

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Advances in RNA metabolic labeling for profiling nascent transcription

Permalink

<https://escholarship.org/uc/item/7w7818dx>

Author

Nainar, Sarah

Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Advances in RNA metabolic labeling for profiling nascent transcription

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Pharmacological Sciences

by

Sarah J. Nainar

Dissertation Committee:
Associate Professor Robert C. Spitale, Chair
Professor Andrej Lupták
Professor Jennifer A. Prescher

2019

DEDICATION

To

my parents, Marlene and Ali, Ash and Noma, my brothers, my family and
friends across the country, and Luke for their unwavering love
and support these past five years and beyond.

“The essence of the independent mind lies not in what it thinks, but in how it thinks.”

-Christopher Hitchens

TABLE OF CONTENTS

	Page
LIST OF FIGURES	v
LIST OF TABLES	ix
ACKNOWLEDGMENTS	x
CURRICULUM VITAE	xi
ABSTRACT OF THE DISSERTATION	xiv
INTRODUCTION	1
CHAPTER 1: Introduction	1
1.1 Background and Motivation	1
1.2 Summary	5
1.3 References	6
CHAPTER 2: Metabolic Incorporation of Azide Functionality into Cellular RNA	9
2.1 Introduction	9
2.2 Results and Discussion	11
2.3 Conclusion	17
2.4 Acknowledgements	19
2.5 Methods	19
2.6 Supplementary Data	33
2.7 References	47
CHAPTER 3: Temporal Labeling of Nascent RNA Using Photo-click Chemistry in Living Cells	49
3.1 Introduction	49
3.2 Results and Discussion	50
3.3 Conclusion	59
3.4 Acknowledgements	60
3.5 Methods	60
3.6 Supplementary Data	77
3.7 References	85
CHAPTER 4: Expanding the Scope of RNA Metabolic Labeling with Vinyl Nucleosides and Inverse Electron-Demand Diels-Alder Chemistry	88
4.1 Introduction	88
4.2 Results and Discussion	91
4.3 Conclusion	103

4.4 Acknowledgments	104
4.5 Methods	105
4.6 Supplementary Data	147
4.7 References	157
CHAPTER 5: An Optimized Chemical-Genetic Method for Cell-Specific Metabolic labeling of RNA	160
5.1 Introduction	160
5.2 Results and Discussion	164
5.3 Conclusion	193
5.4 Acknowledgements	196
5.5 Methods	196
5.6 Supplementary Data	213
5.7 Supplementary Note 1 – Molecular Dynamics Studies	249
5.8 Supplementary Note 2 – Crystallographic Studies	256
5.9 References	257
CHAPTER 6: Conclusions	262
6.1 Conclusions	262
6.2 References	264

LIST OF FIGURES

	Page
Figure 1-1 Schematic of the chemical reporter strategy to install alkynes	2
Figure 1-2 Schematic of ribonucleoside salvage pathways for metabolic labeling	3
Figure 1-3 Cell-specific labeling strategy of 4SU/UPRT	5
Figure 2-1 Screening of azido-nucleosides	11
Figure 2-2 Fluorescent imaging of azido-nucleosides	14
Figure 2-3 Nucleoside analogs are differentially incorporated into cellular RNA	17
Figure 2-S1 LC-MS analysis of analog 1-3 incorporation	33
Figure 2-S2 Cell viability assay following analog treatment	34
Figure 2-S3 Full Northern blots represented in Figure 2-1, C	35
Figure 2-S4 Concentration titration of analogs 1-3	36
Figure 2-S5 Imaging incorporation of analog 1 into nascent RNA	37
Figure 2-S6 Imaging incorporation of analog 2 into nascent RNA	38
Figure 2-S7 Imaging incorporation of analog 3 into nascent RNA	39
Figure 2-S8 RNase A imaging of cells treated with azido-nucleosides	40
Figure 2-S9 Imaging to confirm incorporation into RNA and not DNA	41
Figure 2-S10 UV spectra to confirm SPAAC reactivity of azido-nucleosides	42
Figure 2-S11 DIBO SPAAC labeling of RNA inside living cells	43
Figure 2-S12 DIBO SPAAC imaging of RNA inside cells	44
Figure 2-S13 Selectivity of RNA polymerases and effects on incorporation	45
Figure 2-S14 Crystal structures of purine and pyrimidine kinases	46
Figure 3-1 Establishing photo-click reactivity with azido-nucleoside	51

Figure 3-2	<i>In vitro</i> photo-click SPAAC with azido-RNA	53
Figure 3-3	Photo-ODIBO, 1 is refractory to thiol addition	54
Figure 3-4	Photo-click labeling of RNA in living cells	56
Figure 3-5	Selective irradiation enables cell-specific RNA tagging in co-culture	59
Figure 3-S1	UV-vis spectra of photo-click reaction components	77
Figure 3-S2	ODIBO, 1a tagging of azido-RNA	78
Figure 3-S3	Thiols do not affect azide integrity	79
Figure 3-S4	Second order rate constant determination of GSH with ODIBO, 1a	80
Figure 3-S5	RT-PCR of DMSO-treated control RNA	81
Figure 3-S6	Photo-click imaging of azido-RNA	82
Figure 3-S7	Imaging of 2'-N ₃ -A via photo-ODIBO-rhB is sensitive to RNase A	83
Figure 3-S8	Selective irradiation and photo-click imaging	84
Figure 3-S9	Fluorescent imaging of cell co-culture.	85
Figure 4-1	Schematic of metabolic labeling of RNA	89
Figure 4-2	Initial analysis of vinyl nucleoside incorporation into RNA	92
Figure 4-3	Screening tetrazines for IEDDA reactions with vinyl nucleosides	95
Figure 4-4	Establishing reaction conditions to stabilize tetrazines in IEDDA	99
Figure 4-5	Effects of vinyl-nucleosides on cell viability and use for RNA imaging	102
Figure 4-S1	LC-MS analysis of vinyl-nucleosides in TRIzol reagent	147
Figure 4-S2	HOMO-LUMO energy calculate and reaction kinetics	149
Figure 4-S3	¹ H NMR analysis of Tz-1 hydrolysis	150
Figure 4-S4	MTT assay of vinyl-nucleosides	152
Figure 4-S5	Trypan blue exclusion assay of 5-EU and 5-VU	152

Figure 4-S6	Cell imaging with 2-VA and 7-dVA	154
Figure 4-S7	Mutually orthogonal bioorthogonal reactions	155
Figure 5-1	Schematic of cell-specific metabolic labeling of RNA	162
Figure 5-2	Mammalian cells salvage exogenous uracil analogs for RNA labeling	165
Figure 5-3	Cell screening reveals a UCK2/2'AzUd pair for cell-specific RNA labeling	170
Figure 5-4	Structural analysis of the UCK2/2'AzUd pair reveals selectivity	181
Figure 5-5	Demonstration of cell-specific metabolic labeling with UCK2/2'AzUd	189
Figure 5-S1	CuAAC imaging of 5EU background incorporation in mammalian cells	215
Figure 5-S2	The analog 4SU is incorporated in wildtype cells	215
Figure 5-S3	Structure and salvage activity of human UMPS	216
Figure 5-S4	Primary sequence alignments of UPRT and UMPS homologs	217
Figure 5-S5	Screening incorporation of 2'azido-nucleosides into nascent RNA	219
Figure 5-S6	Expression of UCK2 mutants	220
Figure 5-S7	WT UCK2 expression and 5mAzUd treatment	222
Figure 5-S8	Y112A UCK2 expression and 5mAzUd treatment	223
Figure 5-S9	F114S UCK2 expression and 5mAzUd treatment	224
Figure 5-S10	F83A; F114S UCK2 expression and 5mAzUd treatment	225
Figure 5-S11	Y112S; F114S UCK2 expression and 5mAzUd treatment	226
Figure 5-S12	WT cells and 5mAzUd treatment	227
Figure 5-S13	WT UCK2 expression and 2'AzUd treatment	228
Figure 5-S14	R166G UCK2 expression and 2'AzUd treatment	229
Figure 5-S15	R166A UCK2 expression and 2'AzUd treatment	230
Figure 5-S16	WT cells and 2'AzUd treatment	231

Figure 5-S17 Generation of a UCK2 stable expression cell line	231
Figure 5-S18 Cell proliferation of 5mAzUd and 2'AzUd-treated cells	232
Figure 5-S19 Effects of other RNA metabolic reporters on cell viability	233
Figure 5-S20 Trypan blue exclusion and Alamar blue analysis of RNA reporters	234
Figure 5-S21 Residues monitored in binding assays	235
Figure 5-S22 Kinetic curves of canonical and 2'azido nucleosides with UCK1/2	236
Figure 5-S23 Comparison of UCK1 and UCK2 sequence and structure	237
Figure 5-S24 Comparison of canonical and flipped binding modes of 2'AzCd	238
Figure 5-S25 Structures of UCK2•2'AzCMP	239
Figure 5-S26 Structures of UCK2•2'AzUd represent the flipped binding mode	240
Figure 5-S27 Hydrogen bond contact frequency of 2'AzUd _{flip}	241
Figure 5-S28 Hydrogen bond contact frequency of 2'AzCd _{flip}	242
Figure 5-S29 Structures of UCK2•2'AzUMP represent the canonical binding mode	243
Figure 5-S30 Hydrogen bond contact frequency of 2'AzUd _{canc}	244
Figure 5-S31 Structures of UCK2•2'AzCMP represent the canonical binding mode	245
Figure 5-S32 Hydrogen bond contact frequency of 2'AzCd _{canc}	246
Figure 5-S33 Structural transitions of UCK2	247
Figure 5-S34 Co-expression of UCK1 and UCK2 reduces 2'AzUd RNA incorporation	247

LIST OF TABLES

	Page
Table 2-S1 Relative abundance of canonical and azido-nucleosides in total RNA	46
Table 4-S1 LC-MS/MS triple quadrupole detection parameters	155
Table 4-S2 Lowest energy reaction barriers for each tetrazine-vinyl nucleoside	156
Table 5-1 Binding and kinetic analysis of UCK1/2 with canonical and 2' analogs	248
Table 5-S1 H-bond and contact distances between ligand and UCK2 residue atoms	256

ACKNOWLEDGMENTS

I would first like to thank my early mentors, Professor Mike Leonard, who was my advisor and mentor during my undergraduate career; and Dr. Mike Deal, the program coordinator for my summer REU, which truly inspired me to pursue graduate level research.

I thank my advisor and mentor, Professor Robert Spitale, for his stalwart support of both me and my research throughout my graduate career. His persistent encouragement, high expectations and constant commitment are what inspired and motivated me to succeed. I was never without the resources or support that I needed, and for that I am very grateful.

I thank Professor Jennifer A. Prescher, a member of my committee, for her time in speaking to me about research and careers, and for serving as a role model throughout my graduate studies. I also thank Professor Andrej Lupták, for his help and support throughout the entirety of my time at UCI. Andrej was always there to talk, advise or listen whenever I needed. I am very grateful to have had this support, especially at the start of graduate school.

I would also like to thank Professor Celia Goulding, whose lab was not only our friendly neighbor but a fantastic collaborator. Professor David Mobley too, was a fantastic collaborator as well as the advisor to students in Pharmaceutical Sciences, where he was always willing to talk and provide advice.

I thank past and present members of the Spitale Lab. As a founding member, I have seen the lab grow and change in interesting ways, and each member has made a significant impact on the research and culture. I thank Dr. Miles Kubota for his work on the projects we shared. I thank Dr. Dalen Chan, for his enthusiastic attitude toward science and the great conversations (mostly about food) we have had over the years. I feel fortunate to have worked alongside many people who I not only consider to be my colleagues, but also my friends.

I thank all of my wonderful friends, old and new, who I have met since moving to California. I would like to specifically mention Nathan Lim and Kelly Vu, who have been incredibly supportive and caring, and were my first friends when I moved from Pennsylvania five years ago.

I thank Luke Hanna, who I met here at UCI. Our time together has meant so much to me, and I am incredibly grateful for his unwavering support during graduate school.

Finally, I thank my entire family and specifically my parents. My mother Marlene and Step-father Ali, and my Father, Ash and step-mother Noma, for their unceasing dedication, love and sacrifice that have enabled me to follow my passions and achieve my goals.

Curriculum Vitae

Sarah J. Nainar

Education

- 9/2014 – 6/2019 **University of California Irvine**, Irvine, CA
Ph.D. Pharmaceutical Sciences, Advisor: Professor Robert Spitale
- 9/2010 – 6/2014 **Washington & Jefferson College**, Washington, PA
B.A. Biochemistry, Magna cum laude
- 9/2012 – 2/2013 **Radboud University**, Nijmegen, Netherlands (*overseas study*)

Research Experiences

- 9/2014 – 6/2019 **University of California Irvine**, Irvine, CA, *Dept. Pharmaceutical Sciences*
Graduate Student Researcher, Advisor: Professor Robert Spitale

Expanded the repertoire of chemical reporters for RNA synthesis, as well as developed optimized methods for cell-specific metabolic labeling of RNA.

- Summer 2013 **Stanford University**, Stanford, CA, *Dept. Materials Science and Engineering*
NSF Nanotechnology REU, NNIN
Undergraduate summer researcher with Prof. Reinhold Dauskardt

Worked toward identifying the biomechanical effects of UV-resistant nanoparticles on the stratum corneum harvested from human cadaveric tissue samples.

Publications

(*review article), (‡ Denotes equal authorship)

1. **Nainar, S.**; Cuthbert, B.; Lim, N.M.; England, W.; Sophal, K.; Quechol, R.; Mobley, D.L.; Goulding, C.; Spitale, R.C. “An Optimized Chemical-Genetic Method for Cell-Specific Metabolic Labeling of RNA.” *Submitted*.
2. Kubota, M.‡; **Nainar, S.**‡; Parker, S.; England, W.; Furche, F.; Spitale, R.C. “Expanding the Scope of RNA Metabolic Labeling with Vinyl Nucleosides and Inverse Electron-Demand Diels-Alder Chemistry.” *ACS Chem. Biol.* **2019**. In press.
3. Li, X; Magnan, C.; Wei, W.; Emami, M.; Marshall, P.; Wearick da Silva, L.; Viola, T.; Grassi-Oliviera, M.; **Nainar, S.**; Vagbo, C.B.; Bjoras, M.; Baldi, P.; Spitale, R.C.; Bredy, T.W. “The accumulation of N6-methyl-2'-deoxyadenosine in DNA drives activity-induced gene expression and is required for the extinction of conditioned fear.” *Nat. Neurosci.* **2018**. In Press.

4. Zajackowski, E.; Zhao, Q.; Zhang, Z.; Li, X.; Wei, W.; Marshall, P.; Leighton, L.; **Nainar, S.**; Feng, C.; Spitale, R.C.; Bredy, T. "Bioorthogonal metabolic labelling of nascent RNA in neurons improves the sensitivity of transcriptome-wide profiling." *ACS Chem. Neurosci.* **2018**, *9*, 1858 – 1865.
5. Suire, C.N.; **Nainar, S.**; Fazio, M.; Kreutzer, A.; Yazdi-Paymozd, T.; Topper, C.L.; Thompson, C.R.; Leissring, M.A. "Peptidic inhibitors of insulin-degrading enzyme with potential for dermatological applications discovered *via* phage display." *PLoS ONE* **2018**, *13*, e0193101.
6. **Nainar, S.**; Kubota, M.; McNitt, C.; Tran, C.; Popik, V.V.; Spitale, R.C. "Temporal Labeling of Nascent RNA Using Photo-click Chemistry in Live Cells." *J. Am. Chem. Soc.* **2017**, *139*, 8090 – 8093.
7. Nguyen, K.; Fazio, M.; Kubota, M.; **Nainar, S.**; Feng, C.; Li, X.; Atwood, S.; Bredy, T.; Spitale, R.C. "Cell-Selective Bioorthogonal Metabolic Labeling of RNA." *J. Am. Chem. Soc.* **2017**, *139*, 2148 – 2151.
8. **Nainar, S.**; Beasley, S.; Fazio, M.; Kubota, M.; Spitale, R.C. "Metabolic incorporation of azide functionality into cellular RNA." *ChemBioChem.* **2016**, *16*, 2149 – 2152.
9. **Nainar, S.**; Marshall, P.R.; Tyler, C.R.; Spitale, R.C.; Bredy, T.W. "Evolving insights into RNA modifications and their functional diversity in the brain." *Nat. Neurosci.* **2016**, *19*, 1292 – 1298.*
10. **Nainar, S.**; Feng, C.; Spitale, R.C. "Applying chemical biology to dissect the role of RNA in epigenetic regulation." *ACS Chem. Biol.* **2016**, *11*, 2091 – 2100. *

Presentations

1. **Nainar, S.**; Cuthbert, B.; Lim, N.M.; England, W.; Goulding, C.; Mobley, D.L.; Spitale, R.C. "Optimized methods for cell-type specific labeling of nascent transcription." Janelia Conference: Chemical Tools for Complex Biological Systems II, Asburn, VA (2019).
2. **Nainar, S.**, Kubota, McNitt, C., Popik, V., Spitale, R.C. "Metabolic Incorporation of Nucleoside Analogs into Cellular RNA for Nascent Transcription Profiling." UC Chemical Symposium, Lake Arrowhead, CA (2018).

Mentoring

1/2017 – present **Graduate Student Mentor**, University of California Irvine

1. Mareena Biju, Undergraduate, Pharmaceutical Sciences: 2018 Spring UROP Fellowship winner
2. Kanika Sophal, Undergraduate, Chemistry: 2018 SURP, 2018 Fall UROP Fellowship winner

Teaching Experience

9/2013 – Spring 2015 **Teaching Assistant**, University of California Irvine
Organic Chemistry Laboratory 51LC – Department. of Chemistry
Biopharmaceutics and Nanomedicine Laboratory – Department of Pharmaceutical Sciences

9/2013 – Spring 2014 **Teaching Assistant**, Washington & Jefferson College
Chemistry 160 Laboratory
Biochemistry 333 Laboratory

Honors and Awards

2019 Graduate Student Researcher of the Year, Dept. of Pharmaceutical Sciences
2016 – 2018 NSF Biophotonics Training Grant Fellow, University of California Irvine
2015 – 2016 Vertex Pharmaceuticals Scholar, University of California Irvine
2014 Phi Beta Kappa National Honor's Society; Washington & Jefferson College
2014 Scholar in Biochemistry Award, Washington & Jefferson College
2010 Hearst Foundation Fellowship, Washington & Jefferson College

Technical Skills

Laboratory: Mammalian cell, stem cell (iPSC) culturing, stable expression, transfection, transduction, DNA/RNA isolation, Southern/Northern blot, siRNA knockdown, Protein extraction, Western blot, NGS library preparation, RNA-sequencing, Molecular cloning, site directed mutagenesis, PCR, RT-qPCR, Enzyme kinetic assays, Biomolecule conjugation/functionalization

Software: MS Office, ChemDraw, ImageJ, Pymol, Adobe Illustrator, Prism Graphpad

Instruments: ESI Mass Spectrometry, HPLC, LC-MS/MS, LSM 700 Confocal

ABSTRACT OF THE DISSERTATION

Advances in RNA metabolic labeling for profiling nascent transcription

By

Sarah Nainar

Doctor of Philosophy in Pharmacological Sciences

University of California Irvine, 2019

Professor Robert C. Spitale, Chair

The transcriptional landscape has revealed itself to be highly complex, wherein RNA serves both coding and non-coding functions that are essential for cell differentiation, development and survival. Because the role of RNA has continued to expand, there is increasing need for biochemical methods to label and track RNA transcription, post-transcriptional regulation, and other RNA-based mechanisms. Herein, my doctoral work to expand the scope of metabolic reporters for RNA synthesis details a systematic expansion of chemistries and modifications for nucleoside analogs. This work has also served as a foundation for the development of novel, cell-specific RNA tagging approaches, both chemical and enzymatic. Throughout these studies, I have characterized the limitations of functional group tolerance and in many cases interrogated their effects on cell uptake and metabolism. Furthermore, I have worked to characterize the limitations of current cell-specific methods, with the goal of developing more stringent and robust approaches. These efforts are poised to provide greater context and engineering opportunities for methods to tag, visualize and study RNA gene expression.

Chapter 1: Introduction

1.1 Background and Motivation

In recent years, RNA has come to the forefront of cellular function. Its biological role has rapidly expanded from the previously simple task of acting as the intermediary between DNA and protein, as theorized by the central dogma.¹ Indeed, RNA has revealed itself to be heavily involved in the regulation and homeostasis of nearly all major biological processes—determination of cell fate to cell-cycle regulation—and is implicated in some of the more curious biological phenomena such as cancer metastasis and X-chromosome inactivation.²⁻⁴ While our understanding of RNA's true significance continues to grow, there is still a deficit of methods that can allow us to study and probe its biology in native environments that have high cellular complexity.

Bioorthogonal chemistry has expanded the array of methods to study biomolecules within their cellular contexts.⁵ Instead of relying on antibody affinity approaches to visualize, tag or isolate a target, small molecule handles may be installed in the biomolecule of interest and used for selective functionalization. Metabolic incorporation of chemically-functionalized biomolecular precursors, so named the chemical reporter strategy, has been widely utilized to study glycosylation, protein function and localization, and DNA dynamics inside cells.⁶⁻⁹ These systems rely on endogenous metabolic machinery in cells, which is used to process and install modified monomers (e.g. amino acid, sugar, or deoxynucleotide) where they will be displayed on the surface of the biomolecule in question. The installation of a terminal alkyne has been extensively used for the purposes of tagging biomolecules with fluorophores or enrichment moieties (**Figure 1-1**).¹⁰⁻¹² This

functional group has found widespread utility because of its ability to undergo the copper catalyzed azide-alkyne cycloaddition (CuAAC) reaction with a variety of azido-compounds.¹³

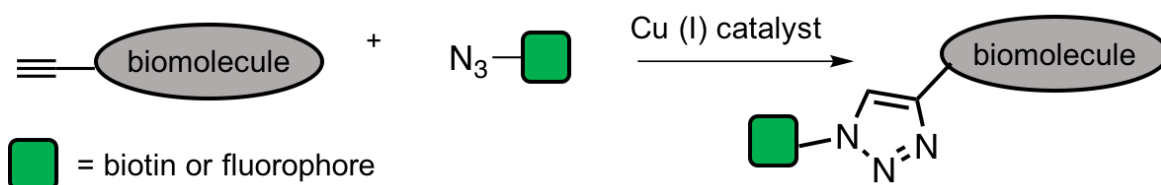


Figure 1-1. Schematic of the chemical reporter strategy to install alkyne functionality. Herein, the biomolecule displaying the alkyne can then be functionalized through a copper-catalyzed azide-alkyne cycloaddition (CuAAC) reaction.

While reporter strategies for proteins and sugars have increasingly become more well-developed and sophisticated,¹⁴⁻¹⁶ the development of new probes and methods for studying RNA has lagged behind. The very first demonstration of using a non-radiometric approach for cellular RNA labeling was reported in 1978, wherein the analog 4-thiouridine was first characterized.¹⁷ Herein, 4-thiouridine can enter into cells and be phosphorylated through the ribonucleoside salvage pathway. It could thereafter be polymerized into nascent RNA by each of the three RNA polymerases (**Figure 1-2**). Following RNA isolation, the thiol group can be functionalized using disulfide exchange chemistry, and biotin can be used for affinity enrichment protocols.¹⁸ This method has been and still is currently used in experiments for profiling nascent transcription, however inefficient chemistry has motivated other researchers to modify nucleosides with alkyne functionality.¹⁹ The first demonstration of cellular RNA labeling with a bioorthogonal functional group was reported

in the work of Jao *et al.*, where uridine was functionalized with an alkyne group at the 5-position of the base.²⁰ Herein, the nucleoside was installed through the same pathway, and was found to be useful for imaging analysis of RNA turnover as well as transcription levels in highly proliferative cells. Other efforts to label RNA have similarly installed alkyne functionality, in several cases using adenosine, which facilitated polyadenylation analyses.^{21,22} Despite these advances, however, the use of terminal alkynes in living systems is limited by the need for copper catalyst, as it is cytotoxic.²³ Furthermore, it can accelerate RNA degradation through generation of reactive oxygen species and resulting strand scission. Thus, it is not ideal when downstream gene expression analysis is desired.

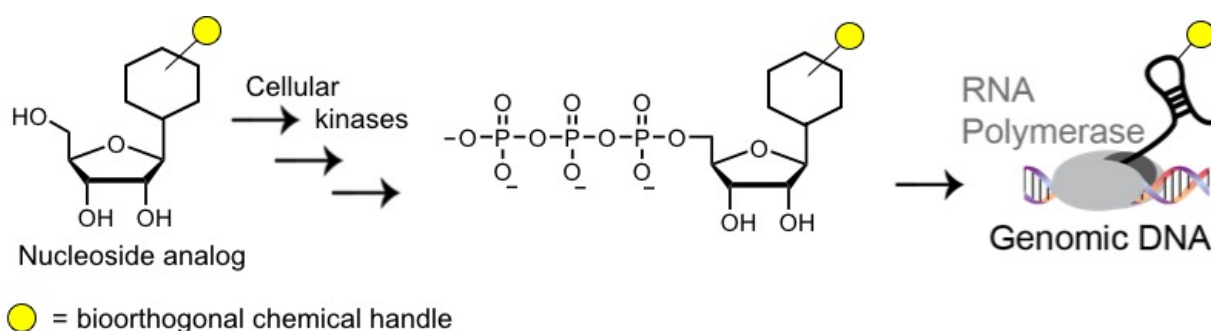


Figure 1-2. Schematic of the ribonucleoside salvage pathway to install and RNA metabolic reporter. Cells are treated with the analog bearing a chemical modification, wherein cellular kinases can process it into the tri-phosphate, after which it can then be polymerized into a nascent RNA transcript.

As an extension of metabolic labeling approaches that tag nascent RNA, researchers have also sought to apply these methods for cell-specific metabolic labeling. In nature, complex tissue environments surround cells, which has a profound impact on their

biological role. As a result, it is often difficult to reconcile gene expression profiles of cells that have been cultured in isolation, from those that have grown and developed in their natural cellular environments and niches.²⁴ This is especially true for the brain, where there are estimated to be hundreds or thousands of neuronal cell types all having unique transcriptional profiles that dictate their biological function.²⁵ A careful and systematic analysis of RNA abundance using *in situ* hybridization has demonstrated that mouse neurons have very specific and highly disparate RNA expression profiles, even between adjacent cells.²⁶ Transcriptome profiling from unique cells in the brain is essential for understanding their individual roles as well as providing greater context to the synergy involved in brain function. This is similarly true for many other heterogeneous organs and tissues, wherein understanding the unique expression profiles of individual cell types will greatly support our ability to study, understand, and even modulate their activity.

While difficult, several methods to assay RNA expression within unique cell types have been developed and used for this purpose. For example, the combination of the base analog 4-thiouracil (4SU) and uracil phosphoribosyl transferase (UPRT), when expressed in a cell-specific manner, can incorporate 4-thiouracil into RNA (**Figure 1-3**). 4-Thiouracil can then be chemically enriched through disulfide purification. Although useful, this method suffers from inefficient chemistry and the resulting inability to image cell specificity, as well as lowered resolution in some cell types.^{27,28} Other parallel methods include RiboTag, involves cell-specific expression of ribosomes tagged with fusion proteins for enrichment of ribosomes, but this technique is hampered by high rate of false-positives. It furthermore cannot enrich for crucial non-coding RNAs, which are not commonly found associated with ribosomes.²⁹ Lastly, polyA binding protein tag fusions can also be

expressed within specific cells, enabling enrichment of polyA RNAs. However, similar to RiboTag, it suffers from high false-positives and it does not enrich for RNAs that do not contain polyA tails—many of which are known to play critical roles in the regulation of cellular pathways.³⁰

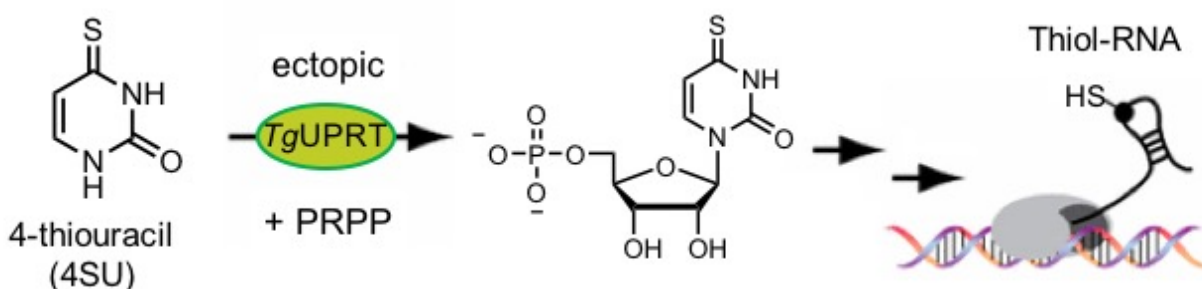


Figure 1-3. Cell-specific labeling strategy of 4SU/UPRT. Ectopic expression of UPRT in a cell type of interest can be used to tag the RNA in these cells with 4SU.

With these challenges in mind, there are many opportunities to further expand the arsenal of metabolic reporters for RNA labeling. Furthermore, there is also great need for advancements in cell-specific RNA profiling, which will serve to improve our understanding of gene expression programs in complex cellular environments.

1.2 Summary

To overcome the aforementioned limitations, I have worked to expand the array of reporters and strategies for profiling nascent transcription. Chapter 2 details novel nucleoside analogs bearing azide functionality that were synthesized and characterized. Through this work, researchers would be able to perform strain promoted azide-alkyne

cycloaddition reactions (SPAAC; sans copper catalyst), which can be used in live cells and *in vivo*.³¹ The findings from this work were then carried over into a novel chemical approach for cell-specific RNA tagging, wherein selective light activation may be used to biotinylate a targeted cell population. Chapter 3 describes the application of this photo-click SPAAC reaction for RNA tagging, which provides spatio-temporal control over reacting azide-modified RNA in living cells.

In Chapter 4, a variety of vinyl modified nucleoside analogs were synthesized and tested for their ability to be incorporated. Concurrently, tetrazines capable of performing inverse electron-demand Diels-Alder chemistry were synthesized and their rate constants were measured with the successful vinyl-nucleosides. Herein, the scope of metabolic reporters was further expanded to encompass a different class of bioorthogonal reaction.

Finally, Chapter 5 details a comprehensive analysis of the challenges and limitations of the UPRT system described above. To reconcile these shortcomings, an optimized cell-specific system was developed and thoroughly characterized. This includes biochemical, computational and structural analysis of the nucleoside-enzyme pair, and ultimately demonstrating its ability to stringently label unique cell types in a co-culture system.

1.3 References

1. Crick, F. *Nature* **1970**, 227, 561-563.
2. Morris, K. V.; Mattick, J. S. *Nat. Rev. Genet.* **2014**, 15, 423-437.
3. Amaral, P. P.; Dinger, M. E.; Mercer, T. R.; Mattick, J. S. *Science* **2008**, 319, 1787-1789.
4. Li, M.; Belmonte, J. C. I. *Nat. Struct. Mol. Biol.* **2015**, 22, 2-4.

5. Lang, K.; Chin, J. W. *ACS Chem. Biol.* **2014**, *9*, 16-20.
6. Rabuka, D.; Hubbard, S. C.; Laughlin, S. T.; Argade, S. P.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2006**, *128*, 12078-12079.
7. Yao, J. Z.; Uttamapinant, C.; Poloukhine, A.; Baskin, J. M.; Codelli, J. A.; Sletten, E. M.; Bertozzi, C. R.; Popik, V. V.; Ting, A. Y. *J. Am. Chem. Soc.* **2012**, *134*, 3720-3728.
8. Salic, A.; Mitchison, T. J. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 2415-2420.
9. Patterson, D. M.; Nazarova, L. A.; Prescher, J. A. *ACS Chem. Biol.* **2014**, *9*, 592-605.
10. Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9482-9487.
11. Chang, P. V.; Chen, X.; Smyrniotis, C.; Xenakis, A.; Hu, T.; Bertozzi, C. R.; Wu, P. *Angew. Chem., Int. Ed.* **2009**, *48*, 4030-4033.
12. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002**, *41*, 2596-2599.
13. Presolski, S. I.; Hong, V. P.; Finn, M. G. *Curr. Protoc. Chem. Biol.* **2011**, *3*, 153-162.
14. Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. *Nature Chem.* **2012**, *4*, 298.
15. Davis, L.; Chin, J. W. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 168-182.
16. Wang, H.; Wang, R.; Cai, K.; He, H.; Liu, Y.; Yen, J.; Wang, Z.; Xu, M.; Sun, Y.; Zhou, X.; Yin, Q.; Tang, L.; Dobrucki, I. T.; Dobrucki, L. W.; Chaney, E. J.; Boppart, S. A.; Fan, T. M.; Lezmi, S.; Chen, X.; Yin, L.; Cheng, J. *Nat. Chem. Biol.* **2017**, *13*, 415-424.
17. Melvin, W. T.; Milne, H. B.; Slater, A. A.; Allen, H. J.; Keir, H. M. *Eur. J. Biochem.* **1978**, *92*, 373-379.
18. Tani, H.; Akimitsu, N. *RNA Biol.* **2012**, *9*, 1233-1238.

19. Burger, K.; Mühl, B.; Kellner, M.; Rohrmoser, M.; Gruber-Eber, A.; Windhager, L.; Friedel, C. C.; Dölken, L.; Eick, D. *RNA Biol.* **2013**, *10*, 1623-1630.
20. Jao, C. Y.; Salic, A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 15779-15784.
21. Grammel, M.; Hang, H.; Conrad, N. K. *ChemBioChem* **2012**, *13*, 1112-1115.
22. Curanovic, D.; Cohen, M.; Singh, I.; Slagle, C. E.; Leslie, C. S.; Jaffrey, S. R. *Nat. Chem. Biol.* **2013**, *9*, 671-673.
23. Kennedy, D. C.; McKay, C. S.; Legault, M. C.; Danielson, D. C.; Blake, J. A.; Pegoraro, A. F.; Stolow, A.; Mester, Z.; Pezacki, J. P. *J. Am. Chem. Soc.* **2011**, *133*, 17993-18001.
24. Wu, Y. E.; Pan, L.; Zuo, Y.; Li, X.; Hong, W. *Neuron* **2017**, *96*, 313-329.e316.
25. Darmanis, S.; Sloan, S. A.; Zhang, Y.; Enge, M.; Caneda, C.; Shuer, L. M.; Hayden Gephart, M. G.; Barres, B. A.; Quake, S. R. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 7285-7290.
26. Mercer, T. R.; Dinger, M. E.; Sunkin, S. M.; Mehler, M. F.; Mattick, J. S. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 716-721.
27. Gay, L.; Karfilis, K. V.; Miller, M. R.; Doe, C. Q.; Stankunas, K. *Nat. Protoc.* **2014**, *9*, 410-420.
28. 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*, e138-e138.
29. Sanz, E.; Yang, L.; Su, T.; Morris, D. R.; McKnight, G. S.; Amieux, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13939-13944.
30. Roy, P. J.; Stuart, J. M.; Lund, J.; Kim, S. K. *Nature* **2002**, *418*, 975-979.
31. Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16793-16797.

Chapter 2: Metabolic Incorporation of Azide Functionality into Cellular RNA

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 expand.² As such, biochemical methods to track RNA transcription, post-transcriptional 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 results in biases and thus limits enrichment of 4SU-labeled RNA.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 imaging facilitated by copper-catalyzed azide-alkyne cycloaddition (CuAAC).⁵ *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.⁶⁻⁸ 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 produces copper-induced radicals that degrade RNA.^{9,10} Such degradation can lead to deleterious effects on downstream analyses such as RNA sequencing^{11,12}. As such, there is a critical need to expand the bioorthogonal toolkit for cellular RNA by endowing 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 CuAAC¹³ and copper-free strain-promoted azide-alkyne cycloadditions (SPAAC),^{14,15} as well as Staudinger ligations.¹⁶ Metabolic labeling with azide-functionalized sugars has been a gold standard for studying glycosylated proteins on the cell surface.¹⁶ Azide-modified unnatural amino acids have been used to track nascent protein synthesis.¹⁷ These examples underscore just a few of the powerful techniques made possible by functionalizing endogenous biomolecules with azide handles. However, installing azide functionality into nascent cellular RNA through metabolic incorporation of azido-nucleosides has yet to be described.

Herein we report, to the best of our knowledge, the first 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-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 designed three adenosine analogs: *N*⁶-ethylazidoadenosine (**1**), *N*⁶-propylazidoadenosine (**2**), and 2'-azidoadenosine (**3**); and one uridine analog, 5-azidomethyl uridine (**4**) (**Methods; Synthesis**). To directly test for their ability to be incorporated through endogenous cell machinery, we first incubated cells with each azido-nucleoside analog for 12 h and then isolated the total RNA (**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 (**Figure 2-1, B**).

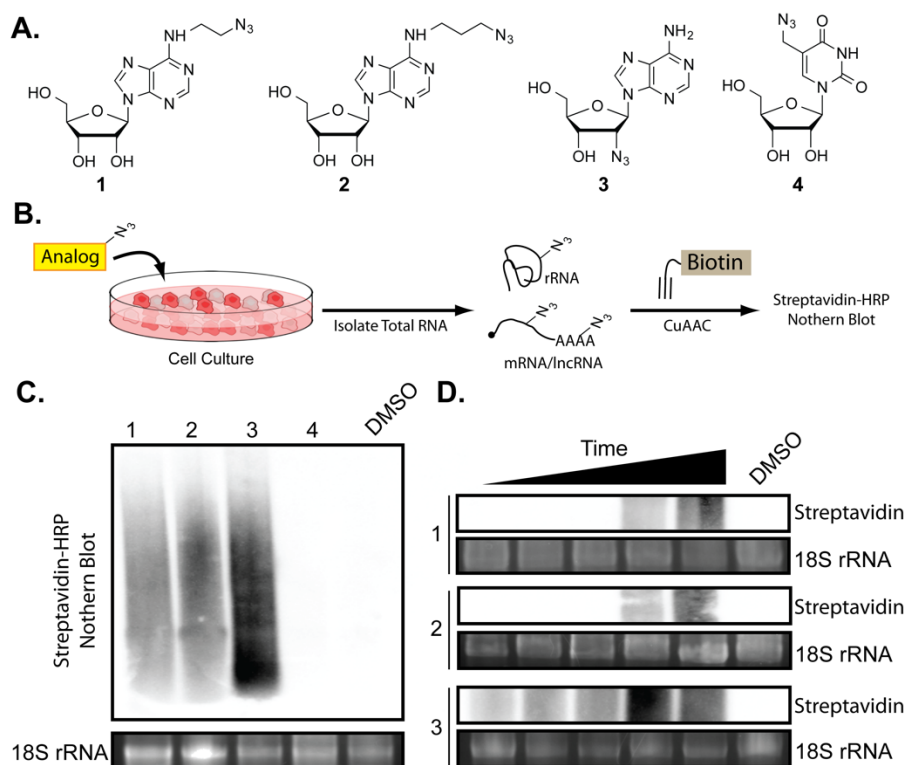


Figure 2-1. Screening of azido nucleosides. **A.** Structures of azido- nucleosides used in this study. **B.** Schematic of HeLa cell treatment with analog, extraction, CuAAC biotinylation

and Northern blot detection through streptavidin-HRP. **C.** Northern blot of RNA from HeLa cells treated with 1mM analog **1-4** for 12 h. **D.** Time course analysis of analog **1-3** incorporation after 1mM incubation.

These results showed that azido-nucleoside analogs **1-3** were robustly incorporated into cellular RNA, while metabolic labelling with **4** was not detected (**Figure 2-1, C**). Incorporation was further confirmed through total RNA digestion and LC-MS analysis (**Figure 2-S1; Table 2-S1**). We examined the cytotoxicity of analogs **1-3** via an MTT assay and observed no significant difference in cell viability of control cells compared to those treated with **1-3** at 1 mM final concentration (**Figure 2-S2**). The data in **Figure 2-1** suggests that the pathway for purine triphosphate biosynthesis may be quite flexible, as our adenosine analogs **1-3** have modifications on the nucleobase and the ribose sugar. In contrast, although 5-ethynyluridine is routinely used for tracking RNA synthesis,⁵ uridine analog **4**, which bears a larger azido-methyl group at the 5 position, showed no detectable incorporation into RNA.

The differences in relative signal prompted us to explore the kinetics of incorporation. We reasoned the lack of incorporation may be due to slower kinetics or simply limited introduction into cellular RNA. We performed a time titration analysis, 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, D; Figure 2-S3**). We also treated cells with a concentration titration (**Figure 2-S4**), and the data demonstrated that 2'-azidoadenosine is the most robustly incorporated at a concentration as low as 10 μ M. 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 a reporter 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, permeabilized and reacted with alkyne-rhodamine via CuAAC (**Figure 2-2; Figure 2-S5 – S7**). 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 through fluorescence for analog **3** at concentrations as low as 10 μ M (**Figure 2-S7**). To confirm that the fluorescent signal we observed was from RNA, cells labeled with analogs **1-3** were lightly fixed and permeabilized before treatment with RNase A (**Figure 2-S8**). After CuAAC with alkyne-rhodamine, cells were imaged and we saw significantly reduced fluorescent signal in the nucleolus, indicating that the incorporation was indeed present in nascent RNA.

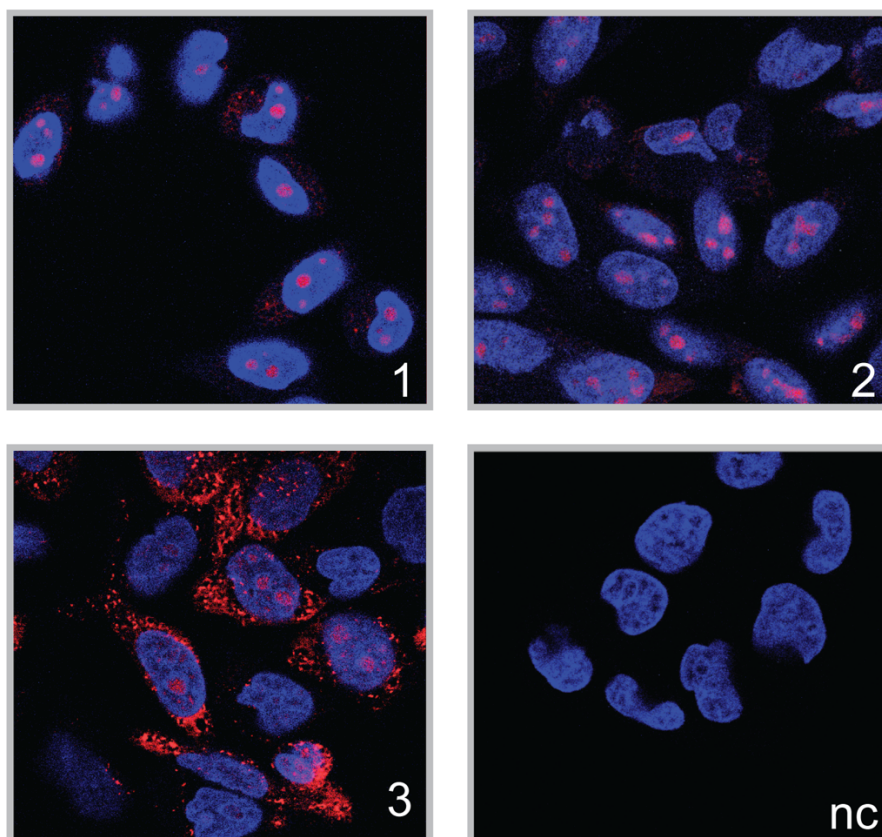


Figure 2-2. Fluorescent imaging of azido-nucleosides. Analogs **1-3** were incubated in HeLa cells for 12 h at 1 mM. Cells were then fixed, permeabilized and CuAAC clicked with alkyne rhodamine. Numbering corresponds to the same as **Figure 2-1**. nc = negative control, no analog

Additionally, we wanted to establish that these analogs were primarily incorporated through transcription and not DNA synthesis. To test this, hydroxyurea, a known inhibitor of ribonucleotide reductase inhibitor, was concurrently incubated in cells treated with analogs **1-3**, and fluorescent imaging was similarly assessed (**Figure 2-S9**). We saw no reductions to the nucleolar staining, and no staining overlapping with the DAPI signal of

the nucleus. Overall, this indicates that these analogs are primarily used for ribonucleoside salvage and RNA metabolism only.

Azides present the unique opportunity to be probed by less-damaging copper-free chemistry, which can support their use for live-cell conjugation. To first confirm that azide-modified adenosine can indeed be functionalized through SPAAC chemistry, we performed UV-vis experiments monitoring the formation of the triazole product. Through this analysis we observed the characteristic bleaching of the 300 nm dibenzocyclooctyne¹⁵ (DIBO) peak once reaction with azide had occurred (**Figure 2-S10**).

To test SPAAC tagging of azido-nucleosides inside living cells, we incubated cells with analogs **1-3** for 12 h to allow for incorporation. After washing the cells and replacing with clean media, DIBO biotin was added to the media and allowed to incubate for 2 h. RNA extracted from the cells was directly subjected to northern blotting, wherein we concluded that bioorthogonal reactions with azide-modified RNA can be performed inside living cells (**Figure 2-S11**). Such an observation demonstrates that nascent RNAs can be tracked in live cells while eliminating the cytotoxic effects of copper-induced radical formation.

To similarly test if the azido-nucleosides can be imaged with SPAAC reagents, we performed the same fluorescent imaging experiment using DIBO-Alexa 488. We saw efficient SPAAC-mediated fluorescent labeling of nascent RNA, indicating that copper free chemistry can likewise be used for imaging experiments (**Figure 2-S12**). While this experiment could not be done in live cells due to cell impermeability of the Alexa dye, other cell-permeable SPAAC imaging reagents should afford the ability to image in living cells. Furthermore, this can have significant applications for live cell RNA tracking in whole

tissues and organisms. Overall, our data supports azide-modified adenosine analogs are useful for labeling RNA by CuAAC as well as SPAAC inside living cells.

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 bearing a wide variety of modifications are well tolerated by a variety of polymerases.¹⁸⁻²⁰ PolyA polymerase selectivity has been investigated *in vitro*,^{8,10} but has yet to be fully explored inside living cells. Adenosine analogs with alkyne groups at the N⁶-, C2- and C7- position of adenosine have been demonstrated to be acceptable substrates.⁶⁻⁸ However, the larger azide modifications at 2'- and N⁶-positions have yet to be explored. To investigate this, we examined the crystal structure of the mammalian poly(A) polymerase bound to adenosine (PDB:1F5A) (**Figure 2-3, A**).²¹ Amino acid residues tightly surround the N⁶-position of adenosine. Conversely, the 2'-position is relatively open and more exposed to the surrounding 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 this selectivity, we challenged the incorporation of all three analogs with actinomycin D (Act D, 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.^{6,7} Analogs **1** and **2**, which contain N⁶-alkyl chains, showed strong inhibition by Act D treatment, while showing limited decrease in signal when treated with Cdy (**Figure 2-3, B; Figure 2-S13, A and B**). 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; Figure 2-S13, C**). 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.

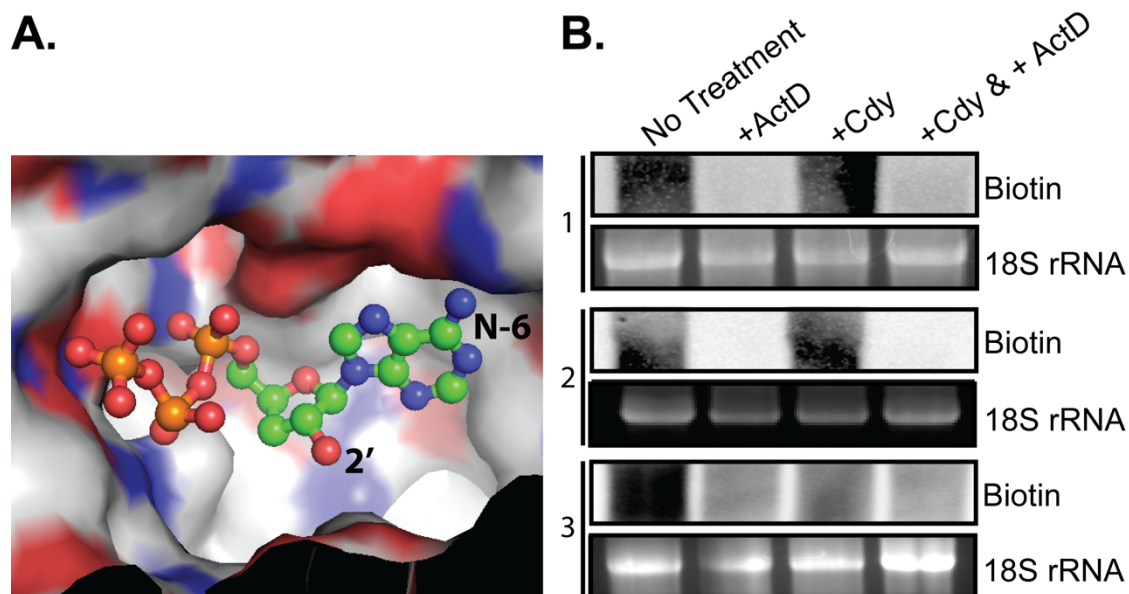


Figure 2-3. Nucleoside analogs are differentially incorporated into cellular RNA. **A.** PDB model of mammalian poly(A) polymerase, PDB: 15FA. **B.** Northern blot analysis of polymerase inhibition and analog incorporation. Act D = Actinomycin D, Cdy = Cordycepin.

2.3 Conclusion

Herein, we have provided the first report of azide-modified nucleoside analogs being metabolically incorporated into cellular RNA. We have also demonstrated 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,³ characterizing the RNA-binding proteome,²² and even tracking viral infection.²⁰ 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-azidomethyluridine, **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. A comparison analysis of the adenosine kinase and uridine/cytidine kinase crystal structures revealed that adenosine with modifications at the 2' and N⁶-positions should indeed be an amenable substrate (**Figure 2-S14**). 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, and this presents an exciting opportunity for enzyme engineering to introduce additional modifications into cellular RNA. These investigations are further detailed in Chapter 5.

2.4 Acknowledgements

Nucleoside analogs **1**, **2** and **4** were synthesized by Samantha Beasley, Michael Fazio and Miles Kubota, respectively. Ivan Correa (New England Biolabs) carried out total RNA digestion and LC-MS quantification of analog incorporation.

2.5 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% CO₂. 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 CuSO₄ 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 (Thermo Fisher 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 (Thermo Fisher 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 μ M actinomycin (VWR International) and/or 10 μ M 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 h. 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 μ L of solution containing 15 μ M alkyne-rhodamine, 760 μ M THPTA ligand, 1.77 mM NaAsc, and 500 μ M CuSO₄, incubated for 1 h 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 1 mM alone (control), or in tandem with 10 mM hydroxyurea and were incubated for 12 h. 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 μ M DIBO-biotin (Life Technologies) for 2 h 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 1 mM and incubated for 12 h. 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 μ L of 50 μ M DIBO-Alexa 488 (Life Technologies) in 1% FBS, DPBS for 2 h 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 d. 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 h 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.

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. 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; N⁶-EtN₃A (1) [M+H]⁺ 337.1 and [M-H]⁻ 335.0; N⁶-PrN₃A (2) [M+H]⁺ 351.1 and [M-H]⁻ 350.0; 2'-N₃A (3) [M+H]⁺ 293.1 and [M-H]⁻ 291.0; and 5-MeN₃U (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; N⁶-EtN₃A (1) 15,588 at 267 nm; N⁶-PrN₃A (2) 15,666 at 267 nm; 2'-N₃A (3) 15,460 at 259 nm; and 5-MeN₃U (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 h. Cells were then fixed, stained, mounted and imaged as previously described.

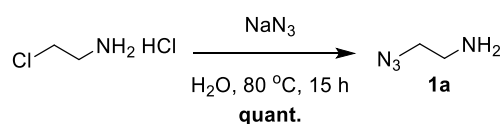
UV spectral analysis of SPAAC reaction of N₃-uridine and DIBO biotin.

Solutions of N₃-uridine, 2'-N₃A, 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 (Thermo Fisher).

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, KMnO₄, FeCl₃, 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 MgSO₄. All reaction vessels were flame dried prior to use. NMR spectra were acquired with Bruker Advanced spectrometers. All spectra were acquired at 298 K. ¹H-NMR spectra were acquired at 400 MHz and 500 MHz. ¹³C-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.

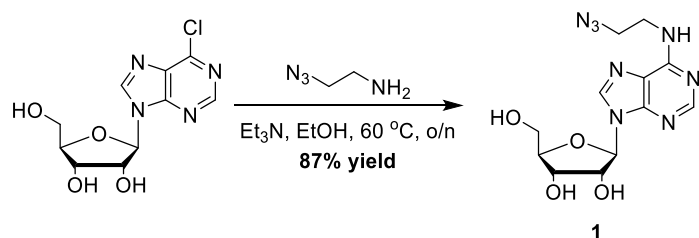
Synthesis



2-azidoethan-1-amine (1a)

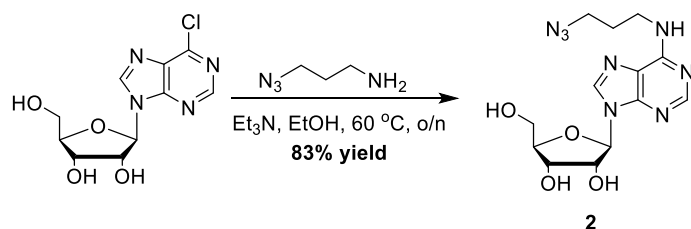
To a stirred solution of 2-chloroethylamine hydrochloride (4.83 mmol, 1 eq) in H₂O (5 mL) was added sodium azide (14.5 mmol, 3 eq) and the suspension was then heated to 80 °C. After 15 h, KOH pellets were added to basify the solution, followed by extraction with

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 $\text{C}_2\text{H}_6\text{N}_4$ $[\text{M}+\text{H}]^+$ 87.07, found 87.0668 $[\text{M}+\text{H}]^+$; ^1H 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.



