UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Synthetic strategies towards the development of new RNA structure and RNA labeling tools.

Permalink

https://escholarship.org/uc/item/5pb14022

Author

Beasley, Samantha

Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Synthetic strategies towards the development of new RNA structure and
RNA labeling tools.

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Pharmacological Sciences

by

Samantha Marie Beasley

Dissertation Committee:
Associate Professor Robert Spitale, Chair
Professor Richard Chamberlin
Professor Celia Goulding

2019

DEDICATION

To

my parents, Tina and Dennis,
my sister, Nikki, and
Michael Fazio for their endless support

"Some people never go crazy. What truly horrible lives they must lead."

"Things get bad for all of us, almost continually, and what we do under
the constant stress reveals who/what we are."

"What matters most is how well you walk through the fire."

Charles Bukowski

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

CURRICULUM VITAE

Samantha Beasley

Department of Pharmaceutical Sciences
University of California, Irvine
Email: smbeasle@uci.edu

Summary

- Hard working and detail-oriented chemical biologist with skills in synthetic chemistry, sample analysis, cell treatment, RNA handling, and project management.
- 5 publications since beginning research work in 2015 (two more In Preparation).
- Strong communication skills, from over two years of working as a supplemental instructor in organic chemistry and STEM tutor in undergrad, and working as both a laboratory and discussion instructor in graduate school.
- Extensive mentorship experience from working as the Pharmaceutical Sciences graduate student ambassador.

Education

- **Doctor of Philosophy,** Pharmaceutical Sciences – emphasis in Chemical biology; University of California, Irvine, July 2019.

- **Master of Science,** Pharmaceutical Sciences; University of California, Irvine, 2017

- **Bachelor of Science, Biochemistry, Summa Cum Laude;** California State University San Marcos, 2014

Publications

1. **S. Beasley** and R.C. Spitale. "Cancer cell-specific metabolic labeling of RNA." *In Preparation*
2. **S. Beasley**, Sakr J., and R.C. Spitale. "Accessing vinyl group reactivity beyond IEDDA Tetrazine reactions." *In Preparation*
3. **S. Beasley,** K. Nguyen, M. Fazio, R.C. Spitale. "Protected pyrimidine nucleosides for cell-specific metabolic labeling of RNA." *Tet Lett*, 2018, 59, 3912-3915.
4. D. Chan*, **S. Beasley***, Y. Zhen, R. C. Spitale. "Facile synthesis and evaluation of a dual-functioning furoyl probe for in-cell SHAPE." *Bioorganic & MedChem Lett*, 2018, 28, 601-605.
   *- Denotes co-first authorship
5. N. Hida, M. Aboukilila, D. Burow, R. Paul, M. Greenberg, M. Fazio, **S. Beasley**, R.C. Spitale, M. Cleary. "EC-tagging allows cell type-specific RNA analysis." *Nucleic Acids Res,* 2017, 45, e138.
6. S. Nainar, **S. Beasley**, M. Fazio, M. Kubota, R.C. Spitale. "Metabolic incorporation of azide functionality into cellular RNA." *ChemBioChem*, 2016, 17, 2149-2152.
7. **S. Beasley**, M. Pilkus, R.C. Spitale, I. Petersen. "The emerging functions of regulatory RNA species in skin biology." *Experimental Dermatology*, 2015, 24, 827-828.

Skills

- Proficient in organic synthesis especially focused on nucleoside analogs.
- Various spectroscopic techniques including NMR, HPLC, MS, IR, among others.
- Proficient in biological techniques including dot blot, gel electrophoresis, cell culture, RNA extraction, PCR, in vitro transcription, reverse transcription, among others.

Awards and Honors

- NIH T32 Cancer Training Grant Fellowship recipient 2017- present

- Pharmaceutical Science's Chair's Lifetime Achievement Award for Service 2019
- American Association for University Women Scholarship 2013

Research Experience
- **Graduate Researcher,** UC Irvine, 2014-current
    - Synthesized various organic small molecules and numerous nucleoside analogs.
    - Performed cell testing and downstream analysis of nucleoside analogs.
    - Reported scientific findings at symposiums and in journal publications, showing oral and written presentation skills.
    - Maintained lab equipment. Organized lab space as well as general lab stocks.
- **Research Assistant,** California State University San Marcos 2012-2013
    - Worked with bioactive natural products produced by marine organisms with the hopes of finding an antibiotic compound that could inhibit the growth of *Mycobacterium marinum,* which is related to *Mycobacterium tuberculosis*, and could therefore be a potential solution to tuberculosis.
    - Prepared and grew cultures, worked with TLC, performed column extraction/separation, bioassay, ran HNMR, HPLC, GMS, as well as other laboratory techniques.

Presentations and Posters
- **S. Beasley**, R.C. Spitale. "Cancer Cell Specific RNA Labeling." ACS National Meeting, 2019
- **S. Beasley**, K. Nguyen, M. Fazio, R.C. Spitale. "Protected Pyrimidine Nucleosides for Cell-specific Metabolic Labeling of RNA." UC Chemical Symposium, 2019
- **S. Beasley**, J. Trischman. "Extraction and Isolation of Bioactive Antibacterial Products in Marine Bacterium strain UA 774 Against *Mycobacterium marinum.*" ACS Western Regional Meeting, 2014.
- S. Habib, **S. Beasley**, J. Trischman. "Isolation of Bioactive Antibacterial Products in a Marine Bacterium Against the Marine Analogue of *M. tuberculosis- M. marinum."* 2013.

Teaching Assistant Experience
- Organic Chemistry Supplemental Instructor 2012-2013
- STEM Tutor 2014
- Chem 151L: General Chemistry Lab 2015
- PharmSci 177L: Medicinal Chemistry and Pharmacology 2015
- PharmSci 120 and 120L: Human Physiology 2016
- PharmSci 173: Pharmacology 2017

Mentorship
- 2015-2019: Mentored the incoming PharmSci graduate student class every year.
- 2015-2018: Mentored rotating students.
- 2018-2019: Mentored undergraduate, Jasmine Sakr, who has been accepted to the PharmSci graduate student program for 2019.
- 2018-2019: Member of the Chemistry Mentor-Mentee Program

ABSTRACT OF THE DISSERTATION

Synthetic strategies towards the development of new RNA structure and
RNA labeling tools.

By

Samantha Marie Beasley

Doctor of Philosophy in Pharmacological Sciences

University of California, Irvine, 2019

Professor Robert C. Spitale, Chair

The central dogma originally postulated that RNA solely played the role of genetic information transfer from DNA to protein. But that has been proven to be an inadequate depiction of this genetic material. RNA plays an integral role in the cell beyond just translation alone. RNA is capable of regulating every step of the central dogma. Because of this realization, the development of the RNA toolkit has been an ever-growing field. This toolkit consists of techniques to study RNA structure, localization, associated binding proteins, and nascent transcription. Herein, my studies aimed to enhance the field of RNA structure probing, as well as the field of labeling nascent RNA with bioorthogonal nucleosides. To that aim, I have synthesized a new probe with increased stability for icSHAPE structure probing of RNA. Additionally, I have synthesized the first azide containing bioorthogonal nucleosides for RNA labeling. Lastly, I extended the PGA toolkit to include pyrimidine nucleosides so that the cell-specific RNA labeling can be biased for either RNA or polyadenylated RNA, based on the nucleoside being used.

Chapter 1: Background

1.1 Introduction

In the 1950s, Watson and Crick proposed the central dogma, which stated that cellular information transfer occurred in the linear process of DNA to RNA to protein.[1] This led the field to believe RNA had no cellular roles beyond translation. However, by 2003 with the human genome project, it was discovered that only 2% of DNA encodes for protein coding genes, and the other 98% of RNA that DNA encodes for was then classified as non-protein coding RNA, better known as noncoding RNA.[2] This was further validated by the ENCODE (encyclopedia of DNA elements) project, thus pushing the field to study this new class of noncoding RNA.[3] With this new research came the discovery that RNA is capable of regulating every aspect of gene regulation.[4] This is done through RNA's unique ability to interact with DNA, RNA, proteins, and other small molecules.[5-7]
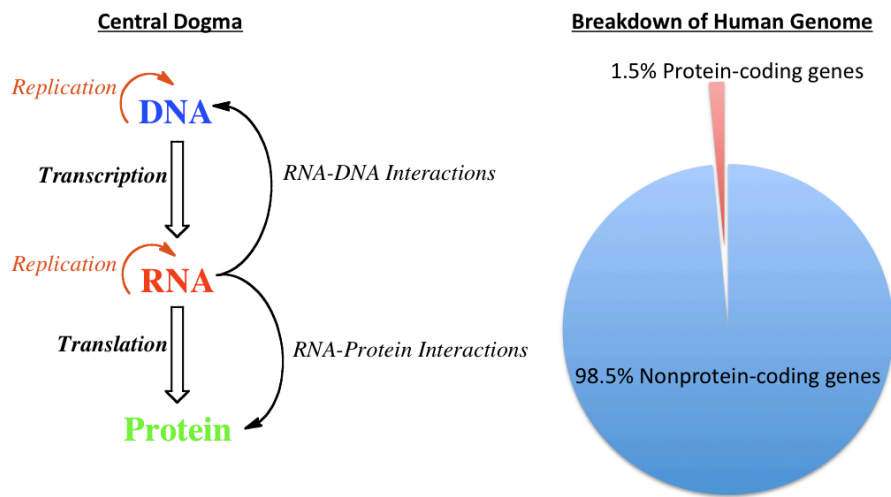
Figure 1. RNA is capable of interacting with every step of the central dogma. The Human Genome project showed found that less than 2% of the genome codes for protein-coding genes.

Chemical similarities in base pair functional groups endows RNA with the ability to base pair with DNA and RNA, which is integral for genetic regulation, such as transcription

and mRNA processing.[8, 9] An example of this is the function of microRNAs, a class of noncoding RNA, that is responsible for binding to mRNA via base-pairing and silencing the gene through degradation or prevention of translation.[9] Additionally, the physical structure of RNA also plays an integral role in its proper function. RNA is normally single stranded, which allows it to fold onto itself and form secondary and tertiary structural motifs, such as hairpins, internal loops, pseudoknots, and bulges.[10] These elements serve as binding sites for DNA, RNA, proteins, and small molecules. A key example of this is the ribosome, which is composed of roughly 60% rRNA and 40% protein, all of which are folded and bound to each other, and without this proper alignment, translation would not be possible.[11]

The importance of RNA structure and function has also been implicated in diseases.[12] Spinal muscular atrophy (SMA) is a neurodegenerative disease that is the direct cause of a key gene, SMN1, being homozygously lost or mutated, resulting in inactive SMN1 RNA and thus significantly reduced amount of survival motor protein-1 (SMN1).[13] SMN1 is essential for the proper assembly of ribonucleoprotein complexes responsible for spinal cord motor neurons survival.[14] An example of an important noncoding RNA is the long-noncoding RNA HOTAIR (HOX antisense intergenic RNA). HOTAIR is an important structural scaffold for binding proteins that make up gene silencing machineries and bringing them to the corresponding gene targets.[15] HOTAIR is a critical lncRNA for proper growth and cell viability. The overexpression of HOTAIR has been linked to a variety of cancers, with overexpression in breast cancer increasing the cancer's metastatic potential and tumor invasiveness.[15] These serve as a few examples of how functional RNAs can play important roles in normal and disease biology.

Unique RNA structure elements can also have important roles in disease biology, where they can fold into structures, such as hairpins, that act as binding sites.[10] Any mutation or alteration to the RNA sequence could then lead to RNA misfolding and a gain or loss of such structures. An example of this is found in the disease known as mytonic muscular dystrophy type 1 (DM1). DM1 is the result of an expansion to the 3' untranslated region of the DM1 gene in the form of a CUG repeat, and it's downstream effect is the weakening of skeletal muscle.[16] The CUG repeat results in a new hairpin formation on the RNA that sequesters the Muscleblind-like 1 protein (MBNL1), a key alternative splicing regulator.[17] This DM1 RNA-MBNL1 complex results in loss of alternative splicing regulation for key pre-mRNAs responsible for muscle development and leads to disease progression.[18]

With the discovery of RNA playing a critical role for cell survival, beyond the once believed solo task of translation, it was necessary to develop better tools to study RNA. This need has led to the development of the RNA toolkit. The toolkit is a plethora of techniques that have been developed to study RNA, covering the aspects of RNA structure, RNA binding proteins, localization, and labeling.[19-22] My thesis will cover both RNA structure and RNA labeling.

RNA structure can be studied *in vitro* through utilizing chemical probes that have very specific chemical selectivity to react with specific nucleobases, sugar functional groups, or the backbone of RNA; with their level of reactivity directly related to functional group accessibility. Examples of these techniques are, but not limited to, dimethylsulfate (DMS) probing, hydroxy radical probing, and selective hydroxyl acylation analyzed by primer extension (SHAPE), as can be seen in Figure 2.[19, 23-24]
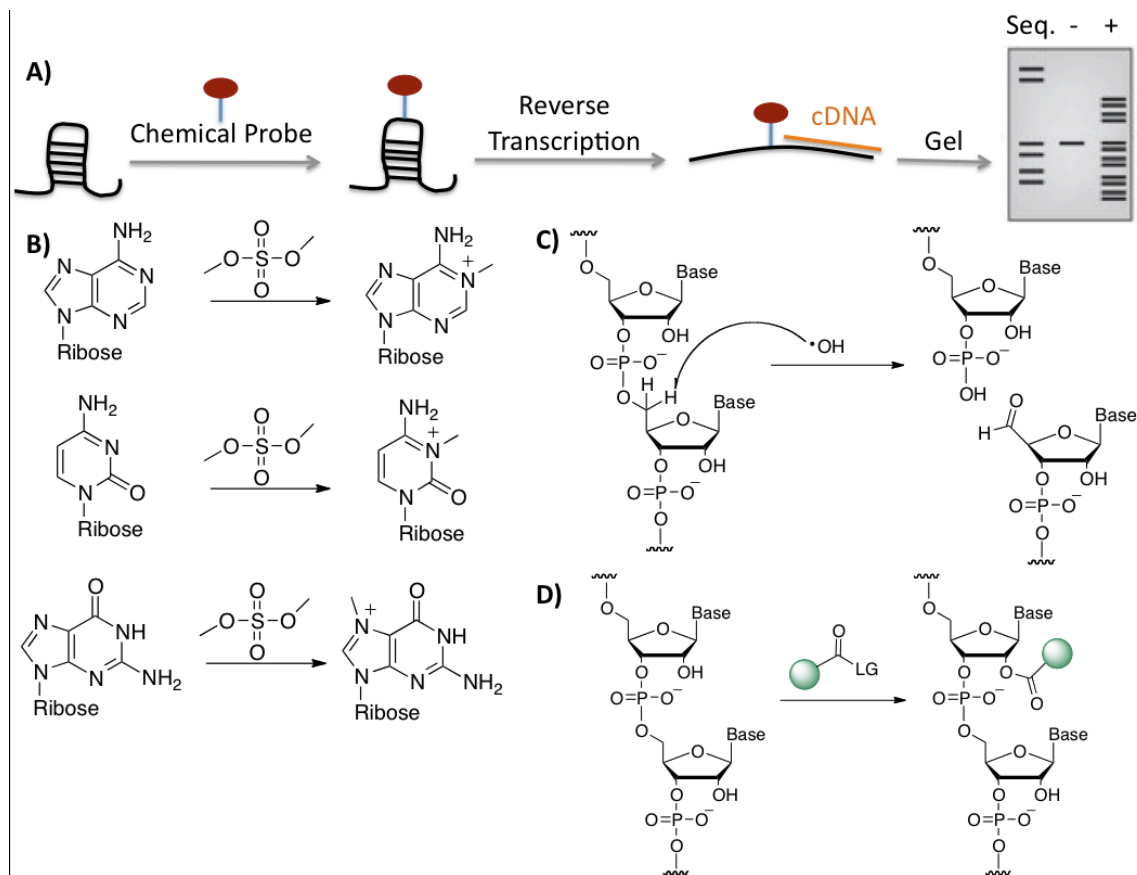
Figure 2. A) Schematic for structure probing of RNA. B) DMS probing. C) Hydroxy radical probing. D) SHAPE structure probing.

DMS probing of RNA results in the methylation of the N7 position of guanosine, the N3 position of cytidine, and the N1 position of adenosine that are sterically accessible.[25] The structure of the DMS treated RNA can be determined by reverse transcription polymerase chain reaction (RT-PCR). Reverse transcriptase will not be able to read through the methylated Watson Crick faces of N3-methyl-cytidine and N1-methyl-adenosine. This will result in truncated cDNAs that can be identified and mapped to the complete cDNA through gel electrophoresis.[26]

Hydroxy radical probing determines RNA structure through cleavage of the RNA backbone at accessible sites.[24] The hydroxy radical attacks the C5' position and results in a

strand cleavage forming a 3'-phosphate and a 5'-aldehyde on the two resulting strands, which can then be analyzed by RT-PCR and gel electrophoresis.[25] This method can be used to determine the position of RNA-binding proteins or tertiary structures in the RNA, which would be seen as a loss of reactivity at those inaccessible sites.[27]

SHAPE relies on using small molecule electrophiles coupled with the nucleophilicity of solvent accessible 2'-hydroxyls of the RNA, and the resulting adduct formed at the 2'-position is read through RT-PCR to reveal sites of single-strandedness in a given RNA sequence.[19] SHAPE has been used to study numerous RNAs, and when combined with structure prediction software, a model of the RNA secondary structure can be developed.[28] SHAPE was further developed to *in vivo* click SHAPE (icSHAPE), where the small molecule electrophile now has an azide appended to it, which allows for the utilization of click chemistry to better purify the modified RNA.[29] This enrichment for the acylated RNA allows for enhanced analysis of RT stops to determine RNA structure.

Overall, DMS probing identifies single stranded adenosine and cytidine nucleosides, which yields a limited glimpse at the overall structure of the RNA. Hydroxy radical probing and SHAPE both go a step further to provide a readout that is a direct depiction of the sugars that are solvent accessible within the complete secondary and tertiary structure of an RNA. icSHAPE further enhances the structural readout of RNA by providing an enrichment handle for affinity purification.

In addition to developing chemical reactions to structurally probe RNA, recent efforts have been focused on utilizing chemically modified nucleosides to track RNA expression. RNA labeling relies on taking advantage of the existing nucleoside salvage pathways inside of cells with the use of bioorthogonal nucleosides. Bioorthogonal nucleosides differ from

the standard uridine, cytidine, adenosine, and guanosine, in that they have been modified

to contain a moiety capable of performing a bioorthogonal reaction. A bioorthogonal

reaction will only react with the moiety inside of cells and be inert to all other cellular

components.[30] Classic examples of bioorthogonal nucleosides are 4-thiouridine and 5-

ethynyluridine, as seen in Figure 3.[22, 31]
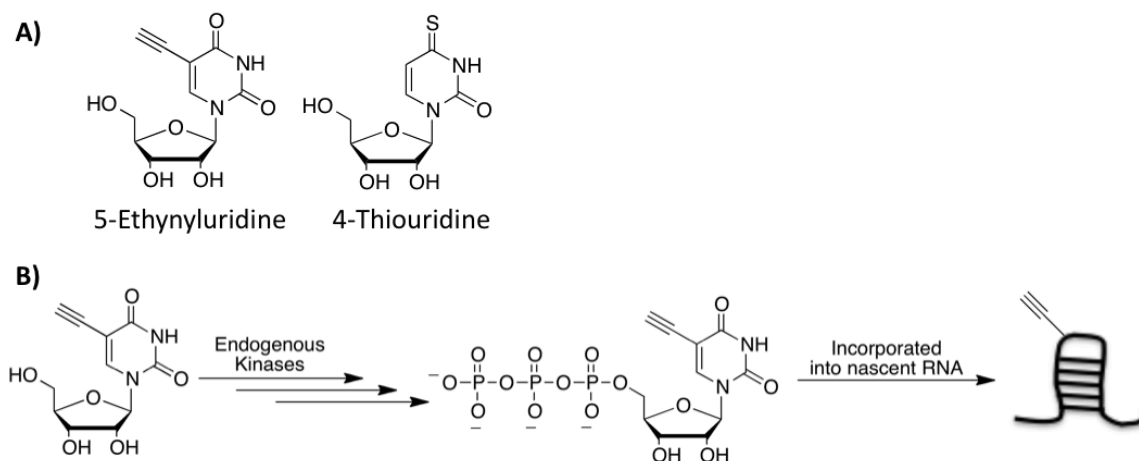
Figure 3. A) Bioorthogonal nucleoside probes 5-ethynyluridine and 4-thiouridine. B) Schematic of 5-
ethynyluridine processing by endogenous kinases and incorporation into nascent RNA. 4-thiouridine follows
the same process.

4-Thiouridine (4-SU) has a sulfur at the 4-position of uridine, and acts as a uridine

mimetic, in that it is recognized and processed by the kinases of the uridine salvage

pathway, and ultimately used by RNA polymerase to be incorporated into nascent RNA.[31]

The thione can then be activated by 365nm wavelength light to generate a sulfur radical,

which is capable of forming cross-links to nearby substrates. This technique, termed PAR-

CLiP (Photoactivatable-Ribonucleoside-Enhances Crosslinking and Immunoprecipitation),

is used to identify RNA binding proteins by crosslinking the 4-SU containing RNA to the

protein bound to it.[32] This technique is limited in that it requires the use of known protein

target so that the crosslinked product can be immunopurified.[31, 32] 5-Ethynyluridine (5-

EUd) is another uridine mimetic capable of being incorporated into nascent RNA, as seen in Figure 3B.[22] The alkyne at the 5-position can be used for the bioorthogonal click chemistry reaction of copper azide-alkyne cycloaddition (CuAAC).[33] By doing so, a fluorophore or biotin handle can be appended to 5-EUd containing RNA for visualization or enrichment, respectively. Since then, a myriad of bioorthogonal RNA nucleosides have been developed and utilized for nascent RNA labeling with functionalities that extend beyond CuAAC reactions. As mentioned previously, the traditional pair is azide alkyne, which undergoes CuAAC. But copper is toxic and can generate radicals that can degrade RNA.[34] Thus strain promoted azide-cyclooctyne cycloaddition (SPAAC) click chemistry was developed, in which the strain energy of the cyclooctyne is enough to promote the cycloaddition without the need of a copper catalyst.[35] This can be paired with azide containing nucleosides, such as 2'-azidoadenosine, and thus avoid the use of toxic copper for RNA click chemistry.[36] In addition to azide and alkyne containing nucleosides, vinyl nucleosides have also been developed, such as 5-vinyluridine and 2-vinyladenosine.[37, 38] The vinyl group can be accessed via inverse-electron-demand Diels-Alder reactions, where the vinyl group acts as the diene and reacts with a dienophile conjugated to a fluorophore or biotin.[38] Dienophiles include norbornenes and tetrazines.[37, 39, 40]

Furthering methods to track RNA expression have led to the additional challenge to control RNA labeling in a cell-specific manner. As probes act as natural nucleoside mimetics, they lack cell-specificity and can be incorporated into the nascent RNA. One approach to control cell specificity is to design, synthesize, and test chemical modifications on nucleoside analogs that limit their processing by kinases in normal cells. Pairing the specific modification with an enzyme that both recognizes and processes it would impart

cell-specificity. The additional enzyme would catalytically activate the chemically modified

nucleoside into the substrate for kinase binding. This system allows for cell-specific

labeling by expressing the enzyme in a species where the enzyme homolog is inactive or

does not exist. This methodology is termed RNA labeling by biosynthetic tagging (RABT).[41,] [42, 43] The basis of this is similar to enzyme prodrug therapy, in which inert prodrugs are

administered, and a corresponding enzyme will recognize and catalytically cleave it into

the active drug.[44] The first example of applying this system to cell-specific RNA labeling

was done with 2-thiouracil, 4-thiouracil and 2,4-thiouracil paired with the enzyme uracil

phosphoribosyltransferase (UPRT) from the bacteria *Toxiplasma gondii,* as can be seen in

Figure 4.[41]
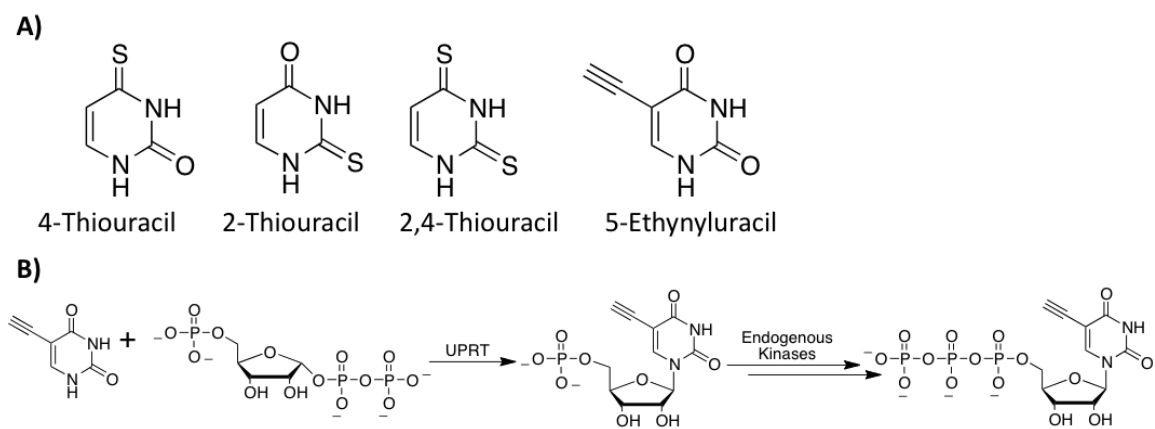
Figure 4. A) Cell-specific bioorthogonal nucleoside probes 4-thiouracil, 2-thiouracil, 2,4-thiouracil, and 5-ethynyluracil. B) Schematic of the UPRT processing of 5-ethynyluracil with phosphoribosylpyrophosphate (PRPP) into 5-ethynyluridine monophosphate, which can be accepted by endogenous kinases. The thiouracil analogs also undergo this processing with UPRT.

UPRT is part of the pyrimidine salvage pathway and acts through converting uracil

into uridine monophosphate (UMP) with phosphoribosylpyrophosphate (PRPP), which can

then be further kinased and incorporated into nascent RNA.[45] The human homolog of UPRT

does not actively play a role in the pyrimidine salvage pathway, thus the bacterial version can be expressed in mammalian cells as an exogenous enzyme, unique to whichever cell expresses it.[42] As with the previous 4-thiouridine acting as a uridine mimetic, 4-thiouracil is a mimetic for uracil, thus allowing it to be recognized by UPRT and converted into 4-thiouridine monophosphate. This system was further developed to utilize 5-ethynyluracil to label cell-specific RNA with a click handle.[42] Similar to 4-thiouracil, 5-ethynyluracil was able to serve as a uracil mimetic for UPRT and be converted into 5-ethynyluridine monophosphate, as can be seen in Figure 4B. For both 4-thiouracil and 5-ethynyluracil, if UPRT was not expressed, then neither probe could be incorporated into nascent RNA. In addition, an alternative version of cell-specific RNA labeling was developed with the bacterial enzyme penicillin G amidase (PGA).[42] PGA is a bacterial enzyme responsible for catalyzing penicillin into 6-aminopenicillanate by specifically recognizing the phenylacetyl group and cleaving the amide bond.[46] The combination of a phenylacetyl protecting group on an amine and the PGA enzyme has been used before as a cell selective system for other biomolecules.[47, 48] In order to apply this system to cell-specific RNA labeling, the phenylacetyl group was appended to the exocyclic amine of 2'-azidoadenosine to yield N^6-phenylacetyl-2'-azidoadenosine (PAC-2'N$_3$A). PAC-2'N$_3$A was too sterically bulky to be recognized by the endogenous kinases when administered to cells and thus was not incorporated into nascent RNA. However, after transfection of the PGA enzyme, the cells were able to catalyze PAC-2'N$_3$A into 2'-N$_3$A and subsequently kinase and incorporate the probe into nascent RNA.[42] These enzyme analog pairs allow for real-time cell-specific transcriptome analysis and they can be applied to better understand how two different cell types talk to one another at the transcriptomic level.

9

My dissertation is divided into two areas of RNA: 1) RNA structure probing and 2) RNA labeling with nucleoside analogs. Chapters 3 will cover the section of RNA structure probing, in which I synthesized an RNA structure probe, 2-azidomethyl-3-furoic acid imidazolide (FAI-N_3) for enhanced icSHAPE. FAI-N_3 was found to form more stable 2'-adducts on RNA than the previous icSHAPE probe NAI-N_3. Chapters 2 and 4 will cover the section of RNA labeling with nucleoside analogs, in which I synthesized bioorthogonal nucleosides capable of being incorporated into nascent RNA; ranging from general cell labeling with the first azide containing RNA nucleosides to extending the PGA and phenylacetyl group system to include pyrimidine nucleosides for cell-specific labeling. Combined, these new methods for RNA structure probing and cell-specific RNA labeling will greatly enhance the RNA toolkit.

1.2 References

1. Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, 227, 561-563.

2. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, 431, 931-945.

3. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, 489, 57-74.

4. K. V. Morris & J. S. Mattick. The rise of regulatory RNA. *Nat. Rev. Genet.* **2014**, 15, 423-437.

5. D. Marchese, et al. Advances in the characterization of RNA-binding proteins. *Wiley Interdiscip. Rev.* **2016**, 7, 793-810.

6. A. V. Sherwood & T. M. Henkin. Riboswitch-mediated gene regulation: Novel RNA architectures dictate gene expression responses. *Annu. Rev. Microbiol.* **2016**, 70, 361-374.

7. J. M. Engreitz, et al. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* **2014**, 159, 188-199.

8. F. Crick. The origin of the genetic code. *J. Mol. Biol.* **1968**, 38, 367-379.

9. Q. Jing, et al. Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell* **2005**, 120, 623-634.

10. V. Bernat ° M. Disney. RNA structures as mediators of neurological diseases and as drug targets. *Neuron* **2015**, 87, 28-46

11. H. Lodish, et al. Molecular Cell Biology, 4th edition **2000**.

12. T. Cooper, L. Wan, & G. Dreyfuss. RNA and disease. *Cell* **2009**, 136, 777-793.

13. L. Cartegni, et al. Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am. J. Hum. Genet.* **2006**, 78, 63-77.

14. G. Meister, C. Eggert, & U. Fischer. SMN-mediated assembly of RNPs: a complex story. *Trends Cell. Biol.* **2002**, 12, 472-478.

15. A. Bhan & S. Mandal. Long noncoding RNAs: emerging stars in gene regulation, epigenetics and human disease. *ChemMedChem* **2014**, 9, 1932-1956.

16. J. Day & L. Ranum. RNA pathogenesis of the myotonic dystrophies. *Neuromuscl. Discord* **2005**, 15, 15-16.

17. A. Pushechnikiv, et al. Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: Application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3. *J. Am. Chem. Soc.* **2009**, 131, 9767-9779.

18. P. Kaliman, et al. Myotonic dystrophy protein kinase phosphorylates phospholamban and regulates calcium uptake in cardiomyocyte sarcoplasmic reticulum. *J. Biol. Chem.* **2005**, 280, 8016-8021.

19. E. Merino, et al. RNA structure analysis at single nucleotide resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* **2005**, 127, 4223-4231.

20. J. Keene, J. Komisarow, & M. Friedersdorf. RIP-Chip: the isolation of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat. Protoc.* **2006**, 1, 302-307.

21. A. Buxbaum, B. Win, & R. Singer. Single β-actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science* **2014**, 343, 419-422.

22. C. Jao & A. Salic. Exploring RNA transcription and turnover *in vivo* by using click chemistry. *Proc. Natl. Acad. Sci.* **2008**, 105, 15779-15784.

23. L. Lempereur, et al. Conformation of yeast 18S rRNA. Direct chemical probing of the 5' domain in ribosomal subunits and in deproteinized RNA by reverse transcriptase mapping of dimethyl sulfate-accessible. *Nucleic Acids Res.* **1985**, 13, 8339-8357.

24. OH

25. M. Kubota, C. Tran, & R. Spitale. Progress and challenges for chemical probing of RNA structure inside of living cells. *Nat. Chem. Biol.* **2015**, 11, 933-941.

26. P. Tijerina, S. Mohr, & R. Russell. DMS footprinting of structured RNAs and RNA-protein complexes. *Nat. Protoc.* **2007**, 2, 2608-2623.

27. R. Karaduman, et al. RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. *J. Mol. Biol.* **2006**, 356, 1248-1262.

28. R. Spitale, et al. RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **2013**, 9, 18-20.

29. R. Spitale, et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **2015**, 519, 486-490.

30. E. Sletten & C. Bertozzi. Bioorthogonal Chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **2009**, 48, 6974-6998.

31. M. Hafner, et al. Trasncriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **2010**, 141, 129-141.

32. M. Hafner, et al. PAR-CliP – A method to identify transcriptome-wide the binding sites of RNA binding proteins. *J. Vis. Exp.* **2010**, 2034.

33. V. Rostovtsev, L. Green, V. Fokin, & B. Sharpless. A stepwise huisgen cycloaddition process: Copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **2002**, 67, 3057-3064.

34. E. Paredes & S. Das. Click chemistry for rapid labeling and ligation of RNA. *ChemBioChem* **2011**, 12, 125-131.

35. N. Agard, et al. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **2006**, 1, 644-648.

36. S. Nainar, et al. Metabolic incorporation of azide functionality into cellular RNA. *ChemBioChem* **2016**, 17, 2149-2152.

37. M. Kubota, et al. Expanding the scope of RNA metabolic labeling with vinyl nucleosides and inverse electron-demand Diels-Alder chemistry. *ACS ChemBio* **2019**, In Press.

38. J. George & S. Srivatsan. Vinyluridine as a Versatile Chemoselective Handle for the Post-transcriptional Chemical Functionalization of RNA. *Bioconjugate Chem*. **2017**, 28, 1529-1536.

39. H. Wu & N. Deveraj. Inverse electron-demand Diels-Alder bioorthogonal reactions. *Topics Curr. Chem.* **2016**, 374.

40. E. Brachet & P. Belmont. Inverse electron demand Diels-Alder (IEDDA) reactions: Synthesis of heterocycles and natural products along with bioorthogonal and material sciences applications. *Curr. Org. Chem.* **2016,** 20, 2136-2160.

41. M. Cleary, et al. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat. Biotech.* **2005**, 23, 232-237.

42. K. Nguyen, et al. Cell-selective bioorthogonal metabolic labeling of RNA. *JACS* **2017**, 139, 2148-2151.

43. N. Hida, et al. EC_tagging allows cell type-specific RNA analysis. *NAR* **2017**, 45, e138.

44. G. Xu & H. McLeod. Strategies for enzyme/prodrug cancer therapy. *Clin. Cancer Res.* **2001**, 7, 3314-3324.

45. M. Iltzsch. Pyrimidine salvage pathways in *Toxoplasma gondii*. *J Eur. Microbiol.* **1993**, 40, 24-28.

46. G. Volpato, R. Rodrigues, & R. Fernandez-Lafuente. Use of enzymes in the production of semi-synthetic penicillins and cephalosporins: drawbacks and perspectives. *Curr. Med. Chem.* **2010**, 17, 3855-3873.

47. D. Kadereit & H. Waldmann. Enzymatic protecting group techniques. *Chem. Rev.* **2001**, 101, 3367.

48. R. Barrett, et al. Cell-specific profiling of nascent proteomes using orthogonal enzyme-mediated puromycin incorporation. *ACS Chem. Biol.* **2016**, 11, 1532-1536.

Chapter 2: Metabolic incorporation of azide containing RNA nucleosides

2.1 Introduction

Recent analysis of complex transcriptomes has revealed that genomes are transcribed to generate diverse coding and non-coding RNAs that are critical for cell survival and identity.[1] The functions of both coding and non-coding RNAs continue to be elucidated.[2] As such, biochemical methods to track RNA transcription, posttranscriptional regulation, and RNA-based mechanisms that control their cellular function are in high demand.

