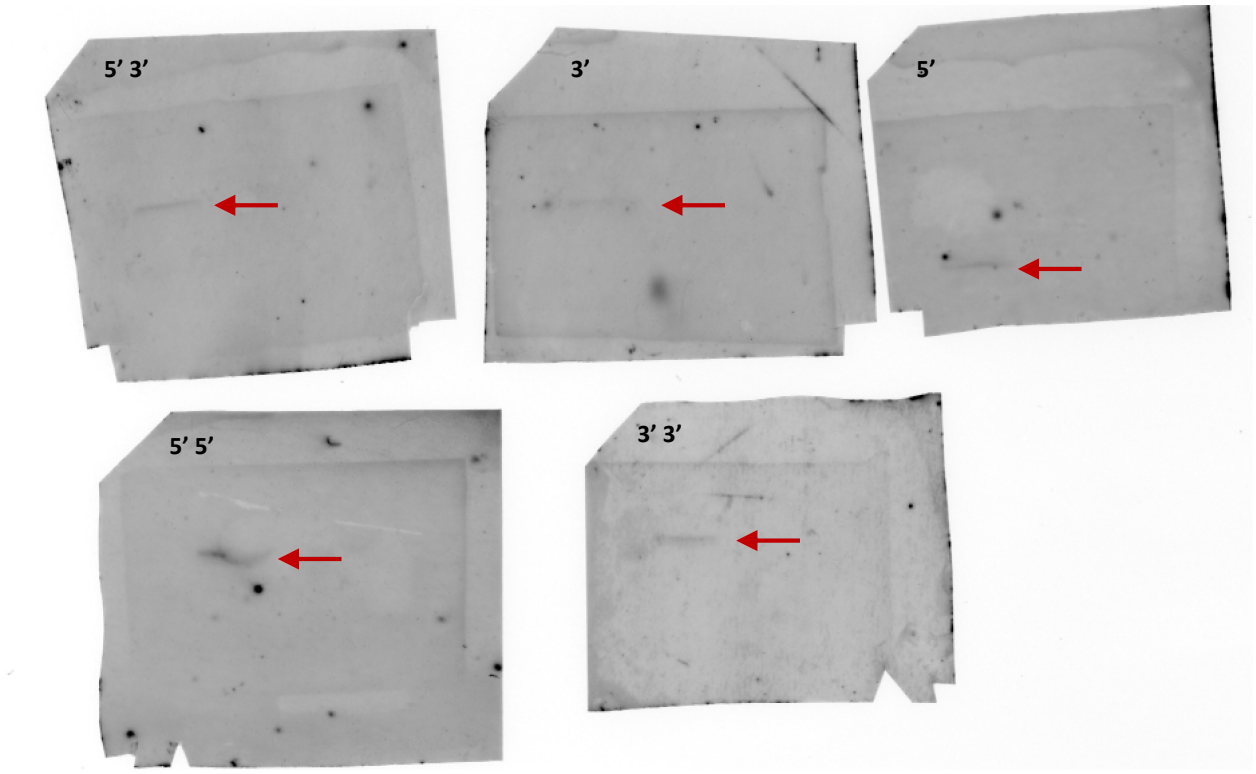



**Figure 3.5** Northern blot detection of U6 using DNA-TAG SiR labeled probes. All probes were able to detect the RNA via fluorescent northern blot. There does not seem to be an appreciable difference when using singly or doubly labeled probes.

The 47 nucleotide 5′ construct was used going forward. Both the cellular extract and IVT U6 RNA samples were separated via urea PAGE in dilution series. The gels were transferred, incubated with the antisense probe, and signal detected via fluorescence scanning and analyzed via linear fit. The DNA-TAG generated U6 northern blot probe

was successful in detecting the U6 RNA in a dose dependent manner with an exceptional $R^2$ value (Figure 3.6.). While further experiments should be done to determine the lower detection limit and the utility of multiple tags, this data shows the utility of DNA-TAG for the generation of fluorescent northern blot probes.

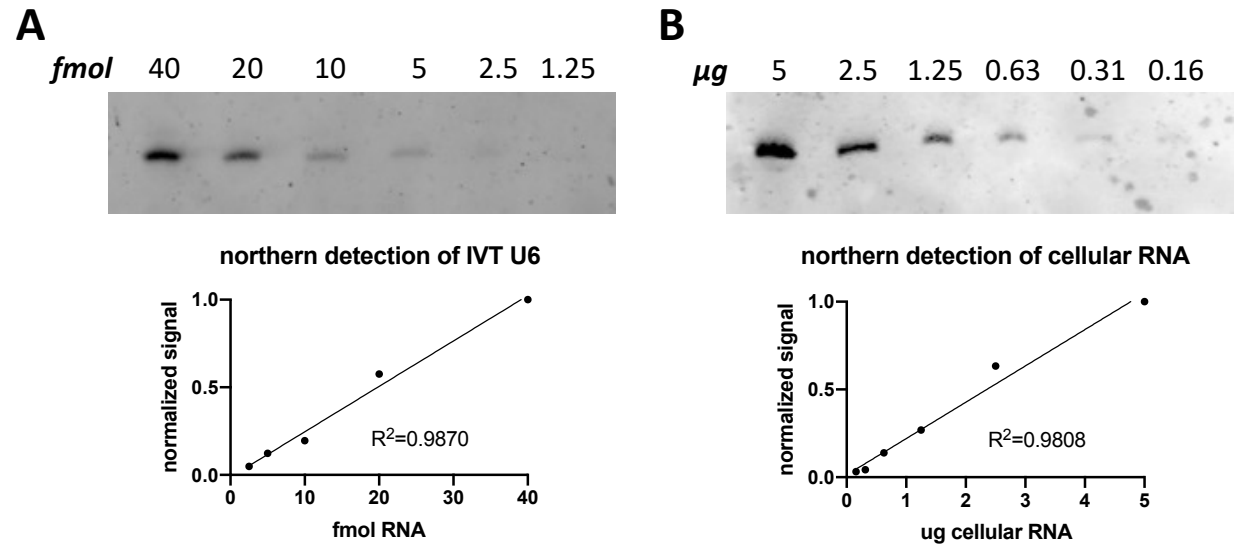



**Figure 3.6** Northern blot detection of U6 RNA. RNA was run on a 15% urea page gel, transferred to a positively charged membrane, fixed, incubated with the 5′ SiR labeled U6 antisense probe, and imaged on a fluorescent gel scanner. Dose dependent fluorescent northern blot detection of (A) IVT U6 RNA and (B) U6 RNA in total RNA extract from U2OS cells.

## 3.3 Fluorescence in situ Hybridization (FISH)

Fluorescence in situ hybridization is used to visualize both DNA and RNA in cells and tissues. Reporting on location and abundance of a specific nucleic sequence can provide valuable insight into the role and prevalence of that sequence for a tissue or cell type of interest. Researchers have developed a number of FISH tools varying in their labeling strategy, antisense probe type (DNA, RNA, PNA), multiplex capability, and approach to amplifying low signal. Enzymatic generation of DNA oligonucleotide probe sets for single molecule FISH is promising as an accessible, flexible, inexpensive tool. The ability to fluorescently modify the probe sets in the lab allows the installation of a number of fluorophores, facilitating multiplex experiments as well as flexibility for use with

complimentary imaging techniques (ICC, IHC, etc.). Single molecule FISH probe sets are generally comprised of 30-48 singly labeled antisense probes. For all probe sets used in FISH experiments in this work I ordered individual oligos from IDT, pooled them, labeled them with DNA-TAG, did a simple spin filter purification, performed FISH experiments using U2OS cells, and imaged them using a fluorescence microscope.

Initial experiments targeted the abundant mRNA βactin. We generated a probe set using Stellaris's probe design tool, chose 25 sequences with the highest melting temperatures, designed constructs with a 3′ TAG hairpin, and ordered the individual oligos from IDT (Table 3.2.). We generated two probe sets, using TdT and TGT, labeled with Cy 5. RNA FISH experiments were conducted using both probe sets at equal concentrations and analyzed in parallel. Both probe sets produced the expected punctate fluorescent signals. Qualitatively, the TGT generated probe set produced a brighter signal (Figure 3.7.).

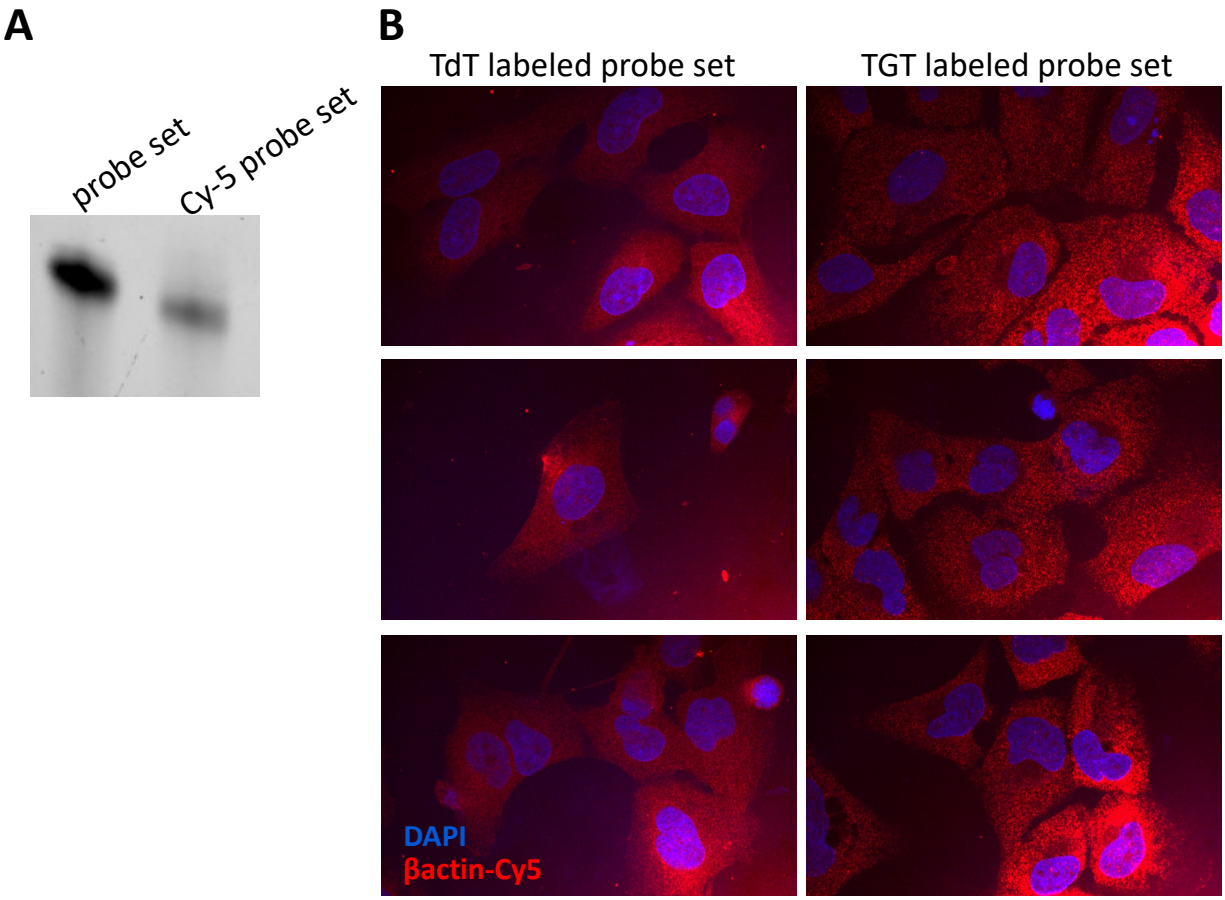



**Figure 3.7** βactin probe set preparation and FISH. (A) DNA-TAG labeling of probe set with PreQ1-Cy5. Equal amounts of starting material and product were loaded. 70% labeling by densitometry difference in starting material and product lanes.

Encouraged by the success of these initial experiments and the apparent advantage TGT labeling seemed to have over TdT, I decided to look at a more interesting long noncoding RNA, MALAT1, and compare our probe set generation to that of the commercially available set from Stellaris. For these experiments, I used all 48 probes generated by the Stellaris probe generator, an equal number to that in the ready to ship probe set sold by Stellaris.

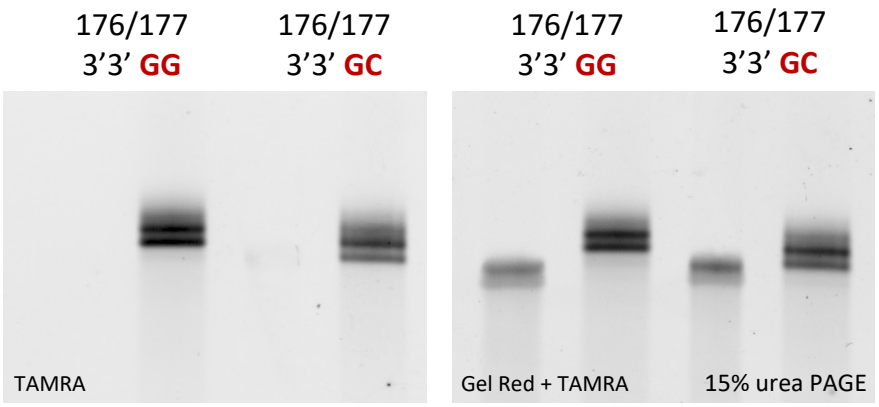



**Figure 3.8** MALAT1 probe set labeling with TAMRA using DNA-TAG. Probe sets are fully labeled as indicated by a lack of starting material in both product lanes with gel red stain.

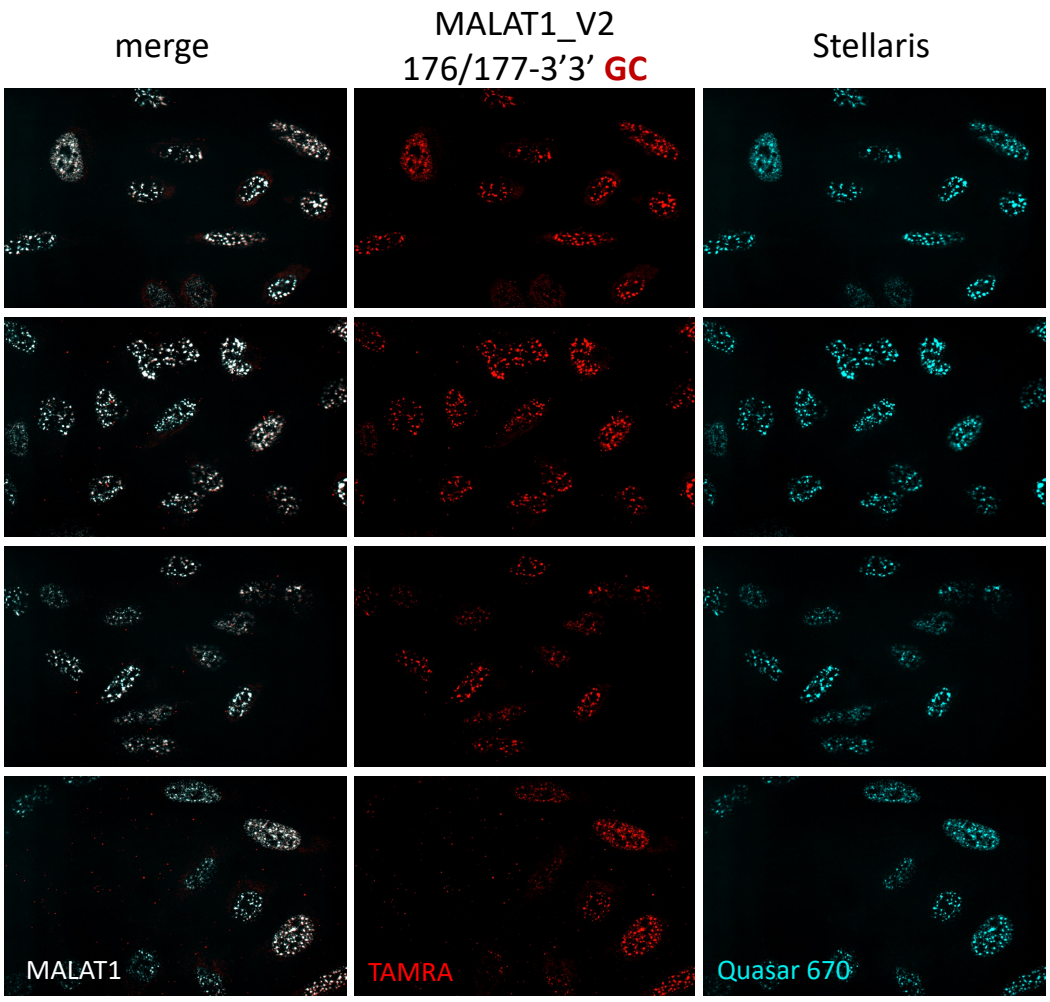



**Figure 3.9** FISH of MALAT1 with singly labeled DNA-TAG generated probe set and Stellaris probe set. The DNA-TAG probe set reliable reports on MALAT1 as seen by colocalization with the commercially available probe set from Stellaris.

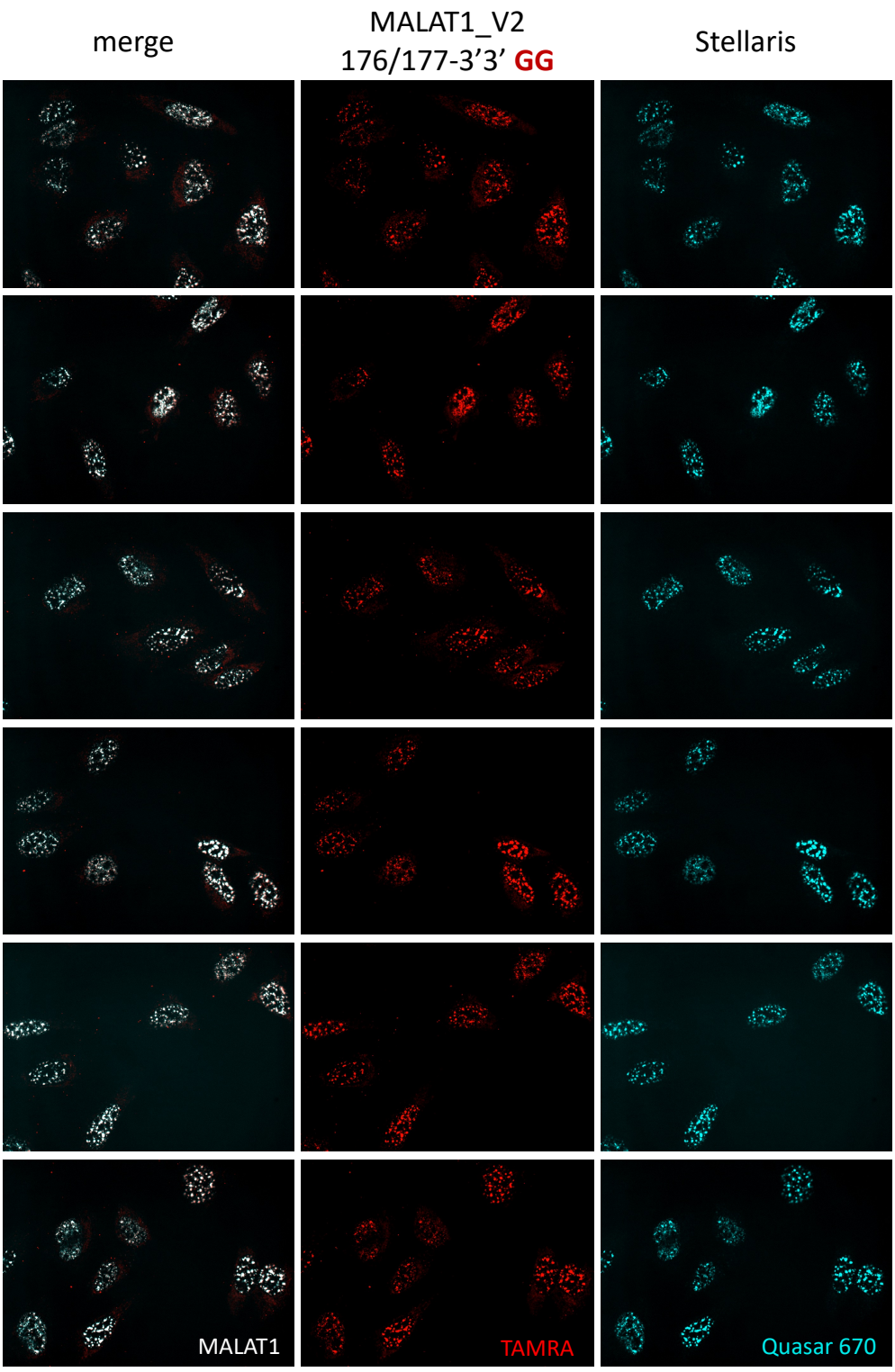



**Figure 3.10** FISH of MALAT1 with doubly labeled DNA-TAG generated probe set and Stellaris probe set. The DNA-TAG probe set reliable reports on MALAT1 as seen by colocalization with the commercially available probe set from Stellaris.

The Stellaris probe sets are labeled with quasar 570 or quasar 670, versions of Cy3 and Cy5, respectively, and the TGT generated probe sets are labeled with TAMRA. Side by side FISH experiments comparing the Stellaris and TGT generated probe sets showed that indeed, the TGT probe set labeled with TAMRA colocalizes with the Stellaris Quasar 670 probe set (Figure 3.9., Figure 3.10.).

Additionally, two sets of probes were designed to determine whether having multiple labels would increase fluorescent signal (Table 3.4., Table 3.5.). Both probe sets were fully labeled as is indicated by the TAMRA signal before gel red staining. Additionally, the signal in the doubly labeled probe set is greater than that in the singly labeled (Figure 3.8.). I compared a probe set with two 3′ hairpins in tandem to one with two hairpins and a single G34→C34 mutation. Qualitatively, the doubly labeled probe set looks to be brighter than the singly labeled counterpart, although not to the degree that would be expected. Both produce a higher signal than the Quasar 570 labeled Stellaris probe set, however, this direct comparison cannot be made due to the presence of two different fluorophores (Figure 3.11.). Each probe sets reliably annealed to MALAT1, as seen in the colocalization experiments, and generated sensitive signal when compared to their Quasar 570 labeled commercial counterpart.