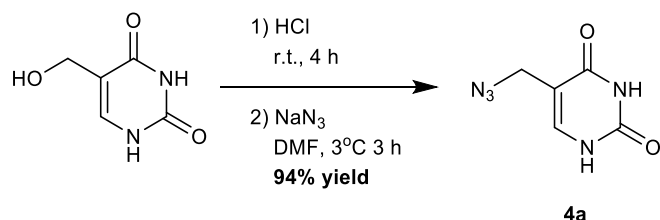
(2R,3R,4S,5R)-2-((6-((2-azidoethyl)amino)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol (1)

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 $\text{C}_{12}\text{H}_{16}\text{N}_8\text{O}_4\text{Na}$ 359.12 $[\text{M}+\text{Na}]^+$, found 359.1201 $[\text{M}+\text{Na}]^+$; ^1H NMR (500 MHz, d_6 -DMSO): δ (ppm) 8.39 (s, 1H), 8.23 (s, 1H), 8.09 (s, 1H), 5.9 (d, $J=6.2$ MHz, 1H), 5.46 (d, $J=6.2$ MHz, 1H), 5.39 (s, 1H), 5.20 (d, $J=4.5$ MHz, 1H), 4.62 (dd, $J=6, 5$ MHz, 1H), 4.15 (dd, $J=4, 3$ MHz), 3.97 (dd, $J=3.5, 3$ MHz, 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.



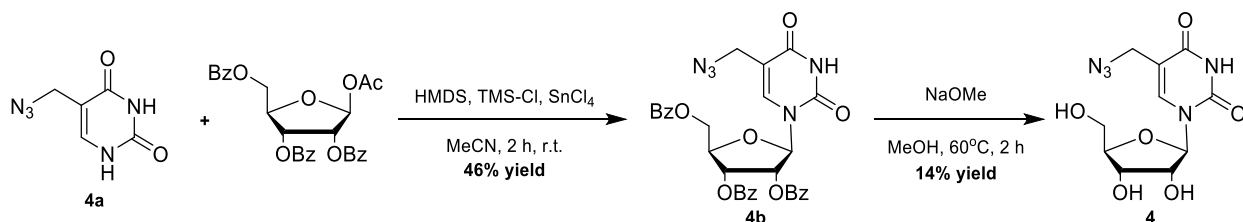
(2*R*,3*R*,4*S*,5*R*)-2-(6-((3-azidopropyl)amino)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol (2)

3-azidopropan-1-amine²³ (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 C₁₃H₁₈N₈O₄Na 373.13 [M+Na]⁺, found 373.1363 [M+Na]⁺; ¹HNMR (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); ¹³C 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**)

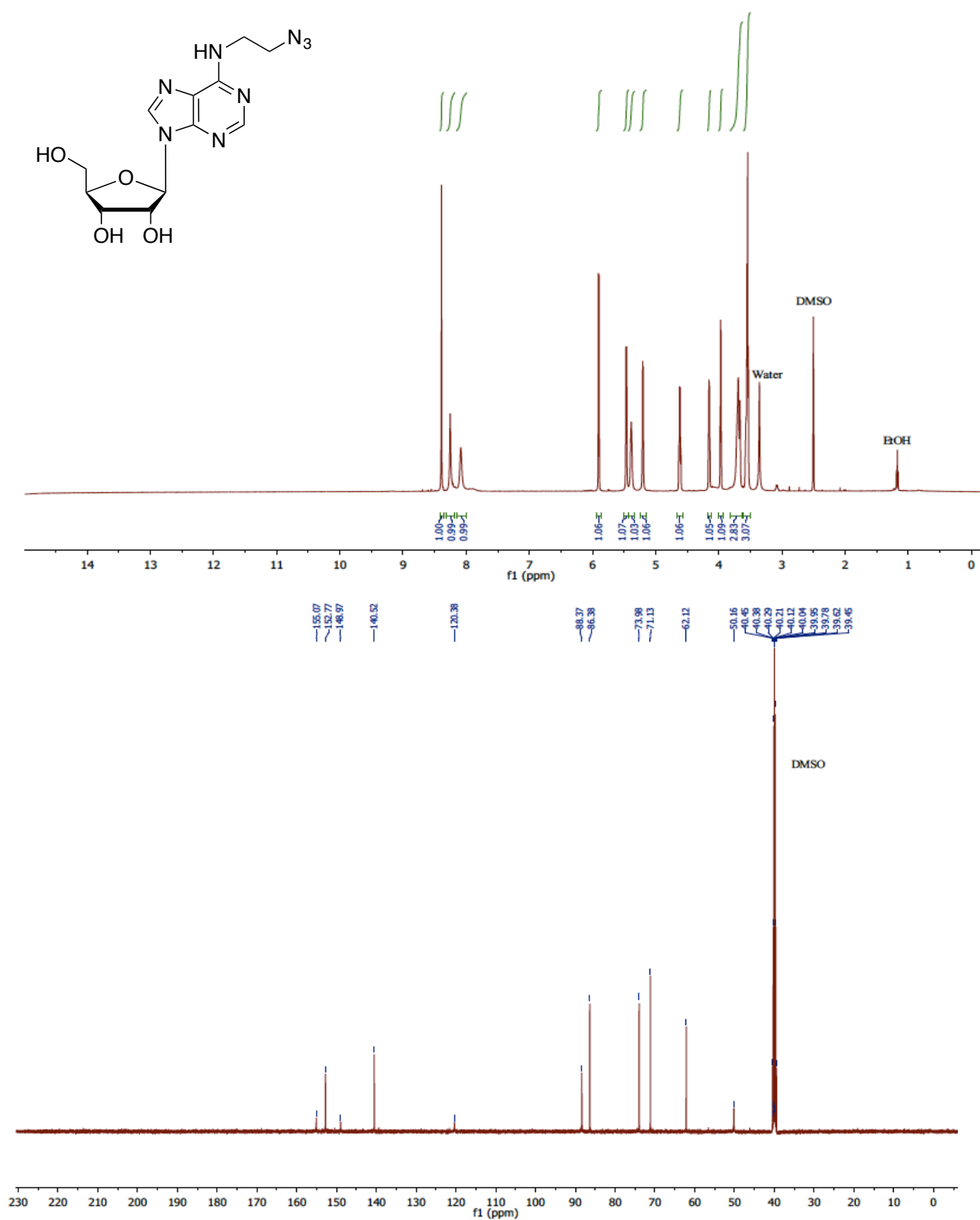
A solution of 5-hydroxymethyluracil (2.00 g, 14.1 mmol) and concentrated HCl (25 mL) was stirred at r.t. for 4 h. The suspension was then put on ice for 30 min and the off white precipitate filtered over a glass fritted filter. The precipitate was allowed to dry and then dissolved in DMF (50 mL). Sodium azide (1.01 g, 15.5 mmol) was added to the solution. The suspension was stirred in the dark at 3 °C 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 C₅H₅N₅O₂ [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.

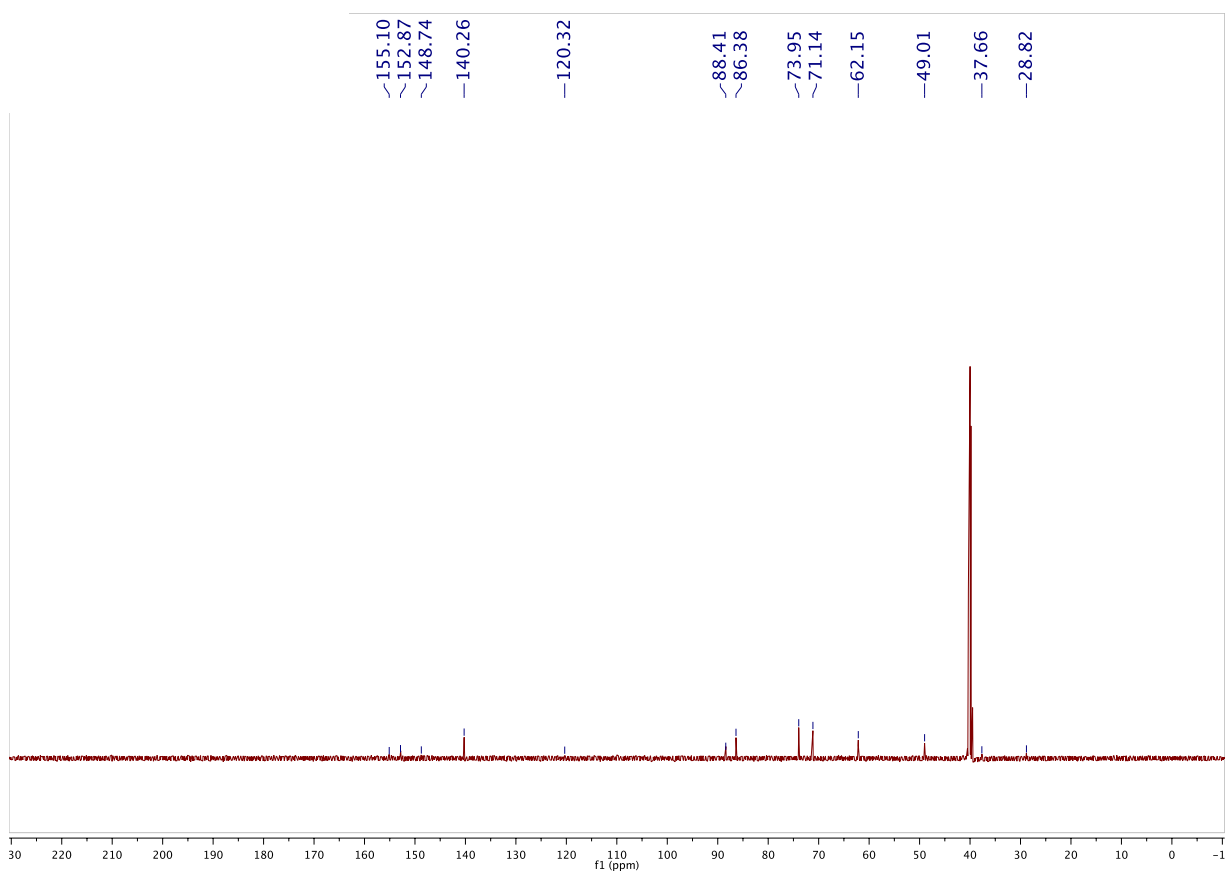
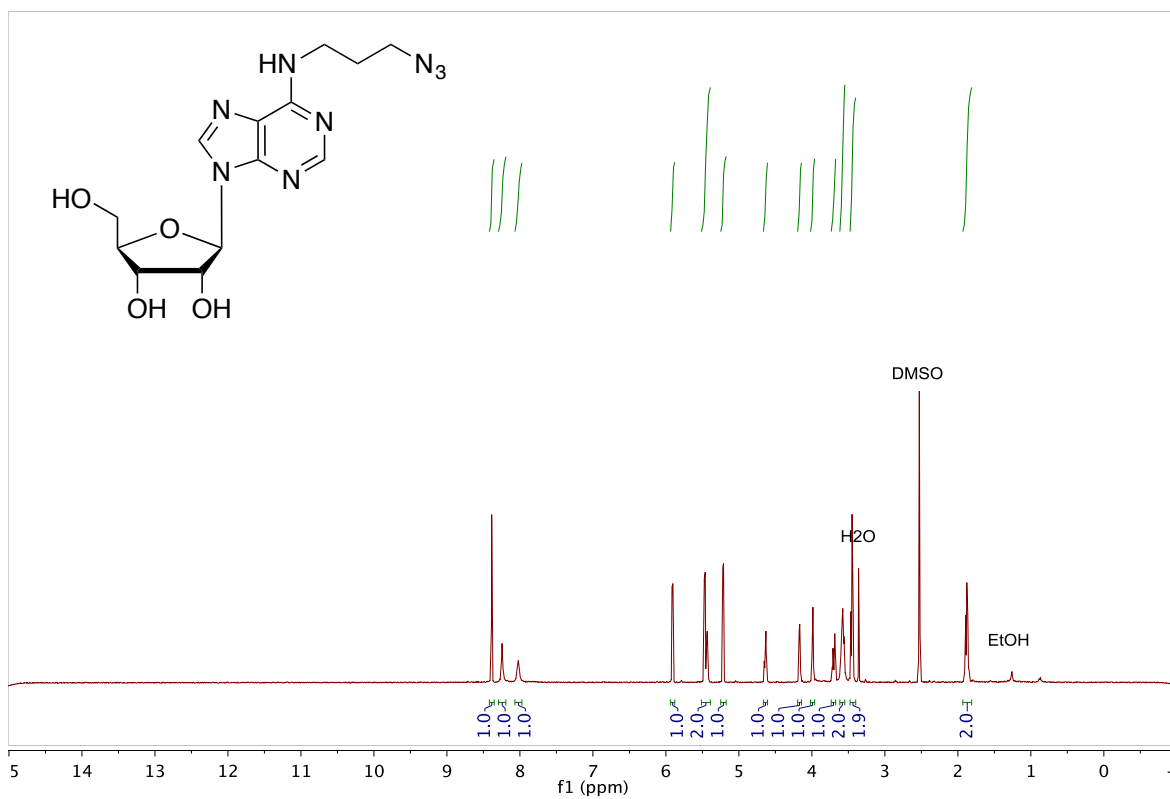


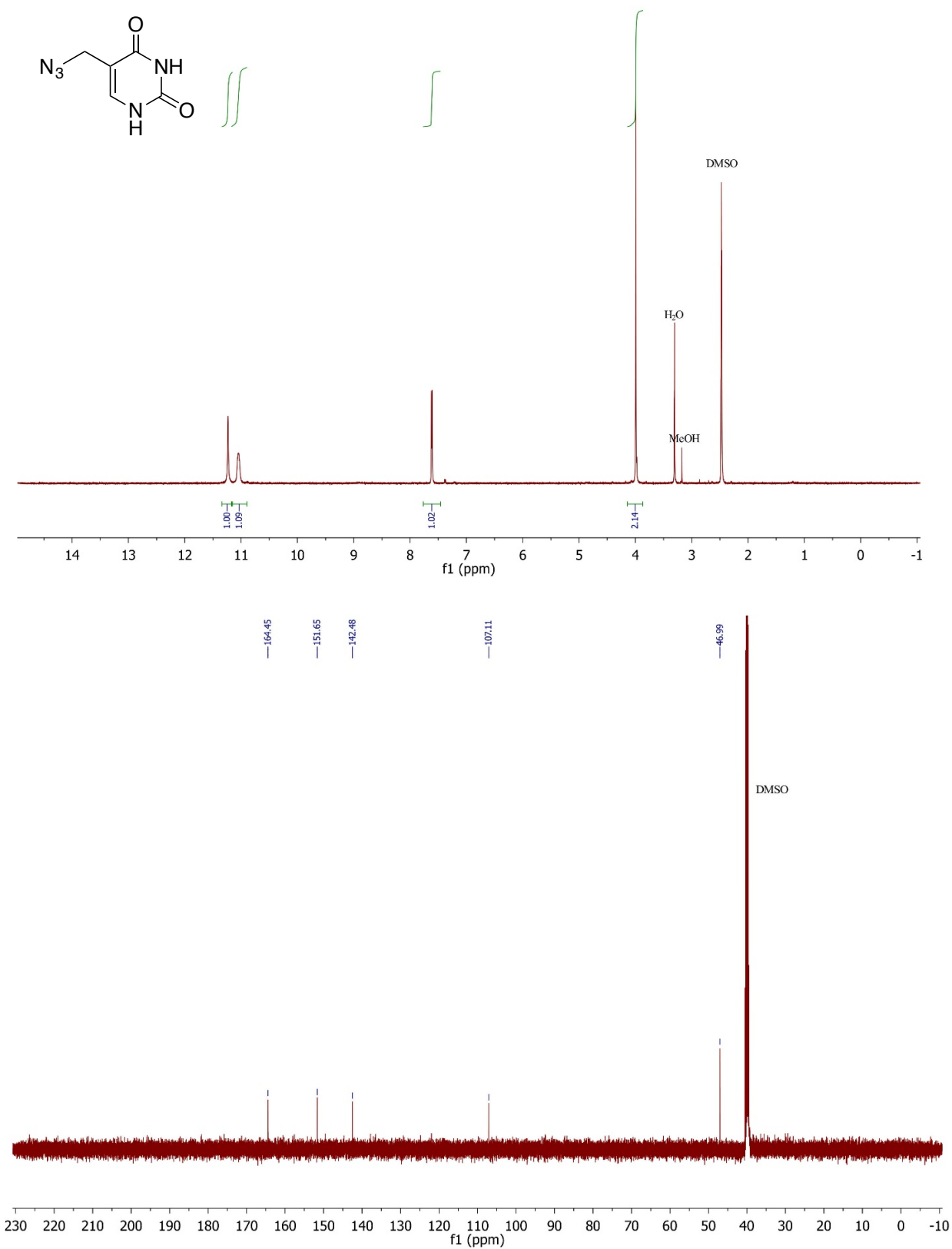
5-(azidomethyl)-1-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)pyrimidine-2,4(1H,3H)-dione (**4**)

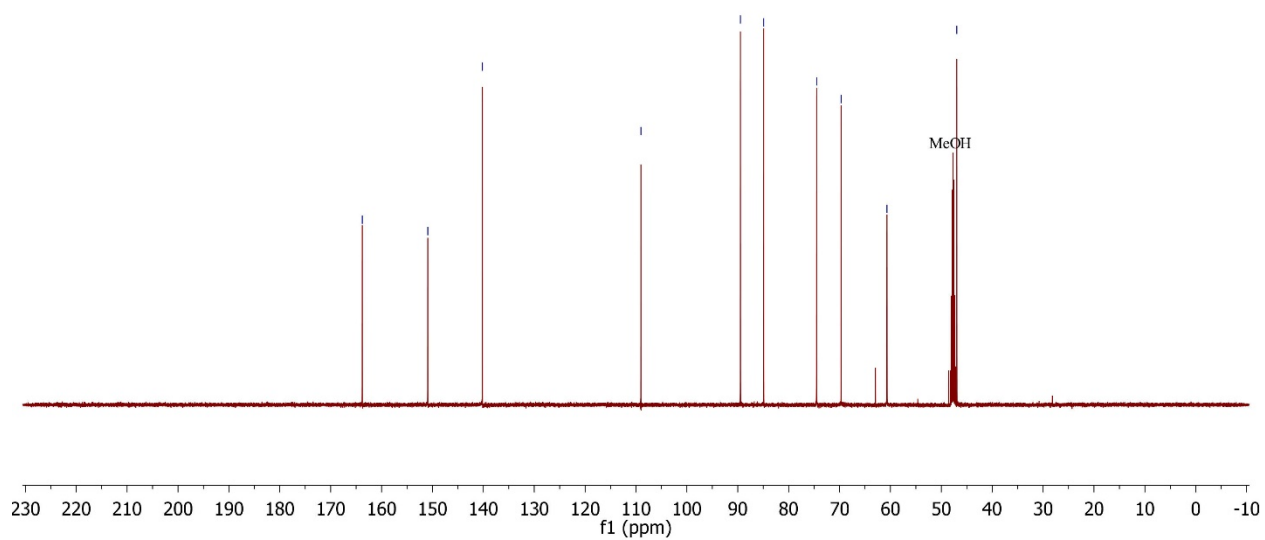
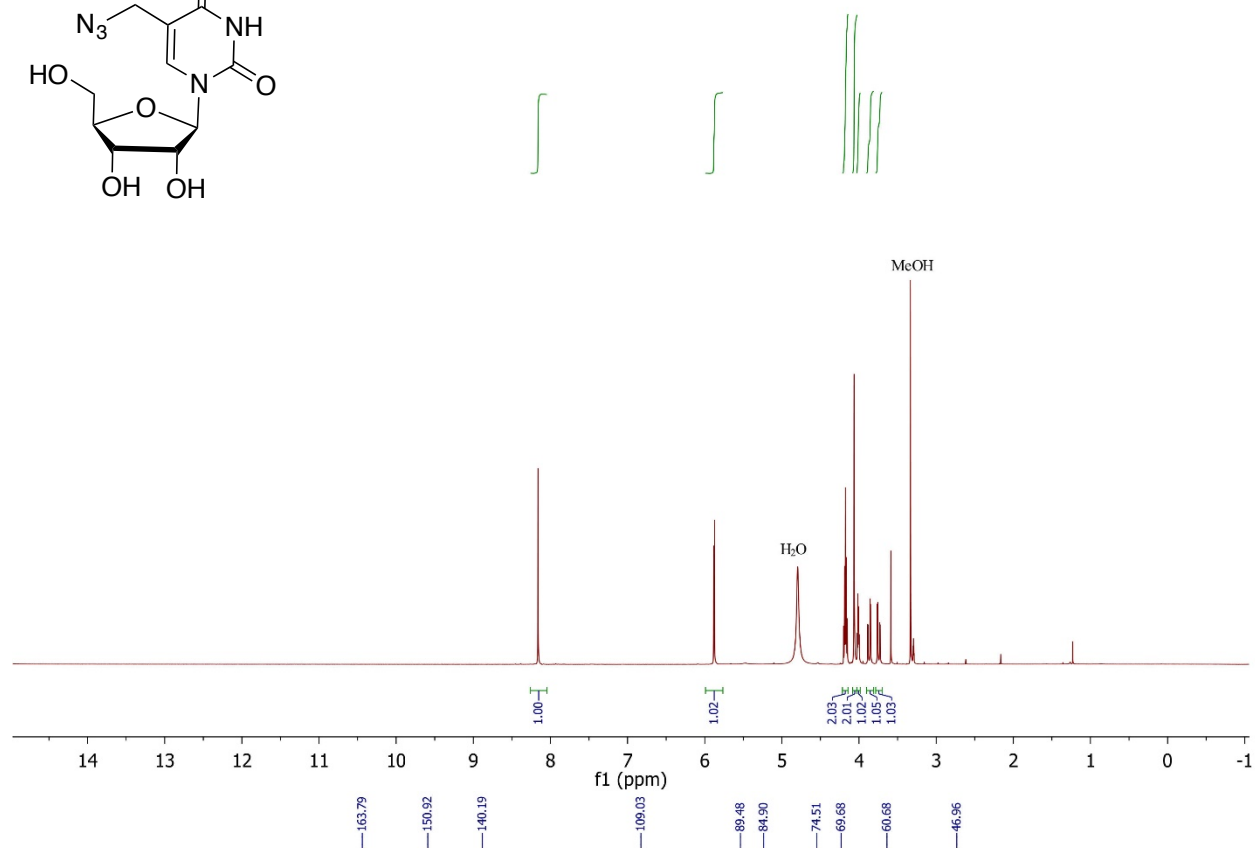
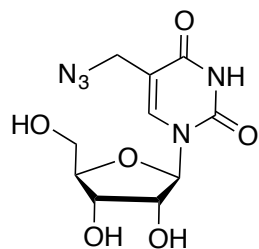
The glycosylated product was prepared according to a modified procedure by Vorbrüggen, H. *et al.* and Rao, H. *et al.*^{20,24} To a suspension containing **4a** (1.00 g, 5.98 mmol), 1-O-acetyl-2,3,5-tri-O-benzoyl- β -D-ribofuranose (3.02 g, 5.98 mmol) in dry MeCN (75 mL) was added HMDS (1.00 mL, 4.78 mmol) and TMS-Cl (0.610 mL, 4.78 mmol) and was stirred for 30 min at r.t. and under Ar. SnCl₄ (0.840 mL, 7.18 mmol) was dissolved in dry MeCN (25 mL). The solution of SnCl₄ 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. NaHCO₃ solution. The organic layer was dried over MgSO₄, 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 C₁₀H₁₃N₅O₆Na [M+Na]⁺ 322.08, found 322.07 [M+Na]⁺; ¹H 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); ¹³C NMR (125 MHz, CD₃OD) δ 163.79, 150.92, 140.19, 109.03, 89.48, 84.90, 74.51, 69.68, 60.68, 46.96.

Spectra









2.6 Supplementary Data

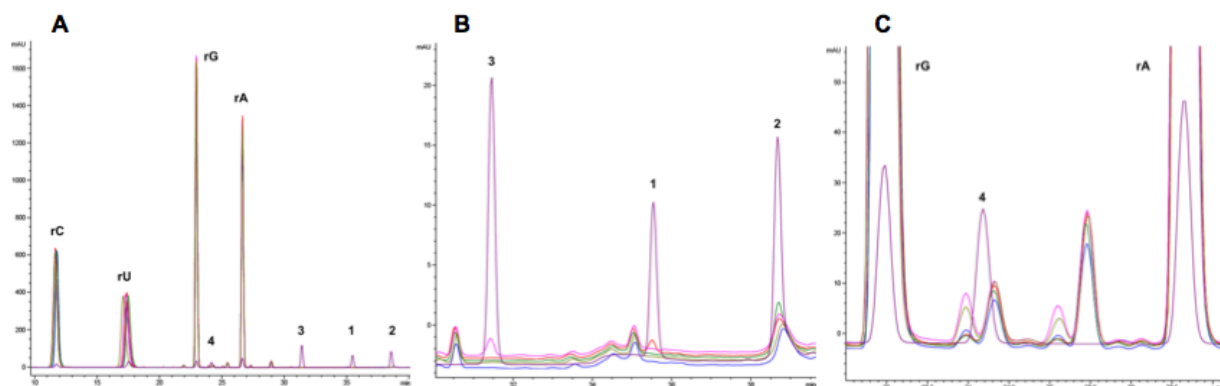


Figure 2-S1. LC-MS data demonstrating that analogs 1-3 are incorporated into cellular RNA. **A.** Ribonucleosides obtained from digestion of total RNA were resolved by reversed-phase HPLC and detected by UV absorption at 260 nm. Digested RNA from cells incubated with N⁶-ethylazido adenosine 1 (red), N⁶-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-methylazidouridine, **4**.

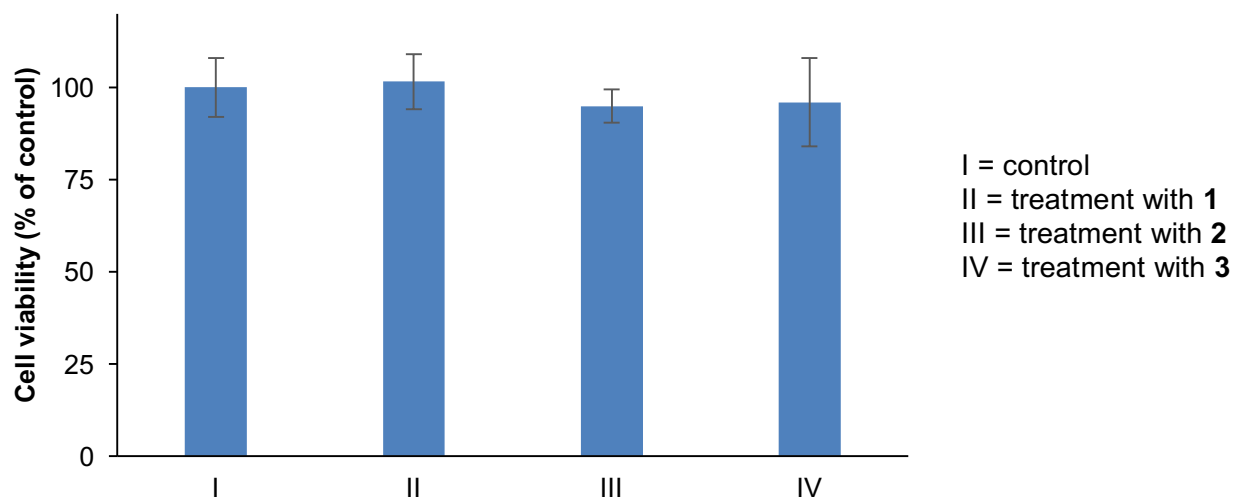


Figure 2-S2. Cell viability assay following analog treatment. HeLa cells were treated with analogs **1-4** at 1 mM concentration for 12 h. Following incubation, MTT reagent was added to the cells and cell proliferation and respiration was assessed by the ability of the cells to reduce MTT to the insoluble formazan. Cell viabilities were normalized to the DMSO-treated control.

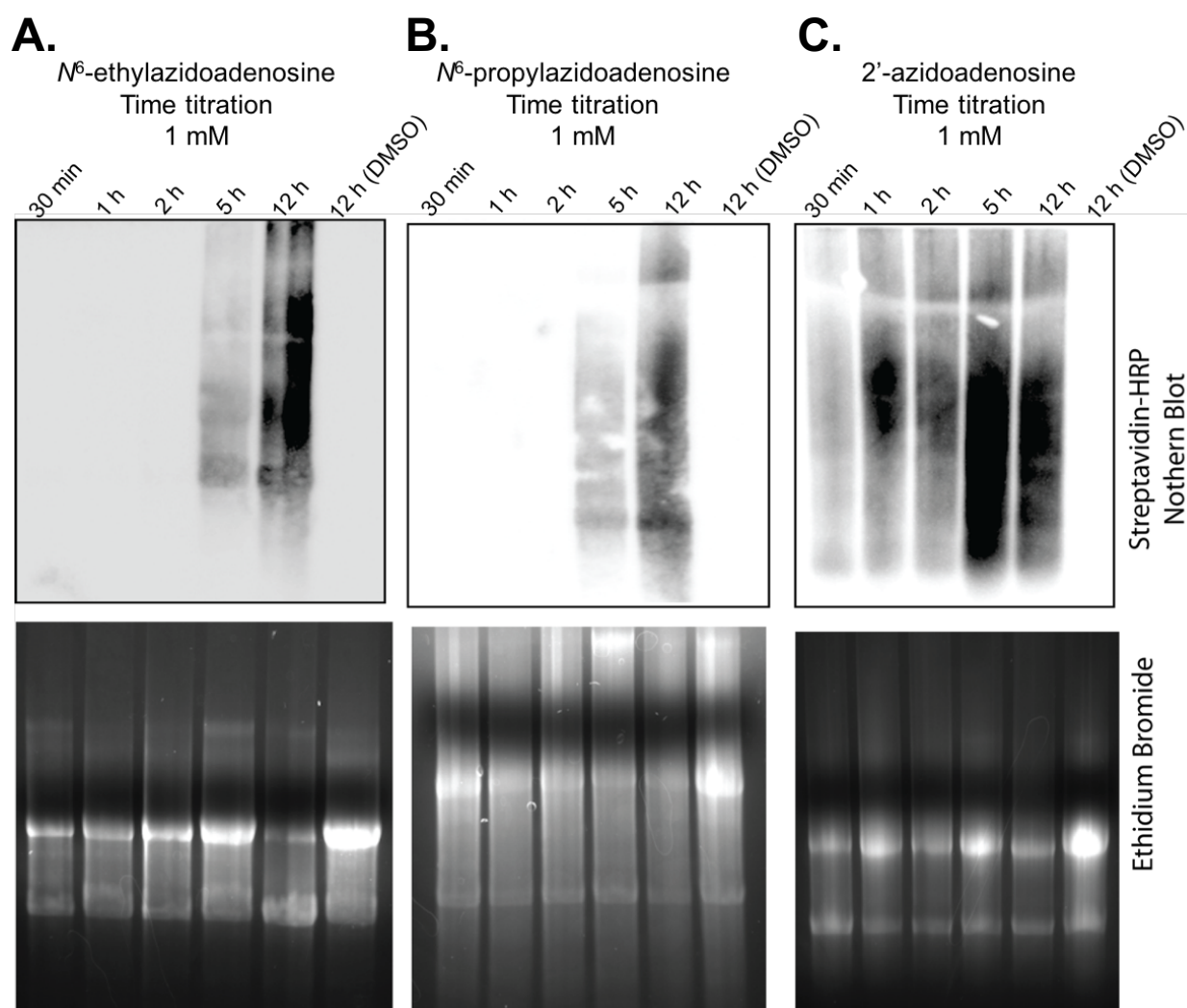


Figure 2-S3. Full Northern blots represented in Figure 2-1, C. **A.** Northern blot for N^6 -ethylazidoadenosine. **B.** Northern blot for N^6 -propylazidoadenosine. **C.** Northern blot for 2'-azidoadenosine. HeLa cells were treated with analog at 1 mM concentrations for 30 min up to 12 h. Following treatment, RNA was extracted and subjected to CuAAC biotinylation following by Streptavidin-HRP northern blot detection.

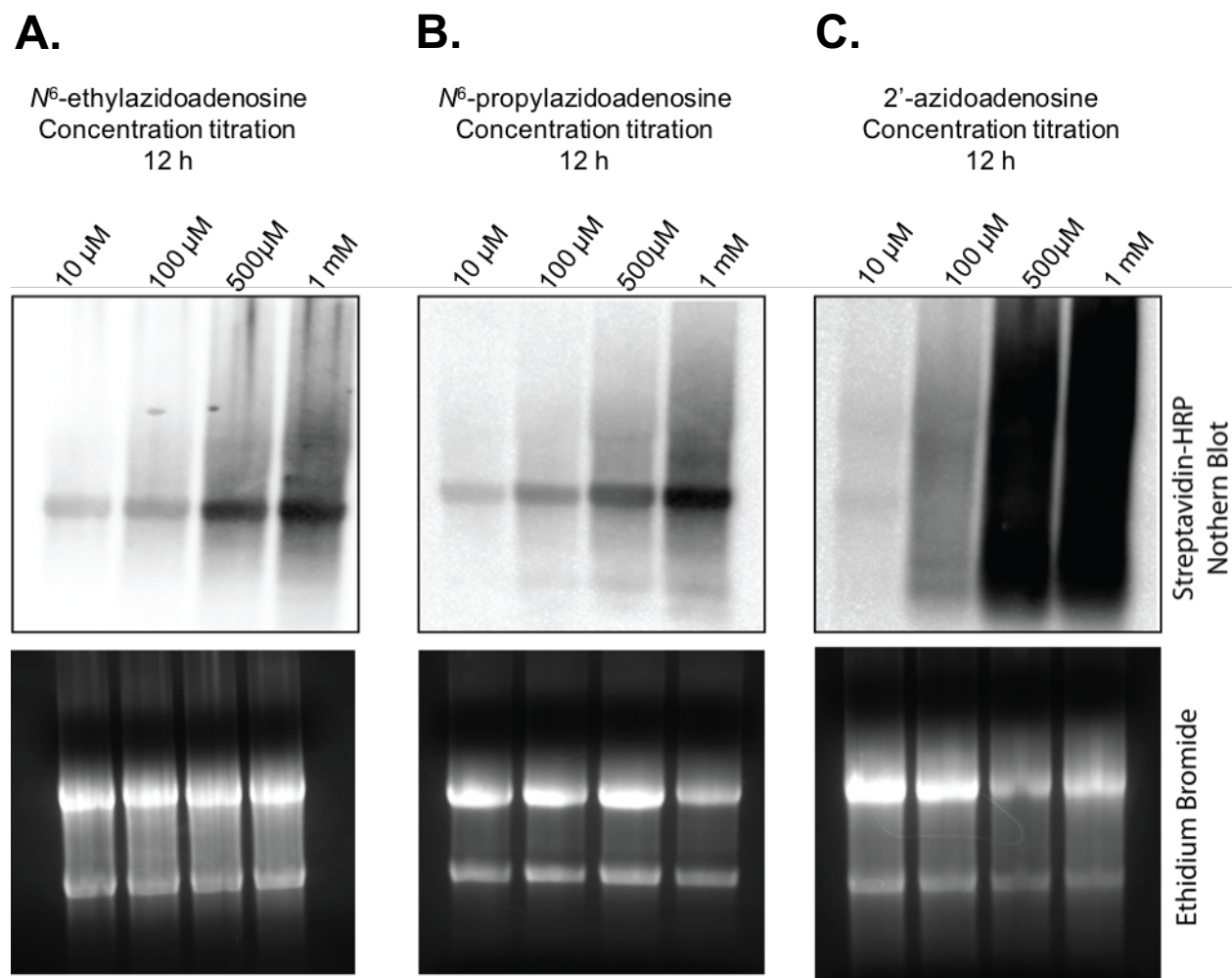


Figure 2-S4. Concentration titration of analogs 1-3. **A.** Northern blot for *N*⁶-ethylazidoadenosine. **B.** Northern blot for *N*⁶-propylazidoadenosine. **B.** Northern blot for 2'-azidoadenosine. HeLa cells were treated with titrated analog for 12 h. Following treatment, RNA was extracted and subjected to CuAAC biotinylation following by Streptavidin-HRP northern blot detection.

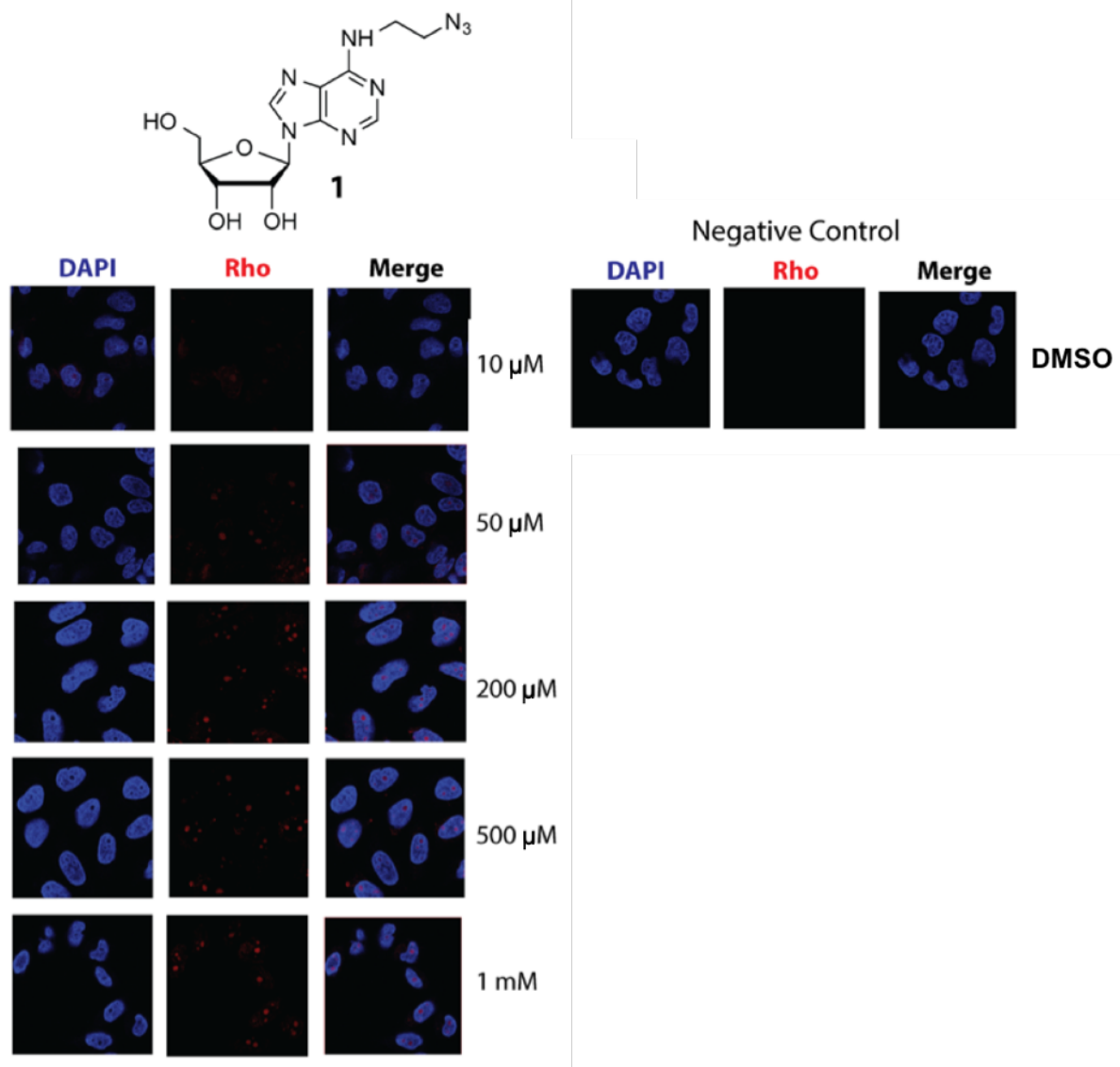


Figure 2-S5. Imaging incorporation of *N*⁶-ethylazidoadenosine, **1 into nascent RNA.**

HeLa cells were treated with titrated *N*⁶-ethylazidoadenosine, **1** for 12 h, followed by fixation, permeabilization and CuAAC functionalization with alkyne-rhodamine.

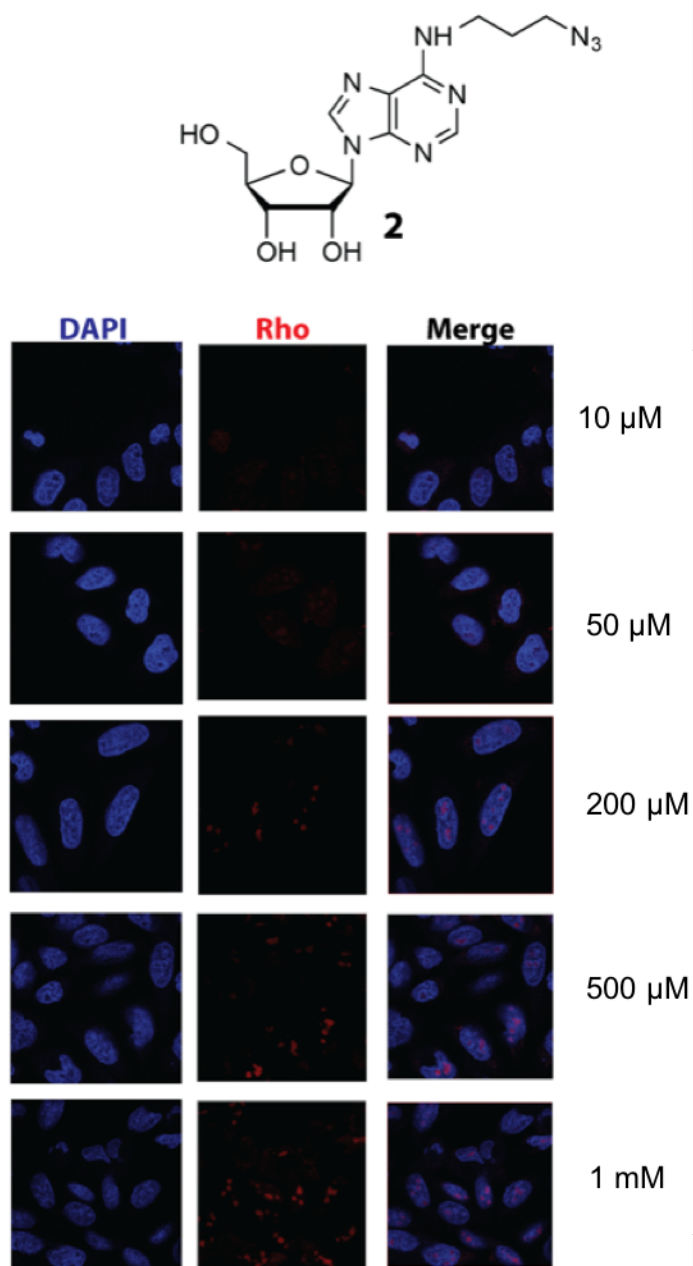


Figure 2-S6. Imaging incorporation of *N*⁶-propylazidoadenosine, **2 into nascent RNA.**

HeLa cells were treated with titrated *N*⁶-propylazidoadenosine, **1** for 12 h, followed by fixation, permeabilization and CuAAC functionalization with alkyne-rhodamine.

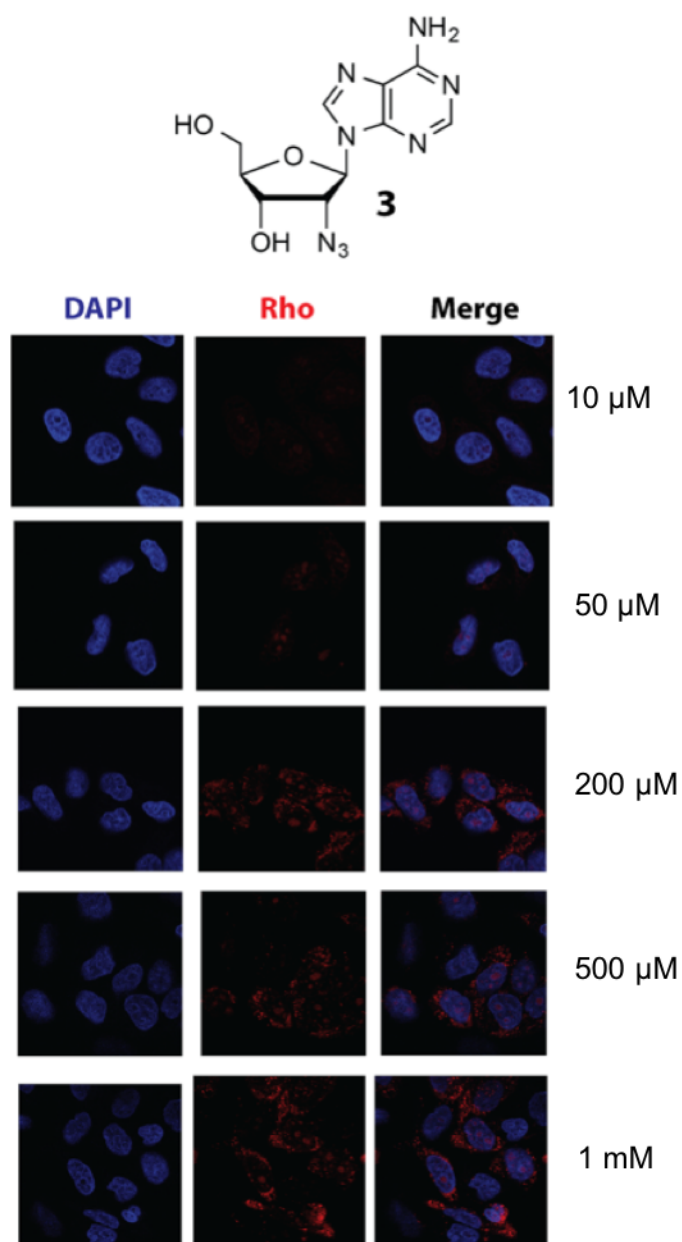


Figure 2-S7. Imaging incorporation of 2'azidoadenosine, **3 into nascent RNA.**

HeLa cells were treated with titrated 2'azidoadenosine, **3** for 12 h, followed by fixation, permeabilization and CuAAC functionalization with alkyne-rhodamine.

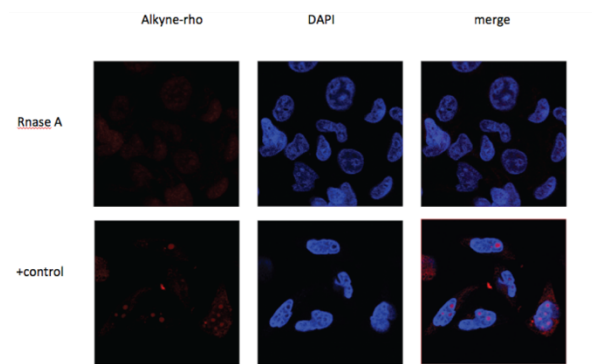
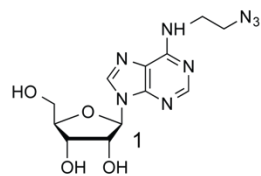
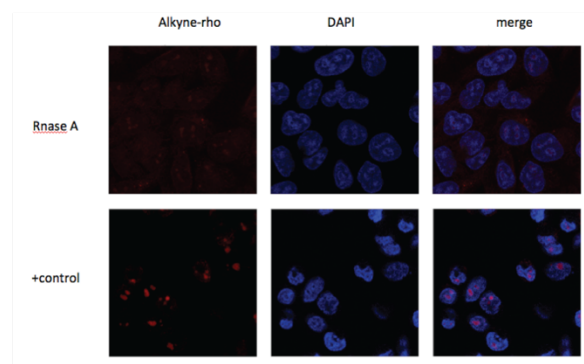
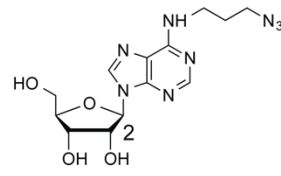
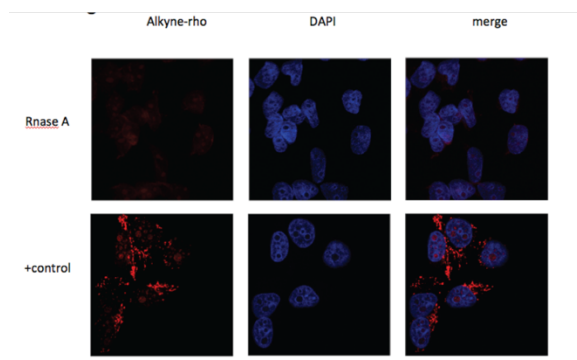
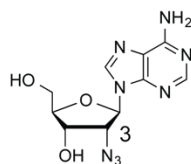
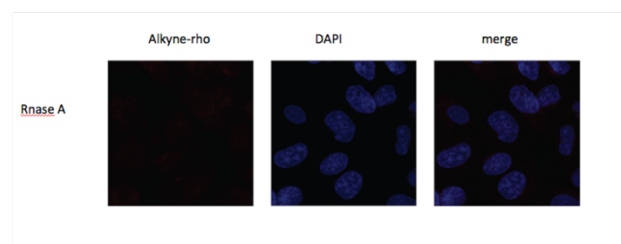
A.**B.****C.****D.**

Figure 2-S8. RNase A imaging of cells treated with azido-nucleosides. HeLa cells were incubated with analogs **1** – **3** (**A**, **B** and **C** respectively) at 1 mM for 12 h. Following incorporation, cells were lightly fixed and permeabilized before treatment with RNase A to digest RNA, and reacted with alkyne-rhodamine via CuAAC.

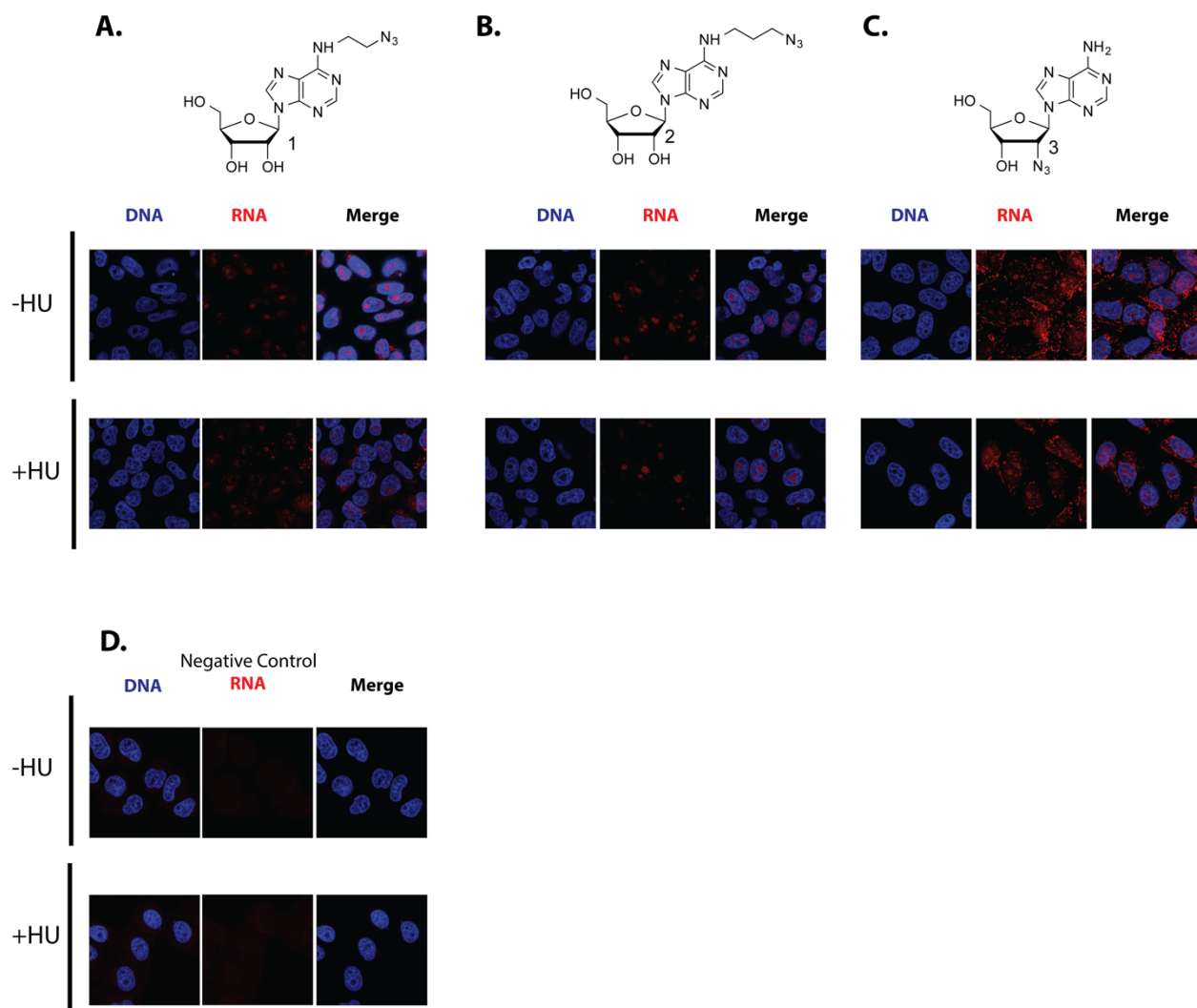


Figure 2-S9. Imaging to confirm incorporation into RNA and not DNA. HeLa cells were treated with 1 mM analog and 10 mM hydroxyurea (HU) for 12 h. Following treatment, cells were fixed, permeabilized and reacted with alkyne-rhodamine followed by fluorescent imaging.