Modified nucleoside analogs have been used to interrogate many facets of RNA biology. 4-thiouridine (4SU) has been employed to track nascent transcription and monitor RNA decay.[3] However, recent evidence has suggested that the transient nature of disulfide bonds can bias RNA enrichment.[4] Extending beyond thiol-modified nucleosides introduces additional analytical properties, such as enrichment with stable covalent chemistry, imaging, and multiplex tracking. This can be accomplished through dosing of analogs containing diverse chemical functionalities. The analog 5-ethynyluridine (5EU) has been used to track transcription and RNA localization via fluorescent imaging facilitated by copper-catalyzed azide-alkyne cycloaddition (CuAAC).[5] 2'-azidonucleosides have proven useful for analysis of RNA produced in vitro by chemical synthesis.[6] N^6-propargyl as well as C2- and C7-ethynyl adenosine have also been demonstrated to be useful probes for metabolic labeling of transcription and polyadenylation.[7] Despite this progress, a holistic description of the types of analogs that can be utilized to track RNA synthesis and processing inside living cells remains to be systematically interrogated.

Although useful, installing alkyne-modified nucleosides into cellular RNA requires the use of CuAAC reactions, which produce copper-induced radicals that degrade RNA.[8]

Such degradation can lead to deleterious effects on downstream analyses such as RNA sequencing.[9] As such, there is a critical need to expand the bioorthogonal toolkit for cellular RNA by endowing substrates with more versatile functionalities.

Azides are perhaps the most widely utilized among the long list of bioorthogonal functional groups used in cells. Azide-containing molecules can be probed through diverse chemical reactions including both CuAAC[10] and copper-free strain-promoted azide-alkyne cycloadditions (SPAAC),[11] as well as Staudinger ligations.[12] Metabolic labeling with azide-functionalized sugars has been a gold standard for studying glycosylated proteins on the cell surface, which has revealed the importance of the glycocalyx in cancer and development.[12] Azide-modified unnatural amino acids have been used to track nascent protein synthesis and have revealed the intricacies of cell-type specific translation.[13] These examples underscore just a few of the powerful techniques made possible by functionalizing endogenous biomolecules with azide handles.

Installing azide functionality into cellular RNA would set the stage for parallel investigations greatly increasing our understanding of RNA biology and function. Nevertheless, the metabolic incorporation of azide functionality into cellular RNA has yet to be explored and reported.

Herein we report evidence that azido-nucleosides can be metabolically incorporated into cellular RNA. We further show there is preference for our adenosine analogs, while an azido-uridine analog is refractory to RNA incorporation. Our data also suggests that depending on the site of azide modification, the adenosine analogs can be selectively utilized for tracking either gene body transcription alone or gene body transcription and polyadenylation. By exploring the limitations and idiosyncrasies of different azido-

16

nucleosides, we can ascertain how they can be leveraged to expand the scope of bioorthogonal reactions for studying RNA biology within living cells.

2.2 Results and Discussion

We first incubated cells with chemically synthesized azido-nucleoside analogs for 12 h (synthetic schemes in supplementary information) and then isolated the total RNA (**Figure 2-1, A; 1-4**). In order to detect the azide group, we appended a biotin-alkyne via CuAAC. We then performed streptavidin Northern blotting to determine incorporation of azido-nucleosides into cellular RNA (**Figure 2-1, B**).

These results showed that azido-nucleoside analogs **2-3** were robustly incorporated into cellular RNA, while metabolic labelling with **4** was not detected (**Figure 2-1, C**). We examined the cytotoxicity of analogs **1-3** via an MTT assay[13] and observed no significant difference in cell viability of control cells as compared to cells treated with **1-3** at 1mM final concentration. (**Supplementary Figure 2-1**). The data in Fig. 1 suggests that the pathway for purine triphosphate biosynthesis may be quite flexible, as the adenosine analogs **1-3** have modifications on the nucleobase and the ribose sugar. Uridine analog **4,** which has a larger methyl-azido group at the 5-position, showed no detectable incorporation into RNA. **4** can smoothly react with biotin-alkyne and propargyl alcohol via CuAAC to form adducts (**Supplementary Figure 2-2**); furthermore, while analogs **1-3** are detected in cellular RNA by LC-MS, **4** is not (**Supplementary Figure 2-3 and Supplementary Table 2-1**). These results strongly suggest that **4** is not metabolically incorporated into cellular RNA. This result is in contrast to 5-ethynyluridine, which is routinely used for tracking RNA synthesis.[14]
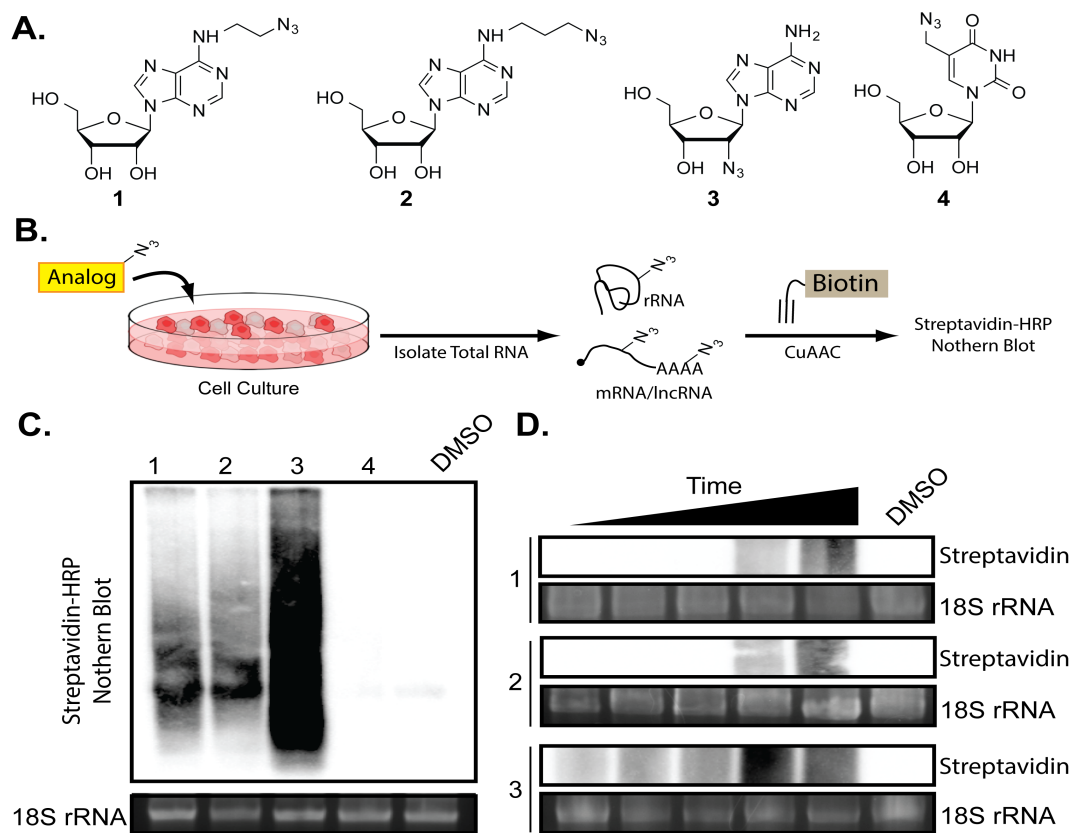
Figure 2-1. Screening of azido nucleosides. A. Structures of azido-nucleosides used in this study. B. Schematic of incubation and RNA processing protocol. C. Northern blot after 12 hr. incubation with **1-4** at 1mM. D. Time titration analysis after 1mM incubation with **1-3**.

The differences in relative signal prompted us to explore the kinetics of incorporation. We reasoned that the lack of incorporation may be due to slower kinetics or simply limited introduction into cellular RNA. We performed an incubation time titration, which revealed that adenosine analogs **1** and **2** were slowly incorporated into RNA, with unobservable signal until 5 h. In contrast, **3** was robustly integrated into cellular RNA within 30 min (**Figure 2-1, C; Supplementary Figure 2-4**). We also treated cells with increasing amounts of each analog (**Supplementary Figure 2-5)**, and the data showed that **3** is the most robustly incorporated at a concentration as low as 10μM. We posit that the differences in incorporation between the azido-nucleosides may be a function of the pathways and

various enzymes involved in ribonucleotide triphosphate synthesis (more discussion below). Overall, these results demonstrate that azido-adenosine analogs are amenable to metabolic incorporation into RNA and therefore present a new set of modified nucleosides that can be utilized for RNA labeling.

A key benchmark for introducing modified molecules into cells is their ability to be used as reporter molecules for imaging experiments. We tested whether incorporation of each analog could be tracked by cellular imaging using fluorophores. Cells were grown in the presence of **1**, **2**, or **3** and were subsequently fixed and reacted with alkyne-rhodamine via CuAAC (**Figure 2-2**). Imaging showed clear staining of the nucleolus for all three analogs, indicating robust incorporation into ribosomal RNA, which accounts for the majority of cellular RNA. Each analog also showed varying amounts of staining in the cytoplasm, likely the result of trafficking, where it was most pronounced for analog **3**. Consistent with our observations by Northern blot, we were also able to observe RNA incorporation for analog **3** at concentrations as low as 10 μM (**Supplementary Figure 2-6**). Treatment with hydroxyurea showed no difference in imaging signal; strongly suggesting that signal was derived from RNA and not DNA (**Supplementary Figure 2-7**). RNase digestion of fixed cells also resulted in loss of signal, further demonstrating that signal is coming from RNA labeling (**Supplementary Figure 2-8**).

Azides present a unique opportunity to be probed by less-damaging copper-free chemistry. To test our azido-nucleosides inside living cells, we incubated cells with analogs **1-3** and dibenzylcyclooctyne (DIBO) biotin for SPAAC reactions.[11] From Northern blotting, we concluded that bioorthogonal reactions with azide-modified RNA can be performed in living cells (**Supplementary Figure 2-9**). Furthermore, SPAAC imaging experiments

19

demonstrated that azide-labeled RNA can be robustly imaged with copper-free cycloadditions (**Supplementary Figure 2-10**). Such observations demonstrate that nascent RNAs can be tracked *in situ* while eliminating the cytotoxic effects of copper-induced radical formation. Overall, our data supports azide-modified adenosine analogs are useful for labeling RNA by CuAAC as well as SPAAC.
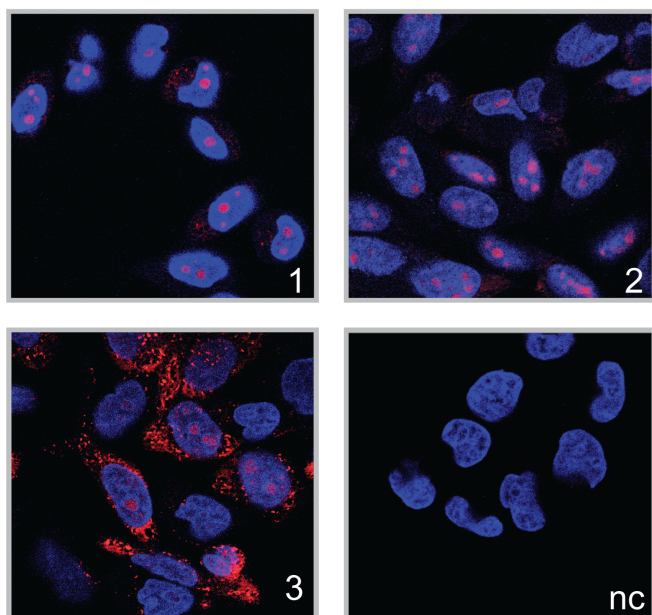
Figure 2-2. Incorporation of adenosine analogs into cellular RNA interrogated by CuAAC fluorescent imaging. Numbering corresponds to the same as Figure 2-1. nc= negative control, no analog.

The differing rates of incorporation observed in Figure 2-1 suggest that there may be selectivity for incorporation into specific types of cellular RNA. It has previously been demonstrated that triphosphate analogs baring a wide variety of modifications are well tolerated by a variety of polymerases.[14] PolyA polymerase selectivity has been investigated *in vitro*,[7c, 8b] but has yet to be fully explored inside living cells. Adenosine analogs with alkyne groups at the N^6-, C2- and C7- position have been demonstrated to be

acceptable substrates.[7a, 7c, 15] However, larger azide modifications at the 2'- and N^6-positions have not been investigated.

To gain insight into the possibility of substrate selectivity by poly(A) polymerase, we analyzed the crystal structure of the mammalian enzyme bound to adenosine triphosphate (PDB:1F5A).[16] Amino acid residues tightly surround the N^6 position of adenosine. By contrast, the 2'-position is relatively open to solvent. We therefore speculated that the different sites of azide modification may impart selectivity for either transcription by polymerases and/or polyadenylation by polyA polymerase.

To further explore polymerase selectivity, we challenged the incorporation of all three analogs with actinomycin D (ActD, an inhibitor of RNA polymerase I and II) and/or cordycepin (Cdy, an inhibitor of RNA polyA polymerase) at concentrations known to inhibit both processes.[7a, 15] Analogs **1** and **2**, which contain N^6 alkyl chains, showed strong inhibition by ActD treatment, while showing limited decrease in signal when treated with Cdy (**Figure 2-3, B, Supplementary Figure 2-11**). In contrast, compound **3** signal decreased but was not completely abolished when challenged with either Cdy and ActD individually, or both inhibitors. This data suggests that analogs **1** and **2** are predominantly incorporated during transcription, while analog **3** is robustly incorporated into RNA by both nascent RNA synthesis and polyadenylation (**Figure 2-3, B**). We anticipate that such differences could be exploited to track nascent RNA synthesis and polyA elongation differentially, solely through analog selection. This will provide unique tracking capabilities for RNA synthesis inside living cells.

A.

B.
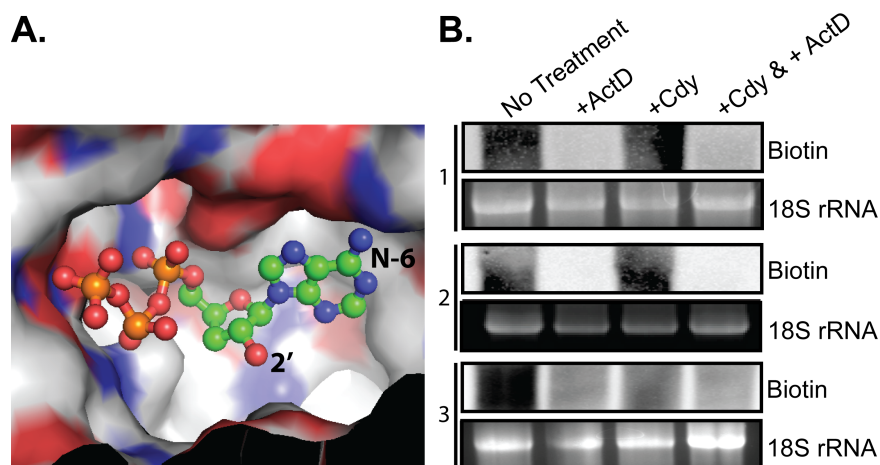
Figure 2-3. Nucleoside analogs are differentially incorporated into cellular RNA. A. PDB model of mammalian poly(A) polymerase. B. Northern blot analysis of polymerase inhibition and analog incorporation. ActD = Actinomycin, Cdy = Cordecypin.

2.3 Conclusion

We have demonstrated that azide-modified nucleoside analogs can be metabolically incorporated into cellular RNA. We have also observed that upon incorporation, our analogs are responsive to bioorthogonal ligations mediated by CuAAC as well as SPAAC, which is amenable to live-cell analysis. Overall, our data supports the use of this novel set of azido-nucleosides for labeling and monitoring nascent RNA. These analogs are sure to find wide-spread utility in the RNA community and beyond. Modified nucleosides have thus far been critical toward understanding the rate of RNA synthesis and turnover,[17] characterizing the RNA-binding proteome,[18] and even tracking viral infection.[19] As such, our azide-modified nucleosides are poised to expand the arsenal of methods used to analyze RNA in cells as well as in complex environments such as tissues and living animals.

Expanding the bioorthogonal toolkit for RNA requires further investigation of the limitations and selectivity for modified nucleosides. Our data has demonstrated there is selectivity in both upstream biosynthesis and downstream synthesis of mature

polyadenylated transcripts. The observation that 5-methylazidouridine (**4**) is not incorporated into cellular RNA compels us to cautiously posit that this difference may arise from the first phosphorylation step in pyrimidine triphosphate synthesis. Indeed, it has been demonstrated with other cytidine analogs that the first phosphorylation step can be rate limiting.[20] A comparison analysis of the adenosine kinase and uracil/cytidine kinase crystal structures revealed that adenosine with modifications at the 2'- and N^6- positions should indeed be an amenable substrate. (**Supplementary Figure 2-12**). In contrast, the 5-position of uridine forms a tight Van der Waals interaction with the enzyme and is relatively compact. This structural difference may account for the observed selectivity, which presents an exciting opportunity for enzyme engineering to introduce additional modifications into cellular RNA. Such investigations are currently underway.

2.4 Acknowledgements

2.5 Methods

1. General

All reagents were purchased from commercial suppliers and were of analytical grade and used without further purification unless otherwise noted. 2'-azidoadenosine was purchased from TriLink BioTechnologies. Reaction progress was monitored by thin-layer chromatography on EMD 60 F254 plates, visualized with UV light, iodine, ninhydrin, $KMnO_4$, $FeCl_3$, p-anisaldehyde, 2,4-DNP, and bromocresol green stains. Compounds were

purified via flash column chromatography using Sorbent Technologies 60 Å 230 x 400 mesh silica gel. Anhydrous solvents acetonitrile (MeCN), dichloromethane (DCM), methanol (MeOH), tetrahydrofuran (THF), dimethylformamide (DMF) were degassed and dried over molecular sieves. Acetone was dried over $MgSO_4$. All reaction vessels were flame dried prior to use. NMR spectra were acquired with Bruker Advanced spectrometers. All spectra were acquired at 298 K. ^{1}H-NMR spectra were acquired at 400 MHz and 500 MHz. ^{13}C-NMR spectra were acquired at 125 MHz. Chemical shifts are reported in ppm relative to residual non-deuterated NMR solvent, and coupling constants (J) are provided in Hz. All NMR spectra was analyzed using MestreNova software. Low and high-resolution electrospray ionization (ESI) mass spectra and Gas Chromatography mass spectra were collected at the University of California-Irvine Mass Spectrometry Facility. IR spectra were acquired from neat samples, unless otherwise noted, with a PerkinElmer Spectrum Two IR Spectrometer.

2. Synthesis

2-azidoethan-1-amine (1a)

To a stirred solution of 2-chloroethylamine hydrochloride (4.83 mmol, 1 eq) in H_2O (5 mL) was added sodium azide (14.5 mmol, 3 eq) and the suspension was then heated to 80 °C. After 15 h, KOH pellets were added to basify the solution, followed by extraction with diethyl ether. The combined organic phases were dried over $MgSO_4$ and concentrated *in vacuo* to give the desired amine as a colorless volatile oil (415 mg, quant.). HRMS Calcd for $C_2H_6N_4$ [M+H]$^+$ 87.07, found 87.0668 [M+H]$^+$; ^{1}H NMR (500 MHz, CDCl$_3$-TMS): δ (ppm) 3.42

(t, J = 5.5 Hz, 2H), 2.94 (t, J = 5.8 Hz, 2H); ^{13}C NMR (125 MHz, CDCl$_3$-TMS): δ (ppm) 54.79, 41.49.
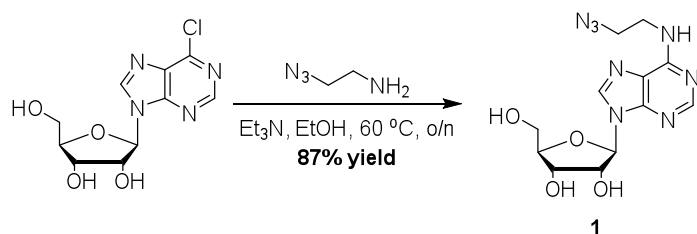
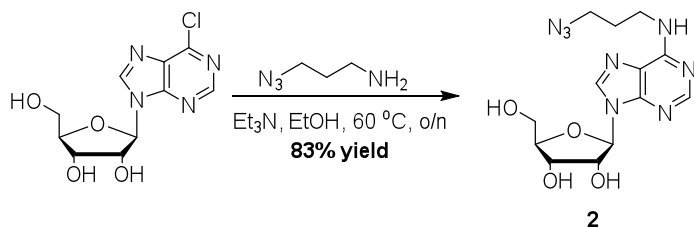**(2*R*,3*R*,4*S*,5*R*)-2-(6-((2-azidoethyl)amino)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol (1)**

1a (4.83 mmol, 4 eq) was dissolved in 2 mL of ethanol followed by addition of TEA (4.83 mmol, 4 eq). The suspension was allowed to stir at room temp for 5 min, a solution of 6-chloro-9-(b-D-ribofuranosyl)purine (1.21 mmol, 1 eq) in 20 mL of ethanol was added. The suspension was then heated to 60 °C and stirred overnight. The reaction was allowed to cool then concentrated *in vacuo* and the crude product was then recrystallized in ethanol as a white solid (354 mg, 87%). HRMS Calcd for C$_{12}$H$_{16}$N$_8$O$_4$Na 359.12 [M+Na]$^+$, found 359.1201 [M+Na]$^+$; ^{1}H NMR (500 MHz, d_6-DMSO): δ (ppm) 8.39 (s, 1H), 8.23 (s, 1H), 8.09 (s, 1H), 5.9 (d, J=6.2MHz, 1H), 5.46 (d, J=6.2MHz, 1H), 5.39 (s, 1H), 5.20 (d, J=4.5MHz, 1H), 4.62 (dd, J=6, 5MHz, 1H), 4.15 (dd, J=4, 3MHz), 3.97 (dd, J=3.5, 3MHz, 1H), 3.8-3.63 (broad, 3H), 3.6-3.5 (broad, 3H); ^{13}C NMR (125 MHz, d_6-DMSO): δ (ppm) 155.07, 152.77, 148.94, 140.52, 120.36, 88.37, 86.38, 71.13, 62.12, 50.17.

N6-propylazido-adenosine (2)

2

(2*R*,3*R*,4*S*,5*R*)-2-(6-((3-azidopropyl)amino)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol (2)

3-azidopropan-1-amine[1] (415 mg, 4.83 mmol) was dissolved in 2 mL EtOH followed by addition of triethylamine (0.67 mL, 4.83 mmol). The suspension was allowed to stir at room temperature for 5 min, and then 6-chloroadenosine (394 mg, 1.21 mmol) dissolved in 20 mL of EtOH was added to the suspension. The suspension was heated to 60 °C and stirred overnight. The reaction was allowed to cool and then concentrated *in vacuo* and the crude product was recrystallized in EtOH to give a white solid (352 mg, 83%). HRMS Calcd for $C_{13}H_{18}N_8O_4Na$ 373.13 [M+Na]$^+$, found 373.1363 [M+Na]$^+$; ^{1}HNMR (500 MHz, *d$_6$*-DMSO): δ (ppm) 8.38 (s, 1H), 8.24 (s, 1H), 8.02 (s, 1H) 5.91 (d, *J* = 6.5 Hz, 1H), 5.47 (d, *J* = 6.5 1H) 5.42 (dd, *J* = 4.5, 2.5 Hz, 1H), 5.21 (d, *J* = 4.5 Hz), 4.63 (q, *J* = 6 Hz, 1H), 4.17 (dd, *J* = 5, 3 Hz, 1H), 3.99 (q, *J* = 3.5 Hz, 1H), 3.70 (m, 1H), 3.58 (m, 2H), 3.45 (t, *J* = 7 Hz, 2H), 1.88 (m, 2H); ^{13}C NMR (125 MHz, *d$_6$*-DMSO): δ (ppm) 155.2, 152.8, 148.8, 140.3, 88.4, 86.4, 73.9, 71.1, 62.15, 49.01, 37.6, 28.8.
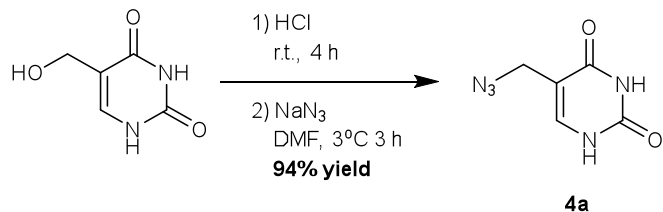
5-(azidomethyl)pyrimidine-2,4(1H,3H)-dione (4a)

A solution of 5-hydroxymethyluracil (2.00 g, 14.1 mmol) and concentrated HCl (25 mL) was stirred at r.t. for 4 h. The suspension was then put on ice for 30 min and the off white precipitate filtered over a glass fritted filter. The precipitate was allowed to dry and then dissolved in DMF (50 mL). Sodium azide (1.01 g, 15.5 mmol) was added to the solution. The suspension was stirred in the dark at 3 ºC for 3 h. Proceed next steps with caution due to formation of hydrazoic acid. Concentrated HCl (1 mL) was added to the suspension and concentrated *in vacuo* to give a white solid. The solid was then washed with cold water and filtered to give **4a** as a white solid (2.22 g, 94%). HRMS Calcd for $C_5H_5N_5O_2$ [M-H]⁻ 166.04, found 166.03 [M-H]⁻; ¹H NMR (400 MHz, DMSO) δ 11.23 (s, 1H), 11.05 (s, 1H), 7.61 (d, *J* = 5.7 Hz, 1H), 3.98 (d, *J* = 7.3 Hz, 2H); ¹³C NMR (125 MHz, DMSO) δ 164.45, 151.65, 142.48, 107.11, 46.99.
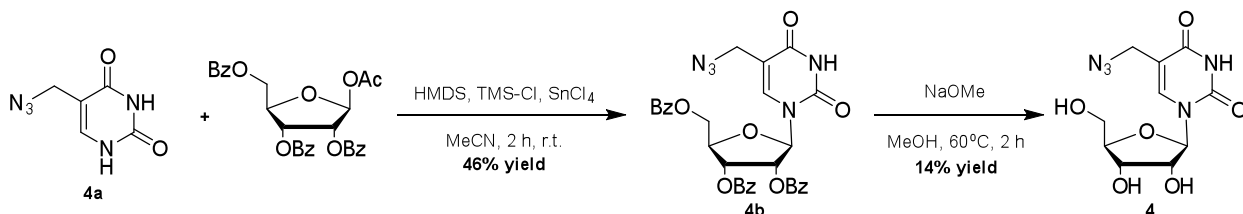
5-(azidomethyl)-1-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)pyrimidine-2,4(1H,3H)-dione (4)

The glycosylated product was prepared according to a modified procedure by Vorbrűggen, H. *et al.* and Rao, H. *et al.*[2,3] To a suspension containing **4a** (1.00 g, 5.98 mmol), 1-O-acetyl-

2,3,5-tri-O-benzoyl-β-D-ribofuranose (3.02 g, 5.98 mmol) in dry MeCN (75 mL) was added

HMDS (1.00 mL, 4.78 mmol) and TMS-Cl (0.610 mL, 4.78 mmol) and was stirred for 30 min

at r.t. and under Ar. $SnCl_4$ (0.840 mL, 7.18 mmol) was dissolved in dry MeCN (25 mL). The

solution of $SnCl_4$ was then cannulated into the suspension containing **4a** and stirred for 2 h.

DCM (30 mL) was added to the suspension and then washed with a sat. $NaHCO_3$ solution.

The organic layer was dried over $MgSO_4$, concentrated *in vacuo*, and purified by flash

column chromatography (2:5 EtOAc:hexanes) to give **4b** as a white foam (1.69 g, 46%). The

crude product was taken into the next reaction without any further purification. A

suspension containing **4b** (1.75 g, 2.87 mmol) and NaOMe (6.66 g, 0.12 mol) in MeOH (250

mL) was heated to 60°C and stirred for 2 h. 2 N HCl was added to the suspension until

neutral. The solution was then concentrated *in vacuo*, dissolved in EtOH, and filtered. The

filtrate was concentrated *in vacuo* and purified by flash column chromatography (1:10

MeOH:DCM) to give **4** as a white solid (0.12 g, 14%). HRMS Calcd for $C_{10}H_{13}N_5O_6Na$

[M+Na]$^+$ 322.08, found 322.07 [M+Na]$^+$; ^{1}H NMR (400 MHz, MeOH) δ 8.16 (s, 1H), 5.87 (t, *J*

= 4.9 Hz, 1H), 4.21 – 4.14 (m, 2H), 4.05 (d, *J* = 14.6 Hz, 2H), 4.03 – 3.98 (m, 1H), 3.87 (dd, *J* =

12.3, 2.6 Hz, 1H), 3.74 (dd, *J* = 12.3, 2.9 Hz, 1H); ^{13}C NMR (125 MHz, CD_3OD) δ 163.79,

150.92, 140.19, 109.03, 89.48, 84.90, 74.51, 69.68, 60.68, 46.96.

3. Biochemical Methods

Azido-nucleoside labeling of cellular RNA.

HeLa cell lines were cultured in DMEM supplemented with 10% FBS, penicillin and

streptomycin and grown at 37°C, 5% CO_2. Azido-nucleoside analogs were added to

complete culture medium from 200 mM stocks with a final concentration of <1% DMSO.

RNA isolation and biotinylation via CuAAC.

After labeling, total cellular RNA was harvested using Trizol Reagent (Invitrogen) following the manufacturers instructions. Click reactions were prepared using 10 μg of total RNA, 1 mM biotin alkyne, and fresh 4.6 mM THPTA to a final concentration of 1 mM, fresh 10.6 mM NaAsc to a final concentration of 1.77 mM, and 12 mM CuSO$_4$ to a final concentration of 200 μM. The reactions were incubated with shaking at room temperature for 30 min. The reactions were purified using Zymo RNA clean and concentrator-5 spin column according to the manufacturers instructions, with RNA eluted in 7 μL of nuclease free water.

HRP-streptavidin Northern blotting.

All gel reagents were from Bio-Rad. Equal amounts of column-purified RNA were loaded using nucleic acid loading dye and were separated on a native 1% agarose gel. RNA was transferred onto Hybond-N+ membrane (GE Healthcare) using standard capillary transfer. Membrane was blocked followed by incubation with high sensitivity streptavidin-HRP (ThermoFisher Scientific). The membrane was washed twice in a 1:10 solution of blocking buffer and twice in Tris-saline buffer. It was then incubated in SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher Scientific) and imaged on a ChemiDoc MP imaging system (Bio-Rad).

Transcription and polyadenylation inhibitor treatments.

Cells were simultaneously treated with compounds **1**, **2** or **3** and 10 μM actinomycin (VWR International) and/or 10 μM cordycepin (Sigma Aldrich) for 5 h. Total RNA was extracted and processed by Northern blot.

RNA fluorescence imaging via CuAAC.

Cells were seeded at 2.5×10^5 and grown to ~50% confluency on glass cover slips. Cells were treated with compounds **1**, **2** or **3** at various concentrations and incubated for 12 hours. After labeling, cells were washed twice with DPBS and fixed for 10 min at room temperature with 4% paraformaldehyde. Cells were then washed and "clicked" upon reaction with 500 μL of solution containing 15μM alkyne-rhodamine, 760μM THPTA ligand, 1.77mM NaAsc, and 500 μM $CuSO_4$, incubated for 1 hour at room temperature in the dark. Cells were washed twice with DPBS for 5 min each, and mounted using VectaShield with DAPI (Vector Labs). Slides were imaged via fluorescence confocal microscopy using a 63x oil immersion objective on a Leica 700 Carl Zeiss microscope.

RNA fluorescence imaging with ribonucleotide reductase inhibitor via CuAAC.

Cells were seeded at 2.5×10^5 and grown to ~50% confluency on glass cover slips. Cells were treated with compounds **1**, **2** or **3** at 1mM alone (control), or in tandem with 10mM hydroxyurea and were incubated for 12 hours. Cells were then fixed, stained, mounted and imaged as previously described.

In-cell RNA labeling via SPAAC.

Media was decanted from cells labeled with compounds **1**, **2** or **3.** Cells were treated with 50 μM DIBO-biotin (Life Technologies) for 2 hours at 37°C. Total RNA was extracted and processed by Northern blot.

RNA fluorescence imaging via SPAAC

Cells were seeded at 2.5×10^5 and grown to ~50% confluency on glass cover slips. Cells were treated with compounds **1**, **2** or **3** at 1mM and incubated for 12 hours. After labeling, cells were washed twice with DPBS and fixed for 10 min at room temperature with 4%

paraformaldehyde. Cells were washed with DPBS and then permeabilized with 0.2% Triton-X in DPBS for 10 min at room temperature. Cells were incubated with 500 μL of 50 μM DIBO-Alexa 488 (Life Technologies) in 1% FBS, DPBS for 2 hours at room temperature. The cells were subsequently washed, mounted and imaged as previously described.

MTT Assay.

HeLa cells were seeded into a 96-well plate and grown for 2 days. Cells were then incubated for 12 hours with compounds **1**, **2** or **3** at 1mM final concentration. Following incubation, cell viability was assessed via the cells' ability to reduce MTT (Life Technologies) to the insoluble formazan salt. Cell media was replaced with DPBS, and cells were incubated for 3 hours at 37°C with 1 mM MTT. Formazan was solubilized in DMSO and absorbance readings were taken at 540 nm.

Mass spec analysis of total RNA.

Experimental Protocol:

After extraction with TRIzol (Thermo Fischer Scientific) and purification with RNEasy column (Qiagen), total RNA samples (20 μg) were digested to nucleosides based on the method described by Hashimoto et al.[1] using a proprietary blend of nuclease(s) and phosphatase(s) (New England Biolabs). LC-MS analysis was performed by injecting a nucleoside standard mixture (rC, rU, rG, rA, and the azido nucleoside analogues 1, 2, 3, and 4) and digested RNA samples on an Agilent LC/MS System 1200 Series equipped with a G1315D diode array detector and a 6120 Single Quadrupole Mass Detector operating in positive (+ESI) and negative (-ESI) electrospray ionization modes. LC was carried out on a Waters Atlantis T3 column (4.6 × 150 mm, 3 μm) with a gradient mobile phase consisting of 10 mM aqueous ammonium acetate (pH 4.5) and methanol. MS data acquisition was

recorded in total ion chromatogram (TIC) mode. Each nucleoside was identified as follows: rC $[M+H]^+$ 244.1 and $[M-H]^-$ 241.9; rU $[M+H]^+$ 245.1 and $[M-H]^-$ 243.1; rG $[M+H]^+$ 284.1 and $[M-H]^-$ 282.0; rA $[M+H]^+$ 268.2 and $[M-H]^-$ 266.1; N^6-EtN$_3$A (1) $[M+H]^+$ 337.1 and $[M-H]^-$ 335.0; N^6-PrN$_3$A (2) $[M+H]^+$ 351.1 and $[M-H]^-$ 350.0; 2'N$_3$A (3) $[M+H]^+$ 293.1 and $[M-H]^-$ 291.0; and 5-MeN$_3$U (4) $[2M+H]^+$ 599.1 and $[2M-H]^-$ 597.2. The relative abundance of each nucleoside was determined by dividing the UV absorbance by the corresponding extinction coefficient (ε, L mol^{-1} cm^{-1}): rC 7,070 at 260 nm; rU 9,660 at 260 nm; rG 12,080 at 260 nm; rA 15,020 at 260 nm; N^6-EtN$_3$A (1) 15,588 at 267 nm; N^6-PrN$_3$A (2) 15,666 at 267 nm; 2'N$_3$A (3) 15,460 at 259 nm; and 5-MeN$_3$U (4) 9,285 at 264 nm. The relative abundance of each nucleoside was normalized to rA (set as 1). For the azido-nucleoside analogues 1, 2, and 3, the normalized relative abundance was further subtracted from the relative wild-type background values.