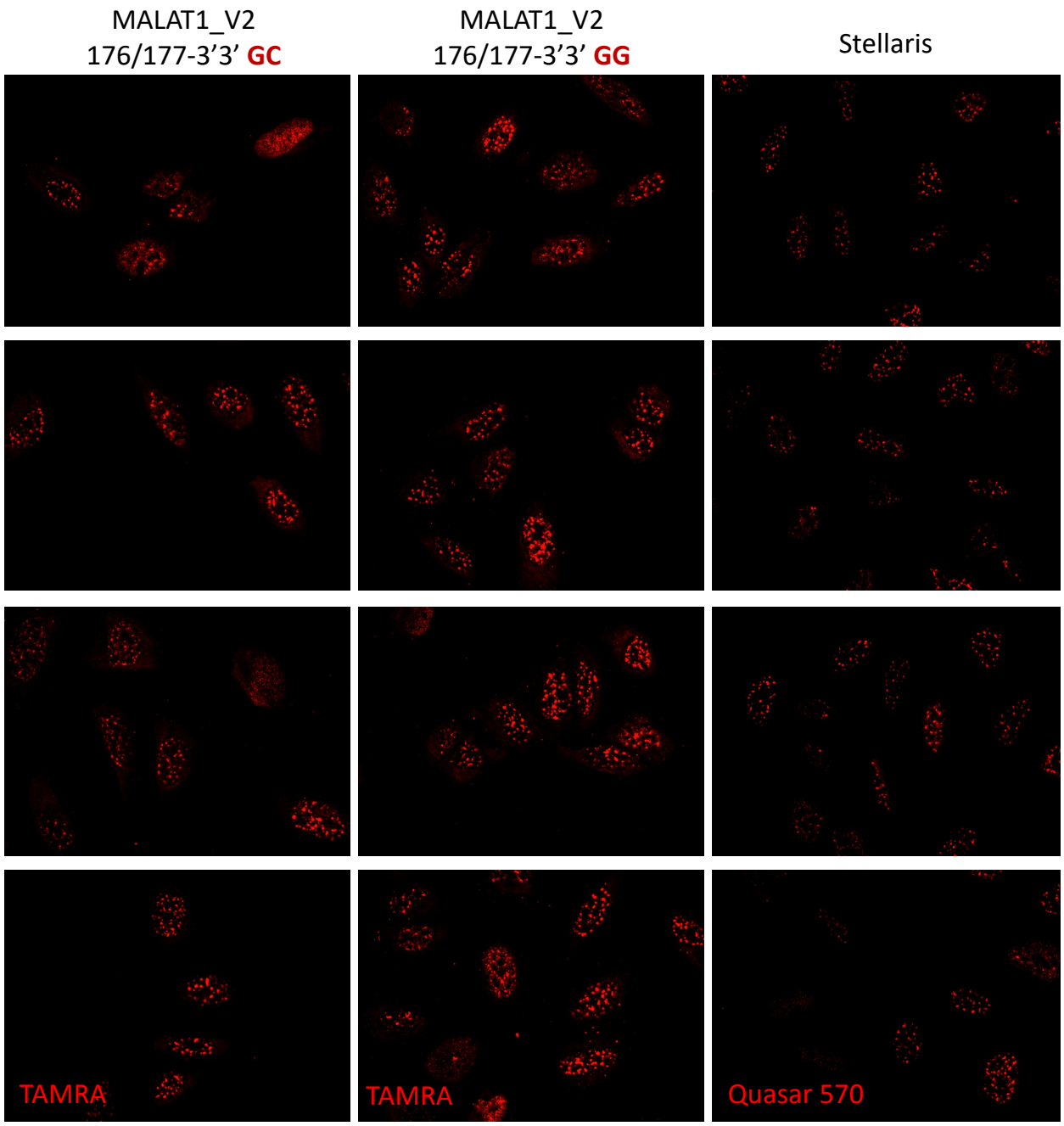



**Figure 3.11** MALAT1 detection with various probe sets. All probe sets generated good signal. The doubly DNA-TAG labeled probe set has the brightest signal. Direct comparisons between the DNA-TAG and Stellaris probe sets cannot be made because they are modified with different fluorophores.

The cost prohibitive nature of state-of-the-art RNA FISH probe sets limits their use. The difference in cost per nanomole to generate the TGT probe sets in comparison to the Stellaris ready ship probe set is on the order of 300X less. Additionally, self-labeled probe

sets offer flexibility in fluorophore choice, multiplexing, and quantity, among other things. TGT labeling offers an advantage over TdT labeling because it requires a single enzymatic installation of a variety of probes (all tested probes were efficiently inserted), while TdT requires a two-step labeling process for most probes. The only downside of TGT labeling is the requirement of the additional 17 nucleotide hairpin sequence, however, based on the MALAT1 colocalization data it does not seem to pose a problem for specificity.

## 3.4 Conclusions

The cost prohibitive nature of functionally modified DNA oligonucleotides is a limiting factor for their use in fundamental research. Developing easy to execute technologies for generating modified oligos from inexpensive commercial sources is imperative for accelerating advancements in the application space for DNA tools. FISH and northern blot analysis are critical tools for understanding the role of an RNA of interest, the work outlined here can provide researchers with easier access to these critical tools. In addition to the fluorophore labeled probes discussed in this work, the ability to install other small molecules such as biotin and digoxin could allow the expansion of these tools to include outputs that are amplifiable, increasing the sensitivity. While this chapter outlines two proof of concept applications that could be powerful, accessible tools for basic research, the potential applications of DNA-TAG are vast. Future work should explore applications like DNA immobilization, DNA barcoding, DNA assembly, DNA conjugation to other biological macromolecules, among others. With the advancement of nucleotide-based tools, researchers are seeking cost effective ways to generate functionally modified nucleic acids. Together, DNA-TAG and RNA-TAG developed in our

lab have the potential to label any ssDNA or RNA of interest. Further work will continue development of these tools and expand our knowledge of what they are capable of.

## 3.5 Materials and methods

### *3.5.1 Buffers*

**In Vitro Transcription 1X**: 40 mM Tris pH 7.5, 5 mM DTT, 25 mM MgCl2, 2 mM spermidine

**TGT reaction buffer 1X**: 100 mM HEPES, pH 7.3, 5 mM DTT, and 20 mM MgCl2

**FISH buffer A**: 2:7:1 Stellaris Buffer A : water : deionized formamide (Biosearch Technologies Cat# SMF-WA1-60)

**FISH buffer B**: Stellaris Buffer B (Biosearch Technologies Cat# SMF-WB1-20)

**FISH hybridization buffer**: 9:1 Stellaris Hybridization Buffer : Deionized Formamide (Biosearch Technologies Cat# SMF-HB1-10)

**Northern Hybridization Buffer**: NorthernMaxTM Prehybridization/Hybridization Buffer (AM8677)

### *3.5.2 Cell culture of U2OS cells*

For all imaging experiments and RNA extraction, I use U2OS cells, a human osteosarcoma derived line. The cells were ordered from ATCC, thawed, expanded, and stocks stored at -80˚C. Cells were cultured according to ATCC guidelines in T25 flasks. Complete growth medium containing 10%FBS and 1% pen/strep was used to maintain and culture the cells. The cells were passaged ~3 times per week at a 1:6 sub cultivation ratio when they reached ~80% confluency. To passage the cells, they were washed 1X with HBSS from Gibco and treated with TRYPLE from Gibco at 37˚C until fully lifted. The trypsin was quenched by adding an equal volume of complete medium and the necessary

volume of cells taken into a new flask. For experiments, cells were seeded in a black glass bottom 24 well plate at a 1:6 ratio calculated to scale with the area of the plate.

### *3.5.3 U2OS sequences*

**Table 3.1** Oligonucleotides and U2OS RNA sequences

| | **sequence** |
|---|---|
| **U6 RNA** | GUGCUCGCUUCGGCAGCACAUAUACUAAAAUUGGAAC GAUACAGAGAAGAUUAGCAUGGCCCCUGCGCAAGGAU GACACGCAAAUUCGUGAAGCGUUCCAUAUUUU |
| **5′ U6 probe 176** | CTGGATTGTCCTTCCAGTTTGCAGGGGCCATGCTAATC TTCTCTGTATCG |
| **3′ U6 probe 176** | GCAGGGGCCATGCTAATCTTCTCTGTATCGTTTCTGGAT TGTCCTTCCAG |
| **5′ 5′ U6 probe 176/177** | CTGGATTGTCCTTCCAGTTTCACCCTTGTCCTGGGTGTT TGCAGGGGCCATGCTAATCTTCTCTGTATCG |
| **5′ 3′ U6 probe 176/177** | CTGGATTGTCCTTCCAGTTTGCAGGGGCCATGCTAATC TTCTCTGTATCGTTTCACCCTTGTCCTGGGTG |
| **3′ 3′ U6 probe 176/177** | GCAGGGGCCATGCTAATCTTCTCTGTATCGTTTCTGGAT TGTCCTTCCAGTTTCACCCTTGTCCTGGGTG |
| **T7 U6 amp fwd** | ATACGACTCACTATAGGTGCTCGCTTCGGCAG |
| **T7 U6 amp rev** | AAAATATGGAACGCTTCACGAATTTGCGTGTC |

### *3.5.4 DNA extraction*

Genomic DNA was extracted from U2OS cells using the Quick-DNA Miniprep Plus Kit and used for the amplification of the T7 U6 RNA template.

### *3.5.5 PCR*

The U6 RNA IVT template was amplified from ~10ng of isolated genomic DNA from U2OS cells using Q5® Hot Start High-Fidelity 2X Master Mix. The PCR product was verified by gel electrophoresis, thoroughly purified using the DNA clean and concentrator from Zymo and taken forward to be used for the IVT of U6 RNA.

### *3.5.6 In vitro Transcription (IVT)*

The 100 nucleotide U6 RNA transcript was *in vitro* transcribed with the following conditions using T7 RNA polymerase previously prepared by another lab member: 1X T7 RNAP Buffer, 5mM MeUTP, 5mM ATP, 5mM CTP, 9mM GTP, 5mM Annealed template, 0.004 U/μL PPTase, 0.05% Triton X, and 0.15 ug/μL T7 RNAP and incubated at 4˚C for 4 hours. Subsequently, 1.2 μL 100 mM CaCl2 and 1 μL Turbo DNAse were added to a 100uL IVT reaction, incubate at 37 C for 20 minutes. The RNA product was then purified by ethanol precipitation and the transcript analyzed by urea PAGE.

### *3.5.7 TGT modification*

TGT modification of DNA substrates was carried out at 37˚C for 4 hours in a thermocycler with a heated lid. Briefly, the reaction conditions are: 10 μM enzyme, 5 μM nucleic acid substrate, and 50 μM PreQ1-biotin (or other small molecule probe), in 1X TGT reaction buffer. Components are mixed in no particular order, gently vortexed, briefly spun down, and incubated on a thermocycler. In the case that the reaction was set up at the end of the day, the thermocycler is set to go to 4˚C indefinitely after the 4-hour reaction time. For labeling reactions of constructs containing 2 hairpins the small molecule substrate concentration was doubled. For analysis purposes reactions were not purified before PAGE which did not affect the gel shift.

### *3.5.8 Urea PAGE*

Urea PAGE was the primary method of analysis to determine TGT activity toward specific substrates. For all gels, I used the 19:1 Acrylamide: Bisacrylamide SequaGel® UreaGel System and the Mini-PROTEAN Tetra Vertical Electrophoresis Cell from Bio-Rad. Briefly, 1.5mm plates assembled on the Bio-Rad gel casting stand, 10 mL of the

desired percentage gel was freshly prepared in a 15 mL centrifuge tube, 14 µL of TEMED was added and gently mixed by inverting the tube, followed by 80 µL of 10% ammonium per sulfate. After briefly mixing the solution is poured between the plates, a comb inserted, and allowed to solidify (~20-30 min). All oligo samples were prepared for gel electrophoresis using 2X RNA loading dye from New England Biosciences. For all constructs, 150-200 ng of the oligonucleotide was loaded onto the gel for visualization. The solutions were heated to 98˚C for 5 minutes on a thermocycler. The electrophoresis cassette was assembled, and the chamber filled with 1X TBE. After thorough washing of the wells to remove leached urea, the samples were loaded into the wells and the gel was run for 90 min at 200V. Afterward, the gel was stained with gel red in TBE (1:1000 dilution) for ~5 min and imaged on a Bio-Rad ChemiDoc gel imager system.

*3.5.9 Northern blot*

Dilution series of *in vitro* transcribed U6 RNA or U2OS total RNA were run on a 15% urea PAGE gel as described above. The RNA was transferred to a positively charged nylon membrane from Invitrogen using the Mini Trans-Blot cell system from Bio-Rad. The transfer was done at room temperature and run at 200 mA for 30 min in 0.5X TBE. The membrane was washed with 10X SCC for 10 minutes with rocking and crosslinked twice on a stratalinker using the auto setting.

*3.5.10 TdT labeling*

The fluorescent probe was prepared by reacting 11-amino-ddUTP with NHS-Cy-5 in a 1:2 equivalent ratio in sodium bicarbonate buffer pH 8.3 for 3 hours at room temperature. The reaction was quenched with Tris-HCL buffer pH 7.4, and the probe used directly for labeling without further purification. The TdT labeling reaction set up according

to the table below and incubated at 37˚C overnight. The product was ethanol precipitated and resuspended in 600 µL water to yield a theoretical concentration of 5 µM.

**Table 3.2** TdT reaction conditions.

| Component | Volume |
|---|---|
| 10X TdT Buffer | 5 µL |
| CoCl2 | 5 µL |
| DNA oligos mix (250 µM total, combined from equal parts from all probes) | 12 µL |
| Dye-ddUTP (5 mM) | 3 µL |
| Nuclease-free Water | 24 µL |
| TdT | 1 µL |

*3.5.11 FISH experiments and analysis*

U2OS cells were seeded in glass bottom black 24 well plates at a ratio of 1:6 and allowed to grow to ~80% confluency. Cells were washed with HBSS and fixed in 4% para formaldehyde (PFA) for 10 minutes at room temperature. PFA was aspirated, the cells were washed twice with HBSS, and permeabilized with 70% ethanol overnight at 4˚C. The ethanol solution was removed, and the cells washed 2X with HBSS. FISH wash buffer A was added and let sit for 5 minutes. Fluorescent probe sets were added to hybridization buffer with a final concentration on 125 nM. Buffer A was aspirated and the probe containing solution added to the well. Parafilm was placed over the wells and around the lid of the plate to minimize evaporation and the plate incubated at 37˚C overnight in the dark. The probe solution was aspirated, and buffer A was added for 30 min at 37˚C in the dark. A solution of 5ng/mL DAPI was prepared in buffer A. Buffer A was aspirated from the cells, the DAPI solution added, and the plate was placed at 37˚C for 30 minutes. The DAPI solution was aspirated, and the cells washed in FISH buffer B for 5 minutes. Buffer B was aspirated and vectashield added to the wells. The cells were imaged on the

confocal microscope in our lab and prepared using FIGI. All images were window leveled and analyzed visually. For wells with multiple probes, equal amounts of each probe were added to a final concentration of 125 nM.

### 3.6 Acknowledgments

Chapter three, in part is currently being prepared for submission for publication. Tota, Ember M.; Harjono, Vince; Devaraj, Neal K. I would like to thank Vince Harjono, who assisted in experiments involving oligonucleotide probe sets labeled with terminal deoxynucleotidyl transferase (TdT). The dissertation author was the primary researcher and author of this material.

# 4 Nucleic Acid – Protein Conjugation using NA-TAG

## 4.1 Introduction

Nucleic acid-protein assemblies are becoming increasingly important in both therapeutic and research contexts. Most current methodologies used to assemble these classes of biomolecules rely on chemical conjugations or non-covalent interactions and enzymatic strategies are generally limited to conjugation at the ends of the oligonucleotide of interest. Bioconjugation strategies are commercially available, requiring chemical modification of both the enzyme and oligonucleotide. Partially chemical, partially enzymatic strategies employing modified oligo nucleotides and self-labeling proteins have also been employed to recruit an enzyme to a transcript of interest via a covalently bound guide sequence.[79] Noncovalent strategies include transient RNA-protein interactions such as the MS2 coat protein which recognizes and associates itself with a stem-loop structure from the phage genome. Here I will utilize an engineered self-labeling protein in conjunction with our RNA-TAG labeling technology to establish a robust, covalent protein-nucleic acid conjugation method. Self-labeling proteins have been engineered to selectively recognize and covalently bond to their small molecule substrate. Three well established self-labeling protein systems are SNAP-tag, CLIP-tag, and HaloTag (Figure 4.1.). While this work uses the SNAP-tag system, it would be easily adapted to the other systems to generate three orthogonal conjugation systems.

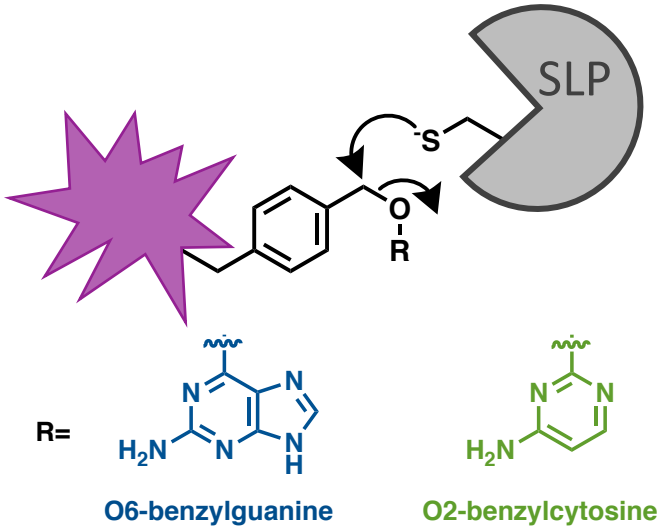



**Figure 4.1** SNAP-tag and CLIP-tag self-labeling protein scheme. A cysteine residue in the active site of the enzyme recognizes its small molecule substrate and covalently attaches to it via an $SN_2$ reaction. Benzylguanine is the substrate for SNAP-tag and benzylcytosine is the substrate for CLIP-tag.

The ability to conjugate nucleic acids with proteins offers the promise of assembling macromolecular machines, simultaneously harnessing the programmability of nucleic acids and the unique functions of proteins. To date there have been a number of methods developed to facilitate these functions, for instance, using a guide sequence covalently bound to a SNAP-ADAR construct to facilitate the conversion of adenosine to inosine, which is effectively recognized as guanosine, or using a sgRNA guided Cas fused (PilT N-terminus) PIN endonuclease to direct RNA degradation.[80] Here we use a bifunctional probe, with one end a substrate for TGT (PreQ1) and the other a substrate for SNAP-tag (benzylguanine) (Figure 4.2.).

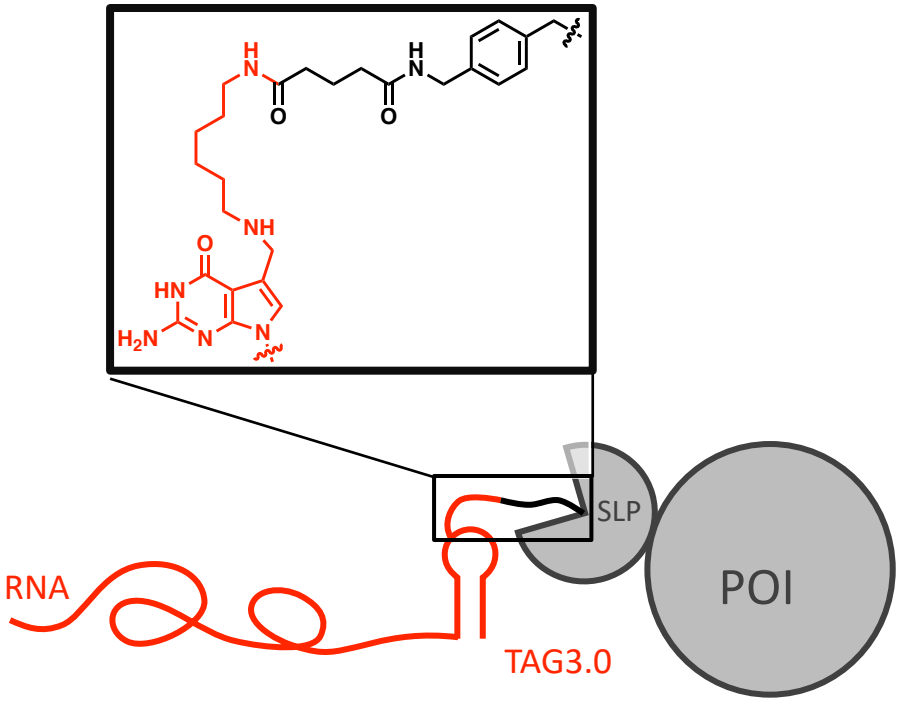



**Figure 4.2** Nucleic acid – protein of interest (POI) conjugate linked by the bifunctional PreQ1-benzylguanine probe. The probe is first inserted into the RNA using TGT then the SNAP-POI is added and attaches itself through benzylguanine.

The PIN domain is a common nuclease domain found in RNA degrading enzymes. To demonstrate the utility of my conjugation method, I designed SNAP-tag PIN fusion enzymes and cloned them into a mammalian expression vector and a bacterial expression vector with a C terminal His tag for purification. The goal with this enzyme was to directly recruit an endonuclease to an RNA of interest to initiate degradation in cells and *in vitro* (Figure 4.3.). The success of this work is grounds for the continued development of this macromolecular assembly technique, expanding it to other self-labeling protein systems and further exploring its use as a recruitment tool for acting on transcripts of interest in live cells.

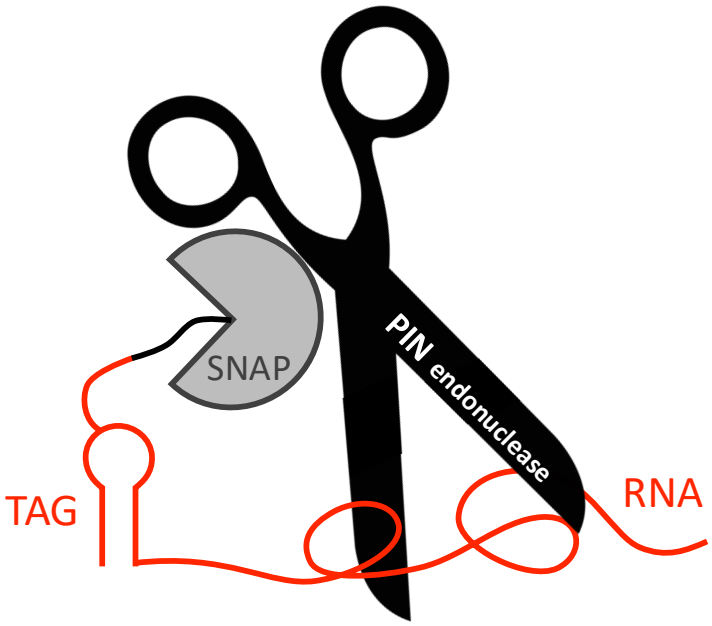



**Figure 4.3** Endonuclease recruitment to degrade a transcript of interest.

## 4.2 PreQ1-benzylguanine labeling and RNA-protein conjugation

The synthesis of PreQ1-benzylguanine was carried out with simple NHS chemistry, linking a benzylguanine NHS ester to the boc deprotected, commercially acquired PreQ1-C6-NH2 PreQ1 probe precursor. The product was purified, and its identity confirmed using LCMS. The resulting small molecule was tested for TGT recognition by inserting it into the 17 nucleotide ECYA1 (TAG2) hairpin and detected via urea PAGE gel shift analysis (Figure 4.4.). The benzylguanine modified TAG2 hairpin was quickly purified via spin column and incubated with commercially available SNAP-tag protein. The conjugated product was confirmed via SDS page gel shift, with the modified protein appearing higher on the gel due to an increase in molecular weight (Figure 4.4.). Both the labeling and conjugation steps go to near completion as evidenced by the disappearance of the starting material in the product lane of the gels.