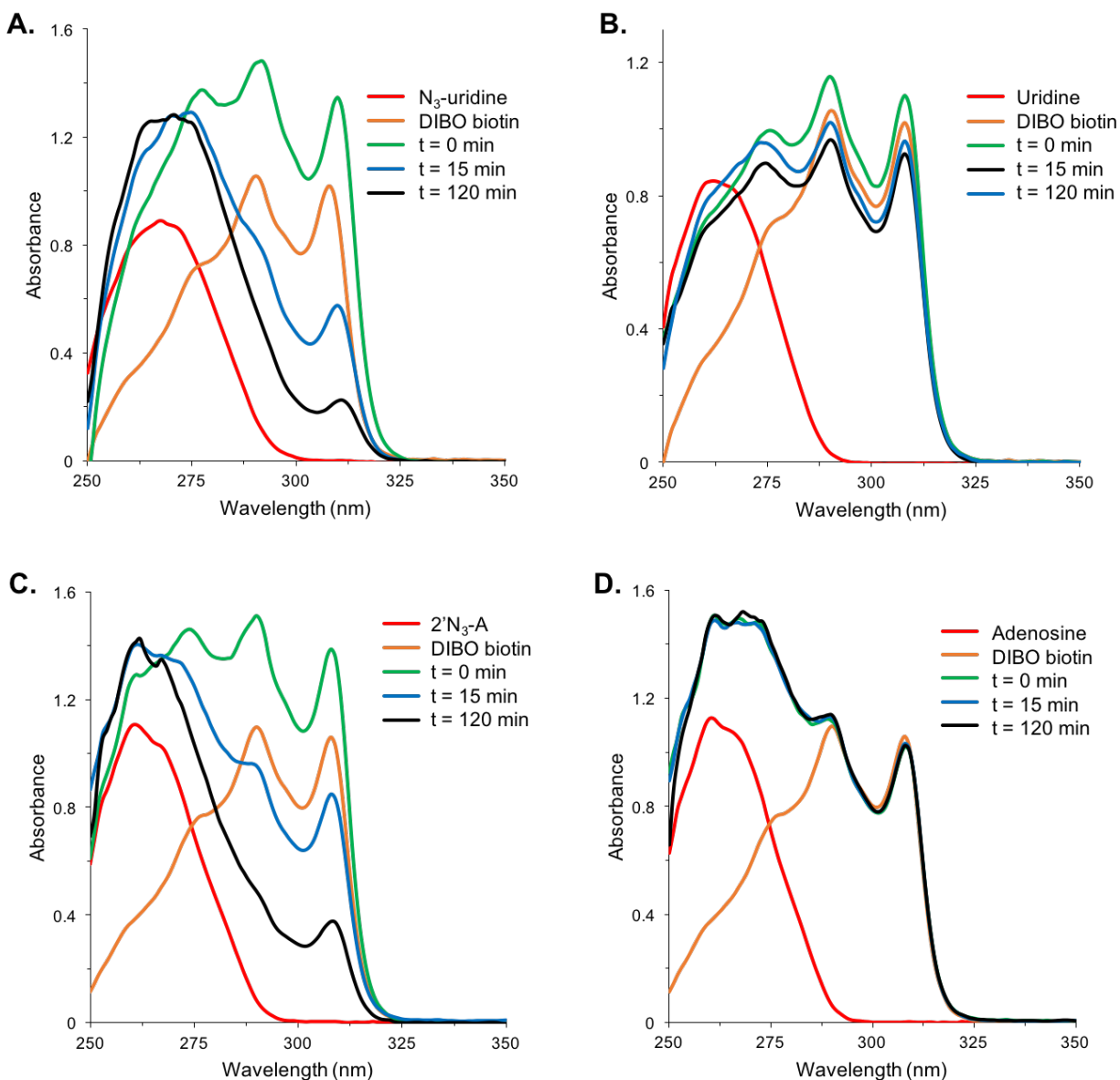


Figure 2-S10. UV spectra to confirm SPAAC reactivity of azido-nucleosides. UV spectra of azido-nucleosides (**A and C**) or nucleosides (**B and D**) (667 μM), DIBO biotin (667 μM) and 1:1 reaction mixture (667 μM final) following incubation for 0, 15 or 120 min. DIBO peaks at 290 nm and 310 nm bleach over time in reactions with azide-containing nucleosides after SPAAC reactions form the triazole product.

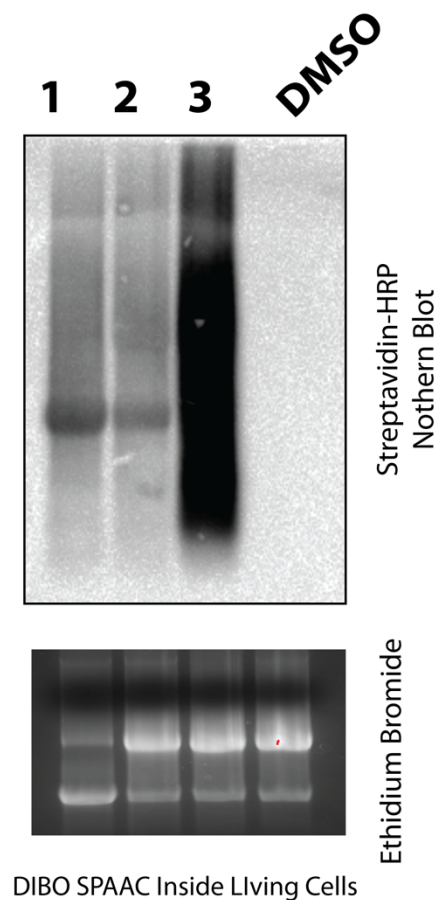


Figure 2-S11. DIBO SPAAC labeling of RNA inside living cells. Analogs **1-3** were incubated in HeLa cells at 1 mM concentration, 12 h. After incorporation and replacing the cells with clean media, 50 μ M DIBO-biotin was added to the media and the SPAAC reaction was allowed to proceed in live cells for 2 h. Cells were subsequently

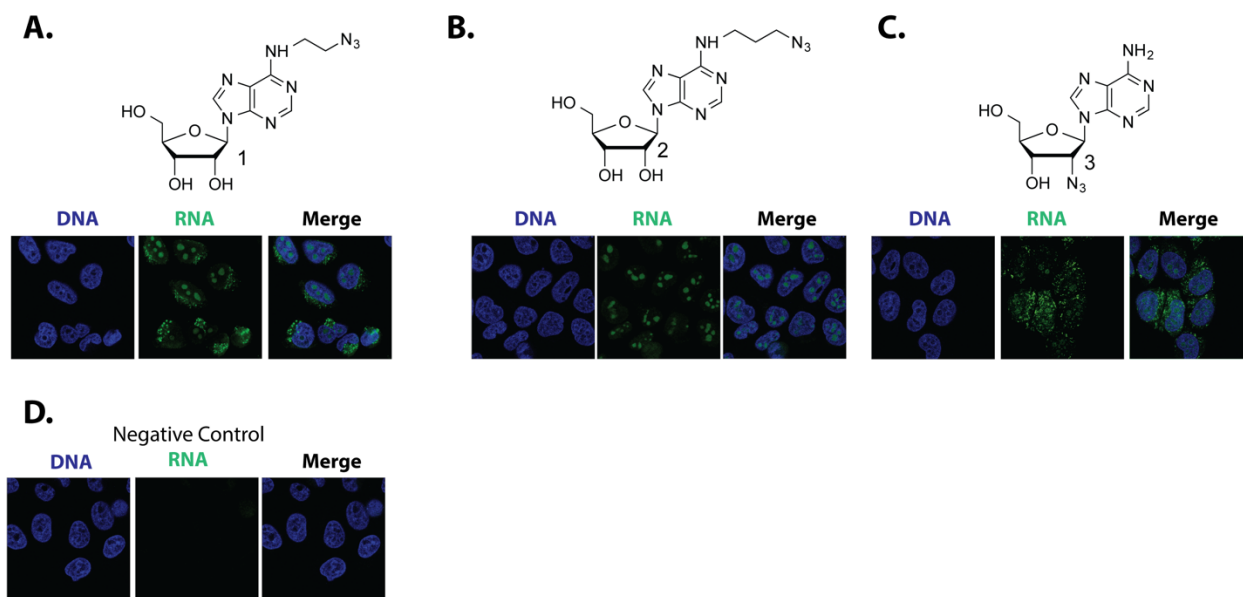


Figure 2-S12. DIBO SPAAC imaging of RNA inside cells. Analogs **1-3** were incubated in HeLa cells at 1 mM for 12 h. After incorporation and replacing the cells with clean media, cells were washed, fixed and permeabilized. 50 μ M DIBO-Alexa 488 was added to the media and the SPAAC reaction was allowed to proceed in live cells for 2 h, after which cells were fluorescently imaged.

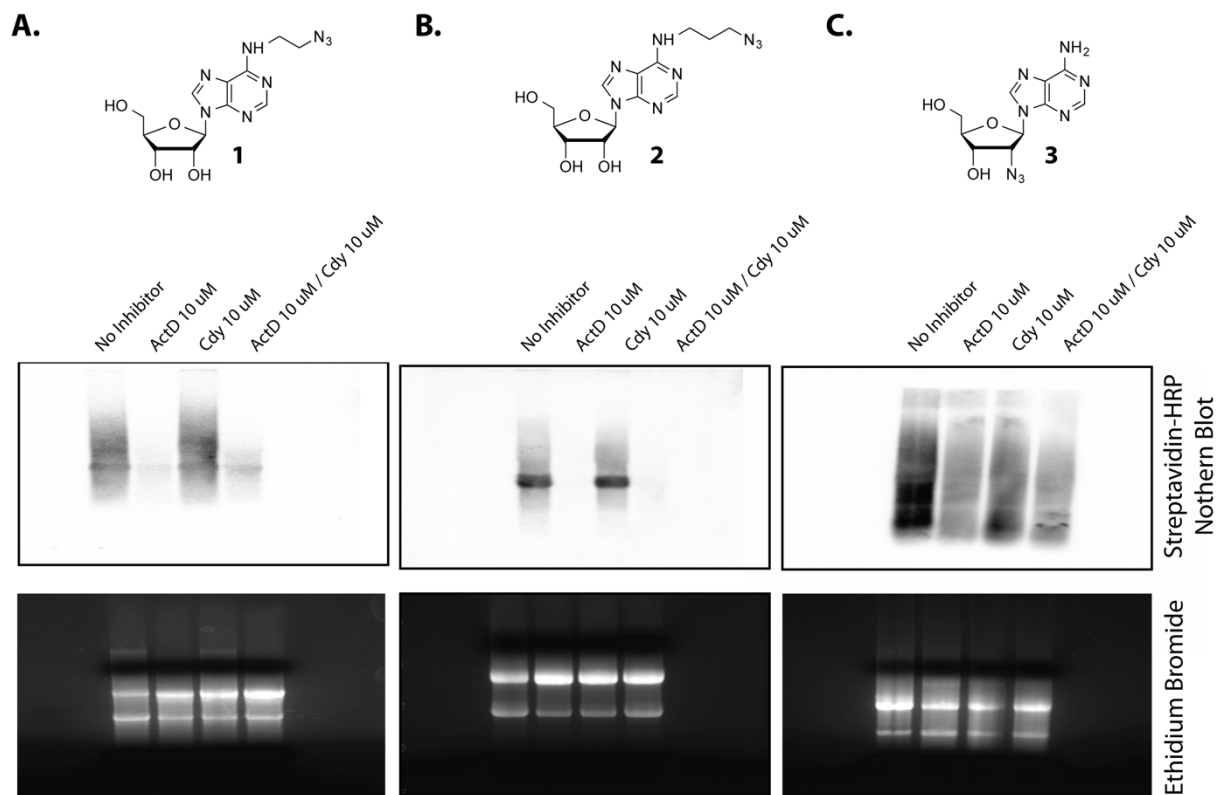


Figure 2-S13. Selectivity of RNA polymerases and their effects on azido-nucleoside incorporation. HeLa cells were concurrently treated with 1 mM analog and 10 μ M Actinomycin D (ActD) and/or cordycepin (Cdy), followed by Northern blot analysis to assess the effects on labeling.

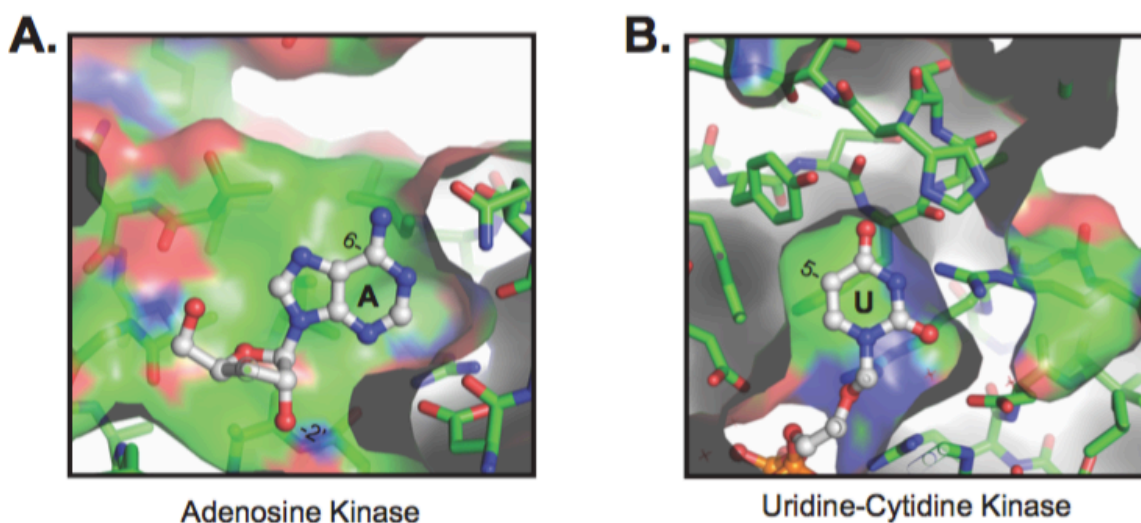


Figure 2-S14. Crystal structures of the purine and pyrimidine monophosphate kinases. **A.** Crystal structure of the adenosine kinase bound to adenosine; PDB: 1BX4 **B.** Crystal structure of uridine/cytidine kinase bound to UTP; PDB: 1UEI.

Table 2-S1. Relative abundance of canonical and azido-nucleoside analogs in total RNA samples. The relative abundance of each nucleoside was normalized to rA (set as 1).

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

2.7 References

1. Morris, K. V.; Mattick, J. S. *Nat. Rev. Genet.* **2014**, *15*, 423-437.
2. Quinn, J. J.; Chang, H. Y. *Nat. Rev. Genet.* **2016**, *17*, 47-62.
3. Rabani, M.; Levin, J. Z.; Fan, L.; Adiconis, X.; Raychowdhury, R.; Garber, M.; Gnirke, A.; Nusbaum, C.; Hacohen, N.; Friedman, N.; Amit, I.; Regev, A. *Nat. Biotechnol.* **2011**, *29*, 436-442.
4. Duffy, E. E.; Rutenberg-Schoenberg, M.; Stark, C. D.; Kitchen, R. R.; Gerstein, M. B.; Simon, M. D. *Mol. Cell* **2015**, *59*, 858-866.
5. Jao, C. Y.; Salic, A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 15779-15784.
6. Grammel, M.; Hang, H.; Conrad, N. K. *ChemBioChem* **2012**, *13*, 1112-1115.
7. Curanovic, D.; Cohen, M.; Singh, I.; Slagle, C. E.; Leslie, C. S.; Jaffrey, S. R. *Nat. Chem. Biol.* **2013**, *9*, 671-673.
8. Zheng, Y.; Beal, P. A. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 1799-1802.
9. Paredes, E.; Das, S. R. *ChemBioChem* **2011**, *12*, 125-131.
10. Winz, M. L.; Samanta, A.; Benzinger, D.; Jaschke, A. *Nucleic Acids Res.* **2012**, *40*, e78.
11. Feng, H.; Zhang, X.; Zhang, C. *Nat. Commun.* **2015**, *6*, 7816.
12. Gallego Romero, I.; Pai, A. A.; Tung, J.; Gilad, Y. *BMC Biol.* **2014**, *12*, 42.
13. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002**, *41*, 2596-2599.
14. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004**, *126*, 15046-15047.
15. Ning, X.; Guo, J.; Wolfert, M. A.; Boons, G. J. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 2253-2255.
16. Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007-2010.

17. Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9482-9487.
18. Sun, H.; Jiang, S.; Caton-Williams, J.; Liu, H.; Huang, Z. *RNA* **2013**, *19*, 1309-1314.
19. Samanta, A.; Krause, A.; Jäschke, A. *Chem. Commun. (Cambridge, U. K.)* **2014**, *50*, 1313-1316.
20. Rao, H.; Tanpure, A. A.; Sawant, A. A.; Srivatsan, S. G. *Nat. Protoc.* **2012**, *7*, 1097-1112.
21. Martin, G.; Keller, W.; Doublié, S. *EMBO J.* **2000**, *19*, 4193-4203.
22. Castello, A.; Horos, R.; Strein, C.; Fischer, B.; Eichelbaum, K.; Steinmetz, L. M.; Krijgsveld, J.; Hentze, M. W. *Nat. Protoc.* **2013**, *8*, 491-500.
23. Neef, A. B.; Pernot, L.; Schreier, V. N.; Scapozza, L.; Luedtke, N. W. *Angew. Chem., Int. Ed.* **2015**, *54*, 7911-7914.
24. Vorbrüggen, H.; Bennua, B. *Chem. Ber.* **1981**, *114*, 1279-1286.

Chapter 3: Temporal Labeling of Nascent RNA Using Photo-click Chemistry in Living Cells

3.1 Introduction

Recent gene expression analysis has revealed that the majority of cellular RNA is non-coding. It has further been elucidated that RNA controls or regulates nearly every step of gene expression, from transcription to translation.¹ The RNA transcriptome is thus highly dynamic and can vary amongst different cell types, even within the same tissue. Despite the importance of RNA and its specificity of expression, there is a deficit of chemical methods to probe the transcriptome of unique cell types with precise spatiotemporal control.

As stated in Chapter 2, bioorthogonal chemistries have, however, provided avenues toward these analyses, as there is currently a robust array of “click” reactions that enable tagging of biomolecules of interest.² The chemical reporter strategy has hastened the development of such reagents, which have found widespread use in studying glycosylation, protein function and localization and DNA dynamics.³⁻⁵

The most widely utilized click-reaction for RNA labeling is the azide-alkyne cycloaddition, of which there are two classes.⁶ While the copper catalyzed azide-alkyne cycloaddition (CuAAC) has been used extensively in many biological contexts, the use of a cytotoxic copper catalyst limits the use of terminal alkynes in living systems.^{7,8} Conversely, azide functionality is more versatile, providing additional utility in metal-free strain promoted azide-alkyne cycloaddition (SPAAC) reactions.⁹ Our lab has recently reported that azide-modified nucleosides can metabolically incorporate into nascent RNA, thus enabling SPAAC reactivity.¹⁰ SPAAC reactions have been widely used for many live cell

applications in addition to *in vivo* imaging.^{4,11} While these methods are highly useful, they do not enable cell-specific tagging, which would endow experimenters with kinetic control of RNA trapping.

It has been reported that short irradiation at 350 nm of cyclopropenone-caged cyclooctyne causes decarbonylation to yield a reactive triple bond.¹² Notably, the products of such decarbonylation reactions can serve as SPAAC reagents with faster reaction kinetics.¹³⁻¹⁶ These light-activated photo-click SPAAC reagents have thus far been used as cross-linkers for surface immobilization and probes for cell surface labeling.^{17,18} However, their intracellular stability and subsequent applications for the selective labeling of many biomolecules have yet to be fully explored. This has limited the use of photo-click SPAAC reactions in cells, and specifically their utility for selective RNA tagging.

Herein, we report the first intracellular use of a photo-click SPAAC reagent for tagging nascent azide-modified RNA. We demonstrate greater *in vitro* stability of the cyclopropenone-caged oxa-dibenzocyclooctyne versus the uncaged molecule in the presence of thiols, which most importantly confers greater reactivity of tagging in living cells. We also demonstrate the ability to use this reagent for light-activated tagging of RNA in live cells, in addition to tagging the RNA of specific cell subpopulations via selective irradiation and RNA fluorescence imaging or enrichment.

3.2 Results and Discussion

With chemically synthesized cyclopropenone-caged oxa-dibenzocyclooctyne (photo-ODIBO **1**) and ODIBO **1a** in hand (**Methods: Synthesis**), we set out to establish their reactivity with the nucleoside analog 2'azidoadenosine (2'N₃-A). Upon 2 min of irradiation

at 350 nm, photo-ODIBO **1** decarbonylates to form ODIBO **1a**, which reacts rapidly with 2'N₃-A to form the resulting triazole product (**Figure 3-1, A**). We were able to fully observe this transformation via UV spectroscopy, where the spectra before irradiation displays a combination of absorbances from both **1** at 350 nm and 2'N₃-A, with a characteristic absorbance at 260 nm. Following decarbonylation, the formation of ODIBO **1a** results in a notable blue shift to 340 nm. Finally, as the 2'N₃-A reacts with **1a**, this band undergoes bleaching as the triazole product is formed. It was seen that after only 15 min, the reaction had gone to near completion (**Figure 3-1, B; Figure 3-S1**). We also measured the rate constant of ODIBO **1a** with 2'N₃-A via UV spectroscopy. Using pseudo-first order conditions in excess of **1a**, we found an exceptionally fast rate constant of $40 \pm 2 \text{ M}^{-1} \text{ s}^{-1}$ (**Figure 3-1, C**).

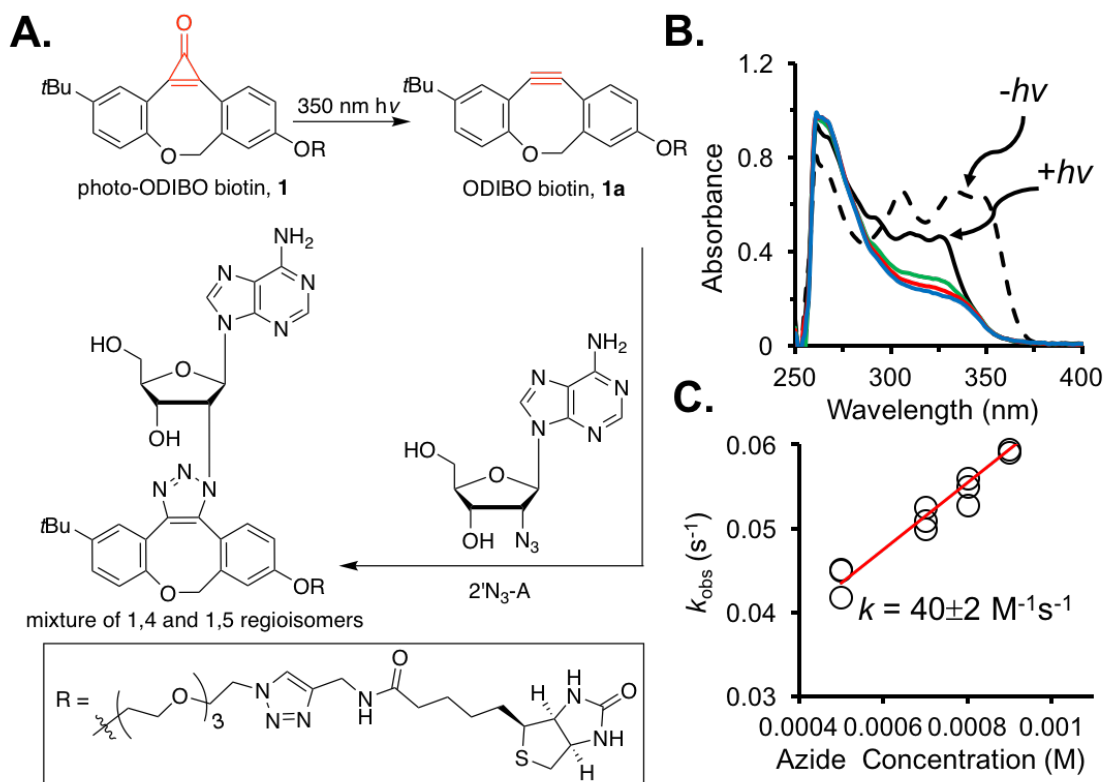


Figure 3-1. Establishing photo-click reactivity with azido-nucleoside. **A.** Scheme of photo-click reaction between photo-ODIBO **1** and 2'azidoadenosine. **B.** UV spectra of 1:1 500 μ M solution of **1** and 2'N₃-A before and after irradiation (solid, and dashed), and subsequent formation of the triazole product after 15, 30 and 60 min (green, red, blue respectively). **C.** Reactivity of ODIBO with azide in PBS, 5% MeOH at 25 °C displaying the linear dependence of the observed rates on 2'N₃-A concentration.

After establishing the kinetics of this reaction, we sought to test the *in vitro* reactivity of photo-ODIBO **1** with extracted cellular RNA that has been metabolically labeled with azide (**Figure 3-2, A**). HeLa cells were treated with 500 μ M 2'N₃-A for 5 h, after which the total RNA was extracted and purified. The azido-RNA was then incubated *in vitro* with 50 μ M of photo-ODIBO **1** or ODIBO **1a** and analyzed via a streptavidin Northern blot (**Figure 3-2, B; Figure 3-S2**). Following a 2 min irradiation with 350 nm light, total RNA was robustly labeled with **1** after only 15 min. Non-irradiated samples showed no labeling even after 60 min of incubation with azido-RNA. DMSO-treated control RNA also showed no labeling.

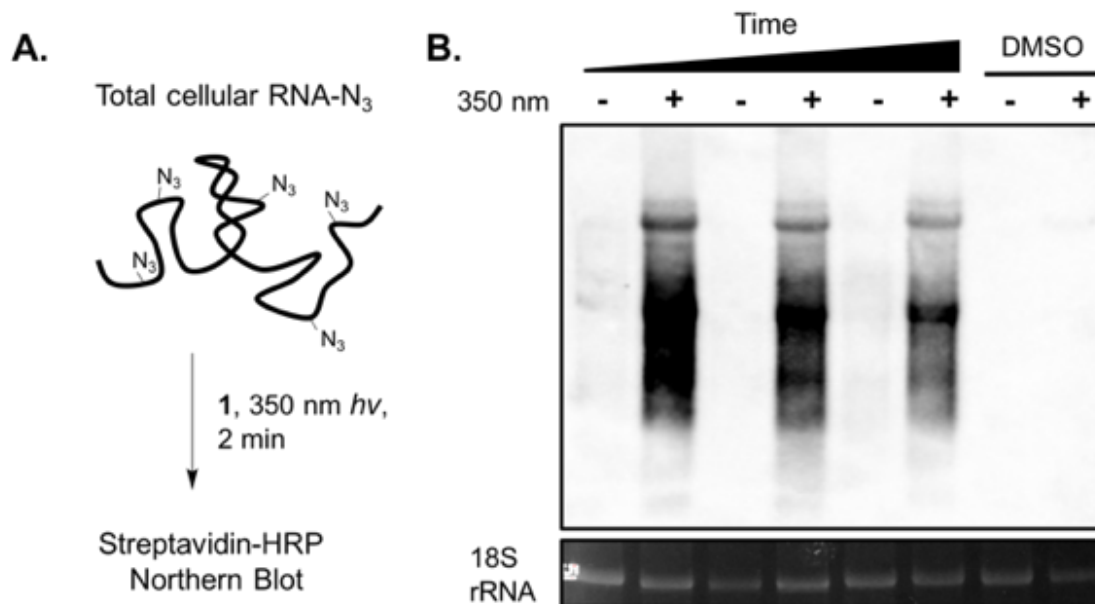


Figure 3-2. *In vitro* photo-click SPAAC with azido-RNA. **A.** Schematic of reaction of azido-RNA with photo-ODIBO **1**. **B.** Streptavidin Northern blot time titration analysis of RNA-N₃ treated with 50 μ M **1**, reacted at 37 °C for 15, 30 and 60 min. Control RNA was extracted from HeLa cells treated with DMSO.

Having demonstrated light-triggered labeling of RNA *in vitro*, we next explored the stability of photo-click SPAAC reagents in the presence of thiols. While cyclooctyne reagents have ample utility for performing ligations in live cells, their robustness can be hampered by thio-yne addition from cellular species such as cysteines and glutathione.¹⁹ Moreover, the cyclopropenone is potentially susceptible to ring opening in the presence of nucleophiles.¹² In order to establish the feasibility of using this reagent intracellularly, we performed *in vitro* competition assays in the presence of titrated thiol. We used β -mercaptoethanol (BME), as its redox potential is comparable to that of glutathione (GSH), the most abundant source of thiol in the cell.²⁰

Upon pre-incubating **1** and **1a** with increasing amounts of thiol, we saw changes in the efficiency of labeling azido-RNA at BME concentrations of 10 and 100 mM. While the signal from adducts formed by photo-ODIBO **1** was reduced at 100 mM, this greatly exceeds physiological conditions and is not likely to be encountered in the native cellular environment. Importantly, labeling by ODIBO **1a** was abolished at the physiologically relevant concentration of 10 mM and greater (**Figure 3-3, A and C**). We also observed no changes in signal when pre-incubating azido-RNA with BME, indicating that thiol-mediated azide reduction was not occurring (**Figure 3-S3**).

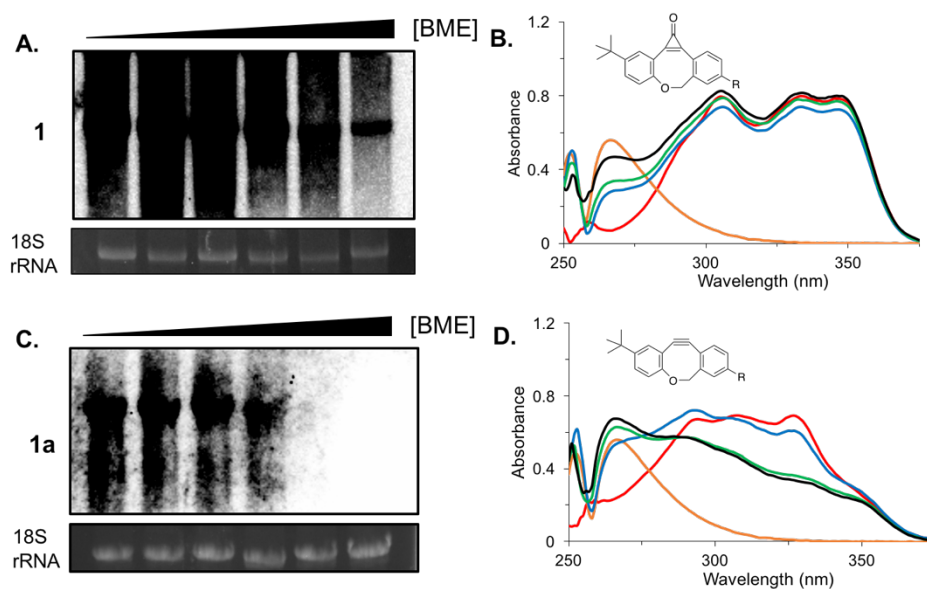


Figure 3-3. Photo-ODIBO, **1 is refractory to thiol addition. A & C.** Streptavidin Northern blot of 50 μM **1** or **1a** pre-incubated with a titration of BME (0, 10 μM, 100 μM, 1 mM, 10 mM, 100 mM) for 30 min at 37 °C, and then reacted with RNA-N₃ for 1 h. **B & D.** 500 μM of **1** or **1a** (red) incubated with 100 mM BME (orange) at 37 °C after 0, 30 and 60 min (blue, green and black, respectively).

To further characterize any side reactivity, we monitored the reaction of **1** and **1a** with BME via UV spectroscopy. We saw that the characteristic peak of photo-ODIBO **1** at 350 nm is unaffected after 1 h, while the blue shifted peak of ODIBO **1a** becomes fully bleached at 30 min (**Figure 3-3, B and D**). From this, we observe that photo-ODIBO **1** is more refractory to thiol addition, demonstrating greater stability of the cyclopropenone-caged cyclooctyne species. We also determined the second order rate constant of ODIBO **1a** with GSH, which was found to be $3.5 \pm 0.2 \times 10^{-4} \text{ M}^{-1} \text{ s}^{-1}$, while there was no measurable reactivity of photo-ODIBO **1** (**Figure 3-S4**). This is of great significance, as free thiol concentrations in cells can exceed 10 mM, concentrations at which the de-caged cyclooctyne species become susceptible to addition before they are able to react.²¹ This data concurs with previous findings that report cyclopropenone stability in aqueous buffers and its ability to be conjugated to thiol ligands.^{22,23} Lastly, this is the first direct demonstration that cyclopropenones with extended conjugation are stable to thiol degradation, further supporting their utility as SPAAC precursors for intracellular ligations.

Encouraged by the observed stability of photo-ODIBO **1a** in the presence of high thiol concentrations, we moved to investigate our ability to tag azido-RNA inside living cells. After treating HeLa cells with 2'N₃-A to install azide functionality (or DMSO control), we allowed them to react with 50 μM of compounds **1** or **1a** for 2 h (**Figure 3-4, A**). Cells treated with photo-ODIBO **1** were pre-incubated for 30 min to allow the probe to diffuse into and distribute throughout cells, after which they were irradiated for 2 min. Following RNA extraction and Northern blot analysis, we found that we were again able to light-dependently label azido-RNA (**Figure 3-4, B**). Interestingly, we saw that the efficiency of

labeling with **1** was reproducibly greater than that of **1a**, supporting our earlier findings of the cyclopropenone conferring chemical inertness to nucleophilic milieu.

Using chemical probes to label RNA is of special interest to the RNA community because it introduces methods to enrich populations of the transcriptome that can be subjected to gene expression analysis. To demonstrate this, we performed semi-quantitative RT-PCR of azido-RNA from cells treated with photo-ODIBO **1** for only 30 min. After extraction and cDNA synthesis, we performed enrichment over streptavidin beads followed by PCR. We were able to successfully recover and enrich two chosen targets: 7SK RNA, a non-coding nuclear transcript, and GAPDH, a cytosolic mRNA. This labeling and subsequent enrichment was also irradiation dependent. RNA from cells not treated with 2'^N₃-A (DMSO control) did not enrich for these targets (**Figure 3-S5**). Herein, we conclude that labeling with **1** is sensitive enough to be used for selective enrichment of transcripts labeled on short timescales.

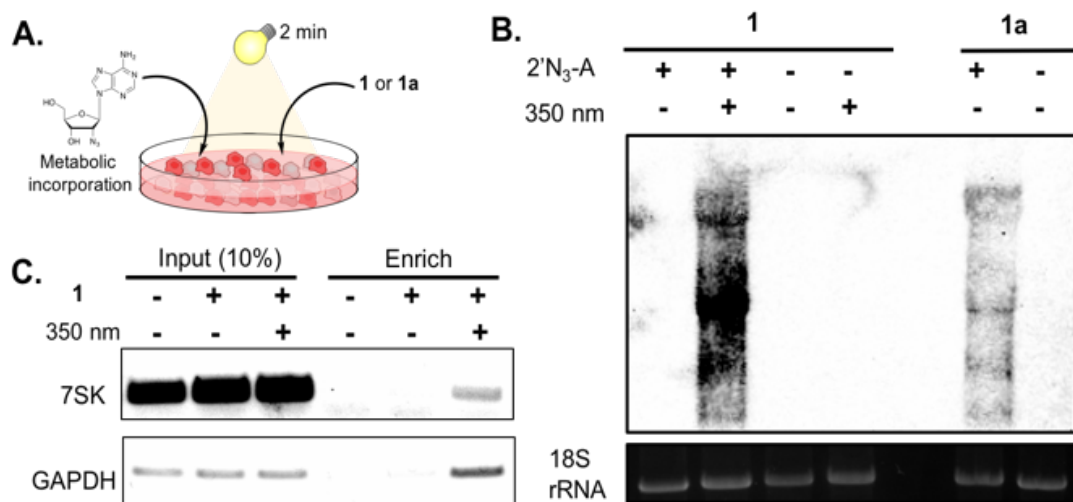


Figure 3-4. Photo-click labeling of RNA in living cells. **A.** Schematic of live cell labeling of HeLa cells with compounds **1** and **1a**. **B.** Streptavidin Northern blot of 2'^N₃-A or DMSO-

treated cells labeled with 50 μ M **1** (plus or minus 350 nm light) or **1a** for 2 h. **C.** RT-PCR of azido-RNA from cells labeled with photo-ODIBO **1** for 30 min. PCR of cDNA successfully synthesized the small non-coding RNA 7SK and the mRNA GAPDH.

Bioorthogonal labeling is also useful for imaging of biomolecules. Thus, we examined our ability to use photo-ODIBO to visualize cellular RNA via fluorescent imaging. To do this, we synthesized photo-ODIBO appended to the fluorophore Rhodamine B (photo-ODIBO-RhB, **2**) (**Figure 3-S6**). After installing azido-functionality, we incubated **2** in cells, after which we were able to visualize strong signal from nascent RNA. In irradiated cells, signal was seen in the nucleolus as well as the cytoplasm, which concurs with our previous findings that 2' N_3 -A is in both nuclear ribosomal RNA and cytosolic mRNA transcripts.¹⁰ DMSO control treated cells showed no fluorescent labeling with **2**, regardless of light irradiation (**Figure 3-S6**). Treatment with RNase A significantly reduced the fluorescent signal from **2**, supporting our findings of RNA labeling (**Figure 3-S7**).

A photo-click SPAAC reagent presents the unique opportunity to label the RNA of specific cells. We posited that the selective irradiation of a targeted region of cells would yield spatio-temporally labeled RNA only in the light-irradiated region of interest. Thus, we created a simple rubber mask through which a window was cut to allow light exposure of half of a cover glass on which HeLa cells are grown (**Figure 3-S8**). Cells treated with 2' N_3 -A were incubated with **2** and one half of the cover glass was irradiated through the exposure window (**Figure 3-S8**). Following incubation, we were able to visualize clear regions of fluorescently labeled and non-labeled cells on opposing sides of the cover glass. These cells also demonstrated the same staining pattern expected from 2' N_3 -A.

We envisioned our approach could be used to selectively label, enrich, and profile the RNA content from unique cells, as current bioorthogonal techniques do not enable this type of application. To further test this, we grew EGFP-expressing cells that were restricted to one half of a tissue culture dish in co-culture with wild-type (WT) HeLa cells (**Figure 3-S9**). After incubation with 2'^N₃-A and pre-incubation with **1**, small regions of the cells were irradiated to specifically tag either EGFP-expressing or WT cells. After labeling for 30 min, RNA was extracted from the whole plate and subjected to enrichment and RT-PCR analysis.

We again observed that we were able to selectively label the RNA of cells based on the location of irradiation (**Figure 3-5**). Enrichment of the EGFP gene only occurred when a region of this subpopulation was exposed to light, while irradiation of the WT population did not enrich for EGFP. GAPDH served as a positive control that could be enriched from both populations. This experiment provides proof of concept for spatiotemporally labeling and enriching the transcriptomes of targeted cell-types, while eliminating the need for specific isolation of these cells prior to RNA extraction.

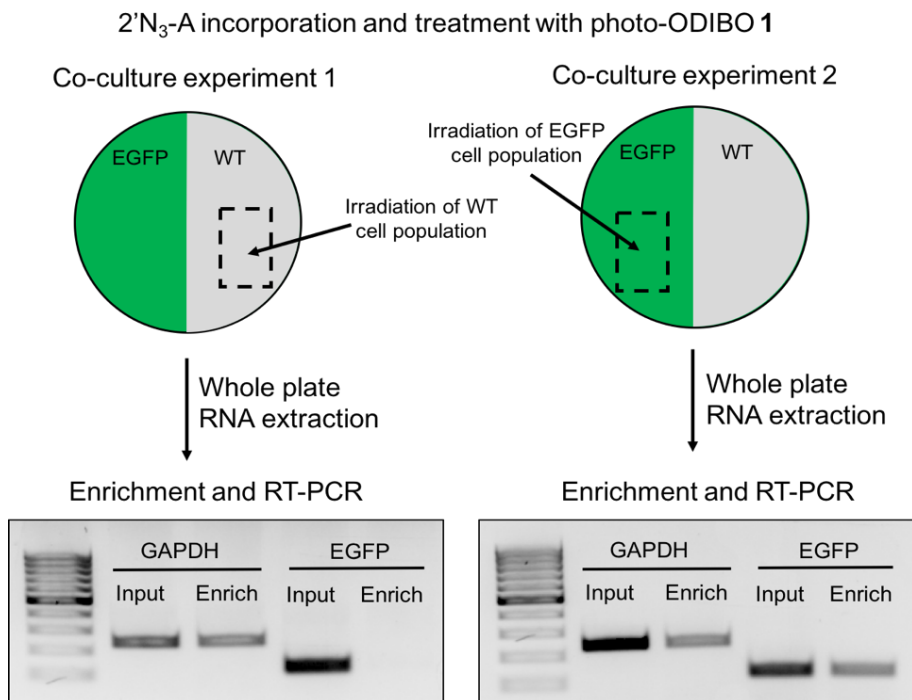


Figure 3-5. Selective irradiation enables RNA labeling of specific cell subpopulations in a co-culture. Cells were grown in co-culture, where EGFP-expressing cells are restricted to one half of the dish. Based on the targeted irradiation of a specific subpopulation, the uniquely expressed EGFP mRNA can be enriched and amplified via RT-PCR.

3.3 Conclusion

Herein, we have demonstrated the efficient use of a photo-click SPAAC reagent for nascent labeling of cellular RNA. We show that the cyclopropenone-caged oxadibenzocyclooctyne remains inert to high concentrations of thiols, while at the same time conferring temporal control to label RNA at will. We are able to use this probe to tag RNA in living cells, and purify various transcripts to be used for downstream gene expression analysis. Furthermore, fluorescent imaging and enrichment followed by RT-PCR

demonstrated light-dependent labeling that can be spatially targeted to a particular cell population. This is an important advance for RNA chemical biology, as the timescales on which RNA expression programs change necessitates the use of reagents that can be spatiotemporally-controlled and used in living cells. We anticipate our approach will be especially valuable, as it endows researchers with the capability to label the transcriptomes of specific cells in complex environments e.g. tissues or tumors, without the need for their specific isolation or cell-sorting. Metabolic labeling can be performed inside living animals, and even the brain, thus our approach could be utilized to label and enrich RNAs from unique brain regions or other sites of interest within complex tissue structures.

3.4 Acknowledgements

Christopher McNitt, supervised by Prof. Vladimir Popik of the University of Georgia, synthesized the photo-O-DIBO and O-DIBO biotin and rhodamine conjugates, and determined the rate constants with the 2'-azidoadenosine analog. Miles Kubota contributed to early syntheses and design of this project.

3.5 Methods

UV spectral analysis of photo-click reaction.

Standards of photo-ODIBO 1, ODIBO 1a and 2'-azidoadenosine were made to concentrations of 500 μ M in DMSO. Following blank with DMSO, spectra were taken to obtain the absorbance signatures of each compound or compound mixture. Solutions of

2'N₃A and photo-ODIBO or ODIBO were made in DMSO to a final concentration of 500 μ M. Spectra of reaction mixtures were taken prior to irradiation, after 2 min of irradiation (in the case of photo-ODIBO) and at incubation times 0, 15, 30 and 60 min. Absorbance readings were taken on a NanoDrop 2000 UV-Vis Spectrophotometer (ThermoFisher)

Kinetic Experiments.

The kinetics of the reaction of ODIBO 1a with 2'-azidoadenosine were studied in PBS (pH 7.4) at 25.0 $\pm$ 0.1 $^{\circ}$ C. The accurate rate measurements were conducted under pseudo-first order conditions, using variable concentrations of the azide at 10-fold or higher excess. Consumption of ODIBO was followed by the decay of the characteristic alkyne peak at 321 nm. The rates were measured in triplicate at each azide concentration. The second order rate constant has been calculated by the least squares analysis of the observed rate constants on the azide concentration.

Cell culture and azido-nucleoside labeling of cellular RNA.

HeLa cell lines (ATCC) were cultured in DMEM supplemented with 10% FBS, penicillin and streptomycin and grown at 37 $^{\circ}$ C, 5% CO₂. The nucleoside analog 2'-azidoadenosine (TriLink Biotechnologies) was added to complete culture medium from a 200 mM DMSO stock to a final concentration of 500 μ M with <1% DMSO and incubated for 5 h.

RNA isolation and biotinylation via SPAAC *in vitro*.

After treating cells with 2'azidoadenosine or DMSO control, total cellular RNA was harvested using Trizol Reagent (Invitrogen) following the manufacturer's instructions and ethanol precipitated. Click reactions were prepared using 3-5 µg of purified azido- or DMSO-treated RNA that was pre-incubated in 20% v/v formamide, heated to 95 °C and snap cooled on ice, followed by incubation with and 50 µM of photo-ODIBO or ODIBO in a final reaction volume of 50 µL. Samples treated with photo-ODIBO **1** were irradiated with a standard hand-held UV lamp (350 nm) for 2 minutes. The reactions were incubated with shaking at room temperature for 15, 30 and 60 min. Biotinylated RNA was purified using Zymo RNA clean and concentrator-5 spin column according to the manufacturer's 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% TAE agarose gel. RNA was then transferred onto Hybond-N+ membrane (GE Healthcare) using vacuum pressure blotting (Biometra, Analytic Jena Company) and UV-crosslinked to a membrane using a UV Stratalinker 1200 (Stratagene). The membrane was blocked in a 10% SDS blocking solution followed by incubation with 1:10,000 dilution of high sensitivity streptavidin-HRP (ThermoFisher Scientific) in blocking solution. The membrane was washed twice in a 1:10 solution of blocking solution and twice in Tris-saline buffer. It was then incubated in

SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher Scientific) according to manufacturer's instructions and imaged on a ChemiDoc MP imaging system (Bio-Rad).

***In vitro* BME competition assay.**

50 μ M of photo-ODIBO **1** or ODIBO **1a** were pre-incubated with thiol concentrations of 0, 10 μ M, 100 μ M, 1 mM, 10 mM, and 100 mM for 30 min at 37 °C. Azido-RNA was subsequently added to the reaction mixture and allowed to react for an additional hour to form adducts with the RNA. Samples were purified, run on a gel and blotted as previously described to assess the extent of labeling at each concentration.

To interrogate thiol-mediated azide reduction, equal amounts of azido-RNA were pre-incubated with thiol concentrations of 0, 10 μ M, 100 μ M, 1 mM, 10 mM, and 100 mM for 30 min at 37 °C. RNA was subsequently ethanol precipitated with 10% 3M sodium acetate and 3x volumes of 100% ethanol at -80 °C for 45 min. Samples were pelleted and washed twice with 70% ethanol, after which they were suspended in nuclease-free water. RNA was then reacted *in vitro* with photo-ODIBO **1** as previously described.

UV spectral analysis of BME competition assay.

Following the determination of the characteristic absorbance of BME, samples of 500 μ M photo-ODIBO or ODIBO **1** were incubated in DMSO with 100 mM BME. Spectra of reaction mixtures were taken after 0, 30 and 60 min had elapsed.

Live cell photo-click SPAAC reaction and RNA isolation.

After nucleoside treatment, cells were incubated with 50 μ M of photo-ODIBO **1** or ODIBO **1a** and pre-incubated for 30 minutes to allow diffusion into the cell. Cells were then irradiated for 2 min and the click reaction was allowed to proceed for 2 h. Cells were washed twice with DPBS after which the RNA was extracted. RNA was then subjected to streptavidin Northern blotting as previously described.

RT-PCR of enriched biotinylated RNA via live cell photo-click SPAAC.

RNA in live cells treated with 2'N₃-A or DMSO was biotinylated via photo-ODIBO **1** (as previously described) for 30 min. Control samples were not incubated with **1** and/or not irradiated. Following extraction, 1 μ g of isolated RNA was treated with DNase (Promega) in 10x DNase reaction buffer for 30 min at 37 °C. Following DNA digestion, total RNA was purified using the Zymo RNA clean and concentrator-5 spin column according to the manufacturer's instructions. Equal amounts of RNA (500 ng) was reverse transcribed using the SuperScript III First-Strand Synthesis kit (Invitrogen) with random hexamer priming via the manufacturer's instructions. After cDNA synthesis, 1/10 of each cDNA reaction was aliquoted, diluted 1:10 and set aside to be used as the input for the polymerase chain reaction. Dynabeads MyOne Streptavidin C1 (Invitrogen) that had been washed x4 times and re-suspended to the original volume with bead binding buffer (100 mM Tris, pH 7.0, 1 M NaCl, 10 mM EDTA, 0.2% (vol/vol) Tween-20) were added to the remaining cDNA reaction, to enrich for biotinylated RNA-cDNA hybrids. The mixture was incubated at room temperature for 20 minutes with end over end rotation. The samples were applied to the DynaMag magnet to retain biotinylated RNA-cDNA hybrid-bound streptavidin beads, and

the flow-through was removed. The bound beads were washed x4 with 4 M wash buffer (100 mM Tris, pH 7.0, 4 M NaCl, 10 mM EDTA, 0.2% (vol/vol) Tween-20) and x2 with nuclease-free water. The beads were then re-suspended in 50 μ L elution mix [1x RNase H buffer, 12.5 mM D-biotin (Life Technologies), 0.1 U/ μ L RNase H (New England Biolabs), RNase A/T1 cocktail at 40 ng/ μ L and 0.1 units/ μ L, respectively (Thermo Fisher Scientific)] for 30 min at 37 °C with shaking. One microliter (1 μ L) of DMSO was added to each RNase-elution reaction and incubated at 95 °C for 5 min. After collecting the elution reaction using the DynaMag, eluted cDNA was subjected to purification using Zymo DNA Clean & Concentration columns according to the manufacturer instructions and eluted in 20 μ L of nuclease free water. Purified cDNA and the corresponding input were subjected to 20 μ L PCR reactions using Dream Taq polymerase (Thermo Fisher Scientific) according to the product manual and either GAPDH primers (Forward 5'-TGTCAGCTCATTTCTGGTAT-3') (Reverse 5'-CTCTCTTCCTCTTGTGC TCTTG-3') or 7SK primers (Forward 5'-CCATTG ATCGCCAGGGTTGA-3') (Reverse 5'- GGAGCGCAGCTACTCGTATAC-3') (Integrated DNA Technologies) for a total of 25 cycles. Half of each completed PCR reaction was loaded onto a 1% TAE agarose gel containing ethidium bromide with 5x nucleic acid loading dye, and the products were separated and then visualized.

RNA fluorescence imaging and selective imaging via photo-click SPAAC.

Cells were seeded at 2.5×10^5 per well and grown to ~50% confluency on sterilized cover glass. Cells were treated 2'-azidoadenosine or DMSO control as previously described. After azide incorporation, cells were treated with compound **2** to a concentration of 50 μ M in 1.1% DMSO and incubated for 2 hours at 37°C in the dark. After labeling, cells were washed

x4 for 5 min with DPBS and fixed for 10 min at room temperature with 4% paraformaldehyde. Cells were then additionally washed with 0.25% Triton-X DPBS x2 for 5 min 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.

For selectively irradiated cells, after treatment with **2**, the cover glass was removed from the well and placed into a humidified 3 cm clear tissue culture dish with 175 μ L of DPBS to keep the cells from dehydrating. A light mask was made from a black rubber sheet with a rectangular window cut to one half the size of a cover glass (22 mm x 11 mm). The light mask was affixed to the lamp and the coverslip in the humidified chamber was aligned and placed over top of the window cell-side up and irradiated for 5 min. The cover glass was removed and a line was drawn on the cell-less side down the center to distinguish the irradiated and non-irradiated sides. Cells were placed into DPBS and allowed to react for only 1 h to prevent lysis from extended incubation in DPBS. The cover glass was then fixed and mounted as previously described. Fluorescent imaging was carried out on a Carl Zeiss Axiovert 200M inverted light microscope on a 10x air objective.

RNA fluorescence imaging with RNase A via photo-click SPAAC.

Cells were grown and treated with azido-analog. Following incubation, cells were washed, permeabilized and lightly fixed with 0.5% paraformaldehyde in 0.5% Triton-X PBS for 45 min. Cells were washed x4 with DPBS and incubated with 400 μ g/mL of RNase A in 0.5% Triton-X PBS at room temperature for 1 hour with light rotation. Cells were then treated with **2**, irradiated and incubated as previously described. Cells were washed x4 with

0.25% Triton-X DPBS for 5 min each, mounted and imaged via fluorescence confocal microscopy.

Population-specific labeling in a co-culture via selective irradiation followed by RT-PCR of enriched biotinylated RNA.

Cells were seeded and grown to ~40% confluency in 60 mm culture dishes. pcDNAs-EGFP plasmid (Addgene) was obtained via transformation of Dh5 α competent *E. coli*, over-night culture inoculation, and plasmid mini-prep (Qaigen). HeLa cells were treated with fresh media and transfected with 750 ng of purified pcDNAs-EGFP plasmid (Polyplus-Transfection). Media was changed after 4 h, and cells were grown for 24 h to allow for EGFP expression. One half of cells on the EGFP-transfected plate were removed using a cell scraper, and fresh WT cells were plated, effectively restricting EGFP expression to one-half of the culture dish. Cells were grown for an additional 24 hours to allow for newly-plated cells to adhere and expand. Plates were imaged through the eye-piece of a Carl Zeiss Axiovert 200M inverted light microscope on a 10x air objective, to see the growth of EGFP-expressing cells restricted to one half of the plate. Very few cells expressing EGFP remained on the re-plated sides, most of which had become detached and could be washed away with DPBS. Cells were treated for 5 h with 2'-azidoadenosine as previously described. After azide incorporation, cells were then treated with photo-ODIBO **1** and pre-incubated for 30 minutes to allow diffusion into the cell. The media was then replaced and the cells were irradiated for 2 min through a mask in pre-determined spots on either half of the plate (irradiation of EGFP cells on plate 1 or irradiation of WT cells on plate 2). The click reaction was allowed to proceed for 30 min, after which the RNA was extracted from each plate

using Trizol Reagent. Precipitated RNA was DNase treated, reverse transcribed and streptavidin enriched as previously described. Purified cDNA and the corresponding input were subjected to 20 μ L PCR reactions using Dream Taq polymerase (Thermo Fisher Scientific) according to the product manual and either GAPDH primers or EGFP primers (Forward 5'-CAAGATCCGCCACAACATCG-3') (Reverse 5'- GACTGGGTGCTCAGGTAGTG-3') (Integrated DNA Technologies) for a total of 25 cycles. Half of each completed PCR reaction was loaded onto an agarose gel, and the products were separated and visualized.

Synthesis

All organic solvents were dried and freshly distilled before use; Tetrahydrofuran was distilled from sodium/benzophenone ketyl and dichloromethane was distilled from CaH₂. Other reagents were obtained from Aldrich or VWR and used as received unless noted. Flash chromatography was performed using 40-63 μ m silica gel. All NMR spectra were recorded in CDCl₃ using 400 MHz instrument (unless otherwise noted). Electronic spectra and scanning kinetics were recorded using a Cary 60 UV. Rate measurements were performed using Carry-300 Bio UV-Vis spectrometer. Photolysis were conducted using a Rayonet photoreactor equipped with sixteen 4W 350 nm fluorescent lamps.

3-(d-Biotinylamido)-1-propyne.²⁴

1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC·HCl) (0.408 mg, 2.12 mmol) was added to a solution of D-biotin (0.400 g, 1.63 mmol) and propargylamine (0.108 g, 1.96 mmol) in CH₃CN:MeOH (3:1, 16 mL). The reaction was stirred at room temperature

for 6.5 hours, and concentrated in vacuo. Then, MeOH was added to the crude residue and the solution was filtered through celite. The filtrate was concentrated in vacuo and purified via flash chromatography (10:1, CHCl₃:MeOH) to afford 3-(D-Biotinylamido)-1-propyne (0.386 g, 84% yield) as a white solid. ¹H-NMR (DMSO-d₆): 8.22-8.25 (t, *J* = 5.4 Hz, 1H), 6.44 (s, 1H), 6.37 (s, 1H), 4.14-4.31 (m, 1H), 4.10-4.14 (m, 1H), 3.82-3.84 (m, 2H), 3.07-3.11 (m, 2H), 2.79-2.84 (dd, *J* = 12.4, 5.1 Hz, 1H), 2.55-2.58 (d, *J* = 12.4 Hz, 1H), 2.05-2.09 (t, *J* = 7.4 Hz, 2H), 1.40-1.65 (m, 4H), 1.24-1.35 (m, 2H). ¹³C-NMR (DMSO-d₆): 171.76, 162.67, 81.34, 72.78, 61.01, 59.16, 55.37, 39.84, 34.84, 28.16, 28.00, 27.69, 25.11.