RNA fluorescence imaging with Rnase A via CuAAC.

Cells were grown and treated with analog as previously described. Following incubation, cells were washed and lightly fixed with 0.5% paraformaldehyde in 0.5% Triton-X PBS. Cells were washed twice with DPBS and incubated with 200 ug/mL of Rnase A in 0.5% Triton-X PBS at room temperature for 1 hour. Cells were then fixed, stained, mounted and imaged as previously described.

UV spectral analysis of SPAAC reaction of N$_3$-uridine and DIBO biotin.

Solutions of N$_3$-uridine, 2'N$_3$A, uridine, adenosine and DIBO biotin were made in DMSO to a final concentration of 667 μM. Reaction mixtures of each nucleoside and DIBO biotin were mixed 1:1 (667 μM) in DMSO and incubated for 0, 15 or 120 min. Absorbance readings were taken on a NanoDrop 2000 UV-Vis Spectrophotometer (ThermoFisher).

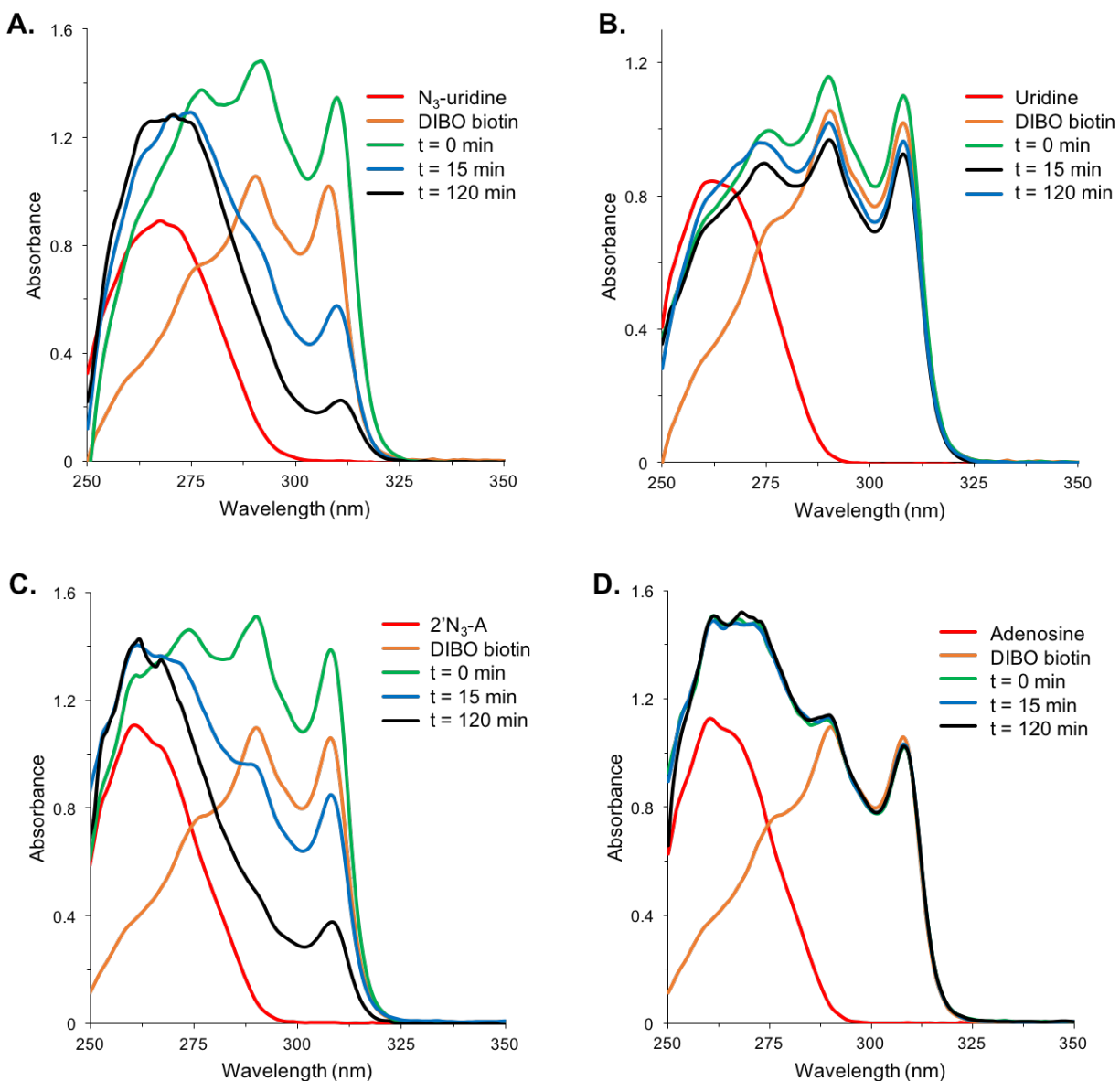**4. References**

[1]   Hashimoto H et al. *Nature* **2014**, *506*, 7488, 391-395.

2.6 Supplementary Data
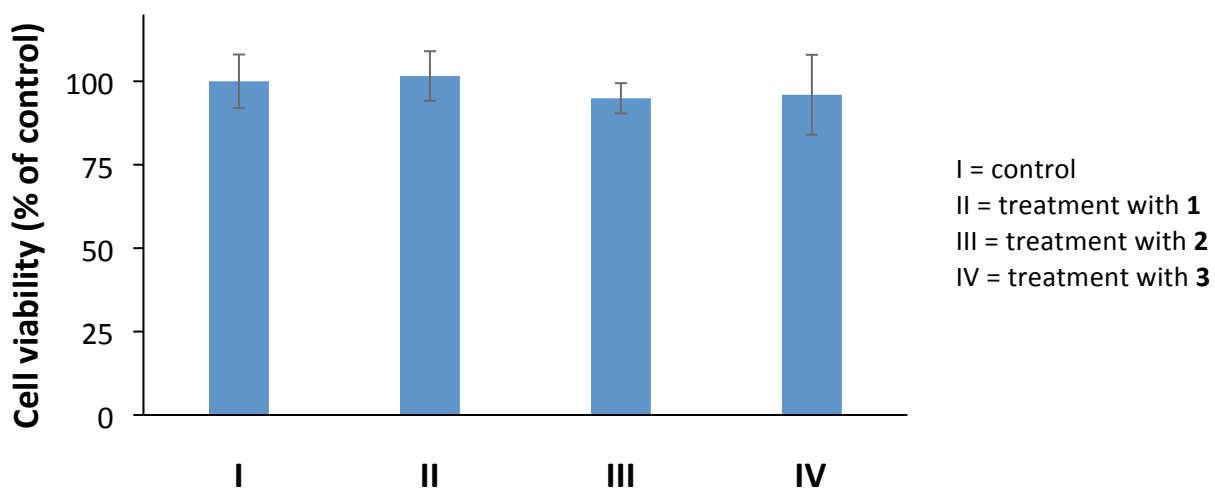
Supplementary Figure 2-1: MTT assay results following cell treatment with analogs 1-3 at

1mM concentration for 12 hours.

Supplementary Figure 2-2: UV spectra to confirm SPAAC reactivity of N_3-uridine. UV spectra of azido-nucleosides (A and C) or nucleosides (B and D) (667 μM), DIBO biotin (667 μM) and 1:1 reaction mixture (667 μM final) following incubation for 0, 15 or 120 min. DIBO peaks at 290 nm and 310 nm disappear in reactions with azide-containing nucleosides after SPAAC reactions form the triazole product.
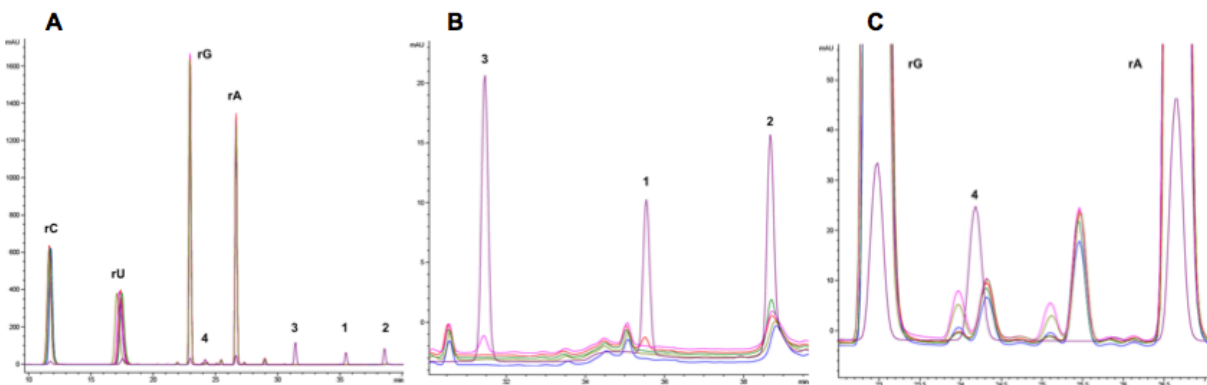
Supplementary Figure 2-3: LC-MS data demonstrating that analogs 1-3 are incorporated into cellular RNA, whereas 4 is not. (A) Ribonucleosides obtained from digestion of total RNA were resolved by reversed-phage HPLC and detected by UV absorption at 260 nm. Digested RNA from cells incubated with N[6]-ethylazido adenosine 1 (red), N[6]-propylazido adenosine 2 (green), and 2'-azido adenosine 3 (magenta), and 5-methylazido uridine 4 (light brown) are overlaid with digested RNA from wild-type cells (blue) and a nucleoside standard mixture containing rC, rU, rG, rA, and the azido-nucleoside analogues 1, 2, 3, and 4 (purple). Expansion of the overlaid chromatograms demonstrating (B) the specific incorporation of the azido-nucleoside analogues 1, 2, and 3 and (C) the lack of incorporation of 5-methylazido uridine 4.

rC	rU	rG	rA	N[6]-EtN$_3$A (1)	N[6]-PrN$_3$A (2)	2'N$_3$A (3)
1.91	1.00	1.66	1.00			
1.98	1.04	1.55	1.00	0.0013		
1.96	1.03	1.59	1.00		0.0033	
2.00	1.05	1.60	1.00			0.0014

Supplementary Table 2-1. Relative abundance of canonical and azido-nucleoside analogues in total RNA samples. The relative abundance of each nucleoside was normalized to rA (set as 1).
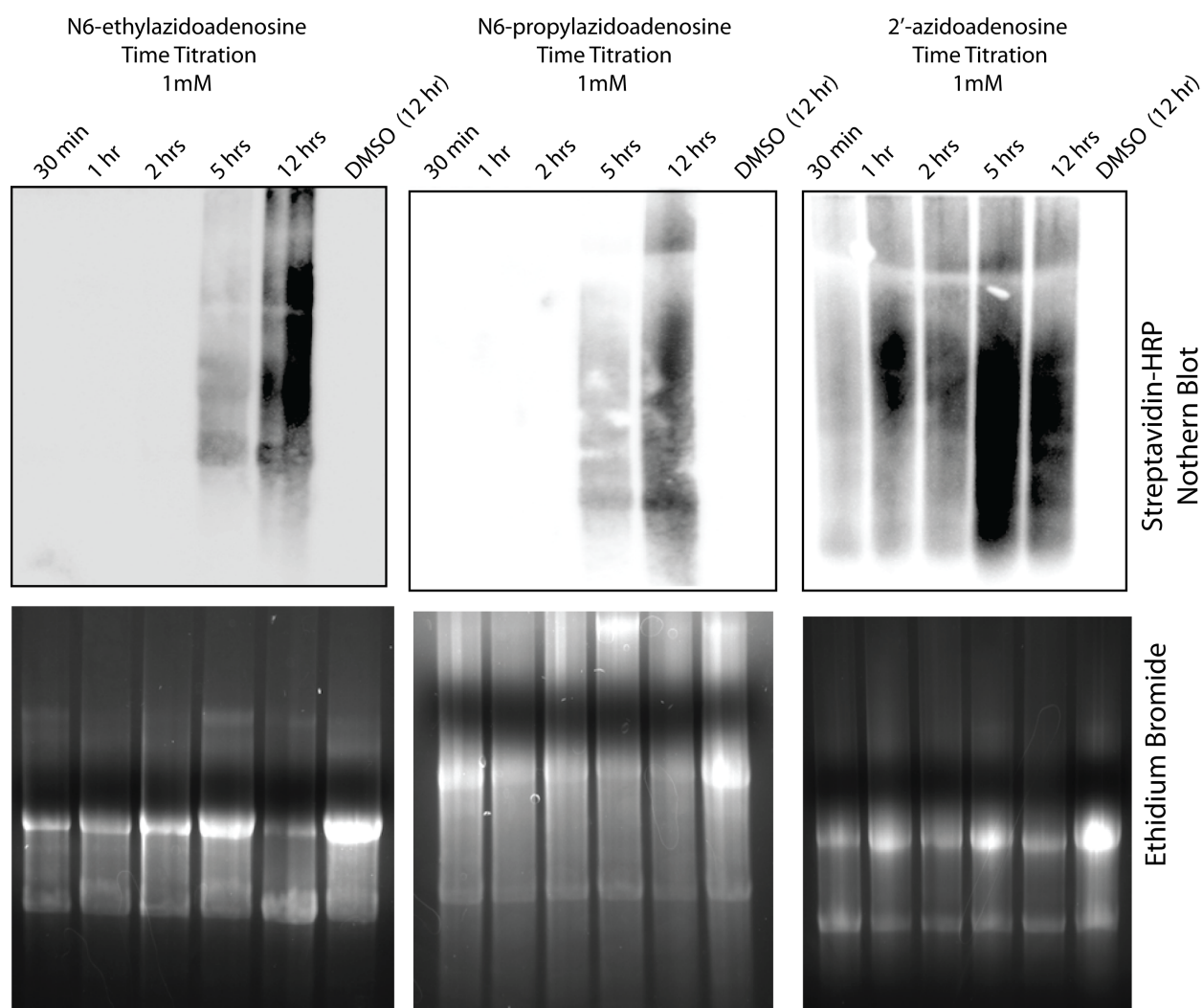
Supplementary Figure 2-4: Full Northern blots represented in Figure 2-1, C. A. Nothern blot for N6-ethylazidoadenosine. B. Nothern blot for N6-propylazidoadenosine. B. Northern blot for 2'-azidoadenosine.
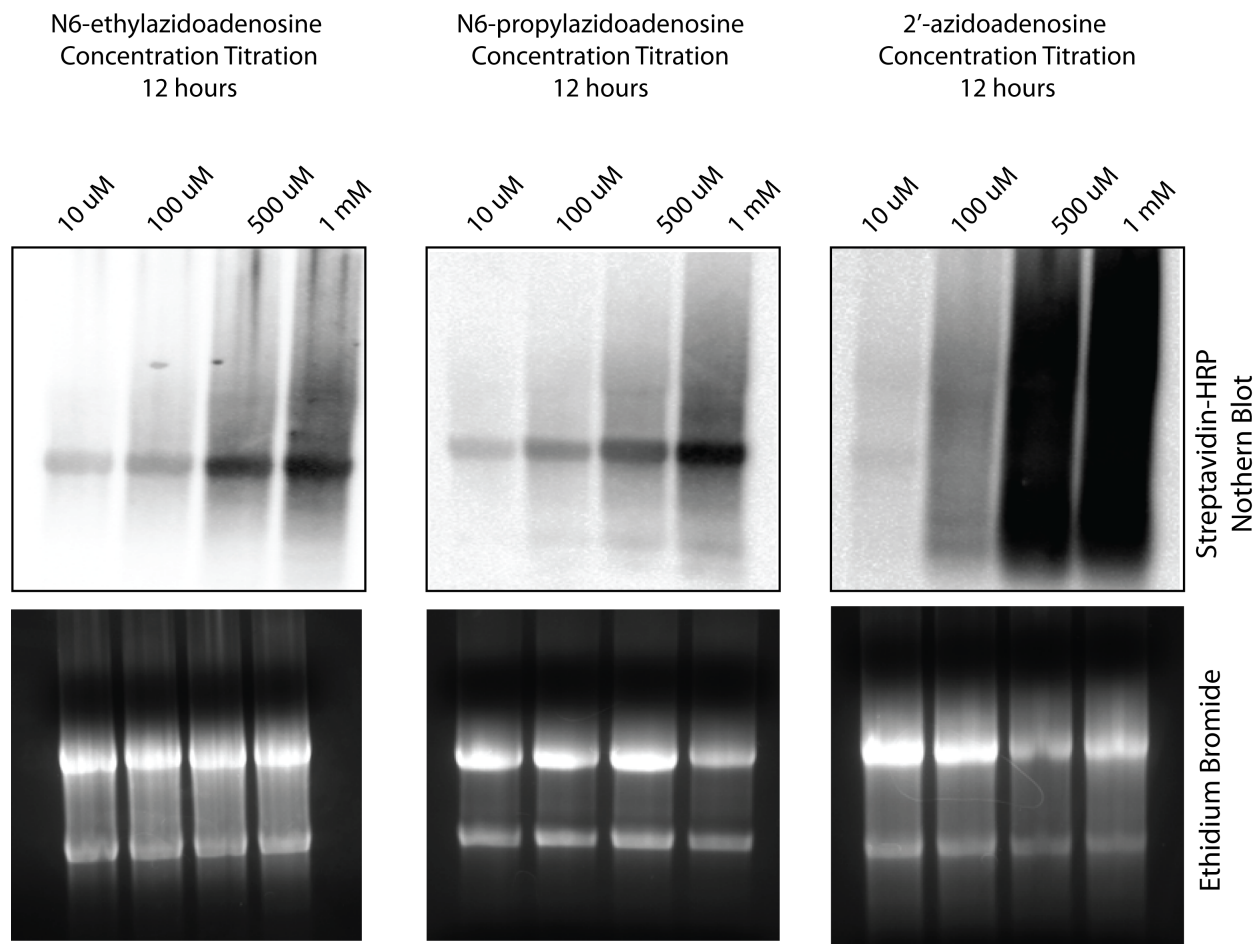
Supplementary Figure 2-5: Full Northern of concentration titration following 12 hour incubation. A. Nothern blot for N6-ethylazidoadenosine. B. Nothern blot for N6-propylazidoadenosine. B. Northern blot for 2'-azidoadenosine.
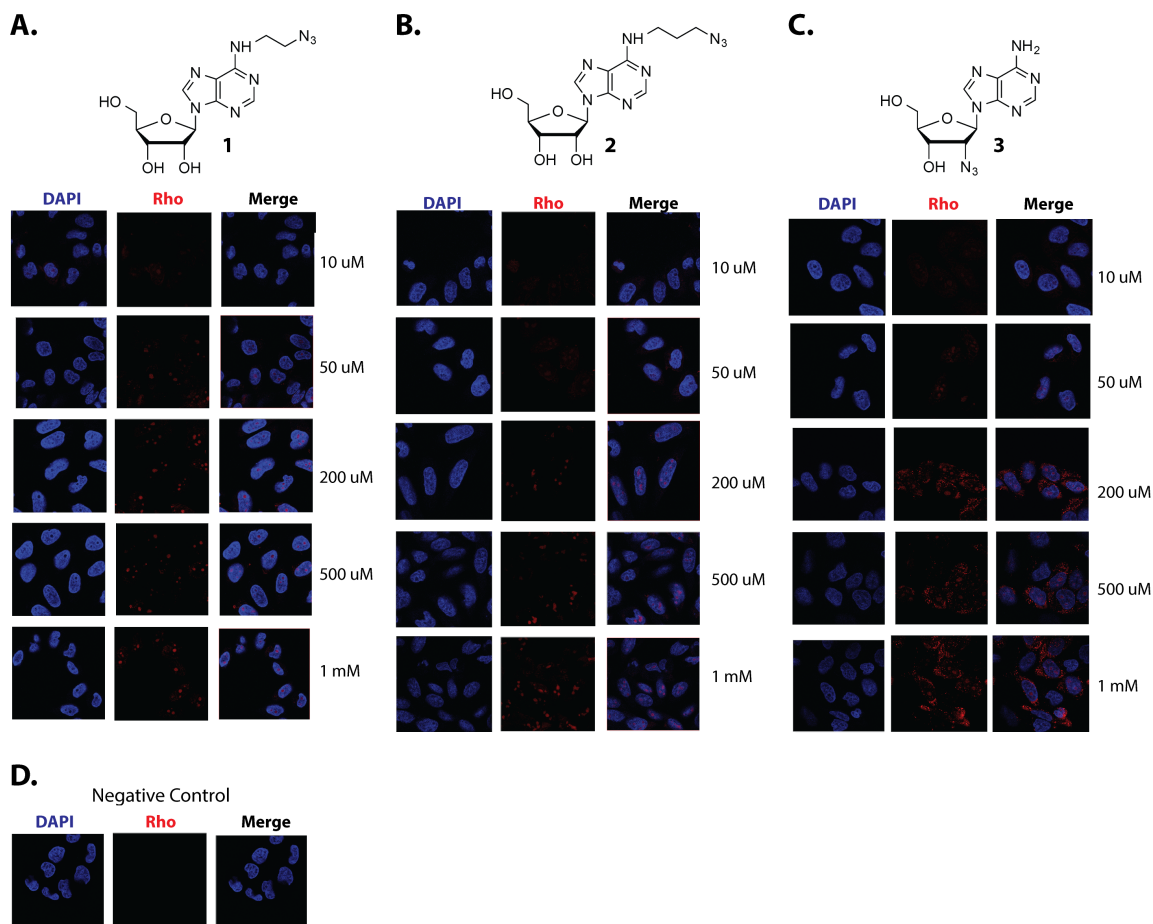
Supplementary Figure 2-6: Imaging RNA incorporation of azido-containing nucleosides. Concentration titration for N6-ethylazidoadenosine imaging (A), N6-propylazidoadenosine (B), and 2'-azidoadenosine (C) following "click chemistry and imaging.
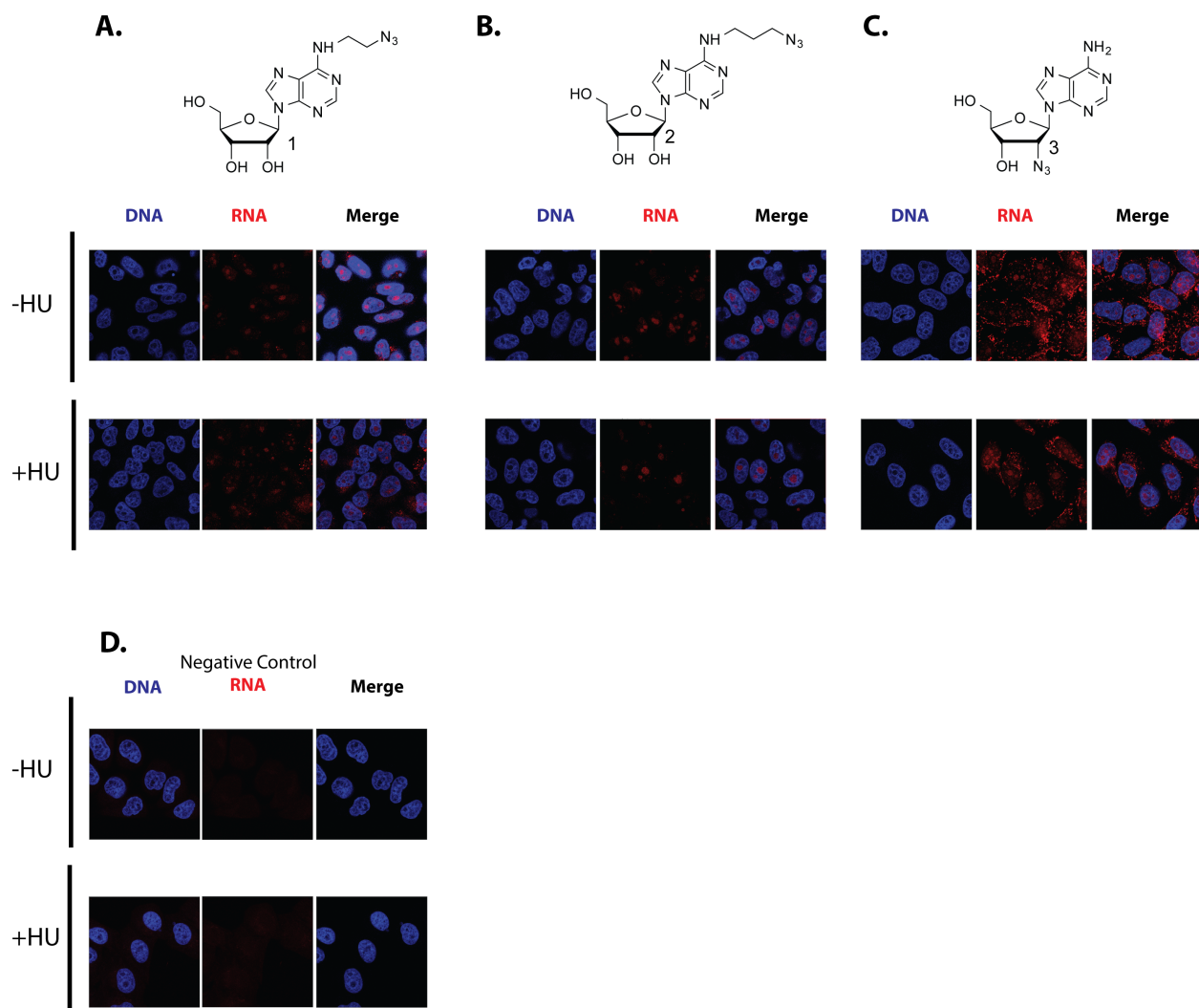
Supplementary Figure 2-7: Imaging RNA incorporation of azido-containing nucleosides in the presence of hydroxyurea.
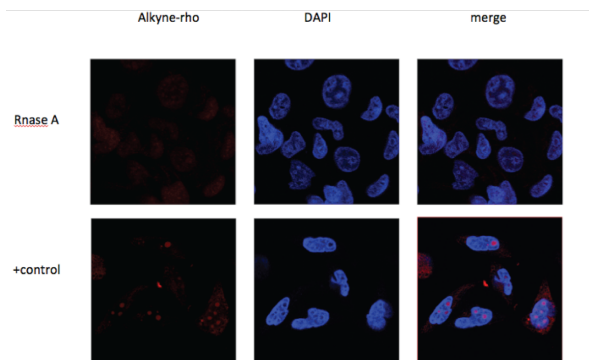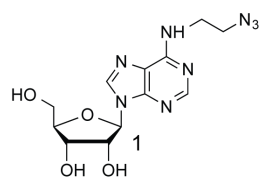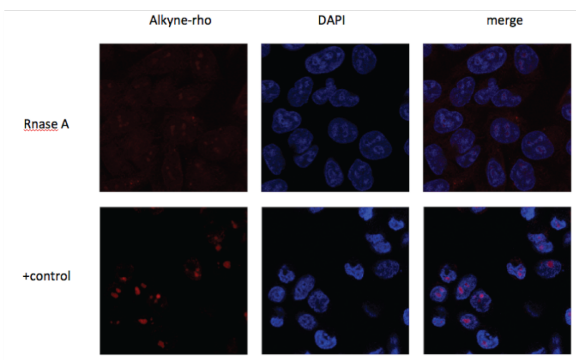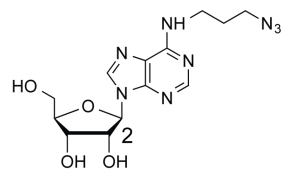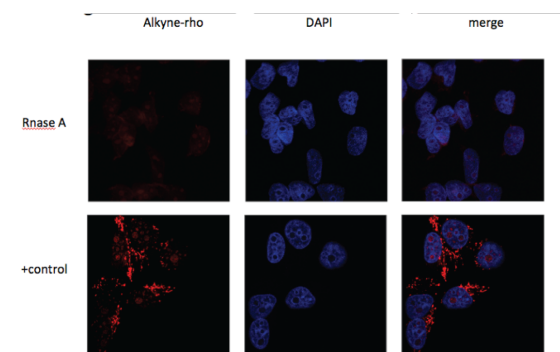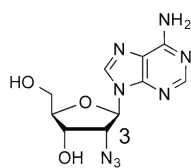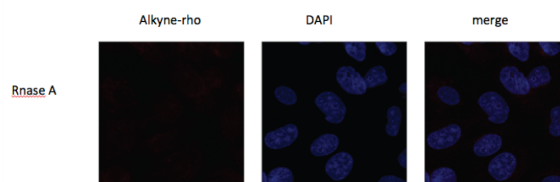
Supplementary Figure 2-8: Images demonstrating loss of fluorescent signal after cells are treated with RNase A.
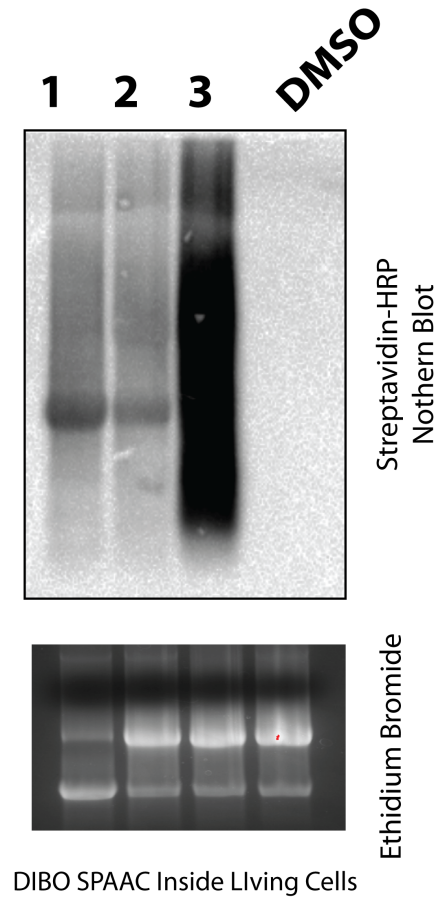
Supplementary Figure 2-9: Northern blot demonstrating DIBO SPAAC labeling of RNA inside living cells. Analog numbers 1-3, correspond to notation from the manuscript.
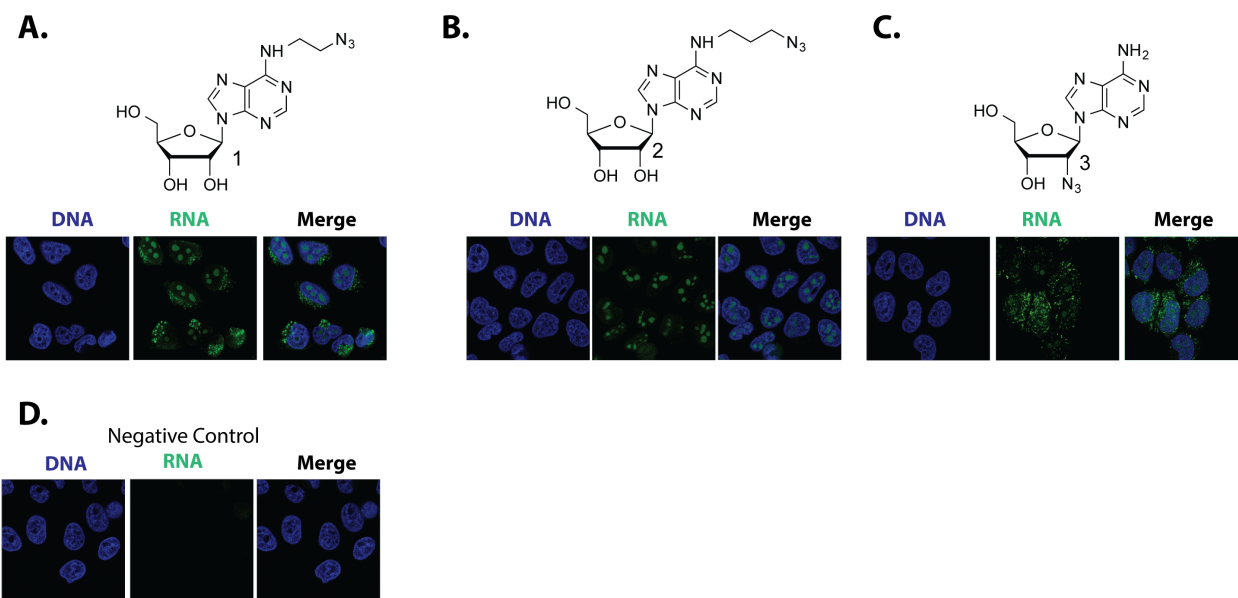
Supplementary Figure 2-10: SPAAC imaging experiments demonstrated that azide-labeled RNA can be robustly imaged with copper-free cycloadditions.
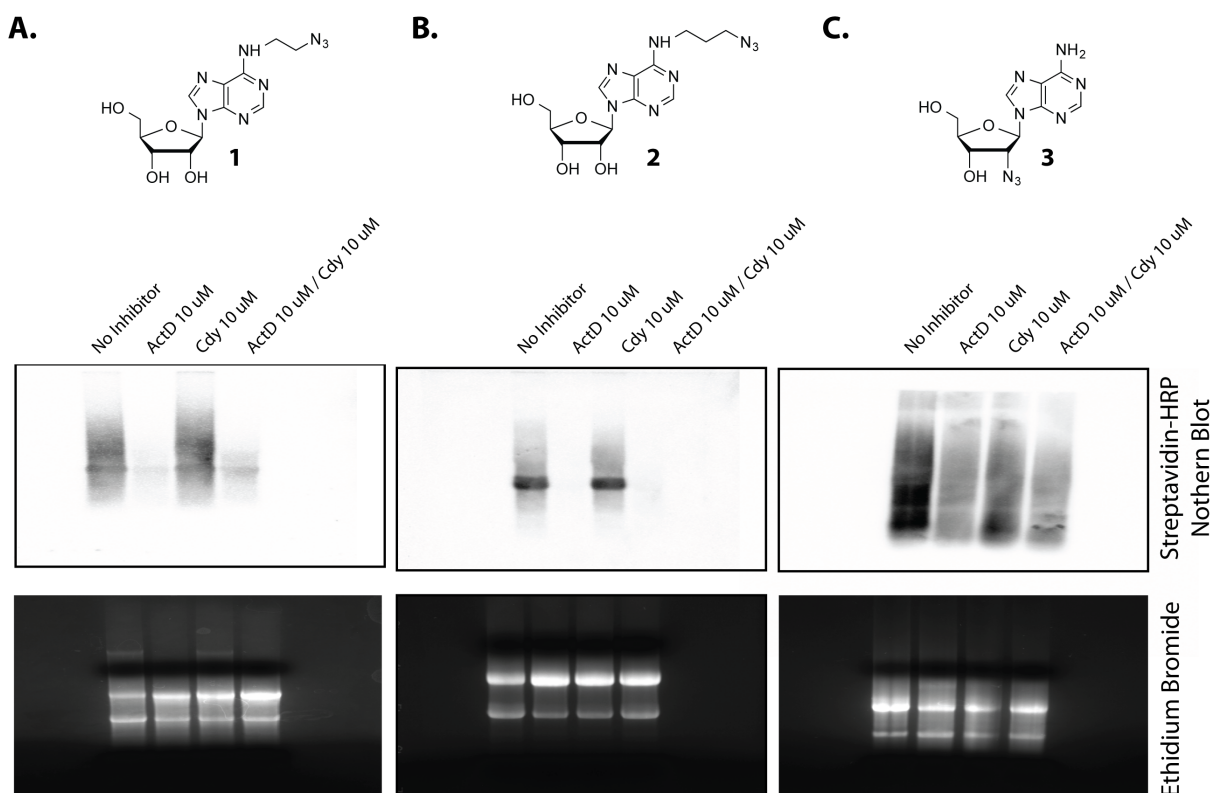
Supplementary Figure 2-11: Full Northern blots represented in Figure 2-3.
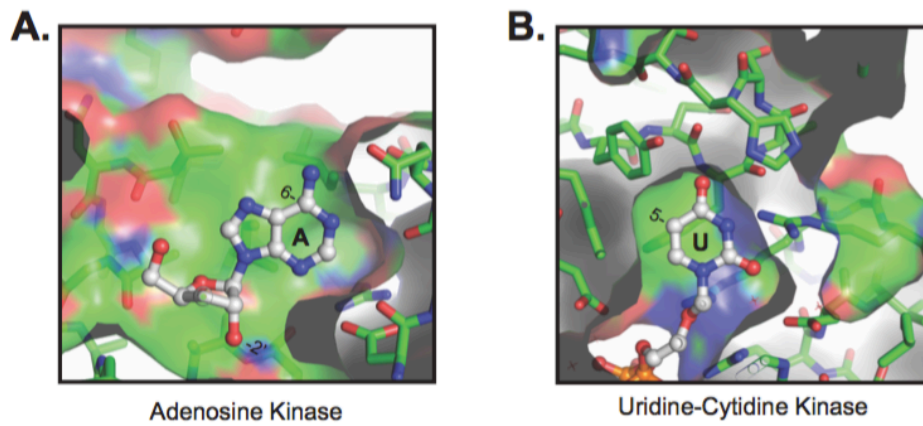
Supplementary Figure 2-12: Crystal structures of the corresponding adenosine and uridine/cytidine kinases. A. Crystal structure of the adenosine kinase bound to adenosine. B. Crystal structure of the uridine/cytidine kinase bound to uridine.