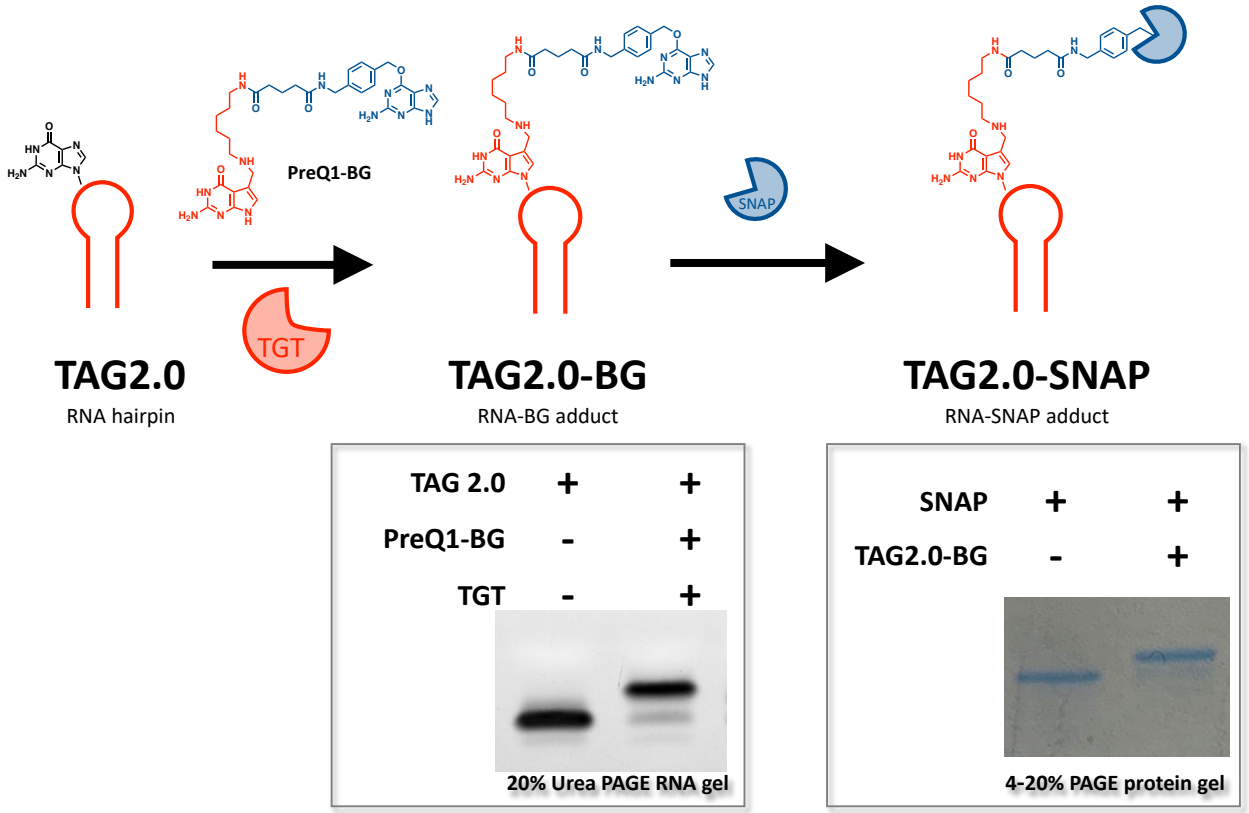



**Figure 4.4** TAG2 benzylguanine labeling and conjugation to SNAP-tag protein. TGT is able to efficiently insert the PreQ1-benzylguanine substrate into as can be seen in the upward shift on urea PAGE. The benzylguanine labeled RNA substrate is treated with commercially available SNAP-tag forming a macromolecular conjugate as seen in the SDS PAGE gel on the right.

Next, I investigated the ability of the SNAP enzyme to conjugate to longer, more interesting RNAs. I transcribed four TAG3 containing RNA constructs of varying sizes: a dummy ~200 nucleotide sequence with the TAG sequence near the 3′ end, the ~400 nucleotide 7SK RNA with the TAG sequence near the 5′ end, the ~500 nucleotide HDAC sequence with a TAG in the 3′ untranslated region (UTR), and the ~1000 nucleotide mCherry transcript also containing a TAG sequence in the 3′ UTR. Each RNA was efficiently conjugated to the SNAP-tag protein, as evidenced by urea PAGE gel shift analysis (Figure 4.5.). Since the insertion of the small molecule into the RNA is not large enough to create a gel shift on its own, the constructs were only analyzed after labeling and conjugation and the apparent efficiency a combination of both steps. The HDAC

transcript is seemingly able to be labeled twice, although with lower efficiency. This result was not further investigated as it was outside the scope of this work. However, it should be noted that there is a chance for off target labeling if a transcript contains a motif that is able to be recognized by TGT. As discussed in the earlier chapters introducing DNA-TAG, TGT is amenable to a number of different nucleic acid substrates, having a loose minimal requirement of a UGU motif.

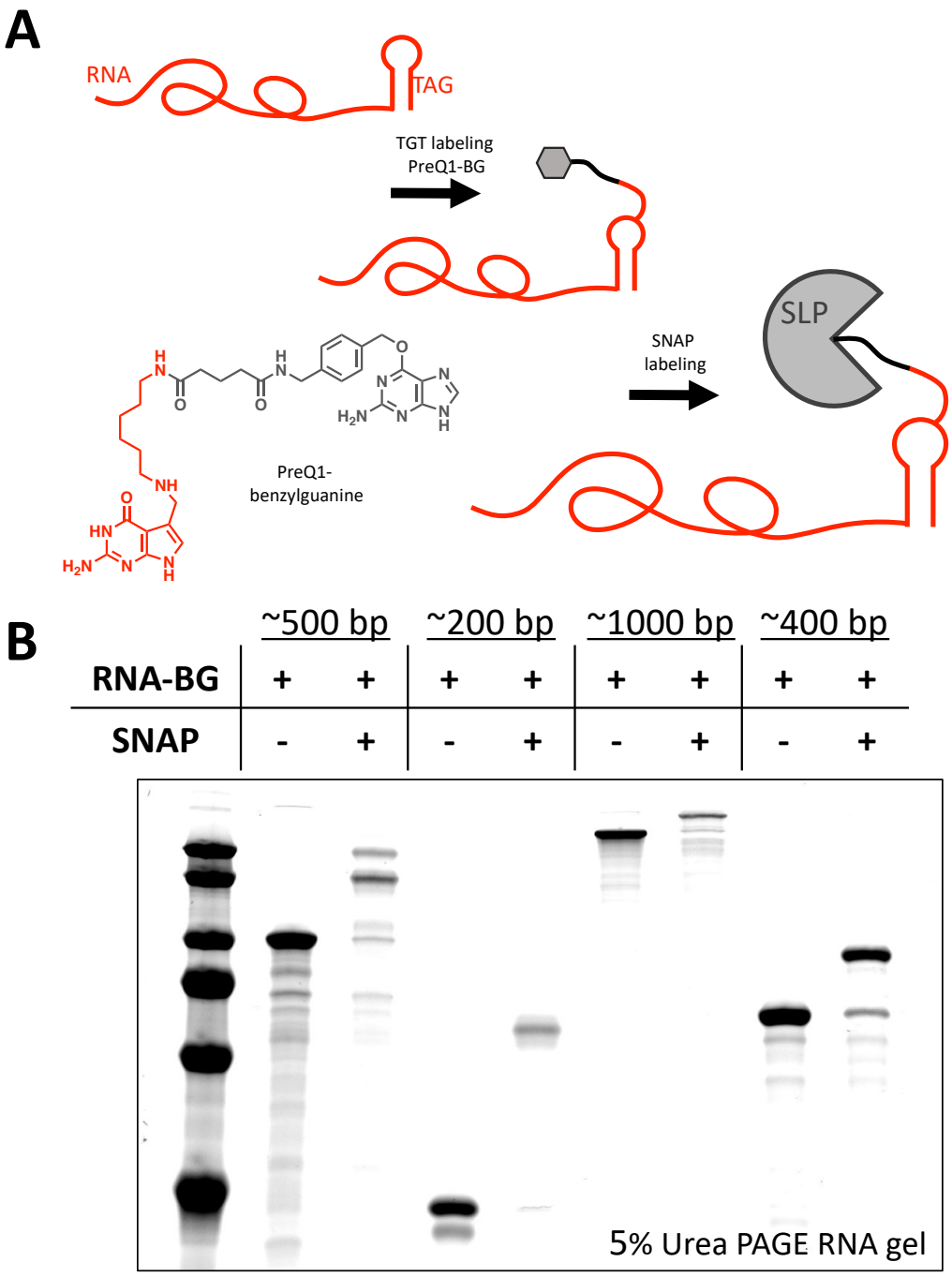



**Figure 4.5** Conjugation of longer RNAs with self-labeling proteins. (A) cartoon schematic of experimental workflow. The TAG3 containing transcript is modified with PreQ1-benzylguanine and subsequently treated with SNAP-tag. (B) 4 constructs of varying lengths are labeled with PreQ1-benzylguanine and treated with SNAP-tag. All reactions go to near completion as evidenced by the gel shift.

## 4.3 SNAP-PIN Dual Enzyme Construct

### *4.3.1 Creating Stable Cell Lines*

Facilitating the targeted degradation of a mRNA transcript in cells requires a particular workflow with different exogenously supplied actors introduced in a step wise manner. Specifically, the SNAP-PIN fusion needs to be expressed in the cells before the addition of the benzylguanine labeled target RNA transcript. In designing the workflow for these experiments, I decided making a stable cell line expressing the SNAP-PIN enzyme fusion was the most straightforward way to go. The cell lines were made using the Sleeping Beauty Transposon system.[81] In short, this unique system is made up of two mammalian expression vectors, one containing the sequence for a transposase and the other containing the sequence of the DNA being inserted into the genome. The plasmid containing the DNA sequence of interest also carries an antibiotic resistance gene for selection. The system has a variety of resistance gene and fluorescent protein reporters to choose from, allowing multiplexed insertion and selection. For this work only one enzyme sequence was necessary, and I chose to use puromycin selection and no fluorescent transfection marker. To make the stable cell lines, the cells are transfected with both plasmids simultaneously, the transposase is expressed in the cell, recognizes the unique cut sites on the second plasmid then cuts and inserts the desired DNA into the genome. Since the actual plasmid DNA is being inserted, you can influence the level of expression by adjusting the ratio of the transposon:gene plasmids. After transfection the media is supplemented with the selection antibiotic of choice for approximately two weeks and all cells not successfully modified are killed, yielding an enriched modified cell population.

Unsure of how introducing an exogenous nuclease would affect the HeLa cells, I created 4 cell lines with increasing amounts of the SNAP-PIN construct (Figure 4.6.). After selection I checked for the presence of SNAP-PIN with genomic PCR and by treating the cells with benzylguanine-siliconrhodamine and looked for fluorescence under the microscope. All of the cell lines exhibited fluorescence, although the 1:5 ratio cells showed the highest fluorescence. All cell lines continued to grow at a normal pace relative to unmodified HeLa cells and were taken forward for initial experiments.

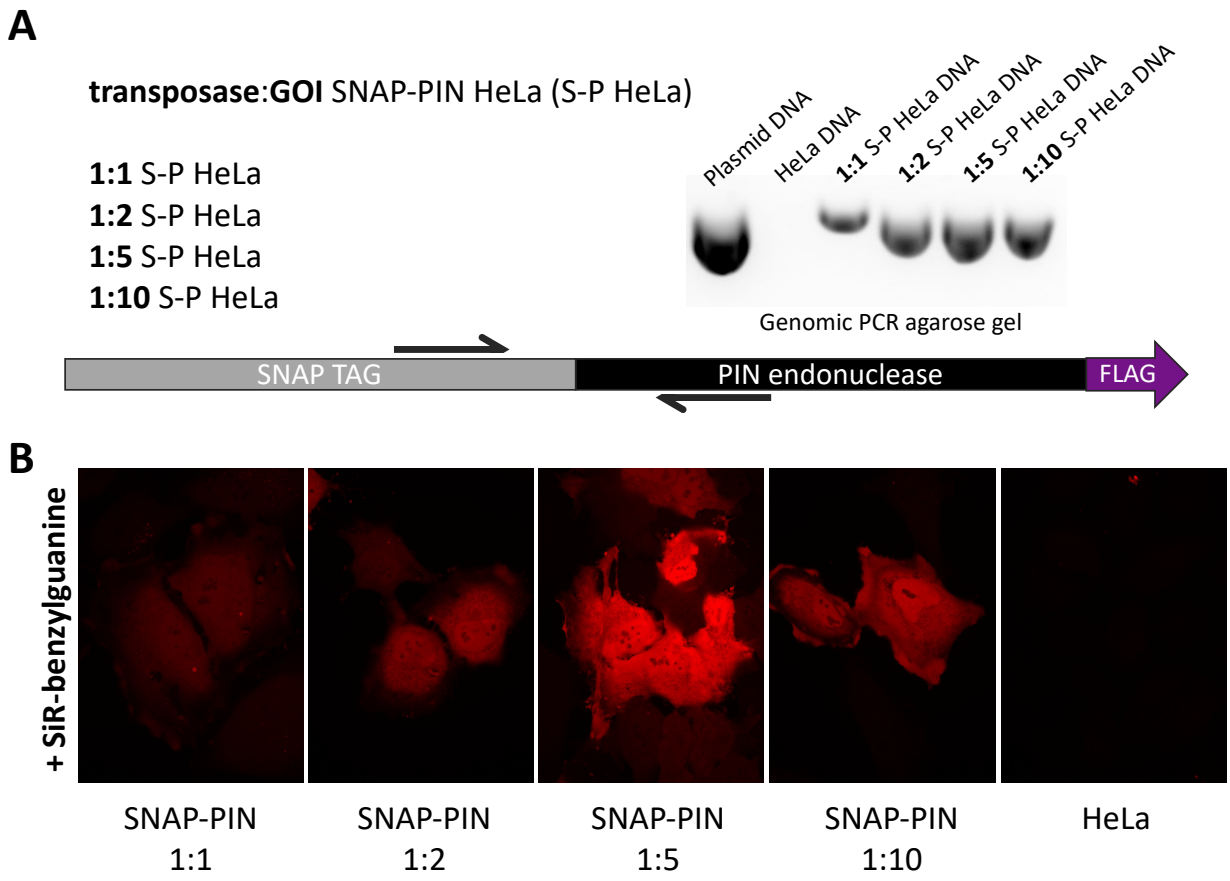



**Figure 4.6** Presence of SNAP-tag in stably cell lines. (A) PCR confirmation of the presence of the gene in extracted genomic DNA. (B) Treatment with siliconrhodamine (SiR)-benzylguanine to visualize SNAP-tag protein in cells.

### *4.3.2 mRNA IVT, Maturation, and Transfection*

To test whether the SNAP-PIN construct was able to degrade a benzylguanine labeled transcript in cells I decided to use mRNA coding for mCherry, a red fluorescent protein. In the case that the SNAP-PIN construct was successfully recruited to the transcript it should degrade the mCherry mRNA I should expect a decrease in fluorescent protein production over time.

When exogenous mRNA is added to cells it requires post transcriptional modifications to be properly expressed, including poly adenylation at the 3′ end and 7meG capping at the 5′ end. A plasmid containing the mCherry gene and a T3 hairpin in the 3′ UTR was used to IVT mCherry mRNA. This construct was then subjected to capping and poly adenylation and finally labeled with PreQ1-benzylguanine by TGT. Benzyl guanine installation was verified by treatment with SNAP-tag protein and urea PAGE gel shift analysis (Figure 4.7.). The stable cell lines were transfected with the benzylguanine modified mature mRNA to test PIN activity. Unfortunately, no difference was seen between the SNAP-PIN and wt HeLa cell lines, either by fluorescence microscopy or flow cytometry (Figure 4.8.).

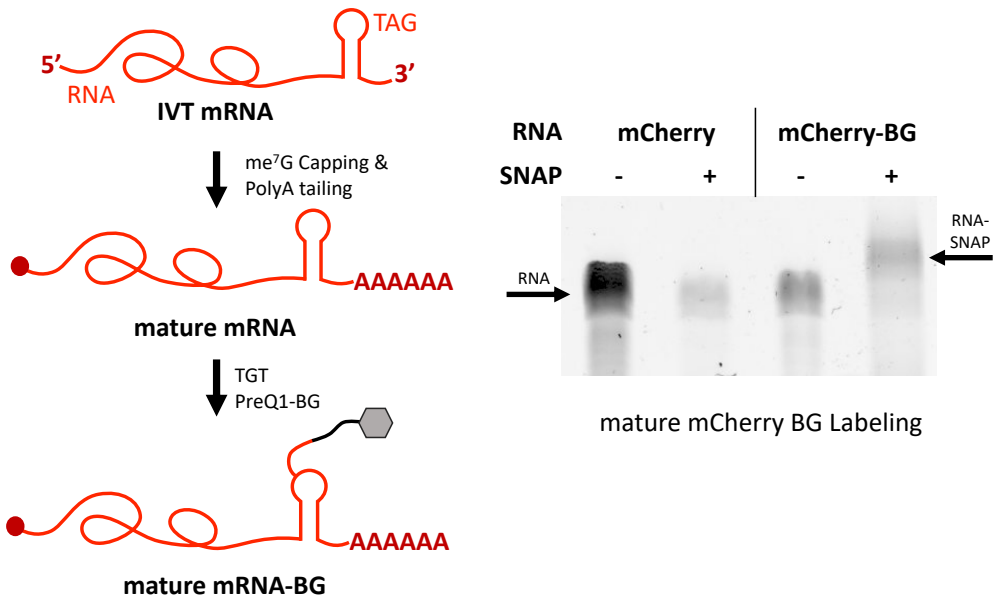
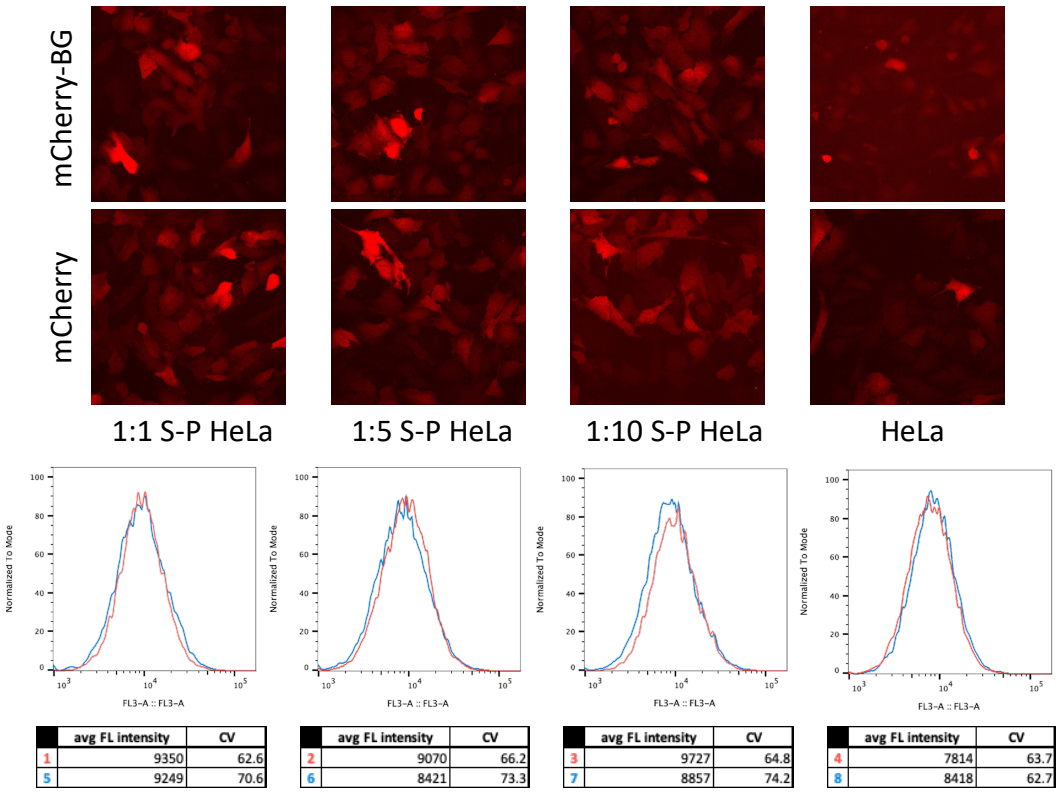



**Figure 4.7** mRNA maturation and benzylguanine modification. mCherry-TAG3 mRNA was IVT, capped, and polyadenylated. Subsequently, it was labeled with PreQ1-benzylguanine which was confirmed by gel shift after treatment with SNAP-tag. This mRNA was used for transfection reactions.




| | avg FL intensity | CV |
|---|---|---|
| 1 | 9350 | 62.6 |
| 5 | 9249 | 70.6 |

| | avg FL intensity | CV |
|---|---|---|
| 2 | 9070 | 66.2 |
| 6 | 8421 | 73.3 |

| | avg FL intensity | CV |
|---|---|---|
| 3 | 9727 | 64.8 |
| 7 | 8857 | 74.2 |

| | avg FL intensity | CV |
|---|---|---|
| 4 | 7814 | 63.7 |
| 8 | 8418 | 62.7 |

**Figure 4.8** mRNA transfection of stable cell lines. Mature mCherry and mCherry-benzylguanine is expressed in cells as evidenced by the fluorescent images. By eye there did not seem to be degradation of the transcript. The lack of degradation was confirmed by flow cytometry. While the SNAP-tag enzyme is active, the PIN domain is not.

### *4.3.3 In vitro Degradation of Benzylguanine Labeled RNA Constructs*

The same construct was cloned into a bacterial expression plasmid and recombinantly expressed and purified. Two benzylguanine labeled IVT RNA constructs, mCherry and HDAC TAG3, were treated with the recombinant enzyme fusion and assessed for degradation by urea PAGE. Similar to the cell experiments, no degradation was observed, however, the enzyme construct did successfully conjugate to the RNA, as evidenced by the clear upward shift of the RNA band in the gel (Figure 4.9.). These results indicated that the PIN domain was inactive. I decided to swap the order of the enzymes in my protein construct, putting the PIN domain at the N-terminus and the SNAP domain at the C terminus.