3-(*tert*-butyl)-9-(2-(2-(2-(2-(azidoethoxy)ethoxy)ethoxy)ethoxy)dibenzo[b,f]cyclopropa [d] oxocin-1(7*H*)-one (photo-ODIBO-EG₄-Azide).

K₂CO₃ (0.866 g, 6.27 mmol) was added to a solution of photo-ODIBO-OH^{22,25} (1.920 g, 6.27 mmol) and a Tos-EG₄-N₃ linker⁴ (3.51 g, 9.41 mmol) in DMF (21 mL). The solution was stirred and heated to 80 °C for 5 hours. The reaction mixture was diluted with ethyl acetate (200 mL), washed with water (5x50 mL), brine (100 mL), and dried over MgSO₄. The organic layer was then filtered, concentrated in vacuum, and purified via flash chromatography (20:1 DCM: MeOH) to afford photo-ODIBO-TEG-Azide (2.405 g, 76% yield) of a slightly yellow oil. ¹H-NMR: 7.91-7.93 (m, 2H), 7.47-7.50 (dd, *J* = 8.5, 2.5 Hz, 1H), 7.18-7.20 (d, *J* = 8.5 Hz, 1H), 7.01-7.05 (m, 2H), 5.23-5.26 (d, *J* = 12.2 Hz, 1H), 4.74-4.77 (d, *J* = 12.1 Hz, 1H), 4.20-4.23 (t, *J* = 4.7 Hz, 2H), 3.87-3.89 (t, *J* = 4.6 Hz, 2H), 3.72-3.73 (m, 2H), 3.64-3.71 (m, 8H), 3.35-3.37 (t, *J* = 5.0 Hz, 2H), 1.33 (s, 9H). ¹³C-NMR: 162.04, 160.44, 152.74, 148.01, 144.01, 142.23, 140.69, 135.50, 131.10, 130.73, 122.05, 117.87, 117.41, 117.05, 114.66,

78.78, 71.00, 70.81, 70.79, 70.76, 70.14, 69.53, 67.99, 50.75, 34.66, 31.44. ESI HRMS: calcd. (M+H⁺): C₂₈H₃₄N₃O₆ 508.2442, found 508.2440. IR: 1845 cm⁻¹ (ν_{C=O}).

Photo-ODIBO-TEG₄-Biotin (1)

CuI (0.014 g, 0.072 mmol) was added to a solution of photo-ODIBO-TEG-N₃ (0.367 g, 0.723 mmol), 3-(D-Biotinylamido)-1-propyne (0.305 g, 1.08 mmol), and triethylamine (0.146 g, 1.44 mmol) in DMF (8 mL). The solution was stirred overnight, diluted with ethyl acetate (100 mL), washed with a saturated solution of NH₄Cl (2x50 mL), brine (50 mL), and dried over MgSO₄. The crude solution was then concentrated in vacuum and purified via flash chromatography (10:1 to 10:2 DCM: MeOH) to afford photo-ODIBO-EG₄-Biotin (0.289 g, 51% yield) as an off-white solid (83-86 °C MP). ¹H-NMR: 7.90-7.94 (m, 3H), 7.75 (s, 1H), 7.49-7.52 (dd, *J* = 8.5, 25 Hz, 1H), 7.24 (s, 1H), 7.19-7.22 (d, *J* = 8.5 Hz, 1H), 7.03-7.06 (m, 2H), 6.67 (s, 1H), 5.25-5.29 (d, *J* = 12.1 Hz, 1H), 4.77-4.80 (d, *J* = 12.1 Hz, 1H), 4.53-4.58 (m, 2 H), 4.47-4.50 (t, *J* = 5.1 Hz, 2H), 4.31-4.37 (m, 2H), 4.22-4.25 (t, *J* = 4.6 Hz, 2H), 3.88-3.90 (t, *J* = 4.6 Hz, 2H), 3.83-3.85 (t, *J* = 5.1 Hz, 2H), 3.72-3.74 (m, 2H), 3.65-3.66 (m, 2H), 3.61 (s, 4H), 3.08-3.13 (m, 1H), 2.88-2.93 (m, 1H), 2.75-2.78 (d, *J* = 12.7 Hz, 1H), 2.14-2.27 (m, 2H), 1.62-1.81 (m, 4H), 1.40-1.47 (m, 2H), 1.35 (s, 9H). ¹³C-NMR: 173.40, 164.79, 162.11, 160.53, 152.82, 148.15, 145.23, 144.12, 142.40, 140.81, 135.61, 131.21, 130.83, 123.68, 122.13, 118.01, 117.51, 117.16, 114.78, 78.86, 71.07, 70.80, 70.77, 70.73, 69.65, 69.54, 68.10, 31.83, 60.45, 55.94, 50.44, 40.90, 35.98, 34.76, 34.42, 31.53, 28.27, 28.11, 25.56. ESI HRMS: calcd. (M+H⁺): C₄₁H₅₃N₆O₈S⁺ 789.3640, found 789.4652. IR: 1844 cm⁻¹ (ν_{C=O}).

ODIBO-TEG₄-Biotin (1a)

Photo-ODIBO-Biotin (0.189 g, 0.240 mmol) was dissolved into MeOH (200 mL) and irradiated for 20 minutes at 350 nm. The solution was concentrated in vacuum and purified by flash chromatography (10:1 DCM: MeOH) to afford ODIBO-EG₄-Biotin (0.088 g, 48% yield) as a white amorphous solid. ¹H-NMR: 7.83-7.86 (m, 1H), 7.74 (s, 1H), 7.20-7.25 (m, 3H), (7.17, s, 1H), 7.09-7.11 (m, 1H), 7.01-7.02 (d, *J* = 2.2 Hz, 1H), 6.88-6.91 (dd, *J* = 8.4, 2.2 Hz, 1H), 6.55 (s, 1H), 5.16-5.19 (d, *J* = 12.0 Hz, 1H), 4.47-4.56 (m, 5H), 4.30-4.37 (m, 2H), 4.15-4.17 (t, *J* = 4.6 Hz, 2H), 3.83-3.88 (m, 4H), 3.71-3.73 (m, 2H), 3.63-3.66 (m, 2H), 3.61 (s, 4H), 3.08-3.12 (m, 1H), 2.87-2.92 (m, 1H), 2.74-2.77 (m, 1H), 2.17-2.25 (m, 2H), 1.73-1.79 (m, 1H), 1.62-1.69 (m, 3H), 1.42-1.49 (m, 2H), 1.31 (s, 9H). ¹³C-NMR: 173.43, 167.29, 164.79, 158.74, 149.13, 147.01, 145.19, 126.92, 125.63, 123.73, 123.65, 121.40, 118.35, 117.90, 117.65, 114.64, 114.17, 110.78, 78.04, 71.04, 70.81, 70.78, 70.73, 69.81, 69.56, 67.90, 61.83, 60.47, 55.88, 50.43, 40.84, 35.98, 34.58, 34.47, 31.58, 28.28, 28.14, 25.58. ESI HRMS: calcd. (M+H⁺): C₄₀H₅₃N₆O₇S⁺ 761.3691, found 761.3706.

3-(*tert*-butyl)-9-((2,2,3,3-tetramethyl-4,7,10,13-tetraoxa-3-silapentadecan-15-yl)oxy)dibenzo[b,f]cyclopropa[d]oxocin-1(7H)-one (photo-ODIBO-EG₄-TBS).

To a solution of photo-ODIBO-OH²⁻³ (1.10 g, 3.59 mmol) in DMF (20 mL) was added TBDMS protected monotosyl tetra(ethyleneglycol)⁵ (2.49 g, 5.39 mmol). Next, K₂CO₃ (0.49 g, 3.59 mmol) was added and the solution was stirred and heated to 80 °C for 5 hours. The reaction mixture was then diluted with ethyl acetate (200 mL), washed with water (5 x 50

ml), brine (100 mL), and dried over MgSO₄. The organic layer was then filtered, concentrated in vacuo, and purified via flash chromatography (3:1 ethyl acetate: hexanes) to afford photo-ODIBO-EG₄-TBS (1.24 g, 58% yield) of a colorless oil. ¹H-NMR: 7.92-7.94 (m, 2H), 7.48-7.51 (dd, *J* = 8.5, 2.5 Hz, 1H), 7.19-7.21 (d, *J* = 8.5 Hz, 1H), 7.02-7.06 (m, 2H), 5.24-5.27 (d, *J* = 12.1 Hz, 1H), 4.76-4.79 (d, *J* = 12.1 Hz, 1H), 4.21-4.23 (t, *J* = 4.7 Hz, 2H), 3.88-3.90 (t, *J* = 4.7 Hz, 2H), 3.72-3.77 (m, 4H), 3.65-3.69 (m, 6H), 3.53-3.56 (t, *J* = 5.4 Hz, 2H), 1.34 (s, 9H), 0.87 (s, 9H), 0.05 (s, 6H). ¹³C-NMR: 162.10, 160.50, 152.79, 148.07, 144.10, 142.32, 140.73, 135.56, 131.17, 130.75, 122.09, 117.96, 117.48, 117.13, 114.72, 78.83, 72.82, 71.09, 70.89, 70.87, 70.81, 69.60, 68.06, 62.87, 34.71, 31.49, 26.10, 18.53, -5.08. ESI HRMS: calcd. (M+H⁺): C₃₄H₄₉O₇Si⁺ 597.3242, found 597.3241. IR: 1846 cm⁻¹ (ν_{C=O}).

3-(*tert*-butyl)-9-(2-(2-(2-(2-(hydroxyethoxy)ethoxy)ethoxy)ethoxy)dibenzo[b,f]cyclopropa [d]oxocin-1(7H)-one (photo-ODIBO-EG₄-OH).

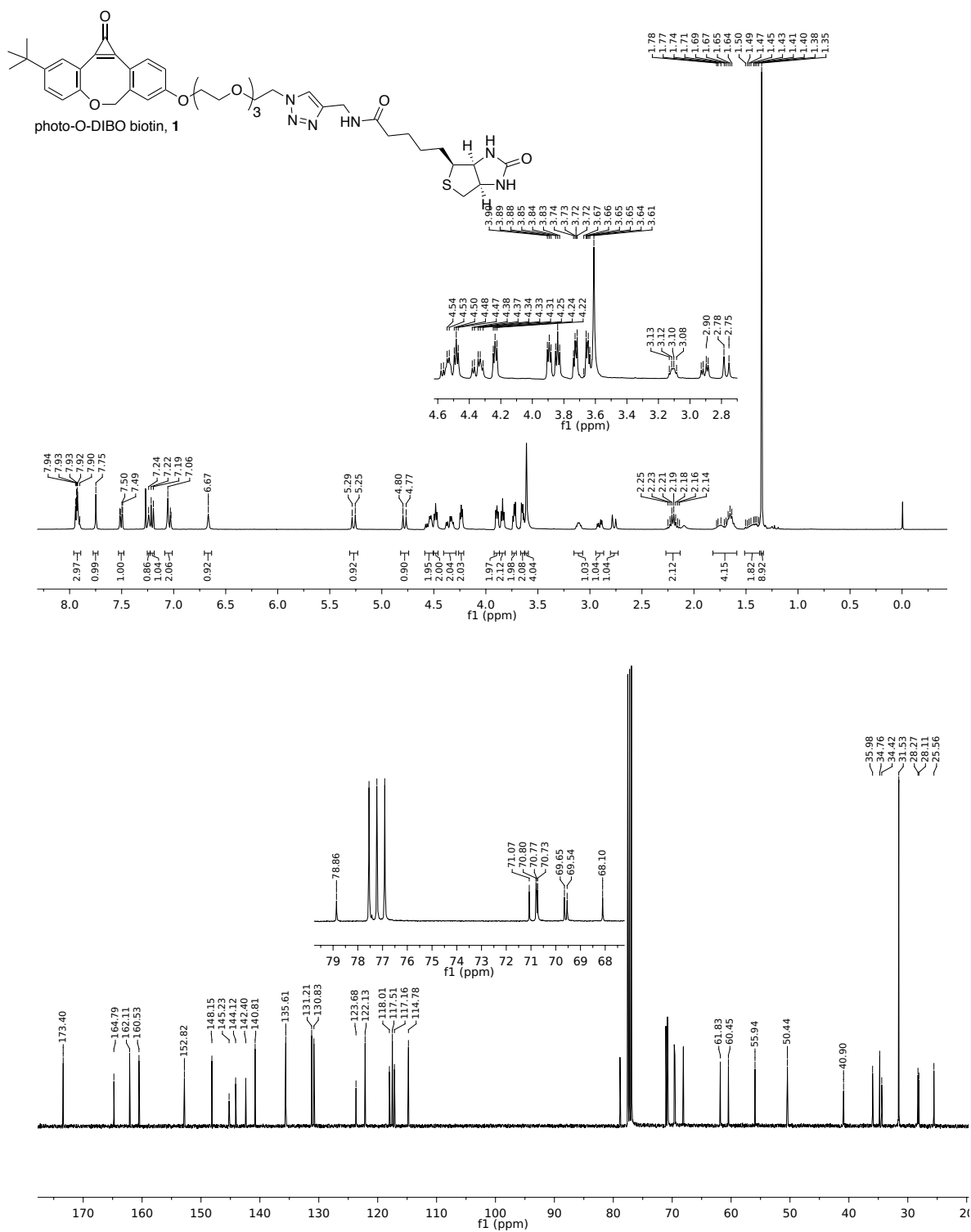
HF (0.123 g, 3.07 mmol, 50%) was added to a solution of photo-ODIBO-EG₄-TBS (1.220 g, 2.04 mmol) in acetonitrile (11 mL). The solution was stirred for 30 minutes, diluted with ethyl acetate (300 mL) and water (100 mL). The organic layer was extracted, dried over MgSO₄, filtered, concentrated in vacuo, and purified via flash chromatography (40:1 CH₂Cl₂: methanol) to afford photo-ODIBO-TEG-OH (0.835 g, 85 % yield) as a colorless oil. ¹H-NMR: 7.92-7.94 (m, 2H), 7.75-7.50 (dd, *J* = 8.5, 2.5 Hz, 1H), 7.18-7.20 (d, *J* = 8.4 Hz, 1H), 7.02-7.05 (m, 2H), 5.24-5.27 (d, *J* = 12.2 Hz, 1H), 4.74-4.77 (d, *J* = 12.1 Hz, 1H), 4.21-4.23 (t, *J* = 4.7 Hz, 2H), 3.86-3.89 (t, *J* = 4.7 Hz, 2H), 3.65-3.73 (m, 10H), 3.58-3.60 (t, *J* = 5.4 Hz, 2H), 2.91 (s, 1H),

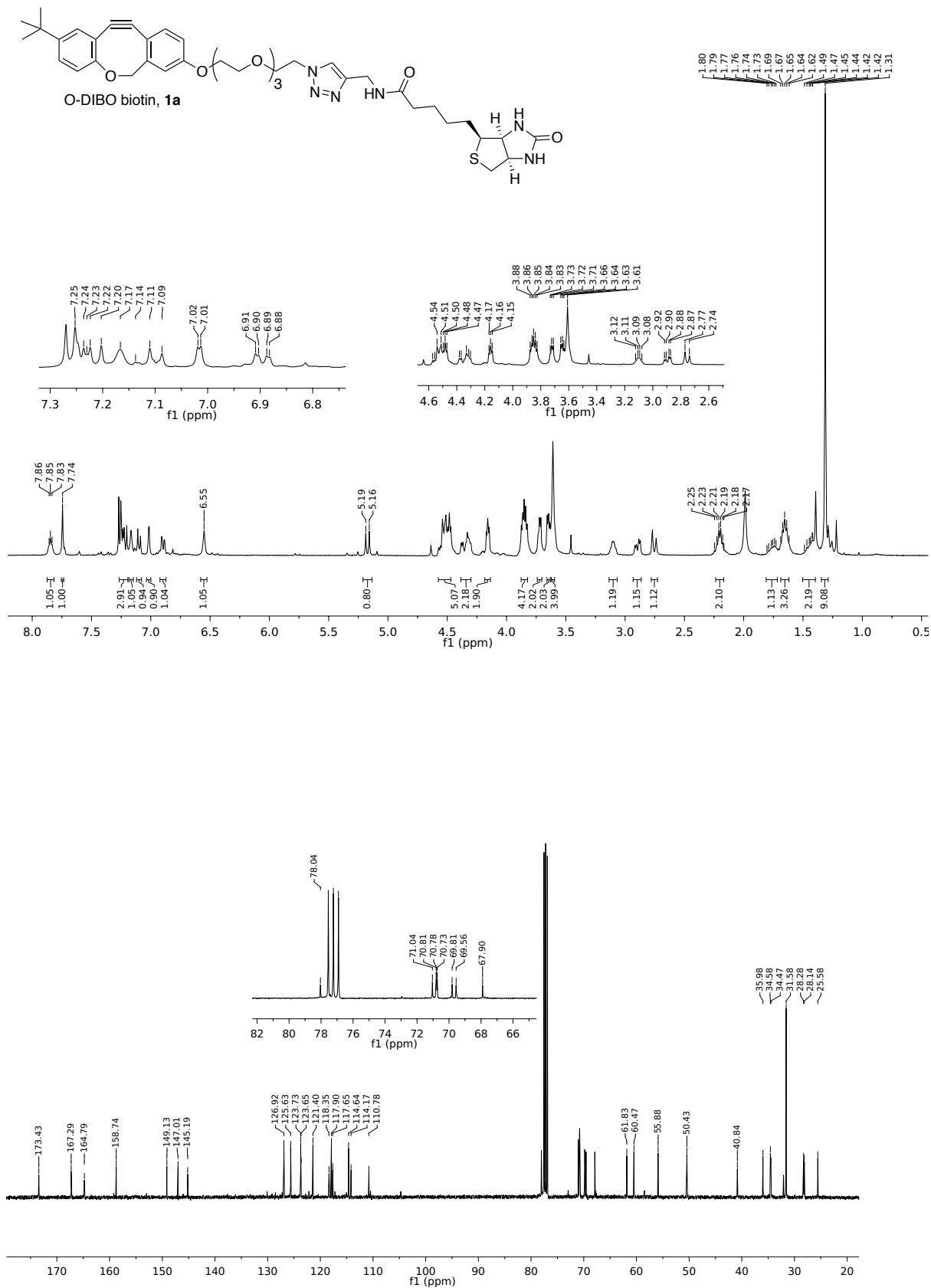
1.33 (s, 9H). ^{13}C -NMR: 162.05, 160.48, 152.81, 148.05, 144.02, 142.23, 140.72, 135.56, 131.14, 130.46, 122.07, 117.91, 117.48, 117.07, 114.72, 78.80, 72.67, 70.98, 70.77, 70.68, 70.34, 69.57, 67.99, 61.81, 34.69, 31.46. $\text{C}_{28}\text{H}_{35}\text{O}_7^+$ 483.2377, found 483.2388. IR: 1846 cm^{-1} ($\nu_{\text{C=O}}$).

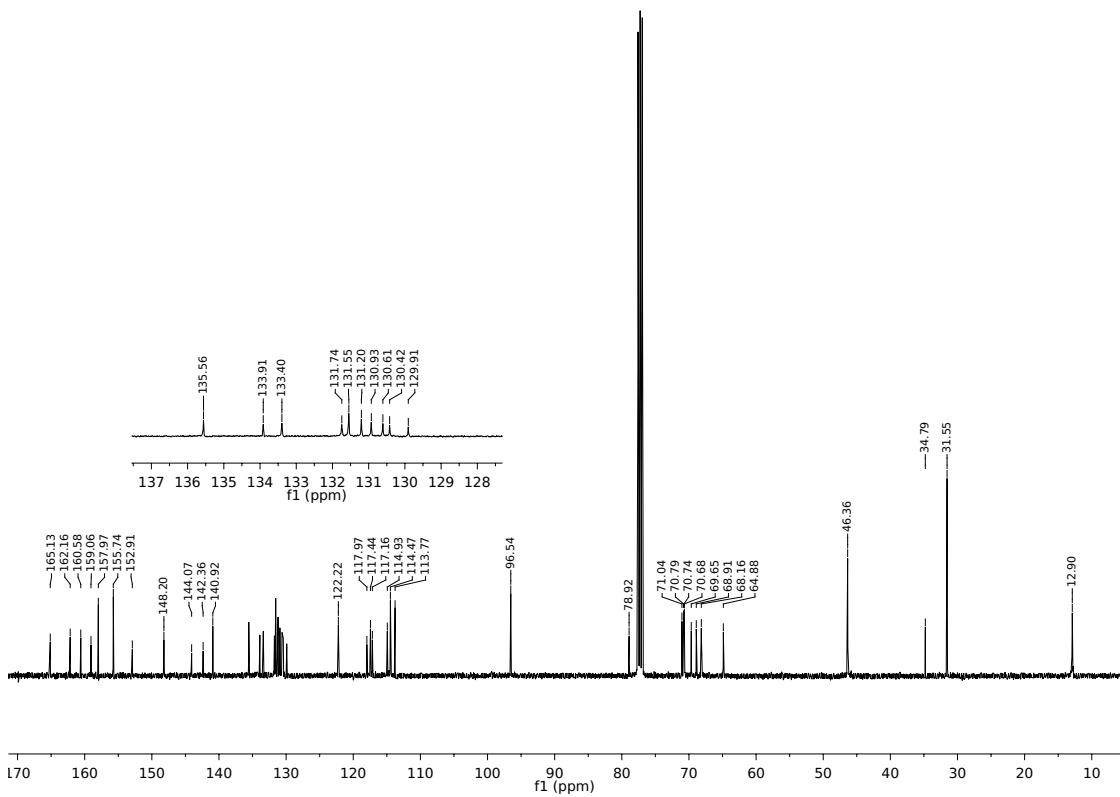
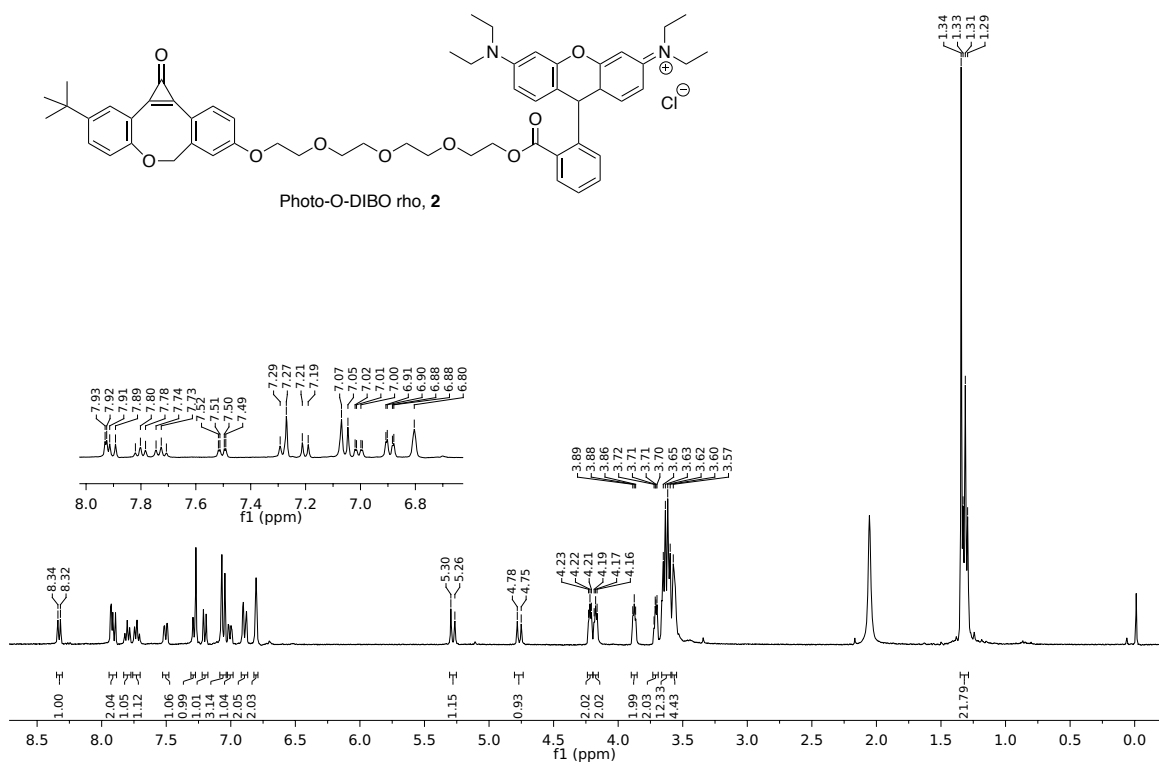
Photo-ODIBO-EG₄-Rhodamine B (2)

DCC (0.141 g, 0.684 mmol) was added to a solution of photo-ODIBO-EG₄-OH (0.330 g, 0.684 mmol), Rhodamine B (0.329 g, 0.684 mmol), and DMAP (CA) in DCM (7 mL). The reaction was stirred overnight, filtered, concentrated in vacuum, and purified via flash chromatography (20:1 to 15:1, DCM: MeOH) to afford photo-ODIBO-EG₄-Rhoadmine B (0.149 g, 24% yield) as a purple solid (131-132 °C MP). ^1H -NMR: 8.32-8.34 (d, J = 7.8 Hz, 1H), 7.89-7.93 (m, 2H), 7.78-7.82 (t, J = 7.4 Hz, 3H), 7.71-7.74 (t, J = 7.4 Hz, 1H), 7.49-7.52 (dd, J = 8.6, 2.2 Hz, 1H), 7.27-7.29 (m, 1H), 7.19-7.21 (d, J = 8.6 Hz, 1H), 7.05-7.07 (m, 3H), 6.99-7.02 (dd, J = 8.5, 2.1 Hz, 1H), 6.88-6.91 (m, 2H), 6.80 (m, 2H), 5.26-5.29 (d, J = 12.2 Hz, 1H), 4.75-4.78 (d, J = 12.2 Hz, 1H), 4.21-4.23 (t, J = 4.5 Hz, 2H), 4.16-4.19 (t, J = 4.5 Hz, 2H), 3.86-3.89 (t, J = 4.4 Hz, 2H), 3.70-3.72 (m, 2H), 3.60-3.65 (m, 12H), 3.56-3.57 (m, 4H), 1.29-1.34 (m, 21H). ^{13}C -NMR: 165.13, 162.16, 160.58, 159.06, 157.97, 155.74, 152.91, 148.20, 144.07, 142.36, 140.92, 135.56, 133.91, 133.40, 131.74, 131.55, 131.20, 130.93, 130.61, 130.42, 129.91, 122.22, 117.97, 117.44, 117.16, 114.93, 114.47, 113.77, 96.54, 78.92, 71.04, 70.79, 70.74, 70.68, 69.65, 68.91, 68.16, 64.88, 46.36, 34.79, 31.55, 12.90. ESI HRMS: calcd. ($\text{M}+\text{H}^+$): $\text{C}_{56}\text{H}_{63}\text{N}_2\text{O}_9^+$ 907.4528, found 907.4525. IR: 1843 cm^{-1} ($\nu_{\text{C=O}}$).

Spectra







3.6 Supplementary Data

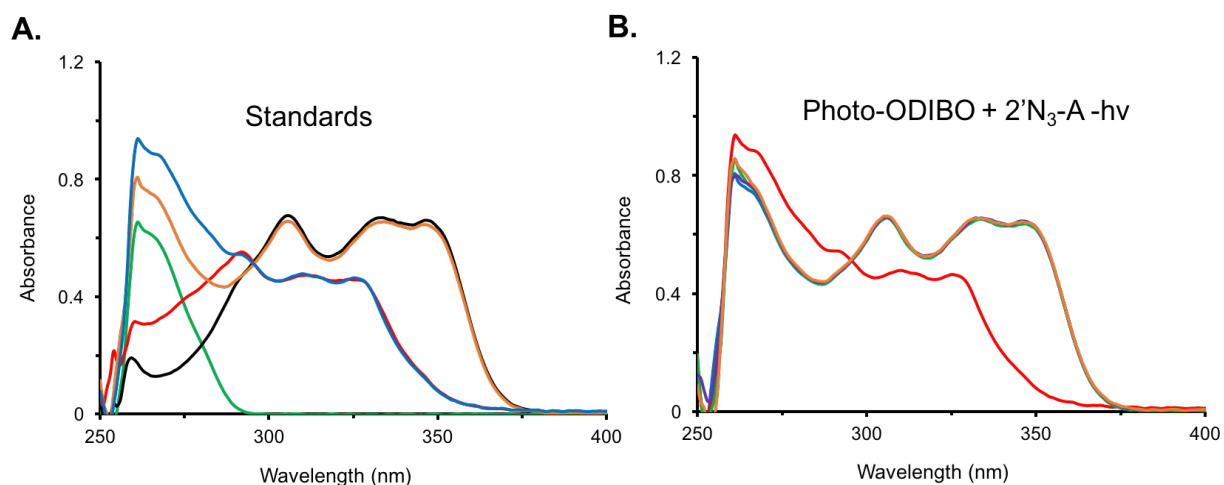


Figure 3-S1. UV-vis spectra of photo-click reaction components. A. Spectra of 500 μM photo-ODIBO **1** (black) showing absorbance at 350 nm, while ODIBO (red) is shifted to 320 nm. The nucleoside analog 2'N₃-A showed an absorbance peak around 260 nm (green), which is characteristic of nucleic acids. The mixture of **1** or **1a** with 2'N₃-A displays a spectral combination of the two compounds (orange and blue, respectively). **B.** UV spectra of 1:1 500 μM solution of **1** and 2'N₃-A time titration (0 min, 15 min, 30 min, 60 min) without irradiation, displaying no absorbance shifts (blue, green, purple and orange). A control spectra of the irradiated reaction at 0 min was included to present the clear difference in absorbances resulting from light de-caging.

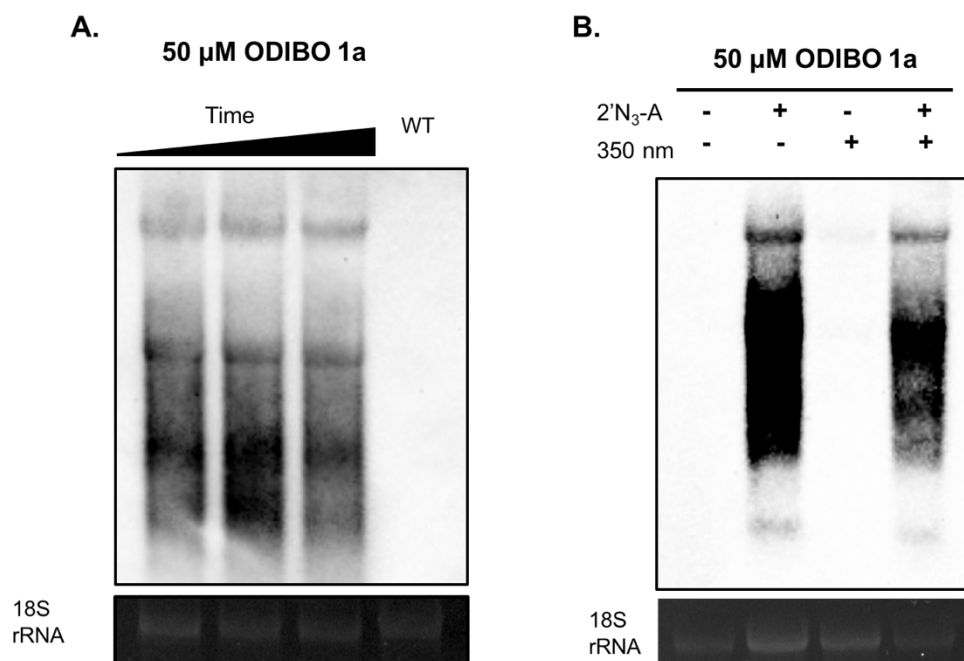


Figure 3-S2. ODIBO, 1a tagging of azido-RNA. **A.** Streptavidin Northern blot *in vitro* time titration analysis of RNA- N_3 treated with 50 μ M ODIBO **1a**, and reacted at 37 °C for 15, 30 and 60 min. Wild-type (WT) control RNA was extracted from HeLa cells treated with DMSO. **B.** Labeling with ODIBO **1a** is independent of light irradiation. WT RNA incubated with **1a** and exposed to light shows no signal, while RNA- N_3 treated with **1a** and exposed to light shows no detectable signal enhancement.

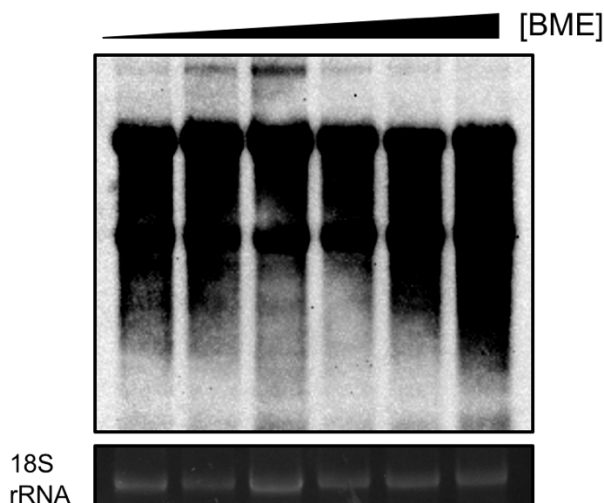


Figure 3-S3. Thiols do not affect azide integrity. Pre-incubation of azido-RNA with increasing thiol to determine the contribution of thiol-mediated azide reduction. Streptavidin Northern blot of equal amounts of azido-RNA pre-incubated with a titration of BME (0, 10 μ M, 100 μ M, 1 mM, 10 mM, 100 mM) for 30 min at 37 $^{\circ}$ C, and then reacted with 50 μ M photo-ODIBO **1** for 1 h. No change in the signal was detected indicating azide-stability when incorporated into cellular RNA.

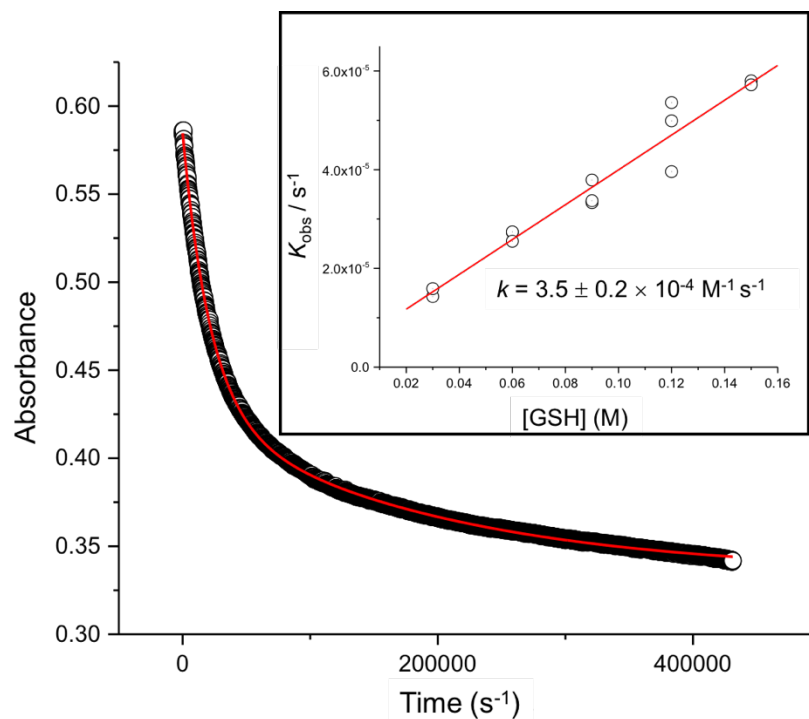


Figure 3-S4. Second order rate constant determination of GSH with ODIBO, 1a. ODIBO reacts with glutathione, but reaction kinetics shows strange multicomponent features and thus, TCEP was added as reducing/radical quenching agent to prevent the formation of GSH oxidation product(s) and stop radical side-reactions. ODIBO does not react with TCEP up to 0.5 M concentrations. The decay of 50 μ M ODIBO **1a** was measured in 5% MeOH PBS in the presence of 100 mM GSH. Scanning kinetics graph shows the initial product of the reaction of ODIBO with GSH undergoes subsequent transformation. Kinetic trace was fit to a double exponential equation.

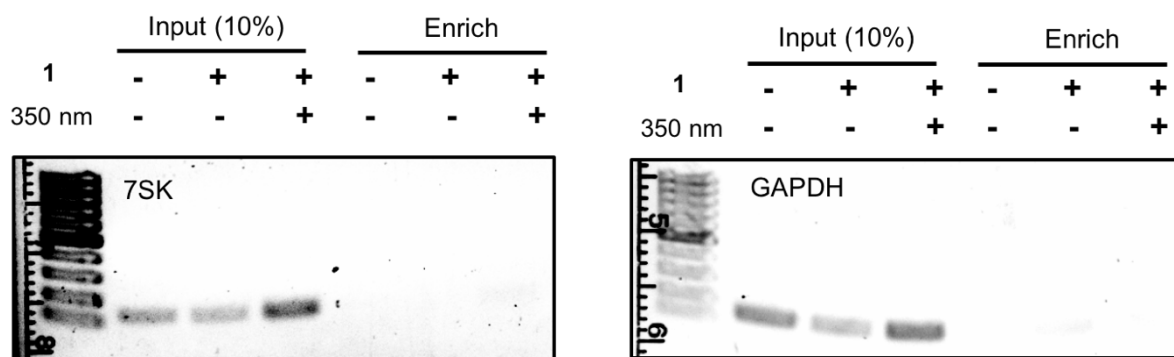


Figure 3-S5. RT-PCR of DMSO-treated control RNA. from cells labeled with photo-ODIBO **1** for 30 min. PCR of cDNA did not enrich in any of the conditions with or without light and photo-ODIBO **1**, thus there was no recovery and synthesis of the small non-coding RNA 7SK and the mRNA GAPDH. Input was aliquoted from cDNA reaction prior to enrichment to serve as a loading control for the PCR reaction.

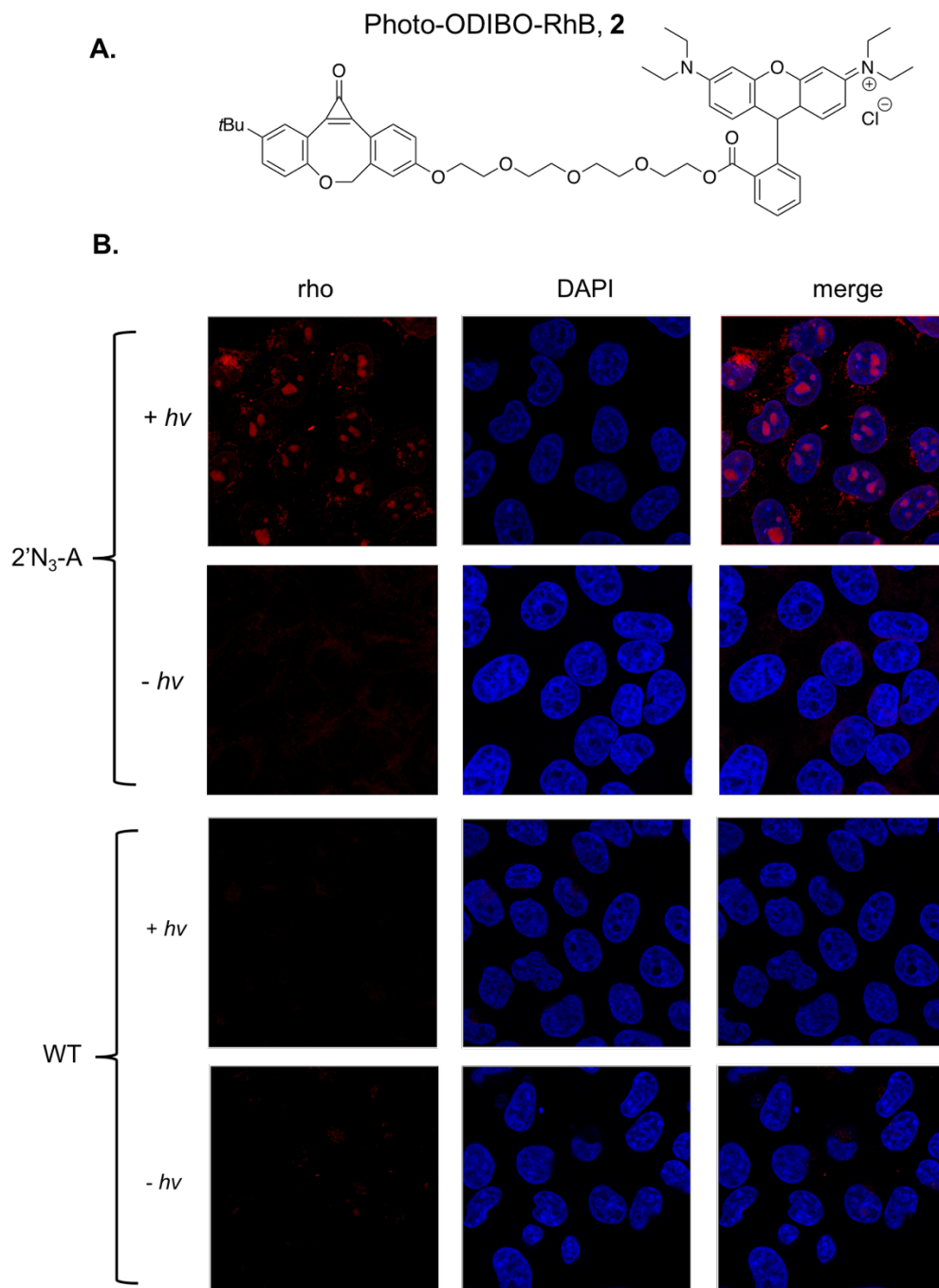


Figure 3-S6. Photo-click imaging of azido-RNA. A. Structure of photo-ODIBO rhodamine B, **2**. **B.** Photo-click SPAAC fluorescent imaging of cellular azido-RNA and DMSO-treated WT RNA. RNA labeled with 2'N₃A can be visualized after incubation with photo-ODIBO rho **2** for 2 h. Labeling was light dependent, as cells that were irradiated displayed the typical

staining pattern in the nucleoli and cytosol from 2' N_3 -A. There was no notable signal from the nucleolus and cytosol in cells not irradiated. Regardless of light irradiation, there was no notable signal from the nucleolus and cytosol in cells treated with DMSO. RNA is displayed in the rhodamine channel, while the nucleus was counterstained with DAPI.

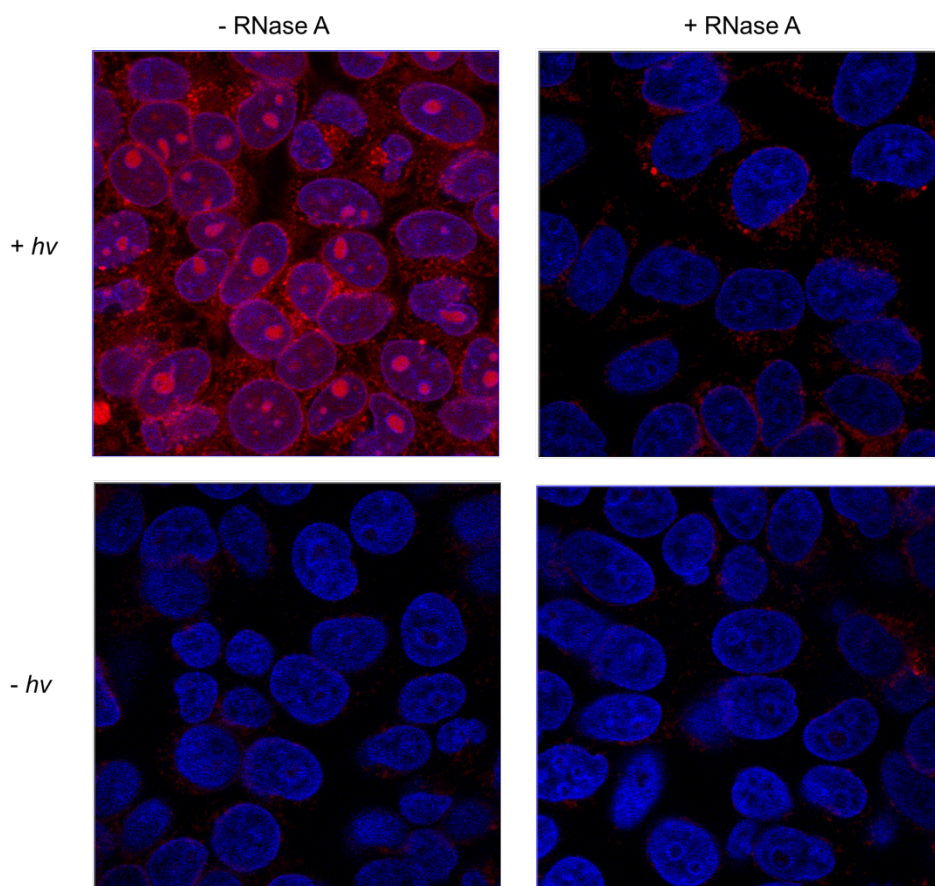


Figure 3-S7. Imaging of 2' N_3 -A via photo-ODIBO-rhB **2 is sensitive to RNase A.** Cells were grown in the presence of 2' N_3 -A, and subsequently permeabilized and lightly fixed before treatment with or without RNase A. Cells were then incubated with **2**, wherein labeling in irradiated cells treated with RNase A was significantly reduced. Non-irradiated

cells displayed no staining, regardless of treatment with RNase A. RNA is visualized via the rhodamine channel, while the nucleus was counterstained with DAPI.

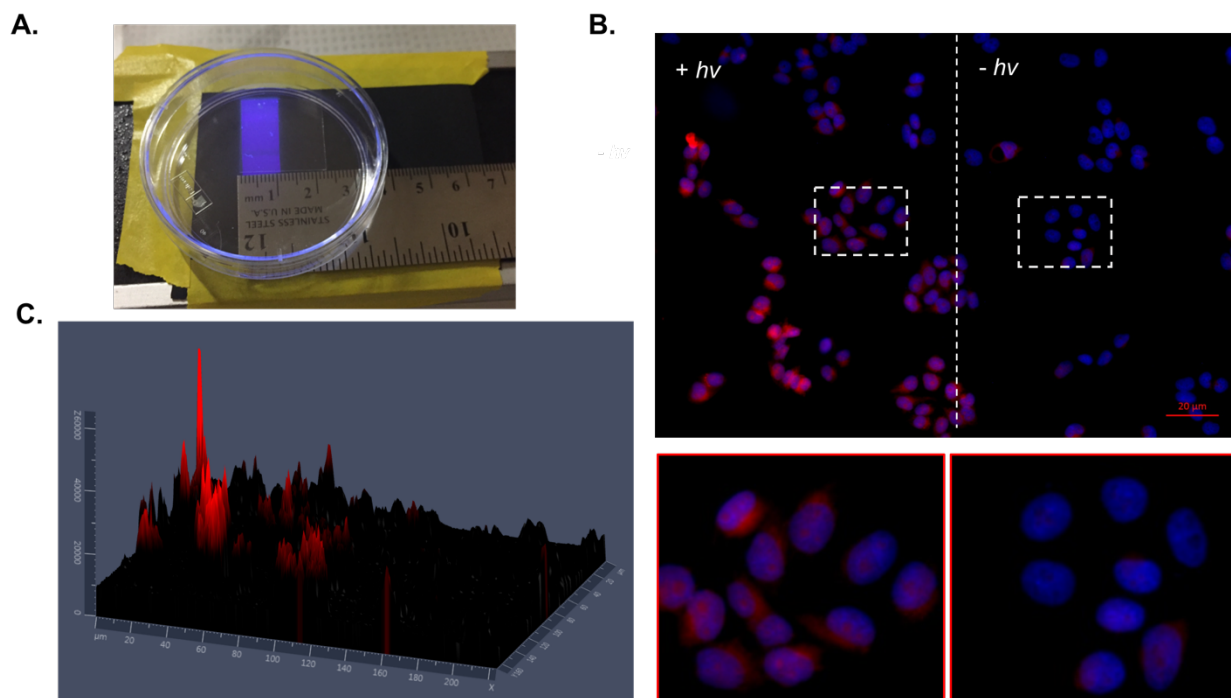


Figure 3-S8. Selective irradiation and photo-click imaging. **A.** Experimental setup for selective irradiation and imaging. Cells grown on cover glass were irradiated via a rectangular window cut to one half the size of the glass. **B.** Fluorescent imaging of cells grown in the presence of 2' N_3 -A, and subsequently treated with 50 μ M of **2**. Selective irradiation (2 min) through a mask exposing one half of the cover glass yielded labeled and non-labeled RNA. **C)** Selective imaging layered-depth display. Fluorescent intensity visualized via depth display shows the difference in irradiated (left) and non-irradiated (right) sides of the cover glass.

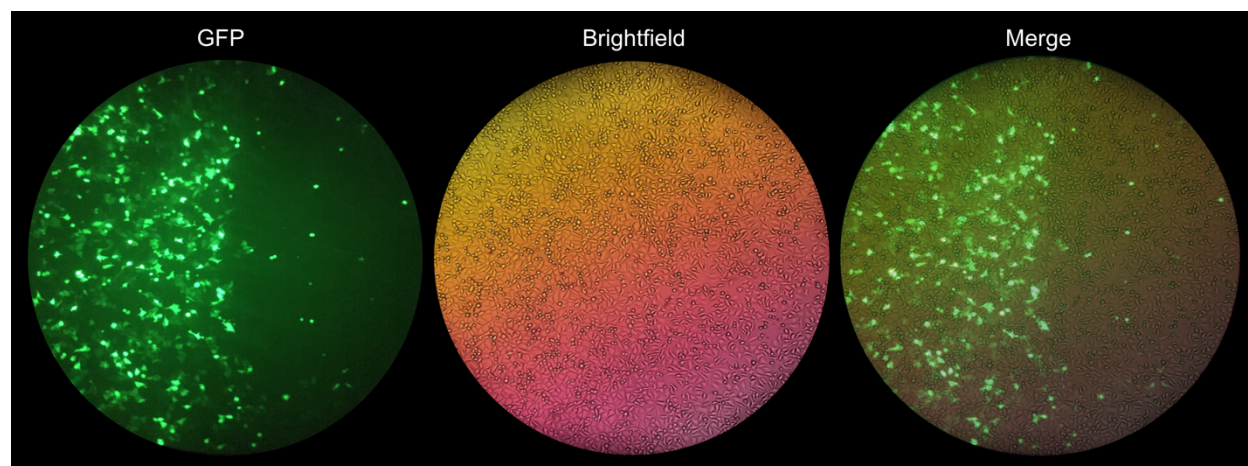


Figure 3-S9. Fluorescent imaging of cell co-culture. EGFP-expressing cells are restricted to one half of the culture dish. Cells were transfected with EGFP plasmid for 24 h, after which one half of the plate was removed and re-plated with wild-type cells.

3.7 References

1. Morris, K. V.; Mattick, J. S. *Nat. Rev. Genet.* **2014**, *15*, 423-437.
2. Prescher, J. A.; Bertozzi, C. R. *Nat. Chem. Biol.* **2005**, *1*, 13-21.
3. Rabuka, D.; Hubbard, S. C.; Laughlin, S. T.; Argade, S. P.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2006**, *128*, 12078-12079.
4. Yao, J. Z.; Uttamapinant, C.; Poloukhine, A.; Baskin, J. M.; Codelli, J. A.; Sletten, E. M.; Bertozzi, C. R.; Popik, V. V.; Ting, A. Y. *J. Am. Chem. Soc.* **2012**, *134*, 3720-3728.
5. Salic, A.; Mitchison, T. J. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 2415-2420.
6. Patterson, D. M.; Nazarova, L. A.; Prescher, J. A. *ACS Chem. Biol.* **2014**, *9*, 592-605.

7. Grammel, M.; Hang, H.; Conrad, N. K. *ChemBioChem* **2012**, *13*, 1112-1115.
8. Chang, P. V.; Chen, X.; Smyrniotis, C.; Xenakis, A.; Hu, T.; Bertozzi, C. R.; Wu, P. *Angew. Chem., Int. Ed.* **2009**, *48*, 4030-4033.
9. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004**, *126*, 15046-15047.
10. Nainar, S.; Beasley, S.; Fazio, M.; Kubota, M.; Dai, N.; Correa, I. R., Jr.; Spitale, R. C. *ChemBioChem* **2016**, *17*, 2149-2152.
11. Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16793-16797.
12. Poloukhine, A.; Popik, V. V. *J. Org. Chem.* **2003**, *68*, 7833-7840.
13. Martinek, M.; Filipova, L.; Galeta, J.; Ludvikova, L.; Klan, P. *Org. Lett.* **2016**, *18*, 4892-4895.
14. Gordon, C. G.; Mackey, J. L.; Jewett, J. C.; Sletten, E. M.; Houk, K. N.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2012**, *134*, 9199-9208.
15. Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2010**, *132*, 3688-3690.
16. Friscourt, F.; Ledin, P. A.; Mbua, N. E.; Flanagan-Steet, H. R.; Wolfert, M. A.; Steet, R.; Boons, G. J. *J. Am. Chem. Soc.* **2012**, *134*, 5381-5389.
17. Sutton, D. A.; Yu, S. H.; Steet, R.; Popik, V. V. *Chem. Commun.* **2016**, *52*, 553-556.
18. Poloukhine, A. A.; Mbua, N. E.; Wolfert, M. A.; Boons, G.-J.; Popik, V. V. *J. Am. Chem. Soc.* **2009**, *131*, 15769-15776.
19. van Geel, R.; Pruijn, G. J.; van Delft, F. L.; Boelens, W. C. *Bioconjugate Chem.* **2012**, *23*, 392-398.
20. Hansen, R. E.; Roth, D.; Winther, J. R. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 422-427.