5. Spectra

DMSO

H2O

EtOH

1.0 1.0 1.0

1.0 2.0 1.0

1.0 1.0 1.0 2.0 1.9

2.0

5 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 –
f1 (ppm)

152.87 148.74
140.26
120.32
88.41 86.38
73.95 71.14
62.15
49.01
37.66
28.82

30 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 −1
f1 (ppm)

45

NMR spectra with labeled peaks:

¹H NMR (top): DMSO, H₂O, MeOH. Integrations: 1.00, 1.09, 1.02, 2.14

¹³C NMR (bottom): 164.45, 151.65, 142.48, 107.11, 46.99, DMSO

46

MeOH

H₂O

1.00 1.02 2.03 2.01 1.02 1.05 1.03

—163.79 —150.92 —140.19 —109.03 —89.48 —84.90 —74.51 —69.68 —60.68 —46.96

MeOH

47

6. References:

1. A. B. Neef, L. Pernot, V. N. Schreier, L. Scapozza, N. W. Luedtke, *Angewandte Chemie* **2015**, *54*, 7911-7914.

2. Vorbrüggen, H.; Bennua, B. *Chem. Ber.* **1981**, *114*, 1279-1286.

3. Rao, H.; Sawant, A.; Tanpure, A.; Srivatsan, S. *Chem. Commun.* **2012**, *48*, 498-500.

2.7 References

[1]   K. V. Morris, J. S. Mattick, *Nature reviews. Genetics* **2014**, *15*, 423-437.

[2]   J. J. Quinn, H. Y. Chang, *Nature reviews. Genetics* **2016**, *17*, 47-62.

[3]   M. Rabani, J. Z. Levin, L. Fan, X. Adiconis, R. Raychowdhury, M. Garber, A. Gnirke, C. Nusbaum, N. Hacohen, N. Friedman, I. Amit, A. Regev, *Nat Biotechnol* **2011**, *29*, 436-442.

[4]   E. E. Duffy, M. Rutenberg-Schoenberg, C. D. Stark, R. R. Kitchen, M. B. Gerstein, M. D. Simon, *Molecular cell* **2015**, *59*, 858-866.

[5]   C. Y. Jao, A. Salic, *Proc Natl Acad Sci U S A* **2008**, *105*, 15779-15784.

[6]   a) K. Fauster, M. Hartl, T. Santner, M. Aigner, C. Kreutz, K. Bister, E. Ennifar, R. Micura, *ACS chemical biology* **2012**, *7*, 581-589; b) T. Santner, M. Hartl, K. Bister, R. Micura, *Bioconjugate chemistry* **2014**, *25*, 188-195.

[7]   a) M. Grammel, H. Hang, N. K. Conrad, *Chembiochem : a European journal of chemical biology* **2012**, *13*, 1112-1115; b) D. Curanovic, M. Cohen, I. Singh, C. E. Slagle, C. S. Leslie, S. R. Jaffrey, *Nat Chem Biol* **2013**, *9*, 671-673; c) Y. Zheng, P. A. Beal, *Bioorganic & medicinal chemistry letters* **2016**, *26*, 1799-1802.

[8]   a) E. Paredes, S. R. Das, *Chembiochem : a European journal of chemical biology* **2011**, *12*, 125-131; b) M. L. Winz, A. Samanta, D. Benzinger, A. Jaschke, *Nucleic acids research* **2012**, *40*, e78.

[9]   a) H. Feng, X. Zhang, C. Zhang, *Nature communications* **2015**, *6*, 7816; b) I. Gallego Romero, A. A. Pai, J. Tung, Y. Gilad, *BMC biology* **2014**, *12*, 42.

[10]   V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angewandte Chemie* **2002**, *41*, 2596-2599.

[11]   N. J. Agard, J. A. Prescher, C. R. Bertozzi, *Journal of the American Chemical Society* **2004**, *126*, 15046-15047.

[12]   E. Saxon, C. R. Bertozzi, *Science* **2000**, *287*, 2007-2010.

[13]   D. C. Dieterich, A. J. Link, J. Graumann, D. A. Tirrell, E. M. Schuman, *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 9482-9487.

[14]   a) H. Sun, S. Jiang, J. Caton-Williams, H. Liu, Z. Huang, *RNA* **2013**, *19*, 1309-1314; b) A. Samanta, A. Krause, A. Jaschke, *Chemical communications* **2014**, *50*, 1313-1316; cH. Rao, A. A. Tanpure, A. A. Sawant, S. G. Srivatsan, *Nature protocols* **2012**, *7*, 1097-1112.

[15]   D. Curanovic, M. Cohen, I. Singh, C. E. Slagle, C. S. Leslie, S. R. Jaffrey, *Nature chemical biology* **2013**, *9*, 671-673.

[16]   G. Martin, W. Keller, S. Doublie, *The EMBO journal* **2000**, *19*, 4193-4203.

[17]   M. Rabani, J. Z. Levin, L. Fan, X. Adiconis, R. Raychowdhury, M. Garber, A. Gnirke, C. Nusbaum, N. Hacohen, N. Friedman, I. Amit, A. Regev, *Nature biotechnology* **2011**, *29*, 436-442.

[18]   A. Castello, R. Horos, C. Strein, B. Fischer, K. Eichelbaum, L. M. Steinmetz, J. Krijgsveld, M. W. Hentze, *Nature protocols* **2013**, *8*, 491-500.
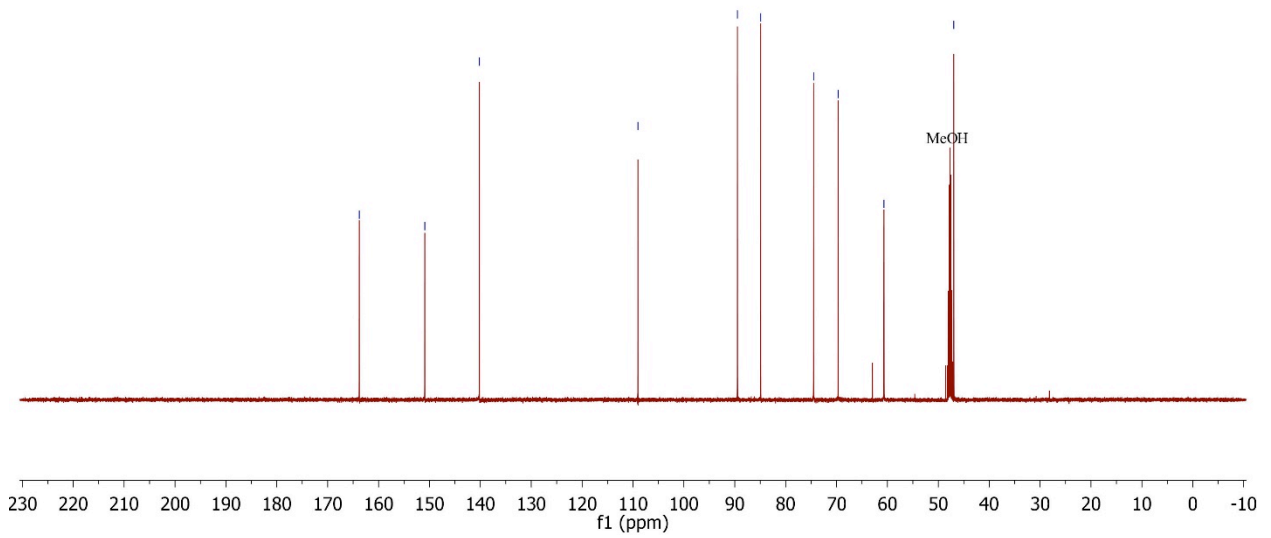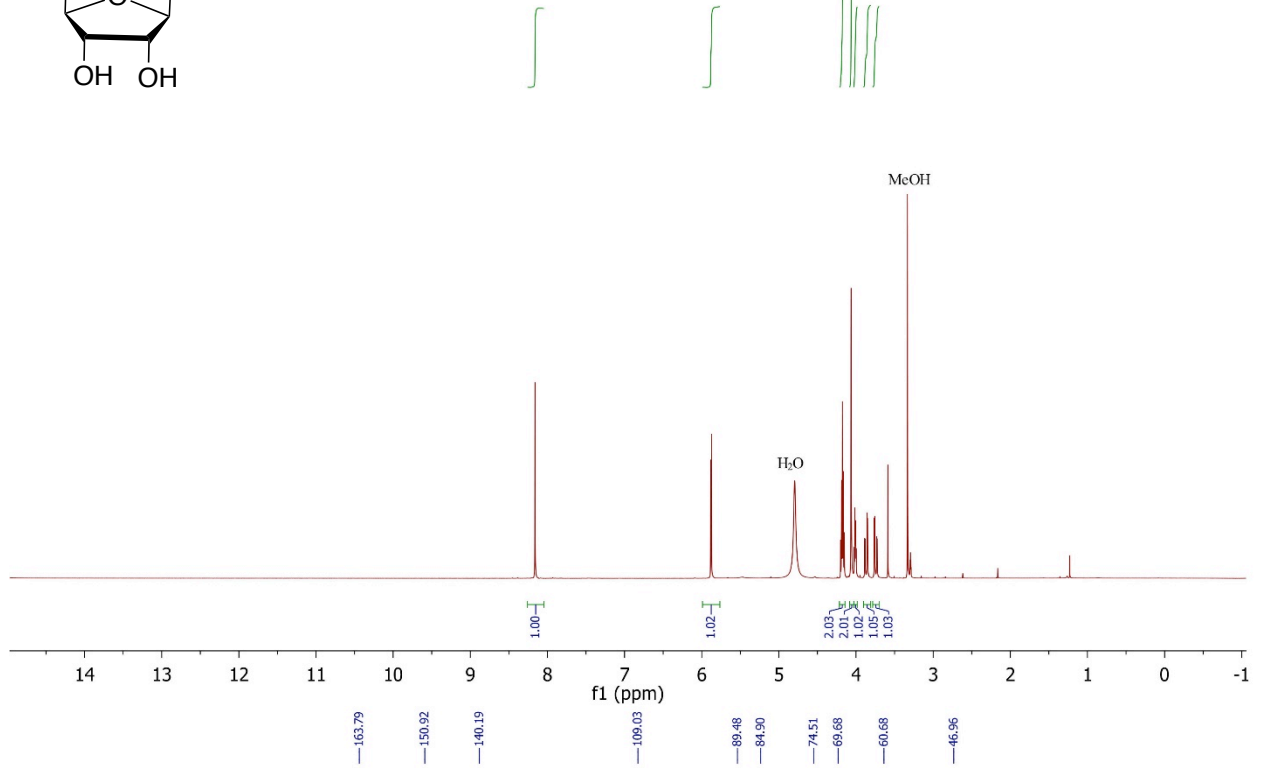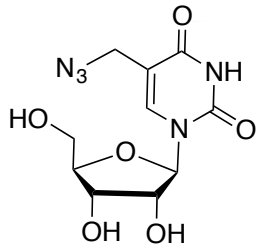
[19]   A. B. Neef, L. Pernot, V. N. Schreier, L. Scapozza, N. W. Luedtke, *Angewandte Chemie* **2015**, *54*, 7911-7914.

[20]   N. L. Golitsina, F. T. Danehy, Jr., R. Fellows, E. Cretton-Scott, D. N. Standring, *Antiviral research* **2010**, *85*, 470-481.

Chapter 3: Synthesis of furoyl probe for in-cell SHAPE

3.1 Introduction

RNA molecules can be found at the heart of many normal biological pathways and are key players in the onset of many diseases.[1] For proper function, RNA molecules must fold into complex secondary and even tertiary structures.[2, 3]

Several chemical methods have been developed for analyzing RNA structure. For example, dimethyl sulfate (DMS) can alkylate A and C residues not involved in Watson-Crick pairing.[4, 5] Selective hydroxyl acylation analyzed by primer extension, or SHAPE, reveals the internucleotide flexibility in RNA by 2'-OH acylation. SHAPE reactions form ester adducts on the 2'-OH (**Figure 3-1**)[6] which is accomplished by incubation with acylation electrophiles such as anhydrides,[6] acyl cyanides,[7] and more recently acyl imidazole reagents.[8]
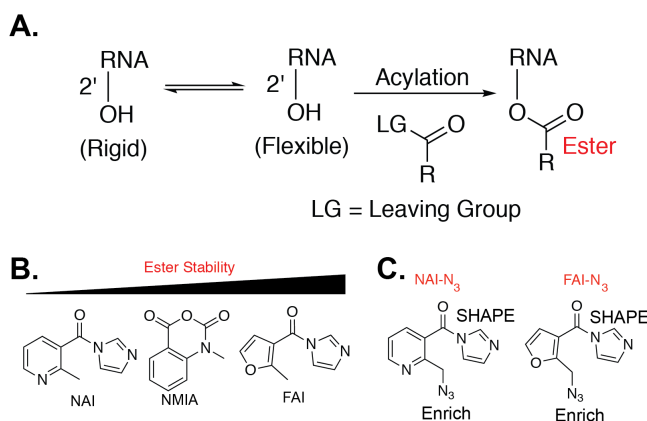
Figure 3-1. The SHAPE reaction and stability of SHAPE reagents. A. The SHAPE reaction. B. Ester stability of different SHAPE electrophiles. C. Structure of NAI-N_3 and FAI-N_3.

A key aspect of SHAPE is the stability of the ester product on the 2'-OH, which is identified by reverse transcription. A recent analysis by our lab demonstrated that furoyl SHAPE reagents form hyper-stable ester products (**Figure 3-1, B**).[9] This critical

observation suggests that the furoyl scaffold would be ideal for downstream *in vivo* SHAPE analysis.

Recently, SHAPE has been extended from selected RNA studies to being used transcriptome-wide. This transition has been accompanied by the development of novel protocols to convert SHAPE adducts into sequencing reads. An important criterion of such protocol development is an enrichment step of SHAPE adducts for achieving high signal-to-noise ratios in sequencing experiments. As such, we have utilized dual-functioning SHAPE reagents. This is tackled through the incorporation of two moieties, an acyl imidazole for the acylation reaction, and an alkyl azide to permit biotin attachment and enrichment through Strain Promoted Azide Alkyne Cycloaddition (SPAAC) reactions (**Figure 3-1, C**).[10]

Our first-generation reagent, NAI-N$_3$, was utilized to generate SHAPE data from the mouse embryonic stem cell transcriptome in both inside and outside cells.[10] However, we previously demonstrated that the nicotinoyl scaffold is highly prone to hydrolysis.[9] To retain the power of dual-functioning reagents coupled with our recent analysis of ester stability, we were prompted to design, synthesize, and test an optimized SHAPE reagent, FAI-N$_3$ (**Figure 3-1, C**), for structural analysis of RNA.

3.2 Results and Discussion

FAI-N$_3$ probe was synthesized in three steps (**Figure 3-2, A**). First, a substitution reaction was done on 2-(bromomethyl)furan-3-carboxylate with sodium azide in dimethylformamide to yield 2-(azidomethyl)furan-3-carboxylate. Then the ester was hydrolyzed to the carboxylic acid using methanol and aqueous sodium hydroxide. Finally, 2-(azidomethyl)furan-3-carboxylic acid was treated with carbonyldiimidazole in

dimethylsulfoxide to convert the carboxylic acid to the imidazole amide. The overall yield

in three steps was 30%.
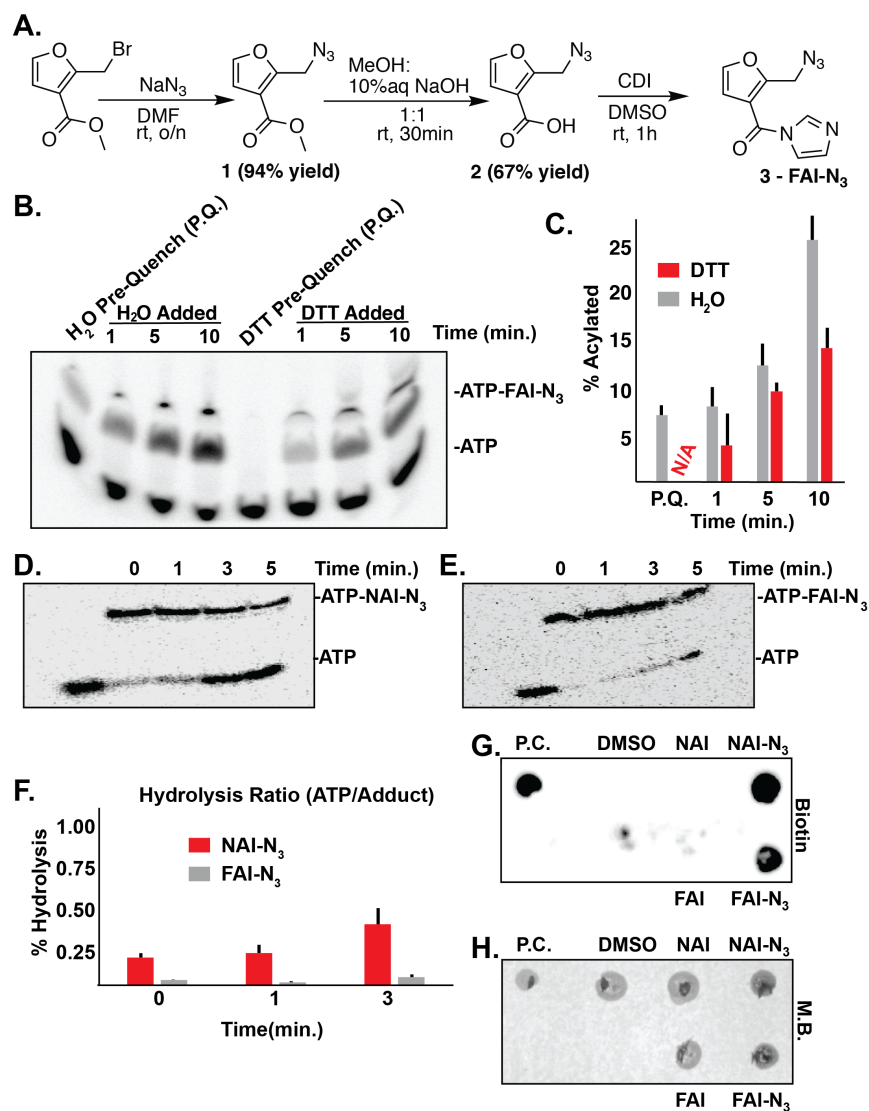
Figure 3-2. Synthesis and evaluation of FAI-N$_3$ for hydroxyl acylation. A). Synthesis of FAI-N$_3$. B). DTT quenching of FAI-N$_3$ acylation. Pre-quench (P.Q.) is designated for addition of either water or DTT to the ATP solution before the addition of FAI-N$_3$ C). Bar-graph detailing the percent acylation (monoacylation) as a function of time in quenching conditions represented in Panel B. Black bars denote error from biological triplicates. N/A is abbreviated for no acylation observed. D). Stability of NAI-N$_3$ at 95 °C. E). Stability of FAI-N$_3$ at 95 °C. F). Bar-graph detailing the percent hydrolysis as represented in Panel D and E. Black bars denote error from biological duplicates. G). Dot blot demonstrating successful "click" for biotinylation of FAI-N$_3$

53

acylated RNA. H). Methylene blue staining of G. Biotin = streptavidin dot blot. M.B. = methylene blue. P.C. = positive control bioinylated oligo.

A common requirement with these long lasting SHAPE probes is the ability for the probes to be quenched with dithiothreitol (DTT). Based on previous findings that FAI can be quenched by DTT, we tested if FAI-N$_3$ could also be quenched with the same method. In **Figure 3-2, B & C (Supplementary Information),** we compared the rates by quenching with DTT as opposed to quenching with water. Preincubation of DTT with ATP followed by addition of FAI-N$_3$ shows no reactivity, demonstrating that DTT is capable of quenching acylation. In subsequent lanes, the addition of DTT following the acylation reaction afforded better control in quenching compared to the addition of water. This shows that DTT is capable of quenching the acylation reactivity of furoyl SHAPE reagents

We recently reported that the furoyl scaffold on SHAPE reagents makes them stable to hydrolysis in conditions used in RNA-seq experiments.[9] We wanted to test whether FAI-N$_3$ adducts were more stable in comparison to NAI-N$_3$ adducts. We subjected isolated ATP-bound adducts with NAI-N$_3$ and FAI-N$_3$ to high-temperature hydrolysis conditions over time (95 °C; temperature used repeatedly for annealing of RNA over several steps including reverse transcription, primer binding, and also linker ligation. These results demonstrated that FAI-N$_3$ esters are much more stable than those from NAI-N$_3$ (**Figure 3-2, D – F; Supplementary Information**). This is consistent with our earlier observations comparing the two SHAPE scaffolds side by side, and demonstrates that the furoyl scaffold is a suitable reagent for RNA structure probing to form stable ester adducts for enrichment and opens the door for stringent washes at elevated temperatures.

The virtue of the alkyl azide is to enrich for SHAPE modifications, which others and we have shown can greatly reduce background of unmodified RNA.[11, 12] We next investigated the possibility if FAI-N$_3$ could be amenable for RNA enrichment with the same aproach. Dot blot analysis of the modified RNA demonstrated that biotin can be attached through SPAAC (**Figure 3-2, G & H**). Successful biotin attachment suggests that SHAPE adducts can be enriched before reverse transcription – an important aspect of applications of SHAPE for transcriptome-wide analysis.
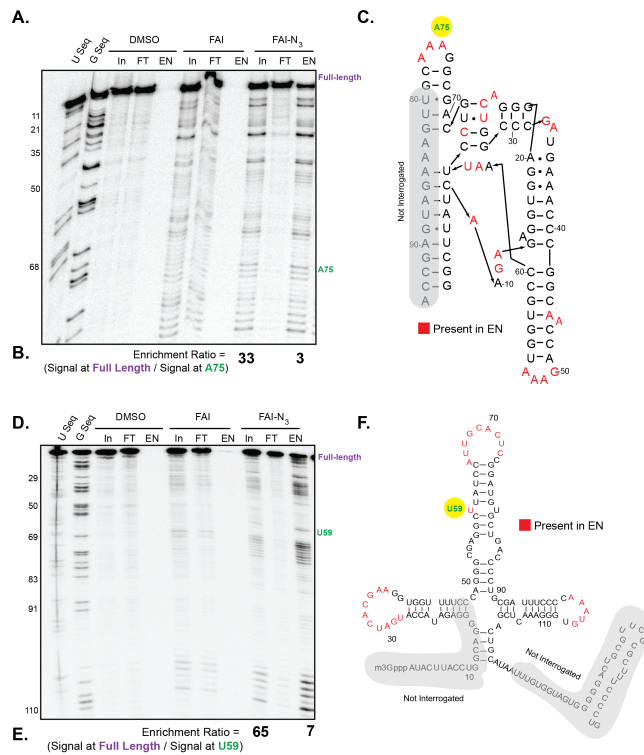
Figure 3-3. FAI-N$_3$ can be used to enrich acylation sites in RNA structure modifications. A). FAI-N$_3$ RNA adducts are capable of undergoing SPAAC for biotinylation and enrichment on SAM-I RNA. In = Input. FT = Flowthrough. EN = enriched. B). Calculation of enrichment of stops against full-length cDNA bands. C). SAM-I riboswitch with sites of FAI-N$_3$ enrichment mapped on the secondary structure. Nucleotides that are denoted in red are enriched in the EN lane of Panel A. D). Same as in Panel A, but for U1 snRNA E). Calculation

of enrichment of stops against full-length cDNA bands. F). U1 snRNA with sites of FAI-N$_3$ enrichment mapped on the secondary structure. Nucleotides that are denoted in red are enriched in the EN lane of Panel D.

We tested if FAI-N$_3$ would still recover viable RT stops for SHAPE on the SAM-I riboswitch and U1 snRNA.[13] We confirmed this by taking FAI-N$_3$ through the protocol of SHAPE modification, biotinylation, enrichment, and then reverse transcription. Due to the low propensity for FAI-N$_3$ adducts to be hydrolyzed, we added a high-temperature wash step. As shown in **Figure 3-3 A & D,** stringent wash steps with the FAI-N$_3$ lane afforded clean RT stops and enrichment at single stranded sites. To determine enrichment, we compared the signal from full-length cDNA from RT to an enriched site from SHAPE (Full-length/SHAPE). We observed a marked reduction in full-length cDNA in the FAI-N$_3$ enriched lane (IP, **Figure 3-3, B & E**) in contrast to observing majority of the full-length cDNA in the flowthrough (FT). The ratio of enriched cDNA bands at stops to full length increased almost 10-fold in the enriched samples. Enriched sites mapped to single-stranded regions of the RNA (**Figure 3-3 C & F**) Overall, these data nicely show that FAI-N$_3$ is a viable probe for SHAPE and enrichment of RT stops with removal of full length cDNAs.

Lastly, we sought to demonstrate that FAI-N$_3$ is a viable probe for in-cell analysis of RNA structure through SHAPE. *In vitro* analysis of RNA structure with SHAPE reagents does not always translate to robust signal for in-cell SHAPE.[14] We compared the *in vitro* SHAPE profiles of FAI and FAI-N$_3$ on 18S rRNA isolated from cells. As shown in **Figure 3-4, A**, all reagents give similar robust SHAPE profiles *in vitro*.

Comparison of reagents in cells demonstrated that FAI-N$_3$ and NAI-N$_3$ can both measure RNA structure robustly inside cells (**Figure 3-4, B**). Comparison of the SHAPE pattern to the published structure of the ribosome confirmed that residues with high SHAPE

reactivity are not base paired and are in orientations that would severely weaken their base pairs to make them flexible, consistent with high SHAPE reactivity (**Figure 3-4, C**).[15]
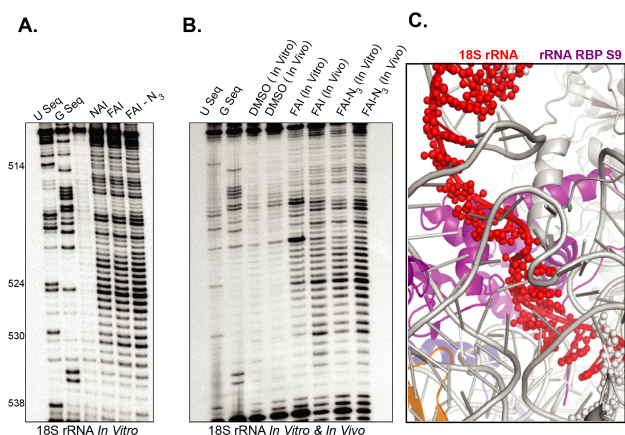
Figure 3-4. FAI-N₃ can be used to measure RNA structure inside living cells. A. Comparison of NAI, FAI, and FAI-N₃ for measuring RNA structure. B. Denaturing gel of adducts comparing in and outside of cell RNA acylation profiles. C. X-Ray structure model of the intact ribosome. (pdb 4V6X) The section in red is the single-stranded residues that have high acylation (SHAPE) in the denaturing gel in panel B.

3.3 Conclusion

RNA molecules are critical to the control of every biological pathway inside cells. Many RNA functions are influenced by unique RNA structures and developing methods to measure RNA structure inside cells is essential toward our understanding of RNA biology. Chemical probing is the go-to method for analyzing RNA structure. Extensions of chemical probing to transcriptome-wide analyses have presented new opportunities for probe and protocol development. Herein we have presented the facile synthesis and evaluation of a dual-functioning furoyl probe for SHAPE, whose function is to measure RNA structure by hydroxyl acylation and also be amenable to enrichment through the attachment of a biotin handle.

We have demonstrated that our novel reagent, FAI-N$_3$ is capable of measuring RNA structure, both inside and outside cells. FAI-N$_3$ reactivity can be controlled through DTT quenching, affording complete experimental control over RNA probing. We have also demonstrated that FAI-N$_3$ adducts can be enriched and such adducts are quite stable to conditions of hydrolysis. These two key observations suggest that FAI-N$_3$ should be an ideal reagent for transcriptome-wide analysis by SHAPE. The recent surge in labs performing SHAPE in cells with acylimidazole reagents further underscores the importance of our findings.[16-19] The facile synthesis and utility of FAI-N$_3$ should make it an ideal reagent for RNA structure probing in living cells.

3.4 Acknowledgements

3.5 Methods

1. General

All reagents were purchased from commercial suppliers and were of analytical grade and used without further purification unless otherwise noted. Methyl 2-(bromomethyl)furan-3-carboxylate was purchased from Santa Cruz Biotechnology. Reaction progress was monitored by thin-layer chromatography on EMD 60 F254 plates, visualized with UV light, iodine, ninhydrin, KMnO$_4$, FeCl$_3$, p-anisaldehyde, 2,4-DNP, and bromocresol green stains. Compounds were purified via flash column chromatography using Sorbent Technologies

60 Å 230 x 400 mesh silica gel. Anhydrous solvents acetonitrile (MeCN), dichloromethane (DCM), methanol (MeOH), tetrahydrofuran (THF), dimethylformamide (DMF) were degassed and dried over molecular sieves. Acetone was dried over $MgSO_4$. All reaction vessels were flame dried prior to use. NMR spectra were acquired with Bruker Advanced spectrometers. All spectra were acquired at 298 K. ^{1}H-NMR spectra were acquired at 400 MHz and 500 MHz. ^{13}C-NMR spectra were acquired at 500 MHz. Chemical shifts are reported in ppm relative to residual non-deuterated NMR solvent, and coupling constants (J) are provided in Hz. All NMR spectra was analyzed using MestreNova software. Low and high-resolution electrospray ionization (ESI) mass spectra and Gas Chromatography mass spectra were collected at the University of California-Irvine Mass Spectrometry Facility. IR spectra were acquired from neat samples, unless otherwise noted, with a PerkinElmer Spectrum Two IR Spectrometer.

2. Synthesis

1 (94% yield)
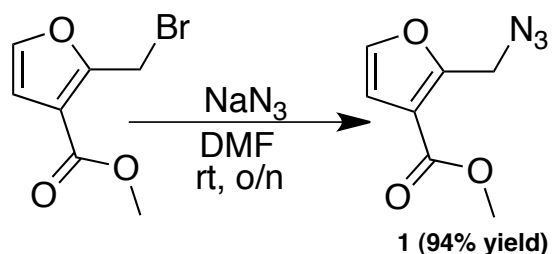
Synthesis of methyl 2-(azidomethyl)furan-3-carboxylate (1).

To a solution of methyl 2-(bromomethyl)furan-3-carboxylate (0.459mmol, 1eq) in dry DMF (5mL) was added NaN_3 (0.918mmol, 2eq). The reaction was left to spin at room temp overnight. Then the reaction was quenched with sat. $NaHCO_3$ and the aqueous layer was extracted with EtOAc. The organics were washed 3x with water and 3x with brine. The

organic layer was dried over MgSO$_4$, filtered, and concentrated to yield the product as a brown liquid (0.078g, 94% yield). ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.39 (d, *J*=2.5Hz, 1H), 6.71 (d, *J*=2.4Hz, 1H), 4.63 (s, 2H), 3.84 (s, 3H). ^{13}C NMR (500 MHz, CDCl$_3$) δ 163.39, 154.61, 142.81, 116.43, 111.05, 51.84, 45.57. HRMS: Theoretical 204.0385 [M+Na$^+$], Observed 204.0393 [M+Na$^+$].
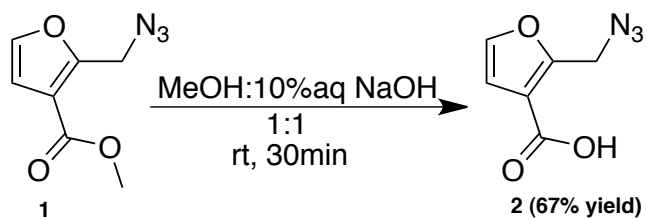
Synthesis of 2-(azidomethyl)furan-3-carboxylic acid (2).

Methyl 2-(azidomethyl)furan-3-carboxylate (0.293mmol, 1eq) was stirred vigorously in a 1:1 solution of MeOH:10% aq. NaOH (1mL:1mL), and monitored by TLC. Once starting material was gone, the reaction was diluted with water, washed with ether only once, then acidified to pH4 with 2N HCl, and extracted 5 times with EtOAc. The organic layer was dried over MgSO$_4$, filtered, and concentrated to yield product as a tan solid (0.033g, 67% yield). ^{1}H NMR (500 MHz, DMSO) δ 13.11 (bs, 1H), 7.83 (d, *J*=1.2Hz, 1H), 6.78 (d, *J*=1.6Hz, 1H), 4.74 (s, 2H).^{13}C NMR (500 MHz, DMSO) δ 164.35, 154.41, 144.24, 117.54, 111.59, 45.36. HRMS: Theoretical 212.0048 [M-H+2Na$^+$], Observed 212.0044 [M-H+2Na$^+$].
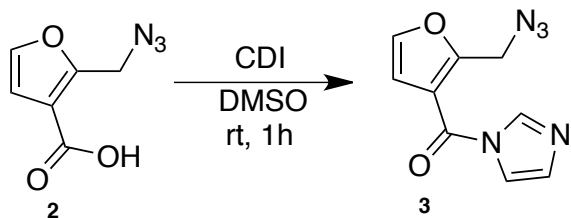
Synthesis of (2-(azidomethyl)furan-3-yl)(1*H*-imidazol-1-yl)methanone (3).

2-(azidomethyl)furan-3-carboxylic acid (0.623mmol, 1eq) was dissolved in anhydrous DMSO (0.208mL) followed by addition of CDI (0.623mmol, 1eq) to yield a final concentration of 3M. The reaction was allowed to stir at room temp for 1h and then used for cell testing. In order to obtain NMR data for compound **3**, the reaction was run as follows: 2-(azidomethyl)furan-3-carboxylic acid (0.171mmol, 1eq) was dissolved in anhydrous DCM (1mL) followed by addition of CDI (0.171mmol, 1eq) and monitored by TLC. Once starting material was gone, the product was purified by column (95% EtOAc/Hex isochratic) to yield 0.0178g (48% yield) and used for spectroscopic data only. ^{1}H NMR (500 MHz, DMSO) δ 8.33 (s, 1H), 8.02 (d, *J*=2Hz, 1H), 7.75 (s, 1H), 7.19 (s, 1H), 7.07 (d, *J*=2Hz, 1H), 4.73 (s, 2H).^{13}C NMR (500 MHz, DMSO) δ 160.68, 156.97, 144.77, 138.29, 131.02, 118.24, 117.27, 111.52, 45.59. HRMS: Theoretical 218.0678 [M+Na[+]], Observed 218.0674 [M+Na[+]].