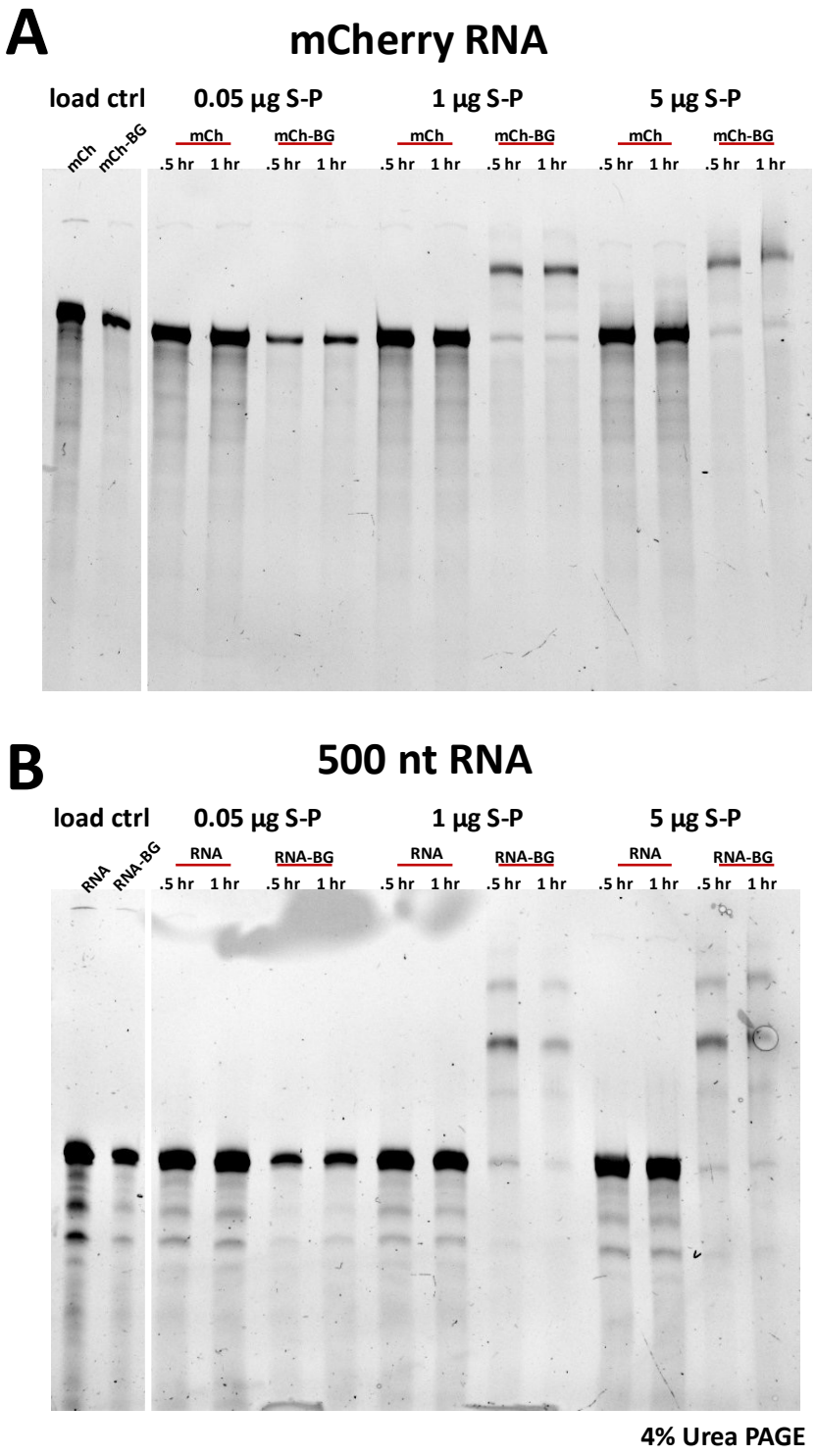



**Figure 4.9** SNAP-PIN degradation of RNA in vitro. The SNAP-PIN construct was cloned into a bacterial expression vector, expressed, and purified then assessed for its activity toward degrading transcripts *in vitro*. 200 ng of each RNA was incubated with the indicated amount of purified protein. Neither the (A) mCherry mRNA or (B) HDAC RNA were degraded by the SNAP-PIN construct but were both labeled with benzylguanine and able to conjugate to the SNAP-PIN construct.

## 4.4 PIN-SNAP Dual Enzyme Construct

### *4.4.1 Creating Stable Cell Lines*

The inactivity of the PIN domain in earlier experiments was likely due to sensitivity of placement in the protein fusion construct. I used Gibson assembly to insert clone the N-terminal PIN C-terminal SNAP construct into the sleeping beauty puromycin selection vector and verified the sequence via sanger sequencing. Simultaneously, the construct was cloned into a bacterial expression vector with a C terminal His tag. As described above, I generated stable HeLa cell lines expressing the PIN-SNAP construct in different ratios, this time only using three ratios of transposon:gene plasmids: 1:2, 1:5, and 1:10. All cell lines showed expression of the construct as determined by SNAP labeling with benzylguanine-Oregon green treatment. Before conducting more complex cellular based experiments I wanted to test whether the recombinantly expressed construct could successfully degrade a benzylguanine tagged RNA *in vitro*.

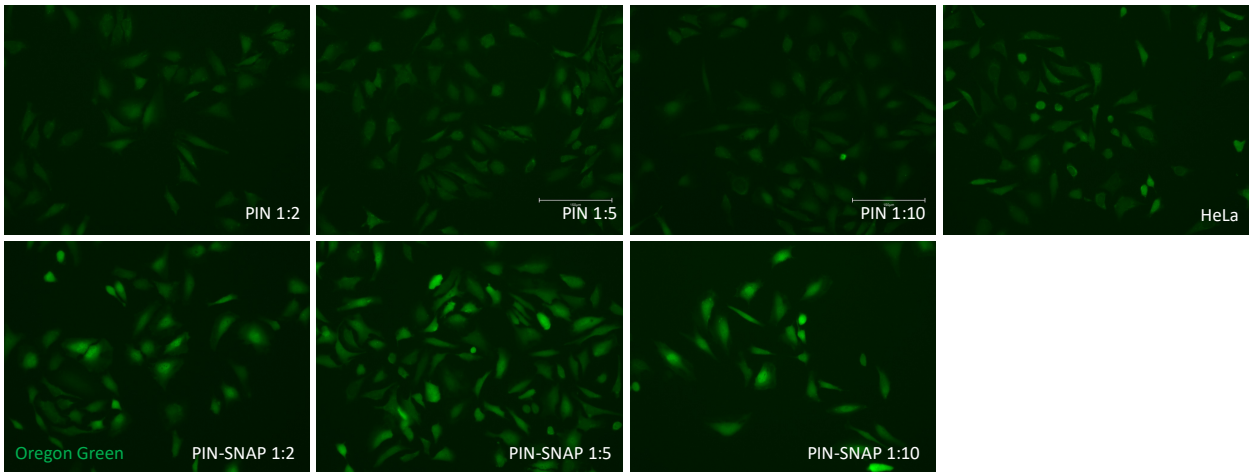



**Figure 4.10** Presence of SNAP-tag in stable cell lines. Treatment with Oregon Green-benzylguanine confirmed the presence of functional SNAP-tag in the PIN-SNAP transfected cell lines.

### *4.4.2 In vitro Degradation of Benzylguanine Labeled RNA Constructs*

*In vitro* transcribed mCherry-TAG3 and mCherry-BG transcripts were incubated with 0, 0.05, 1, or 5 µg of either the PIN domain recombinant enzyme or the PIN-SNAP construct for 4 hours at 37˚C. While some degradation of the RNA was seen for all of the reactions, a very clear increase in degradation was seen for the reactions containing both a benzylguanine labeled construct as well as the PIN-SNAP enzyme (Figure 4.11.). It should be noted that the experiment was designed to use equal µg amounts of both enzymes, however, the difference in molecular weight between the two creates a molar disparity. In fact, far more PIN only enzyme is needed to see similar degradation, further confirming the ability of recruitment to an RNA of interest by benzylguanine insertion and SNAP-tag recognition can direct enzymatic activity to a transcript of interest. Efficient targeted degradation occurs in the 1.3uM PIN-SNAP reaction, in the reaction with 2.5uM PIN only degradation is not efficient (Figure 4.11.). This piece of data is promising for the ability of the PIN-SNAP construct to work in live cells.

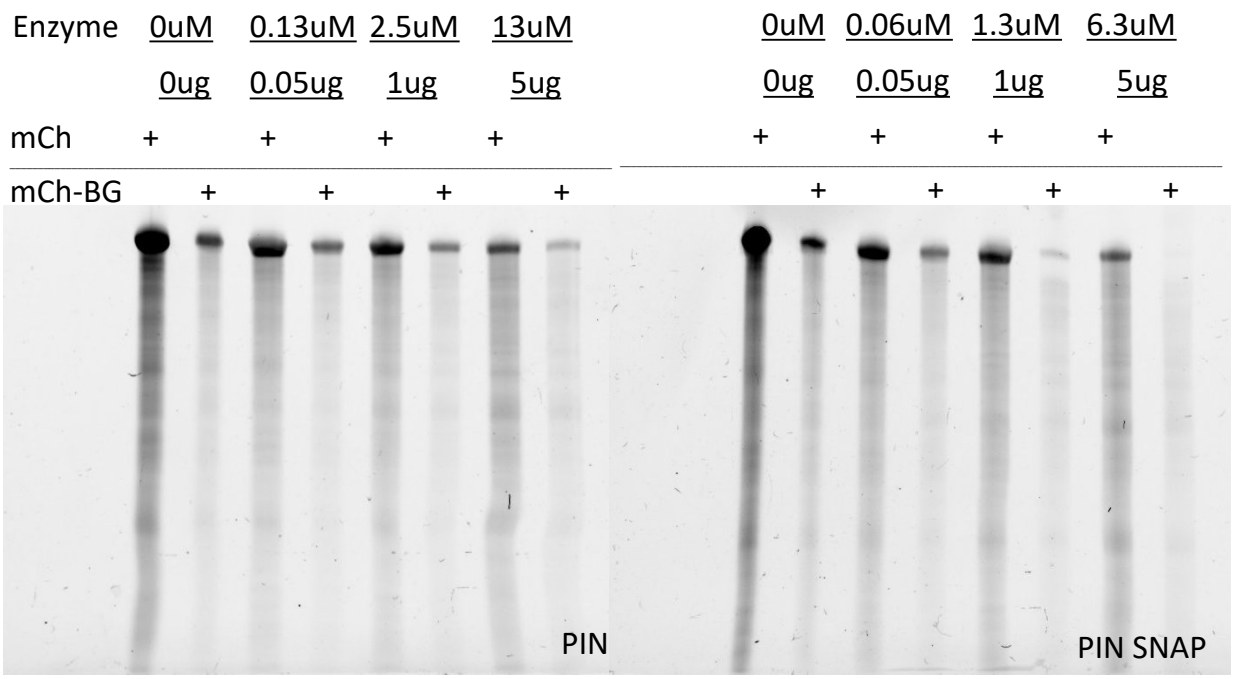



**Figure 4.11** PIN-SNAP degradation of RNA in vitro. The PIN-SNAP construct was cloned into a bacterial expression vector, expressed, and purified then assessed for its activity toward degrading transcripts *in vitro*. The reactions were carried out with the indicated amount of enzyme and 0.65 μM RNA. The PIN endonuclease is active in both the PIN only and PIN-SNAP constructs, but degradation is highly improved with recruitment of the SNAP conjugate directly to the transcript.

## 4.5 Conclusions

Assembling macromolecular constructs composed of different classes of biomolecules has the potential for great utility in fundamental research as well as therapeutics. The conjugation system outlined here is efficient and simple to execute. I successfully modified a variety of RNAs with a benzylguanine handle and using a well characterized self-labeling protein system was able to create RNA-enzyme conjugates. To show the utility of this system I designed a functional enzyme construct containing a nuclease domain, that when recruited to an RNA of interest was able to efficiently induce degradation. This demonstration of use only scratches the surface of what could be achieved using this conjugation strategy. Ongoing work should revisit the ability of the construct to degrade mRNA in live cells with the active PIN-SNAP construct and explore

the use of guide sequences to recruit the functional protein to RNA of interest, adding the ability to target endogenous RNA transcripts for degradation without direct modification. Additionally, this tool can be easily morphed into a multiplex capable technology by expanding it to include both CLIP-tag and HaloTag self-labeling protein systems.

## 4.6 Materials and methods

### *4.6.1 Urea PAGE*

For all gels, I used the 19:1 Acrylamide:Bisacrylamide SequaGel® UreaGel System and the Mini-PROTEAN Tetra Vertical Electrophoresis Cell from Bio-Rad. Briefly, 1.5mm plates assembled on the Bio-Rad gel casting stand, 10 mL of the desired percentage gel was freshly prepared in a 15 mL centrifuge tube, 14 µL of TEMED was added and gently mixed by inverting the tube, followed by 80 µL of 10% ammonium per sulfate. After briefly mixing the solution is poured between the plates, a comb inserted, and allowed to solidify (~20-30 min). All oligo samples were prepared for gel electrophoresis using 2X RNA loading dye from New England Biosciences. For all constructs, 150-200 ng of the oligonucleotide was loaded onto the gel for visualization. The solutions were heated to 98˚C for 5 minutes on a thermocycler. The electrophoresis cassette was assembled, and the chamber filled with 1X TBE. After thorough washing of the wells to remove leached urea, the samples were loaded into the wells and the gel was run for 90 min at 200V. Afterward, the gel was stained with gel red in TBE (1:1000 dilution) for ~5 min and imaged on a Bio-Rad ChemiDoc gel imager system.

### *4.6.2 SDS PAGE*

Denaturing protein gels were run using the Bio-Rad Mini Protean system. Precast 4-20% PAGE gels were used to analyze purified recombinant proteins and to analyze

SNAP-tag conjugates. Approximately 500 ng of protein was run for analysis gels and voltage and time recommendations were taken from the mini protean protocol. Gels were stained with instant blue and imaged after at least one hour.

### *4.6.3 SNAP imaging in cells*

Cells were incubated with commercially available SNAP-cell probes from NEB according to the manufacturers protocol. In short, the probe was diluted 1:200 in full media and added to the cells for 30 minutes at 37˚C. wash the cells twice with full media and replace once more, incubate another 30 minutes at 37˚C to allow unreacted benzylguanine-fluorophore to wash out. The cells were fixed in 4% PFA for 10 minutes at room temperature, washed with HBSS, and imaged on a fluorescent microscope.

### *4.6.4 TGT labeling*

TGT modification of RNA substrates was carried out at 37˚C for 4 hours in a thermocycler with a heated lid. Briefly, the reaction conditions are: 10 μM enzyme, 5 μM nucleic acid substrate, and 50 μM PreQ1-benzylguanine, 1X TGT reaction buffer, and 1-unit RNAsin from Promega. Components are mixed in no particular order, gently vortexed, briefly spun down, and incubated on a thermocycler. In the case that the reaction was set up at the end of the day, the thermocycler is set to go to 4˚C indefinitely after the 4-hour reaction time. For analysis purposes reactions were not purified before PAGE which did not affect the gel shift. For multistep reactions, the modified oligonucleotide product was purified away from excess small molecule using either the oligo clean and concentrator or RNA clean and concentrator kits from Zymo.

### *4.6.5 SNAP conjugation experiments*

SNAP-tag protein conjugation experiments were conducted for 1 hour at room temperature with 1.5-fold excess SNAP-tag protein for hairpin labeling experiments and 4 hours at 37˚C with 10-fold excess for longer RNA transcript labeling.

### *4.6.6 Cell culture*

HeLa cells were maintained according to ATCC guidelines in full media which consists of high glucose DMEM supplemented with 10% FBS and 1% pen strep. In short, cells were split at a 1:6 ratio every 2-3 days when they were ~80% confluent and maintained in tissue culture treated T25 flasks. To passage the cells, they were washed 1X with HBSS from Gibco and treated with TRYPLE from Gibco at 37˚C until fully lifted. The trypsin was quenched by adding an equal volume of complete medium and the necessary volume of cells taken into a new flask. For experiments, cells were seeded in a black glass bottom 24 well plate at a 1:6 ratio calculated to scale with the area of the plate.

### *4.6.7 Cloning*

A plasmid containing the PIN domain was ordered from Addgene (#104184). Another lab member gifted me a plasmid containing the SNAP-tag sequence for mammalian expression and a bacterial expression vector was ordered from Addgene (#101137). The sleeping beauty transfection system plasmids were ordered from Addgene (#60523). All constructs were designed using the NEBuilder online tool and assembled using the Gibson assembly strategy and the 2X HiFi Assembly master mix from NEB. The product was transformed into NEB® 5-alpha Competent *E. coli* cells and spread on a carbenicillin agar plate. The plate was incubated over night at 37˚C. The next

day, several colonies were selected, and 12 mL cultures grown in carbenicillin containing LB overnight. The new plasmids were extracted with the Biomiga mini prep kit and submitted to Eton Biosciences for sanger sequencing.

*4.6.8 Sleeping beauty stable transfection*

Stable cell lines expressing the PIN-SNAP and PIN constructs were generated using HeLa cells and the Sleeping Beauty transposon system. In 6 well plates with cells at ~80% confluency 1.5 ug of total DNA were transfected with 3 µL of lipofectamine 3000 from TThermofisher. Different ratios of SB100X enzyme:DNA template were used as described in the text above. After two days, each 6 well was expanded into a T25 flask and incubated with full media containing 250 ng/mL puromycin. Cells were split every two to three days and selection carried out for two weeks.

*4.6.9 Bacterial expression*

Briefly, the purified plasmids were transformed into BL21(DE3) Competent *E. coli* cells from New England Biosciences using the protocol described on the Addgene website. A starter culture was grown overnight in the presence of carbenicillin for selection. Approximately 5mL of starter culture was transferred into 200 mL of carbenicillin containing LB and allowed to grow to an optical density of ~0.6. At this point expression was induced with by adding IPTG to a final concentration of 1mM and shaken at 37˚C for 4 hours. The bacteria were pelleted 6,000xg for 20 minutes at 4˚C. Cell pellets can be resuspended and purified immediately or stored at -20˚C if desired.

*4.6.10 In vitro transcription of RNAs*

All plasmids used for *in vitro* transcription were cloned by another member of the lab. The empty plasmid encoding the dummy sequence is available on Addgene

(#138209). Templates were prepared by linearization using XbaI for mCherry, HDAC, and the dummy sequence. PCR amplification was used to amplify the 7SK construct and attach a T7 promoter with the following primers: forward-AAGCTGTAATACGACTCACTATAGGGGATCCCCGGGAGCAGAC, and reverse-AAAAGAAAGGCAGACTGCCACATG. The cut templates were purified by phenol chloroform extraction to ensure removal of RNAses and subsequently ethanol precipitated. Transcription reactions were set up with rNTPs (5 mM each ATP, CTP, UTP, 9 mM GTP) from NEB, 0.004 U/μL Thermostable Inorganic Pyrophosphatase from NEB 0.15 μg/μL T7-RNAP made in house, and 0.05% Triton X-100 from Sigma Aldrich in 1X T7 Reaction Buffer. For each reaction, ~4 μg of cut plasmid or 1 μg of purified PCR product was used for a 100 μL transcription reaction. The reactions were run at 37°C for 4 hours. The transcripts were purified via lithium chloride precipitation by addition of LiCl to a final concentration of 2.5 M and refrigeration at -20°C for overnight. The RNA was quantified via absorbance at 260 nm, and verified via 4% urea PAGE (4% polyacrylamide in TBE with 8M urea) and kept frozen at -20 °C.

*4.6.11 Capping and poly A tailing*

mRNA capping was carried out according to the table below. The RNA and water were added to a PCR tube and heated at 65˚C for 5 minutes then cooled on ice for an additional 5 minutes. All other components were added, and the reaction was allowed to proceed for 2 hours at 37˚C. The product was taken directly to the polyadenylation step without purification. All components for the polyadenylation reaction were added to the capping reaction and the reaction allowed to proceed for 90 minutes at 37 ˚C. The mature mRNA was ethanol precipitated and resuspended un ultra-pure water.

**Table 4.1** mRNA in vitro capping reaction conditions.

| **component** | **stock conc** | **conditions** | **µL** |
|---|---|---|---|
| buffer | 10X | 1X | 6 |
| GTP | 10 mM | .5 mM | 3 |
| SAM | 8 µM | .4 µM | 3 |
| RNAsin | | | 1.5 |
| Capping enzyme | | | 3 |
| RNA | | 30 ug | |
| **Reaction Vol:** | **60 µL** | | |

**Table 4.2** mRNA in vitro poly A tailing reaction conditions.

| **component** | stock conc | conditions | **µL** |
|---|---|---|---|
| **Buffer** | | 1X | 8 |
| **ATP** | 10 mM | 1mM | 8 |
| **RNAsin** | | | 2 |
| **Tailing enzyme** | | | 3 |
| **RNA** | Total capping rxn | 30 ug | 60 |
| **Reaction vol:** | **80 µL** | | |

*4.6.12 Genomic PCR*

The gene fragment was amplified from ~50ng of isolated genomic DNA from the stable cell lines cells using Q5® Hot Start High-Fidelity 2X Master Mix and the following primers: Fwd: GAAACTGCTGAAAGTGGTGAAGTTCGGAGAG Rev: GATGCTTTTTCG CGCCTTCTCTTGGAC. The PCR product was verified by running 5uL of the crude PCR reaction on a native agarose gel.

*4.6.13 Flow cytometry*

Fluorescence was measured using the Accuri C6 flow cytometer. In short, cells from Figure 4.6. were washed twice with HBSS, trypsinized, fixed in 4% PFA for 10 minutes, washed 2X with HBSS, and resuspended in 0.1% BSA/HBSS for flow cytometry.

*4.6.14 In vitro RNA cleavage*

Components from Figures 4.9. and 4.11 were added in no particular order at the concentrations indicated and allowed to run fun 1 hour and 4 hours for the SNAP-PIN and PIN-SNAP experiments, respectively. Afterward ~100-200ng of RNA were loaded on 4% urea PAGE, stained with gel red and imaged on a Bio-Rad ChemiDoc imager.

## 4.7 Acknowledgements

Chapter four, in part is currently being prepared for submission for publication. Tota, Ember M.; Devaraj, Neal K. The dissertation author was the primary researcher and author of this material.

# 5 Conclusions

## 5.1 Driving Innovations with Nucleic Acid-Based Technologies

The power of nucleic acids has extended far beyond its biological role as the blueprint for life. The unmatched versatility and programmability of nucleic acids, DNA in particular, has made them a critical tool for the advancements of many technologies across several fields. Employing their unique coding ability to direct desired function has enabled groundbreaking technologies like CRISPR-Cas gene editing,[43,44] RNA base editing,[82] antisense therapies,[83] aptamer delivery, DNA barcoding,[84] DNA origami,[85] data storage, in situ hybridization (ISH),[67] innovative sequencing strategies, and a variety of *in vitro* tools. Most of these technologies require their respective oligonucleotides to be modified in some way, most often by covalent attachment of a reporter or affinity ligand. Tools to generate these modifications are critical yet lacking, specifically in the realm of single stranded DNA modification. The drive to develop new innovative nucleic acid modification strategies is fueled by the shortcomings of current technologies as well as the promise of their role in enabling yet to be conceptualized tools.

This promise is what has driven the development of nucleic acid transglycosylation at guanine (NA-TAG) technologies in our lab which have proven to be quite powerful in the applications we have explored thus far. The ability to site specifically modify RNA has enabled applications like RNA centric proteomic studies, light activated control of Cas9 editing in cells, mRNA translation control, wash free RNA imaging, enzymatic labeling of modRNA substrates, and fluorogenic mRNA detection.[18–23,57,86] The expansion of RNA-TAG to recognize and modify DNA substrates takes it beyond its use to study RNA and manipulate cellular processes and into the realm of a powerful *in vitro* tool. The higher stability of DNA oligonucleotides in comparison to RNA makes it the NA of choice for

many of the applications described above. With its new expansion to include DNA substrates, the power of NA-TAG lies in its versatility, a single enzyme that can insert a wide variety of functional small molecule probes into virtually any single stranded ribose based nucleic acid substrate.