21. Tian, M.; Guo, F.; Sun, Y.; Zhang, W.; Miao, F.; Liu, Y.; Song, G.; Ho, C. L.; Yu, X.; Sun, J. Z.; Wong, W. Y. *Org. Biomol. Chem.* **2014**, *12*, 6128-6133.
22. McNitt, C. D.; Popik, V. V. *Org. Biomol. Chem.* **2012**, *10*, 8200-8202.
23. Luo, W.; Gobbo, P.; McNitt, C. D.; Sutton, D. A.; Popik, V. V.; Workentin, M. S. *Chem. Eur. J.* **2017**, *23*, 1052-1059.
24. Aoki, S.; Matsuo, N.; Hanaya, K.; Yamada, Y.; Kageyama, Y. *Bioorg. Med. Chem.* **2009**, *17*, 3405-3413.
25. Brooks, K.; Yatvin, J.; McNitt, C. D.; Reese, R. A.; Jung, C.; Popik, V. V.; Locklin, J. *Langmuir* **2016**, *32*, 6600-6605.

Chapter 4: Expanding the Scope of RNA Metabolic Labeling with Vinyl Nucleosides and Inverse Electron-Demand Diels-Alder Chemistry

4.1 Introduction

RNA plays critical roles in the regulation of nearly all biological processes: from controlling chromatin state, to directing transcription, to regulation of cell proliferation.^{1,2} Thus, profiling nascent RNA expression can allow researchers to identify potentially functional RNAs, and this is aided by the availability of metabolic chemical reporters.³ As such, several chemical functionalities have been adapted for RNA nucleoside analogs. 4-thiouridine is a conventionally used probe, which allows for enrichment of nascent RNA through disulfide exchange chemistry.⁴ More recently, alkynyl-⁵⁻⁷ and azido-containing⁸ nucleosides characterized by our lab have been introduced into cells, wherein nascent RNA transcripts can be tracked through click chemistry. Both the Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) and strain-promoted azide-alkyne cycloaddition (SPAAC) can be utilized for enrichment as well as imaging.⁹

Despite the utility of these analogs, their chemistries present several drawbacks. 4-thiouridine enrichment is often inefficient due to the transient nature of disulfide bonds, which further impedes RNA imaging in cells.¹⁰ Likewise, CuAAC reactions are known to dramatically compromise the integrity of RNA, where spurious fragmentation of transcripts has altered the results of downstream applications such as RNA sequencing.^{11,12} Methods that are damaging to cells or RNA integrity can confound interpretation of biologically meaningful results.^{13,14} Lastly, discovering novel analogs for labeling is further hindered by

the selectivity of metabolic enzymes, which often have limited functional group tolerance.^{8,15} These challenges demonstrate the need to further expand the scope of different bioorthogonal chemistries applicable for RNA metabolic labeling (**Figure 4-1, A**).

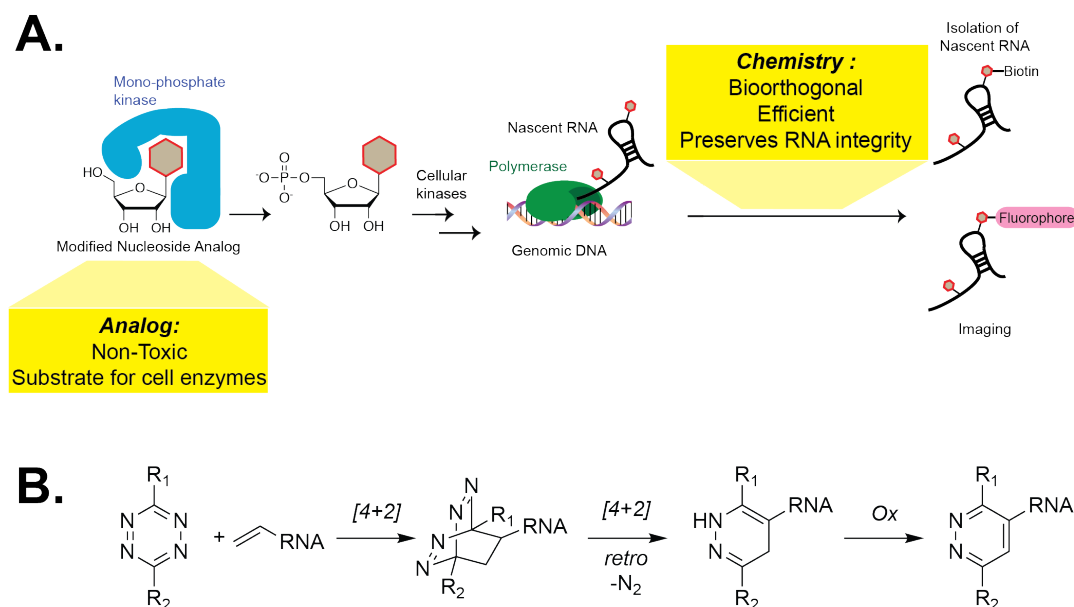


Figure 4-1. Schematic of metabolic labeling of RNA. A. Key steps that require optimization for efficient metabolic labeling of nascent RNA. **B.** Schematic of the inverse electron-demand Diels-Alder reaction with vinyl-modified RNA.

The inverse electron-demand Diels-Alder (IEDDA) reaction presents an attractive alternative to the currently adopted chemistries. It has been widely used for labeling biomolecules due to its fast kinetics, high yields and exclusion of metal catalysts (**Figure 4-1, B**).^{16,17} Traditionally, installation or genetic encoding of a strained alkene such as norbornene, *trans*-cyclooctene and cyclopropene have facilitated this rapid labeling.¹⁸⁻²⁰ However, this presents a significant challenge to reporter strategies that rely solely on cellular machinery for installation, wherein the enzymes in these metabolic pathways may

not accommodate large or even moderately-sized functional groups. We have observed that the size and placement of biorthogonal handles on nucleosides often dictates their utility as metabolic reporters.^{8,15} Furthermore, there is the possibility for cross-reactivity with nucleophiles in the cellular milieu.²¹ Thus, applying this chemistry requires smaller and more stable alkenes for nucleoside functionalization. These challenges present a unique opportunity to explore IEDDA reaction pairs that could be used to form cycloadducts on acceptable timescales, but at the same time are stable enough to provide selectivity.

While the application of vinyl-deoxynucleosides and vinyl-nucleotides has been explored, there has been no investigation of vinyl-nucleosides for RNA metabolic labeling in living cells.^{22,23} Furthermore, there has been no analysis of pairing vinyl-nucleosides with optimized tetrazines that balance the required reactivity with stability for labeling.

Herein, we report a holistic analysis of vinyl-nucleoside reactivity with a larger subset of tetrazines. We examine which nucleoside analogs can be incorporated into nascent RNA following addition to cellular media. We then synthesized and characterized optimal tetrazine/vinyl-nucleoside pairs based on kinetics and molecular orbital analyses. Optimized pairs enable tagging and imaging of nascent RNA without compromising RNA integrity. The methods described herein are also orthogonal to conventional CuAAC reactions used for RNA analysis. Overall, this analysis greatly expands our understanding of IEDDA reactions with nucleoside analogs and underscores the utility of biorthogonal methods that are amenable to RNA labeling.

4.2 Results and Discussion

Metabolic Incorporation of vinyl-nucleoside analogs.

Chemical reporter strategies that rely on host cell machinery naturally also depend on the modified analog being an acceptable substrate. In many cases, such as for sugars and amino acids, modifications can be placed or sized more liberally; that is, the enzymes in these metabolic pathways accept many configurations.²⁴ However, the design of successful nucleoside analog reporters has proven more challenging, requiring specific analysis of enzyme structures.

It is reported that mono-phosphate nucleoside kinases utilized in nucleoside salvage pathways are the most selective for their substrates. These enzymes have the highest level of substrate specificity due to very specific active site structures and substrate binding contacts.²⁵ Therefore, to identify positions that could accommodate vinyl functionality, we examined the crystal structures of uridine-cytidine kinase (PDB: 1BX4)²⁶ and adenosine kinase (PDB: 1UEJ)²⁷. Uridine-cytidine kinase revealed an open space at position C5 (**Figure 4-2, A**). Previous work has demonstrated that 5-ethynyluridine is well tolerated and incorporated into RNA, thus we synthesized 5-vinyluridine (**5-VU**) and 5-vinylcytidine (**5-VC**) as possible substrates.^{5,28} Adenosine kinase has cavernous pockets about C2, N6 and N7 of the canonical substrate. Alkyl modifications of adenosine at all three of these positions have been previously demonstrated.^{6,7,29} However, we did not wish to perturb the Watson-Crick face of the nucleoside, which is responsible for base pairing, thus we synthesized only 2-vinyladenosine (**2-VA**) and 7-deazavinyladenosine (**7-dVA**).³⁰ There have been no successful reports of incorporation of C8-functionalized purine nucleosides, thus we synthesized 8-vinyladenosine (**8-VA**) and 8-vinylguanosine (**8-VG**) to test the limits

of modification. This brought the final total to six vinyl-modified nucleosides, as shown in **Figure 4-2, B**.

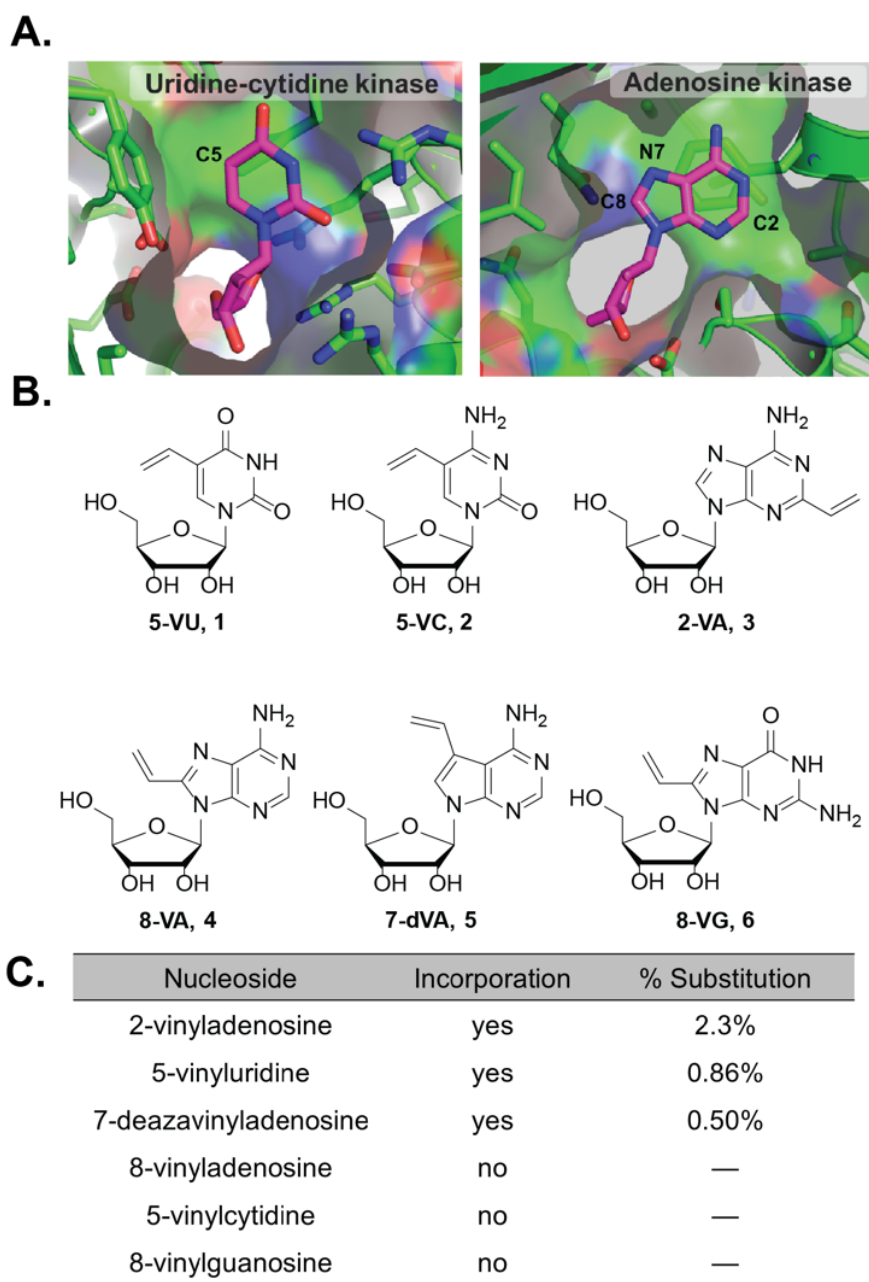


Figure 4-2. Initial analysis of vinyl nucleoside incorporation into RNA. **A.** Crystal structures of uridine-cytidine kinase and adenosine kinase bound to cytidine and adenosine, respectively. PDB: 1UEI, 1BX4 **B.** Structures of vinyl-nucleosides **C.**

Incorporation levels of vinyl-nucleosides. HEK293T cells were treated for 12 h at 1 mM final concentration, after which RNA was isolated. Percent incorporation was determined through total RNA digestion and LC-MS/MS analysis of mono-nucleoside mixtures.

To understand the extent of nucleoside substitution, we performed LC-MS/MS to obtain quantitative analysis of percent incorporation for each analog (**Table 4-S1**). RNA was isolated from cells treated with each analog, followed by digestion to mono-nucleosides. LC-MS/MS analysis showed that **2-VA** had the highest incorporation at 2.3% (**Figure 4-2, C**). **5-VU** showed 0.86% incorporation, while **7-dVA** showed 0.50% incorporation. These percent incorporation are similar to other metabolic probes, such as 5-ethynyridine.⁵ The markedly higher incorporation of **2-VA** is likely consistent with other reports that adenosine analogs with small modifications at C2 are highly incorporated into messenger RNA through polyadenylation.^{6,8}

We were unable to detect incorporation of **8-VA**, **8-VG**, and **5-VC**. Aside from the structure of the kinases, the solution-structure of modified nucleosides could be a reason for lack of incorporation. Two of the analogs (**8-VA** and **8-VG**) may adopt a primarily *syn* conformation about their N9-C1 glycosidic bond. Analysis of other C8 modified nucleosides has demonstrated that modifications at this position can dramatically alter *syn-anti* ratios of purine nucleosides.^{31,32} Therefore, these analogs may not be suitable substrates for initial kinase steps due to the inability of the base to bind. A greater surprise was the lack of **5-VC** incorporation. LC-MS analysis of the **5-VC** incubated with Trizol reagent revealed **5-VC** was completely degraded, and the Michael adduct with thiocyanate is observed (**Figure 4-S1**).

Therefore, we believe **5-VC** is susceptible to other Michael donors in the cellular environment.

Our initial analysis provides an extended framework for understanding the design and successful incorporation of vinyl-modified nucleosides in cells. Furthermore, ascertaining the limitations of modification placement and size will afford greater insights into the potential design of completely orthogonal, nucleoside-enzyme systems.

Kinetic analysis of vinyl-nucleosides and tetrazines.

Having established the structures of vinyl nucleosides tolerated by the cellular machinery, we turned our attention to identifying and characterizing paired tetrazines for IEDDA reactions. Acceptable chemical structures and functionalities of nucleoside analogs are dictated by host-cell enzymes in the metabolic pathways. As such, modifying them to achieve faster IEDDA rates through HOMO-LUMO gap narrowing can be challenging because terminal alkenes, the least reactive dienophiles, were installed. Therefore, we reasoned that modifying the structure of tetrazines with electron withdrawing groups would lower the LUMO and thereby narrow the HOMO-LUMO energy gap. After characterization of IEDDA kinetics, we can identify nucleoside-tetrazine pairs.

We worked toward the synthesis of six tetrazines (**Figure 4-3, A**), each of them having unique structures that could systematically lower the LUMO+1 orbital energy. **Tz-1**, **Tz-3**, **Tz-5**, and **DP-Tz** have been previously synthesized for use in IEDDA reactions with dienophile-bearing amino acids in proteins.^{20,33-35} **Tz-2** and **Tz-4** were pursued due to the examination of the LUMO+1 orbital energies calculated with B3LYP, 6-31+G* (**Figure 4-S2**). From these calculations, **Tz-2** showed a similar but slightly lower LUMO+1 orbital energy

than **Tz-1**. Structurally, **Tz-1** and **Tz-2** differ in placement of the nitrogen of the pyridine ring, which is *ortho*- or *meta*- to the tetrazine, respectively. Overall, we surmised this broad range of LUMO+1 orbital energies and structures would allow us to identify tetrazine-nucleoside pairs for IEDDA reactions.

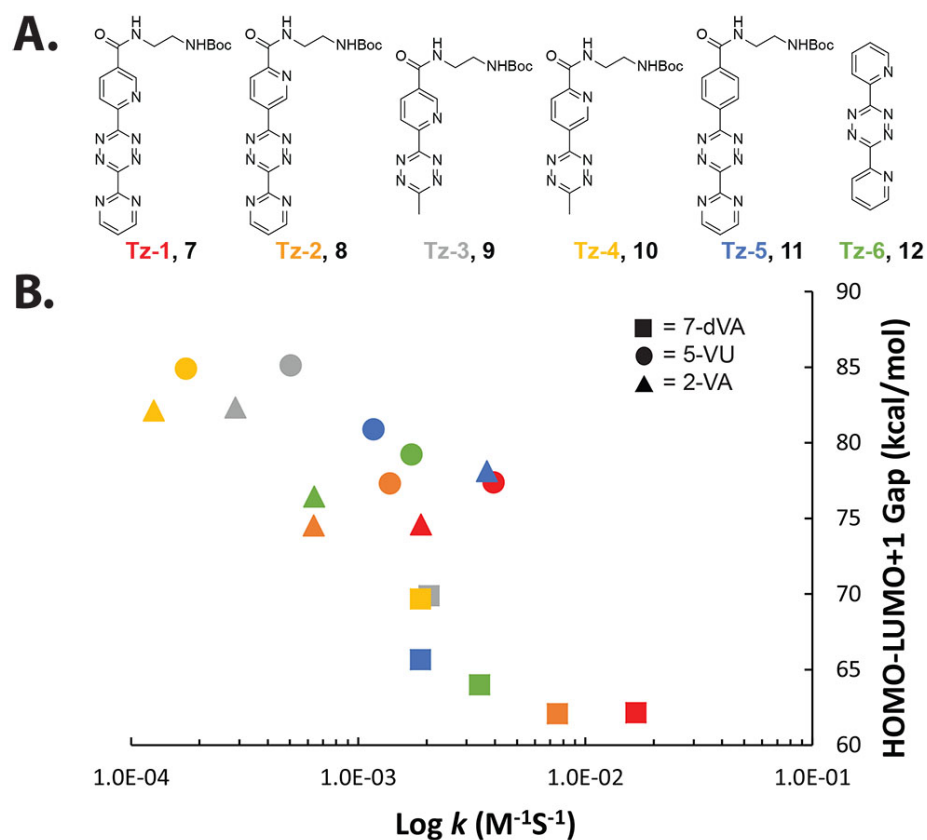


Figure 4-3. Screening tetrazines for IEDDA reactions with vinyl nucleosides **A.** Structures of tetrazines synthesized for this study. **B.** Comparison between HOMO-LUMO+1 energy gap and IEDDA rate constants. Second-order rate constants were determined from reactions containing 1 mM tetrazine and 10, 12.5, 15, 17.5 and 20 mM vinyl nucleoside or nucleobase.

Kinetic analysis was performed by measuring the change in absorbance of tetrazine over time. The reactions exhibited a color change of red-to-yellow, which was used to determine second-order rate constants through the initial rates method (**Figure 4-S2**).³⁶ As expected, addition of a polar solvent (water) resulted in faster reaction kinetics. Rates of these reactions were found to be between $10^{-4} - 10^{-2} \text{ M}^{-1}\text{s}^{-1}$. **Tz-1** and **7-dVA** showed the fastest rate with $1.67 \pm 0.68 \times 10^{-2} \text{ M}^{-1}\text{s}^{-1}$.

We plotted the HOMO-LUMO+1 gap against the rate constants for each tetrazine-nucleoside pair to ascertain any correlations between these features (**Figure 4-3, B**). Generally, we noticed a modest trend wherein narrower HOMO-LUMO+1 gaps resulted in faster IEDDA reaction kinetics. Notable outliers were observed with **Tz-1** and **Tz-2** and their reaction kinetics. The rate of reactions with each vinyl nucleoside showed that **Tz-1** were 2 to 10 factors faster than that of **Tz-2** despite having similar LUMO+1 energies (**Figure 4-S2**).

Overall, we have provided a comprehensive analysis of reaction rates between the vinyl nucleosides and tetrazine pairs, which underscores the spectrum of reactivity amongst these reagents. The rates of reaction were determined to be comparable to SPAAC reactions ($\sim 10^{-2} \text{ M}^{-1}\text{s}^{-1}$)³⁷. This result is especially exciting as it suggests that IEDDA reactions could be a sufficient replacement for RNA metabolic reporters that rely on CuAAC chemistry.

Establishing reaction conditions to stabilize tetrazines during IEDDA with modified RNA.

Our kinetic analysis suggested that reactions between **Tz-1** and the vinyl-nucleosides would be the most efficient for IEDDA labeling of metabolic RNA. However, we questioned its stability in contexts where it would be used for cell imaging and RNA enrichment. A major challenge to utilizing reactive tetrazines for IEDDA chemistry is their susceptibility to be hydrolyzed, restricting their utility in biological applications.³⁸ We note that addition of water to each of the tetrazines consistently resulted in a color shift of red-to-yellow, indicating disruption of the extended pi-conjugation. This will naturally reduce the labeling efficiency of RNA labeling reactions, which are typically performed in aqueous buffers. While there were changes noted for all tetrazines, we found that **Tz-1** was the most susceptible to hydrolysis (**Figure 4-S3**). In light of this finding, we surmised it would not be amenable to RNA tagging reactions, despite its fast IEDDA kinetics. Further attempted syntheses of a biotinylated version failed under multiple conditions due to degradation by nucleophilic addition (data not shown).

We next sought to identify reaction conditions that could reduce hydrolysis of the tetrazines, and thus improve the efficiency. We hypothesized that the rate determining step of the hydrolysis pathway is expulsion of nitrogen after addition of water. With the addition of water being reversible, we hypothesized that the equilibrium of the addition could be shifted through lowering the pH (**Figure 4-4, A**). Hydrolysis was tested with 10% acetic acid added to reactions of **Tz-1** in a DMSO:H₂O (1:1) mixture (**Figure 4-4, B**). A striking reduction of hydrolysis was observed in reactions with the addition of acetic acid (pH 5.0).

Kinetic analysis was then performed on reactions of **Tz-1** and vinyl-nucleosides and vinyl-nucleobases in the presence of 10% acetic acid (**Figure 4-4, C; Figure 4-S2**). Interestingly, increased rates are observed in **5-VU** under acidic conditions, which may be due to protonation of the nucleobase, thereby lowering the HOMO of the dienophile. These results suggest that lowered pH conditions could provide greater tetrazine stability and in turn, efficiency, while at the same time enabling robust reactivity for labeling.

We then compared RNA integrity and IEDDA biotinylation at pH 7.4 and 5.0. We observed that lowering the pH did not cause spurious RNA degradation, and may indeed be protective to RNA degradation through preventing cleavage of the phosphodiester bonds via the 2' hydroxyl (**Figure 4-4, D**). We also note that at pH 5.0, the rate of depurination is $\sim 10^{-8} \text{ M}^{-1}\text{s}^{-1}$,³⁹ which is far slower than the rate of IEDDA, making this a viable method for preserving RNA stability and conjugation.

After comparing reaction kinetics and tetrazine hydrolysis, we decided to pursue synthesis of a tetrazine conjugate, **Tz-4 biotin (Methods: Synthesis)** due to its stability and ease of synthesis. Further, kinetic analyses have demonstrated that this tetrazine as well as **Tz-3** exhibit reasonable rate constants with the all nucleosides in aqueous conditions (**Figure 4-S2**). With a biotin conjugate in hand, we utilized our modified reaction conditions to perform dot-blot analysis of vinyl nucleoside incorporation (**Figure 4-4, E**). Nucleosides were incubated in cells for 5 h, after which total RNA was extracted and reacted with **Tz-4 biotin** to enable detection of nascent transcripts through streptavidin-HRP. Notably, reactions with vinyl-modified RNA were completed in 2 h, in contrast to previous reports of IEDDA reactions taking >24 h for completion.^{22,23} Dot blot analysis confirmed that **7-dVA**, **2-VA**, and **5-VU** were incorporated into nascent RNA.

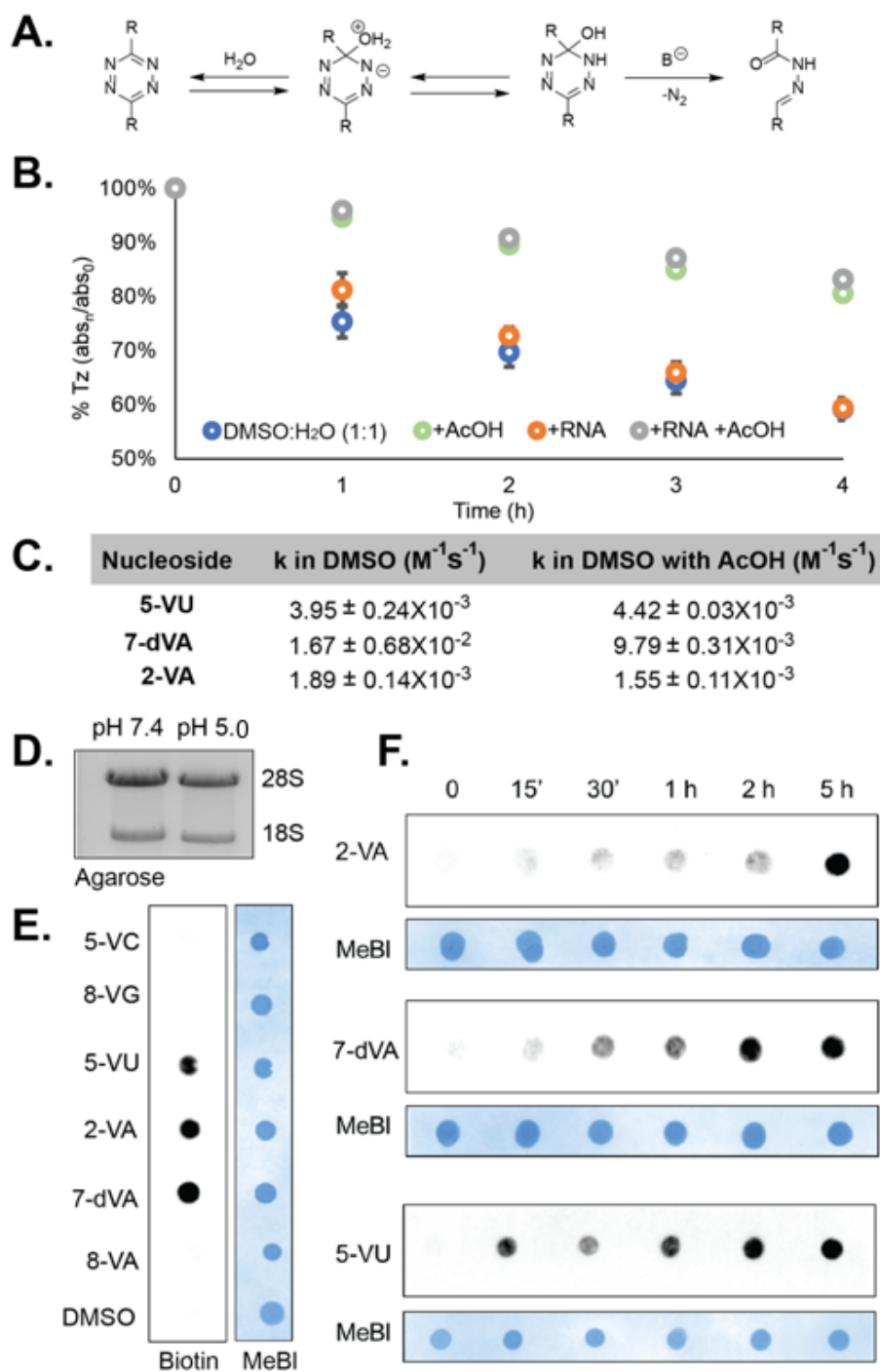


Figure 4-4. Establishing reaction conditions to stabilize tetrazines in IEDDA. A. Proposed hydrolysis reaction of tetrazines **B.** Analysis of tetrazine stability in different buffer conditions. **C.** Comparison of rate constants in DMSO and DMSO with acetic acid. **D.** Agarose gel demonstrating RNA integrity is maintained post IEDDA labeling at lower pH. **E.**

Streptavidin dot blot detection of analogs **2-VA**, **7-dVA** and **5-VU** demonstrates biotinylation through IEDDA. HEK293T cells were treated with each analog for 5 h at 1 mM concentrations. **F.** Time course analysis of **2-VA**, **7-dVA** and **5-VU** incorporation at times 0 min, 15 min, 30 min, 1 h, 2 h and 5 h treated. At 1 mM concentrations. All IEDDA biotinylation reactions were performed with 50 μ M **Tz-4 biotin** incubated for 2 h at 37 °C.

We also analyzed the time-dependence of incorporation through performing a time course of analog treatment followed by dot blot analysis (**Figure 4-4, F**). Notably, biotinylation signal was visible after only 15 min, and robust incorporation was noted at 5 h. While **2-VA** was the most highly incorporated according to the LC-MS/MS analysis, it is not the most reactive. We found that **5-VU**, which had second highest incorporation, also demonstrated efficient biotinylation and was more easily detected at lower time points.

The addition of acid to IEDDA markedly reduced tetrazine hydrolysis, which serves to improve conjugation. To the best of our knowledge, IEDDA reactions with tetrazines and olefins have typically been performed at neutral pH. As such, the possibility of augmenting IEDDA reactions through lowering the pH offers a unique improvement to contexts where the required tetrazine is highly reactive. It is also exciting to note that reactions between modified RNA and tetrazines can be performed under conditions that preserve RNA integrity.

Utilizing vinyl nucleosides for in-cell analyses.

We have observed that many of our vinyl nucleosides can be incorporated into RNA. To further test their utility in tracking RNA synthesis, we aimed to establish: (1) vinyl

analogs are non-toxic and compare them to other more widely-used analogs for RNA labeling, (2) vinyl analogs can be imaged through IEDDA, and (3) that the IEDDA conditions developed herein are indeed orthogonal to other biorthogonal methods of labeling.

Metabolic labeling of RNA requires that the modified analogs be substrates for a wide variety of enzymes and pathways. As such, there is always the possibility that analogs could exert stress upon cells and cause cytotoxicity or inhibited growth. This is especially concerning when incubating analogs in cells for longer timescales. Furthermore, this limits their use in living animals where it can take several days for analogs to be incorporated at rates suitable for downstream analysis.

Other analogs such as 4-thiouridine⁴⁰ have been shown to induce significant amounts of toxicity in cells; however, analysis of other conventional analogs has not been extensively explored. MTT assays demonstrated that all incorporated vinyl analogs had no significant effect on cell proliferation after incubations as long as 12 h (**Figure 4-S4**). We compared this to 5-ethynyluridine (**5-EU**), a well-established, commercially available nucleoside used in metabolic labeling of RNA.⁵ Interestingly, we found that it had modest effects on cell proliferation at 12 h, but after 48 h, cell proliferation had decreased by nearly 50% (**Figure 4-5, A**). Even more striking, a direct comparison with **5-VU** at these same conditions showed no significant changes in cell proliferation. Cell death assays indicated that growth inhibition, and not direct cytotoxicity was the primary effect of extended **5-EU** incubations (**Figure 4-S5**). This striking difference has also been observed with some alkynyl analogs of DNA.⁴¹ We also observed changes to cell morphology, including cell detachment and aggregation (**Figure 4-5, B**), which is consistent with problems in cellular respiration and toxicity. These results demonstrate that vinyl analogs are a safer option for

applications *in vivo*, which require longer incubation times. This is an incredibly important result for researchers focused on performing RNA labeling experiments in animals.

We next investigated using vinyl-analogs for imaging of nascent transcription. We synthesized a fluorophore-conjugated tetrazine (**Tz-3 TAMRA**; **Methods: Synthesis**). As shown in **Figure 4-5, C**, robust RNA labeling is observed in the nucleolus for **5-VU**, which is the location of rRNA biosynthesis. This pattern has been observed with other analogs that can be used for RNA imaging. We saw similar imaging results for **2-VA** and **7-dVA**, and treatment with actinomycin D, an RNA polymerase I/II inhibitor, significantly diminished nucleolar staining for all three analogs (**Figure 4-S6**).

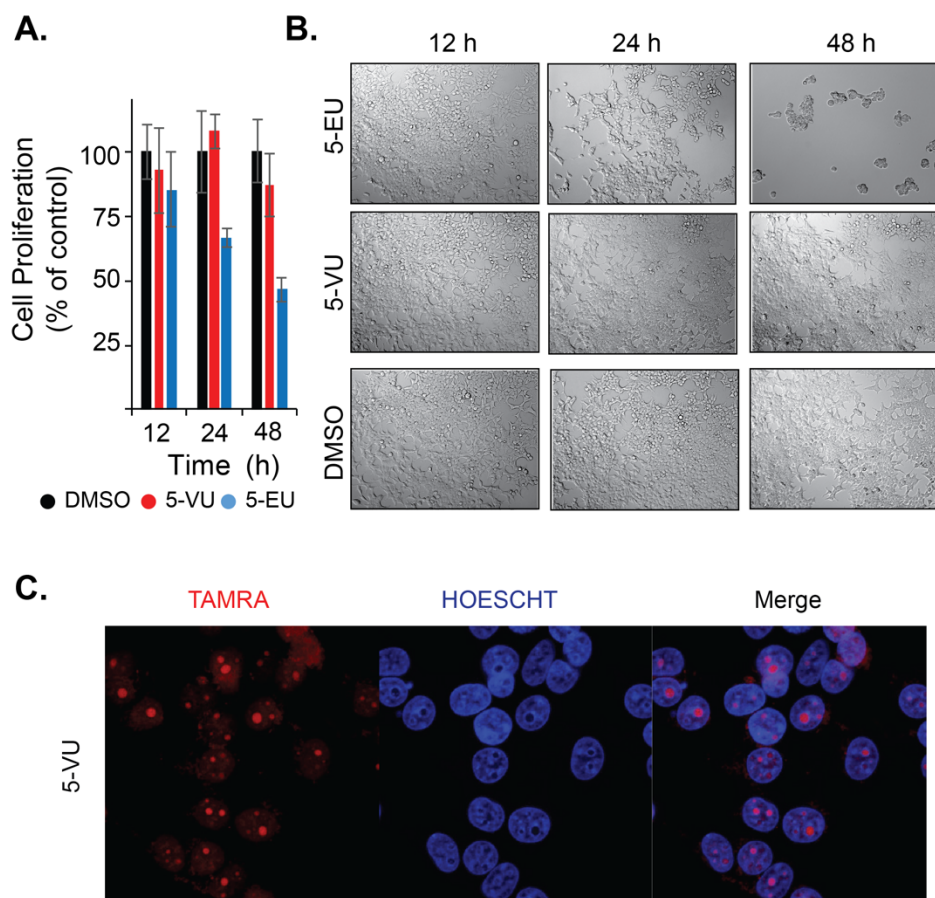


Figure 4-5. Effects of vinyl-nucleosides on cell viability and use for RNA imaging. A.

MTT assays demonstrate that 5-VU does not have an effect on cell growth, whereas 5-EU produces significant growth inhibition relative to the DMSO-treated control at 1 mM for 12, 24 and 48 h. **B.** DIC imaging of cells after various incubation lengths confirms the effects on cell growth. **C.** IEDDA-based imaging demonstrates that 5-VU can be used for visualizing nascent transcription. HEK293T cells were treated with 1 mM 5-VU for 5 h, following by fixation, permeabilization and incubation with **Tz-3 TAMRA**.

Finally, we wanted to test if our IEDDA reactions would be mutually orthogonal to a CuAAC reaction with alkyne-modified RNA. That is, we wished to ascertain that the tetrazine was selective only for the alkene and the azide only for the alkyne. After incubation of **5-VU** or **5-EU** in cells and isolation of the vinyl- or alkynyl-RNA, reactions with biotin-N₃ and **Tz-4 biotin** were carried out. Dot blot analysis showed no cross talk between functional groups (**Figure 4-S7**). To our knowledge, this is the first example of mutually orthogonal reactions performed on RNA transcripts, offering the ability to perform multiple bioorthogonal reactions in the same experiment for the study of RNA. This approach may be useful on shorter time scale experiments performed *in vitro*.

4.3 Conclusion

Elucidating the utility and limitations of biorthogonal reactions is critical to expanding their use for studying all classes of biomolecules. We have established that vinyl nucleosides can be utilized to metabolically label nascent cellular RNA. Our data demonstrates the successful metabolic incorporation of **2-VA**, **7-dVA**, and **5-VU**, which were

intelligently designed through enzyme structural examination. We have also performed an exhaustive evaluation of IEDDA kinetics between all of our synthesized analogs and accompanying tetrazines. Our study highlights the importance of performing these context-dependent analyses to identify optimized reagents and conditions. Using these comparisons, we have identified fast-acting tetrazines that could provide great utility for IEDDA reactions with vinyl analogs of any biomolecule.

Through our studies, we are also the first to report the use of acidic reaction conditions to supplant tetrazine hydrolysis, without severely compromising reaction kinetics or altering RNA integrity. Using our modified protocols, we have demonstrated that vinyl-nucleosides and tetrazines can be used to detect and image nascent transcripts. Vinyl nucleosides do not inhibit cellular proliferation like the most common nucleosides used for *in vivo* work, such as 4-thiouridine and 5-EU. In addition, vinyl-nucleosides can be used in tandem with alkyne functionalized nucleosides and for multicomponent experiments to study RNA.

We envision these vinyl nucleosides will be used extensively both *in vitro* and *in vivo*, as they are easily handled, non-toxic, are able to be functionalized through selective chemistry, and can preserve RNA integrity. Future work in our lab will be to use these exciting new tools to evaluate differences in RNA expression and biology within living systems, such as whole animals.

4.4 Acknowledgements

Synthesis of vinyl-nucleosides and tetrazines, kinetic analysis and hydrolysis analysis of tetrazines was carried out by Miles Kubota.

4.5 Methods

Vinyl-nucleoside TRizol Reaction Analysis by LC-MS

To analyze the stability of vinyl-nucleosides in Trizol, **5-VU**, **5-VC**, **7-dVA**, and **2-VA** were placed in TRizol and analyzed by LC-MS. 20 μL of 2 mM stock solution of Vinyl-nucleosides in DMSO were added to 80 μL of Trizol (Life Technologies) in a HPLC vials. The mixtures were placed on a shaker at r.t. for 5 h and then analyzed by LC-MS using a C18 column and eluted with a gradient of $\text{H}_2\text{O}/\text{MeCN}$ 5-95% with 0.1% formic acid. The mass spectrums were searched for the exact mass of each vinyl-nucleoside and the subsequent thiocyanate addition product.

HOMO-LUMO Energy Calculation

Vinyl-nucleoside HOMO energy and tetrazine LUMO+1 energies were calculated using density functional theory (DFT). Calculations were performed with Spartan 14, using the B3LYP level of theory and the basis set 6-31+G*.

Rate Studies

The reactions between vinyl-nucleosides and tetrazines were monitored by the change in the tetrazine absorbance at 530 nm. Reactions were initiated in a 96-well plate by mixing 100 μL of 2 mM tetrazine solution in DMSO with 100 μL of vinyl-nucleoside in the proposed solvent (DMSO , 1:1 $\text{H}_2\text{O}:\text{DMSO}$, and 1:5 $\text{AcOH}:\text{DMSO}$). The concentration of vinyl-nucleoside ranged from 10-20 mM, while tetrazine concentrations were held at 1 mM. All rate studies were performed in triplicate under pseudo-first order conditions. Absorbance measurements were recorded every min over a time period of 3 h time interval using a

BioTek Epoch plate reader equipped with Gen5 software. Pseudo-first order rate constants (k_{obs}) were calculated by plotting natural log of [tetrazine] versus time (in s). Second-order rate constants were determined by plotting k_{obs} vs nucleoside concentration.

Tetrazine Hydrolysis Analysis

Analysis of tetrazine hydrolysis was performed by NMR. 10 mg of Tz-1 was added to an NMR tube and dissolved in DMSO- D_6 (0.5 mL). An initial ^1H NMR was performed to establish a baseline. H_2O (0.5 mL) was then added to the NMR tube and vortexed. The NMR tube was then heated to 37 °C for 24 h. ^1H NMR was then performed on the sample and hydrolysis products were determined.

Tetrazine Stability with AcOH

Evaluation of the stability of Tetrazine and RNA was performed by monitoring the change of the tetrazine absorbance at 530 nm. Reactions were performed within in a 96-well plate by mixing 300 μL of 2 mM tetrazine solution in DMSO with 300 μL of the proposed solvent (H_2O , 20 μg of RNA in H_2O , 20 μg of RNA in 20% AcOH in H_2O , and in 20% AcOH in H_2O). The concentration of tetrazines were held at 1 mM, AcOH concentrations were held at 10%, and RNA were held at 10 μg . All stability studies were performed in triplicate at 23 °C and absorbance measurements were recorded every hour over a time period of 4 h time interval using a BioTek Epoch plate reader equipped with Gen5 software.

Vinyl-nucleoside labeling of cellular RNA

HEK293T cells (ATCC) were cultured in DMEM (Corning) supplemented with 10% FBS, 1% penicillin and streptomycin and grown at 37 °C, 5% CO₂. Analogs were added to complete culture medium from 200 mM stocks with a final concentration of <1% DMSO.

RNA isolation and biotinylation via IEDDA

After labeling, total cellular RNA was harvested using Trizol Reagent (Invitrogen) following the manufacturer's instructions or RIPA lysis buffer (150 mM NaCl, 5 mM EDTA pH 8.0, 50 mM Tris, pH 8.0, 1.0 % NP-40, 0.5 % sodium deoxycholate, 0.1% SDS) spiked with DNase Turbo and RNaseOut (Thermo Fisher). Briefly, ~20 million cells were scraped and suspended in 2 mL of lysis buffer, and incubated on ice for 15 min. Lysates were mixed 1:1 with phenol/chloroform pH 5.2 (Fisher) vortexed briefly and spun down at 4 °C to separate the layers. The aqueous layer was then applied to a Zymo RNA clean and concentrator IIC column (Zymo) and total RNA was isolated according to the manufacturer's instructions, with RNA eluted in 50 µL of nuclease free water. IEDDA reactions were prepared using 10 µg of total RNA in solutions of 6% DMSO:AcOH (pH 5.0) water, and 1 mM Tz-4 biotin to a final volume of 50 µL. Reactions were incubated with at 37 °C for 2 h at 400 RPM. Biotinylation reactions were purified using Zymo RNA clean and concentrator-5 spin column according to the manufacturer instructions, with RNA eluted in 10 µL of nuclease free water.

HRP-streptavidin Dot blotting

Equal amounts of column-purified RNA were spotted onto 2x SSC equilibrated Hybond-N+ membrane (GE Healthcare), allowing spots to dry in between. The membrane was

crosslinked using the auto crosslink function at 254 nm on a UV Stratalinker (Stratagene). The membrane was then blocked in a 10% SDS blocking solution followed by incubation with 1:10,000 dilution of high sensitivity streptavidin-HRP (ThermoFisher Scientific) in blocking solution. The membrane was washed x2 in a 1:10 solution of blocking solution and x2 in Tris-saline buffer. It was then incubated in SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher Scientific) according to manufacturer's instructions and imaged on a ChemiDoc MP imaging system (Bio-Rad) under the chemiluminescence channel. Following imaging, the blot was stained in a 0.04% methylene blue, 0.3 M sodium acetate solution to visualize all RNA as a loading control. The blot was imaged again using the colorimetric channel.

RNA fluorescence imaging via IEDDA

Coverslips in 6-well tissue culture plates were treated with 1x poly-D lysine solution for 3 h at 37 °C. HEK293T cells were seeded at densities of 2.5×10^5 and grown to ~50% confluency on glass cover slips. Cells were treated with vinyl nucleosides to final concentrations of 1 mM and incubated for 5 h. After labeling, cells were washed twice with DPBS and fixed for 10 min at room temperature with 3.7% paraformaldehyde. Cells were quenched with 50 mM glycine, DPBS for 5 min, and then washed twice more. Cells were permeabilized in 0.5% Triton-X DPBS for 15 min, and then washed x2 with DPBS. Cells were then treated with solutions of 5 μ M Tz-3 TAMRA in DPBS (6% DMSO:AcOH (pH 5)), and incubated for 3 h at 37 °C in the dark. Cells were washed x2 with 0.25% Triton-X DPSB for 5 min, and then x5 with DPBS for 5 min each on an orbital shaker. Cells were stained with a solution of 1:3000 Hoechst 333242 for 10 min (Thermo Fisher). Coverslips were briefly washed and

then mounted using VectaShield mounting medium (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 Actinomycin D

Cells were grown and treated with analog as previously described, with transcriptional inhibitor actinomycin D (Sigma) being co-administered to a final concentration of 20 μ M. Following incubation, cells were washed x3 with DPBS and lightly fixed with 0.5% paraformaldehyde in 0.5% Triton-X DPBS. Cells were then fixed, stained, mounted and imaged as previously described.

MTT Assay

HEK293T cells were seeded at densities of 10,000 cells into a 96-well plate and grown for 1 day. Cells were then incubated for various lengths of time with compounds **2-VA**, **7-dVA**, **5-VU**, and **5-VC** at 1 mM final concentration, 0.5% DMSO or 0.5% DMSO only. 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 unsupplemented RPMI minus phenol red, and cells were incubated for 3 h at 37 °C with 1 mM MTT. Formazan solid was solubilized in DMSO for 30 min at 37 °C and absorbance readings were taken at 540 nm.

Trypan Blue Assay

HEK293T cells were seeded at densities of $1.5 - 2.5 \times 10^5$ into 6 well plates and grown for 18 - 24 h. Cells were then incubated for 12, 24 and 48 h with compounds **5-VU** and **5-EU** at 1 mM final concentration. Following incubation, cells were trypsinized, removed from the plate and spun down. The supernatant was aspirated and the cells were resuspended in 1 mL of culture medium, after which 10 μ L was aliquoted for Trypan staining. To each aliquot, 10 μ L of 0.4% Trypan Blue solution (Gibco) was added and 10 μ L of this solution was applied to a hemocytometer for assessment of cell viability. Total and dead cell counts was recorded for each of the four corners of the grid, and repeated for biological duplicates of two.

Mass spec analysis of total RNA

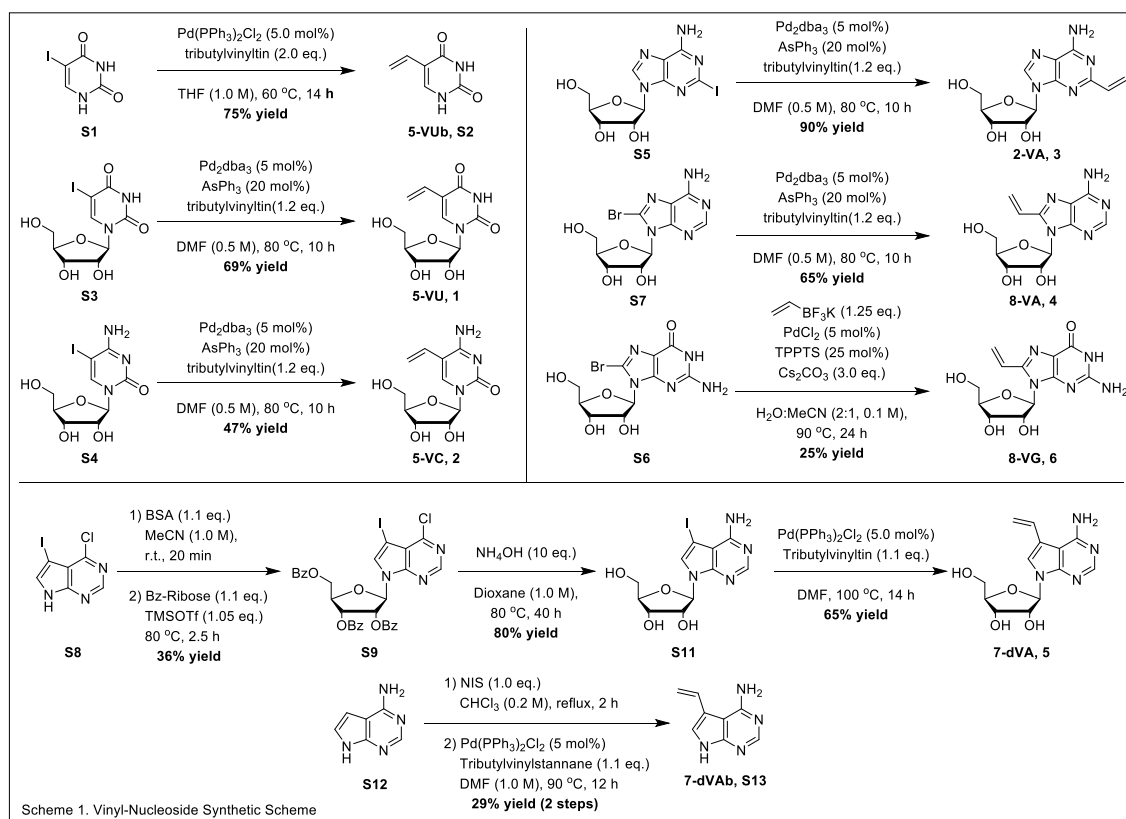
Cells were treated with 1 mM analog for 12 h as previously described. After RNA extraction with TRIzol (Thermo Fischer Scientific) and purification with RNA clean & concentrator-5 columns (Zymo), 1 μ g total RNA samples containing each analog were digested to nucleosides using a nucleoside digestion mix (New England Biolabs). A calibration curve of standard solutions containing Adenosine, Uridine, Guanosine, Uridine, vinyl-modified nucleoside analogues 5-VU, 5-VC, 7-dVA, 2-VA and 5-iodocytidine as an internal standard were injected on a Waters Micromass Quattro Premier XE LC-MS system equipped with a triple quadrupole (QQQ) mass detector operating in multiple reaction monitoring (MRM) and positive (+ESI) electrospray ionization modes. Liquid chromatography was carried out on a Waters Acuity UPLC equipped with a C18 column (4.6×150 mm, 5 μ m) with a gradient mobile phase consisting of water (+0.2% acetic acid) : acetonitrile. Digested RNA

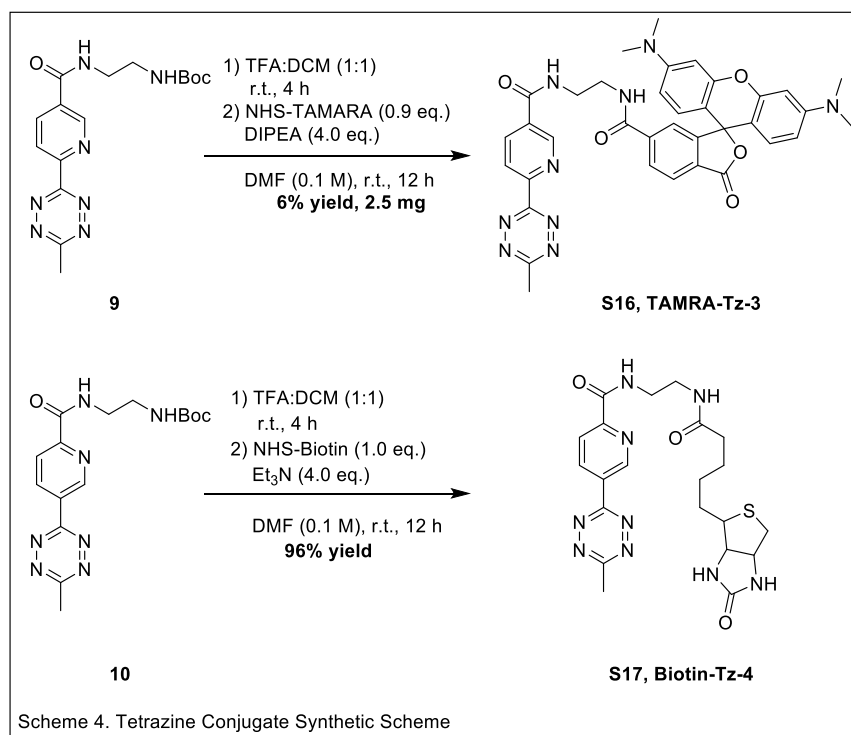
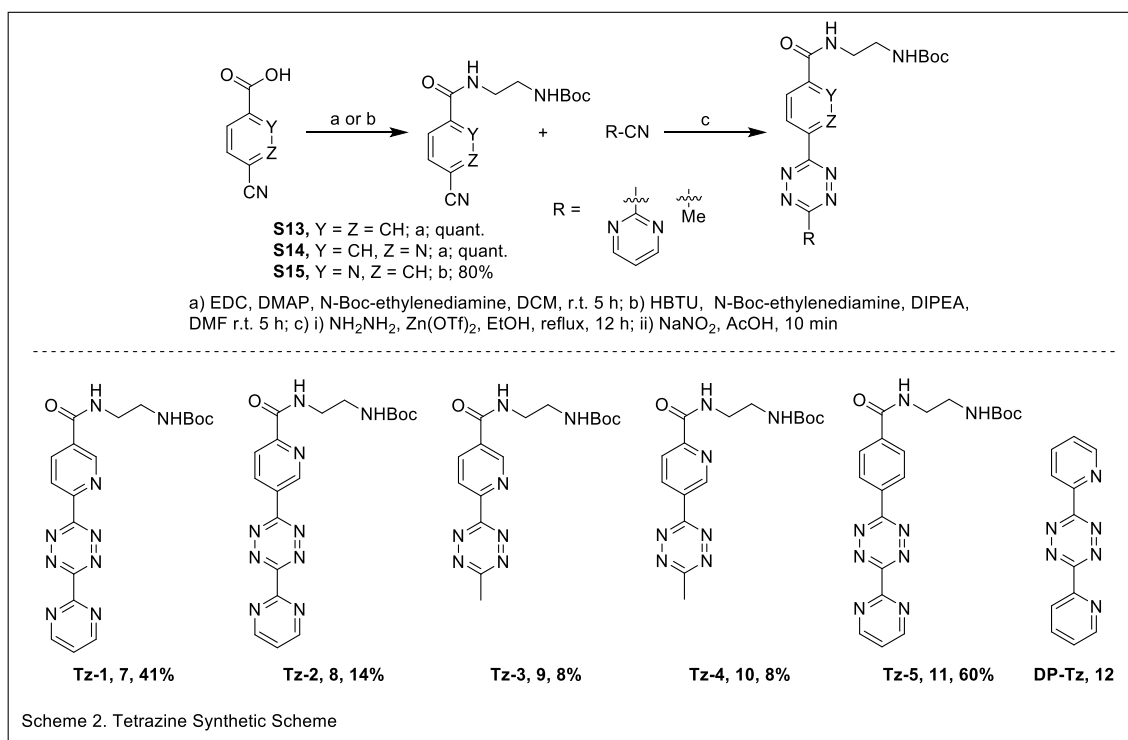
sample reactions were diluted 1:10 and 10 uL was directly injected. Mass spectrometry data acquisition was recorded in MRM mode and analyzed on Waters MassLynx 4.1 software. Each nucleoside was identified according to the precursor and product ions found through QuanOpt parameters. The percent substitution of each modified nucleoside was determined by dividing the spectral counts of each analog by the total spectral counts for the canonical and modified analog.