Characterization of ATP deacetylation.

Radiolabeled ATP was incubated with 1 μL SHAPE reagent for 1 hour at 300 mM final concentration with FAI-N$_3$ (10% final volume) or 100 mM final concentration with 1 M NAI-N$_3$ (10% final volume) in 100 mM HEPES buffer, pH 8.0, containing 6 mM MgCl$_2$ and 100 mM NaCl to a final volume of 10 μL. Reaction was incubated at 37 °C, and then placed

on ice. Products were resolved on 25% native polyacrylamide gel. (29:1 acrylamide and bisacrylamide and 1% Tris-borate-EDTA (TBE)) and visualized by phosphorimaging (Typhoon, GE healthcare). ATP supershifts were identified and excised and gel slices were crushed and added to buffer containing a solution of 400 mM KCl and incubated overnight at 4 °C. The following day, ATP conjugates were isolated by spinning down gel slices and isolating the supernatant. For deacylation reactions, ATP conjugates (1,000 counts per minute, cpm) were added to volume up to 20 µL in ultrapure RNase-free water. The reactions were incubated at appropriate times and temperatures and the reactions were stopped by placing the solutions on ice for immediate loading into the gels. Deacylated products were loaded and resolved on a 25% native polyacrylamide gel as shown above with 10 µL of GLB2. Percent acylation was calculated as the ratio of intensity between ATP and adduct bands. Band intensity were quantified using imageJ.

Demonstrating SHAPE reactivity with DTT quench.

Radiolabeled ATP was incubated with each SHAPE electrophile with procedures described above at a final concentration of 300 mM FAI. Following addition of each SHAPE electrophile, the reaction was incubated at 37 °C until the indicated time. An aliquot of reaction was removed and quenched in either an equal volume of Gel Loading Buffer II (Ambion, Inc.), or Gel Loading Buffer II supplanted with 666 mM DTT to a final concentration of 333 mM DTT. For the pre-incubation control, the SHAPE electrophile was added to a solution of Gel Loading Buffer II supplanted with 666 mM DTT. Products were resolved on 28% native polyacrylamide gel and visualized by a gel imager (Typhoon, GE

healthcare). Percent acylation was calculated as the intensity from acylated products over the total intensity of ATP signal. Bands intensity were quantified using imageJ.

Characterization of NAI-N₃ and FAI-N₃ reactivity with RNA.

Reverse Transcription primers:

SAM-I (5', ATTTAGGTGACACTATAGTT, 3')

HeLa 18s (5', CCAATTACAGGGCCTCGAAA, 3')

U1 Primer (5', CCCACTACCACAAATTATGCAG, 3')

In vitro transcription of SAM-I construct. A 94 nucleotide construct consisting of the sequence for the SAM riboswitch from the metF-metH2 operon of *T. tencongensis* was designed into a plasmid with IDT.[1] One Shot Top 10 chemically competent cells were transformed by SAM-I plasmid and plated on lysogeny broth (LB) supplemented with ampicillin (100 mg/mL) agar plates. A single colony was selected in 3 mL culture and grown overnight. The resulting plasmid was isolated according to conditions using QIAprep Miniprep. Transcription template was prepared by PCR using primers directed against the T7 promoter (5', TAATACGACTCACTATAGGG, 3') and an adaptor sequence for reverse transcription (5', ATTTAGGTGACACTATAGTT, 3'). RNA was transcribed in 40 μL reactions according to conditions using Ribomax Large Scale RNA production systems kit from promega. The transcription reaction was allowed to proceed for three hours at 37 °C. The resulting RNA was treated with 1 μL of RQ1 DNase for 15 min. The resulting RNA was phenol chloroform extracted, and excess nucleotides were removed using a G25 column. Resulting concentrations of RNA was determined through NanoDrop.

In vitro transcription of U1 construct. A 164 nucleotide construct for the U1 snRNA was graciously donated to us from the Hertel Lab (UC Irvine).[2] Transcription template was prepared by PCR forward (5', TAATACGACTCACTATAGGGATACTTACCTGGCAGGGGA, 3') and reverse (5', CAGGGGAAAGCGCGA, 3') primers. RNA was transcribed in 40 μL reactions according to conditions using Ribomax Large Scale RNA production systems kit from promega. The transcription reaction was allowed to proceed for three hours at 37 °C. The resulting RNA was treated with 1 μL of RQ1 DNase for 15 min. The resulting RNA was phenol chloroform extracted, and excess nucleotides were removed using a G25 column. Resulting concentrations of RNA was determined through NanoDrop.

[32]P End labeling for reverse transcription. 200 pmol of primer DNA was kinased with Optikinase according to manufacturer's conditions by Affymetrix. The reaction was allowed to proceed for two hours at 37 °C. Reactions were stopped by the addition of equal amounts of Gel Loading Buffer II (Ambion, Inc.). The reactions were loaded onto a 15% denaturing PAGE gel. The band of interest was visualized by gel imager (Typhoon, GE healthcare). The resulting band was excised and eluted overnight in 300 mM KCl. Resulting solution was EtOH precipitated and dissolved to 8,000 cpm/μL for further use in reverse transcription.

Acylation of RNA in vitro. In a typical *in vitro* modification protocol, 5 μg total RNA (isolated from HeLa cells) or 2 pmol of *in vitro* transcribed *RNA* was heated in 6 μL metal-free water for 2 min at 95 °C. The RNA was then flash cooled on ice. 3 μL of 3× SHAPE buffer (333 mM HEPES, pH 8.0, 20 mM $MgCl_2$ and 333 mM NaCl) was added, and the RNA was allowed to

equilibrate at 37 °C for 5 min. To this mixture, 1 µL of 3 M FAI-N$_3$ or 1 M NAI-N$_3$ in DMSO (+) or DMSO alone (−) was added. The reaction was permitted to continue for 15 min. Reactions were extracted once with acid phenol and chloroform (pH 4.5 ± 0.2 (s.d.)) and twice with chloroform. RNA was ethanol precipitated with 20 µL of 3 M sodium acetate buffer (pH 5.2) and 1 µL of glycogen (20 µg/µL). Pellets were washed twice with 70% ethanol and resuspended in 5 µL RNase-free water.

Copper free click chemistry of acylated RNA. In a typical reaction, acylated RNA (1 pmol) was reacted with 100 equivalents of Dibenzocyclooctyne-PEG4-biotin (DBCO-Biotin, sigma) for 1 h at 37 °C in 1x PBS. Reactions were extracted once with acid phenol:chloroform (pH 4.5) and twice with chloroform. RNA was ethanol precipitated with sodium acetate (pH 5.2) and 1 µL glycoblue. Pellets were washed twice with 70% ethanol and resuspended in 17 µL RNase-free water.

Enrichment of Modified RNA. To 50 pmol of precipitated and biotinylated RNA (in 700 µL binding buffer: 50 mM Tris-HCl pH 7.0 and 1 mM EDTA) was added with 50 µL of prewashed Dynabeads MyOne C1 beads. The reaction mixture was then mixed at room temperature for 1 h. The beads were collected on a magnetic plate and flowthrough was saved. The beads were then washed three times with 700 µL of Biotin Wash buffer (10 mM Tris-HCl, pH 7.0, 1 mM EDTA, 4 M NaCl, 0.2% Tween). The first wash was saved and combined with the flowthrough for further analysis. Samples were later washed twice with RNase-free water. FAI-N$_3$ adducts undergoing harsher wash conditions was subjected to two 700 µL Biotin wash buffer for 5 min along with two washes with RNase-free water at

70 °C. Samples were eluted twice with 44 μL formamide, 1 μL of 0.5 M EDTA, and 5 μL of 50 mM of free D-Biotin at 95 °C for 5 min. Eluted samples were diluted with 600 μL RNase-free water. All samples were purified using RNA Clean and Concentrator Kit (Zymo). Samples were eluted in 6 μL of RNase-free water and used for subsequent reverse transcription. SAM-I and U1 RNA RT stops were put through biological duplicates. The enrichment factor was calculated by the signal at full-length RT/SHAPE signal at a chosen band, as denoted in the figures.

Dot blot Analysis of Enriched modified RNA. Hybond N+ membranes (GE) were pre-incubated in 10X SSC. Precipitated biotinylated total RNA was dissolved in 15 μL of RNase free water. RNA was loaded onto the Hybond membrane and crosslinked using 254 nm ultraviolet light. The membrane was incubated with blocking solution (120 mM NaCl, 16 mM Na_2HPO_4, 8 mM NaH_2PO_4, 170 mM SDS) for 30 min. To the membrane was added 1 μL streptavidin (Pierce High Sensitivity Streptavidin-HRP) in blocking solution. The membrane was washed twice with wash buffer A (1:10 blocking solution) for 30 min, and twice with wash buffer B (100 mM Tris pH 9.5, 100 mM NaCl, 20 mM $MgCl_2$) for 5 min. Membrane was incubated with Pierce Western blotting substrate (Thermo) and visualized on the ChemiDoc (Biorad) under chemiluminesence hi sensitivity.

Acylation of RNA in HeLa cells. HeLa cells were grown in DMEM (high glucose) culture medium supplemented with 10% FBS, 1% penicillin streptomycin. Cells were washed three times with Dulbecco's phosphate-buffered saline (DPBS) and then scraped and spun down at 1000 r.p.m. for 5 min. Cells (~3-6 x 10^7) were resuspended in 45 μL DPBS. 5 μL DMSO(-),

10% final concentration or 5 µL 10x electrophilic stock in DMSO (+) was added to the desired final concentration. Cell suspensions were incubated at 37 °C for 10 min. Cells were pelleted by centrifugation at 1000 r.p.m. for 5 min and resuspended in 1 mL Trizol Reagent. RNA was harvested using Trizol Reagent following the manufacturer's instructions.

Reverse transcription of modified RNA (in vitro and in vivo). ^{32}P-end-labeled DNA primers were annealed to modified RNA by incubating 95 °C for two minutes, then 25 °C for two minutes, and 4 °C for 2 minutes. To the reaction, first strand buffer, DTT, and dNTP's were added. The reaction was preincubated at 52°C for 1 min, then superscript III (2 units/µL final concentration) was added. Extensions were performed for 15 min. To the reaction, sodium hydroxide was added to a final concentration of 400 mM and allowed to react for 5 min at 95 °C. The resulting complementary DNA (cDNA) was snapped cooled on ice, and ethanol precipitated according to above procedures. Purified cDNA was resuspended in 2 µL of nuclease-free water and 2 µL of Gel Loading Buffer II was added. cDNA products were resolved on 10% denaturing polyacrylamide gel, and visualized by a gel imager (Typhoon, GE healthcare).

3.6 Supplementary Data
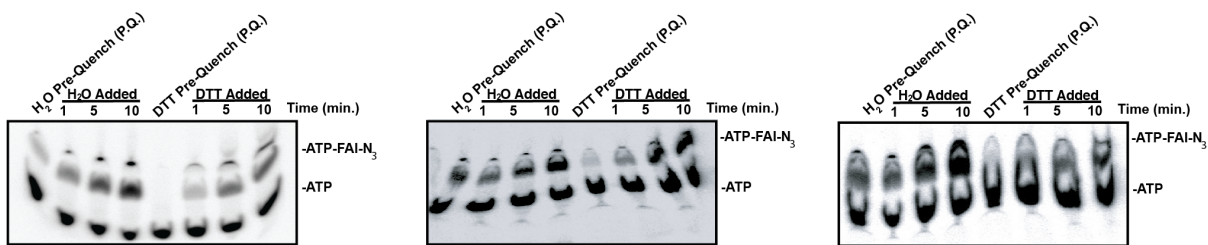
Supporting Information Figure 3-1. Additional experiments supporting Figure 3-2 and DTT quenching of acylation. The images are replicates of data that was used for Figure 3-2 B and C.
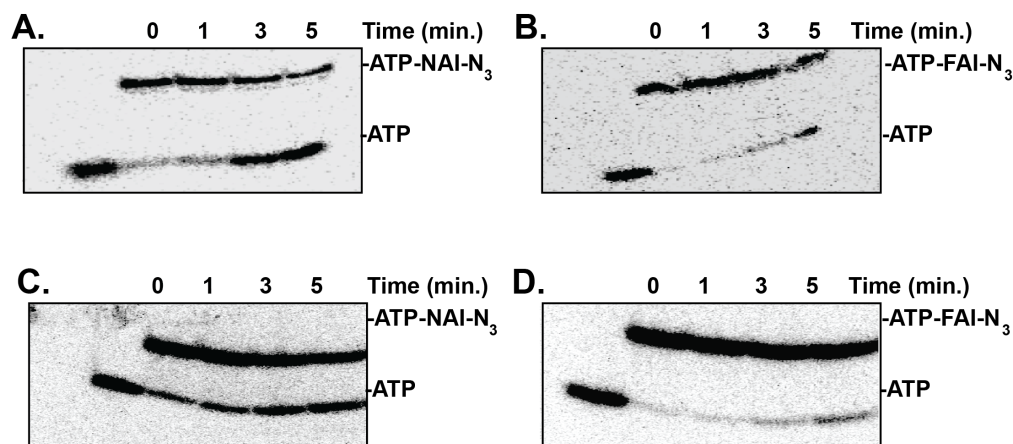
Supporting Information Figure 3-2. Additional experiments supporting Figure 3-2 and Hydrolysis of NAI-N_3 and FAI-N_3 acylation products. This is in support of Figure 3-2 D-F. A) Deacylation of NAI-N_3 acylation adducts with ATP, REP 1. B) Deacylation of NAI-N_3 acylation adducts with ATP, REP 1. C) Deacylation of NAI-N_3 acylation adducts with ATP, REP 2. D) Deacylation of NAI-N_3 acylation adducts with ATP. REP 2.
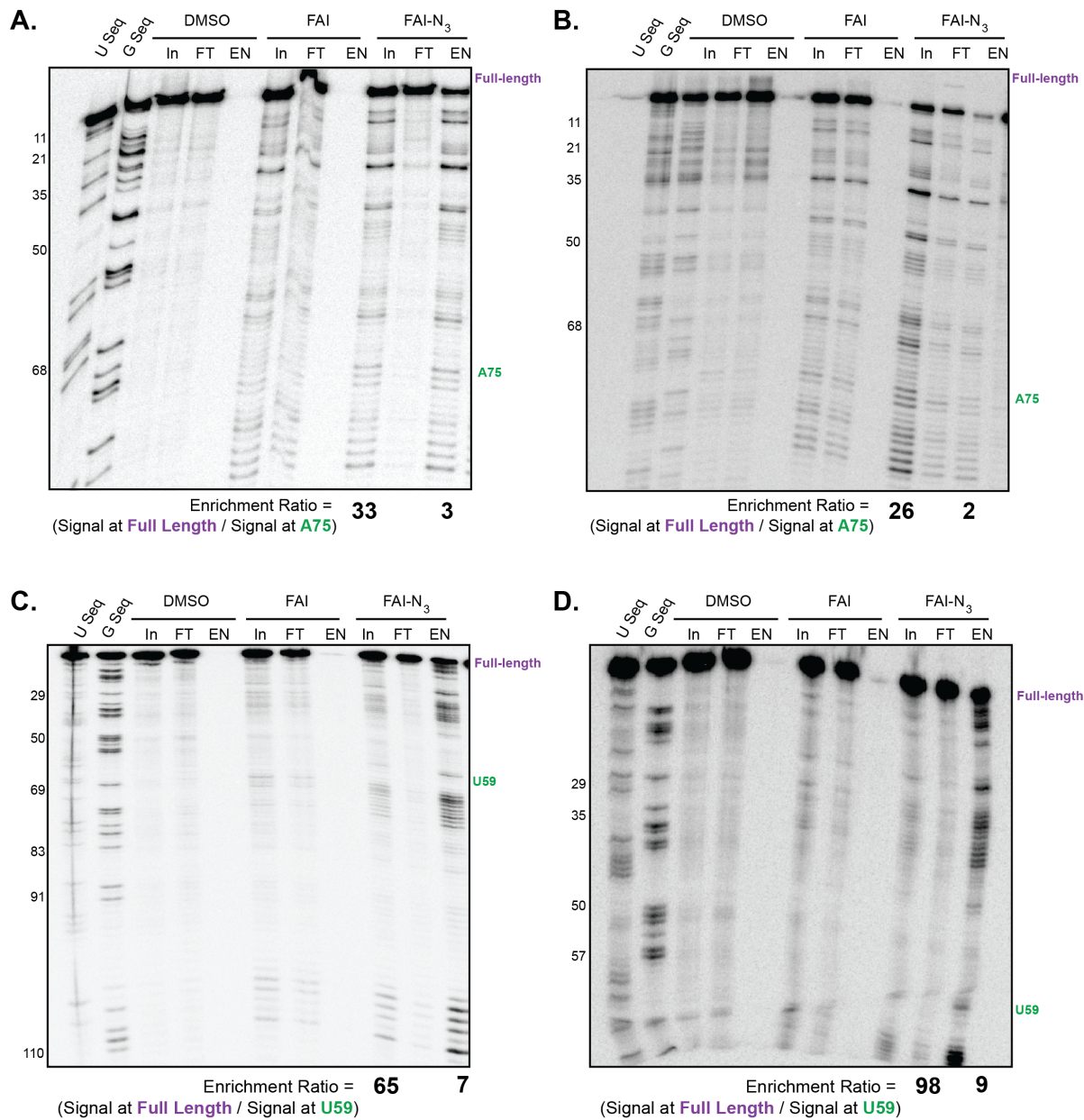
Supporting Information Figure 3-3. Additional experiments supporting Figure 3-2 and Hydrolysis of NAI-N$_3$ and FAI-N$_3$ acylation products. This is in support of Figure 3-3 A) Enrichment RT gel for SAM-I, Rep 1. B) Enrichment RT gel for SAM-I, Rep 2. C) Enrichment RT gel for U1, Rep 1. D) Enrichment RT gel for U1, Rep 2.

Spectra

Methyl 2-(azidomethyl)furan-3-carboxylate (1)

^{1}H NMR (500MHz, CDCl$_3$)
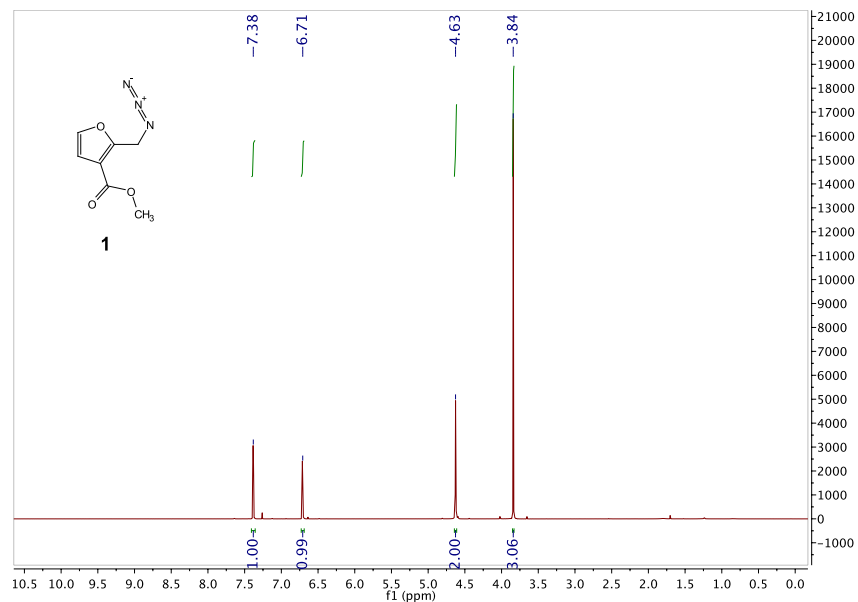
^{13}C NMR (500MHz, CDCl$_3$)
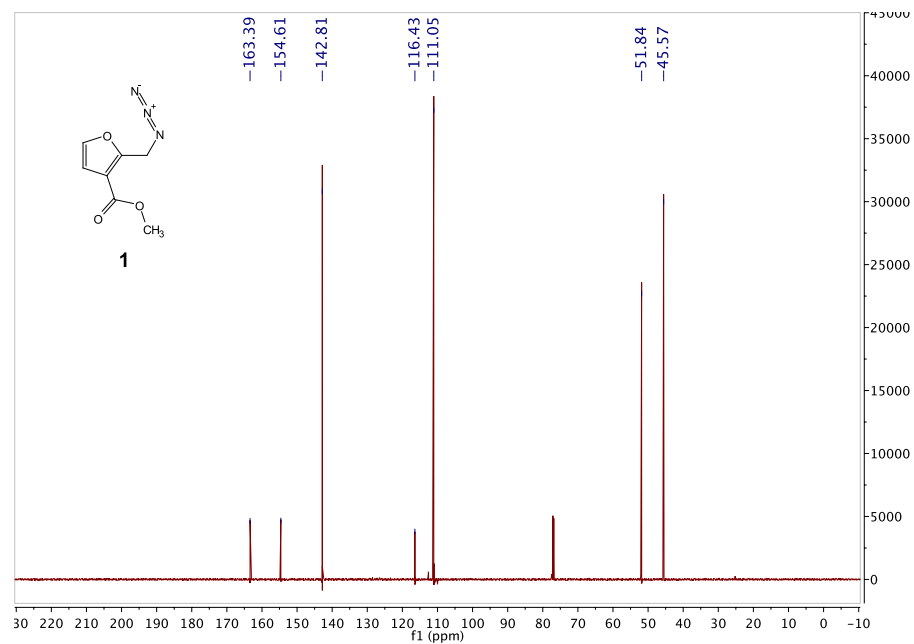
2-(azidomethyl)furan-3-carboxylic acid (2)

¹H NMR (500MHz, DMSO)
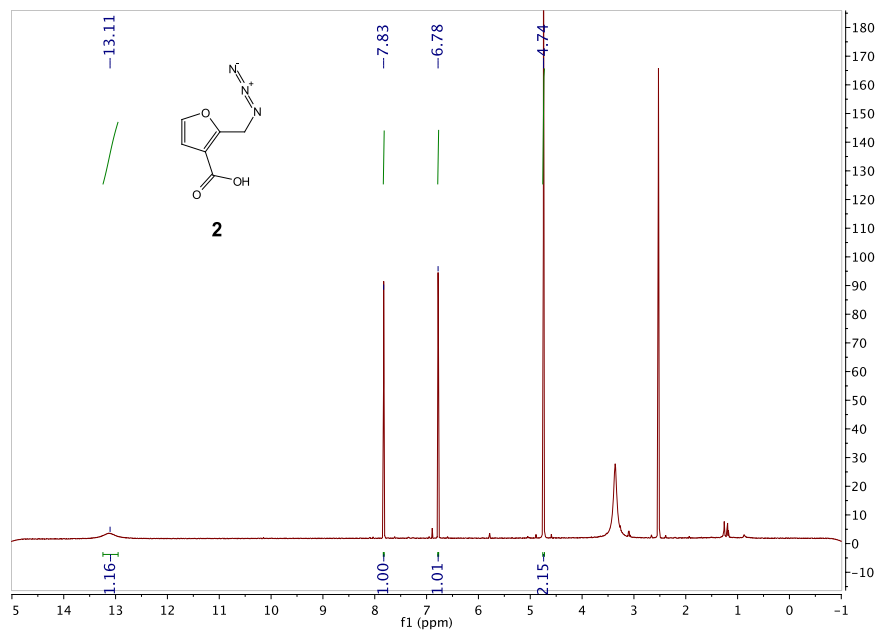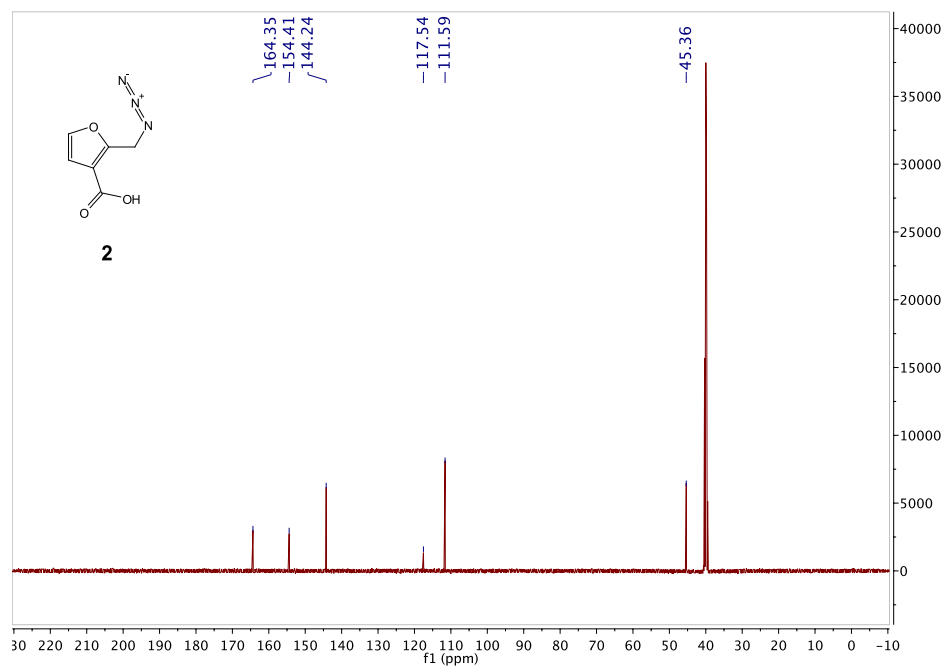
¹³C NMR (500MHz, DMSO)

(2-(azidomethyl)furan-3-yl)(1*H*-imidazol-1-yl)methanone (3)

¹H NMR (500MHz, DMSO)
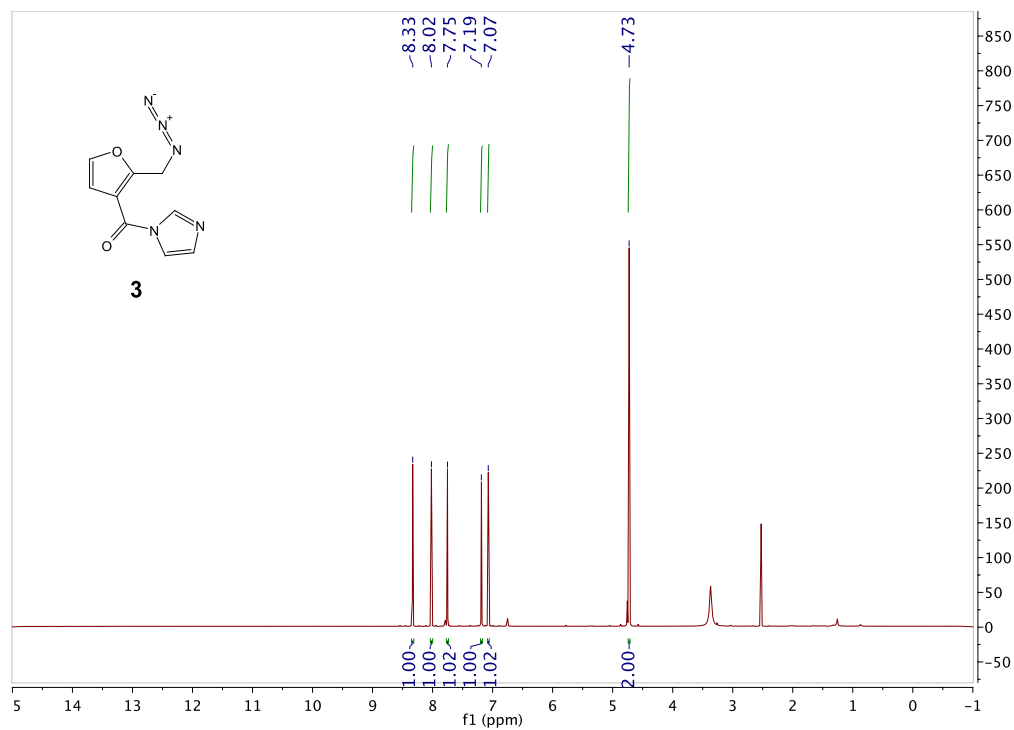
¹³C NMR (500MHz, DMSO)
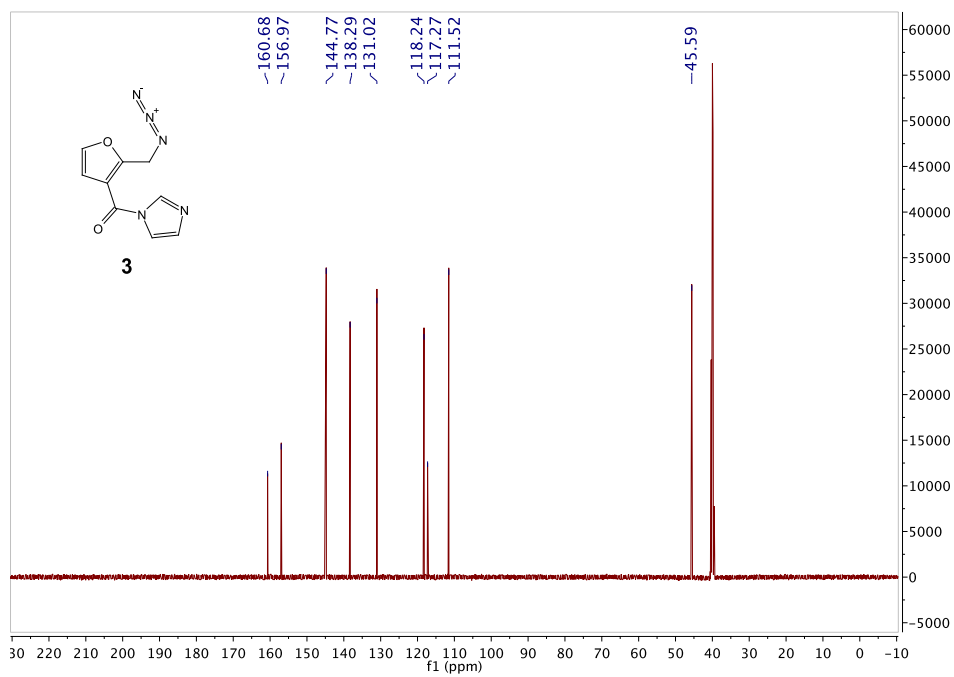
References

1.   Montange, R. K.; Batey, R. T. *Nature* **2006,** *441* (7097), 1172-1175.

2.   Bernstein, L. B.; Manser, T.; Weiner, A. M. *Mol. Cell Biol.* **1985**, 5 (9), 2159-2171

4.7 References

1. Sharp, P. A. (2009) The centrality of RNA, *Cell 136*, 577-580.

2. Serganov, A., and Nudler, E. (2013) A decade of riboswitches, *Cell 152*, 17-24.

3. Bevilacqua, P. C., Ritchey, L. E., Su, Z., and Assmann, S. M. (2016) Genome-Wide Analysis of RNA Secondary Structure, *Annu. Rev. Genet. 50*, 235-266.

4. Weeks, K. M. (2010) Advances in RNA structure analysis by chemical probing, *Curr. Opin. Struct. Biol. 20*, 295-304.

5. Brunel, C., and Romby, P. (2000) Probing RNA structure and RNA-ligand complexes with chemical probes, *Methods Enzymol 318*, 3-21.

6. Merino, E. J., Wilkinson, K. A., Coughlan, J. L., and Weeks, K. M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE), *J. Am. Chem. Soc. 127*, 4223-4231.

7. Mortimer, S. A., and Weeks, K. M. (2008) Time-resolved RNA SHAPE chemistry, *J. Am. Chem. Soc. 130*, 16178-16180.

8. Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T., and Chang, H. Y. (2013) RNA SHAPE analysis in living cells, *Nat. Chem. Biol. 9*, 18-20.

9. Chan, D., Feng, C., Zhen, Y., Flynn, R. A., and Spitale, R. C. (2017) Comparative Analysis Reveals Furoyl in Vivo Selective Hydroxyl Acylation Analyzed by Primer Extension Reagents Form Stable Ribosyl Ester Adducts, *Biochemistry*.

10. Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J. W., Kuchelmeister, H. Y., Batista, P. J., Torre, E. A., Kool, E. T., and Chang, H. Y. (2015) Structural imprints in vivo decode RNA regulatory mechanisms, *Nature 519*, 486-490.

11. Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J. W., Kuchelmeister, H. Y., Batista, P. J., Torre, E. A., Kool, E. T., and Chang, H. Y. (2015) Structural imprints in vivo decode RNA regulatory mechanisms, *Nature 519*, 486-+.

12. Poulsen, L. D., Kielpinski, L. J., Salama, S. R., Krogh, A., and Vinther, J. (2015) SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data, *RNA 21*, 1042-1052.

13. Montange, R. K., and Batey, R. T. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element, *Nature 441*, 1172-1175.

14. Lee, B., Flynn, R. A., Kadina, A., Guo, J. K., Kool, E. T., and Chang, H. Y. (2017) Comparison of SHAPE reagents for mapping RNA structures inside living cells, *RNA 23*, 169-174.

15. Anger, A. M., Armache, J. P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D. N., and Beckmann, R. (2013) Structures of the human and Drosophila 80S ribosome, *Nature 497*, 80-+.

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

17. Kwok, C. K., Ding, Y., Tang, Y., Assmann, S. M., and Bevilacqua, P. C. (2013) Determination of in vivo RNA structure in low-abundance transcripts, *Nat. Commun. 4*, 2971.

18. Hector, R. D., Burlacu, E., Aitken, S., Le Bihan, T., Tuijtel, M., Zaplatina, A., Cook, A. G., and Granneman, S. (2014) Snapshots of pre-rRNA structural flexibility reveal eukaryotic 40S assembly dynamics at nucleotide resolution, *Nucleic Acids Res. 42*, 12138-12154.

19. Pirakitikulr, N., Kohlway, A., Lindenbach, B. D., and Pyle, A. M. (2016) The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures, *Mol. Cell 62*, 111-120.

Chapter 4: Protected pyrimidine nucleosides for cell-specific RNA labeling

4.1 Introduction

Cells exist in complex environments in nature. Understanding the molecular signatures of individual cell types has emerged as a key challenge in biology and medicine. However, precise chemical tools to do so are still overall lacking.[1] RNA molecules, once regarded as transient messengers between the genome and proteome, have now emerged as key players in regulating cell-fate decisions, development, cancer and neurological disorders.[2-4] Cell-specific profiling of RNA molecules, in the form of gene expression analysis, is extremely challenging. Such analyses rely on cell-sorting techniques, which have been demonstrated to be both notoriously dirty and often detrimental to the stability of RNA molecules.[5, 6] Such obstacles manifest themselves in high false-positive rates in data analysis that can often mislead biologists aiming to understand changes in gene expression or the function of RNA.[7] Cell-specific metabolic labeling has emerged as a powerful technique in understanding the proteome and metabolome, yet, with few exceptions, analogous methods for RNA are still lacking. We recently reported the utility of protected nucleoside and nucleobase analogs for cell-specific metabolic labeling of RNA (**Figure 4-1, A**).[8, 9] In such a case an exogenous enzyme can be expressed, within a cell-type of interest, to reveal RNA metabolic intermediates containing bio-orthogonal chemical handles. This approach allows the imaging and isolation of nascent and cell-type specific RNA. Such an approach is extremely powerful as it opens the door for cell-specific RNA profiling with the potential to reveal novel functions and gene expression signatures.