## 5.2 DNA-TAG

Currently there are two popular ways to site specifically modify DNA oligonucleotides: Chemical modification during solid phase synthesis, and 3′ insertion of probes using terminal deoxynucleotidyl transferase (TdT). Both strategies have proven critical for the use of DNA technologies, however, they are both limited in their utility. Chemical synthesis of modified oligonucleotides is cost prohibitive and time consuming as they often require multiple rounds of HPLC purification. TdT offers researchers an in-house option for generating modified oligos, however, the probe tolerance of the enzyme is narrow, and most probes require a two-step assembly process. Additionally, this enzymatic strategy can only insert the labeled nucleotide into the three prime end and is somewhat limited to inserting a single label or multiple with no control over the actual number being inserted.

## 5.3 Macromolecular assemblies

The combination of biological molecules to assemble macromolecular machines is one area that has proven fruitful. Proteins, lipids, glycans, and nucleic acids all play an important and unique role in biological processes. Combining their abilities to generate useful new tools will have an impact on biomedical applications, fundamental research, engineering, and clinical tools and therapeutics, among other things. Enzymes have proven to be a powerful tool to be harnessed by scientists both for fundamental research applications as well as by industry as innovative technologies with novel applications.

Joining enzymatic functionality with programmability of nucleic acids has led to critical and transformative technology developments, for instance, gene editing and RNA modification strategies. By combining two powerful enzymatic tools and a simple bifunctional small molecule I have developed a strategy to effortlessly conjugate a nucleic acid and a protein of interest, with the potential for multiplexing applications and without the need to purchase cost prohibitive modified oligonucleotides. This strategy can be used with both RNA-TAG and DNA-TAG.

### 5.4 Future outlooks for NA-TAG

The application space of nucleic acid modification and labeling is vast and mostly unexplored. Purchasing modified oligonucleotides is cost prohibitive, limiting the availability and therefore advancements that could otherwise be developed. The development and characterization of DNA-TAG described here offers a cost effective, efficient alternative to purchasing chemically modified constructs. The versatility of NA-TAG technologies offers a single enzyme capable of inserting a variety of small molecule probes into nearly any single stranded nucleic acid construct of interest. With the optimization and further characterization of the technology described in this work, DNA-TAG has the potential to offer an even further minimized recognition element that would be suitable for short constructs in which a 17-nucleotide hairpin might be too perturbing for the intended function.

This work also discusses using NA-TAG to facilitate the functional conjugation of two macromolecules. From functional recruitment to guide directed enzymatic activity, this technology has immense potential. Expansion to include other self-labeling proteins and their small-molecule counterparts will offer the ability for multiplexed recruitment events.

Rather than modifying an existing nucleotide, this class of enzymes swaps one base for another. This mechanism of action offers unparalleled potential for developing technologies like base exchange and insertion of novel nucleobases.[15,33] The applications described here just scratch the surface of what is possible with NA-TAG technologies. *E. coli* TGT has proven to be very versatile, amenable to mutations, and tolerant of changes to both its small molecule and nucleic acid substrates.[33] NA-TAG technologies using the wildtype version of the enzyme offer a window into what could be possible through the engineering and evolution of TGTs with novel targets.

# Appendix for Chapter 2: dTAG Hairpins

**Table 2.1** dTAG mutant hairpin sequences.

| | Construct | Sequence |
|---|---|---|
| **45** | **TYR-HIS** | GCAGATTGTGATTCTGC |
| **46** | **HIS-TYR** | CTGGACTGTAAATCCAG |
| **47** | **TYR-ASP** | GCAGACTGTCACTCTGC |
| **48** | **ASP-TYR** | CCTGCCTGTAAAGCAGG |
| **49** | **TYR-ASN** | GCAGACTGTTAATCTGC |
| **50** | **ASN-TYR** | GCGGACTGTAAATCCGT |
| **51** | **HIS-ASP** | CTGGACTGTCACTCCAG |
| **52** | **ASP-HIS** | CCTGCTTGTGATGCAGG |
| **53** | **HIS-ASN** | CTGGACTGTTAATCCAG |
| **54** | **ASN-HIS** | GCGGATTGTGATTCCGT |
| **55** | **ASP-ASN** | CCTGCCTGTTAAGCAGG |
| **56** | **ASN-ASP** | GCGGACTGTCACTCCGT |
| **64** | **TYR** | GCAGACTGTAAATCTGC |
| **65** | **HIS** | CTGGATTGTGATTCCAG |
| **66** | **ASP** | CCTGCCTGTCACGCAGG |
| **67** | **ASN** | GCGGACTGTTAATCCGT |
| **68** | **TYR T3** | GGGAGCAGACTGTAAATCTGCTCCC |
| **69** | **HIS T3** | GGGACTGGATTGTGATTCCAGTCCC |
| **70** | **ASP T3** | GGGACCTGCCTGTCACGCAGGTCCC |
| **71** | **ASN T3** | GGGAGCGGACTGTTAATCCGTTCCC |
| **72** | **UGU T3** | GGGAGCAGACUGUAAATCTGCTCCC |
| **73** | **UGT T3** | GGGAGCAGACUGTAAATCTGCTCCC |
| **74** | **TGU T3** | GGGAGCAGACTGUAAATCTGCTCCC |
| **75** | **TGA T3** | GGGAGCAGACTGAAAATCTGCTCCC |
| **76** | **TGC T3** | GGGAGCAGACTGCAAATCTGCTCCC |
| **77** | **TGG T3** | GGGAGCAGACTGGAAATCTGCTCCC |
| **78** | **AGU T3** | GGGAGCAGACAGUAAATCTGCTCCC |
| **79** | **CGU T3** | GGGAGCAGACCGUAAATCTGCTCCC |
| **80** | **GGU T3** | GGGAGCAGACGGUAAATCTGCTCCC |
| **81** | **TCT T3** | GGGAGCAGACTCTAAATCTGCTCCC |
| **94** | **ASP M1** | CCTCGCTGTCACCGAGG |
| **95** | **ASP M2** | GGACGCTGTCACCGTCC |
| **96** | **ASP M3** | GGAGCCTGTCACGCTCC |
| **97** | **ASP M4** | CCGCGCTGTCACCGCGG |
| **98** | **ASP M5** | CCTGCCTGTCCCGCAGG |

| | | |
|---|---|---|
| **99** | **ASP M6** | CCTGCCTGTCTCGCAGG |
| **100** | **ASP M7** | CCTCGCTGTCCCCGAGG |
| **101** | **ASP M8** | CCTCGCTGTCTCCGAGG |
| **102** | **ASP M9** | CCTGACTGTCACTCAGG |
| **103** | **ASP M10** | CCTCACTGTCACTGAGG |
| **104** | **TYR M1** | GCAGACTGTAAATCTGT |
| **105** | **TYR M2** | GCAAACTGTAAATTTGC |
| **106** | **TYR M3** | ACAGACTGTAAATCTGT |
| **107** | **TYR M4** | CCAGACTGTAAATCTGG |
| **108** | **TYR M5** | GCAGACTGTAATTCTGC |
| **109** | **TYR M6** | GCAGACTGTAACTCTGC |
| **110** | **TYR M7** | GCAGACTGTTAATCTGC |
| **111** | **TYR M8** | GCAGACTGTCAATCTGC |
| **112** | **HIS M1** | CTCGATTGTGATTCGAG |
| **113** | **HIS M2** | CTAGATTGTGATTCTAG |
| **114** | **HIS M3** | CTGGATTGTTATTCCAG |
| **115** | **HIS M4** | CTGCATTGTGATTGCAG |
| **116** | **HIS M5** | CTGAATTGTGATTTCAG |
| **117** | **HIS M6** | GTGGATTGTGATTCCAC |
| **118** | **HIS M7** | GTGGATTGTGATTCCAT |
| **119** | **HIS M8** | ATGGATTGTGATTCCAT |
| **120** | **HIS M9** | CTGGATTGTCATTCCAG |
| **121** | **ASN M1** | GCGGACTGTTAATCCGC |
| **122** | **ASN M2** | ACGGACTGTTAATCCGT |
| **123** | **ASN M3** | GCCGACTGTTAATCGGT |
| **124** | **ASN M4** | GCGCACTGTTAATGCGT |
| **125** | **ASN M5** | GCGGACTGTTATTCCGT |
| **126** | **ASN M6** | GCGGACTGTTATTCCGC |
| **127** | **ASN M7** | GCGGACTGTTACTCCGT |
| **128** | **ASN M8** | GCGGACTGTTACTCCGC |
| **129** | **ASN M9** | GTGGACTGTTAATCCAT |
| **130** | **HIS M10** | CTGGATTGTCCTTCCAG |
| **131** | **HIS M11** | CTGGATTGTTTTTCCAG |
| **132** | **HIS M12** | CTGGATTGTTATTCCAG |
| **133** | **HIS M13** | CTGGATTGTCTTTCCAG |
| **134** | **HIS M14** | CTGGATTGTCGTTCCAG |
| **135** | **HIS M15** | CTGGATTGTTGTTCCAG |
| **136** | **HIS M16** | CTGGATTGTTCTTCCAG |
| **137** | **HIS M17 (HIS-ASP)** | CTGGACTGTCACTCCAG |

| | | |
|---|---|---|
| **138** | **HIS M18** | CTGGACTGTCCCTCCAG |
| **139** | **HIS M19** | CTGGACTGTTTCTCCAG |
| **140** | **HIS M20** | CTGGACTGTTACTCCAG |
| **141** | **HIS M21** | CTGGACTGTCTCTCCAG |
| **142** | **HIS M22** | CTGGACTGTCGCTCCAG |
| **143** | **HIS M23** | CTGGACTGTTGCTCCAG |
| **144** | **HIS M24** | CTGGACTGTTCCTCCAG |
| **145** | **HIS M25** | CTGGATTGTCACTCCAG |
| **146** | **HIS M26** | CTGGATTGTCCCTCCAG |
| **147** | **HIS M27** | CTGGATTGTTTCTCCAG |
| **148** | **HIS M28** | CTGGATTGTTACTCCAG |
| **149** | **HIS M29** | CTGGATTGTCTCTCCAG |
| **150** | **HIS M30** | CTGGATTGTCGCTCCAG |
| **151** | **HIS M31** | CTGGATTGTTGCTCCAG |
| **152** | **HIS M32** | CTGGATTGTTCCTCCAG |
| **153** | **HIS M33** | CTGGACTGTCATTCCAG |
| **154** | **HIS M34** | CTGGACTGTCCTTCCAG |
| **155** | **HIS M35** | CTGGACTGTTTTTCCAG |
| **156** | **HIS M36** | CTGGACTGTTATTCCAG |
| **157** | **HIS M37** | CTGGACTGTCTTTCCAG |
| **158** | **HIS M38** | CTGGACTGTCGTTCCAG |
| **159** | **HIS M39** | CTGGACTGTTGTTCCAG |
| **160** | **HIS M40** | CTGGACTGTTCTTCCAG |
| **161** | **161_Ala1_H10** | CTTGCTTGTCCTGCAAG |
| **162** | **162_Ala2_H10** | CCTGCTTGTCCTGCAGG |
| **163** | **163_arg1_H10** | AGCCCTTGTCCTGGGCT |
| **164** | **164_arg2_H10** | CTCGGTTGTCCTCCGAG |
| **165** | **165_arg3_H10** | CTGCCTTGTCCTGGCAG |
| **166** | **166_arg4_H10** | ACGACTTGTCCTGTCGT |
| **167** | **167_asn1_H10** | GCGGATTGTCCTTCCGT |
| **168** | **168_asp1_H10** | CCTGCTTGTCCTGCAGG |
| **169** | **169_cys1_H10** | GCGGATTGTCCTTCCGT |
| **170** | **170_gln1_H10** | CCGGATTGTCCTTCCGG |
| **171** | **171_gln2_H10** | CCGGTTTGTCCTACCGG |
| **172** | **172_glu1_H10** | CCGCCTTGTCCTGGCGG |
| **173** | **173_gly1_H10** | AGAGCTTGTCCTGCTCT |
| **174** | **174_gly2_H10** | TCAGCTTGTCCTGCTGA |
| **175** | **175_gly3_H10** | CGACCTTGTCCTGGTCG |
| **176** | **176_his1_H10** | CTGGATTGTCCTTCCAG |

| | | |
|---|---|---|
| **177** | **177_ile1_H10** | CACCCTTGTCCTGGGTG |
| **178** | **178_leu1_H10** | GTTGATTGTCCTTCAAC |
| **179** | **179_leu2_H10** | CTAGCTTGTCCTGTTAG |
| **180** | **180_leu3_H10** | CTACCTTGTCCTGGTAG |
| **181** | **181_leu4_H10** | CCAGATTGTCCTTCTGG |
| **182** | **182_leu5_H10** | AGGGATTGTCCTTCCCT |
| **183** | **183_lys1_H10** | GTTGATTGTCCTTCAAT |
| **184** | **184_met1_H10** | GGCGATTGTCCTTCGCT |
| **185** | **185_met2_H10** | CATCATTGTCCTTGATG |
| **186** | **186_met3_H10** | GGCGATTGTCCTTCGCT |
| **187** | **187_met4_H10** | TCGGGTTGTCCTCCCGA |
| **188** | **188_phe1_H10** | GGGGATTGTCCTTCCCC |
| **189** | **189_pro1_H10** | ACTGGTTGTCCTCCAGT |
| **190** | **190_pro2_H10** | CCGTCTTGTCCTGTCGG |
| **191** | **191_pro3_H10** | CTTCGTTGTCCTCGAGG |
| **192** | **192_ser1_H10** | CCGGTTTGTCCTACCGG |
| **193** | **193_ser2_H10** | CCGGTTTGTCCTACCGG |
| **194** | **194_ser3_H10** | CTCCCTTGTCCTGGGAG |
| **195** | **195_ser4_H10** | CACGCTTGTCCTGTGTG |
| **196** | **196_thr1_H10** | CACCCTTGTCCTGGGTG |
| **197** | **197_thr2_H10** | TCGCATTGTCCTTGCGA |
| **198** | **198_thr3_H10** | CGCATTTGTCCTATGCG |
| **199** | **199_thr4_H10** | ACTGATTGTCCTTCAGT |
| **200** | **200_thr5_H10** | CACCCTTGTCCTGGGTG |
| **201** | **201_trp1_H10** | CCGGTTTGTCCTACCGG |
| **202** | **202_tyr1_H10** | GCAGATTGTCCTTCTGC |
| **203** | **203_val1_H10** | CCACCTTGTCCTGGTGG |
| **204** | **204_val2_H10** | CCTCCTTGTCCTGGAGG |
| **205** | **His M10 T2 int** | ACCATCATTTTCATATCCTCCACTGGATTGTCCTTCCAGACCACCATCATTTGCAATGA |
| **206** | **His M10 T2 3′** | ACCATCATTTTCATATCCTCCAACCACCATCATTTGCAATGACTGGATTGTCCTTCCAG |
| **207** | **His M10 T2 5′** | CTGGATTGTCCTTCCAGACCATCATTTTCATATCCTCCAACCACCATCATTTGCAATGA |
| **208** | **His M10 T3** | GGGACTGGATTGTCCTTCCAGTCCC |
| **209** | **His M10 T3 dC** | GGGACTGGATTCTCCTTCCAGTCCC |
| **210** | **His M10 T2 dC** | CTGGATTCTCCTTCCAG |
| **211** | **His M10 T2 dT** | CTGGATTTTCCTTCCAG |
| **212** | **His M10 T2 dA** | CTGGATTATCCTTCCAG |

| | | |
|---|---|---|
| **226** | **dualHP_5/3_ 177/176_GG** | CACCCTTGTCCTGGGTGCATCATTTTCATATCCTC CACTGGATTGTCCTTCCAG |
| **227** | **dualHP_5/3_ 177/176_GC** | CACCCTTGTCCTGGGTGCATCATTTTCATATCCTC CACTGGATTCTCCTTCCAG |
| **228** | **dualHP_5/3_ 177/176_CG** | CACCCTTCTCCTGGGTGCATCATTTTCATATCCTC CACTGGATTGTCCTTCCAG |
| **229** | **dualHP_5/3_ 177/177_GG** | CACCCTTGTCCTGGGTGCATCATTTTCATATCCTC CACACCCTTGTCCTGGGTG |
| **230** | **dualHP_5/3_ 176/176_GG** | CTGGATTGTCCTTCCAGCATCATTTTCATATCCTC CACTGGATTGTCCTTCCAG |
| **231** | **dualHP_5/3_ 176/177_GG** | CTGGATTGTCCTTCCAGCATCATTTTCATATCCTC CACACCCTTGTCCTGGGTG |
| **232** | **dualHP_5/3_ 176/177_GC** | CTGGATTGTCCTTCCAGCATCATTTTCATATCCTC CACACCCTTCTCCTGGGTG |
| **233** | **dualHP_5/3_ 176/177_CG** | CTGGATTCTCCTTCCAGCATCATTTTCATATCCTC CACACCCTTGTCCTGGGTG |
| **234** | **dualHP_5/5_ 176/177_GG** | CTGGATTGTCCTTCCAGTTTCACCCTTGTCCTGGG TGCATCATTTTCATATCCTCCA |
| **235** | **dualHP_5/5_ 176/177_GC** | CTGGATTGTCCTTCCAGTTTCACCCTTCTCCTGGG TGCATCATTTTCATATCCTCCA |
| **236** | **dualHP_5/5_ 176/177_CG** | CTGGATTCTCCTTCCAGTTTCACCCTTGTCCTGGG TGCATCATTTTCATATCCTCCA |
| **237** | **dualHP_3/3_ 176/177_GG** | CATCATTTTCATATCCTCCACTGGATTGTCCTTCC AGTTTCACCCTTGTCCTGGGTG |
| **238** | **dualHP_3/3_ 176/177_GC** | CATCATTTTCATATCCTCCACTGGATTGTCCTTCC AGTTTCACCCTTCTCCTGGGTG |
| **239** | **dualHP_3/3_ 176/177_CG** | CATCATTTTCATATCCTCCACTGGATTCTCCTTCC AGTTTCACCCTTGTCCTGGGTG |
| **240** | **free 5′ loop extended** | TTGTCCTTCCAGCACCC |
| **241** | **free 3′ loop extended** | GGGTGCTGGATTGTCCT |
| **242** | **free 5′ loop** | TTGTCCTTCCAG |
| **243** | **free 3′ loop** | CTGGATTGTCCT |
| **244** | **HP 176 T3 5′I** | GGGACTGIATTGTCCTTCCAGTCCC |
| **245** | **HP 177 T3 3′I** | GGGACACCCTTGTCCTIGGTGTCCC |

# Appendix for Chapter 3: FISH Probe Set Sequences

**Table 3.3** βActin FISH probe set sequences

| Probe | | **Sequence** |
|---|---|---|
| **1** | **actinB-81** | CGATATCATCATCCATGGTGCCTGCTTGTCCTGCAGG |
| **2** | **actinB-170** | CACGATGGAGGGGAAGACGGCCTGCTTGTCCTGCAGG |
| **3** | **actinB-226** | ACATAGGAATCCTTCTGACCCCTGCTTGTCCTGCAGG |
| **4** | **actinB-273** | GGTACTTCAGGGTGAGGATGCCTGCTTGTCCTGCAGG |
| **5** | **actinB-322** | CAGATTTTCTCCATGTCGTCCCTGCTTGTCCTGCAGG |
| **6** | **actinB-352** | ACACGCAGCTCATTGTAGAACCTGCTTGTCCTGCAGG |
| **7** | **actinB-435** | ACATGATCTGGGTCATCTTCCCTGCTTGTCCTGCAGG |
| **8** | **actinB-488** | GGATAGCACAGCCTGGATAGCCTGCTTGTCCTGCAGG |
| **9** | **actinB-524** | CATCACGATGCCAGTGGTACCCTGCTTGTCCTGCAGG |
| **10** | **actinB-565** | TCGTAGATGGGCACAGTGTGCCTGCTTGTCCTGCAGG |
| **11** | **actinB-639** | TCTTCATGAGGTAGTCAGTCCCTGCTTGTCCTGCAGG |
| **12** | **actinB-702** | TAATGTCACGCACGATTTCCCCTGCTTGTCCTGCAGG |
| **13** | **actinB-745** | ATCTCTTGCTCGAAGTCCAGCCTGCTTGTCCTGCAGG |
| **14** | **actinB-773** | CAGGGAGGAGCTGGAAGCAGCCTGCTTGTCCTGCAGG |
| **15** | **actinB-902** | GAAGGTAGTTTCGTGGATGCCCTGCTTGTCCTGCAGG |
| **16** | **actinB-924** | CGTCACACTTCATGATGGAGCCTGCTTGTCCTGCAGG |
| **17** | **actinB-946** | TACAGGTCTTTGCGGATGTCCCTGCTTGTCCTGCAGG |
| **18** | **actinB-993** | CAATGCCAGGGTACATGGTGCCTGCTTGTCCTGCAGG |
| **19** | **actinB-1045** | ATCTTCATTGTGCTGGGTGCCCTGCTTGTCCTGCAGG |
| **20** | **actinB-1067** | CTCAGGAGGAGCAATGATCTCCTGCTTGTCCTGCAGG |
| **21** | **actinB-1089** | CGATCCACACGGAGTACTTGCCTGCTTGTCCTGCAGG |
| **22** | **actinB-1185** | ATTTGCGGTGGACGATGGAGCCTGCTTGTCCTGCAGG |
| **23** | **actinB-1231** | GTCAAGAAAGGGTGTAACGCCCTGCTTGTCCTGCAGG |
| **24** | **actinB-1420** | GTGCAATCAAAGTCCTCGGCCCTGCTTGTCCTGCAGG |
| **25** | **actinB-1492** | CTTTTAGGATGGCAAGGGACCCTGCTTGTCCTGCAGG |
| **26** | **actinB-1516** | TCTCCTTAGAGAGAAGTGGGCCTGCTTGTCCTGCAGG |
| **27** | **actinB-1542** | GTGGACTTGGGAGAGGACTGCCTGCTTGTCCTGCAGG |
| **28** | **actinB-1719** | ACTCCCAGGGAGACCAAAAGCCTGCTTGTCCTGCAGG |
| **29** | **actinB-1761** | GTCTCAAGTCAGTGTACAGGCCTGCTTGTCCTGCAGG |