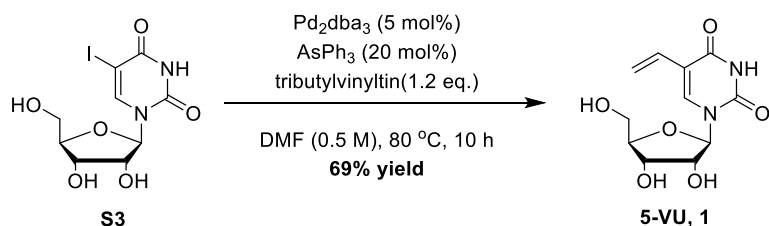
Synthesis

Compounds **S2**⁴² and **12** were synthesized as previously reported. Compounds **7-11** were synthesized following previously reported methods.^{18,33,35} All reagents were purchased from commercial suppliers and were of analytical grade and used without further purification unless otherwise noted. Reaction progress was monitored by thin-layer chromatography on EMD 60 F254 plates, visualized with UV light. Compounds were purified via flash column chromatography using Sorbent Technologies 60 Å 230 x 400 mesh silica gel or Teledyne ISCO CombiFlash System with RediSep Rf Media. Preparative reverse-phase HPLC was performed on a Rainin Dynamax equipped with a ZORBAX SB-C18 column (Agilent) with HPLC grade acetonitrile and 18 MΩ deionized water, each containing 0.1% trifluoroacetic acid. Anhydrous solvents dichloromethane (DCM), methanol (MeOH), dimethylformamide (DMF), tetrahydrofuran (THF), toluene (PhMe) were degassed and dried over molecular sieves. All reaction vessels were flame dried prior to use. NMR spectra were acquired with Bruker Advanced spectrometers. All reactions were carried in dark unless otherwise noted. All spectra were acquired at 298 K. ¹H-NMR spectra were

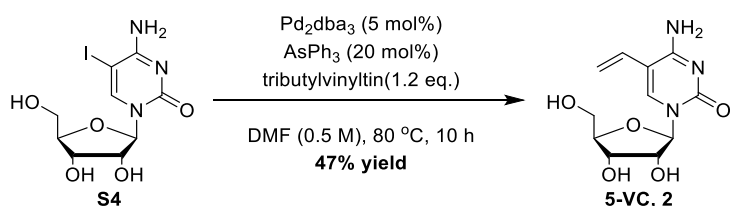
acquired at 400 MHz and 500 MHz. ^{13}C -NMR spectra were acquired at 125 MHz. Chemical shifts are reported in ppm relative to residual non-deuterated NMR solvent, and coupling constants (J) are provided in Hz. All NMR spectra was analyzed using MestreNova software. Low and high-resolution electrospray ionization (ESI) mass were collected at the University of California-Irvine Mass Spectrometry Facility. The abbreviations used can be found in the document JOC standard Abbreviations and Acronyms, <http://pubs.acs.org/paragonplus>





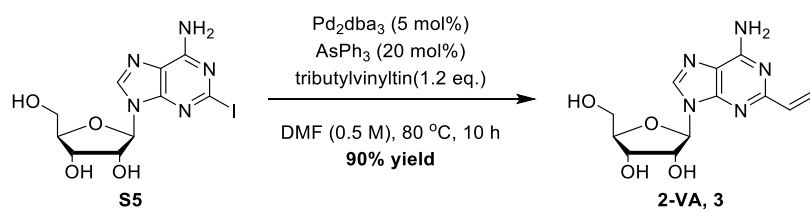


5-Vinyluridine was prepared according to a modified procedure by Farina, V. *et al.*⁴² 5-Iodouridine, **S3** (0.50 g, 1.35 mmol, $\text{Pd}_2(\text{dba})_3$ (67.5 mg, 67.6 μmol), AsPh_3 (82.7 mg, 0.27 mmol) was dissolved in DMF (3 mL). tributylvinylstannane (0.47 mL, 1.62 mmol) was added to the mixture and stirred for 10 h. Upon completion, the mixture was concentrated *in vacuo* and purified by silica flash column chromatography (EtOAc followed by 0-10% MeOH/DCM) to give **5-VU, 1** (0.25 mg, 69%) as a white solid. HRMS Calcd for $\text{C}_{11}\text{H}_{14}\text{N}_2\text{O}_6\text{Na}$ $[\text{M}+\text{Na}^+]$ 293.09, found 293.0750 $[\text{M}+\text{Na}^+]$; ^1H NMR (400 MHz, CDCl_3) δ 8.21 (s, 1H), 6.39 (dd, $J = 17.7, 11.7$ Hz, 1H), 5.88 (ddd, $J = 8.1, 5.6, 1.6$ Hz, 2H), 5.12 (dd, $J = 11.5, 1.6$ Hz, 1H), 4.19 – 4.14 (m, 2H), 4.01 – 3.97 (m, 1H), 3.89 – 3.83 (m, 1H), 3.76 – 3.70 (m, 1H); ^{13}C NMR (126 MHz, CD_3OD) δ 151.77, 139.10, 129.07, 115.31, 113.29, 90.87, 86.18, 75.96, 70.93, 61.87.

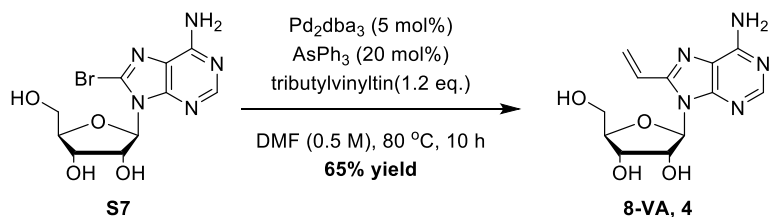


5-Iodocytidine, **S4** (0.50 g, 1.35 mmol, $\text{Pd}_2(\text{dba})_3$ (62.0 mg, 67.7 μmol), AsPh_3 (82.9 mg, 0.27 mmol) was dissolved in DMF (3 mL). tributylvinylstannane (0.48 mL, 1.63 mmol) was added to the mixture and stirred for 10 h. Upon completion, the mixture was concentrated *in vacuo* and purified by silica flash column chromatography (EtOAc followed by 10%

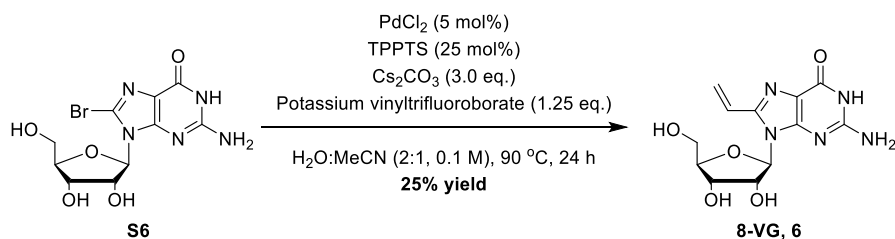
H₂O/MeCN) to give **5-VC, 2** (172 mg, 47%) as an off-white solid. HRMS Calcd for C₁₁H₁₆N₃O₅ [M+H⁺] 270.11, found 270.1090 [M+H⁺]; ¹H NMR (400 MHz, D₂O) δ 8.12 (s, 1H), 6.43 (dd, *J* = 17.2, 11.0 Hz, 1H), 5.89 (d, *J* = 3.0 Hz, 1H), 5.55 (d, *J* = 17.2 Hz, 1H), 5.35 (d, *J* = 11.0 Hz, 1H), 4.31 (dd, *J* = 5.1, 3.0 Hz, 1H), 4.29 – 4.24 (m, 1H), 4.19 – 4.12 (m, 1H), 4.02 (dd, *J* = 12.9, 2.6 Hz, 1H), 3.87 (dd, *J* = 12.9, 3.4 Hz, 1H); ¹³C NMR (126 MHz, D₂O) δ 164.00, 156.83, 138.23, 126.79, 126.77, 117.29, 117.27, 107.49, 90.71, 83.56, 74.50, 68.70, 60.04.



2-Iodoadenosine, **S5** (500 mg, 1.27 mmol), Pd₂dba₃ (58.2 mg, 63.5 μmol), and AsPh₃ (77.9 mg, 0.25 mmol) was dissolved in DMF (1.3 mL) and stirred at 100 °C. Tributylvinyltin (0.45 mL, 1.53 mmol) was added to the mixture and stirred for 10 h. The reaction was then concentrated *in vacuo* and purified by silica flash column chromatography (10% MeOH/DCM) to give **2-VA, 3** (335 mg, 90%) as a tan foam. HRMS Calcd for C₁₂H₁₅N₅O₄Na [M+Na⁺] 316.11, found 316.1022 [M+Na⁺]; ¹H NMR (400 MHz, DMSO) δ 8.27 (s, 1H), 7.23 (s, 2H), 6.55 (dd, *J* = 17.2, 10.4 Hz, 1H), 6.31 (dd, *J* = 17.2, 1.8 Hz, 1H), 5.84 (d, *J* = 6.4 Hz, 1H), 5.54 – 5.46 (m, 1H), 5.37 (d, *J* = 6.2 Hz, 1H), 5.33 (d, *J* = 4.6 Hz, 1H), 5.13 (d, *J* = 4.2 Hz, 1H), 4.60 (dd, *J* = 11.3, 5.9 Hz, 1H), 4.11 (d, *J* = 2.5 Hz, 1H), 3.92 (d, *J* = 2.8 Hz, 1H), 3.70 – 3.58 (m, 1H), 3.56 – 3.46 (m, 1H); ¹³C NMR (126 MHz, DMSO) δ 157.80, 155.75, 149.73, 140.42, 137.10, 121.21, 118.62, 87.74, 85.98, 73.31, 70.84, 61.82.

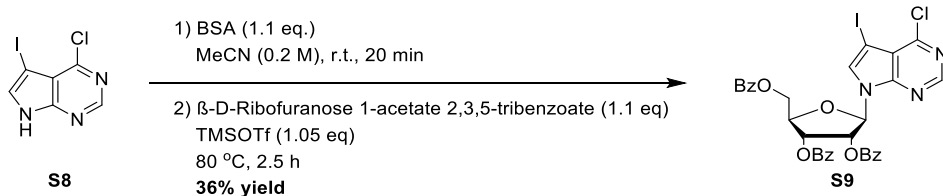


8-Bromoadenosine, **S7** (50 mg, 0.14 mmol), AsPh_3 (8.80 mg, 28.9 μmol) and Pd_2dba_3 (6.60 mg, 7.22 μmol) were added to a flask and dissolved in DMF (0.2 mL) and heated to 80 °C. Tributylvinyltin (51.0 μL , 173 μmol) was added to the flask and stirred for 10 h. The mixture was concentrated *in vacuo* and purified by silica flash column chromatography (EtOAc then 10% MeOH/DCM) to give **8-VA, 4** (21.7 mg, 65%) as a white solid. HRMS Calcd for $\text{C}_{12}\text{H}_{15}\text{N}_5\text{O}_4\text{Na}$ $[\text{M}+\text{Na}^+]$ 316.11, found 316.1022 $[\text{M}+\text{Na}^+]$; ^1H NMR (400 MHz, DMSO) δ 8.10 (s, J = 8.4 Hz, 1H), 7.39 (s, 2H), 7.08 (dd, J = 17.0, 11.1 Hz, 1H), 6.35 (dd, J = 17.0, 1.8 Hz, 1H), 5.95 (d, J = 7.1 Hz, 1H), 5.72 – 5.59 (m, 2H), 5.34 (d, J = 6.8 Hz, 1H), 5.17 (d, J = 4.6 Hz, 1H), 4.81 (dd, J = 12.5, 6.8 Hz, 1H), 4.16 (td, J = 5.1, 2.4 Hz, 1H), 3.98 (d, J = 2.6 Hz, 1H), 3.68 (dt, J = 12.1, 3.5 Hz, 1H), 3.56 (ddd, J = 11.9, 8.3, 3.5 Hz, 1H); ^{13}C NMR (126 MHz, DMSO) δ 155.91, 152.02, 149.74, 147.71, 124.14, 123.02, 118.90, 87.78, 86.50, 72.21, 70.64, 61.96.



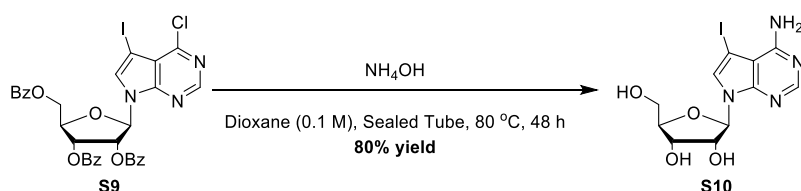
8-VG was prepared according to a modified procedure by Zilbershtein, L. *et al.*⁴³ PdCl_2 (12.2 mg, 0.07 mmol) and TPPTS (0.19 g, 0.35 mmol) was dissolved in a mixture of $\text{H}_2\text{O}:\text{MeCN}$ (2:1, 14 mL) and stirred for 10 min at r.t. 8-Bromoguanosine, **S6** (0.50 g, 1.38 mmol), Potassium vinyltrifluoroborate (0.23 g, 1.73 mmol), and Cs_2CO_3 (1.35 g, 4.14 mmol) were added to the mixture and heated to 90 °C and stirred for 24 h. The mixture was then

allowed to cool to r.t. and concentrated *in vacuo*. The crude mixture was then purified by C18 reverse phase column chromatography (5-95% MeCN/H₂O) to give **8-VG, 6** (0.11 g, 25% yield) as a green solid. HRMS Calcd for C₁₂H₅N₅O₅Na [M+Na⁺] 332.11, found 332.0971 [M+Na⁺]; ¹H NMR (499 MHz, DMSO) δ 10.90 (s, 1H), 6.96 (dd, *J* = 17.1, 11.1 Hz, 1H), 6.52 (s, 2H), 6.12 (dd, *J* = 17.0, 1.8 Hz, 1H), 5.80 (d, *J* = 7.1 Hz, 1H), 5.45 – 5.38 (m, 1H), 5.31 (s, 1H), 5.18 (s, 1H), 5.04 (s, 1H), 4.54 (s, 1H), 4.09 (s, 1H), 3.86 (d, *J* = 3.3 Hz, 1H), 3.65 (d, *J* = 11.3 Hz, 1H), 3.57 (d, *J* = 11.8 Hz, 1H); ¹³C NMR (126 MHz, DMSO) δ 153.66, 152.03, 144.10, 124.81, 119.33, 116.44, 86.78, 85.60, 72.05, 70.04, 61.61.

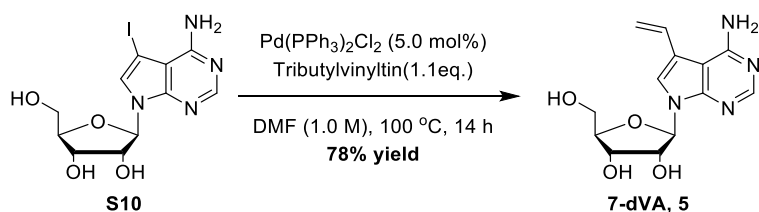


6-Chloro-7-iodo-7-deazapurine, **S8** (2.00 g, 7.16 mmol) was suspended in MeCN (36 mL). BSA (1.93 mL, 7.87 mmol) was added to the suspension and stirred for 20 min at r.t. β-D-Ribofuranose 1-acetate 2,3,5-tribenzoate (3.97 g, 7.87 mmol) and TMSOTf (1.36 mL, 7.51 mmol) were added to the mixture and heated to 80 °C and stirred for 2.5 h. Upon completion, sat. NaHCO₃ solution (40 mL) was added to the mixture, extracted with EtOAc (3 x 50 mL), and dried of MgSO₄. The crude mixture was then concentrated *in vacuo* and purified by silica flash column chromatography (0-50% EtOAc/Hex) to give **S9** (2.61 g, 50%) as a clear foam. HRMS Calcd for C₃₂H₂₃ClIN₃O₇Na [M+Na⁺] 746.06, found 746.0167 [M+Na⁺]; ¹H NMR (400 MHz, CDCl₃) δ 8.57 (s, 1H), 8.14 – 8.08 (m, 2H), 8.03 – 7.95 (m, 2H), 7.95 – 7.89 (m, 2H), 7.65 – 7.46 (m, 6H), 7.42 – 7.32 (m, 4H), 6.68 (d, *J* = 5.4 Hz, 1H), 6.17 (t, *J* = 5.6 Hz, 1H), 6.12 (dd, *J* = 5.7, 4.6 Hz, 1H), 4.90 (dd, *J* = 12.2, 3.1 Hz, 1H), 4.81 (dd, *J* = 7.7,

3.5 Hz, 1H), 4.73 – 4.63 (m, 1H). ^{13}C NMR (126 MHz, CDCl_3) δ 166.16, 165.42, 165.13, 153.16, 151.30, 151.05, 133.86, 133.83, 133.63, 132.17, 129.90, 129.89, 129.77, 129.35, 128.88, 128.75, 128.63, 128.59, 128.45, 117.86, 86.92, 80.73, 74.24, 71.52, 63.56, 53.73.

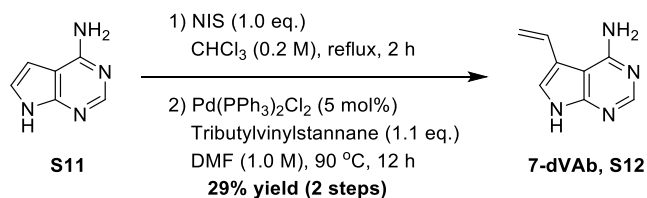


S9 (500 mg, 691 μmol) was placed in a sealed tube with a 1:1 mixture of strong NH_4OH solution and 1,4-Dioxane (12 mL). The tube was then sealed, heated to 80 °C, and stirred for 48 h. The mixture was then cooled to r.t. and concentrated *in vacuo*. The crude material was then purified by silica flash column chromatography (0-10% MeOH/DCM/1% Et_3N) to give **S10** (217 mg, 80%) as a white solid. HRMS Calcd for $\text{C}_{11}\text{H}_{13}\text{IN}_4\text{O}_4\text{Na}$ [$\text{M}+\text{Na}^+$] 414.99, found 414.9879 [$\text{M}+\text{Na}^+$]; ^1H NMR (400 MHz, DMSO) δ 8.10 (s, 1H), 7.67 (s, 1H), 6.66 (s, 2H), 6.02 (d, J = 6.2 Hz, 1H), 5.29 (d, J = 6.4 Hz, 1H), 5.14 (t, J = 5.5 Hz, 1H), 5.09 (d, J = 4.7 Hz, 1H), 4.35 (dd, J = 11.6, 6.1 Hz, 1H), 4.06 (dd, J = 8.1, 4.9 Hz, 1H), 3.88 (q, J = 3.6 Hz, 1H), 3.67 – 3.58 (m, 1H), 3.53 (ddd, J = 11.9, 5.9, 3.9 Hz, 1H); ^{13}C NMR (126 MHz, DMSO) δ 157.26, 151.99, 150.20, 127.18, 103.26, 86.79, 85.23, 73.91, 70.57, 61.59, 52.02.



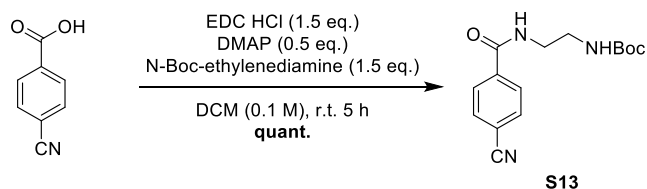
7-dVA was synthesized following a modified procedure by Hottin, A., *et al.*³⁰ **S10** (206 mg, 524 μmol) and $\text{Pd}(\text{PPh}_3)_2\text{Cl}_2$ (18.4 mg, 26.2 μmol) was added to a flask and dissolved in DMF

(0.5 mL) and heated to 100 °C. Tributylvinyltin (169 μ L, 577 μ mol) was added to the flask and stirred for 14 h. The mixture was cooled to r.t. and concentrated *in vacuo*. The crude material was then purified by silica flash column chromatography (0-10% MeOH/DCM/1% Et₃N) and then treated with AcOH and filtered over a pad of silica (10% MeOH/DCM) to give **7-dVA, 5** (161 mg, 78%) as a white solid. HRMS Calcd for C₁₃H₁₆IN₄O₄Na [M+Na⁺] 315.12, found 315.1069 [M+Na⁺]; ¹H NMR (400 MHz, DMSO) δ 8.05 (s, *J* = 24.1 Hz, 1H), 7.66 (s, 1H), 7.11 (dd, *J* = 17.2, 10.9 Hz, 1H), 6.71 (s, 2H), 6.04 (d, *J* = 6.2 Hz, 1H), 5.56 (d, *J* = 17.2 Hz, 1H), 5.24 (d, *J* = 24.6 Hz, 2H), 5.18 – 5.01 (m, 2H), 4.40 (t, *J* = 5.4 Hz, 1H), 4.08 (d, *J* = 3.6 Hz, 1H), 3.89 (d, *J* = 3.4 Hz, 1H), 3.64 (d, *J* = 11.6 Hz, 1H), 3.54 (d, *J* = 11.8 Hz, 1H); ¹³C NMR (126 MHz, DMSO) δ 157.66, 151.48, 150.72, 129.04, 119.15, 114.05, 113.01, 100.79, 87.02, 85.05, 73.68, 70.54, 61.67.

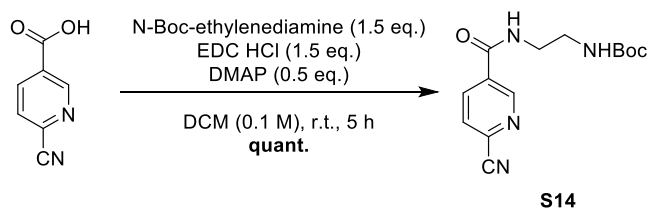


6-Amino-7-deazapurine, **S11** (0.50 g, 3.73 mmol) and NIS (0.84 g, 3.73 mmol) was refluxed in CHCl₃ (15 mL) for 2 h. Upon completion, the mixture was cooled to r.t., filtered and washed with DCM (30 mL) and the precipitate was placed in a flask. Pd(PPh₃)₂Cl₂ (131 mg, 0.19 mmol), tributylvinylstannane (1.19 mL, 4.10 mmol) was added to the flask and dissolved in DMF (4 mL). The mixture was heated to 90 °C and stirred for 12 h. Upon completion, the reaction was allowed to cool to r.t., filtered over celite and concentrated *in vacuo*. The crude precipitate was then purified by silica flash column chromatography (5-10% MeOH/DCM) to give **7-dVAb S12** (173 mg, 29%) as a cream solid. HRMS Calcd for

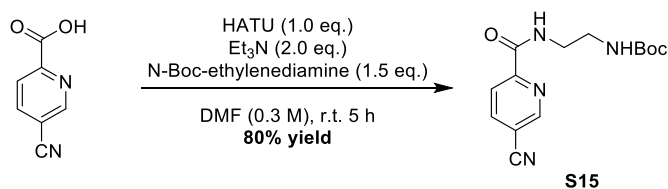
$C_8H_9N_4$ $[M+H]^+$ 161.08, found 161.0827 $[M+H]^+$; 1H NMR (400 MHz, DMSO) δ 11.61 (s, 1H), 8.01 (s, J = 12.1 Hz, 1H), 7.37 (s, 1H), 7.09 (dd, J = 17.2, 10.9 Hz, 1H), 6.53 (s, 2H), 5.53 (d, J = 17.2 Hz, 1H), 5.06 (d, J = 10.9 Hz, 1H); ^{13}C NMR (126 MHz, DMSO) δ 157.51, 151.53, 151.35, 129.63, 118.42, 113.53, 111.95, 100.05.



N-Boc-ethylenediamine (1.61 mL, 10.2 mmol) was added to a mixture of 4-cyanobenzoic acid (1.00 g, 6.79 mmol) and DMAP (415 mg, 3.39 mmol) was dissolved in DCM (70 mL) and stirred for 5 min. EDC HCl (1.95 g, 10.2 mmol) was added to the mixture and stirred at r.t. for 5 h. Upon completion, 1 M citric acid (50 mL) was added to the mixture and separated. The organic layer was then washed with sat. $NaHCO_3$ solution (50 mL) and dried over $MgSO_4$. The crude organic layer was then concentrated *in vacuo* and purified by silica flash column chromatography (5% MeOH/DCM) to give **S13** (2.04 g, quant.) as a white solid. HRMS Calcd for $C_{15}H_{19}N_3O_3Na$ $[M+Na]^+$ 312.14, found 312.1324 $[M+Na]^+$; 1H NMR (400 MHz, $CDCl_3$) δ 7.94 (d, J = 8.3 Hz, 2H), 7.72 (d, J = 8.4 Hz, 2H), 7.64 (s, 1H), 5.02 (s, 1H), 3.56 (dd, J = 10.4, 4.8 Hz, 2H), 3.42 (dd, J = 10.8, 5.9 Hz, 2H), 1.43 (s, 9H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.83, 138.21, 132.48, 127.88, 123.32, 123.31, 118.26, 115.11, 80.57, 43.24, 39.82, 28.45.

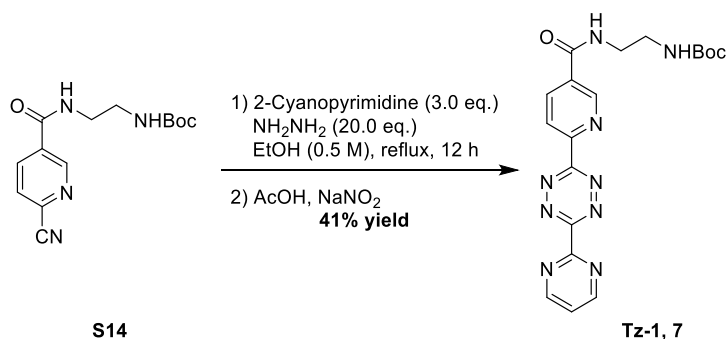


6-cyanonicotinic acid (1.00 g, 6.76 mmol), DMAP (0.49 g, 3.38 mmol), N-Boc-ethylenediamine (1.6 mL, 10.1 mmol) was stirred in DCM (60 mL) for 5 min. EDC HCl (1.94 g, 10.1 mmol) was added to the mixture and stirred for 5 h. The mixture was then added into 1 M citric acid (60 mL) and extracted and washed with sat. NaHCO₃ solution (60 mL). The organic layer was separated, dried over MgSO₄ and concentrated *in vacuo*. The crude product was then purified by silica flash column chromatography (5-10% MeOH/DCM) to give **S14** (1.96 g, quantitative yield) as a white solid. HRMS Calcd for C₁₂H₁₈N₄O₃Na [M+Na⁺] 313.14, found 313.1292 [M+Na⁺]; ¹H NMR (400 MHz, CDCl₃) δ 9.15 (s, 1H), 8.30 (dd, *J* = 8.1, 2.2 Hz, 1H), 8.03 (s, 1H), 7.78 (d, *J* = 8.1 Hz, 1H), 5.03 (s, 1H), 3.58 (dd, *J* = 10.1, 4.6 Hz, 2H), 3.44 (dd, *J* = 10.6, 6.1 Hz, 2H), 1.44 (s, 9H); ¹³C NMR (126 MHz, CDCl₃) δ 149.63, 136.02, 135.77, 132.30, 128.18, 116.77, 80.86, 43.70, 39.63, 28.35.



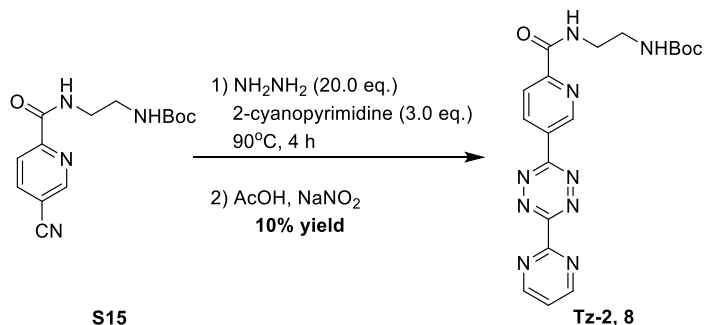
5-cyanopyridine-2-carboxylic acid (1.00 g, 6.75 mmol) was dissolved in DMF (23 mL). N-Boc-ethylenediamine (1.07 mL, 6.75 mmol) and Et₃N (1.88 mL, 13.5 mmol) were added to the mixture and stirred for 5 min. HATU (2.82 g, 7.43 mmol) was added to the mixture and the mixture was stirred at r.t. for 5 h. Upon completion, sat. NaHCO₃ solution (20 mL) was added and extracted with EtOAc (40 mL). The organic layer was then washed with water (3

x 100 mL), dried over MgSO_4 and concentrated *in vacuo*. The crude material was then purified by silica flash column chromatography (50-100% EtOAc/Hex) to give **S15** (1.56 g, 80%) as a white solid. HRMS Calcd for $\text{C}_{14}\text{H}_{18}\text{N}_4\text{O}_3\text{Na}$ $[\text{M}+\text{Na}^+]$ 313.14, found 313.1277 $[\text{M}+\text{Na}^+]$; ^1H NMR (400 MHz, CDCl_3) δ 8.82 (dd, $J = 2.0, 0.9$ Hz, 1H), 8.31 (dd, $J = 8.1, 0.8$ Hz, 2H), 8.13 (dd, $J = 8.1, 2.0$ Hz, 1H), 4.90 (s, 1H), 3.60 (dd, $J = 11.7, 5.9$ Hz, 2H), 3.40 (dd, $J = 11.5, 5.8$ Hz, 2H), 1.41 (s, 9H); ^{13}C NMR (126 MHz, CDCl_3) δ 163.30, 156.60, 152.41, 150.98, 141.03, 122.26, 116.16, 112.32, 79.88, 40.64, 40.28, 28.43.

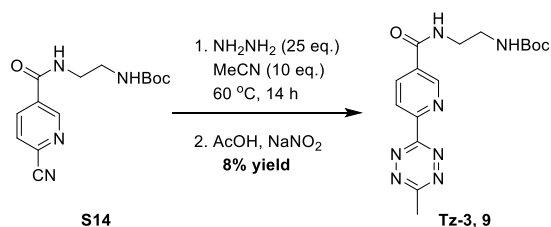


S14 (500 mg, 1.72 mmol) and 2-cyanopyrimidine (1.10 g, 5.17 mmol) was dissolved in EtOH (3 mL). Hydrazine (1.08 mL, 34.4 mmol) was added to the mixture and refluxed for 12 h. The mixture was then allowed to cool to r.t and concentrated *in vacuo*. 1 M NaNO_2 solution (15 mL) was added to the flask and AcOH was added to the mixture slowly until a pH of 5 was reached. The mixture stirred for 20 min and then neutralized with saturated NaHCO_3 solution, extracted with DCM (3 x 25 mL), and dried over MgSO_4 . The organic mixture was then concentrated *in vacuo* and purified by silica flash column chromatography (0-5% MeOH/DCM). The pink residue was then further purified with recrystallization in EtOH to give **Tz-1, 7** (298 mg, 41%) as magenta crystals. HRMS Calcd for $\text{C}_{19}\text{H}_{21}\text{N}_9\text{O}_3\text{Na}$ $[\text{M}+\text{Na}^+]$ 446.18, found 446.1665 $[\text{M}+\text{Na}^+]$; ^1H NMR (499 MHz, CDCl_3) δ

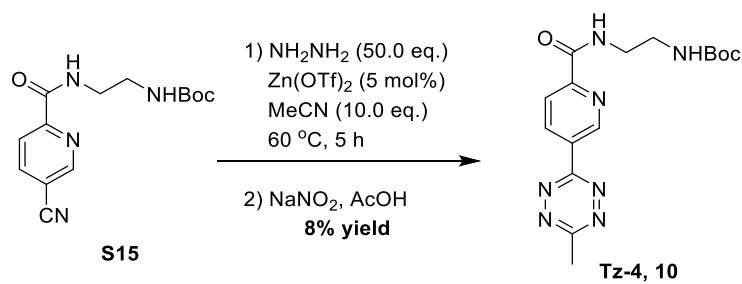
9.43 (s, 1H), 9.17 (d, $J = 4.8$ Hz, 2H), 8.87 (d, $J = 8.2$ Hz, 1H), 8.49 (dd, $J = 8.2, 2.2$ Hz, 1H), 7.98 (s, 1H), 7.62 (t, $J = 4.9$ Hz, 1H), 5.04 (s, 1H), 3.71 – 3.61 (m, 2H), 3.48 (s, 2H), 1.45 (s, $J = 4.3$ Hz, 9H); ^{13}C NMR (126 MHz, CDCl_3) δ 164.80, 163.76, 163.66, 159.48, 158.65, 157.64, 151.83, 149.78, 136.76, 132.37, 124.72, 122.83, 80.63, 43.30, 39.95, 28.47.



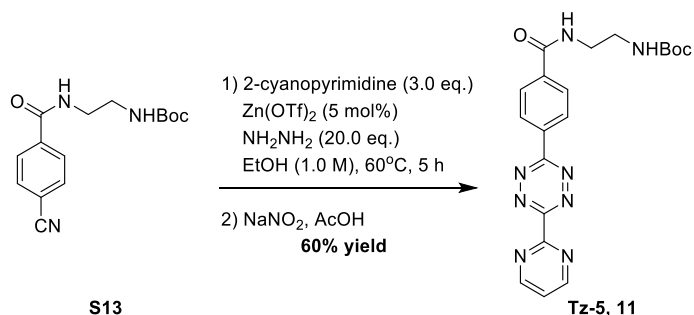
S15 (500 mg, 2.38 mmol) and 2-cyanopyrimidine (750 mg, 7.13 mmol) was dissolved in hydrazine (1.49 mL, 47.5 mmol) and stirred at 90 °C for 4 h. The mixture was then allowed to cool to r.t. and 20 mL of 1 M NaNO_2 solution was added to the mixture. AcOH was added to the mixture slowly until a pH of 5 was reached. The mixture stirred for 10 min and then neutralized with saturated NaHCO_3 solution, extracted with DCM (3 x 25 mL), and dried over MgSO_4 . The organic mixture was then concentrated *in vacuo* and purified by silica flash column chromatography (0-50% Acetone/DCM). The pink residue was then further purified with recrystallization in EtOH to give **Tz-2, 8** (72.9 mg, 10%) as purple crystals. HRMS Calcd for $\text{C}_{19}\text{H}_{21}\text{N}_9\text{O}_3\text{Na}$ [$\text{M}+\text{Na}^+$] 446.18, found 446.1665 [$\text{M}+\text{Na}^+$]; ^1H NMR (400 MHz, CDCl_3) δ 9.84 (s, 1H), 9.19 – 9.09 (m, $J = 5.8$ Hz, 3H), 8.45 (d, $J = 8.2$ Hz, 2H), 7.62 (t, $J = 4.8$ Hz, 1H), 4.99 (s, 1H), 3.65 (m, $J = 5.8$ Hz, 2H), 3.44 (m, $J = 5.4$ Hz, 2H), 1.43 (s, 9H); ^{13}C NMR (126 MHz, CDCl_3) δ 163.52, 163.36, 159.50, 159.29, 158.56, 153.04, 148.45, 137.34, 129.59, 122.79, 122.60, 122.36, 79.68, 40.31, 29.32, 28.39.



S14 (1.14 g, 3.92 mmol) was dissolved in MeCN (2.05 mL, 39.2 mmol) and heated to 60 °C. Hydrazine (3.08 mL, 98.0 mmol) was added to the mixture and stirred for 14 h. The mixture was then allowed to cool to r.t. and AcOH (4 mL) was added to the mixture. A 3 M NaNO₂ solution (30 mL) was added slowly to the mixture. The pH was adjusted to 5 using AcOH and stirred until bubbles ceased. The mixture was then diluted with DCM (40 mL) and washed with sat. NaHCO₃ solution. The organic layer was separated, dried over MgSO₄, and concentrated *in vacuo*. The impure residue was then purified by silica flash column chromatography (0-40% Acetone/DCM) give **Tz-3, 9** (0.11 g, 8% yield) as a pink solid. HRMS Calcd for C₁₆H₂₂N₇O₃ [M+H⁺] 360.18, found 360.1798 [M+H⁺]; ¹H NMR (400 MHz, CDCl₃) δ 9.36 (s, 1H), 8.72 (d, *J* = 8.2 Hz, 1H), 8.44 (dd, *J* = 8.2, 2.1 Hz, 1H), 7.91 (s, 1H), 5.02 (s, 1H), 3.63 (dd, *J* = 10.1, 4.7 Hz, 2H), 3.51 – 3.41 (m, 2H), 3.18 (s, 3H), 1.45 (s, 9H); ¹³C NMR (126 MHz, CDCl₃) δ 173.67, 172.25, 168.34, 163.32, 149.42, 136.55, 134.25, 131.88, 123.54, 80.62, 53.46, 39.85, 28.38, 21.45.

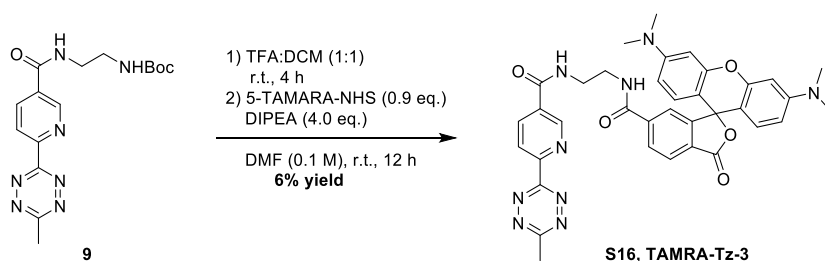


S15 (500 mg, 1.72 mmol), Zn(OTf)₂ (31.8 mg, 86.1 μmol) were dissolved in MeCN (0.89 mL, 17.2 mmol) and heated to 60 °C. Hydrazine (2.70 mL, 86.1 mmol) was added to the mixture and stirred for 5 h. The mixture was then allowed to cool to r.t. and 40 mL of 1 M NaNO₂ solution was added to the mixture. AcOH was added to the mixture slowly until a pH of 5 was reached. The mixture stirred for 10 min and was then neutralized with saturated NaHCO₃ solution, extracted with DCM (3 x 40 mL), and dried over MgSO₄. The organic mixture was then concentrated *in vacuo* and purified by silica flash column chromatography (50-80% EtOAc/Hex) to give **Tz-4, 10** (33.7 mg, 8%) as a pink solid. HRMS Calcd for C₁₆H₂₁N₇O₃Na [M+Na⁺] 383.17, found 382.1604 [M+Na⁺]; ¹H NMR (400 MHz, CDCl₃) δ 9.72 (d, *J* = 2.1 Hz, 1H), 9.01 (dd, *J* = 8.2, 2.1 Hz, 1H), 8.41 (d, *J* = 8.2 Hz, 1H), 4.94 (s, 1H), 3.64 (dd, *J* = 11.5, 5.8 Hz, 3H), 3.44 (dd, *J* = 11.1, 5.6 Hz, 3H), 3.16 (s, 3H), 1.43 (s, 9H); ¹³C NMR (151 MHz, CDCl₃) δ 168.30, 164.38, 162.86, 156.50, 152.63, 147.81, 136.66, 130.17, 122.60, 79.80, 40.62, 40.37, 28.49, 21.50.



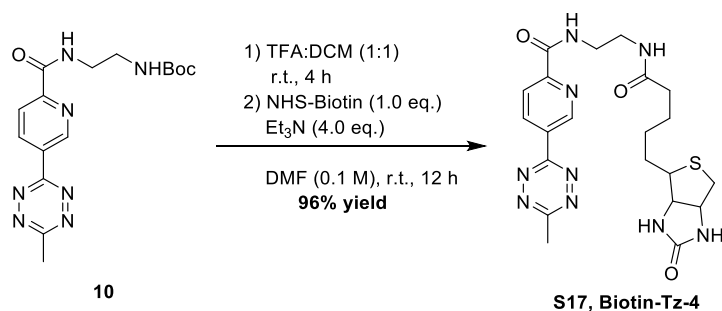
S13 (500 mg, 1.73 mmol), 2-cyanopyrimidine (545 mg, 5.18 mmol), Zn(OTf)₂ (31.4 mg, 0.09 mmol) were dissolved in EtOH (1.7 mL). Hydrazine (1.09 mL, 34.6 mmol) was added to the mixture and heated to 60 °C and stirred for 5 h. The mixture was then allowed to cool to r.t. and 15 mL of 1 M NaNO₂ solution was added to the mixture. AcOH was added to the mixture slowly until a pH of 5 was reached. The mixture stirred for 10 min and then

neutralized with saturated NaHCO₃ solution, extracted with DCM (3 x 20 mL), and dried over MgSO₄. The organic mixture was then concentrated *in vacuo* and purified by silica flash column chromatography (0-5% MeOH/DCM) to give **Tz-5, 11** (0.44 g, 60%) as a purple solid. HRMS Calcd for C₂₀H₂₂N₈O₃Na [M+Na⁺] 445.18, found 445.1713 [M+Na⁺]; ¹H NMR (400 MHz, CDCl₃) δ 9.14 (d, *J* = 4.9 Hz, 2H), 8.79 (d, *J* = 8.6 Hz, 2H), 8.09 (d, *J* = 8.4 Hz, 2H), 7.59 (t, *J* = 4.9 Hz, 1H), 5.08 (s, *J* = 10.7 Hz, 1H), 3.61 (dd, *J* = 10.6, 4.9 Hz, 2H), 3.46 (d, *J* = 4.7 Hz, 2H), 1.44 (s, *J* = 5.8 Hz, 9H); ¹³C NMR (126 MHz, CDCl₃) δ 166.80, 164.23, 163.33, 159.63, 158.58, 158.03, 138.56, 133.95, 129.05, 128.18, 122.70, 80.42, 42.86, 39.95, 28.49.



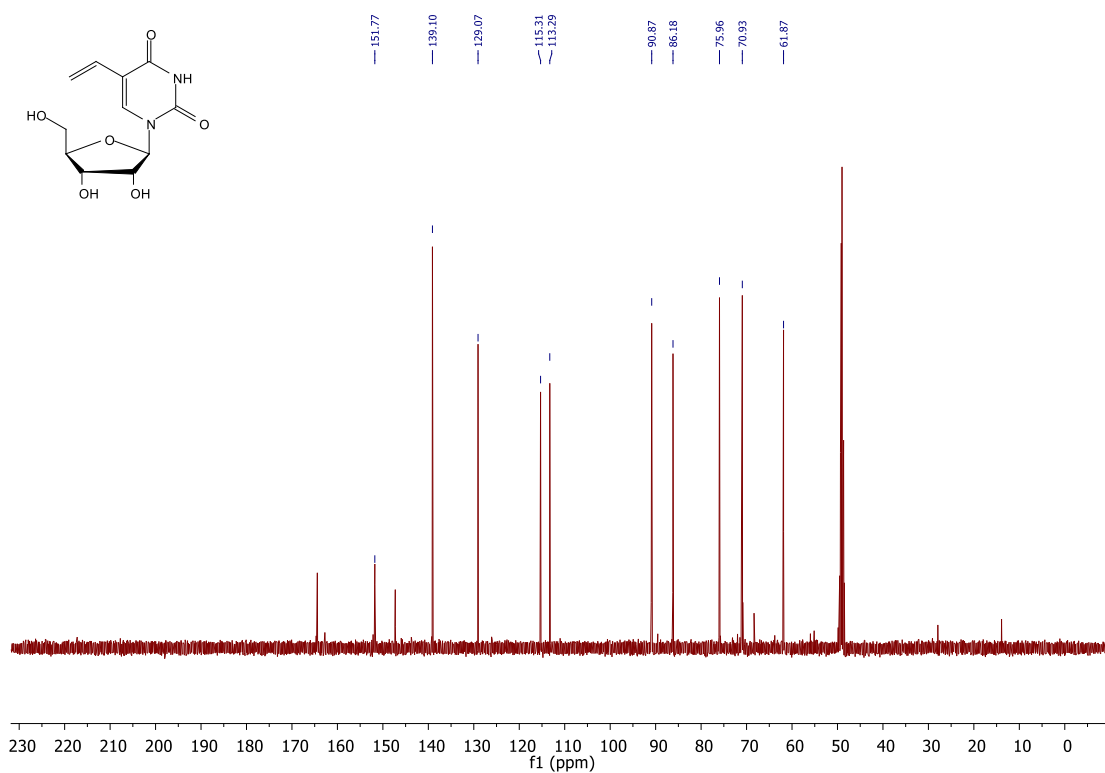
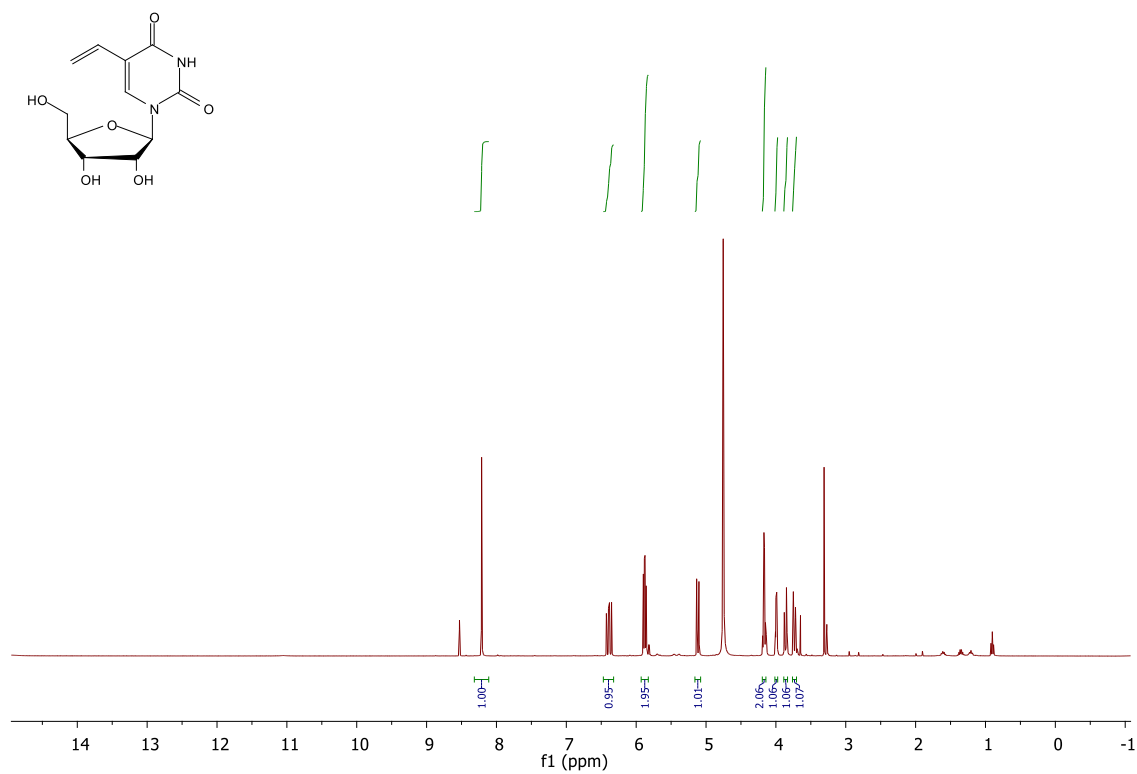
Tz-3, 9 (25.0 mg, 69.6 μmol) was dissolved in a 1:1 TFA:DCM solution (1 mL) and stirred at r.t., for 4 h. The mixture was then concentrated *in vacuo*. 5-TAMARA-NHS ester (66.0 mg, 62.6 μmol) was dissolved in DMF (0.6 mL) and stirred at r.t. DIPEA (48.5 μL, 27.8 μmol) was added to the mixture and stirred for 12 h. The mixture was concentrated *in vacuo* and the crude product was purified by reverse phase C18 column chromatography (30-70% MeCN/H₂O with 0.1% TFA) to give **S16, TAMRA-Tz-3** (2.5 mg, 6% yield) as a red solid. HRMS Calcd for C₃₆H₃₄N₉O₅ [M+H⁺] 672.27, found 672.2689 [M+H⁺]; ¹H NMR (400 MHz, CD₃OD) δ 9.23 (s, 1H), 9.00 (s, 1H), 8.76 (s, 1H), 8.74 (s, 1H), 8.50 (dd, *J* = 8.3, 2.2 Hz, 1H), 8.25 (dd, *J* = 7.9, 1.8 Hz, 1H), 7.51 (d, *J* = 8.0 Hz, 1H), 7.12 (d, *J* = 9.5 Hz, 2H), 7.05 (dd, *J* = 9.5, 2.4 Hz, 2H), 6.97 (d, *J* = 2.4 Hz, 2H), 3.80 – 3.71 (m, 4H), 3.10 (s, 3H); ¹³C NMR (126 MHz,

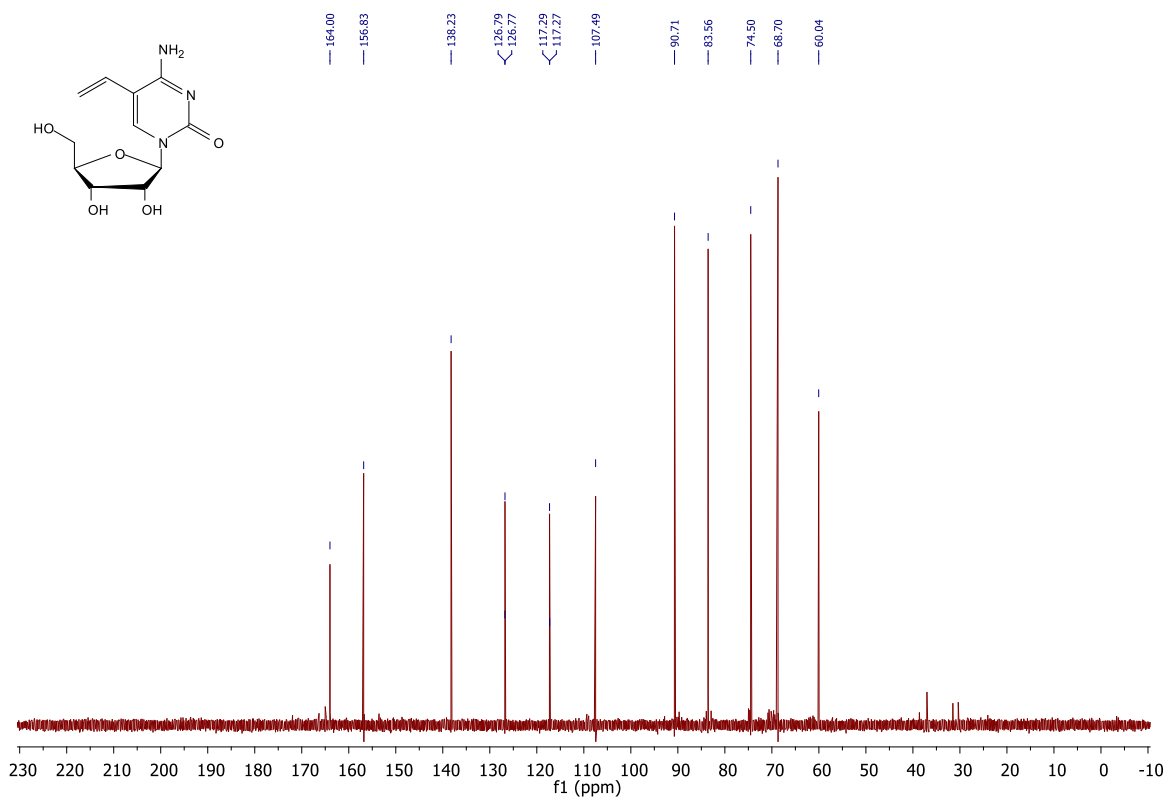
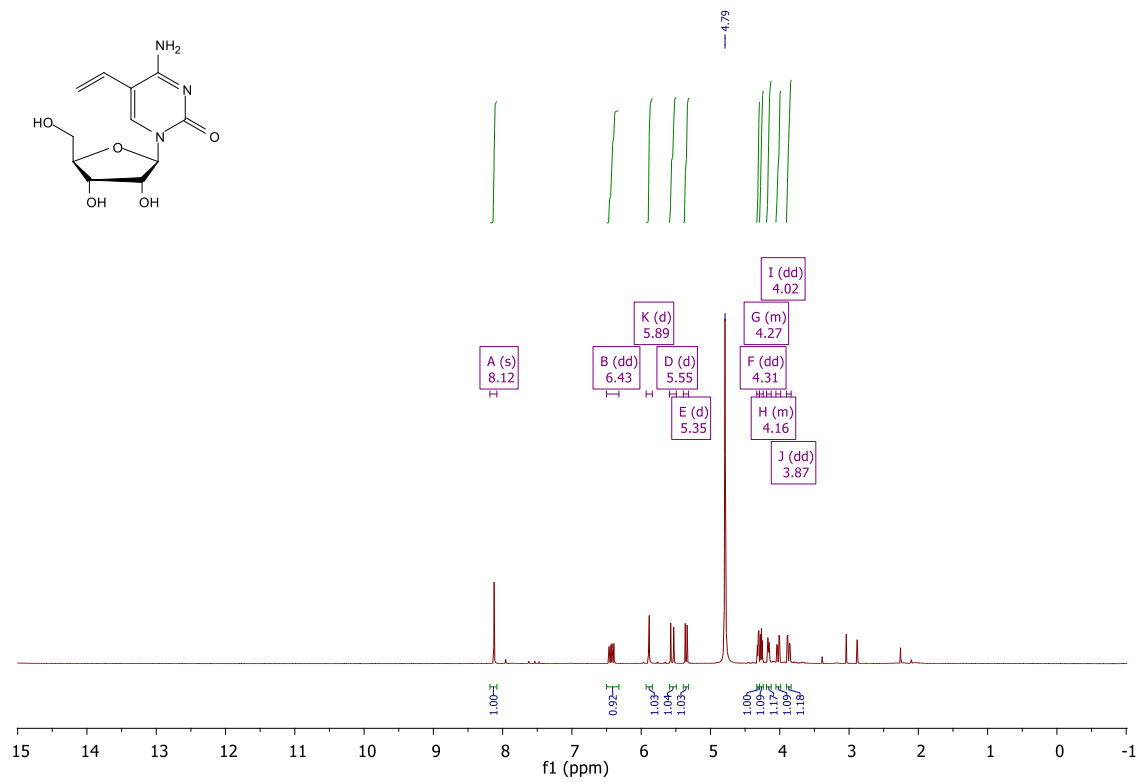
CD₃OD) δ 169.96, 168.73, 167.76, 164.19, 160.67, 159.08, 159.01, 153.64, 150.31, 138.25, 138.15, 137.72, 133.90, 132.97, 132.33, 131.93, 131.37, 124.65, 115.59, 114.74, 97.45, 40.93, 21.31.