We previously demonstrated the enzyme uracil phosphoribosyl transferase (UPRT) can be used to incorporate 5-ethynyluracil into nascent RNA.[8] Although useful, UPRT is expressed in many lower organisms and as such the host pyrimidine salvage pathway compromises its utility inside these animals.[10] As a viable alternative, we matched a *N*6-phenylacetyladenosine analog with a 2'-azido handle (2'-AzA), with an enzyme, penicillin G amidase (PGA).[8] PGA is a bacterial-specific enzyme that cleaves an amide bond.[11] In our work, cellular RNA is endowed with azido functionality only when cells express PGA. While useful, we reasoned extending the toolset of the PGA enzymatic removal paradigm could further reveal additional nucleosides that can be used for metabolic labeling of RNA.

We reasoned that the azido moiety of 2'-AzA could be prone to reduction inside the cell, and more complicated environments *in vivo*, which would negate its ability to be utilized for affinity and purification experiments. Azide reduction has been observed with many other azido-containing analogs, and modified nucleosides, and can have a gross effect on their activity in cells and animals.[12] We also reasoned that adenosine analogs would be highly biased towards RNAs that have long polyA tails, a hallmark of messenger RNAs and long non-coding RNAs. Our own work with 2'-Az-A has demonstrated that it is largely incorporated into polyA tails.[13] This is a well-documented problem when understanding RNA expression levels through RNA sequencing – the current state of the art method of characterizing RNA expression.[14] Together, these matters underscore the need to expand the scope of novel analogs for cell-specific metabolic labeling of RNA.

4.2 Results and Discussion

We sought to utilize a different nucleobase structure for metabolic labeling of RNA. Penicillin G amidase is our current enzyme of choice and as such we decided to design and synthesize an alkynyl cytidine and uridine, which would be refractory to incorporation into nascent RNA due to the placement of the PGA-specific protecting group (**Figure 4-1, B**). Cytidine and uridine are more reasonable choices for metabolic labeling as they are evenly distributed throughout the genome, and as we have seen before bulky substituents can prevent processing by endogenous kinase enzymes, such as uridine/cytidine kinase 1, thus preventing production of the triphosphate.[8] Furthermore, functionalizing the 5-position with an alkyne is not predicted to alter the sugar orientation or the *anti*-position of the nucleobase to disrupt RNA structure.
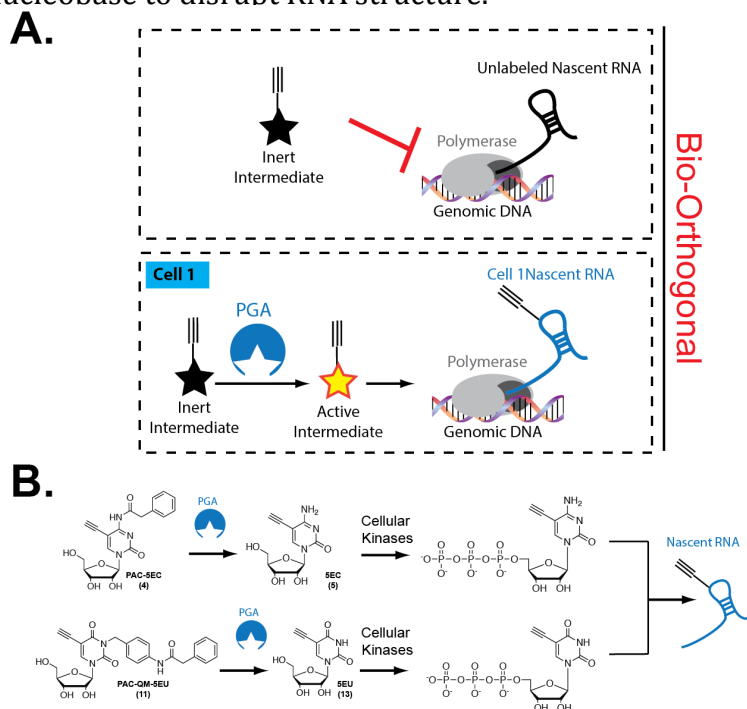
Figure 4-1. Protected analogs for cell-specific metabolic labeling of RNA. (A) Schematic of cell-specific metabolic labeling of RNA. (B) Structures of analogs synthesized herein and their proposed reactions with PGA.

PAC-5EC was synthesized in five steps starting from commercially available 5-iodocytidine (5-IC; ESI). 5-IC was subjected to a Sonogashira coupling to install the trimethylsilyl alkyne handle at the 5-position, followed by protection of the sugar alcohols with *tert*-butyldimethylsilyl groups. To synthesize the PAC portion, 2-phenylacetic acid was treated with DCC to yield 2-phenylacetic anhydride. The exocyclic amine of the silyl-protected cytidine was then reacted with 2-phenylacetic anhydride to install the phenylacetyl group, followed by tetrabutylammonium fluoride deprotection of the silyl groups to yield the final product, PAC-5EC, in an overall yield of 19%. PACQM-5EU was synthesized in 13 steps starting from commercially available uridine (ESI). Iodination of uridine yielded 5-iodouridine, which was then subjected to a Sonogashira coupling to install the trimethylsilyl alkyne handle at the 5 position. To synthesize the PAC portion, *para*-aminobenzylalcohol was coupled to phenylacetic acid using 2-ethoxy-1-ethoxycarbonyl-1,2-dihyrdoquinoline to generate *N*-(4-(hydroxymethyl)phenyl)-2-phenylacetamide. The benzyl alcohol was then brominated with PBr$_3$, followed by Finkelstein reaction to install the iodine. The iodo-intermediate was then coupled with the TMS-protected uridine using potassium carbonate, followed by tetrabutylammonium fluoride deprotection of the TMS group to yield the final product, PACQM-5EU, in an overall yield of 11%.

With PAC-5EC and PACQM-5EU in hand, we next focused on testing whether these analogs would display selectivity for metabolic labeling inside cells (**Figure 4-2**). Briefly, HEK293T cells were grown in culture either with or without PGA transfected with plasmid. To the cells, 200uM of PAC-5EC (5EC as positive control) or PACQM-5EU (5EU as positive
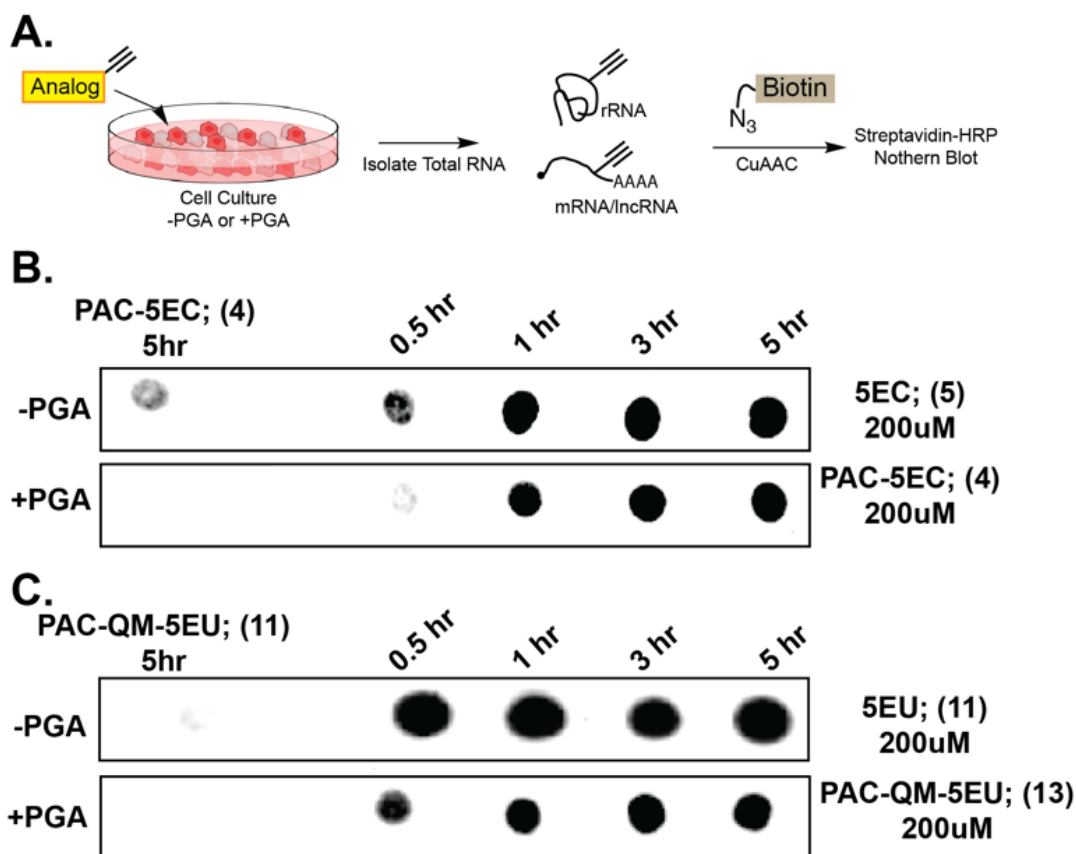
Figure 4-2. **Incorporation of protected analogs into total RNA in HEK cells expressing PGA.** (A) Schematic of experiments to test incorporation. (B) Demonstration of PGA-dependent incorporation of PAC-5EC. (C) Demonstration of PGA-dependent incorporation of PAC-QM-5EU.control) was added. Total RNA was isolated using TriZol. RNA was then appended with bitoin using Cu mediated Alkyne-Azide cycloaddition (CuAAC) and biotin-azide. Dot blot analysis demonstrated that PAC-5EC and PACQM-5EU can both be robustly deprotected within 1 h in PGA-expressing cells. In contrast, when PAC-5EC and PACQM-5EU were incubated in cells not expressing transfected PGA there was nearly undetectable background incorporation (Figure 4-2, B and C). These results demonstrated that PAC-5EC and PACQM-5EU are selectively incorporated into cells expressing PGA.

A major advantage to using analogs that contain alkyne functionality is the ability to image them and their incorporation into cellular RNA; this is in contrast to utilizing 4-thiouridine, which cannot be imaged. We wanted to test whether our novel analogs could

be imaged for RNA incorporation in a PGA-dependent manner. As shown in Figure 4-3, we observed punctate staining within the nucleolus only in PGA transfected cells with the addition of PAC-5EC to the media. A similar labeled pattern was observed with PACQM-5EU (**Figure 4-S1, ESI**). This staining is consistent with what has been observed with other analogs and is known to represent rRNA synthesis within the nucleolus. Staining in the nucleolus nicely demonstrates that our analogs are being robustly incorporated into RNA.
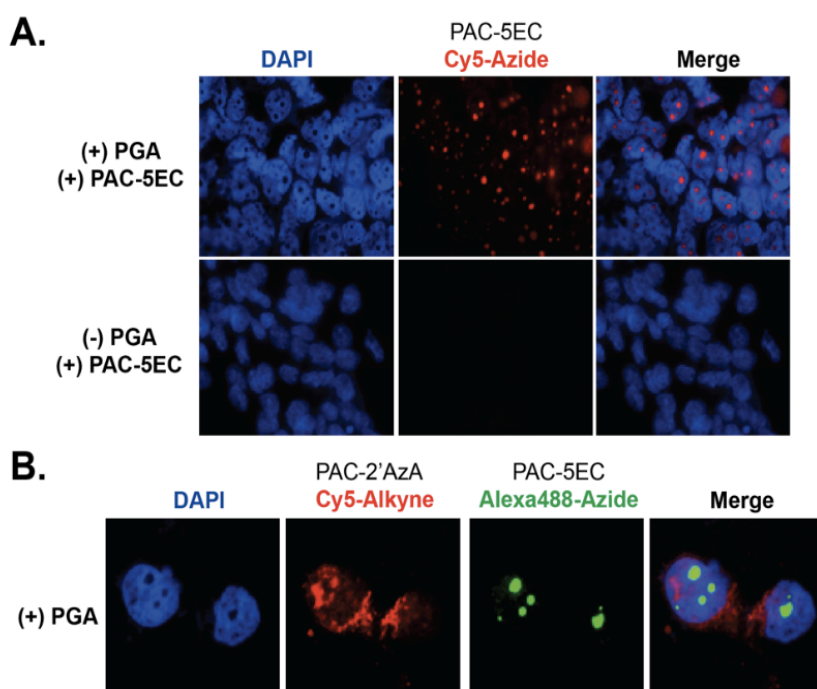
Figure 4-3. Imaging of sub-RNA population-specific incorporation in HEK cells. (A) Imaging demonstrates that PAC-5EC is incorporated into cellular RNA in a PGA-dependent manner. (B) Imaging demonstrates differential RNA incorporation for PAC-5EC and PAC-2'-AzA mediated by PGA.

To test our analogs further, we aimed to demonstrate that multiple analogs could be differentiated from each other using orthogonal chemistries, and we reasoned our work herein with PAC-5EC and PGA could be used to satisfy this goal. In our previous work we demonstrated that PAC-2'-AzA is incorporated into cellular RNA only in the presence of PGA, but that was mostly incorporated into polyA tails, by polyA polymerase. As such, PAC-

2'-AzA staining is almost exclusively cytoplasmic. We compared PAC-2'-AzA and PAC-5EC staining within the same cells and demonstrated their localization patterns are indeed different, further demonstrating that the two analogs have different distributions in RNA. Together, these data further suggest that PAC-5EC (or PACQM-5EU) are likely better analogs for analyzing incorporation into all types of RNAs.

4.3 Conclusion

In conclusion, we have reported the design, synthesis, and utilization of protected alkynyl analogs for cell-specific metabolic labeling of RNA. We have demonstrated that PAC-protected analogs are incorporated into cellular RNA only in the presence of PGA. Furthermore, we have utilized imaging to further demonstrate their utility in imaging RNA, and also that the pyrimidine nucleoside analogs herein are incorporated into a different pool of RNA from our previously reported PAC-2'-AzA. Importantly, we envision these analogs will be used widely for understanding changes in RNA expression in a cell-specific manner, and could also find utility for such efforts *in vivo*.
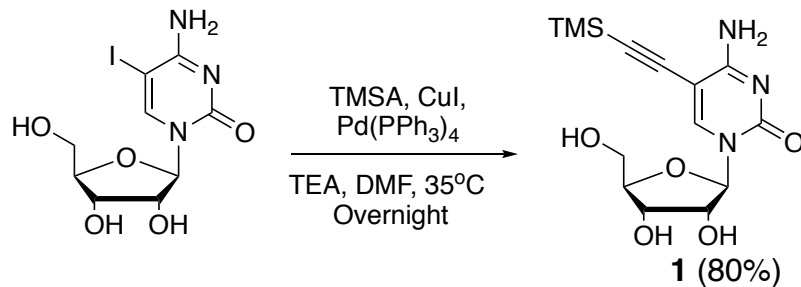
4.4 Acknowledgements

4.5 Methods

1. General

All reagents were purchased from commercial suppliers and were of analytical grade and used without further purification unless otherwise noted. 5-Iodocytidine was purchased from AE Chem Scientific Corporation. Uridine was purchased from Fisher Scientific. Reaction progress was monitored by thin-layer chromatography on EMD 60 F254 plates, visualized with UV light, iodine, ninhydrin, $KMnO_4$, $FeCl_3$, p-anisaldehyde, 2,4-DNP, and bromocresol green stains. Compounds were purified via flash column chromatography using Sorbent Technologies 60 Å 230 x 400 mesh silica gel. Anhydrous solvents acetonitrile (MeCN), dichloromethane (DCM), methanol (MeOH), tetrahydrofuran (THF), dimethylformamide (DMF) were degassed and dried over molecular sieves. Acetone was dried over $MgSO_4$. All reaction vessels were flame dried prior to use. NMR spectra were acquired with Bruker Advanced spectrometers. All spectra were acquired at 298 K. ^{1}H-NMR spectra were acquired at 400 MHz and 500 MHz. ^{13}C-NMR spectra were acquired at 125 MHz. Chemical shifts are reported in ppm relative to residual non-deuterated NMR solvent, and coupling constants (J) are provided in Hz. All NMR spectra was analyzed using MestreNova software. Low and high-resolution electrospray ionization (ESI) mass spectra and Gas Chromatography mass spectra were collected at the University of California-Irvine Mass Spectrometry Facility. IR spectra were acquired from neat samples, unless otherwise noted, with a PerkinElmer Spectrum Two IR Spectrometer.
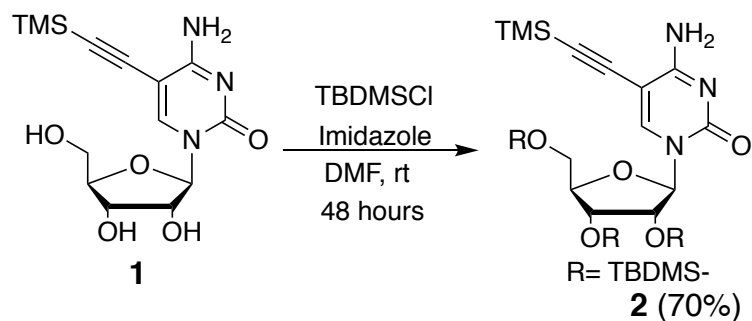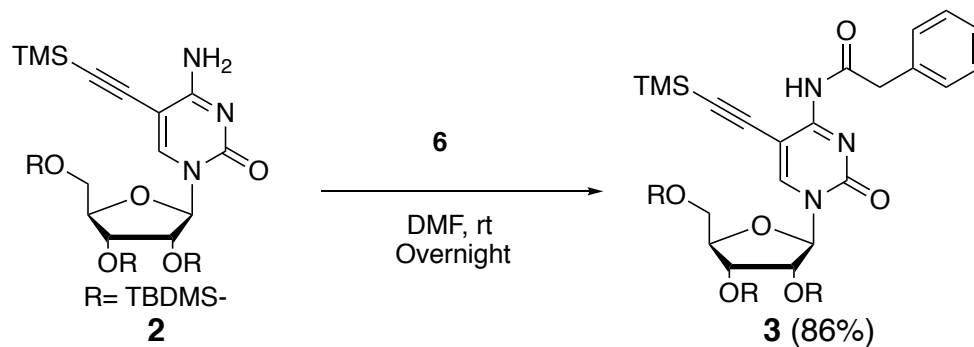
2. Synthesis

1 (80%)

Synthesis of 4-amino-1-((2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5 (hydroxymethyl)tetra-hydrofuran-2-yl)-5-((trimethylsilyl)ethynyl)pyrimidin-2(1*H*)-one (1)

5-iodocytidine (2.71mmol, 1eq), CuI (1.84mmol, 0.68eq), and Pd(PPh$_3$)$_4$ (0.379mmol, 0.14eq) were dissolved in dry DMF (45mL) and dry TEA (35mL). Then TMSA (4.90mmol, 0.698mL, 1.81eq) was added and the reaction was heated to 35°C and left to stir overnight. Volatiles were evaporated and the residue was treated with methanol and filtered over a pad of Celite. The filtrate was concentrated and the residue was purified by silica FC (0 to 10% MeOH/DCM) to yield 0.733g (80% yield) of product **1** as a tan solid. ^{1}H NMR (500MHz, DMSO) δ 8.34 (s, 1H), 7.80 (s, 1H), 6.64 (s, 1H), 5.78 (d, *J*=3.4 Hz, 1H), 5.37 (d, *J*=4.8 Hz, 1H), 5.18 (t, *J*=5 Hz, 1H), 5.00 (d, *J*=5.2 Hz, 1H), 3.99-3.96 (m, 2H), 3.87-3.85 (m, 1H), 3.75-3.71 (m, 1H), 3.61-3.57 (m, 1H), 0.24 (s, 9H). ^{13}C NMR (500MHz, DMSO) δ 164.36, 153.97, 146.26, 100.03, 97.37, 89.84, 84.53, 74.67, 69.29, 60.42, 0.40. HRMS: Theoretical 362.1148 [M+Na[+]], Observed 362.1152 [M+Na[+]].

Synthesis of 4-amino-1-((2*R*,3*R*,4*R*,5*R*)-3,4-bis((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-5-((trimethylsilyl)ethynyl)pyrimidin-2(1*H*)-one (2)

To a solution of **1** (2.16mmol, 1eq) and imidazole (10.8mmol, 5eq) in DMF (15mL) was added TBDMSCl (8.64mmol, 4eq) and the reaction was left to spin overnight at rt for 2 days. Then all volatiles were removed and the residue was resuspended in EtOAc and washed with saturated NH$_4$Cl (aq). Organics were then washed with brine, dried over MgSO$_4$, filtered and concentrated. The residue was purified by silica FC (0 to 100% EtOAc/Hex) to yield 1.01g (70% yield) of product **2** as a while solid. ^{1}H NMR (500MHz, DMSO) δ 7.94 (s, 1H), 7.88 (s, 1H), 6.68 (s, 1H), 5.90 (d, *J*=5.5 Hz, 1H) 4.12 (t, *J*=5 Hz, 1H), 3.97 (d, *J*=2.7 Hz, 1H), 3.89 (dd, *J*=2.2 Hz, 1H), 3.74 (dd, *J*=1.8 Hz, 1H), 0.93 (s, 9H), 0.88 (s, 9H), 0.81 (s, 9H), 0.20 (s, 9H), 0.15 (d, *J*=2.6 Hz, 6H), 0.08 (d, *J*=5.8 Hz, 6H), -0.01 (s, 3H), -0.06 (s, 3H). ^{13}C NMR (500MHz, DMSO) δ 163.73, 153.38, 143.79, 100.17, 96.53, 90.42, 87.47, 84.58, 75.80, 71.68, 62.30, 26.03, 25.70, 25.63, 18.20, 17.76, 17.63, 0.17, -4.60, -4.87, -4.96, -5.38. HRMS: Theoretical 704.3743 [M+Na$^+$], Observed 704.3761 [M+Na$^+$].

Synthesis of *N*-(1-((2*R*,3*R*,4*R*,5*R*)-3,4-bis((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-2-oxo-5-((trimethylsilyl)ethynyl)-1,2-dihydropyrimidin-4-yl)-2-phenylacetamide (3)

To **2** (1.394mmol, 1eq) in dry DMF (12mL) was added 2-phenylacetic anhydride (**6**) (3.067mmol, 2.2eq) and the reaction was allowed to stir at rt for 24h. Then the reaction was diluted with EtOAc and washed with brine. The aqueous layer was back extracted with EtOAc and the combined organic layers were dried over MgSO$_4$, filtered, and concentrated. The residue was purified by silica FC (0 to 100% EtOAc/Hex) to yield 0.958g (86% yield) of product **3** as a white solid. ^{1}H NMR (500MHz, DMSO) δ 9.69 (s, 1H), 8.31 (s, 1H), 7.26-8.30 (m, 5H), 5.79 (d, *J*=3.5 Hz, 1H), 4.25 (t, *J*=4 Hz, 1H), 4.09 (t, *J*=4.5 Hz, 1H), 4.02-4.06 (m, 2H), 3.98 (s, 2H), 3.79 (d, *J*=10.5 Hz), 0.96 (s, 9H), 0.90 (s, 9H), 0.87 (s, 9H), 0.21 (s, 9H), 0.18 (s, 6H), 0.11 (d, *J*=10 Hz, 6H), 0.06 (s, 6H). ^{13}C NMR (500MHz, DMSO) δ 170.36, 161.68, 153.13, 146.87, 135.11, 129.96, 128.76, 127.23, 102.19, 96.53, 94.42, 89.95, 84.39, 76.25, 70.59, 61.86, 43.53, 26.62, 26.19, 26.15, 18.84, 18.20, 18.14, 0.16, -3.92, -4.11, -4.56, -4.62, -4.70, -4.98. HRMS: Theoretical 823.3 [M+Na$^+$], Observed 823.4189 [M+Na$^+$].
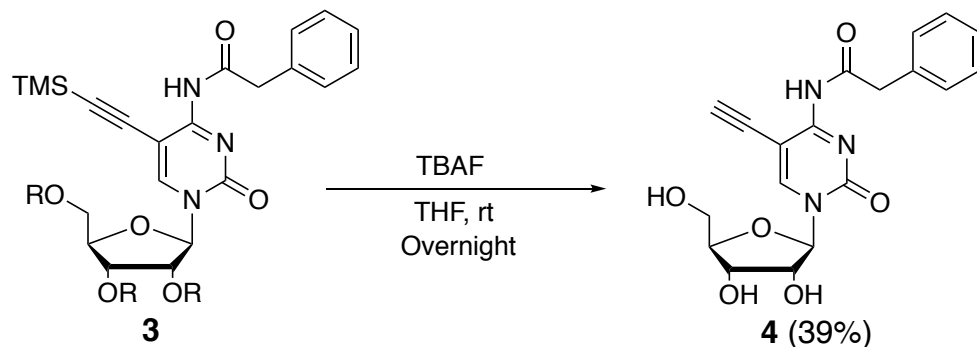
Synthesis of *N*-(1-((2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-ethynyl-2-oxo-1,2-dihydropyrimidin-4-yl)-2-phenylacetamide (N-PAC-5EC, 4)

3 (1.158mmol, 1eq) was dissolved in dry THF (75mL) at 0°C followed by addition of TBAF (5.79mmol, 5.79mL, 5eq) that was previously dried over sieves. The reaction was allowed to warm to rt and spin overnight. Then the reaction was quenched with 1.0M NH_4Cl (aq) by pouring the reaction into the NH_4Cl solution and waiting for the color change (orange to clear, ~10min). Then the mix was extracted with EtOAc multiple times, dried over $MgSO_4$, filtered and concentrated. Then the residue was purified by silica FC (0 to 10% MeOH/DCM) to yield crude product that was suspended in $CHCl_3$ and filtered. The precipitate was collected and dried to yield 0.174g (39% yield) of product **4** as a white solid. ^1H NMR (500MHz, DMSO) δ 9.73 (s, 1H), 8.83 (s, 1H), 7.35-7.23 (m, 5H), 5.72 (d, *J*=1.9 Hz, 1H), 5.58 (d, *J*=4.1 Hz, 1H), 5.34 (t, *J*=4.5 Hz, 1H), 5.03 (d, *J*=5.6 Hz, 1H), 4.39 (s, 1H), 4.02-3.94 (m, 2H), 3.94-3.88 (m, 3H), 3.84-3.76 (m, 1H), 3.63-3.58 (m, 1H). ^{13}C NMR (500MHz, DMSO) δ 169.99, 161.25, 152.76, 148.75, 134.79, 129.54, 128.35, 126.77, 92.91, 90.73, 86.85, 84.00, 75.07, 74.53, 67.94, 59.11, 43.15. HRMS: Theoretical 408.1180 [M+Na$^+$], Observed 408.1172 [M+Na$^+$].
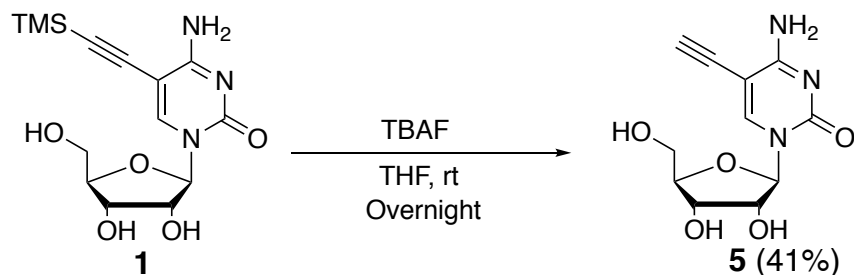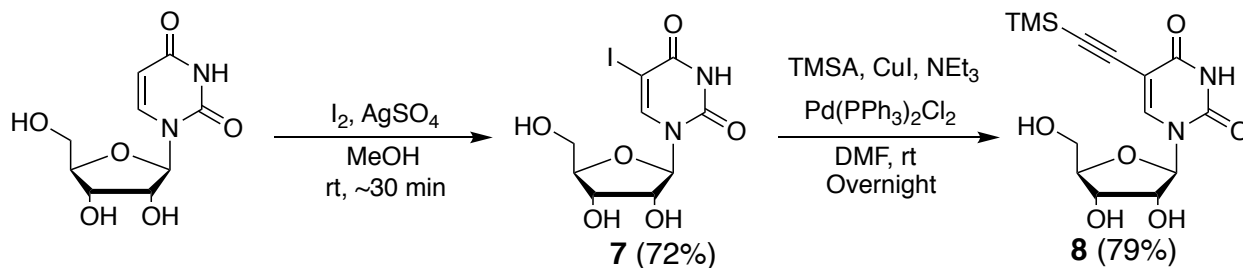
Synthesis of 4-amino-1-((2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5-(hydroxymethyl)tetra-hydrofuran-2-yl)-5-ethynylpyrimidin-2(1*H*)-one (5EC, 5)

2 (0.442mmol, 1eq) was dissolved in dry THF (7mL) followed by dropwise addition of TBAF (0.531mmol, 0.531mL, 1.2eq). The reaction was allowed to spin overnight at room temp and a brown precipitate formed. Then a small amount of methanol was added to dissolve the precipitate and the solution was then dried down and the residue was purified by silica FC (0 to 15% MeOH/DCM). The fractions containing product were dried down and the solid was triturated with DCM to remove any excess TBAF to yield the product **5** as an off-white solid (0.048g, 41% yield). ^{1}H NMR (500MHz, DMSO) δ 8.36 (s, 1H), 7.70 (s, 1H), 6.82 (s, 1H), 5.75 (d, *J*=3.4 Hz, 1H), 5.36 (d, *J*=5 Hz, 1H), 5.18 (t, *J*=5 Hz, 1H), 4.98 (d, *J*=5.4 Hz, 1H), 4.32 (s, 1H), 3.97-3.92 (m, 2H), 3.85-3.83 (m, 1H), 3.72-3.68 (m, 1H), 3.58-3.55 (m, 1H). ^{13}C NMR (500MHz, DMSO) δ 164.16, 153.60, 145.84, 89.45, 88.85, 85.86, 84.02, 75.86, 74.37, 68.80, 59.91. HRMS: Theoretical 290.0753 [M+Na$^+$], Observed 290.0767 [M+Na$^+$].
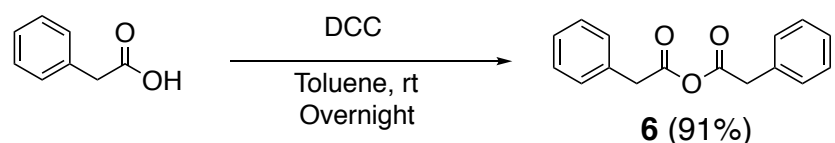
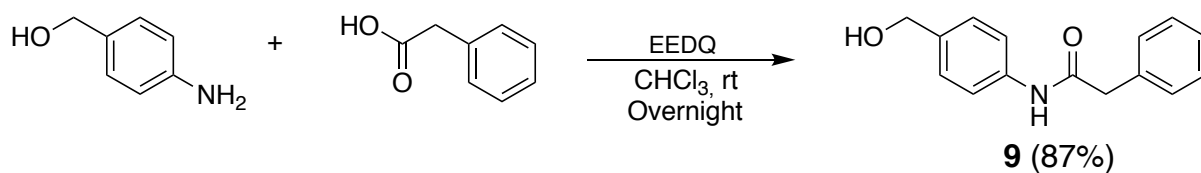**Synthesis of 2-Phenylacetic anhydride (6)**

To a solution of phenylacetic acid (7.34mmol, 1eq) in toluene (20mL) at rt was added DCC (3.67mmol, 0.5eq) and the reaction stirred for 15min. The reaction was then filtered through a pad of Celite (eluent: toluene) and concentrated under reduced pressure to yield final product **6** as a white solid, 1.70g (91% yield). ^{1}H NMR (500MHz, CDCl$_3$) δ 7.36-7.30 (m, 6H), 7.23-7.21 (m, 4H), 3.74 (s, 4H). Spectroscopic data are in accordance with the literature.[1] HRMS: Theoretical 277.0841 [M+Na[+]], Observed 277.0849 [M+Na[+]].

Synthesis of 1-((2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-((trimethylsilyl)ethynyl)pyrimidine-2,4(1*H*,3*H*)-dione (8).

Iodine (2.0mmol, 1eq) in 2mL of methanol was added to a suspension of silver sulfate (2.0mmol, 1eq) and uridine (2.0mmol, 1eq) in 10mL of methanol. The reaction was stirred at room temp until the color disappeared (~30min). The precipitate was removed by filtration over celite. The filtrate was concentrated, giving a pale yellow solid as product, which was used immediately in the next reaction. Product **7** (2.0mmol, 1eq), Pd(PPh$_3$)$_2$Cl$_2$ (0.2mmol, 0.1eq), and CuI (0.4mmol, 0.2eq) were dissolved in dry DMF followed by addition of TEA (10mmol, 1.42mL, 5eq) and TMSA (10mmol, 1.39mL, 5eq). The reaction was stirred overnight at room temp. Then the solvents were removed and the residue was dissolved in MeOH and filtered over celite. The solvent was then removed and the crude

residue was purified by column (2-10% MeOH/DCM) to yield the product **8** as a tan solid (0.537g, 79% yield). ^{1}H NMR (500MHz, DMSO) δ 11.65 (s, 1H), 8.35 (s, 1H), 5.73 (d, *J*=4.6Hz, 1H), 5.43 (d, *J*=5.4Hz, 1H), 5.21 (t, *J*=4.7Hz, 1H), 5.08 (d, 5.4Hz, 1H), 4.04 (q, *J*=5.0Hz, 1H), 3.98 (q, *J*=5.0Hz, 1H), 3.86 (m, 1H), 3.68 (m, 1H), 3.57 (m, 1H), 0.18 (s, 9H). ^{13}C NMR (500MHz, DMSO) δ 161.41, 149.64, 144.92, 98.31, 97.89, 97.08, 88.41, 84.72, 73.86, 69.24, 60.17, -0.02. HRMS: Theoretical 363.0988 [M+Na$^+$], Observed 363.0988 [M+Na$^+$].

Synthesis of *N*-(4-(hydroxymethyl)phenyl)-2-phenylacetamide (9).

To a solution of p-aminobenzyl alcohol (4.06mmol, 1eq) and phenylacetic acid (4.06mmol, 1eq) in chloroform (70mL) was added EEDQ (4.47mmol, 1.1eq). The reaction was left to spin overnight at room temp. Then the solvent was evaporated and the residue was resuspended in EtOAc and washed with 0.1M HCl solution. The aqueous layer was extracted three times with EtOAc. Organics were washed with brine, dried over MgSO$_4$, filtered and concentrated. The residue was purified by column (0 to 100% EtOAc/Hex) to yield the product **9** as a white solid (0.8554g, 87% yield). ^{1}H NMR (500MHz, DMSO) δ 10.15 (s, 1H), 7.57 (d, *J*=8.3Hz, 2H), 7.36 (m, 4H), 7.26 (d, *J*=8.3Hz, 3H), 5.12 (t, *J*=5.7Hz, 1H), 4.45 (d, *J*=5.6Hz, 2H), 3.65 (s, 2H). ^{13}C NMR (500MHz, DMSO) δ 169.40, 138.28, 137.82, 136.53, 129.57, 128.77, 127.39, 126.98, 119.32, 63.06, 43.78. HRMS: Theoretical 264.1000 [M+Na$^+$], Observed 264.0993 [M+Na$^+$].
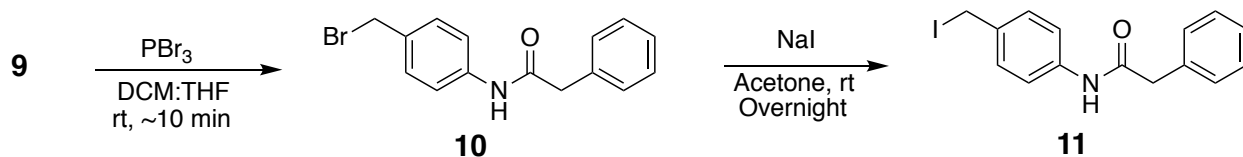
Synthesis of *N*-(4-(bromomethyl)phenyl)-2-phenylacetamide (10) and *N*-(4-(iodomethyl)phenyl)-2-phenylacetamide (11).