**Table 3.4** MALAT1 dual FISH probe set sequences.

| **Probe** | **Sequence** |
|---|---|
| **MALAT1_V2_1_176/177_3′3′_GG** | GCCCTTCTATTGGTATTAATTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |

| | |
|---|---|
| **MALAT1_V2_2_176/177_3′3′_GG** | CTCCCAATTAATCTTTCCATTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_3_176/177_3′3′_GG** | TCTCCAAATTGTTTCATCCTTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_4_176/177_3′3′_GG** | TTAGCTTTTTGTTTCCTAGCTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_5_176/177_3′3′_GG** | ATCTTCTCAAGCTTTACCTTTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_6_176/177_3′3′_GG** | ACTATATTTAAGGCCTTCCATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_7_176/177_3′3′_GG** | TACTTCCGTTACGAAAGTCCTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_8_176/177_3′3′_GG** | CTGGGTCAGCTGTCAATTAATTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_9_176/177_3′3′_GG** | TCAGTCCTAGCTTCATCAAATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_10_176/177_3′3′_GG** | AACAACATATTGCCGACCTCTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_11_176/177_3′3′_GG** | AGTCATTTGCCTTTAGGATTTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_12_176/177_3′3′_GG** | AACTGTAAACCTGTGGTGGTTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_13_176/177_3′3′_GG** | CTCAACGTGAGAACTGCTCTTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_14_176/177_3′3′_GG** | CTCTAACCCAGTTTGTCAATTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_15_176/177_3′3′_GG** | CCAAGGATAAAAGCAGCTCCTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_16_176/177_3′3′_GG** | TGAACCAAAGCTGCACTGTGTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_17_176/177_3′3′_GG** | ACTGCCAACTAATTGCCAATTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_18_176/177_3′3′_GG** | CCAGTGGCTCATATTTAACTTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_19_176/177_3′3′_GG** | ACTTTCCTTGCCCAAATTAATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_20_176/177_3′3′_GG** | CCCAATGGAGGTATGACATATTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_21_176/177_3′3′_GG** | CATGCAATACTGCAGATGCATTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_22_176/177_3′3′_GG** | AGGAGAAAGTGCCATGGTTGTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |

| | |
|---|---|
| **MALAT1_V2_23_176/177_3′3′_GG** | TTTCTCAATCCTGAAATCCCTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_24_176/177_3′3′_GG** | TGAAGTGTACTATCCCATCATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_25_176/177_3′3′_GG** | TCCTGATCTGGTCCATTAAATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_26_176/177_3′3′_GG** | TCTTTCCTGCCTTAAAGTTATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_27_176/177_3′3′_GG** | TGTCAATTTATAGACCCCTGTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_28_176/177_3′3′_GG** | AAACATTGCCTACCACTCTATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_29_176/177_3′3′_GG** | CCTGAATGGCTTCATGAAGGTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_30_176/177_3′3′_GG** | TGCATTTACTTGCCAACAGATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_31_176/177_3′3′_GG** | GTCGTTTCACAATGCATTCTTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_32_176/177_3′3′_GG** | CAACACTCAGCCTTTATCACTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_33_176/177_3′3′_GG** | TTTTTTCTTACTGGGTCTGGTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_34_176/177_3′3′_GG** | ACTGGAAGCTCCTTCTATAGTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_35_176/177_3′3′_GG** | TCATTTTGTCCACTGGTGAATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_36_176/177_3′3′_GG** | TTGTCCCATAACTGATCTGATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_37_176/177_3′3′_GG** | AACACAGTTTGCTCACATGCTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_38_176/177_3′3′_GG** | TGACACTTCTCTTGACCTTATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_39_176/177_3′3′_GG** | CACTCCAGAAAGAGGGAGTTTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_40_176/177_3′3′_GG** | AGATCAAAAGGCACGGGGTGTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_41_176/177_3′3′_GG** | ACAAGGATCCAAGCTACTGGTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_42_176/177_3′3′_GG** | CATCGTTACCTTGAAACCGATTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_43_176/177_3′3′_GG** | TAACATAGTTCAACCCACCATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |

| | |
|---|---|
| **MALAT1_V2_44_176/177_3′3′_GG** | TTGCAGGCAAATTAATGGCCTTTCTGGATTGTCCT TCCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_45_176/177_3′3′_GG** | TGTGGTTTTAATACCCTTCTTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_46_176/177_3′3′_GG** | TTTCTCCACTTACTGGTTTATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_47_176/177_3′3′_GG** | CTATCTTCTCCAGTCTACAATTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |
| **MALAT1_V2_48_176/177_3′3′_GG** | CAAGTCTGTTATGTTCACCTTTTCTGGATTGTCCTT CCAGTTTCACCCTTGTCCTGGGTG |

**Table 3.5** MALAT1 single FISH probe set sequences.

| Probe | Sequence |
|---|---|
| **MALAT1_V2_1_176/177_3′3′_GC** | GCCCTTCTATTGGTATTAATTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_2_176/177_3′3′_GC** | CTCCCAATTAATCTTTCCATTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_3_176/177_3′3′_GC** | TCTCCAAATTGTTTCATCCTTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_4_176/177_3′3′_GC** | TTAGCTTTTTGTTTCCTAGCTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_5_176/177_3′3′_GC** | ATCTTCTCAAGCTTTACCTTTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_6_176/177_3′3′_GC** | ACTATATTTAAGGCCTTCCATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_7_176/177_3′3′_GC** | TACTTCCGTTACGAAAGTCCTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_8_176/177_3′3′_GC** | CTGGGTCAGCTGTCAATTAATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_9_176/177_3′3′_GC** | TCAGTCCTAGCTTCATCAAATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_10_176/177_3′3′_GC** | AACAACATATTGCCGACCTCTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_11_176/177_3′3′_GC** | AGTCATTTGCCTTTAGGATTTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_12_176/177_3′3′_GC** | AACTGTAAACCTGTGGTGGTTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_13_176/177_3′3′_GC** | CTCAACGTGAGAACTGCTCTTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_14_176/177_3′3′_GC** | CTCTAACCCAGTTTGTCAATTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_15_176/177_3′3′_GC** | CCAAGGATAAAAGCAGCTCCTTTCTGGATTGTCCT TCCAGTTTCACCCTTCTCCTGGGTG |

| | |
|---|---|
| **MALAT1_V2_16_176/177_3′3′_GC** | TGAACCAAAGCTGCACTGTGTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_17_176/177_3′3′_GC** | ACTGCCAACTAATTGCCAATTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_18_176/177_3′3′_GC** | CCAGTGGCTCATATTTAACTTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_19_176/177_3′3′_GC** | ACTTTCCTTGCCCAAATTAATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_20_176/177_3′3′_GC** | CCCAATGGAGGTATGACATATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_21_176/177_3′3′_GC** | CATGCAATACTGCAGATGCATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_22_176/177_3′3′_GC** | AGGAGAAAGTGCCATGGTTGTTTCTGGATTGTCCT TCCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_23_176/177_3′3′_GC** | TTTCTCAATCCTGAAATCCCTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_24_176/177_3′3′_GC** | TGAAGTGTACTATCCCATCATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_25_176/177_3′3′_GC** | TCCTGATCTGGTCCATTAAATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_26_176/177_3′3′_GC** | TCTTTCCTGCCTTAAAGTTATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_27_176/177_3′3′_GC** | TGTCAATTTATAGACCCCTGTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_28_176/177_3′3′_GC** | AAACATTGCCTACCACTCTATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_29_176/177_3′3′_GC** | CCTGAATGGCTTCATGAAGGTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_30_176/177_3′3′_GC** | TGCATTTACTTGCCAACAGATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_31_176/177_3′3′_GC** | GTCGTTTCACAATGCATTCTTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_32_176/177_3′3′_GC** | CAACACTCAGCCTTTATCACTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_33_176/177_3′3′_GC** | TTTTTTCTTACTGGGTCTGGTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_34_176/177_3′3′_GC** | ACTGGAAGCTCCTTCTATAGTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_35_176/177_3′3′_GC** | TCATTTTGTCCACTGGTGAATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_36_176/177_3′3′_GC** | TTGTCCCATAACTGATCTGATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |

| | |
|---|---|
| **MALAT1_V2_37_176/177_3′3′_GC** | AACACAGTTTGCTCACATGCTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_38_176/177_3′3′_GC** | TGACACTTCTCTTGACCTTATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_39_176/177_3′3′_GC** | CACTCCAGAAAGAGGGAGTTTTTCTGGATTGTCCT TCCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_40_176/177_3′3′_GC** | AGATCAAAAGGCACGGGGTGTTTCTGGATTGTCCT TCCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_41_176/177_3′3′_GC** | ACAAGGATCCAAGCTACTGGTTTCTGGATTGTCCT TCCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_42_176/177_3′3′_GC** | CATCGTTACCTTGAAACCGATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_43_176/177_3′3′_GC** | TAACATAGTTCAACCCACCATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_44_176/177_3′3′_GC** | TTGCAGGCAAATTAATGGCCTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_45_176/177_3′3′_GC** | TGTGGTTTTAATACCCTTCTTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_46_176/177_3′3′_GC** | TTTCTCCACTTACTGGTTTATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_47_176/177_3′3′_GC** | CTATCTTCTCCAGTCTACAATTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |
| **MALAT1_V2_48_176/177_3′3′_GC** | CAAGTCTGTTATGTTCACCTTTTCTGGATTGTCCTT CCAGTTTCACCCTTCTCCTGGGTG |

# Appendix for Chapter 4: RNA Sequences and PIN Construct Plasmid Sequences

## RNA Sequences

**Table 4.3** RNA construct sequences used for SNAP-tag conjugation.

| construct | sequence |
|---|---|
| 7skT3 | GGAUCCCCGGGAGCAGACUGUAAAUCUGCUCCCACCACCGGAUGU<br>GAGGGCGAUCUGGCUGCGACAUCUGUCACCCCAUUGAUCGCCAG<br>GGUUGAUUCGGCUGAUCUGGCUGGCUAGGCGGGUGUCCCCUUCC<br>UCCCUCACCGCUCCAUGUGCGUCCCUCCCGAAGCUGCGCGCUCG<br>GUCGAAGAGGACGACCAUCCCCGAUAGAGGAGGACCGGUCUUCGG<br>UCAAGGGUAUACGAGUAGCUGCGCUCCCCUGCUAGAACCUCCAAA<br>CAAGCUCUCAAGGUCCAUUUGUAGGAGAACGUAGGGUAGUCAAGC<br>UUCCAAGACUCCAGACACAUCCAAAUGAGGCGCUGCAUGUGGCAG<br>UCUGCCUUUCUUUUU |
| HDAC T3 | GAGACCCAAGCUGGCUAGCGUUUAAACUUAAGCUUGGUACCGAGC<br>UCGGAUCCUGACAGUCUCACCAAUUUCAGAAAAUCAUUAAAAAGAA<br>AAUAUUGAAAGGAAAAUGUUUUCUUUUUGAAGACUUCUGGCUUCA<br>UUUUAUACUACUUUGGCAUGGACUGUAUUUAUUUUCAAAUGGCUU<br>UUUCGUUUUUGUUUUUCUUGGCAAGUUUUAUUGUGAGUUUUUCUA<br>AUUAUGAAGCAAAAUUUCUUUUCUCCACCAUGCUUUAUGUGAUAG<br>UAUUUAAAAUUGAUGUGAGUUAUUAUGUCAAAAAAACUGAUCUAUU<br>AAAGAAGUAAUUGGCCUUUCUGAGCUGAUAUCUAACACAGACUCU<br>CGGUACCAUCAUUUUCAUAUCCCCGGGAGCAGACUGUAAAUCUGC<br>UCCCACCACCAUCAUUUAAUGAAUUCCAUCAGGAAUCCCUCACUUC<br>UGCAGACUGGCCGUCGUUUUACACUCGAGU |
| dummy T3 | GAGACCCAAGCUGGCUAGCGUUUAAACUUAAGCUUGGUACCGAGC<br>UCGGAUCCUCUAACACAGACUCUCGGUACCAUCAUUUUCAUAUCC<br>CCGGGAGCAGACUGUAAAUCUGCUCCCACCACCAUCAUUUAAUGA<br>AUUCCAUCAGGAAUCCCUCACUUCUGCAGACUGGCCGUCGUUUUA<br>CACUCGAGU |
| mChT3 | GAGACCCAAGCUUGGUACCGAGCUCGGAUCCACUAGUAACGGCCG<br>CCAGUGUGCUGGAAUUCUGCAGAUAUCGUUGACCUUGCAGAAGGA<br>GAUAUAAUGGUGAGCAAGGGCGAGGAGGAUAACAUGGCCAUCAUC<br>AAGGAGUUCAUGCGCUUCAAGGUGCACAUGGAGGGCUCCGUGAAC<br>GGCCACGAGUUCGAGAUCGAGGGCGAGGGCGAGGGCCGCCCCUA<br>CGAGGGCACCCAGACCGCCAAGCUGAAGGUGACCAAGGGUGGCCC<br>CCUGCCCUUCGCCUGGGACAUCCUGUCCCCUCAGUUCAUGUACGG<br>CUCCAAGGCCUACGUGAAGCACCCCGCCGACAUCCCCGACUACUU |

| | |
|---|---|
| | GAAGCUGUCCUUCCCCGAGGGCUUCAAGUGGGAGCGCGUGAUGA<br>ACUUCGAGGACGGCGGCGUGGUGACCGUGACCCAGGACUCCUCC<br>CUGCAAGACGGCGAGUUCAUCUACAAGGUGAAGCUGCGCGGCACC<br>AACUUCCCCUCCGACGGCCCCGUAAUGCAGAAGAAGACGAUGGGC<br>UGGGAGGCCUCCUCCGAGCGGAUGUACCCCGAGGACGGCGCCCU<br>GAAGGGCGAGAUCAAGCAGAGGCUGAAGCUGAAGGACGGCGGCCA<br>CUACGACGCUGAGGUCAAGACCACCUACAAGGCCAAGAAGCCCGU<br>GCAGCUGCCCGGCGCCUACAACGUGAACAUCAAGUUGGACAUCAC<br>CUCCCACAACGAGGACUACACCAUCGUGGAACAGUACGAACGCGC<br>CGAGGGCCGCCACUCCACCGGCGGCAUGGACGAGCUGUACAAGUA<br>AUAUCUAACACAGACUCUCGGUACCAUCAUUUUCAUAUCCCCGGG<br>AGCAGACUGUAAAUCUGCUCCCACCACCAUCAUUUAAUGAAUUCCA<br>UCAGGAAUCCCUCACUUAAAGCCCGCCGAAAGGCGGGCUUUUCUG<br>UGUCUGCAGACUGGCCGUCGUUUUACACUCGAGCAUGCAU |