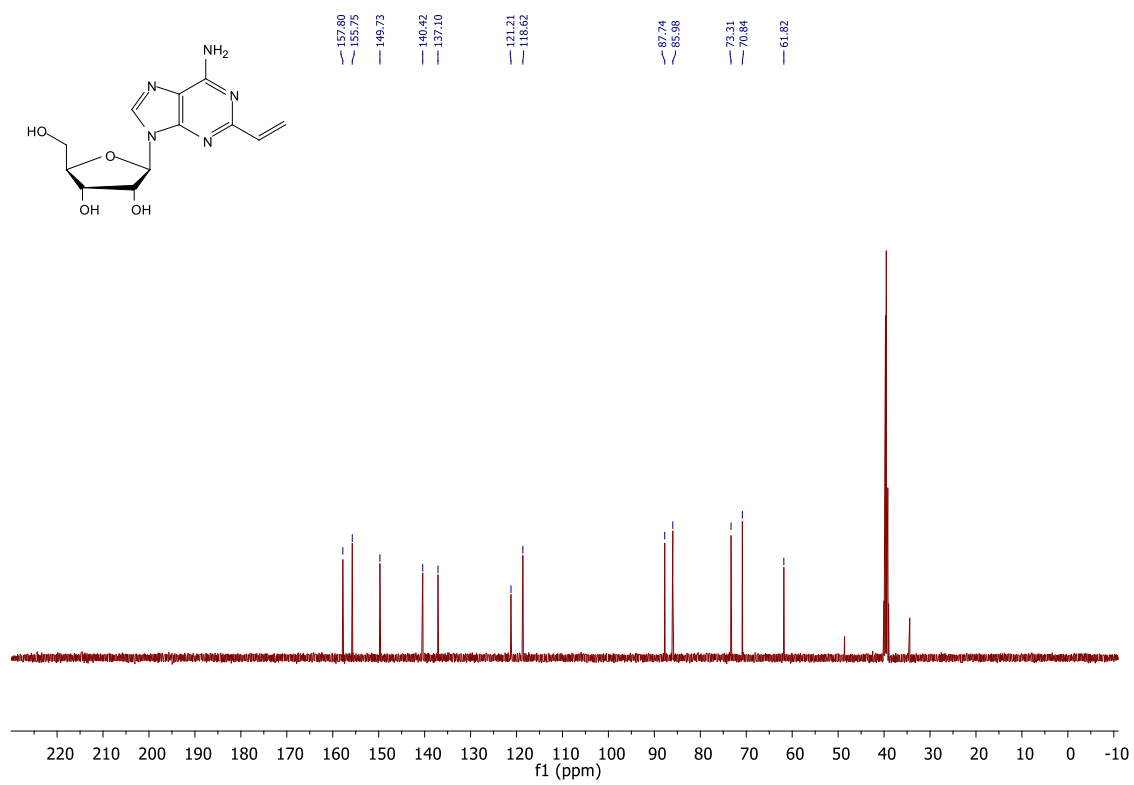
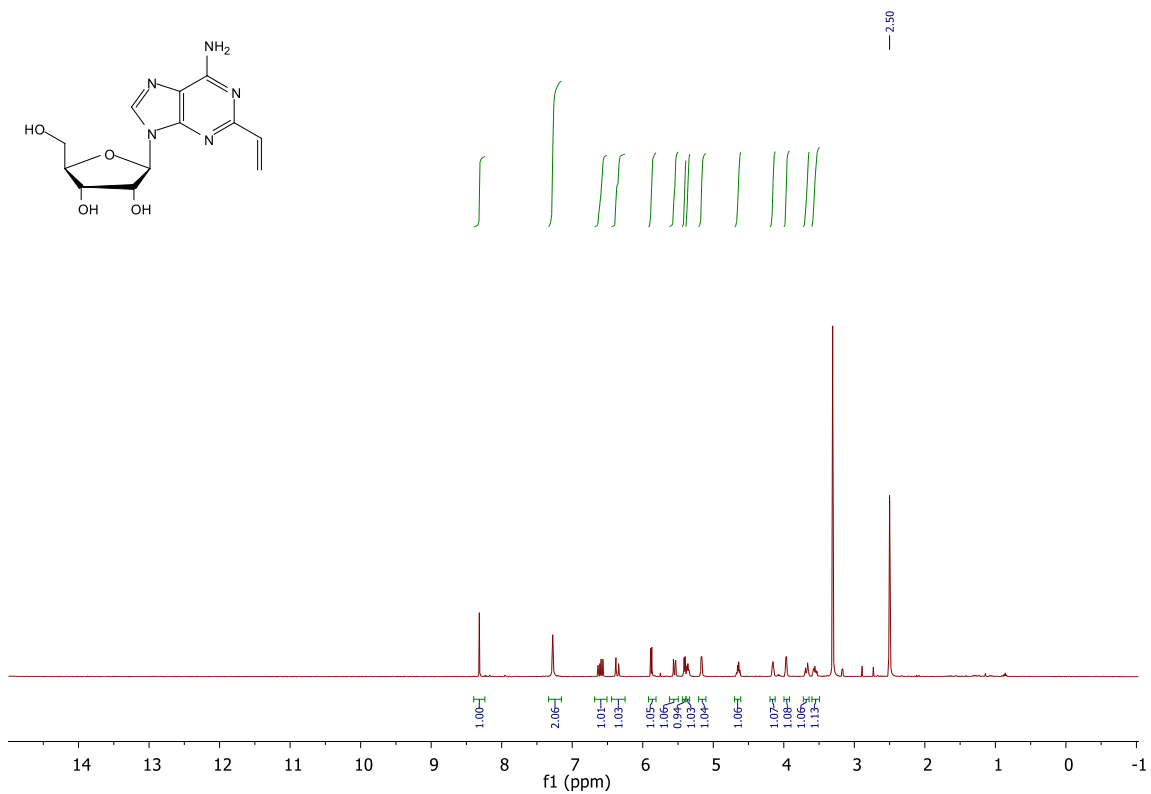


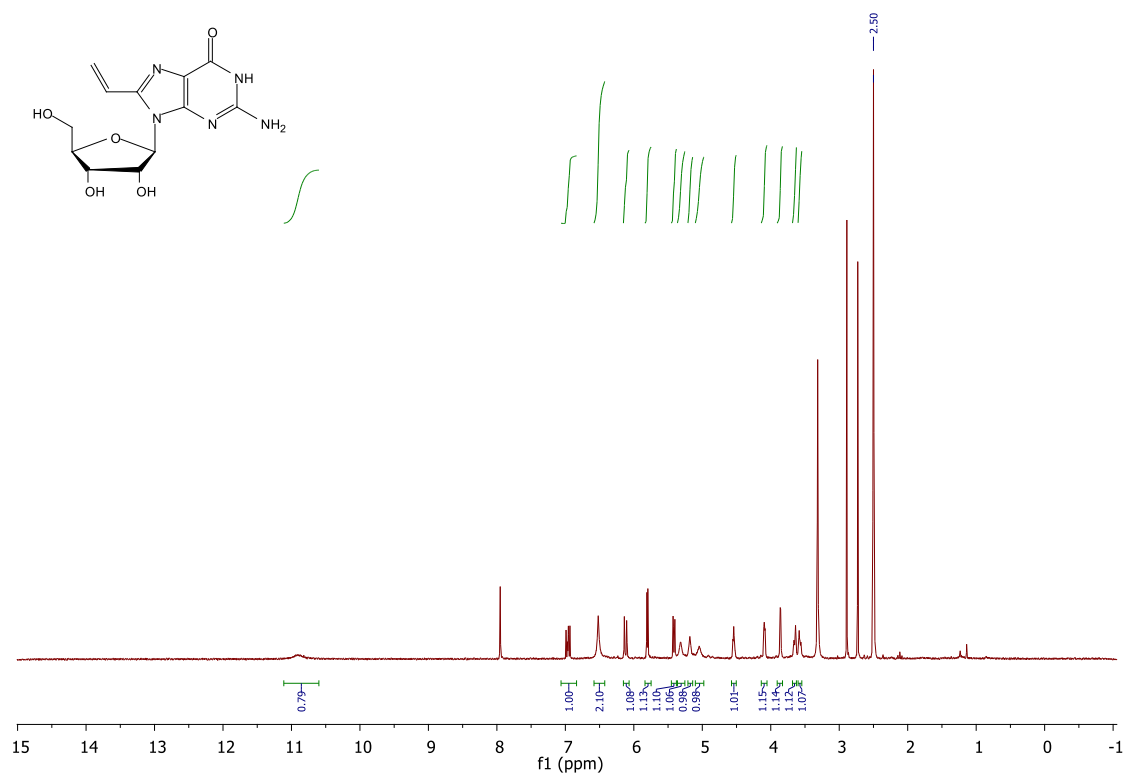
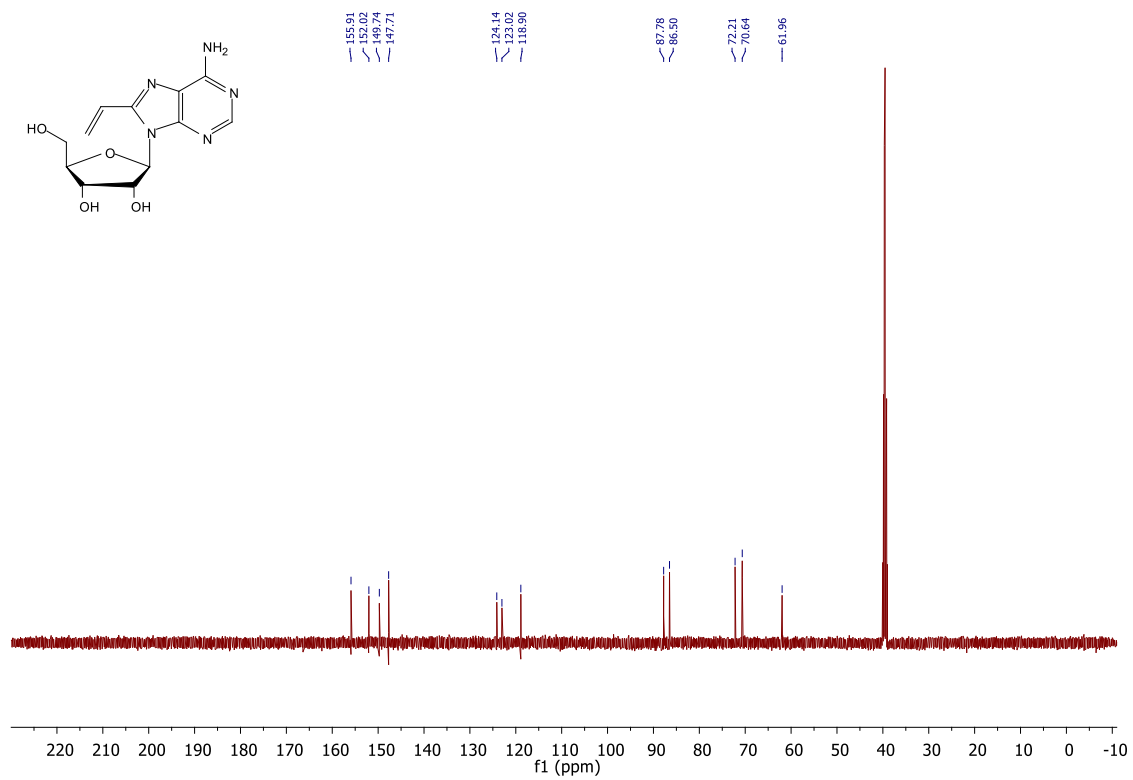
Tz-4, 10 (19 mg, 52.9 μ mol) was dissolved in TFA:DCM (1:1, 1 mL) and stirred at r.t. for 4 h. The mixture was then concentrated *in vacuo*. NHS-Biotin (18 mg, 52.9 μ mol) was added to the flask and dissolved in DMF (0.5 mL). Et₃N (30 μ L, 0.21 mmol) was added to the mixture and stirred at r.t. for 12 h. The mixture was then concentrated *in vacuo* and purified by C18 reverse phase column chromatography (0-50% MeCN/H₂O with 0.1% TFA) to give **S17, Biotin-Tz-4** (24.7 mg, 96% yield) as a pink solid. HRMS Calcd for C₂₁H₂₇N₉O₃SNa [M+Na⁺] 508.18, found 508.1850 [M+Na⁺]; ¹H NMR (400 MHz, CD₃OD) δ 9.68 (s, 1H), 9.00 (d, *J* = 8.2 Hz, 1H), 8.30 (s, 1H), 7.25 (d, *J* = 8.9 Hz, 1H), 6.84 (d, *J* = 8.9 Hz, 1H), 4.42 (dd, *J* = 7.7, 4.7 Hz, 2H), 4.22 (dd, *J* = 7.8, 4.5 Hz, 2H), 3.56 (s, 2H), 3.45 (d, *J* = 5.2 Hz, 2H), 3.09 (s, 1H), 3.06 (s, 3H), 2.89 – 2.80 (m, 1H), 2.63 (d, *J* = 12.7 Hz, 1H), 2.19 (t, *J* = 7.3 Hz, 2H), 1.73 – 1.43 (m, 6H), 1.37 (dd, *J* = 15.1, 7.3 Hz, 3H); ¹³C NMR (126 MHz, CD₃OD) δ 169.58, 158.80, 127.17, 119.44, 117.18, 115.62, 115.43, 115.14, 114.92, 112.65, 63.28, 61.62, 56.88, 41.01, 39.87, 36.79, 31.95, 29.61, 29.44, 26.80, 21.26.

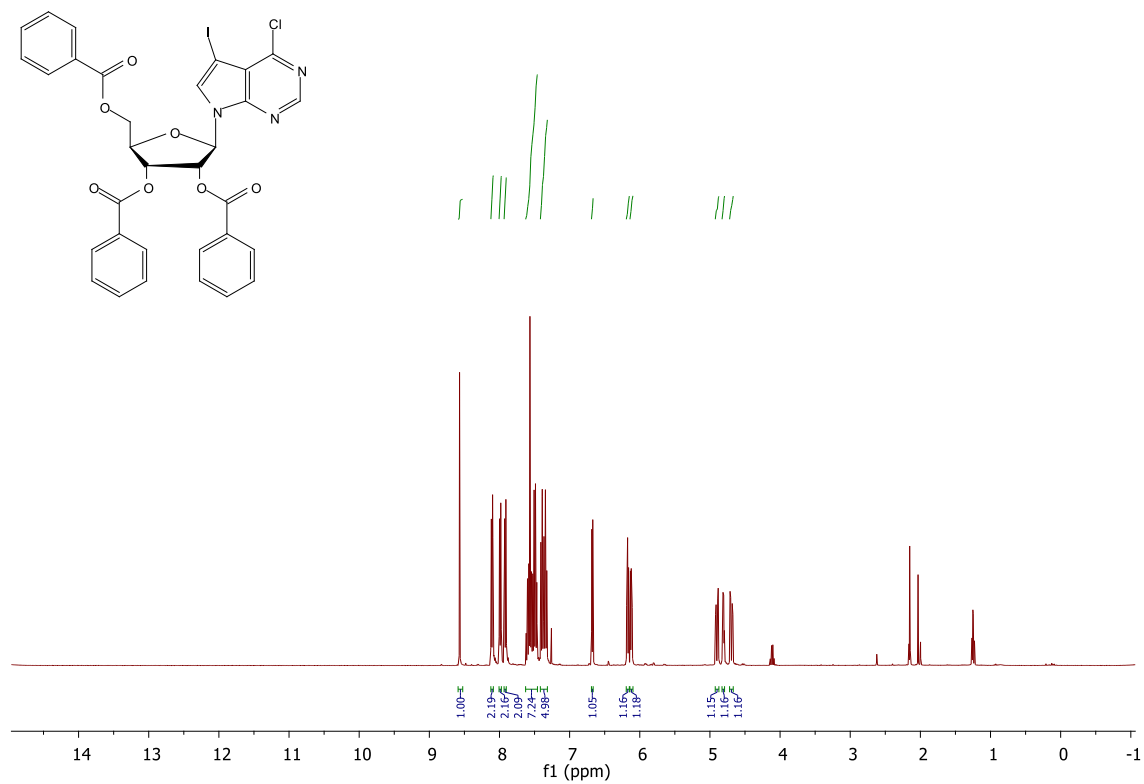
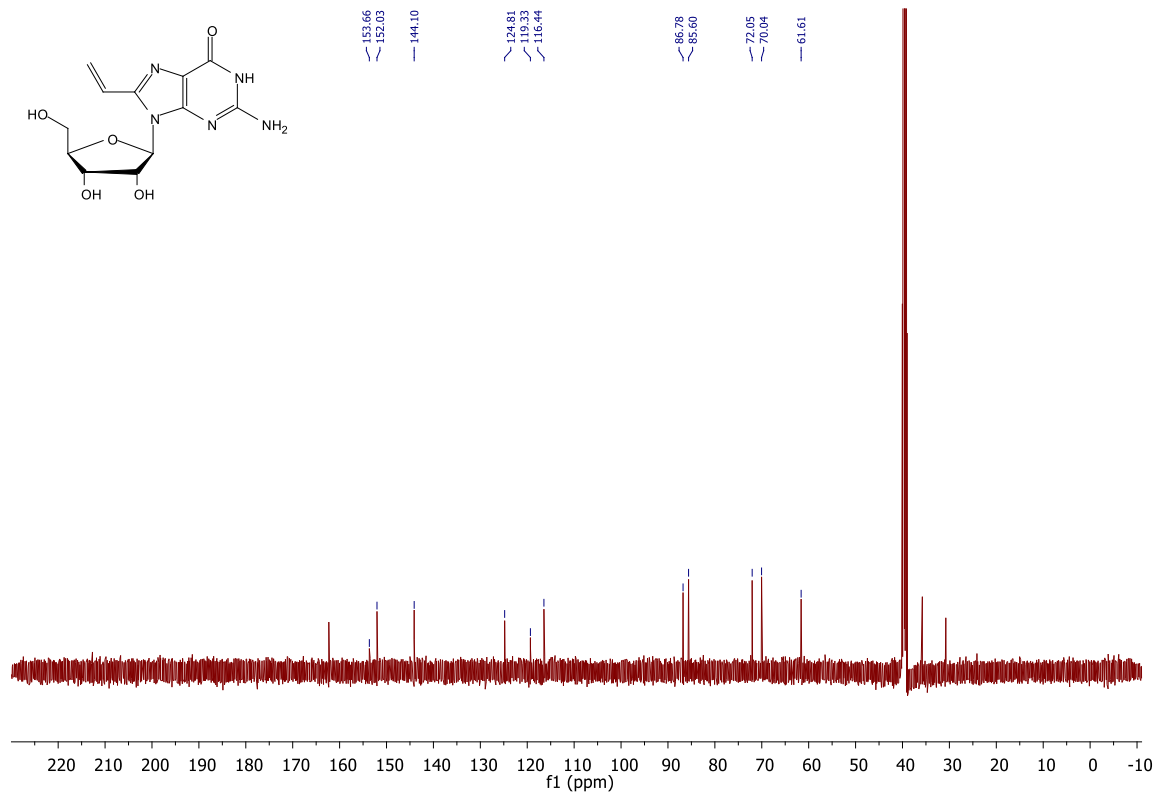
Spectra

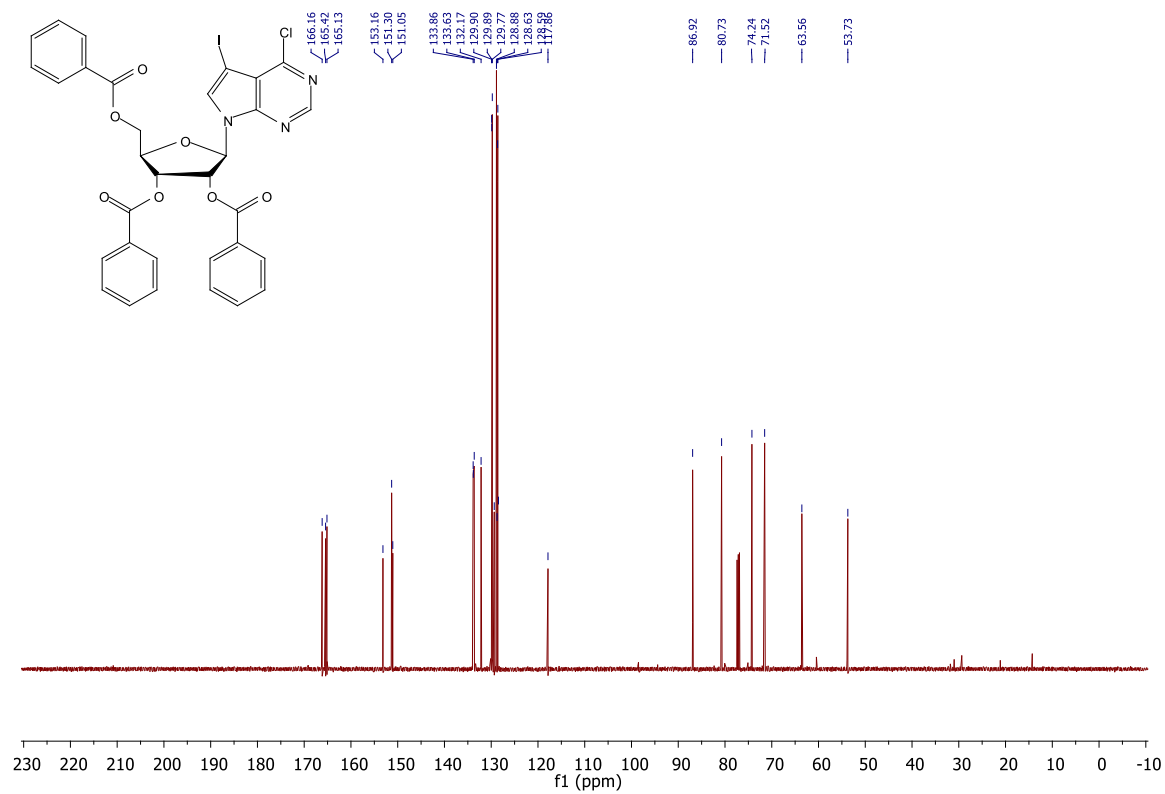
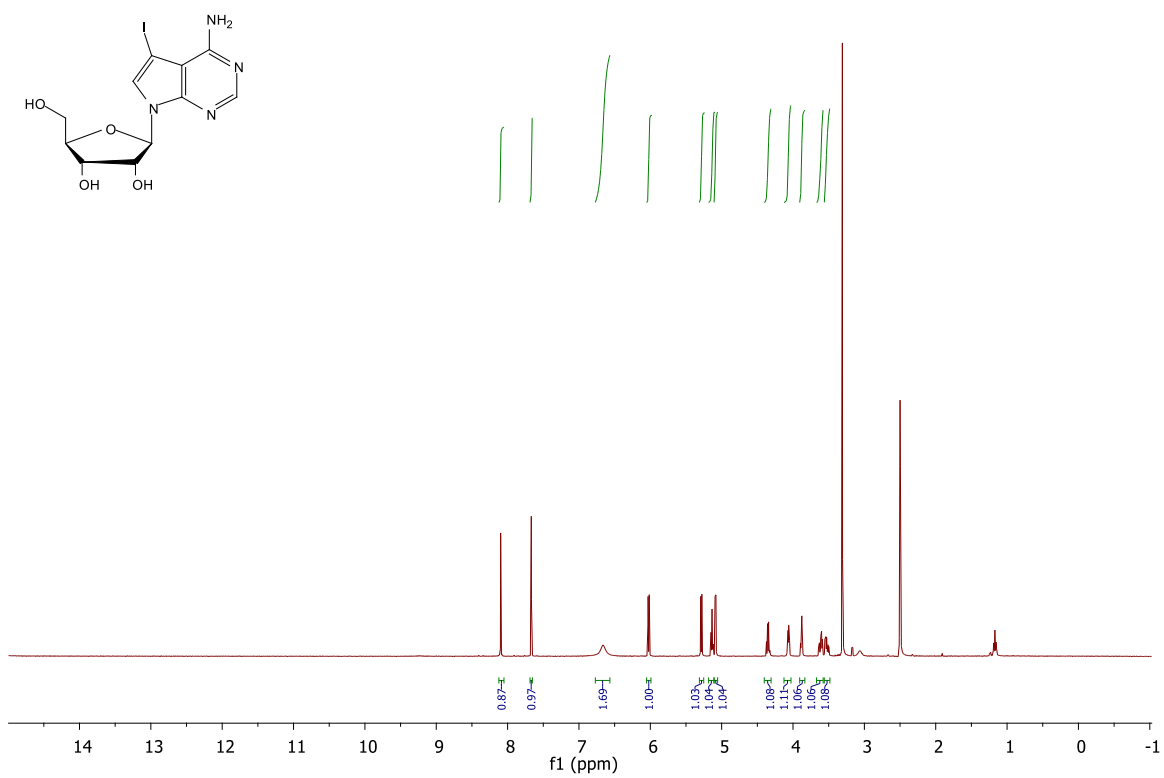


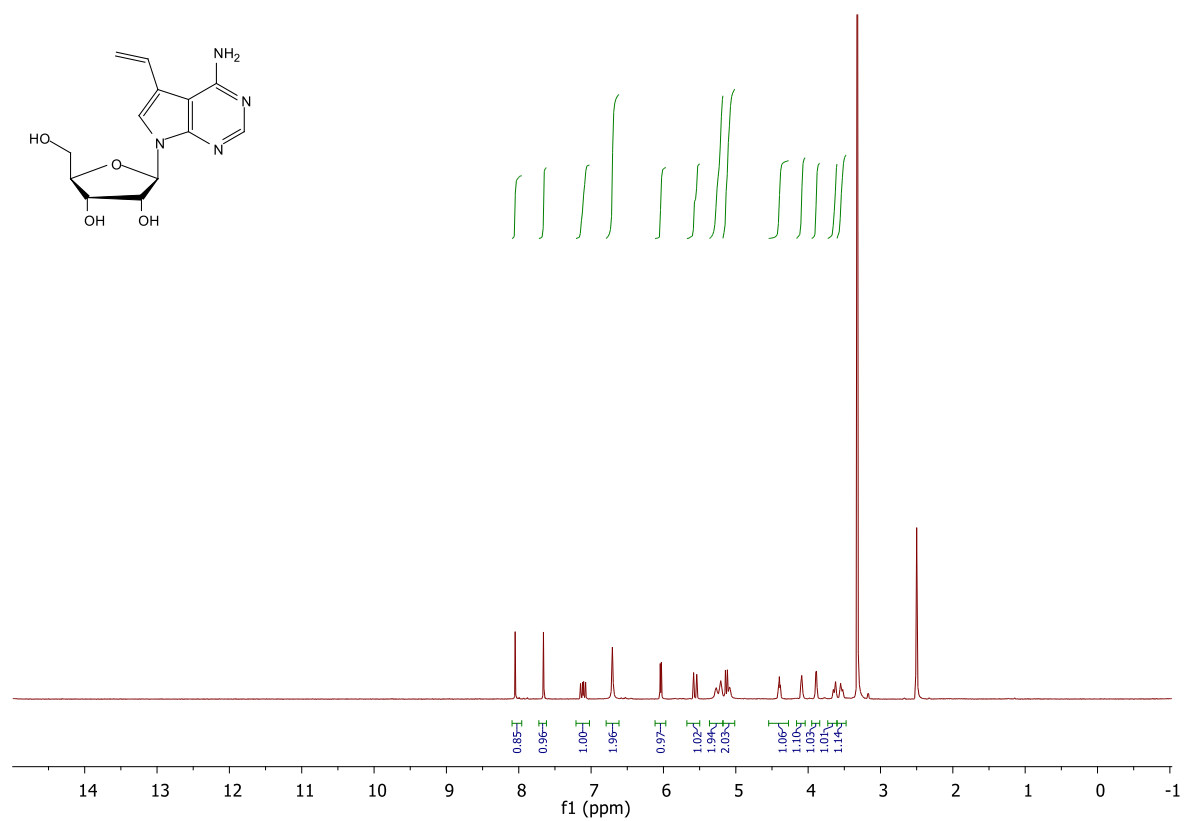
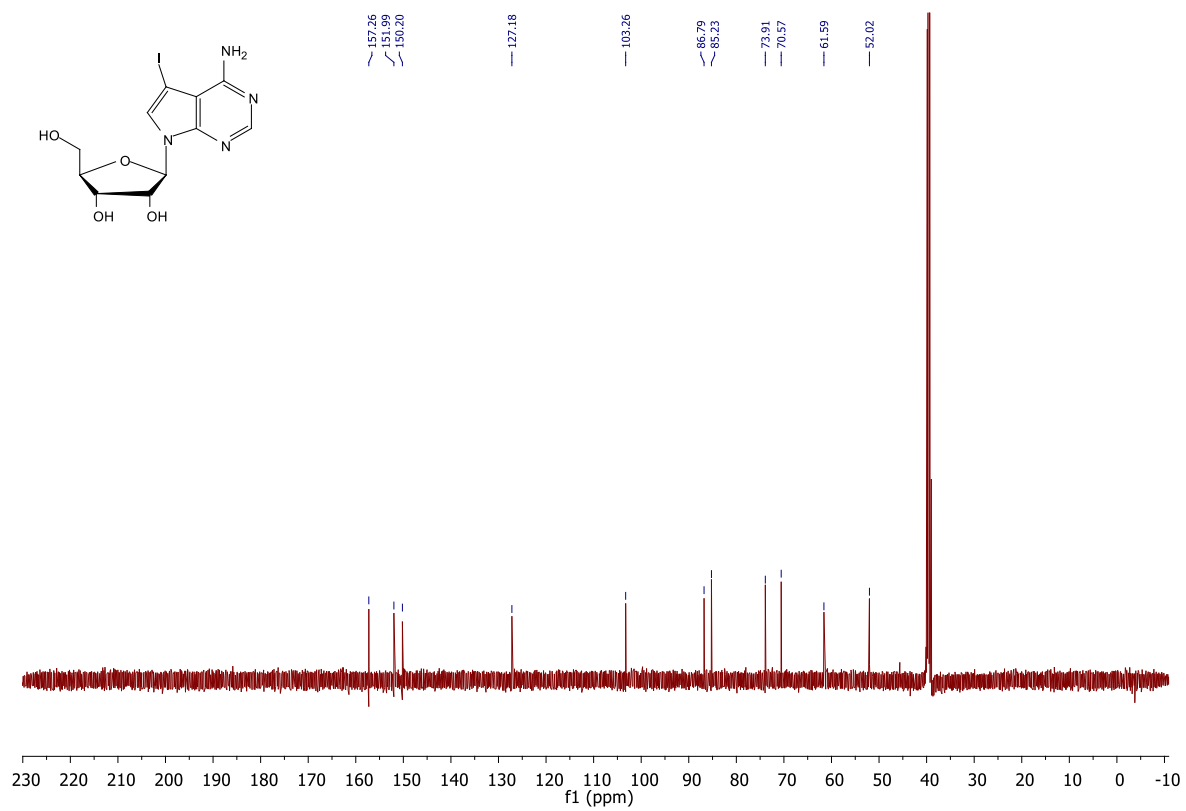


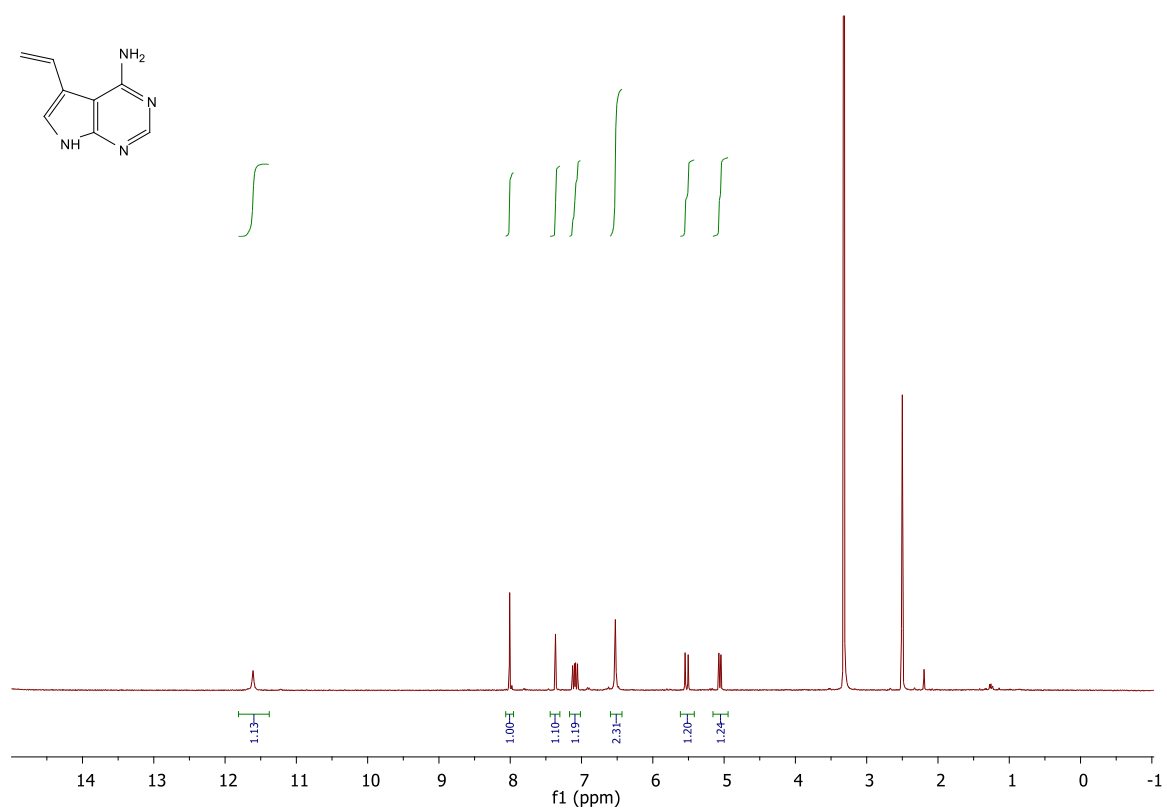
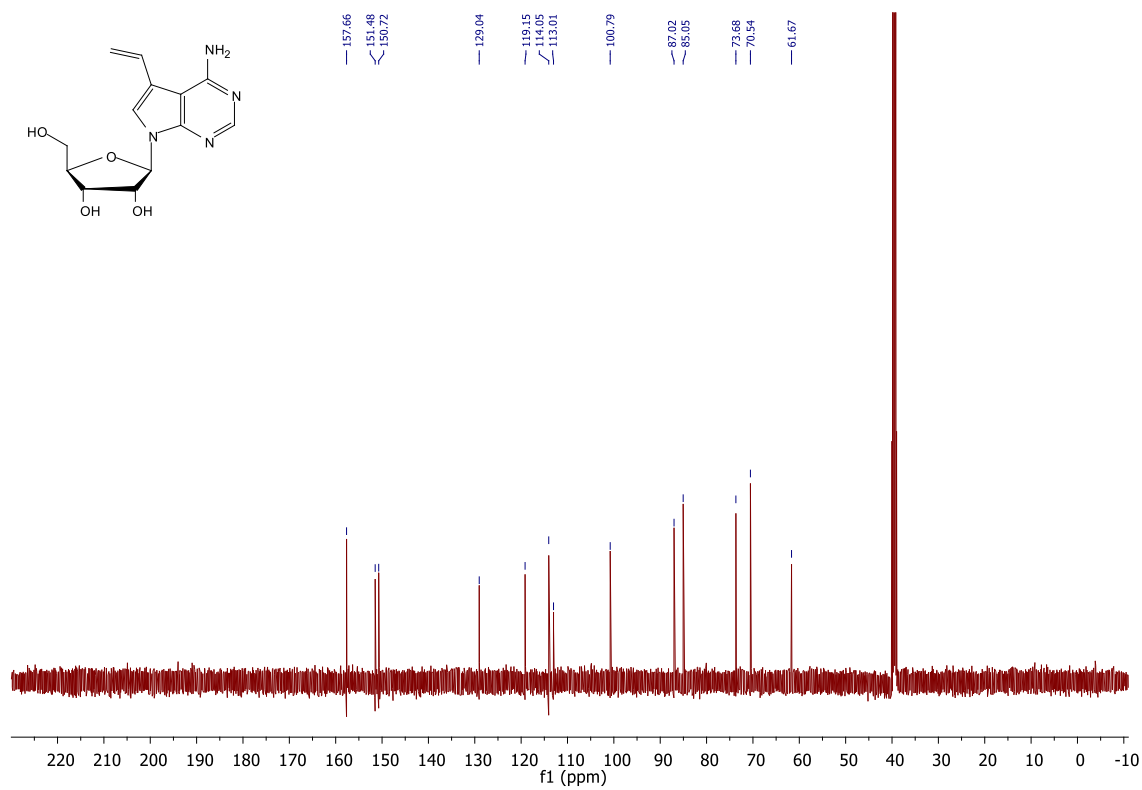


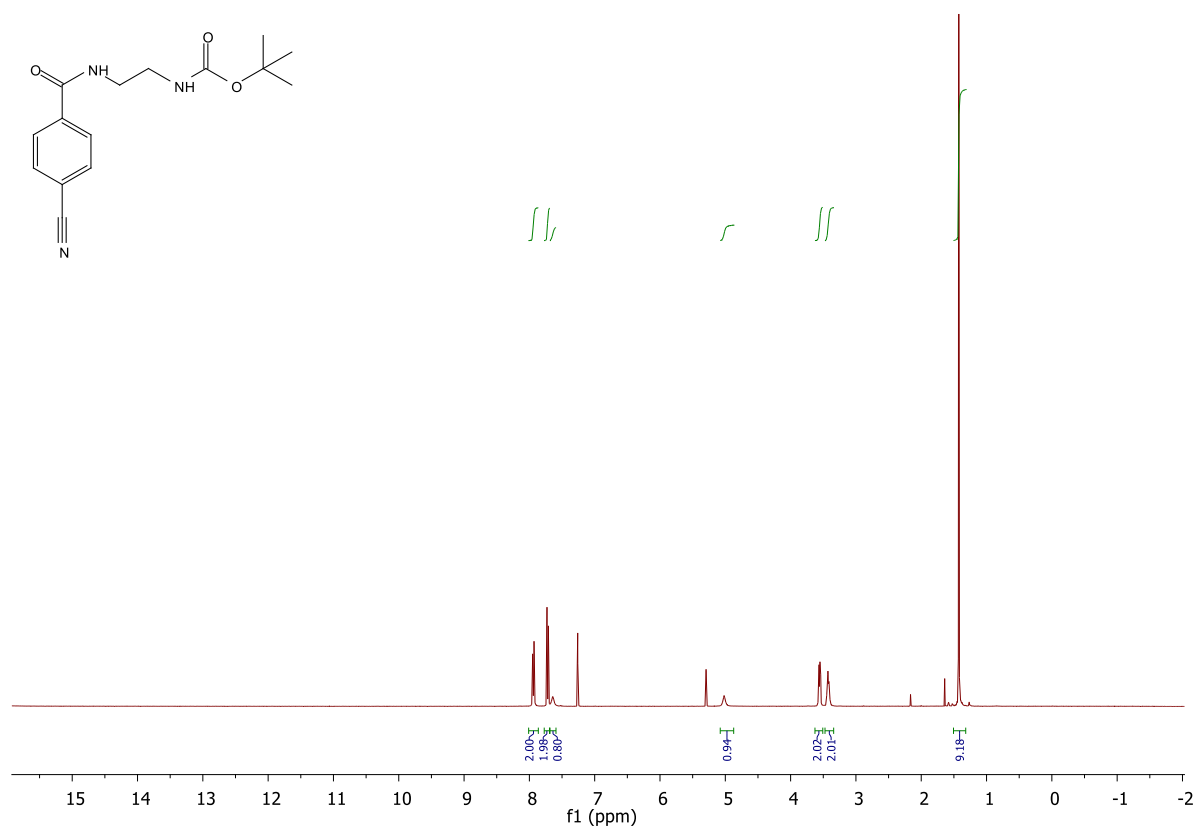
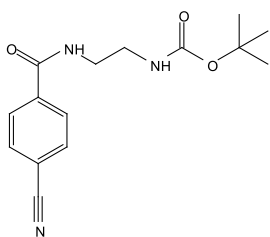
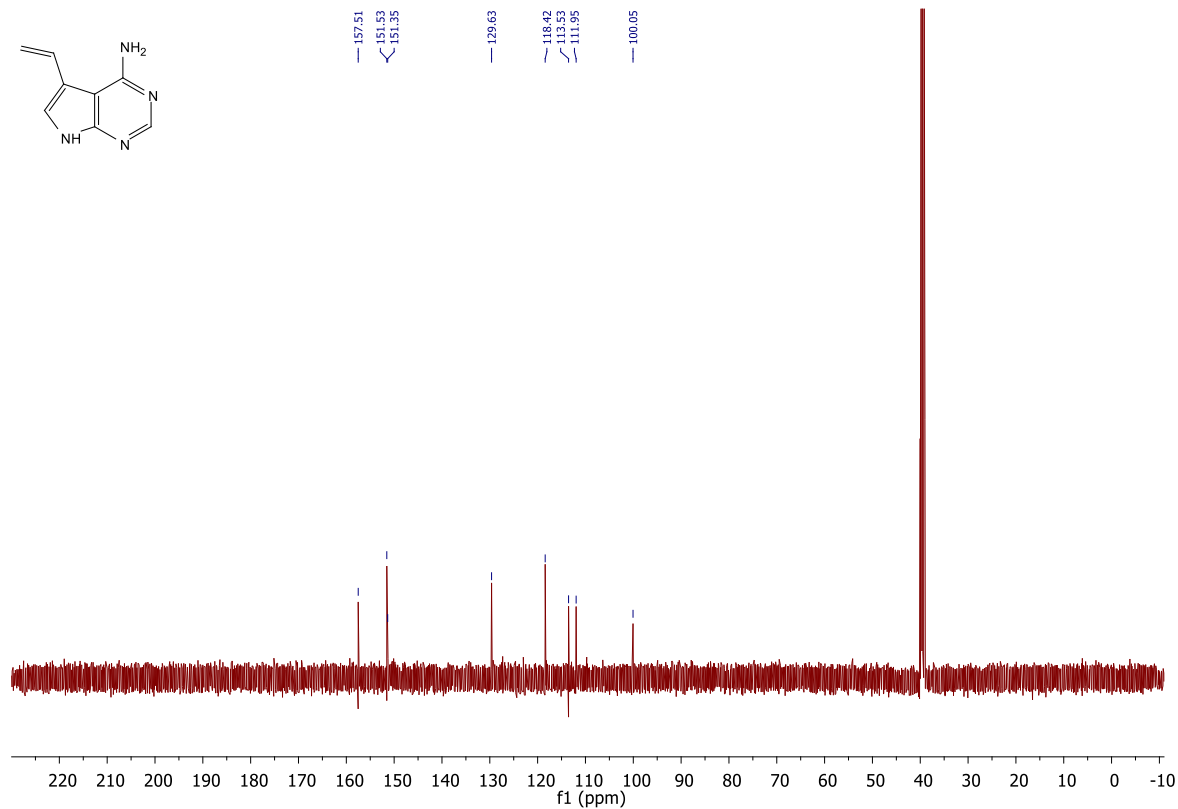
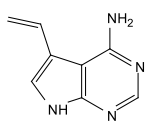


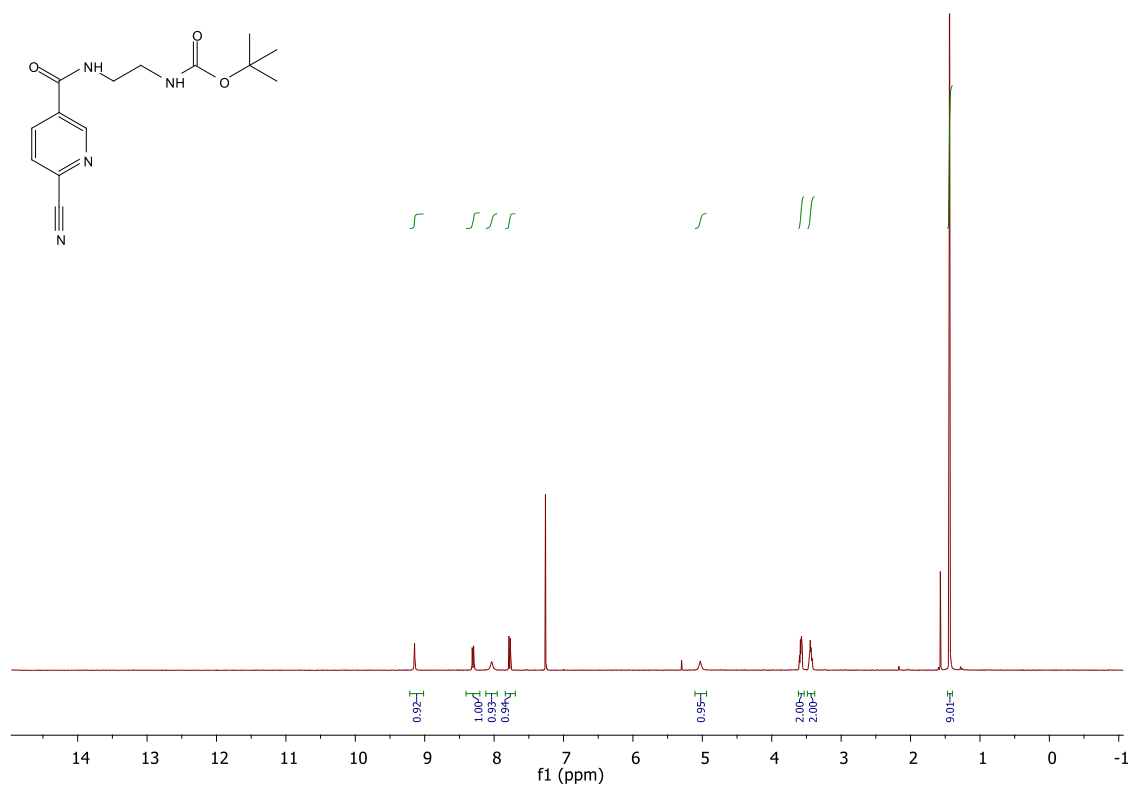
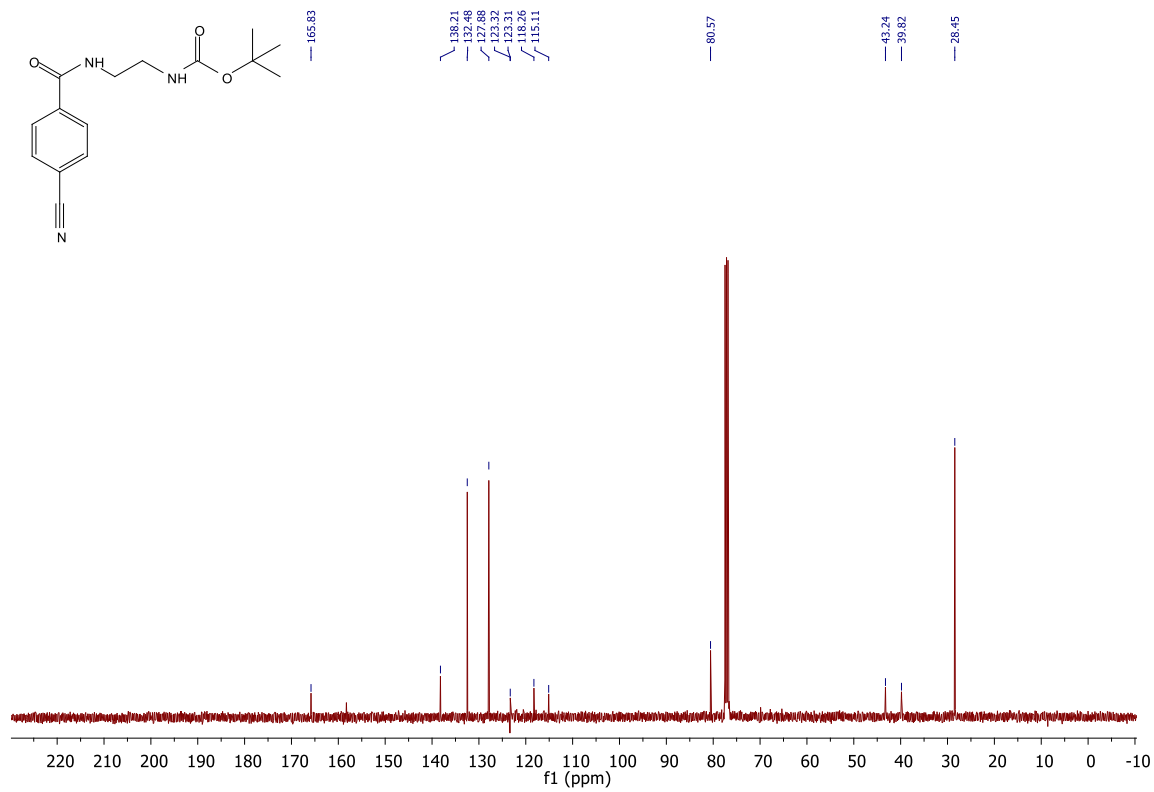


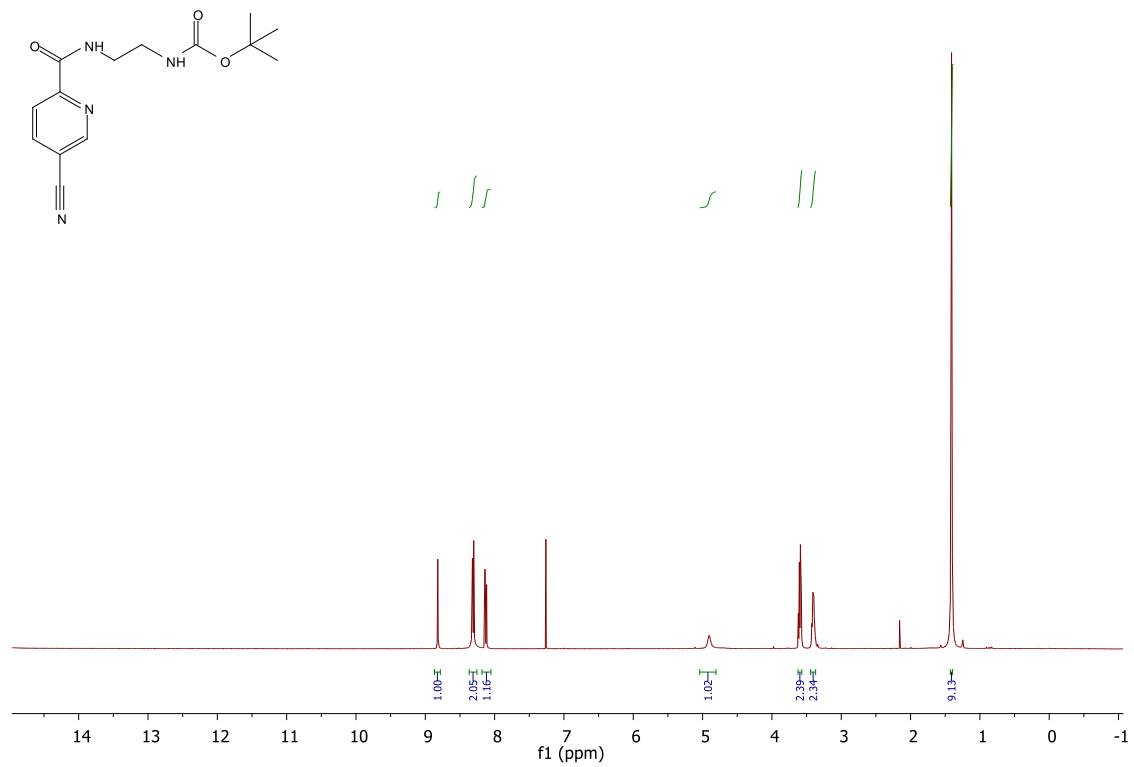
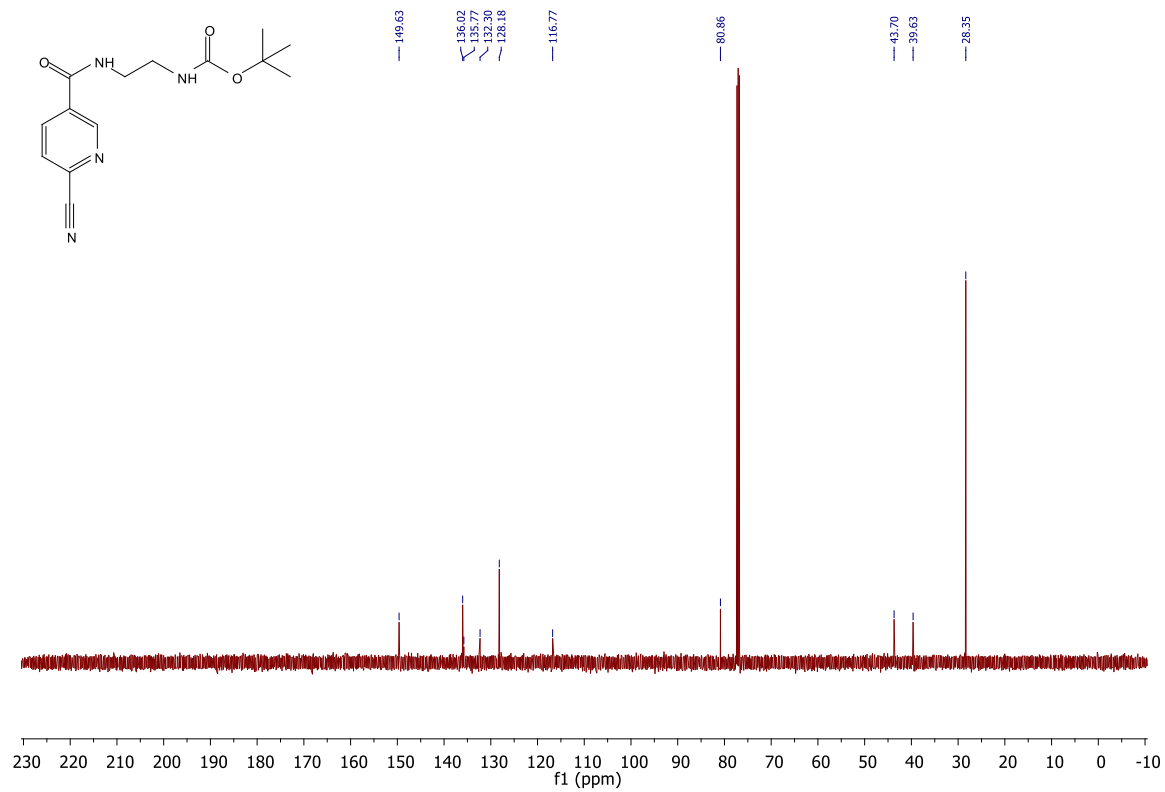