N-(4-(hydroxymethyl)phenyl)-2-phenylacetamide (1.32mmol, 1eq) was dissolved in a 1:1 solution of DCM:THF (20mL:20mL) at 0°C. Then PBr$_3$ (0.436mmol, 0.041mL, 0.33eq) was added dropwise. TLC showed complete consumption of the starting material after 10min. The reaction was then quenched with ice water, extracted with CHCl$_3$, dried over MgSO$_4$, filtered and concentrated. The residue was purified by column (50/50 EtOAc/Hex) to yield product **10** as a white solid. HRMS: Theoretical: 326.0157 [M+Na$^+$]and 328.0137 [M+Na$^+$+2], Observed: 326.0157 [M+Na$^+$]and 328.0139 [M+Na$^+$+2]. The product **10** (1.32mmol, 1eq) was then dissolved in dry acetone (5mL) followed by addition of NaI (1.32mmol, 1eq). The reaction was left to spin overnight at room temp. The NaBr precipitate was filtered off and the filtrate was dried down to yield a yellow solid, TLC showed complete consumption of starting material, and the product **11** was used in the coupling reaction immediately. HRMS: Theoretical: 374.0018 [M+Na$^+$], Observed: 374.0004 [M+Na$^+$].
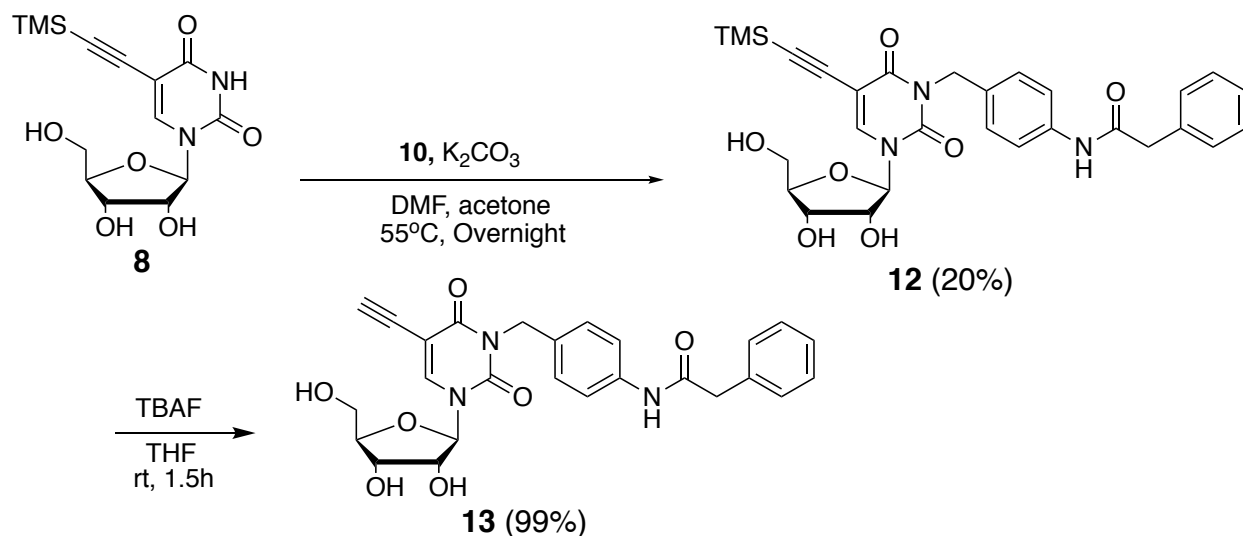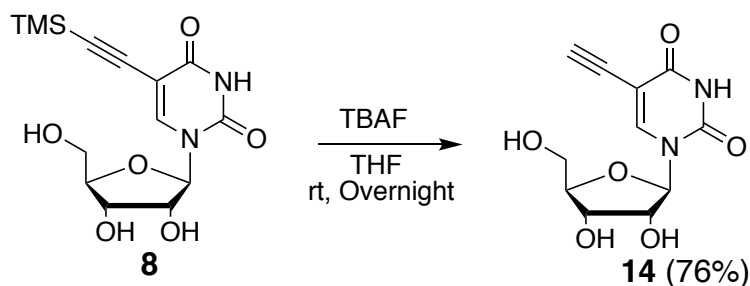
Synthesis of N-(4-((3-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetra-hydrofuran-2-yl)-5-ethynyl-2,6-dioxo-2,3-dihydropyrimidin-1(6H)-yl)methyl)phenyl)-2-phenylacetamide (13; PAC-QM-5EU).

To the residue of *N*-(4-(iodomethyl)phenyl)-2-phenylacetamide (1.32mmol, 1.5eq) was added the solution of DMF (3mL) and acetone (3mL) followed by the addition of **8** (0.88mmol, 1eq) and K_2CO_3 (1.50mmol, 1.7eq). The reaction was then heated to 55°C and left to stir overnight. The reaction was then filtered and the filtrate was then evaporated and the resulting residue was purified by column (1 to 5% MeOH/DCM) to yield a mixture of TMS and no TMS products (0.090g, 20% yield). The combined products (0.160mmol, 1eq) were dissolved in dry THF (2.5mL) followed by addition of 1.0M TBAF in THF (0.192mmol, 0.192mL, 1.2eq). The reaction was left to spin for 1.5h at room temp. Then solvents were evaporated and the residue was purified by column (0 to 5% MeOH/DCM) to yield the product **13** as a white powder (0.078g, 99% yield). ^{1}H NMR (500MHz, DMSO) δ 10.17 (s, 1H), 8.50 (s, 1H), 7.52 (d, *J*=8.5Hz, 2H), 7.32 (d, *J*=4.4Hz, 4H), 7.23 (m, 3H), 5.76 (d, *J*=3.9Hz, 1H), 5.48 (d, *J*=5.2Hz, 1H), 5.28 (t, *J*=9.4Hz, 1H), 5.07 (d, *J*=7.4Hz, 1H), 4.91 (d,

J=3.1Hz, 2H), 4.14 (s, 1H), 4.05 (q, *J*=4.7Hz, 1H), 3.99 (q, *J*=5.3Hz, 1H), 3.87 (m, 1H), 3.71 (m, 1H), 3.61-3.58 (m, 3H). ^{13}C NMR (500MHz, DMSO) δ 169.05, 160.79, 149.77, 143.38, 138.39, 136.00, 131.38, 129.09, 128.41, 128.31, 126.53, 119.04, 96.84, 89.76, 84.58, 83.57, 76.42, 74.09, 68.86, 59.83, 43.72, 43.28. HRMS: Theoretical 514.1590 [M+Na$^+$], Observed 514.1591 [M+Na$^+$].

Synthesis of 1-((2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-ethynylpyrimidine-2,4(1*H*,3*H*)-dione (14).

Compound **8** (1.47mmol, 1eq) was dissolved in dry THF (10mL) followed by addition of TBAF (1.0M in THF, 1.47mmol, 1.47mL, 1eq). The reaction was left to stir overnight at room temp. Then the reaction was dried down and the residue was purified by column (0-10% MeOH/DCM). The corresponding tubes were dried down and the solid was triturated with DCM and dried to yield product **14** as a white solid (0.299g, 76% yield). ^{1}H NMR (500MHz, DMSO) δ 11.67 (s, 1H), 8.41 (s, 1H), 5.77 (d, *J*=4.6Hz, 1H), 5.46 (d, *J*=5.3Hz, 1H), 5.27 (t, *J*=4.7Hz, 1H), 5.10 (d, *J*=5.3Hz, 1H), 4.13 (s, 1H), 4.07 (q, *J*=4.9Hz, 1H), 4.01 (q, *J*=4.8Hz, 1H), 3.89 (m, 1H), 3.70 (m, 1H), 3.60 (m, 1H). ^{13}C NMR (500MHz, DMSO) δ 162.10, 150.14, 145.14, 98.09, 88.87, 85.20, 84.14, 76.81, 74.45, 69.78, 60.66. HRMS: Theoretical 291.0593 [M+Na$^+$], Observed 291.0598 [M+Na$^+$].

3. Biochemical Methods

Cloning and plasmids.

PWIG-1xFlag-PGA-IRES-eGFP, abbreviated pPGA, (a kind gift from Dr. Michael S. Cohen) was digested with XmaI and NotI flanking IRES-eGFP sequence (New England BioLabs, Inc., MA). The backbone was gel purified (Qiaquick Gel Extraction Kit; Cat# 28704) and ligated (Thermo Scientific Rapid DNA Ligation Kit, Cat# K1423) to double stranded oligos (IDT, Inc.; oKN101FL: 5'- CCGGG AAGCTT GATA GC-'3 and oKN102RL: 5'- GGCCGC TATC AAGCTT C-'3) resulting plasmid pKN317-1xFlag-PGA. The plasmid was verified by Sanger sequencing analysis (Genewiz, Inc.).

Cell lines and culture conditions.

HEK cell lines were cultured in DMEM supplemented with 10% FBS, penicillin and streptomycin and grown at 37°C, 5% CO_2. PWIG-1xFlag-PGA-IRES-eGFP, abbreviated pPGA, (a kind gift from Dr. Michael S. Cohen) was used to transiently transfect HEK cells.

Azido-nucleoside and ethynyl-nucleoside labeling of cellular RNA.

Azido-nucleoside and ethynyl-nucleoside analogs were added to complete culture medium from 400 mM and 200 mM stocks with a final concentration of 1 mM and 200 μM at <1% DMSO, respectively. For time course study, analog was exposed at 0h, 0.5, 1, 3 and 5 hours. For titration of analog concentration, analog was exposed for 5 hours.

RNA isolation and biotinylation via CuAAC.

After labeling, total cellular RNA was harvested using Trizol Reagent (Invitrogen) following the manufacturers instructions. Click reactions were prepared using 30 μg of total RNA, 1 mM biotin alkyne or biotin azide, and fresh 4.6 mM THPTA to a final concentration of 1 mM, fresh 10.6 mM NaAsc to a final concentration of 1.77 mM, and 12 mM $CuSO_4$ to a final

94

concentration of 200 μM. The reactions were incubated at room temperature on shaker for 30 min. The reactions were purified using Qiagen RNAeasy Mini Purification kit according to the manufacturer instructions (Qiagen), and RNA was eluted in 47 μL of nuclease free water.

HRP-streptavidin dot blot analysis.

All gel reagents were from Bio-Rad. Equal amounts of column-purified RNA were applied onto Hybond-N+ membrane (GE Healthcare) as individual dots or loaded using nucleic acid loading dye and were separated on a native 1% agarose gel. RNA was transferred onto Hybond-N+ membrane using a standard vacuum blotter (Biometra, Analytic Jena Company), and UV-crosslinked to a membrane (Stratalinker UV crosslinker). Membranes were blocked followed by incubation with high sensitivity streptavidin-HRP (ThermoFisher Scientific). The membrane was washed twice in a 1:10 solution of blocking buffer and twice in Tris-saline buffer. It was then incubated in ECL Chemiluminescent Substrate (ThermoFisher Scientific) and imaged on a ChemiDoc MP imaging system (Bio-Rad).

RNA fluorescence imaging via CuAAC.

HEK cells were seeded at 2.5×10^5. HEK cells were transiently transfected with 5 μg pWIG-1xFLAG-PGA-IRES-eGFP on the following day using *jet*PRIME transfection reagent according to manufacturer's manual (Polyplus Transfection, France) and grown on glass cover slips 67 hours post-transfection. Cells were treated with 1 mM PAC-5EC or PAC-QM-5EU, and incubated for 18 hours. After labeling, cells were washed three times with DPBS, and fixed and permeabilized for 30 min at room temperature with 3.7% paraformaldehyde and 0.15% Triton-X100. Cells were then washed three times (7 min/each) on orbital

shaker with DPBS, blocked with BSA (1 mg/ml in DPBS, 0.45% NaCl and 0.025% NaN$_3$) for 30 min at room temperature, washed twice with DPBS, and incubated with 500 μL of click solution (1 mM CuSO$_4$, 2 mM THPTA ligand, 10 mM NaAsc, and 15 μM Azide-Alexa 488 or Azide-Cy5) for 1 hour at 37°C in the dark. Cells were washed three times for 5 min/each on an orbital shaker: twice with DPBS-0.1% Triton-X100 and one with DPBS, and mounted using VectaShield with DAPI, 4',6-diamidino-2-phenylindole (Vector Labs). Slides were imaged via fluorescence confocal microscopy using a 40x oil immersion objective on a Leica 700 Carl Zeiss microscope.

RNA fluorescence imaging of PGA transiently transfected cell lines exposed to both PAC-5EC and PAC-2'-AzA via CuAAC.

Cell seeding and transfection parameters were similar to single analog exposure mentioned above. However, in this experiment a different construct of PGA (pKN317 = 1xFLAG-PGA) which does not have GFP, was used for transfection. To transfected cells, both PAC-5EC and PAC-2'-AzA were added at the same time to 1mM final concentration. After 18 h of analog exposure (i.e. 85 h post-transfection), cells were washed three times with DPBS buffer and fixed-permeabilized for 30 min at room temperature with 3.7% paraformaldehyde-0.15% Triton-X100. Cells were then washed three times (7 min/each) on an orbital shaker with DPBS, blocked with BSA (1 mg/ml in DPBS, 0.45% NaCl and 0.025% NaN$_3$) for 30 min at room temperature, washed twice with DPBS, and incubated with 250 μL of click solution (1 mM CuSO$_4$, 2 mM THPTA ligand, 10 mM NaAsc, and 15 μM azide-Alexa 488) for 1 hour at 37°C in the dark. After first click reaction, cells were washed once with DPBS-0.1% Triton-X100 for 10 min and twice with DPBS for 20 min/each. The cells were subjected to a second click reaction with 250 μl of the second click solution (1 mM CuSO$_4$, 2 mM THPTA

96

ligand, 10 mM NaAsc, and 15 μM alkyne-Cy5) and incubated at 37°C for 1 hour in the dark. After second labeling, cells were washed three times with DPBS-0.1% Triton-X100 and once with DPBS for 5 min/each. The processed cover slides were mounted using VectaShield with DAPI, 4',6-diamidino-2-phenylindole (Vector Labs). Slides were imaged via fluorescence confocal microscopy using a 40x oil immersion objective on a Leica 700 Carl Zeiss microscope.

4.6 Supplementary Data
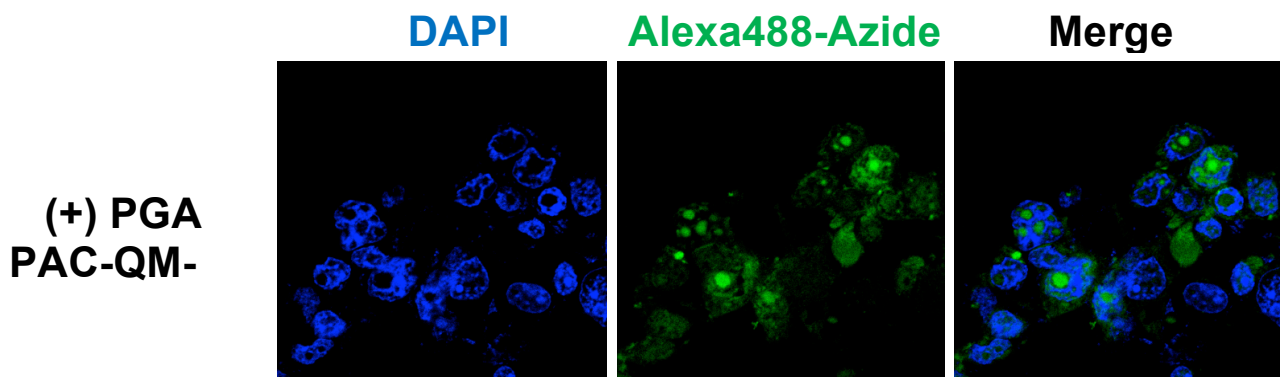
Supplementary Figure 4-1. Microscopy imaging of RNA incorporation of caged PAC-QM-5EU in PGA-transiently transfected cells. DAPI stained DNA, Alexa488-Azide appended into RNA containing 5EU, which is a decaged product of PAC-QM-5EU mediated by PGA via CuAAC. Merge is a composite of DAPI and 5EU-stained by Alexa488-azide
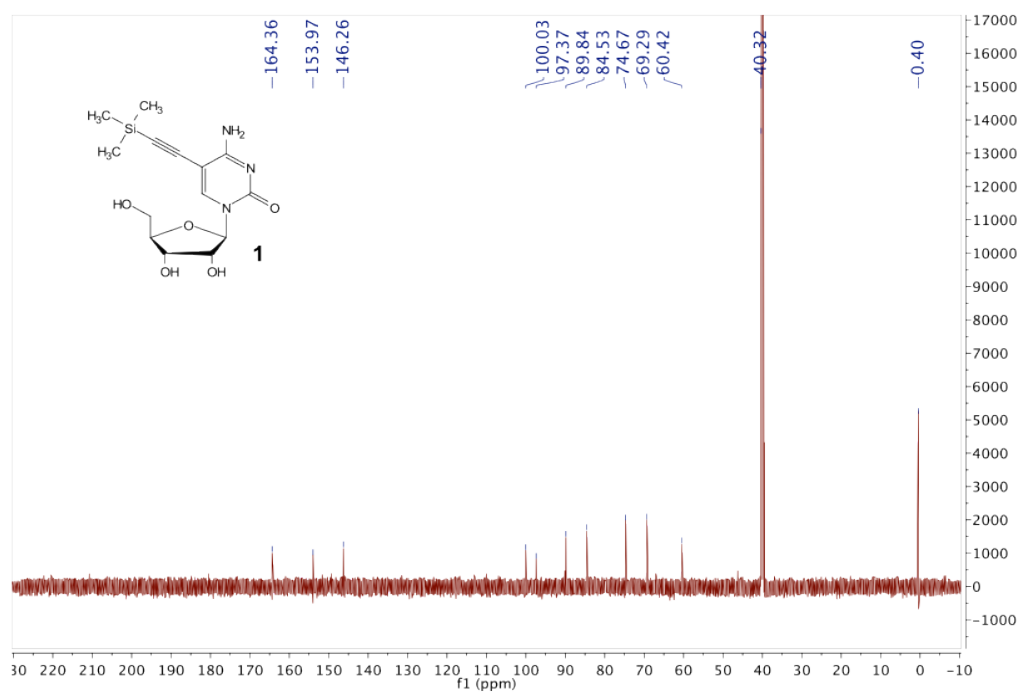
5. Spectra:

4-amino-1-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-

5-((trimethylsilyl)ethynyl)pyrimidin-2(1H)-one(1)

¹H NMR (500MHz, DMSO)
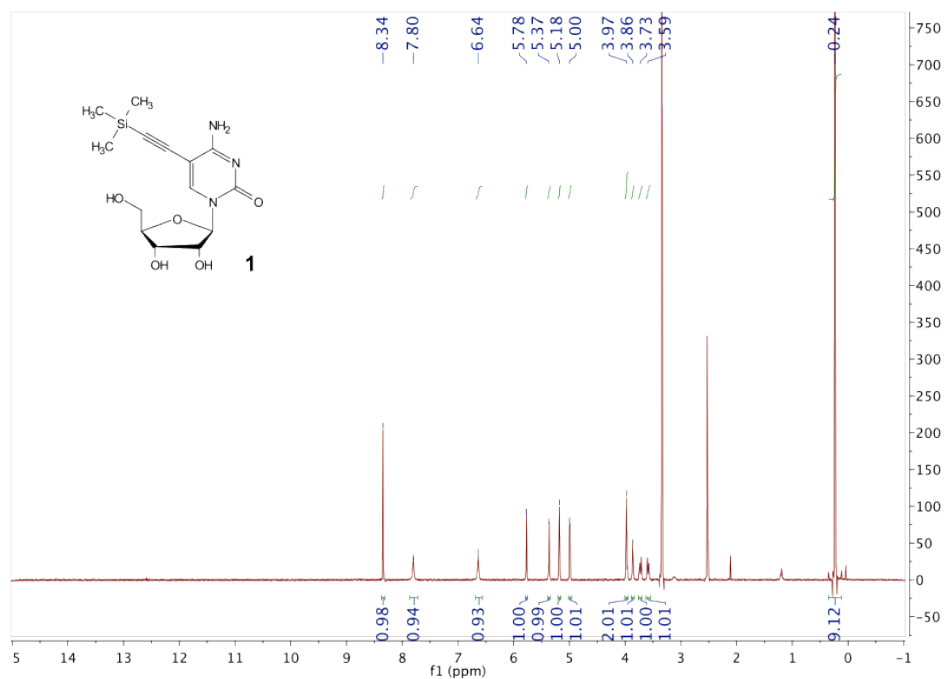
¹³C NMR (500MHz, DMSO)

4-amino-1-((2*R*,3*R*,4*R*,5*R*)-3,4-bis((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-5-((trimethylsilyl)ethynyl)pyrimidin-2(1*H*)-one (2)

¹H NMR (500MHz, DMSO)

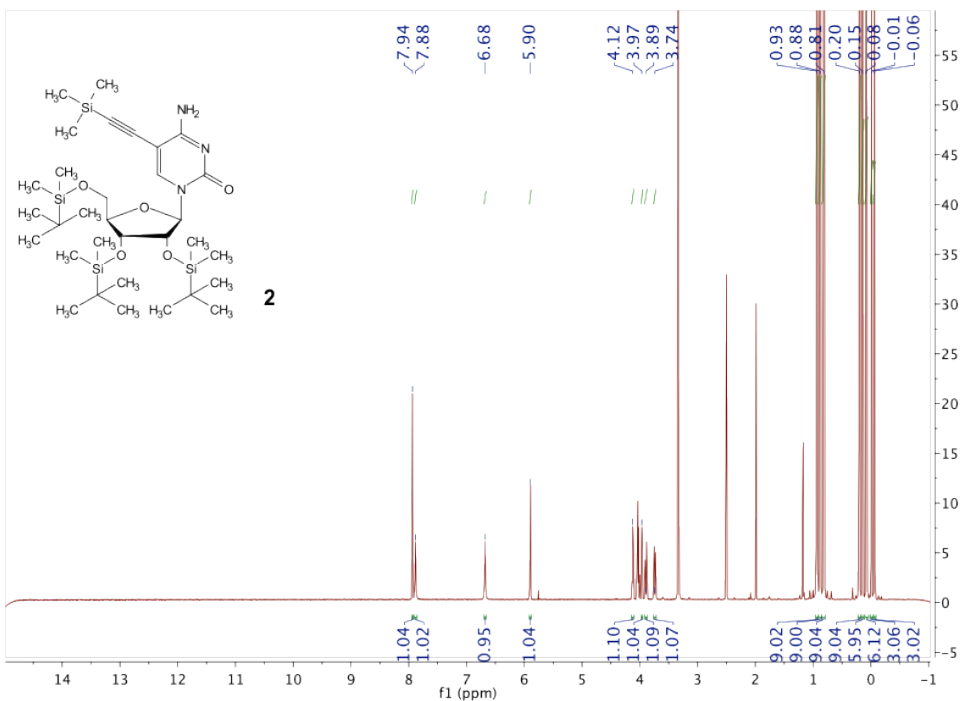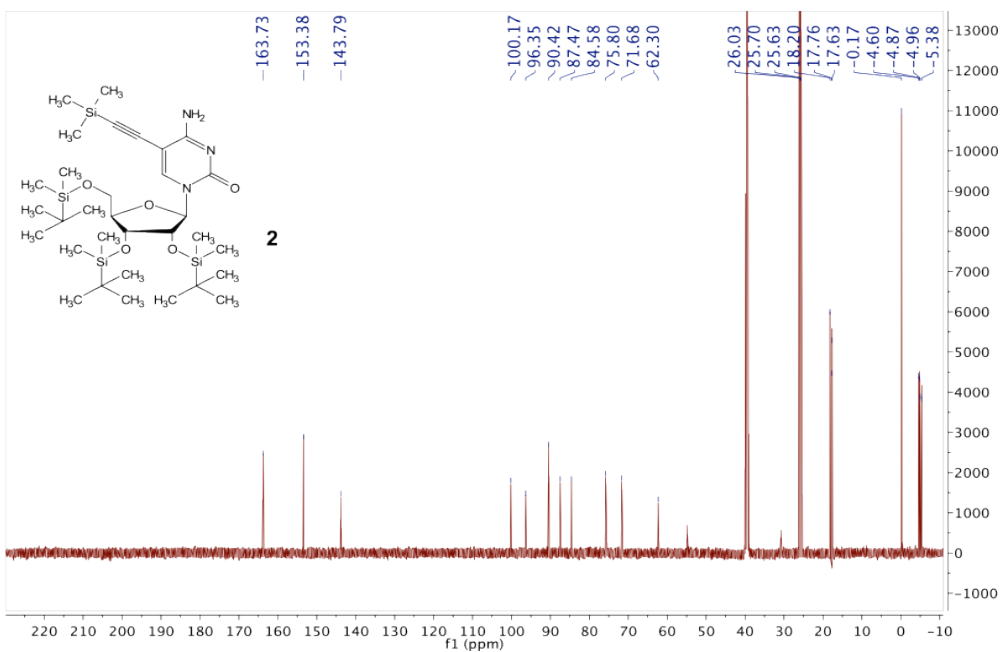¹³C NMR (500MHz, DMSO)

N-(1-((2R,3R,4R,5R)-3,4-bis((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-2-oxo-5-((trimethylsilyl)ethynyl)-1,2-dihydropyrimidin-4-yl)-2-phenylacetamide (3)

¹H NMR (500MHz, DMSO)
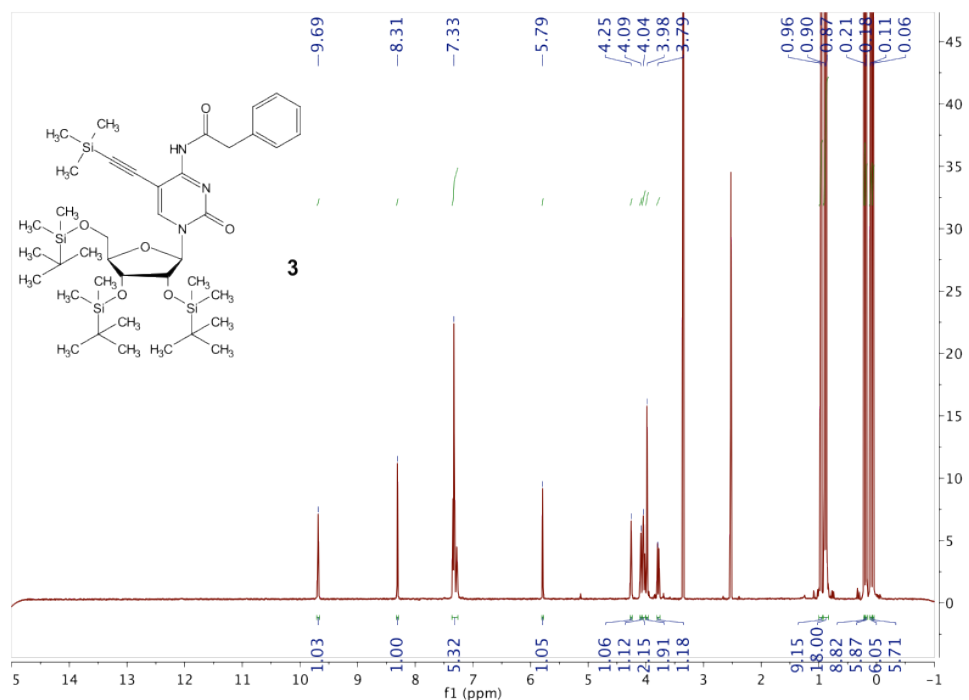
¹³C NMR (500MHz, DMSO)

Synthesis of *N*-(1-((2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-ethynyl-2-oxo-1,2-dihydropyrimidin-4-yl)-2-phenylacetamide (N-PAC-5EC, 4)

^{1}H NMR (500MHz, DMSO)
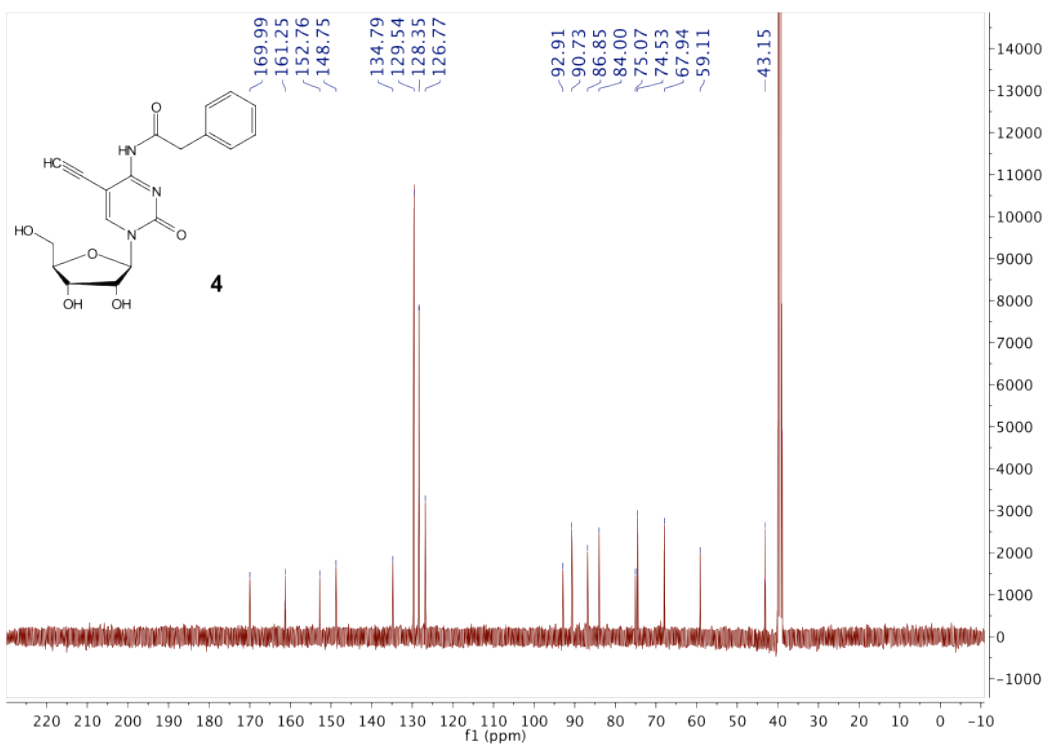
^{13}C NMR (500MHz, DMSO)

4-amino-1-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-ethynylpyrimidin-2(1H)-one (5EC, 5)

¹H NMR (500MHz, DMSO)

¹³C NMR (500MHz, DMSO)

2-Phenylacetic anhydride (6)

Spectroscopic data are in accordance with the literature.[1]

^{1}H NMR (500MHz, CDCl$_3$)

1-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-((trimethylsilyl)ethynyl)pyrimidine-2,4(1H,3H)-dione (8).

^{1}H NMR (500MHz, DMSO)
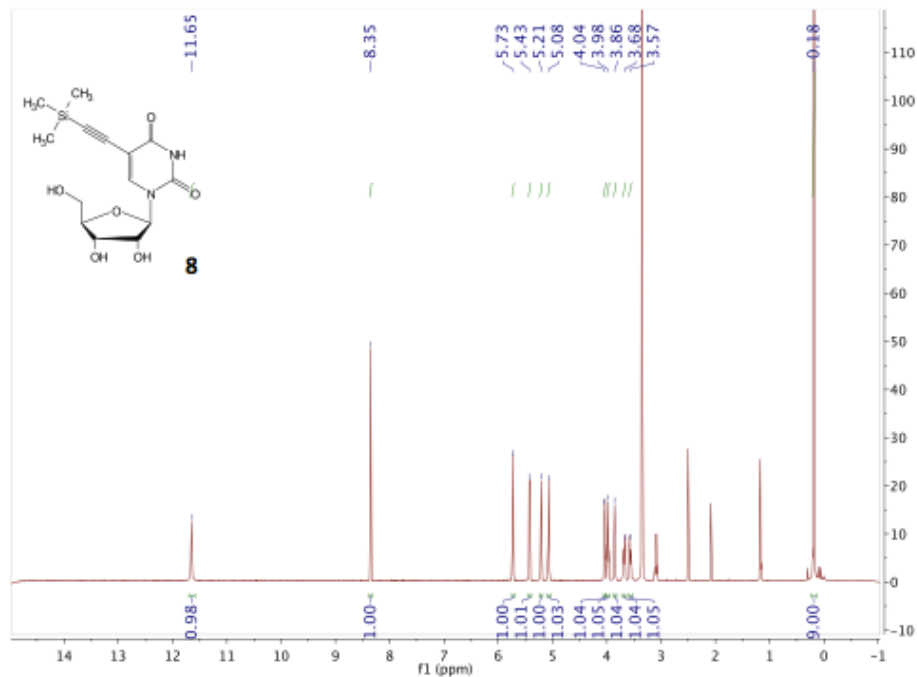
^{13}C NMR (500MHz, DMSO)
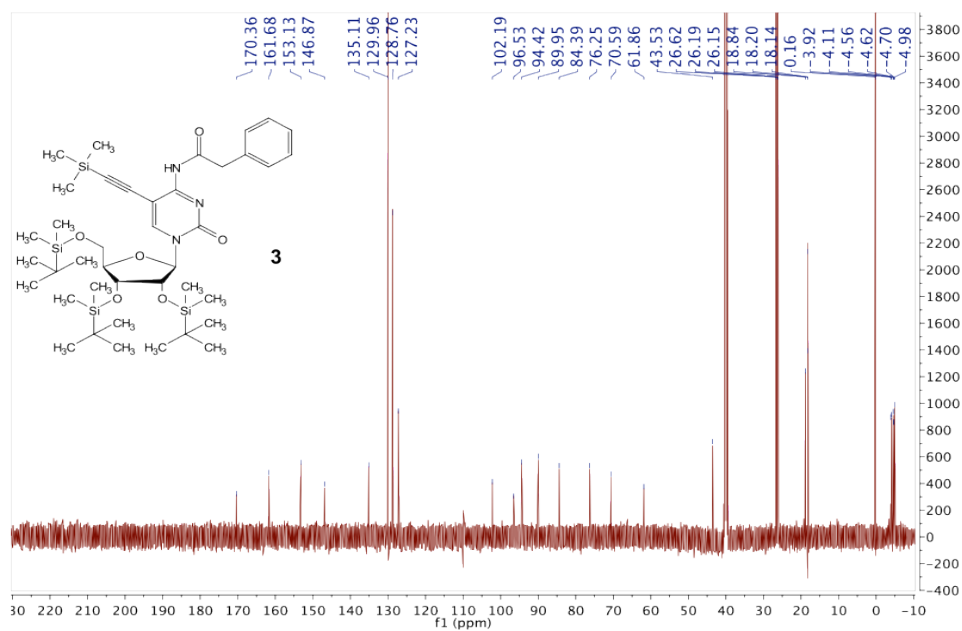
N-(4-(hydroxymethyl)phenyl)-2-phenylacetamide (9).