**PIN Construct Plasmid Sequences**

*Sleeping beauty sequences*

**SNAP-PIN:**

AATTTTAACAAAATATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCA

TCTGTGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGGGAAATGTGCGCGG

AACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAA

TAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACAT

TTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC

CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGG

GTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAA

GAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCC

CGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATG

ACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTA

AGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACT

TCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGG

GATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA

CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTA

TTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGA

GGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTT

ATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCAC

TGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCA

GGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTA

AGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACT

TCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAA

ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAA

AGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA

ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC

CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAG

CCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCT

GCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGG

TTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG

GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACC

TACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACA

GGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAG

GGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG

CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCA

ACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTC

CTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGAT

ACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC

GGATCCCTATACAGTTGAAGTCGGAAGTTTACATACACCTTAGCCAAATACATTTAA

ACTCACTTTTTCACAATTCCTGACATTTAATCCTAGTAAAAATTCCCTGTCTTAGGTC

AGTTAGGATCACCACTTTATTTTAAGAATGTGAAATATCAGAATAATAGTAGAGAGA

ATGATTCATTTCAGCTTTTATTTCTTTCATCACATTCCCAGTGGGTCAGAAGTTTACA

TACACTCGACTCCTCTGCAGAATGCGGCGATGTTTCGGTAAGGGGTCCCGAGTAG

CTAGTTCATGGCAGCCAGCATATGGCATATGTTGCCAAACTCTAAACCAAATACTC

ATTCTGATGTTTTAAATGATTTGCCCTCCCATATGTCCTTCCGAGTGAGAGACACAA

AAAATTCCAACACACTATTGCAATGAAAATAAATTTCCTTTATTAGCCAGAAGTCAG

ATGCTCAAGGGGCTTCATGATGTCCCCATAATTTTTGGCAGAGGGAAAAAGATCTC

AGTGGTATTTGTGAGCCAGGGCATTGGCCACACCAGCCACCACCTTCTGATAGGC

AGCCTGCACCTGAGGAGTGAATTCACGACAGGCCTTCGAATCAGGCACCGGGCTT

GCGGGTCATGCACCAGGTGCGCGGTCCTTCGGGCACCTCGACGTCGGCGGTGAC

GGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCA

GGAAGGCGGGCACCCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACG

GCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGGCCAG

GAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTG

CTGCGCGGCCAGCCGGGAACCGCTCAACTCGGCCATGCGCGGGCCGATCTCGG

CGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGG

CGCCGTCGTCCGCGACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGA

AGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGACGGTGT

GGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTG

GGGACGTCGTCGCGGGTGGCGAGGCGCACCGTGGGCTTGTACTCGGTCATGGTG

GCCTCAGGTGCAGAGGTTTCTACAGGGAAAGAGAGAAAGGCTACAGTTAGTTTGC

TGTGATGATGCTAAGTGCCGAGGTCAGATGTCTATGGATTGTACACGGAATCCCAA

CGCAGGAACCAAAAAACACTCACCCAAAAAAATGGCGCTAAGGCCGAGAGAAAGG

CGAAGTTCCGTCTACGGCTATATACCTTGAAACAGGGGCGGCAGAAGTGACGCAA

CGCGGTTGCGCGGACGGCTTCCTGTGTTTTCATTGGTCGAAGCGTTCTGCGCGTG

CGCAGAGATTCCCCAGGGAAGGCGGGCCCTCTTCCGAATGTCCAATCCGGACATT

CTTCCGGTTGGACCTCCGCGGTCAGCGCTGTGCCCGGCGAGAGATCACGTGGGG

CGCGGAGGCGGTGCTGCTGGGGCACGGCCGTCCAGCCTCGGCGGCCATATTTTT

GAGGGGCTGTTCATCTCGTTCACACGCTCTGTCCGCCATGTTTGTGAGTGGAAGC

GCCATTACCCCTTCAAGCGACTGAAGGCTGCAGGGCCTCTGGTGGCCCGCATGG

GGAGACCAGACCCGCCAGGCCCGCCTTTCCGCACTCAGTCCGGGCTTACTTTATT

TTGTGAGACAGGGTCTCGCCTAGAGGCTCCGGTGCCCGTCAGTGGGCAGAGCGC

ACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGT

GCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTC

CGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGA

ACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGG

TTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTT

CCACGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTG

GGTGGGAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGT

TGAGGCTTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCG

CGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGC

TGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACA

CTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGC

GCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGG

GGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGT

ATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGC

GGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGACGCG

GCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCC

GTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCGTCCAGGCA

CCTCGATTAGTTCTCGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGGGGGAGGGG

TTTTATGCGATGGAGTTTCCCCATACTGAGTGGGTGGAGACTGAAGTTAGGCCAGC

TTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTTGGATCTTGGTT

CATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTAAGGTGTCGTG

AAAACTACCCCAAGCTGGCCTCTGAGGCCACCATGGACAAAGACTGCGAAATGAA

GCGCACCACCCTGGATAGCCCTCTGGGCAAGCTGGAACTGTCTGGGTGCGAACA

GGGCCTGCACCGTATCATCTTCCTGGGCAAAGGAACATCTGCCGCCGACGCCGTG

GAAGTGCCTGCCCCAGCCGCCGTGCTGGGCGGACCAGAGCCACTGATGCAGGCC

ACCGCCTGGCTCAACGCCTACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCCTG

TGCCAGCCCTGCACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCGCCAGGTGCT

GTGGAAACTGCTGAAAGTGGTGAAGTTCGGAGAGGTCATCAGCTACAGCCACCTG

GCCGCCCTGGCCGGCAATCCCGCCGCCACCGCCGCCGTGAAAACCGCCCTGAGC

GGAAATCCCGTGCCCATTCTGATCCCCTGCCACCGGGTGGTGCAGGGCGACCTG

GACGTGGGGGGCTACGAGGGCGGGCTCGCCGTGAAAGAGTGGCTGCTGGCCCA

CGAGGGCCACAGACTGGGCAAGCCTGGGCTGGGTCCTGCAGGCGGAGGCGCGC

CAGGGTCTGGCGGCGGCAGTGAGCTCGAAATCAGGCCGCTGTTCCTCGTGCCGG

ACACTAATGGTTTTATAGATCACTTGGCGTCCTTGGCTAGACTTCTGGAAAGCCGA

AAGTATATATTGGTAGTGCCGTTGATTGTAATTAACGAATTGGATGGGTTGGCGAA

AGGACAAGAGACTGATCACAGAGCAGGAGGCTACGCGAGGGTCGTCCAAGAGAA

GGCGCGAAAAAGCATCGAGTTCCTGGAGCAGCGATTTGAGAGCAGGGACTCATGC

CTGAGAGCCCTCACGTCACGGGGGAACGAGCTGGAGTCCATCGCTTTCCGAAGT

GAAGACATTACGGGCCAACTTGGGAATAATGATGACCTCATCTTGTCCTGCTGCCT

GCACTACTGCAAGGACAAGGCTAAGGACTTCATGCCTGCCTCCAAGGAGGAGCCT

ATCCGATTGTTGAGGGAAGTAGTACTTTTGACGGACGACCGCAACCTCCGGGTAA

AGGCGCTGACTCGAAATGTCCCAGTAAGGGATATACCGGCGTTCCTTACATGGGC

TCAAGTAGGGGATTATAAGGATCATGATGGGGATTACAAGGACCATGACATCGATT

ATAAAGATGACGACGATAAGTAGGGCCTGTCAGGCCAAGCTTGGGGATCAATTCT

CTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTT

TGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTT

CCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGG

GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGG

CATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGG

GCCTACTAGATAGCGGACCCCTTACCGAAACATCGCCGCATTCTGCAGAGGAGTC

GAACTTAGTGTATGTAAACTTCTGACCCACTGGAATTGTGATACAGTGAATTATAAG

TGAAATAATCTGTCTGTAAACAATTGTTGGAAAAATGACTTGTGTCATGCACAAAGT

AGATGTCCTAACTGACTTGCCAAAACTATTGTTTGTTAACAAGAAATTTGTGGAGTA

GTTGAAAAACGAGTTTTAATGACTCCAACTTAAGTGTATGTAAACTTCCGACTTCAA

CTGTATAGGGATCC

**PIN-SNAP:**

AATTTTAACAAAATATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCA

TCTGTGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGGGAAATGTGCGCGG

AACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAA

TAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACAT

TTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC

CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGG

GTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAA

GAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCC

CGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATG

ACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTA

AGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACT

TCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGG

GATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA

CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTA

TTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGA

GGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTT

ATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCAC

TGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCA

GGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTA

AGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACT

TCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAA

ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAA

AGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA

ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC

CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAG

CCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCT

GCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGG

TTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG

GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACC

TACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACA

GGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAG

GGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG

CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCA

ACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTC

CTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGAT

ACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC

GGATCCCTATACAGTTGAAGTCGGAAGTTTACATACACCTTAGCCAAATACATTTAA

ACTCACTTTTTCACAATTCCTGACATTTAATCCTAGTAAAAATTCCCTGTCTTAGGTC

AGTTAGGATCACCACTTTATTTTAAGAATGTGAAATATCAGAATAATAGTAGAGAGA

ATGATTCATTTCAGCTTTTATTTCTTTCATCACATTCCCAGTGGGTCAGAAGTTTACA

TACACTCGACTCCTCTGCAGAATGCGGCGATGTTTCGGTAAGGGGTCCCGAGTAG

CTAGTTCATGGCAGCCAGCATATGGCATATGTTGCCAAACTCTAAACCAAATACTC

ATTCTGATGTTTTAAATGATTTGCCCTCCCATATGTCCTTCCGAGTGAGAGACACAA

AAAATTCCAACACACTATTGCAATGAAAATAAATTTCCTTTATTAGCCAGAAGTCAG

ATGCTCAAGGGGCTTCATGATGTCCCCATAATTTTTGGCAGAGGGAAAAAGATCTC

AGTGGTATTTGTGAGCCAGGGCATTGGCCACACCAGCCACCACCTTCTGATAGGC

AGCCTGCACCTGAGGAGTGAATTCACGACAGGCCTTCGAATCAGGCACCGGGCTT

GCGGGTCATGCACCAGGTGCGCGGTCCTTCGGGCACCTCGACGTCGGCGGTGAC

GGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCA

GGAAGGCGGGCACCCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACG

GCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGGCCAG

GAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTG

CTGCGCGGCCAGCCGGGAACCGCTCAACTCGGCCATGCGCGGGCCGATCTCGG

CGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGG

CGCCGTCGTCCGCGACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGA

AGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGACGGTGT

GGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTG

GGGACGTCGTCGCGGGTGGCGAGGCGCACCGTGGGCTTGTACTCGGTCATGGTG

GCCTCAGGTGCAGAGGTTTCTACAGGGAAAGAGAGAAAGGCTACAGTTAGTTTGC

TGTGATGATGCTAAGTGCCGAGGTCAGATGTCTATGGATTGTACACGGAATCCCAA

CGCAGGAACCAAAAAACACTCACCCAAAAAAATGGCGCTAAGGCCGAGAGAAAGG

CGAAGTTCCGTCTACGGCTATATACCTTGAAACAGGGGCGGCAGAAGTGACGCAA

CGCGGTTGCGCGGACGGCTTCCTGTGTTTTCATTGGTCGAAGCGTTCTGCGCGTG

CGCAGAGATTCCCCAGGGAAGGCGGGCCCTCTTCCGAATGTCCAATCCGGACATT

CTTCCGGTTGGACCTCCGCGGTCAGCGCTGTGCCCGGCGAGAGATCACGTGGGG

CGCGGAGGCGGTGCTGCTGGGGCACGGCCGTCCAGCCTCGGCGGCCATATTTTT

GAGGGGCTGTTCATCTCGTTCACACGCTCTGTCCGCCATGTTTGTGAGTGGAAGC

GCCATTACCCCTTCAAGCGACTGAAGGCTGCAGGGCCTCTGGTGGCCCGCATGG

GGAGACCAGACCCGCCAGGCCCGCCTTTCCGCACTCAGTCCGGGCTTACTTTATT

TTGTGAGACAGGGTCTCGCCTAGAGGCTCCGGTGCCCGTCAGTGGGCAGAGCGC

ACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGT

GCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTC

CGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGA

ACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGG

TTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTT

CCACGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTG

GGTGGGAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGT

TGAGGCTTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCG

CGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGC

TGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACA

CTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGC

GCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGG

GGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGT

ATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGC

GGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGACGCG

GCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCC

GTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCGTCCAGGCA

CCTCGATTAGTTCTCGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGGGGGAGGGG

TTTTATGCGATGGAGTTTCCCCATACTGAGTGGGTGGAGACTGAAGTTAGGCCAGC

TTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTTGGATCTTGGTT

CATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTAAGGTGTCGTG

AAAACTACCCCAAGCTGGCCTCTGAGGCCACCATGGAGCTCGAAATCAGGCCGCT

GTTCCTCGTGCCGGACACTAATGGTTTTATAGATCACTTGGCGTCCTTGGCTAGAC

TTCTGGAAAGCCGAAAGTATATATTGGTAGTGCCGTTGATTGTAATTAACGAATTGG

ATGGGTTGGCGAAAGGACAAGAGACTGATCACAGAGCAGGAGGCTACGCGAGGG

TCGTCCAAGAGAAGGCGCGAAAAAGCATCGAGTTCCTGGAGCAGCGATTTGAGAG

CAGGGACTCATGCCTGAGAGCCCTCACGTCACGGGGGAACGAGCTGGAGTCCAT

CGCTTTCCGAAGTGAAGACATTACGGGCCAACTTGGGAATAATGATGACCTCATCT

TGTCCTGCTGCCTGCACTACTGCAAGGACAAGGCTAAGGACTTCATGCCTGCCTC

CAAGGAGGAGCCTATCCGATTGTTGAGGGAAGTAGTACTTTTGACGGACGACCGC

AACCTCCGGGTAAAGGCGCTGACTCGAAATGTCCCAGTAAGGGATATACCGGCGT

TCCTTACATGGGCTCAAGTAGGGAGTGGAAGTGAGACACCGGGAACCTCAGAGAG

CGCCACGCCAGAAAGCGACAAAGACTGCGAAATGAAGCGCACCACCCTGGATAG

CCCTCTGGGCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCTGCACCGTATCATC

TTCCTGGGCAAAGGAACATCTGCCGCCGACGCCGTGGAAGTGCCTGCCCCAGCC

GCCGTGCTGGGCGGACCAGAGCCACTGATGCAGGCCACCGCCTGGCTCAACGCC

TACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCCTGTGCCAGCCCTGCACCACC

CAGTGTTCCAGCAGGAGAGCTTTACCCGCCAGGTGCTGTGGAAACTGCTGAAAGT

GGTGAAGTTCGGAGAGGTCATCAGCTACAGCCACCTGGCCGCCCTGGCCGGCAA

TCCCGCCGCCACCGCCGCCGTGAAAACCGCCCTGAGCGGAAATCCCGTGCCCAT

TCTGATCCCCTGCCACCGGGTGGTGCAGGGCGACCTGGACGTGGGGGGCTACGA

GGGCGGGCTCGCCGTGAAAGAGTGGCTGCTGGCCCACGAGGGCCACAGACTGG

GCAAGCCTGGGCTGGGTGATTATAAGGATCATGATGGGGATTACAAGGACCATGA

CATCGATTATAAAGATGACGACGATAAGTGAGGCCTGTCAGGCCAAGCTTGGGGA

TCAATTCTCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCAT

CTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACT

GTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCT

ATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAAT

AGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACC

AGCTGGGCCTACTAGATAGCGGACCCCTTACCGAAACATCGCCGCATTCTGCAGA

GGAGTCGAACTTAGTGTATGTAAACTTCTGACCCACTGGAATTGTGATACAGTGAA

TTATAAGTGAAATAATCTGTCTGTAAACAATTGTTGGAAAAATGACTTGTGTCATGC

ACAAAGTAGATGTCCTAACTGACTTGCCAAAACTATTGTTTGTTAACAAGAAATTTG

TGGAGTAGTTGAAAAACGAGTTTTAATGACTCCAACTTAAGTGTATGTAAACTTCCG

ACTTCAACTGTATAGGGATCC

**PIN:**

AATTTTAACAAAATATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCA

TCTGTGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGGGAAATGTGCGCGG

AACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAA

TAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACAT

TTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC

CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGG

GTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAA

GAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCC

CGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATG

ACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTA

AGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACT

TCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGG

GATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA

CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTA

TTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGA

GGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTT

ATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCAC

TGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCA

GGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTA

AGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACT

TCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAA

ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAA

AGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA

ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC

CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAG

CCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCT

GCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGG

TTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG

GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACC

TACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACA

GGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAG

GGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG

CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCA

ACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTC

CTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGAT

ACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC

GGATCCCTATACAGTTGAAGTCGGAAGTTTACATACACCTTAGCCAAATACATTTAA

ACTCACTTTTTCACAATTCCTGACATTTAATCCTAGTAAAAATTCCCTGTCTTAGGTC

AGTTAGGATCACCACTTTATTTTAAGAATGTGAAATATCAGAATAATAGTAGAGAGA

ATGATTCATTTCAGCTTTTATTTCTTTCATCACATTCCCAGTGGGTCAGAAGTTTACA

TACACTCGACTCCTCTGCAGAATGCGGCGATGTTTCGGTAAGGGGTCCCGAGTAG

CTAGTTCATGGCAGCCAGCATATGGCATATGTTGCCAAACTCTAAACCAAATACTC

ATTCTGATGTTTTAAATGATTTGCCCTCCCATATGTCCTTCCGAGTGAGAGACACAA

AAAATTCCAACACACTATTGCAATGAAAATAAATTTCCTTTATTAGCCAGAAGTCAG

ATGCTCAAGGGGCTTCATGATGTCCCCATAATTTTTGGCAGAGGGAAAAAGATCTC

AGTGGTATTTGTGAGCCAGGGCATTGGCCACACCAGCCACCACCTTCTGATAGGC

AGCCTGCACCTGAGGAGTGAATTCACGACAGGCCTTCGAATCAGGCACCGGGCTT

GCGGGTCATGCACCAGGTGCGCGGTCCTTCGGGCACCTCGACGTCGGCGGTGAC

GGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCA

GGAAGGCGGGCACCCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACG

GCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGGCCAG

GAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTG

CTGCGCGGCCAGCCGGGAACCGCTCAACTCGGCCATGCGCGGGCCGATCTCGG

CGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGG

CGCCGTCGTCCGCGACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGA

AGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGACGGTGT

GGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTG

GGGACGTCGTCGCGGGTGGCGAGGCGCACCGTGGGCTTGTACTCGGTCATGGTG

GCCTCAGGTGCAGAGGTTTCTACAGGGAAAGAGAGAAAGGCTACAGTTAGTTTGC

TGTGATGATGCTAAGTGCCGAGGTCAGATGTCTATGGATTGTACACGGAATCCCAA

CGCAGGAACCAAAAAACACTCACCCAAAAAAATGGCGCTAAGGCCGAGAGAAAGG

CGAAGTTCCGTCTACGGCTATATACCTTGAAACAGGGGCGGCAGAAGTGACGCAA

CGCGGTTGCGCGGACGGCTTCCTGTGTTTTCATTGGTCGAAGCGTTCTGCGCGTG

CGCAGAGATTCCCCAGGGAAGGCGGGCCCTCTTCCGAATGTCCAATCCGGACATT

CTTCCGGTTGGACCTCCGCGGTCAGCGCTGTGCCCGGCGAGAGATCACGTGGGG

CGCGGAGGCGGTGCTGCTGGGGCACGGCCGTCCAGCCTCGGCGGCCATATTTTT

GAGGGGCTGTTCATCTCGTTCACACGCTCTGTCCGCCATGTTTGTGAGTGGAAGC

GCCATTACCCCTTCAAGCGACTGAAGGCTGCAGGGCCTCTGGTGGCCCGCATGG

GGAGACCAGACCCGCCAGGCCCGCCTTTCCGCACTCAGTCCGGGCTTACTTTATT

TTGTGAGACAGGGTCTCGCCTAGAGGCTCCGGTGCCCGTCAGTGGGCAGAGCGC

ACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGT

GCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTC

CGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGA

ACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGG

TTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTT

CCACGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTG

GGTGGGAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGT

TGAGGCTTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCG

CGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGC

TGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACA

CTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGC

GCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGG

GGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGT

ATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGC

GGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGACGCG

GCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCC

GTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCGTCCAGGCA

CCTCGATTAGTTCTCGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGGGGGAGGGG

TTTTATGCGATGGAGTTTCCCCATACTGAGTGGGTGGAGACTGAAGTTAGGCCAGC

TTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTTGGATCTTGGTT

CATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTAAGGTGTCGTG

AAAACTACCCCAAGCTGGCCTCTGAGGCCACCATGGAGCTCGAAATCAGGCCGCT

GTTCCTCGTGCCGGACACTAATGGTTTTATAGATCACTTGGCGTCCTTGGCTAGAC

TTCTGGAAAGCCGAAAGTATATATTGGTAGTGCCGTTGATTGTAATTAACGAATTGG

ATGGGTTGGCGAAAGGACAAGAGACTGATCACAGAGCAGGAGGCTACGCGAGGG

TCGTCCAAGAGAAGGCGCGAAAAAGCATCGAGTTCCTGGAGCAGCGATTTGAGAG

CAGGGACTCATGCCTGAGAGCCCTCACGTCACGGGGGAACGAGCTGGAGTCCAT

CGCTTTCCGAAGTGAAGACATTACGGGCCAACTTGGGAATAATGATGACCTCATCT

TGTCCTGCTGCCTGCACTACTGCAAGGACAAGGCTAAGGACTTCATGCCTGCCTC

CAAGGAGGAGCCTATCCGATTGTTGAGGGAAGTAGTACTTTTGACGGACGACCGC

AACCTCCGGGTAAAGGCGCTGACTCGAAATGTCCCAGTAAGGGATATACCGGCGT

TCCTTACATGGGCTCAAGTAGGGGATTATAAGGATCATGATGGGGATTACAAGGAC

CATGACATCGATTATAAAGATGACGACGATAAGTGAGGCCTGTCAGGCCAAGCTTG

GGGATCAATTCTCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAG

CCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCC

CACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCA

TTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGA

CAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAG

AACCAGCTGGGCCTACTAGATAGCGGACCCCTTACCGAAACATCGCCGCATTCTG

CAGAGGAGTCGAACTTAGTGTATGTAAACTTCTGACCCACTGGAATTGTGATACAG

TGAATTATAAGTGAAATAATCTGTCTGTAAACAATTGTTGGAAAAATGACTTGTGTC

ATGCACAAAGTAGATGTCCTAACTGACTTGCCAAAACTATTGTTTGTTAACAAGAAA

TTTGTGGAGTAGTTGAAAAACGAGTTTTAATGACTCCAACTTAAGTGTATGTAAACT

TCCGACTTCAACTGTATAGGGATCC

*Bacterial Expression Sequences*

**SNAP-PIN:**

AACTACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTAT

TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCT

TCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTAT

TCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAA

AGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGAT

CTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTCCCAATGAT

GAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGC

AAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCA

CCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGC

TGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAG

GACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTT

GATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCA

CGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTT

ACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAG

GACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA

GCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAG

CCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAAC

GAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCA

GACCAAGTTTACTCATATATACTTTAGATTGATTTACCCCGGTTGATAATCAGAAAA

GCCCCAAAAACAGGAAGATTGTATAAGCAAATATTTAAATTGTAAACGTTAATATTTT

GTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAA

ATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGT

TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGC

GAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAG

TTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCC

CGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAG

AAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGC

GTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTAAAAGGATC

TAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCG

TTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTT

TTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGG

TTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGC

AGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTT

CAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG

CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTT

ACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAG

CTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAA

AGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGG

GTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTT

TATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTC

GTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTT

CCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTC

TGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGA

ACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCTATGGTGCACTCTCAGTAC

AATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGA

CTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG

GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGC

TGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGT

AAAGCTCATCAGCGTGGTCGTGCAGCGATTCACAGATGTCTGCCTGTTCATCCGC

GTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGG

CCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGA

TTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACG

GGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTG

GCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCCG

AACGCCAGCAAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCGATAATGGCC

TGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGG

GCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGC

GAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGA

GTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGC

CCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCC

CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTT

TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG

GAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGAC

GGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCG

GTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGC

GGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGC

ACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTG

ATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATG

GTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCT

GAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGA

GACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACC

AGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGAT

GGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCT

TCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGAC

GCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGT

TCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCG

CCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAA

TCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTC

AGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGG

CCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGAC

ATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGC

GCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTC

GACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAG

GCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCCGGCATGCCGCCCTTTC

GTCTTCAAGAATTAATTCCCAATTCCCCAGGCATCAAATAAAACGAAAGGCTCAGT

CGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGT

AGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGG

TGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAATTCCCCAGGCATCAAA

TAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCG

GTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGA

AGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATT

AATTCCCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGT

TTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGC

GGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGC

CATAAACTGCCAGGAATTAATTCCCCAGGCATCAAATAAAACGAAAGGCTCAGTCG

AAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAG

GACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTG

GCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAATTCCCCAGGCATCAAATA

AAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGT

GAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAG

CAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTGG

GGATCGGAATTAATTCCCGGTTTAAACCGGGGATCTCGATCCCGCGAAATTAATAC

GACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGT

TTAACTTTAAGAAGGAGATATACATATGGACAAAGACTGCGAAATGAAGCGCACCA

CCCTGGATAGCCCTCTGGGCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCTGC

ACCGTATCATCTTCCTGGGCAAAGGAACATCTGCCGCCGACGCCGTGGAAGTGCC

TGCCCCAGCCGCCGTGCTGGGCGGACCAGAGCCACTGATGCAGGCCACCGCCTG

GCTCAACGCCTACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCCTGTGCCAGCC

CTGCACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCGCCAGGTGCTGTGGAAAC

TGCTGAAAGTGGTGAAGTTCGGAGAGGTCATCAGCTACAGCCACCTGGCCGCCCT

GGCCGGCAATCCCGCCGCCACCGCCGCCGTGAAAACCGCCCTGAGCGGAAATCC

CGTGCCCATTCTGATCCCCTGCCACCGGGTGGTGCAGGGCGACCTGGACGTGGG

GGGCTACGAGGGCGGGCTCGCCGTGAAAGAGTGGCTGCTGGCCCACGAGGGCC

ACAGACTGGGCAAGCCTGGGCTGGGTCCTGCAGGCGGAGGCGCGCCAGGGTCT

GGCGGCGGCAGTGAGCTCGAAATCAGGCCGCTGTTCCTCGTGCCGGACACTAAT

GGTTTTATAGATCACTTGGCGTCCTTGGCTAGACTTCTGGAAAGCCGAAAGTATAT

ATTGGTAGTGCCGTTGATTGTAATTAACGAATTGGATGGGTTGGCGAAAGGACAAG

AGACTGATCACAGAGCAGGAGGCTACGCGAGGGTCGTCCAAGAGAAGGCGCGAA

AAAGCATCGAGTTCCTGGAGCAGCGATTTGAGAGCAGGGACTCATGCCTGAGAGC

CCTCACGTCACGGGGGAACGAGCTGGAGTCCATCGCTTTCCGAAGTGAAGACATT

ACGGGCCAACTTGGGAATAATGATGACCTCATCTTGTCCTGCTGCCTGCACTACTG

CAAGGACAAGGCTAAGGACTTCATGCCTGCCTCCAAGGAGGAGCCTATCCGATTG

TTGAGGGAAGTAGTACTTTTGACGGACGACCGCAACCTCCGGGTAAAGGCGCTGA

CTCGAAATGTCCCAGTAAGGGATATACCGGCGTTCCTTACATGGGCTCAAGTAGG

GGATTATAAGGATCATGATGGGGATTACAAGGACCATGACATCGATTATAAAGATG

ACGACGATAAGCACCATCACCACCATCACCCCGGGCTCGAGGTTAATTAAGCGGC

CGCATTGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCA

CCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGG

TTTTTTGCTGAAAGGAGGAACTATATCCGGAT

**PIN-SNAP:**

AACTACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTAT

TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCT

TCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTAT

TCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAA

AGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGAT

CTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTCCCAATGAT

GAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGC

AAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCA

CCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGC

TGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAG

GACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTT

GATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCA

CGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTT

ACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAG

GACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA

GCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAG

CCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAAC

GAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCA

GACCAAGTTTACTCATATATACTTTAGATTGATTTACCCCGGTTGATAATCAGAAAA

GCCCCAAAAACAGGAAGATTGTATAAGCAAATATTTAAATTGTAAACGTTAATATTTT

GTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAA

ATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGT

TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGC

GAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAG

TTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCC

CGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAG

AAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGC

GTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTAAAAGGATC

TAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCG

TTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTT

TTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGG

TTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGC

AGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTT

CAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG

CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTT

ACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAG

CTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAA

AGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGG

GTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTT

TATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTC

GTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTT

CCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTC

TGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGA

ACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCTATGGTGCACTCTCAGTAC

AATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGA

CTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG

GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGC

TGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGT

AAAGCTCATCAGCGTGGTCGTGCAGCGATTCACAGATGTCTGCCTGTTCATCCGC

GTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGG

CCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGA

TTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACG

GGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTG

GCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCCG

AACGCCAGCAAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCGATAATGGCC

TGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGG

GCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGC

GAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGA

GTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGC

CCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCC

CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTT

TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG

GAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGAC

GGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCG

GTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGC

GGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGC

ACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTG

ATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATG

GTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCT

GAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGA

GACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACC

AGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGAT

GGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCT

TCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGAC

GCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGT

TCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCG

CCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAA

TCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTC

AGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGG

CCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGAC

ATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGC

GCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTC

GACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAG

GCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCCGGCATGCCGCCCTTTC

GTCTTCAAGAATTAATTCCCAATTCCCCAGGCATCAAATAAAACGAAAGGCTCAGT

CGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGT

AGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGG

TGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAATTCCCCAGGCATCAAA

TAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCG

GTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGA

AGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATT

AATTCCCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGT

TTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGC

GGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGC

CATAAACTGCCAGGAATTAATTCCCCAGGCATCAAATAAAACGAAAGGCTCAGTCG

AAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAG

GACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTG

GCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAATTCCCCAGGCATCAAATA

AAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGT

GAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAG

CAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTGG

GGATCGGAATTAATTCCCGGTTTAAACCGGGGATCTCGATCCCGCGAAATTAATAC

GACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGT

TTAACTTTAAGAAGGAGATATACATATGGAGCTCGAAATCAGGCCGCTGTTCCTCG

TGCCGGACACTAATGGTTTTATAGATCACTTGGCGTCCTTGGCTAGACTTCTGGAA

AGCCGAAAGTATATATTGGTAGTGCCGTTGATTGTAATTAACGAATTGGATGGGTT

GGCGAAAGGACAAGAGACTGATCACAGAGCAGGAGGCTACGCGAGGGTCGTCCA

AGAGAAGGCGCGAAAAAGCATCGAGTTCCTGGAGCAGCGATTTGAGAGCAGGGA

CTCATGCCTGAGAGCCCTCACGTCACGGGGGAACGAGCTGGAGTCCATCGCTTTC

CGAAGTGAAGACATTACGGGCCAACTTGGGAATAATGATGACCTCATCTTGTCCTG

CTGCCTGCACTACTGCAAGGACAAGGCTAAGGACTTCATGCCTGCCTCCAAGGAG

GAGCCTATCCGATTGTTGAGGGAAGTAGTACTTTTGACGGACGACCGCAACCTCC

GGGTAAAGGCGCTGACTCGAAATGTCCCAGTAAGGGATATACCGGCGTTCCTTAC

ATGGGCTCAAGTAGGGAGTGGAAGTGAGACACCGGGAACCTCAGAGAGCGCCAC

GCCAGAAAGCGACAAAGATTGCGAAATGAAACGTACCACCCTGGATAGCCCGCTG

GGCAAACTGGAACTGAGCGGCTGCGAACAGGGCCTGCATGAAATTAAACTGCTGG

GTAAAGGCACCAGCGCGGCCGATGCGGTTGAAGTTCCGGCCCCGGCCGCCGTGC

TGGGTGGTCCGGAACCGCTGATGCAGGCGACCGCGTGGCTGAACGCGTATTTTCA

TCAGCCGGAAGCGATTGAAGAATTTCCGGTTCCGGCGCTGCATCATCCGGTGTTT

CAGCAGGAGAGCTTTACCCGTCAGGTGCTGTGGAAACTGCTGAAAGTGGTTAAAT

TTGGCGAAGTGATTAGCTATCAGCAGCTGGCGGCCCTGGCGGGTAATCCGGCGG

CCACCGCCGCCGTTAAAACCGCGCTGAGCGGTAACCCGGTGCCGATTCTGATTCC

GTGCCATCGTGTGGTTAGCTCTAGCGGTGCGGTTGGCGGTTATGAAGGTGGTCTG

GCGGTGAAAGAGTGGCTGCTGGCCCATGAAGGTCATCGTCTGGGTAAACCGGGT

CTGGGACCTGCAGGCCACCATCACCACCATCACCCCGGGCTCGAGGTTAATTAAG

CGGCCGCATTGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCT

GCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGA

GGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

**PIN:**

AACTACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTAT

TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCT

TCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTAT

TCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAA

AGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGAT

CTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTCCCAATGAT

GAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGC

AAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCA

CCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGC

TGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAG

GACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTT

GATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCA

CGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTT

ACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAG

GACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA

GCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAG

CCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAAC

GAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCA

GACCAAGTTTACTCATATATACTTTAGATTGATTTACCCCGGTTGATAATCAGAAAA

GCCCCAAAAACAGGAAGATTGTATAAGCAAATATTTAAATTGTAAACGTTAATATTTT

GTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAA

ATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGT

TCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGC

GAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAG

TTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCC

CGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAG

AAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGC

GTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTAAAAGGATC

TAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCG

TTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTT

TTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGG

TTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGC

AGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTT

CAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG

CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTT

ACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAG

CTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAA

AGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGG

GTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTT

TATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTC

GTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTT

CCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTC

TGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGA

ACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCTATGGTGCACTCTCAGTAC

AATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGA

CTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG

GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGC

TGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGT

AAAGCTCATCAGCGTGGTCGTGCAGCGATTCACAGATGTCTGCCTGTTCATCCGC

GTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGG

CCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGA

TTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACG

GGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTG

GCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCCG

AACGCCAGCAAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCGATAATGGCC

TGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGG

GCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGC

GAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGA

GTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGC

CCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCC

CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTT

TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG

GAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGAC

GGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCG

GTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGC

GGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGC

ACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTG

ATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATG

GTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCT

GAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGA

GACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACC

AGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGAT

GGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCT

TCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGAC

GCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGT

TCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCG

CCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAA

TCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTC

AGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGG

CCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGAC

ATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGC

GCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTC

GACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAG

GCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCCGGCATGCCGCCCTTTC

GTCTTCAAGAATTAATTCCCAATTCCCCAGGCATCAAATAAAACGAAAGGCTCAGT

CGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGT

AGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGG

TGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAATTCCCCAGGCATCAAA

TAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCG

GTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGA

AGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATT

AATTCCCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGT

TTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGC

GGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGC

CATAAACTGCCAGGAATTAATTCCCCAGGCATCAAATAAAACGAAAGGCTCAGTCG

AAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAG

GACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTG

GCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAATTCCCCAGGCATCAAATA

AAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGT

GAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAG

CAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTGG

GGATCGGAATTAATTCCCGGTTTAAACCGGGGATCTCGATCCCGCGAAATTAATAC

GACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGT

TTAACTTTAAGAAGGAGATATACATATGGAGCTCGAAATCAGGCCGCTGTTCCTCG

TGCCGGACACTAATGGTTTTATAGATCACTTGGCGTCCTTGGCTAGACTTCTGGAA

AGCCGAAAGTATATATTGGTAGTGCCGTTGATTGTAATTAACGAATTGGATGGGTT

GGCGAAAGGACAAGAGACTGATCACAGAGCAGGAGGCTACGCGAGGGTCGTCCA

AGAGAAGGCGCGAAAAAGCATCGAGTTCCTGGAGCAGCGATTTGAGAGCAGGGA

CTCATGCCTGAGAGCCCTCACGTCACGGGGGAACGAGCTGGAGTCCATCGCTTTC

CGAAGTGAAGACATTACGGGCCAACTTGGGAATAATGATGACCTCATCTTGTCCTG

CTGCCTGCACTACTGCAAGGACAAGGCTAAGGACTTCATGCCTGCCTCCAAGGAG

GAGCCTATCCGATTGTTGAGGGAAGTAGTACTTTTGACGGACGACCGCAACCTCC

GGGTAAAGGCGCTGACTCGAAATGTCCCAGTAAGGGATATACCGGCGTTCCTTAC

ATGGGCTCAAGTAGGGCCTGCAGGCCACCATCACCACCATCACCCCGGGCTCGA

GGTTAATTAAGCGGCCGCATTGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTG

AGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAA

ACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

## References


1. Alexander, S. C. & Devaraj, N. K. Developing a Fluorescent Toolbox To Shed Light on the Mysteries of RNA. *Biochemistry* **56**, 5185–5193 (2017).

2. Hayashi, G. & Okamoto, A. Probe Design for the Effective Fluorescence Imaging of Intracellular RNA. (2013) doi:10.1002/tcr.201200026.

3. Rath, A. K. & Rentmeister, A. Genetically encoded tools for RNA imaging in living cells. *Curr. Opin. Biotechnol.* **31**, 42–49 (2015).

4. Babendure, J. R., Adams, S. R. & Tsien, R. Y. Aptamers Switch on Fluorescence of Triphenylmethane Dyes. *J. AM. CHEM. SOC* **125**, 10 (2003).

5. Dolgosheina, E. V, Jeng, S. C. Y., Shyam, S., Panchapakesan, S., Cojocaru, R., Chen, P. S. K., Wilson, P. D., Hawkins, N., Wiggins, P. A. & Unrau, P. J. RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. (2014) doi:10.1021/cb500499x.

6. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science (80-. ).* **333**, 642–646 (2011).

7. Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H. & Long, R. M. Localization of ASH1 mRNA Particles in Living Yeast. *Mol. Cell* **2**, 437–445 (1998).

8. Daigle, N. & Ellenberg, J. l N-GFP: an RNA reporter system for live-cell imaging. (2007) doi:10.1038/NMETH1065.

9. Holstein, J. M., Schulz, D. & Rentmeister, A. Bioorthogonal site-specific labeling of the 5′-cap structure in eukaryotic mRNAs. *Chem. Commun.* **50**, 4478–4481 (2014).

10. Holstein, J. M., Stummer, D. & Rentmeister, A. Enzymatic modification of 5′-capped RNA with a 4-vinylbenzyl group provides a platform for photoclick and inverse electron-demand Diels–Alder reaction. *Chem. Sci.* **6**, 1362–1369 (2015).

11. Gillingham, D. & Shahid, R. Catalysts for RNA and DNA modification. *Curr. Opin. Chem. Biol.* **25**, 110–114 (2015).

12. Büttner, L., Javadi-Zarnaghi, F. & Höbartner, C. Site-Specific Labeling of RNA at Internal Ribose Hydroxyl Groups: Terbium-Assisted Deoxyribozymes at Work. *J. Am. Chem. Soc.* **136**, 8131–8137 (2014).

13. El Yacoubi, B., Bailly, M. & De Crécy-Lagard, V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.* **46**, 69–95 (2012).

14. Li, F., Dong, J., Hu, X., Gong, W., Li, J., Shen, J., Tian, H. & Wang, J. A Covalent Approach for Site-Specific RNA Labeling in Mammalian Cells. *Angew. Chemie - Int. Ed.* **54**, 4597–4602 (2015).

15. Garcia, G. A. & Kittendorf, J. D. Transglycosylation: A mechanism for RNA modification (and editing?). *Bioorganic Chemistry* vol. 33 229–251 (2005).

16. Alexander, S. C., Busby, K. N., Cole, C. M., Zhou, C. Y. & Devaraj, N. K. Site-specific covalent labeling of RNA by enzymatic transglycosylation. *J. Am. Chem. Soc.* **137**, 12756–12759 (2015).

17. Busby, K. N. & Devaraj, N. K. Enzymatic covalent labeling of RNA with RNA transglycosylation at guanosine (RNA-TAG). in *Methods in Enzymology* vol. 641 373–399 (Elsevier Inc., 2020).

18. Zhang, D., Jin, S., Liu, L., Tota, E., Li, Z., Piao, X. & Devaraj, N. K. RNA-CLAMP Enables Photo-activated Control of CRISPR-Cas9 Gene Editing by Site-specific Intramolecular Cross-linking of the sgRNA. *bioRxiv* 2021.04.22.441030 (2021) doi:10.1101/2021.04.22.441030.

19. Ehret, F., Zhou, C. Y., Alexander, S. C., Zhang, D. & Devaraj, N. K. Site-Specific Covalent Conjugation of Modified mRNA by tRNA Guanine Transglycosylase. *Mol. Pharm.* **15**, 737–742 (2018).

20. Busby, K. N., Fulzele, A., Zhang, D., Bennett, E. J. & Devaraj, N. K. Enzymatic RNA Biotinylation for Affinity Purification and Identification of RNA − Protein Interactions. 0–11 (2020) doi:10.1021/acschembio.0c00445.

21. Zhang, D., Jin, S., Piao, X. & Devaraj, N. K. Multiplexed Photoactivation of mRNA with Single-Cell Resolution. (2020) doi:10.1021/acschembio.0c00205.

22. Zhang, D., Zhou, C. Y., Busby, K. N., Alexander, S. C. & Devaraj, N. K. Light-Activated Control of Translation by Enzymatic Covalent mRNA Labeling Angewandte. **92093**, 2822–2826 (2018).

23. Zhou, C. Y., Alexander, S. C. & Devaraj, N. K. Fluorescent turn-on probes for wash-free mRNA imaging via covalent site-specific enzymatic labeling†. *Chem. Sci.* 7169–7173 (2017) doi:10.1039/c7sc03150e.

24. Stengl, B., Reuter, K. & Klebe, G. Mechanism and substrate specificity of tRNA-guanine transglycosylases (TGTs): tRNA-modifying enzymes from the three different kingdoms of life share a common catalytic mechanism. *ChemBioChem* **6**, 1926–1939 (2005).

25. Chen, Y. C., Brooks, A. F., Goodenough-Lashua, D. M., Kittendorf, J. D., Showalter, H. D. & Garcia, G. A. Evolution of eukaryal tRNA-guanine transglycosylase: Insight gained from the heterocyclic substrate recognition by the wild-type and mutant human and Escherichia coli tRNA-guanine

transglycosylases. *Nucleic Acids Res.* **39**, 2834–2844 (2011).

26. Curnow, A. W., Kung, F. L., Koch, K. A. & Garcia, G. A. tRNA-Guanine Transglycosylase from Escherichia coli: Gross tRNA Structural Requirements for Recognition. *Biochemistry* **32**, 5239–5246 (1993).

27. Nakanishi, S., Ueda, T., Hori, H., Yamazaki, N., Watanabe, N. & Okada, K. A UGU sequence in the anticodon loop is a minimum requirement for recognition by Escherichia coli tRNA-guanine transglycosylase. *J. Biol. Chem.* **269**, 32221–32225 (1994).

28. Todorov, K. A., Tan, X. J., Nonekowski, S. T., Garcia, G. A. & Carlson, H. A. The role of aspartic acid 143 in E. coli tRNA-guanine transglycosylase: Insights from mutagenesis studies and computational modeling. *Biophys. J.* **89**, 1965–1977 (2005).

29. Stengl, B., Reuter, K. & Klebe, G. Mechanism and substrate specificity of tRNA-guanine transglycosylases (TGTs): tRNA-modifying enzymes from the three different kingdoms of life share a common catalytic mechanism. *ChemBioChem* **6**, 1926–1939 (2005).

30. Romier, C., Meyer, J. E. W. & Suck, D. Slight sequence variations of a common fold explain the substrate specificities of tRNA-guanine transglycosylases from the three kingdoms. *FEBS Lett.* **416**, 93–98 (1997).

31. Goodenough-Lashua, D. M. & Garcia, G. A. tRNA–guanine transglycosylase from E. coli: a ping-pong kinetic mechanism is consistent with nucleophilic catalysis. *Bioorg. Chem.* **31**, 331–344 (2003).

32. Farkas, W. R., Jacobson, K. B. & Katze, J. R. Substrate and inhibitor specificity of tRNA-guanine ribosyltransferase. *BBA - Gene Struct. Expr.* **781**, 64–75 (1984).

33. Abold Todorov, K. 1 & Garcia, G. A. Role of aspartate 143 in Escherichia coli tRNA-guanine transglycosylase: alteration of heterocyclic substrate specificity. *Biochemistry* **45**, 617–625 (2006).

34. Nonekowski, S. T., Kung, F. L. & Garcia, G. A. The Escherichia coli tRNA-guanine transglycosylase can recognize and modify DNA. *J. Biol. Chem.* **277**, 7178–7182 (2002).

35. Kittendorf, J. D., Sgraja, T., Reuter, K., Klebe, G. & Garcia, G. A. An Essential Role for Aspartate 264 in Catalysis by tRNA-Guanine Transglycosylase from Escherichia coli. *J. Biol. Chem.* **278**, 42369–42376 (2003).

36. Stengl, B., Meyer, E. A., Heine, A., Brenk, R., Diederich, F. & Klebe, G. Crystal Structures of tRNA-guanine Transglycosylase (TGT) in Complex with Novel and Potent Inhibitors Unravel Pronounced Induced-fit Adaptations and Suggest Dimer Formation Upon Substrate Binding. *J. Mol. Biol.* (2007)

doi:10.1016/j.jmb.2007.04.008.

37. Romier, C., Reuter, K., Suck, D. & Ficner, R. Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange. *EMBO J.* **15**, 2850–2857 (1996).

38. Ishitani, R., Nureki, O., Fukai, S., Kijimoto, T., Nameki, N., Watanabe, M., Kondo, H., Sekine, M., Okada, N., Nishimura, S. & Yokoyama, S. Crystal structure of archaeosine tRNA-guanine transglycosylase. *J. Mol. Biol.* **318**, 665–677 (2002).

39. Grädler, U., Ficner, R., Garcia, G. A., Stubbs, M. T., Klebe, G. & Reuter, K. Mutagenesis and crystallographic studies of Zymomonas mobilis tRNA-guanine transglycosylase to elucidate the role of serine 103 for enzymatic activity. *FEBS Lett.* **454**, 142–146 (1999).

40. Barandun, L. J., Immekus, F., Kohler, P. C., Ritschel, T., Heine, A., Orlando, P., Klebe, G. & Diederich, F. High-affinity inhibitors of Zymomonas mobilis tRNA-guanine transglycosylase through convergent optimization. *Acta Crystallogr. Sect. D* (2013) doi:10.1107/S0907444913014509.

41. Xie, W., Liu, X. & Huang, R. H. Chemical trapping and crystal structure of a catalytic tRNA guanine transglycosylase covalent intermediate. *Nat. Struct. Biol.* **10**, 781–788 (2003).

42. Biela, I., Tidten-Luksch, N., Immekus, F., Glinca, S., Nguyen, T. X. P., Gerber, H. D., Heine, A., Klebe, G. & Reuter, K. Investigation of Specificity Determinants in Bacterial tRNA-Guanine Transglycosylase Reveals Queuine, the Substrate of Its Eucaryotic Counterpart, as Inhibitor. *PLoS One* **8**, (2013).

43. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-. ).* **337**, 816–821 (2012).

44. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. & Zhang, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (80-. ).* **339**, 819–823 (2013).

45. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

46. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.* **109**, E2579–E2586 (2012).

47. Xu, Y. & Li, Z. CRISPR-Cas systems: Overview, innovations and applications in human disease research and gene therapy. *Comput. Struct. Biotechnol. J.* **18**, 2401 (2020).

48. Esposito, C. L., Catuogno, S., Condorelli, G., Ungaro, P. & Franciscis, V. de. Aptamer Chimeras for Therapeutic Delivery: The Challenging Perspectives. *Genes (Basel).* **9**, (2018).

49. Li, X., Feng, K., Li, L., Yang, L., Pan, X., Yazd, H. S., Cui, C., Li, J., Moroz, L., Sun, Y., Wang, B., Li, X., Huang, T. & Tan, W. Lipid–oligonucleotide conjugates for bioapplications. *Natl. Sci. Rev.* **7**, 1933–1953 (2020).

50. Tesz, G. J., Aouadi, M., Prot, M., Nicoloro, S. M., Boutet, E., Amano, S. U., Goller, A., Wang, M., Guo, C. A., Salomon, W. E., Virbasius, J. V., Baum, R. A., O'Connor, M. J., Soto, E., Ostroff, G. R. & Czech, M. P. Glucan particles for selective delivery of siRNA to phagocytic cells in mice. *Biochem. J.* **436**, 351–362 (2011).

51. Juillerat, A., Gronemeyer, T., Keppler, A., Gendreizig, S., Pick, H., Vogel, H. & Johnsson, K. Directed Evolution of O6 -Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules In Vivo. *Chem. Biol.* **10**, 53–60 (2003).

52. Los, G. V, Encell, L. P., Mcdougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., Wood, M. G., Learish, R., Friedman Ohana, R., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D. H., Bulleit, R. F., *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol. Downloaded* **3**, (2021).

53. Gautier, A., Juillerat, A., Heinis, C., Corrêa, I. R., Kindermann, M., Beaufils, F. & Johnsson, K. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **15**, 128–136 (2008).

54. Curnow, A. W. & Garcia, G. A. tRNA-guanine transglycosylase from Escherichia coli: Minimal tRNA structure and sequence requirements for recognition. *Journal of Biological Chemistry* vol. 270 17264–17267 (1995).

55. Ono, K. Inhibitory effects of various 2′, 3′-dideoxynucleoside 5′-triphosphates on the utilization of 2′-deoxynucleoside 5′-triphosphates by terminal deoxynucleotidyltransferase from calf thymus. *BBA - Gene Struct. Expr.* **1049**, 15–20 (1990).

56. Zearfoss, N. R. & Ryder, S. P. End-Labeling Oligonucleotides with Chemical Tags After Synthesis. *Methods Mol. Biol.* **941**, 181 (2012).

57. Busby, K. N. & Devaraj, N. K. *Enzymatic covalent labeling of RNA with RNA transglycosylation at guanosine (RNA-TAG)*. *Methods in Enzymology* (Elsevier Inc., 2020). doi:10.1016/bs.mie.2020.03.009.

58. Zhang, D., Zhou, C. Y., Busby, K. N., Alexander, S. C. & Devaraj, N. K. Light-Activated Control of Translation by Enzymatic Covalent mRNA Labeling. *Angew. Chemie - Int. Ed.* **57**, 2822–2826 (2018).

59. Zhou, C. Y., Alexander, S. C. & Devaraj, N. K. Fluorescent turn-on probes for wash-free mRNA imaging: Via covalent site-specific enzymatic labeling. *Chem. Sci.* **8**, 7169–7173 (2017).

60. He, S. L. Northern blot. *Methods Enzymol.* **530**, 75 (2013).

61. Miller, B. R., Wei, T., Fields, C. J., Sheng, P. & Xie, M. Near-infrared fluorescent northern blot. *RNA* **24**, 1871–1877 (2018).

62. Levsky, J. M. & Singer, R. H. Fluorescence in situ hybridization: Past, present and future. *Journal of Cell Science* vol. 116 2833–2838 (2003).

63. Shaffer, S. M., Wu, M.-T., Levesque, M. J. & Raj, A. Turbo FISH: A Method for Rapid Single Molecule RNA FISH. *PLoS One* **8**, e75120 (2013).

64. Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J. & Luo, Y. RNAscope: A Novel in Situ RNA Analysis Platform for Formalin-Fixed, Paraffin-Embedded Tissues. *J. Mol. Diagn.* **14**, 22 (2012).

65. Levesque, M. J. & Raj, A. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nat. Methods 2013 103* **10**, 246–248 (2013).

66. Levesque, M. J., Ginart, P., Wei, Y. & Raj, A. Visualizing SNVs to quantify allele-specific expression in single cells. *Nat. Methods 2013 109* **10**, 865–867 (2013).

67. Gall, J. G. & Pardue, M. L. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc. Natl. Acad. Sci. U. S. A.* **63**, 378–383 (1969).

68. Bauman, J. G. J., Wiegant, J., Borst, P. & van Duijn, P. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. *Exp. Cell Res.* **128**, 485–490 (1980).

69. Singer, R. H. & Ward, D. C. Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 7331–7335 (1982).

70. Vargas, D. Y., Raj, A., Marras, S. A. E., Kramer, F. R. & Tyagi, S. Mechanism of mRNA transport in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17008–17013 (2005).

71. Tyagi, S. & Kramer, F. R. Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nat. Biotechnol. 1996 143* **14**, 303–308 (1996).

72. Kishi, J. Y., Lapan, S. W., Beliveau, B. J., West, E. R., Zhu, A., Sasaki, H. M., Saka, S. K., Wang, Y., Cepko, C. L. & Yin, P. SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat. Methods 2019 166*

**16**, 533–544 (2019).

73. Xia, C., Babcock, H. P., Moffitt, J. R. & Zhuang, X. Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci. Reports 2019 91* **9**, 1–13 (2019).

74. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science (80-. ).* **280**, 585–590 (1998).

75. Raj, A., Bogaard, P. van den, Rifkin, S. A., Oudenaarden, A. van & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods 2008 510* **5**, 877–879 (2008).

76. Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., Emili, A. & Sunney Xie, X. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science (80-. ).* **329**, 533–538 (2010).

77. Paré, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y. & McGinnis, W. Transcriptional analysis of the Hox gene Scr at single molecule resolution yields evidence for transcriptional bursting during Drosophila embryogenesis. *Curr. Biol.* **19**, 2037 (2009).

78. Gaspar, I., Wippich, F. & Ephrussi, A. Enzymatic production of single-molecule FISH and RNA capture probes. *RNA* **23**, 1582–1591 (2017).

79. Hanswillemenke, A., Kuzdere, T., Vogel, P., Jékely, G. & Stafforst, T. Site-Directed RNA Editing in Vivo Can Be Triggered by the Light-Driven Assembly of an Artificial Riboprotein. *J. Am. Chem. Soc.* **137**, 15875–15881 (2015).

80. Batra, R., Nelles, D. A., Pirie, E., Blue, S. M., Marina, R. J., Wang, H., Chaim, I. A., Thomas, J. D., Zhang, N., Nguyen, V., Aigner, S., Markmiller, S., Xia, G., Corbett, K. D., Swanson, M. S. & Yeo, G. W. Elimination of Toxic Microsatellite Repeat Expansion RNA by RNA-Targeting Cas9. *Cell* **170**, 899 (2017).

81. Ivics, Z., Izsvák, Z., Medrano, G., Chapman, K. M. & Hamra, F. K. Sleeping Beauty transposon mutagenesis in rat spermatogonial stem cells. *Nat. Protoc.* **6**, 1521 (2011).

82. Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J. & Zhang, F. RNA editing with CRISPR-Cas13 David. *Science (80-. ).* **0180**, 1–15 (2017).

83. Crooke, S. T., Liang, X.-H., Baker, B. F. & Crooke, R. M. Antisense technology: A review. *J. Biol. Chem.* **296**, 100416 (2021).

84. Goodnow, R. A., Dumelin, C. E. & Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov. 2016 162* **16**, 131–147 (2016).

85. Hong, F., Zhang, F., Liu, Y. & Yan, H. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem. Rev.* **117**, 12584–12640 (2017).

86. Alexander, S. C., Busby, K. N., Cole, C. M., Zhou, C. Y. & Devaraj, N. K. Site-specific covalent labeling of RNA by enzymatic transglycosylation. *J. Am. Chem. Soc.* **137**, 12756–12759 (2015).