¹H NMR (500MHz, DMSO)
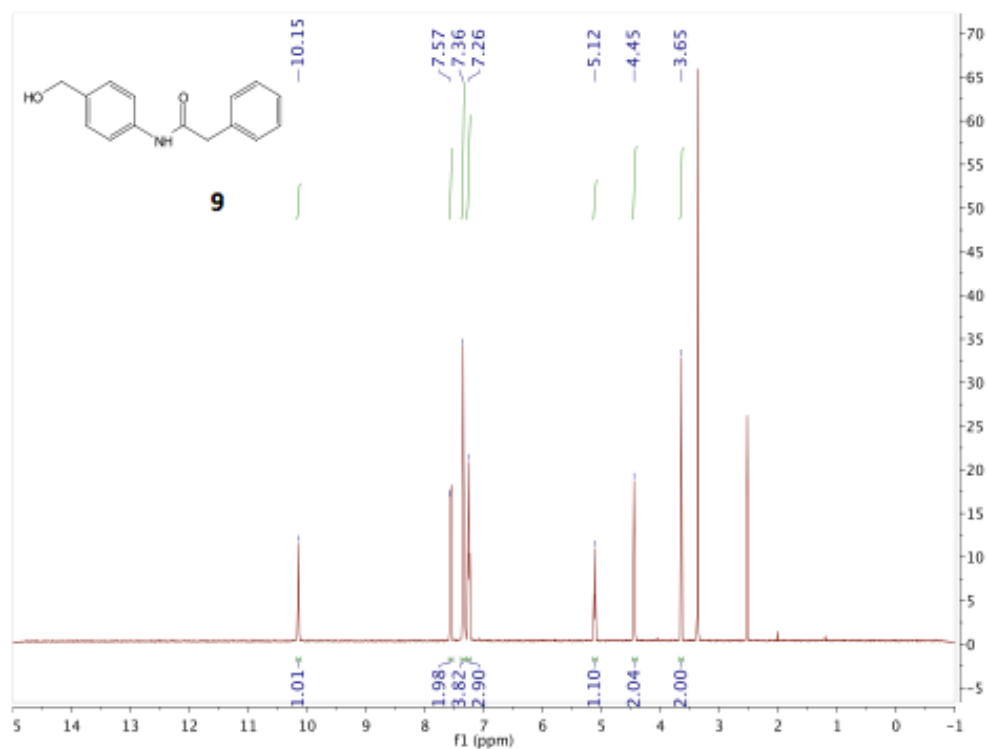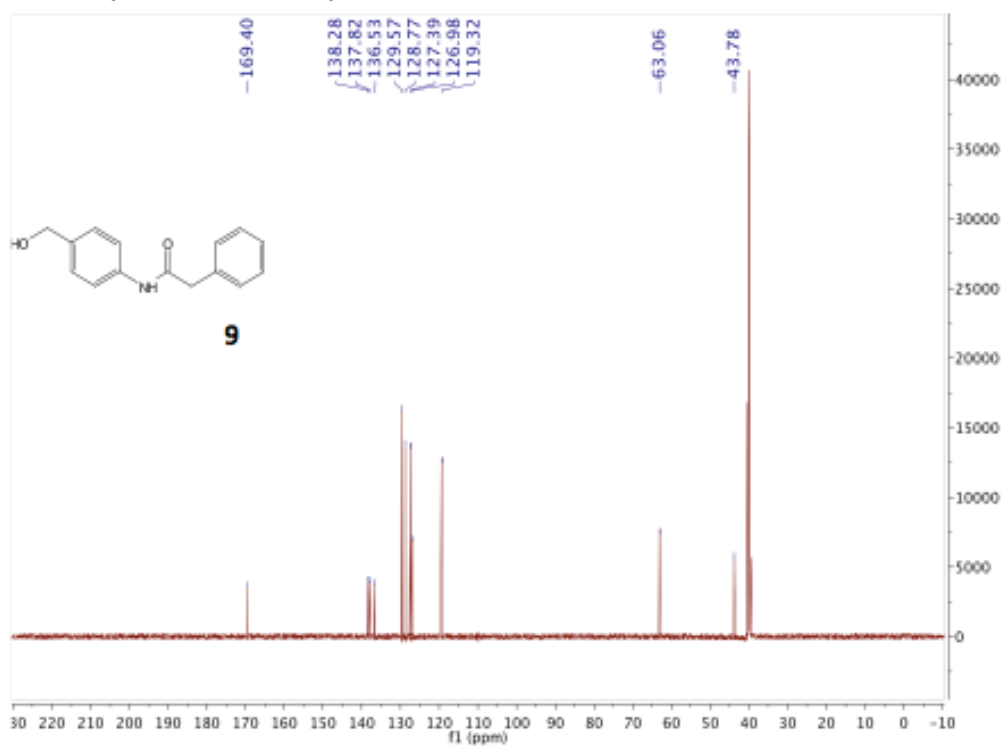
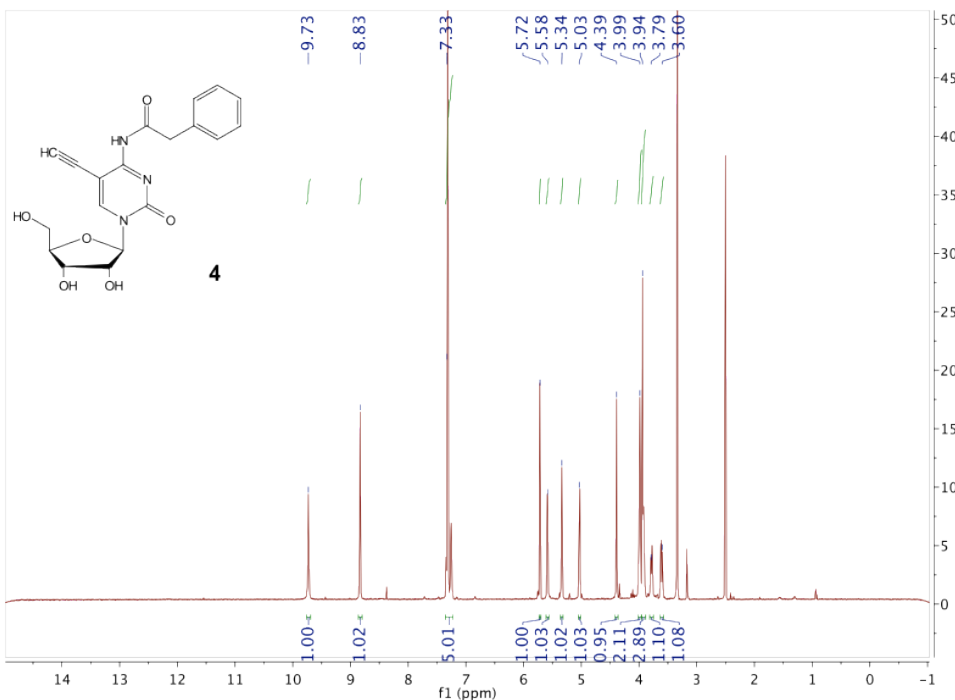¹³C NMR (500MHz, DMSO)

N-(4-((3-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-ethynyl-2,6-dioxo-2,3-dihydropyrimidin-1(6H)-yl)methyl)phenyl)-2-phenylacetamide (13; PAC-QM-5EU).

¹H NMR (500MHz, DMSO)
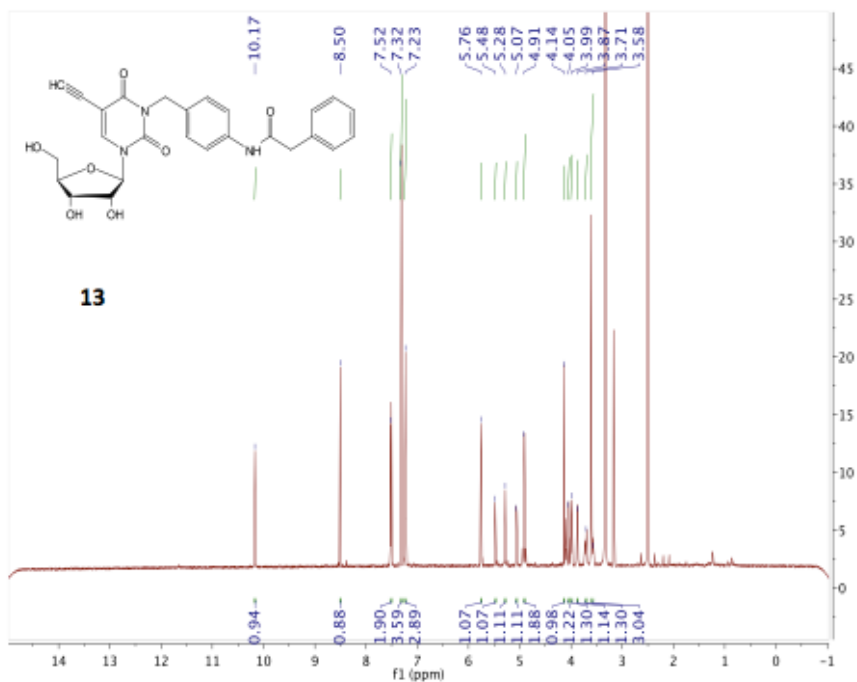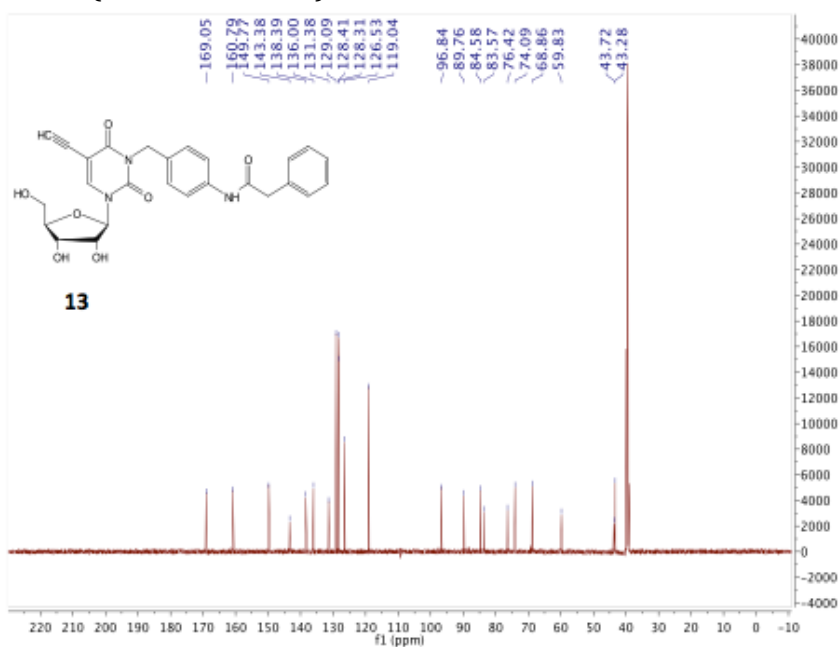
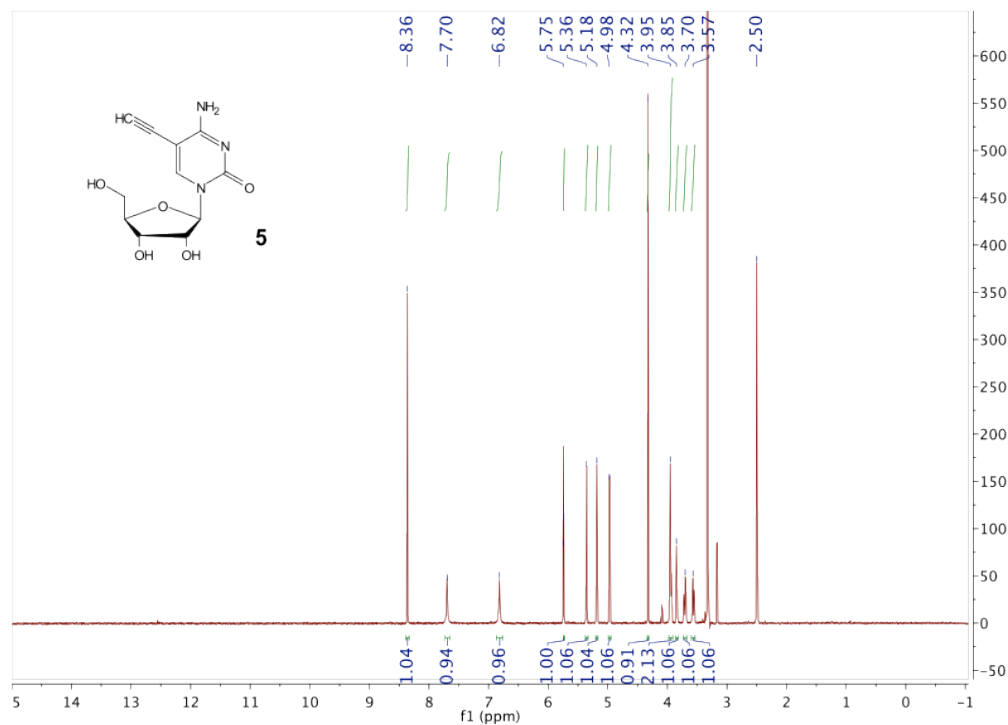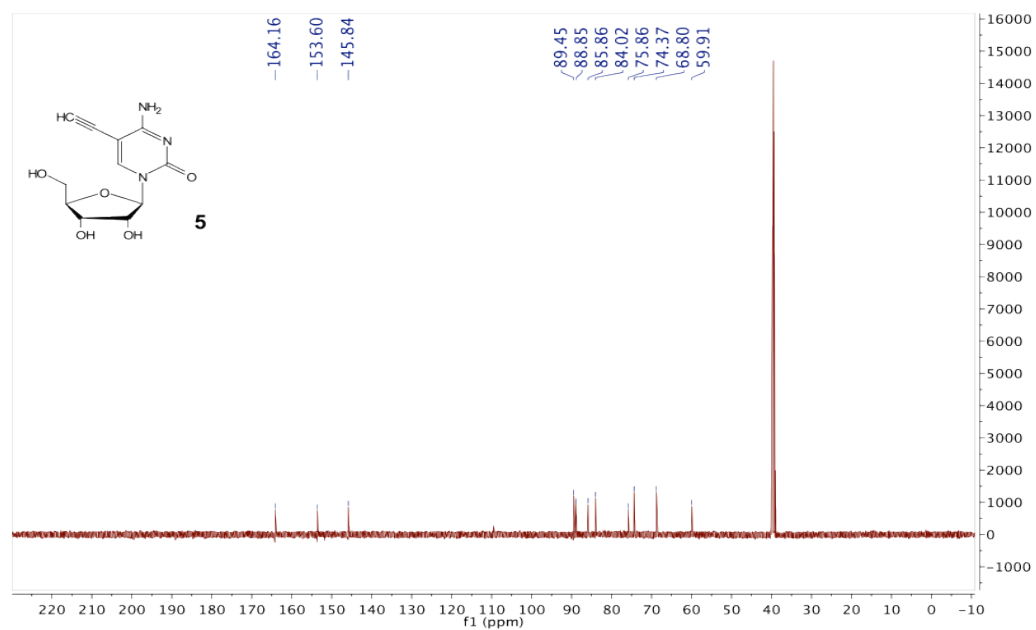¹³C NMR (500MHz, DMSO)

1-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-ethynylpyrimidine-2,4(1H,3H)-dione (14).
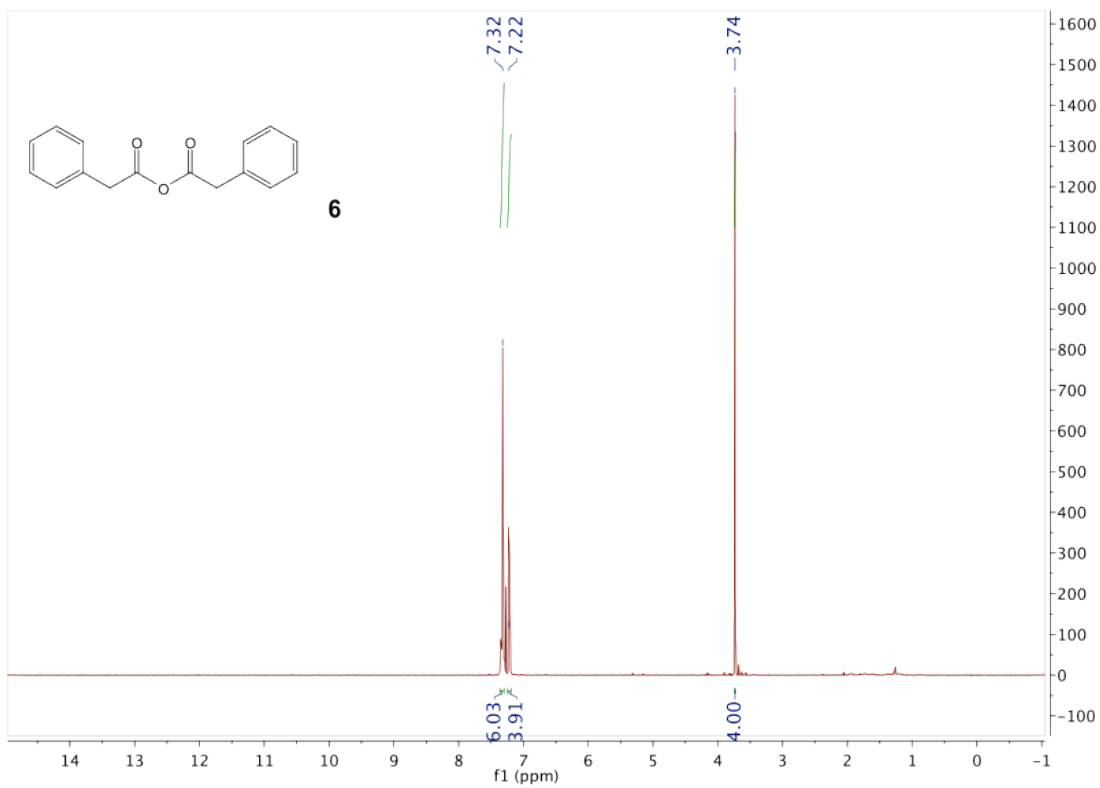
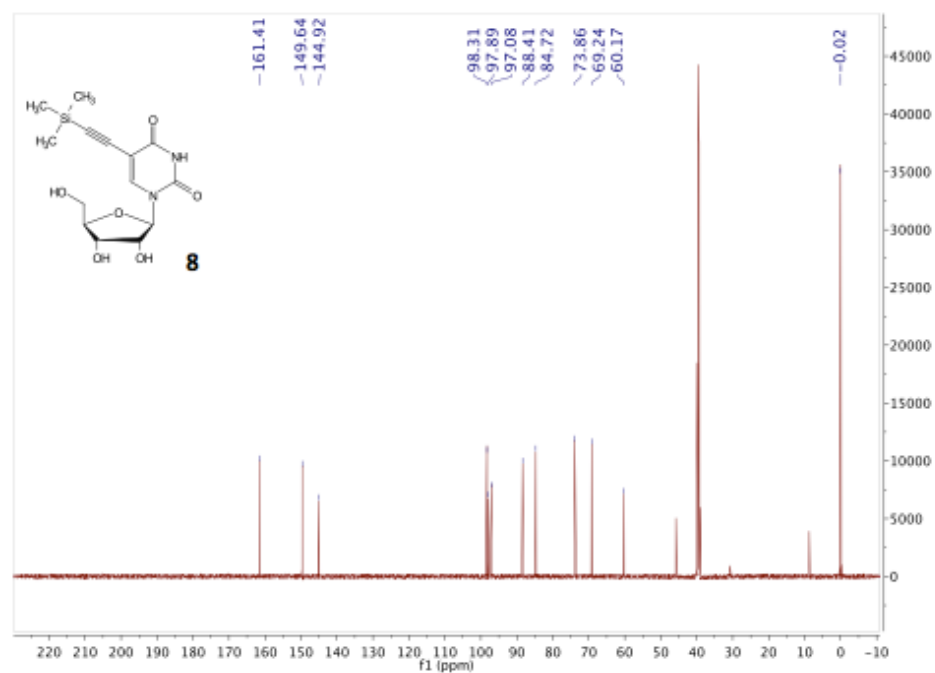¹H NMR (500MHz, DMSO)
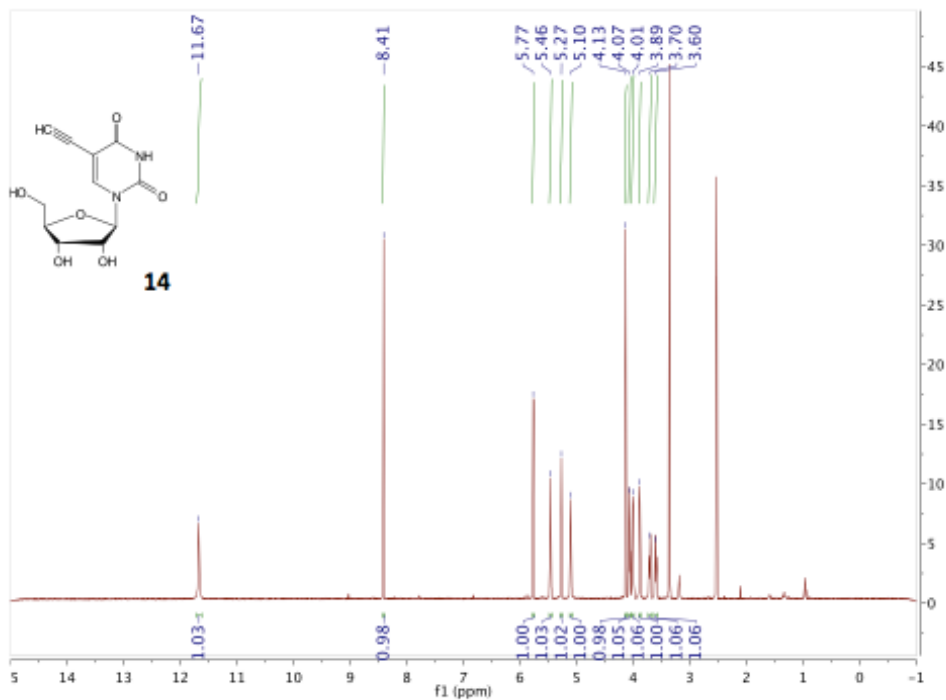
¹³C NMR (500MHz, DMSO)
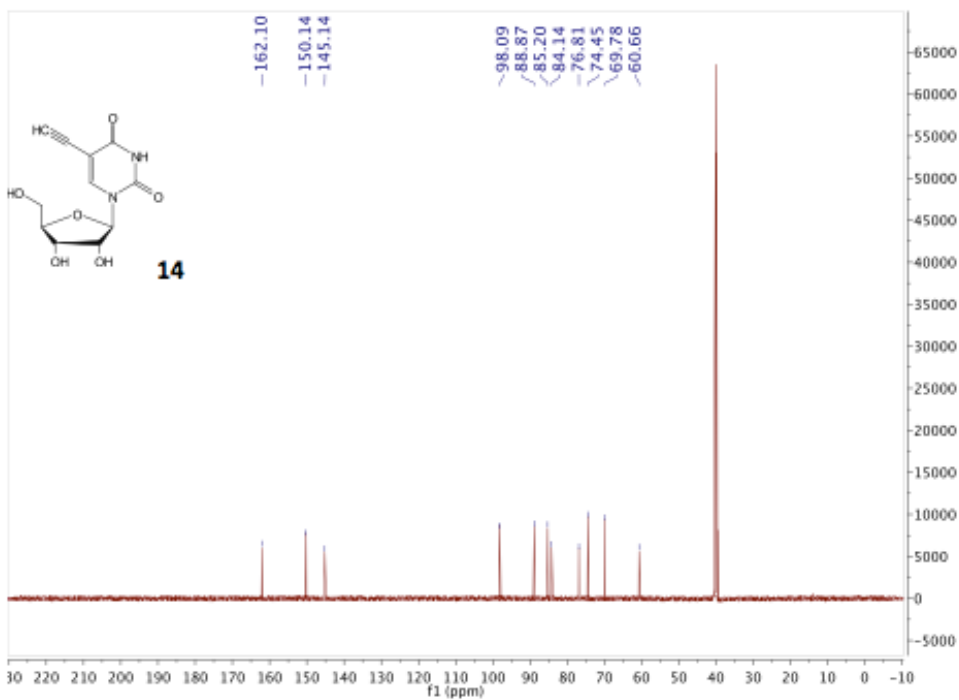
6. Supplementary Information References:

1) L. C. Morrill, L. A. Ledingham, J-P. Couturier, J. Bickel, A. D. Harper, C. Fallan, A. D. Smith,

Org. Biomol. Chem. **2014**, *12,* 624-636.

4.7 References

1. Handley, A.; Schauer, T.; Ladurner, A. G.; Margulies, C. E. *Mol. Cell* **2015**, *58*, 621.

2. Multherjee, N.; Calviello, L.; Hirsekorn, A.; de Pretis, S.; Pelizzola, M.; Ohler, U. *Nat. Struct. Mol. Biol.* **2017**, *24*, 86.

3. Ribeiro, D. M.; Zanzoni, A.; Cipriano, A.; Delli Ponti, R.; Spinelli, L.; Ballarino, M.; Bozzoni, I.; Tartaglia, G. G.; Brun, C. *Nucleic Acids Res.* **2018**, *46*, 917.

4. Schmitt, A. M.; Chang, H. Y. *Cancer Cell* **2016**, *29*, 452.

5. Hrvatin, S.; Deng, F.; O'Donnell, C. W.; Gifford, D. K.; Melton, D. A. *PLoS One* **2014**, *9*.

6. Mack, E.; Neubauer, A.; Brendel, C. *Cytometry A* **2007**, *71a*, 404.

7. Feng, H. J.; Zhang, X. G.; Zhang, C. L. *Nature Communications* **2015**, *6*.

8. Nguyen, K.; Fazio, M.; Kubota, M.; Nainar, S.; Feng, C.; Li, X.; Atwood, S. X.; Bredy, T. W.; Spitale, R. C. *J. Am. Chem. Soc.* **2017**, *139*, 2148.

9. Hida, N.; Aboukilila, M. Y.; Burow, D. A.; Paul, R.; Greenberg, M. M.; Fazio, M.; Beasley, S.; Spitale, R. C.; Cleary, M. D. *Nucleic Acids Res.* **2017**, *45*.

10. Ghosh, A. C.; Shimell, M.; Leof, E. R.; Haley, M. J.; O'Connor, M. B. *Scientific Reports* **2015**, *5*.

11. Forney, L. J.; Wong, D. C. L. *Appl. Environ. Microbiol.* **1989**, *55*, 2556.

12. Handlon, A. L.; Oppenheimer, N. J. *Pharm. Res.* **1988**, *5*, 297.

13. Nainar, S.; Beasley, S.; Fazio, M.; Kubota, M.; Dai, N.; Correa, I. R.; Spitale, R. C. *ChemBioChem* **2016**, *17*, 2149.

14. Sigurgeirsson, B.; Emanuelsson, O.; Lundeberg, J. *PLoS One* **2014**, *9*.

Chapter 5: Conclusions and Future Directions

5.1 Conclusions and Future Directions

In this dissertation, I have worked toward the synthesis of a new RNA structure probe and new RNA labeling probes.

In Chapter 3, I synthesized FAI-N$_3$, which proved to be an improvement for performing icSHAPE structure probing on RNA. The original probe, NAI-N$_3$, while great, had some drawbacks. NAI-N$_3$ was shown to be unstable at the higher temperatures RNA endures during the downstream analysis of icSHAPE. Thus the modified RNA could be deacetylated during the process of fragmentation, the quenching after ligations, or the priming from reverse transcription. The loss of the adduct at any of these steps would result in a false positive RT stop, hindering the robustness of icSHAPE. FAI-N$_3$ was thus developed to improve the adduct stability. The two probes were compared side by side by subjecting NAI-N$_3$ treated RNA and FAI-N$_3$ treated RNA to heat treatments as high as 95°C and the level of deacetylation was measured. FAI-N$_3$ showed higher adduct stability than NAI-N$_3$. The icSHAPE toolkit has now been extended to include two available probes for structure probing of RNA.

Previous to the research discussed in Chapter 2, there had never been an azide containing nucleoside used for RNA labeling. I developed the synthetic strategy to make the N^6-alkyl-azide modified adenosine analogs. As could be seen through western blot analysis, N^6-ethylazidoadenosine and N^6-propylazidoadenosine were both capable of being incorporated into nascent RNA. The modification to the N^6 position did not seem to perturb endogenous kinase binding. Interestingly, 2'-azidoadenosine showed the highest level of incorporation into nascent RNA. We showed that this is in fact due to the ability of 2'-

110

Chapter 5: Conclusions and Future Directions

5.1 Conclusions and Future Directions

In this dissertation, I have worked toward the synthesis of a new RNA structure probe and new RNA labeling probes.

In Chapter 3, I synthesized FAI-N$_3$, which proved to be an improvement for performing icSHAPE structure probing on RNA. The original probe, NAI-N$_3$, while great, had some drawbacks. NAI-N$_3$ was shown to be unstable at the higher temperatures RNA endures during the downstream analysis of icSHAPE. Thus the modified RNA could be deacetylated during the process of fragmentation, the quenching after ligations, or the priming from reverse transcription. The loss of the adduct at any of these steps would result in a false positive RT stop, hindering the robustness of icSHAPE. FAI-N$_3$ was thus developed to improve the adduct stability. The two probes were compared side by side by subjecting NAI-N$_3$ treated RNA and FAI-N$_3$ treated RNA to heat treatments as high as 95°C and the level of deacetylation was measured. FAI-N$_3$ showed higher adduct stability than NAI-N$_3$. The icSHAPE toolkit has now been extended to include two available probes for structure probing of RNA.

Previous to the research discussed in Chapter 2, there had never been an azide containing nucleoside used for RNA labeling. I developed the synthetic strategy to make the N^6-alkyl-azide modified adenosine analogs. As could be seen through western blot analysis, N^6-ethylazidoadenosine and N^6-propylazidoadenosine were both capable of being incorporated into nascent RNA. The modification to the N^6 position did not seem to perturb endogenous kinase binding. Interestingly, 2'-azidoadenosine showed the highest level of incorporation into nascent RNA. We showed that this is in fact due to the ability of 2'-

azidoadenosine to be recognized by poly A polymerase and incorporated into polyA tails. Thus these new azido modified adenosine analogs were capable of labeling two different classes of RNA. The N^6-alkyl-azide analogs were able to label nascent non-polyadenylated RNA, while 2'-azidoadenosine was able to label nascent polyadenylated RNA. Another advantage to the three new bioorthogonal adenosine analogs is that alternative click chemistries can be performed on the RNA. Because the RNA now contains an azide, either CuAAC or SPAAC can be performed. Since it's known that CuAAC has the potential to generate radicals, which can degrade RNA, SPAAC can be performed instead in order to protect the integrity of the RNA. Additionally, while CuAAC is limited to *in vitro* click chemistry, SPAAC can be used to do both *in vitro* and *in vivo* click chemistry. Overall, these new azide modified adenosine analogs have greatly enhanced the RNA toolkit to provide type-specific RNA labeling probes and extending the available chemistries to perform on modified RNA.

As mentioned previously, PGA and N^6-phenylacetyl-2'-azidoadenosine (PAC-2'N_3A) were developed as an enzyme-analog pair for cell-specific RNA labeling in our lab. Because the system is ultimately based on labeling RNA with 2'-azidoadenosine, after PGA activation, the RNA being labeled was biased towards polyadenylated RNA. Thus I sought to design and synthesize alternative phenylacetyl protected bioorthogonal nucleosides, as covered in Chapter 4. I synthesized PAC-5EC and PAC-QM-5EU, both are pyrimidine based nucleoside analogs. With this extension to the PGA system, it is now possible to label both classes of RNA at the same time in a cell-specific manner by using one of these new pyrimidine probes in tandem with PAC-2'N_3A.

Cell specific RNA labeling is a great tool for directly assaying the transcriptome of a cell of interest. However, the process of reaching the end goal of testing an enzyme analog pair is a laborious one. First, an exogenous enzyme that catalyzes a specific chemical reaction has to be found. Second, that modification has to be synthetically appended to a bioorthogonal nucleoside. Third, the exogenous enzyme has to be transfected into the cell of interest for every labeling experiment. Fourth, for studies planning to be moved into mice, a stably expressing cell line has to be developed. Fifth, a stably expressing mouse line has to be bred and raised. Just making the stably expressing cell line and mouse line is easily a year worth of time. An alternative to finding an exogenous enzyme would be to utilize an endogenous enzyme that is differentially expressed across different cell lines. This eliminates the need to make a stably expressing cell and mouse line completely. A prime example of this system is cancer. Cancer cells have been found to have genes that are over-expressed, in comparison to normal cells, that are key to cancer cell progression.[1] Two enzymes that fit this are histone deacetylases (HDAC) and cathepsin L proteases (CatL).[2, 3] HDAC is responsible for deacetylating the lysine tails of histones, a fundamental process for gene silencing.[4] HDAC overexpression has become a trademark of cancer and now HDAC inhibitors are emerging as a new class of anti-cancer drugs.[5] CatL, in normal cells, is found in the lysosome and is key regulator of protein degradation.[6] In cancer cells, CatL is found not only in the lysosome, but also the cytoplasm, nucleus, at the cell surface and even secreted. They have been found to cleave the lysine tails off of histones and they play a major role in cleaving the basal and extracellular matrix to allow cancer cells to leave the tumor and form metastases.[6, 7] Because of this, overexpression of CatL is now a trademark of metastatic cancer.[7, 8] It is important to note that not all cancer cells have both

proteins over-expressed, and some may have one more preferentially expressed over the other.

These observations have led to the hypothesis that appending a terminally acetylated lysine chain to bioorthogonal nucleosides will allow for cancer-cell specific RNA labeling, *in vitro* and *in vivo,* as can be seen in Figure 1*.* The over-expressed HDAC will
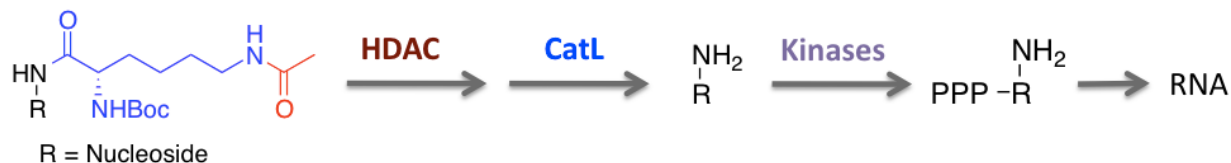

Figure 1. General enzymatic pathway for cancer targeted bioorthogonal nucleoside probes.

deacetylate the lysine tail and the over-expressed CatL will subsequently cleave the free lysine tail. Such new probes could provide insight into RNA profiles during all stages of tumor development, including the formation of metastases, as well as interactions with the surrounding tissue. This technology would not only permit analysis for tumor cell RNA dynamics, but would also allow for analysis of tumor mutations that occur in real time due to drug challenge or programmed changes to the tumor niche. Overall, this would extend the RNA toolkit to now include a cell specific way to target cancer RNA without the need of an exogenous enzyme. My future research direction is to complete this cancer cell-specific RNA labeling with a new class of lysine protected bioorthogonal nucleosides paired with the over-expressed HDAC and CatL and demonstrate that cancer cell-specific RNA labeling is achieved *in vitro* and *in vivo.*

5.2 References

1. Bartonicek, N. et al. Long noncoding RNAs in cancer: mechanisms of action and technological advancements. *Molecular Cancer* **2016**, 15, 1-10.

2. Ueki, N. et al Selective cancer targeting with prodrugs activated by histone deacetylases and a tumour-associated protease. *Nat. Comm.* **2013** 4, 2735.

3. Wang, H. et al. Selective *in vivo* metabolic cell-labeling-mediated cancer targeting system. *Nat. Chem. Bio.* **2017** 13, 415-424.

4. Witt, O. et al. HDAC family: what are the cancer relevant targets? *Cancer Lett.* **2009** 277, 8-21.

5. H. Lee Moffitt Cancer Center & Research Institute. Histone deacetylases and cancer. *Oncogene* **2007** 26, 5420-5432.

6. Duncan, E. et al. Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. *Cell* **2008** 135, 284-294.

7. Gocheva, V. & Joyce, J. Cysteine cathepsins and the cutting edge of cancer invasion. *Cell Cycle* **2007** 6, 60-64.

8. Goulet, B. et al. Increased expression and activity of nuclear cathepsin L in cancer cells suggest a novel mechanism of cell transformation. *Mol. Cancer Res.* **2004** 5, 207-219.