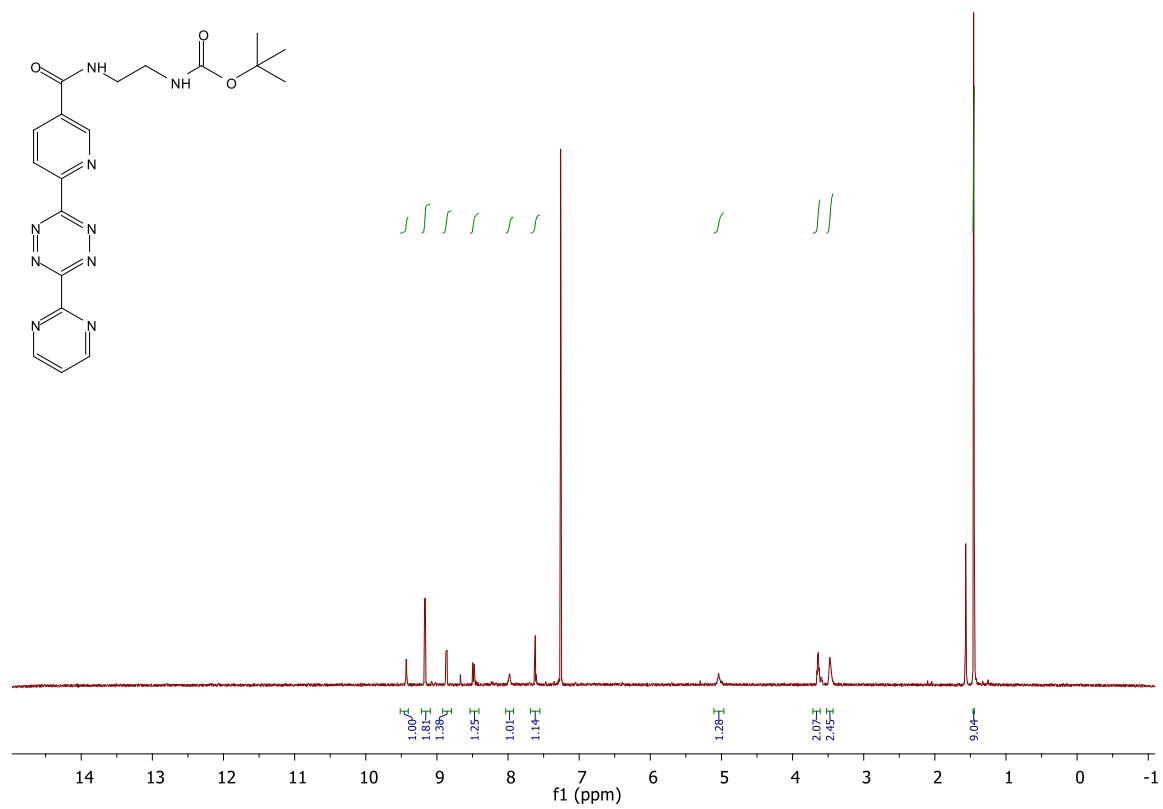
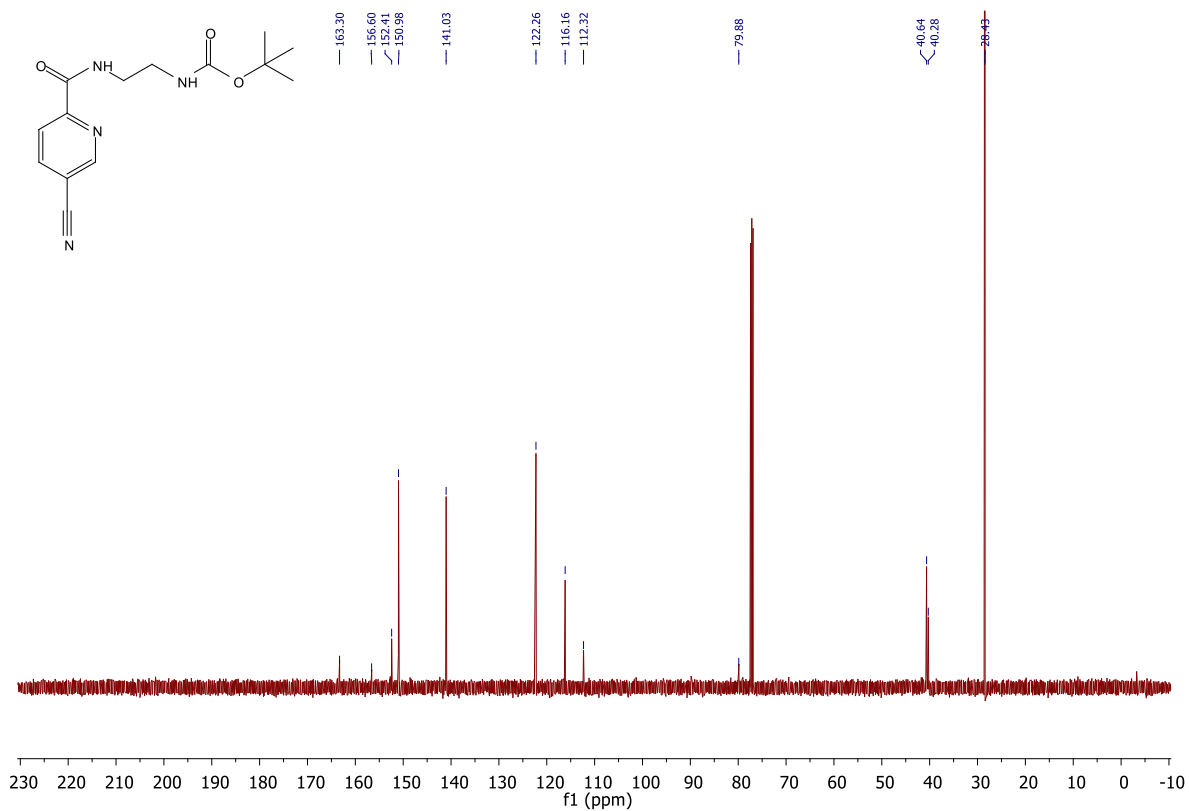


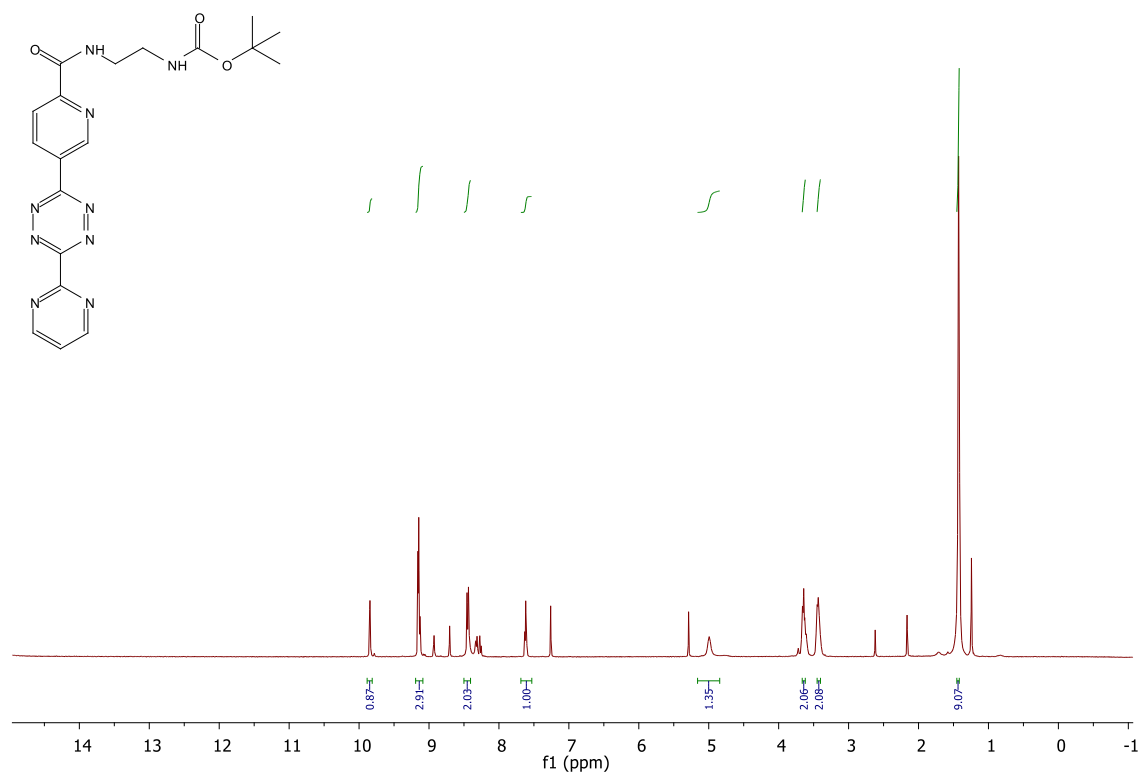
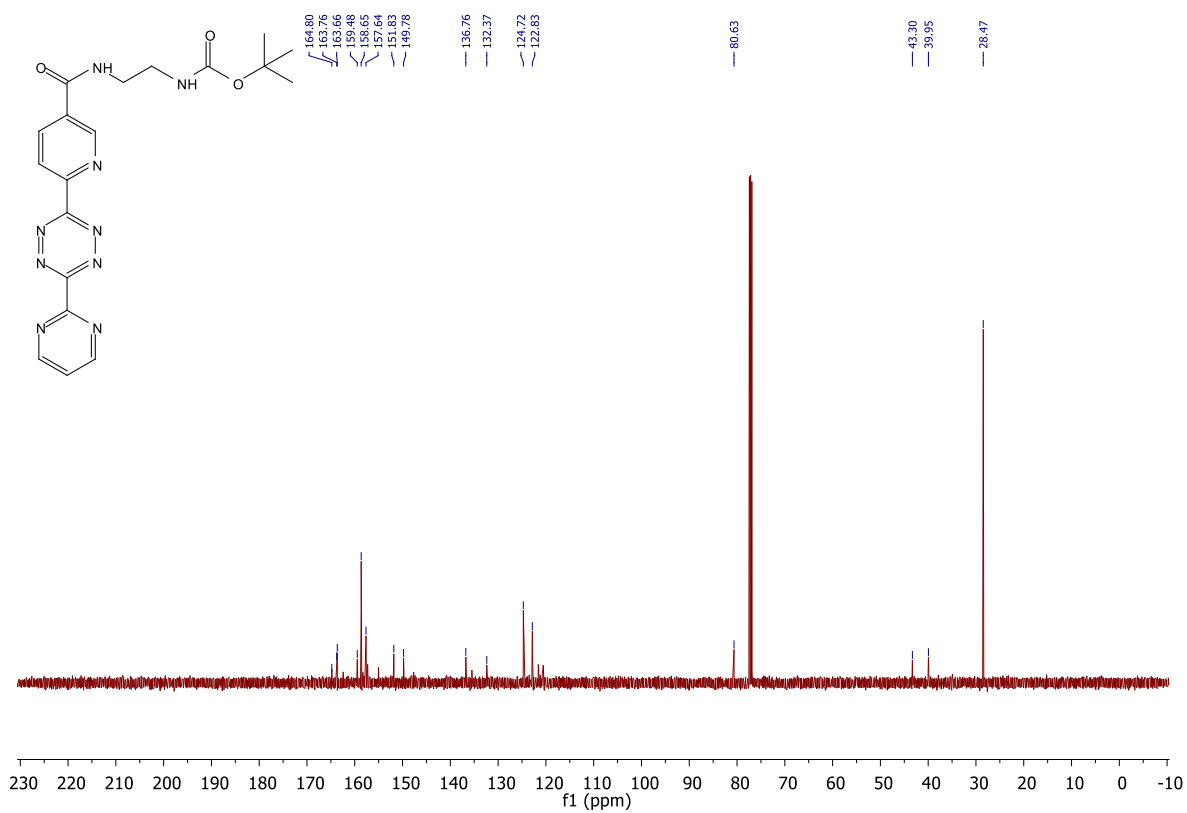


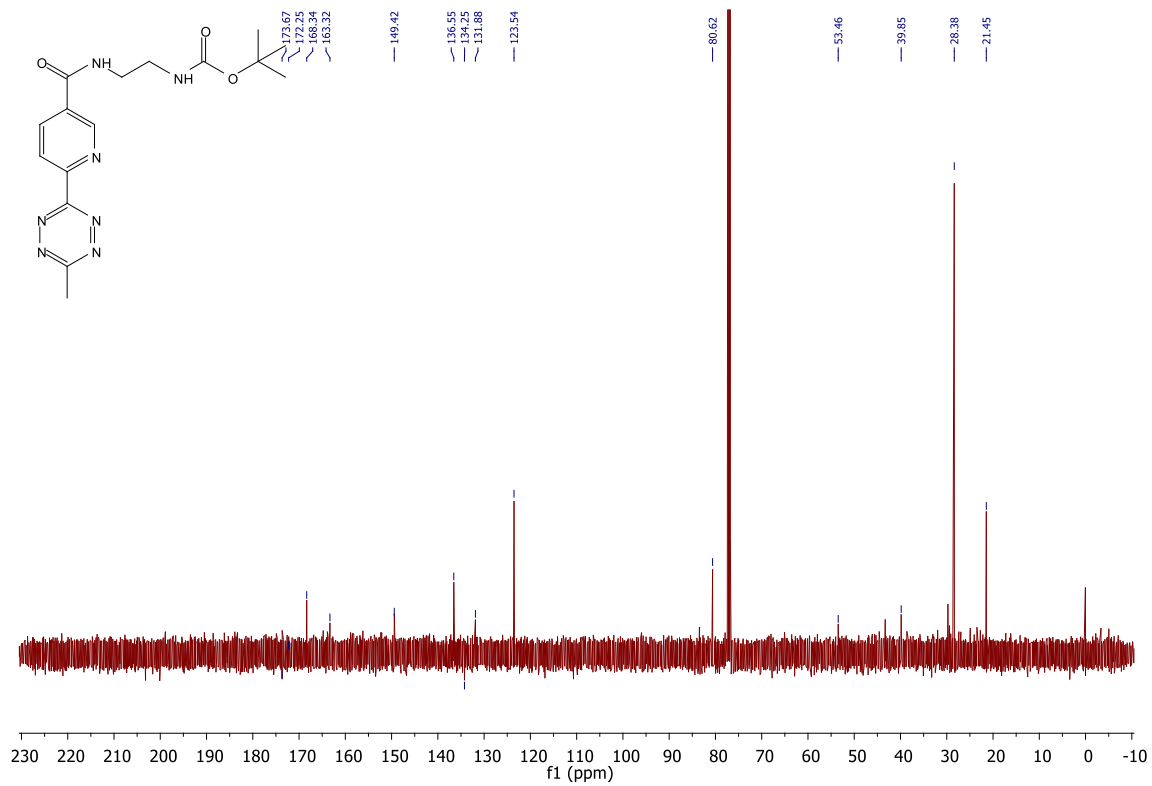
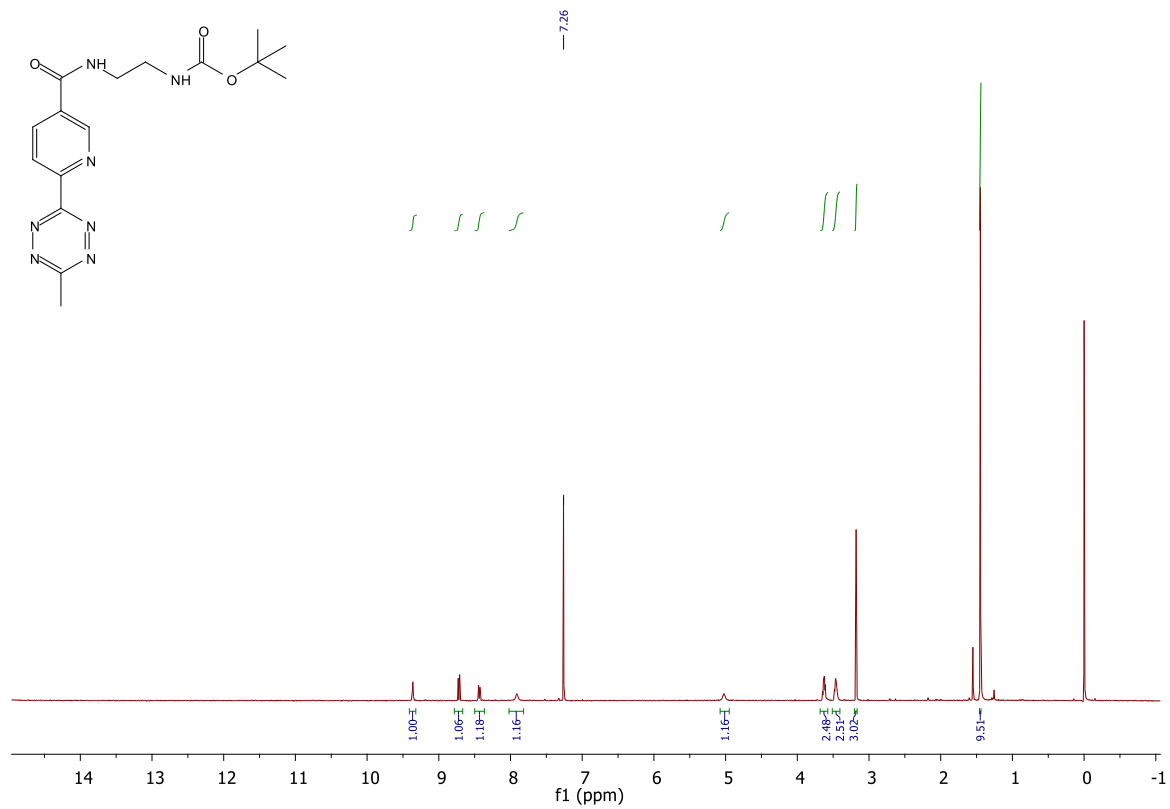


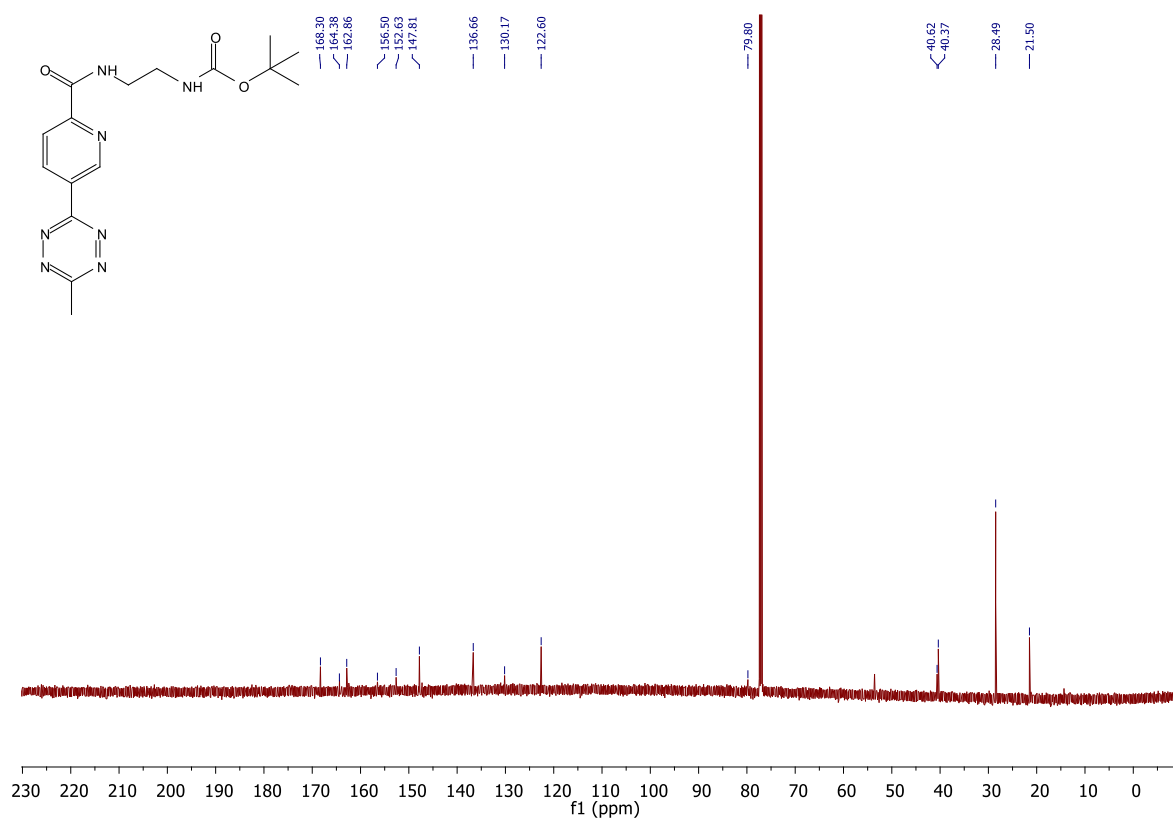
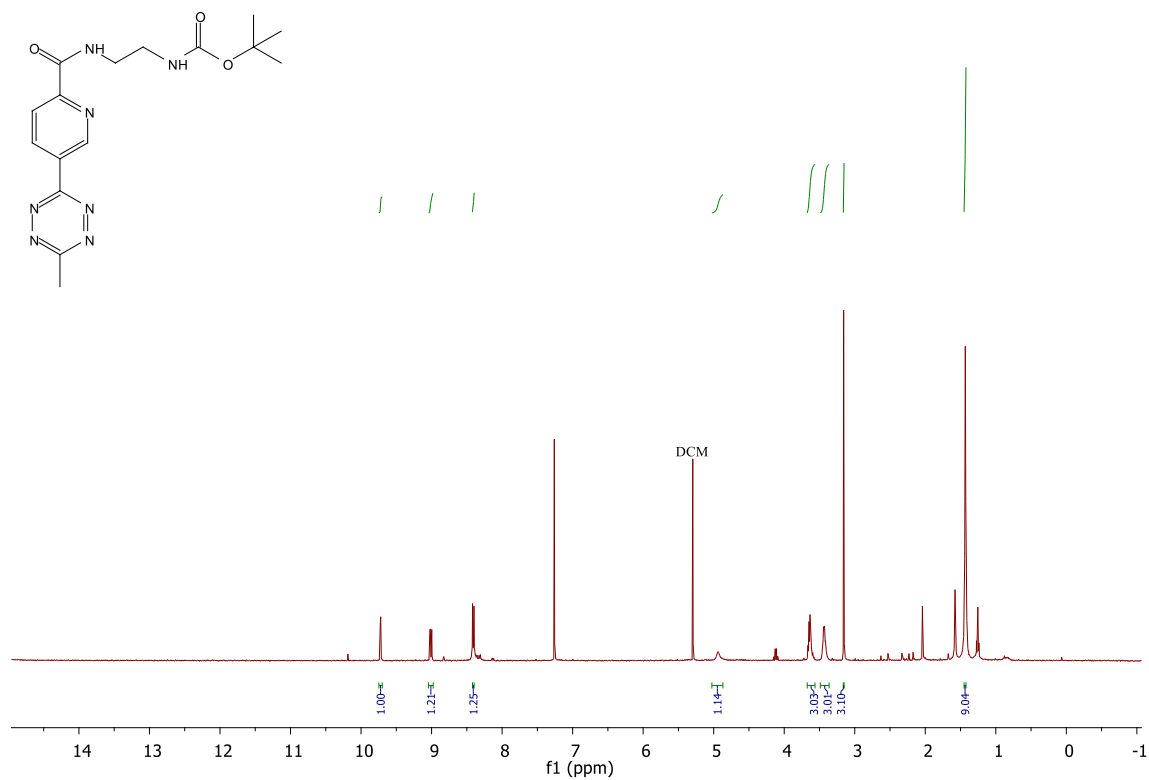


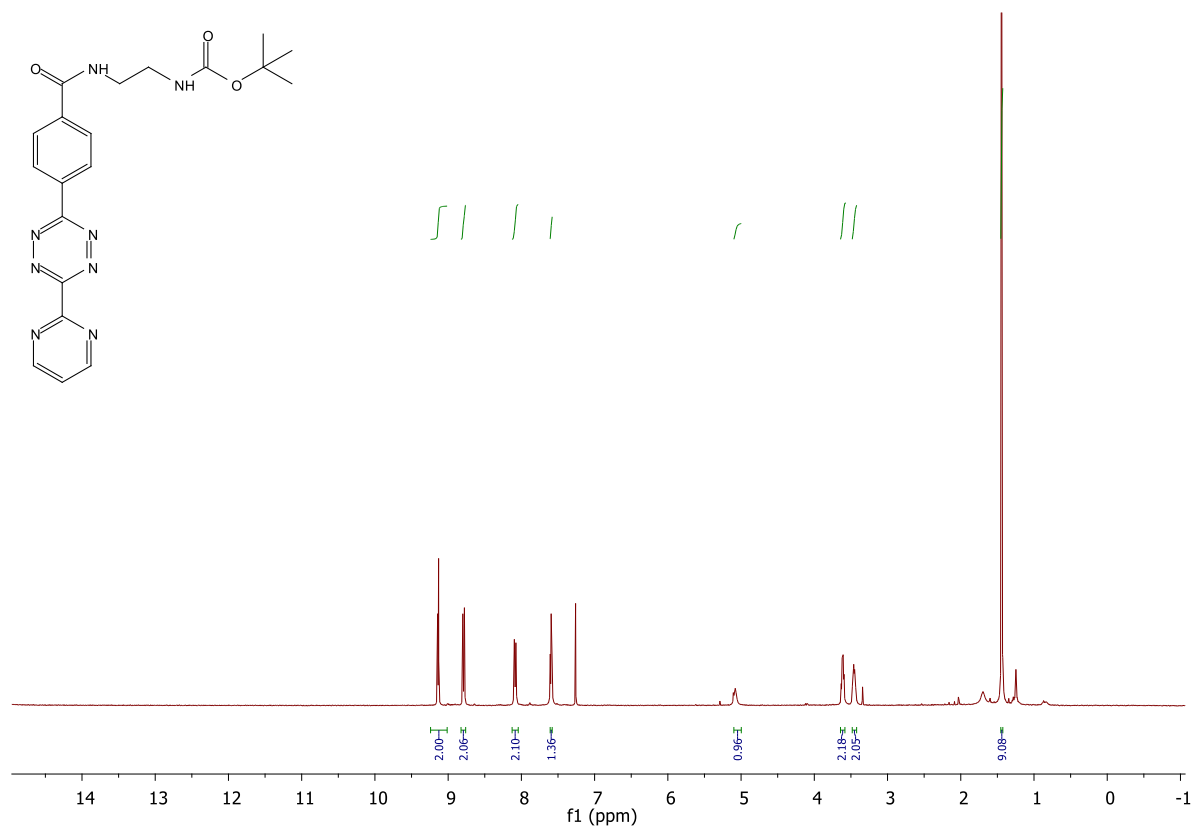
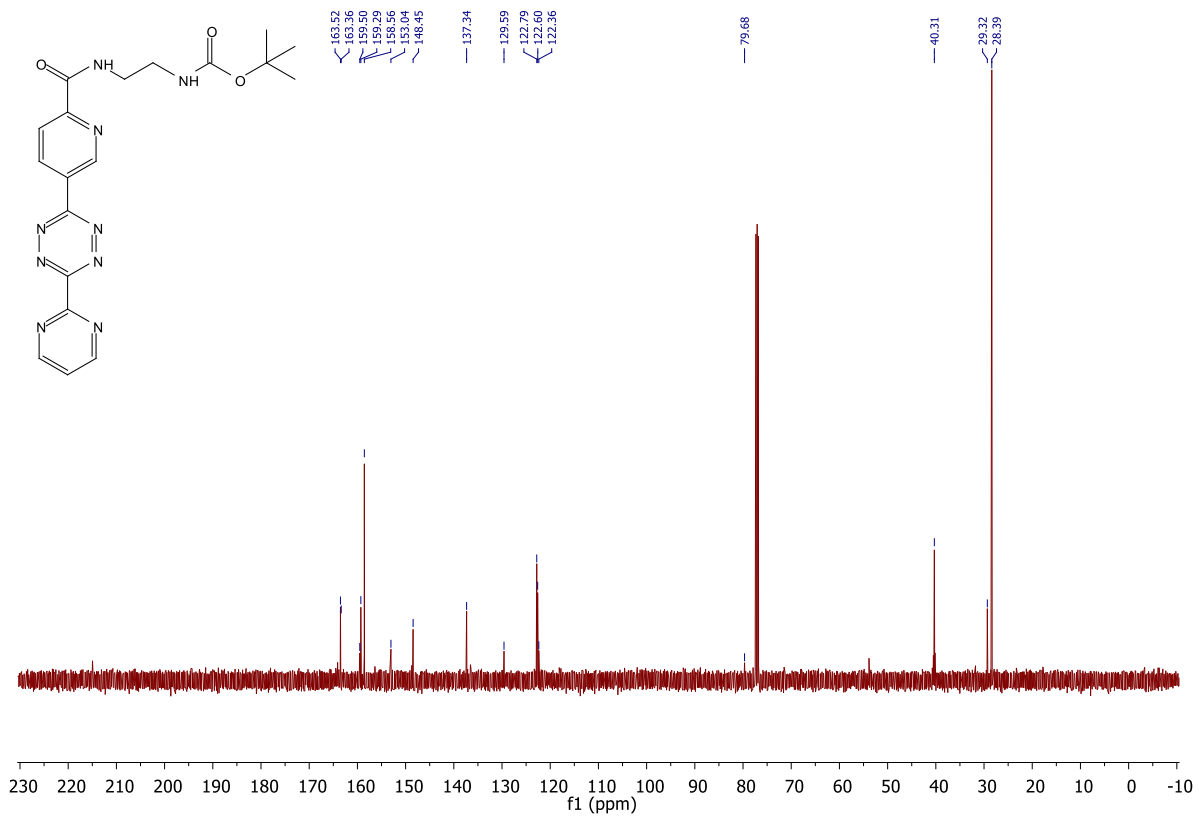


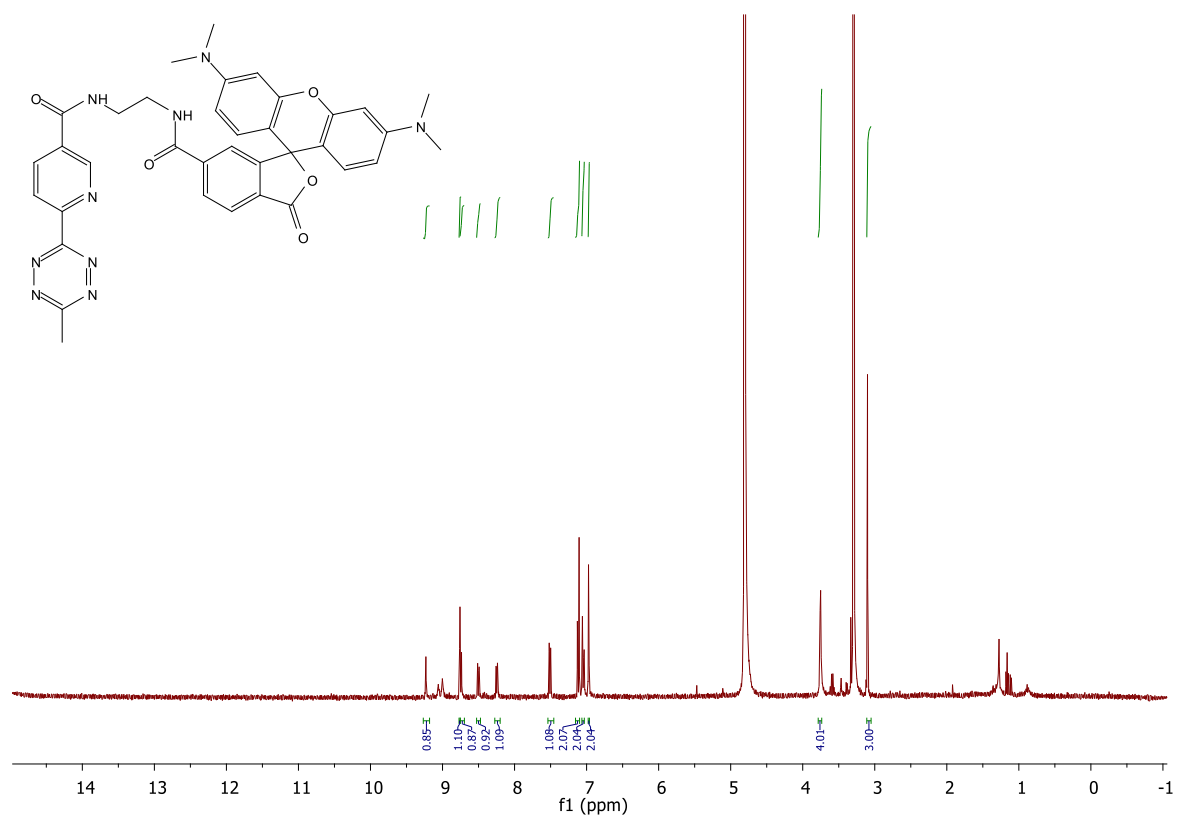
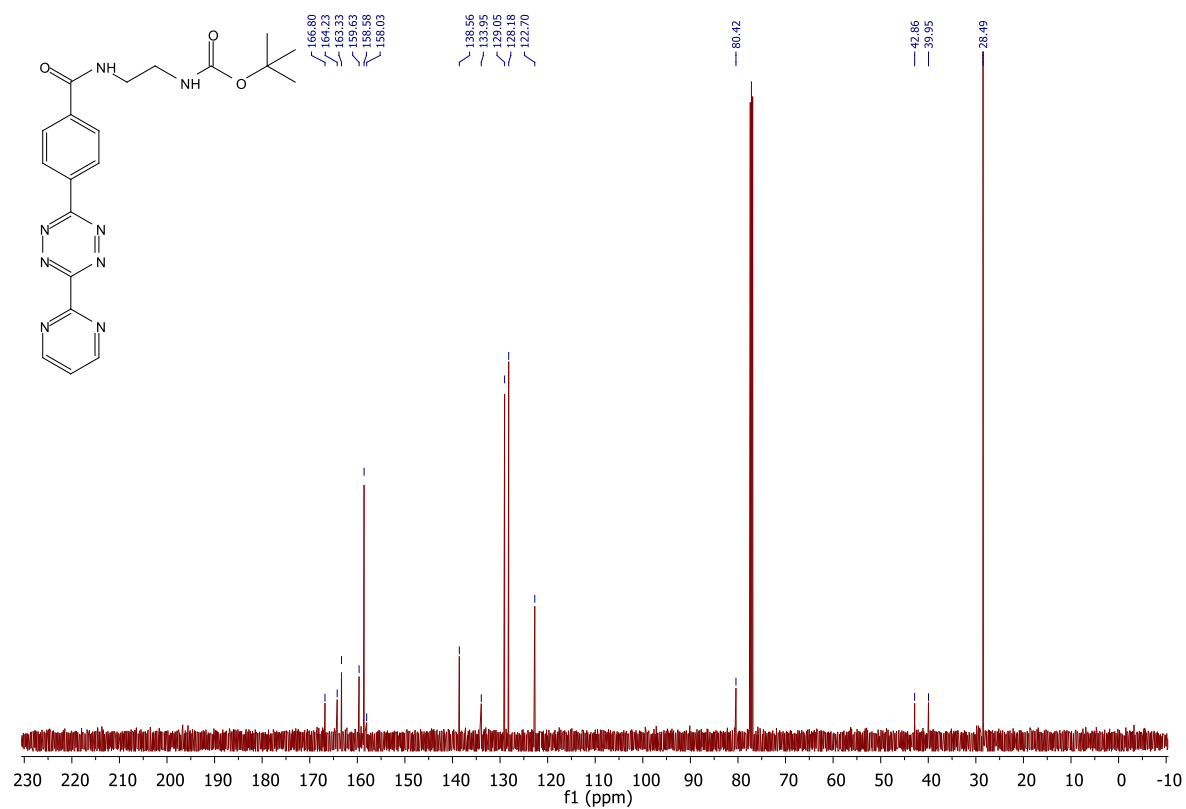


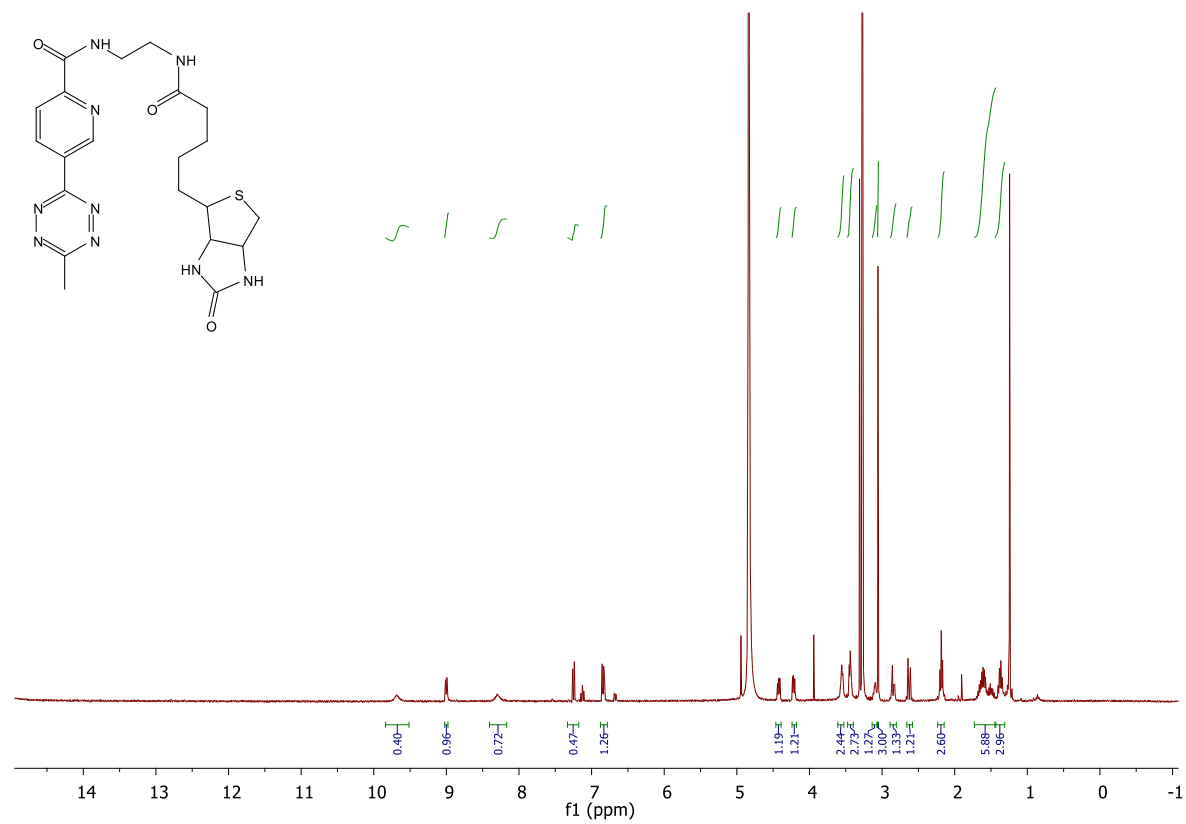
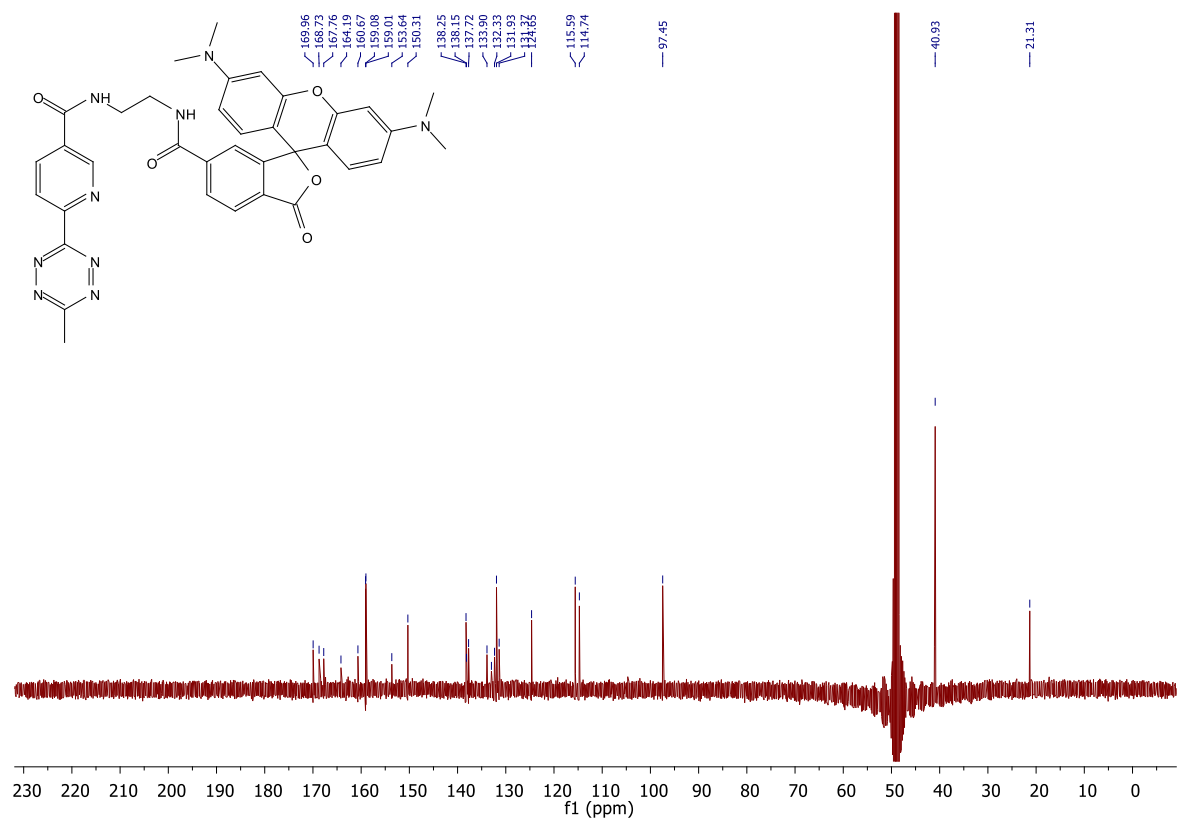


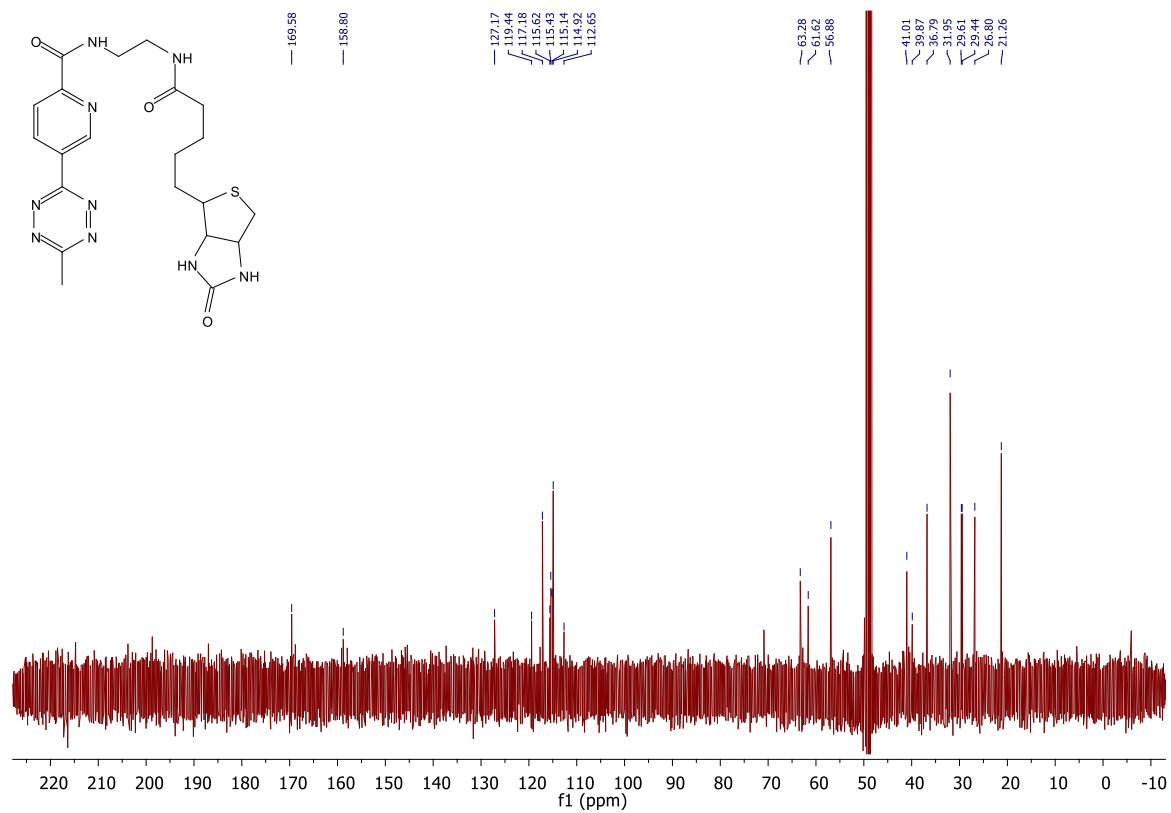




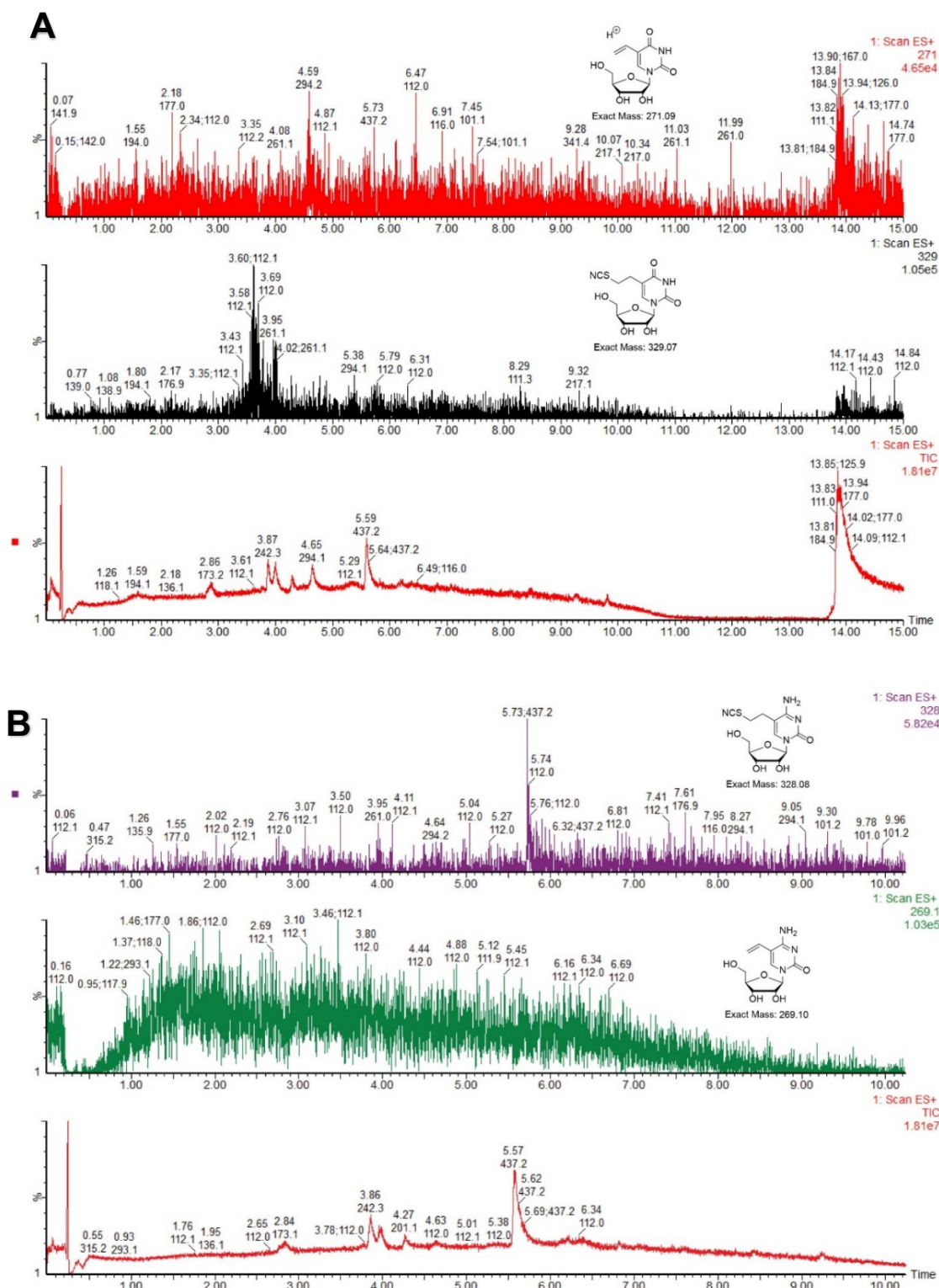








4.6 Supplementary Data



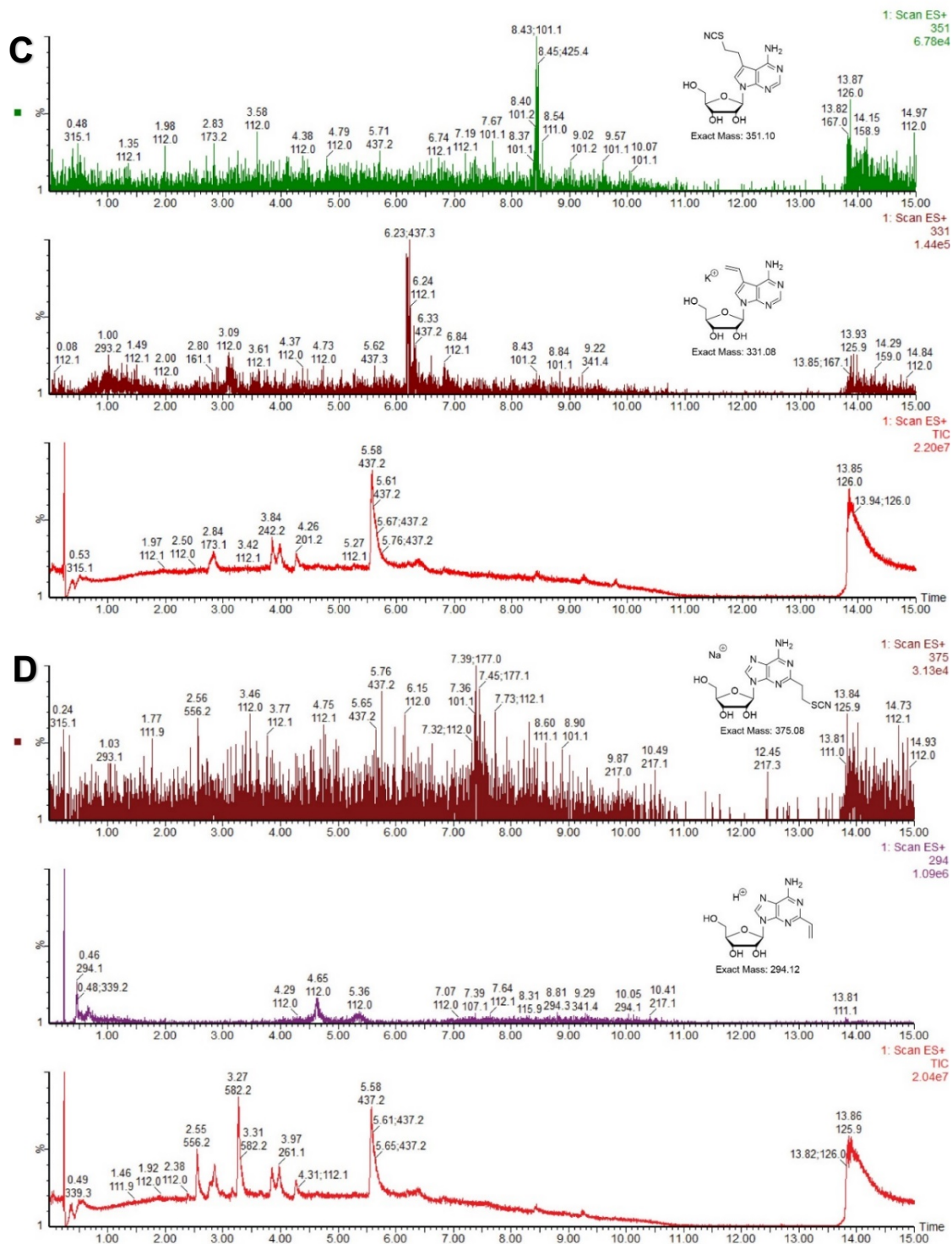


Figure 4-S1. LC-MS analysis of nucleosides in TRizol reagent. A. 5-VU B. 5-VC C. 7-dVA and D. 2-VA. Top two panels display peaks correlated to the labeled mass. Bottom panel

displays the chromatogram. **5-VC** showing little-to-no detection after TRIzol incubation and thiocyanate adduct found. **5-VC** is most susceptible to Michael addition.

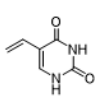
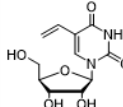
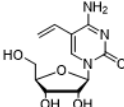
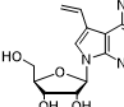
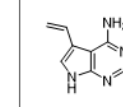
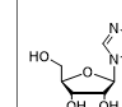
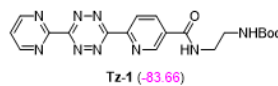
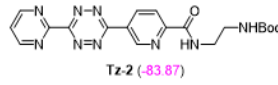
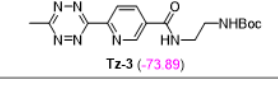
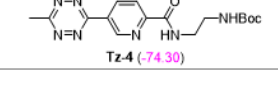
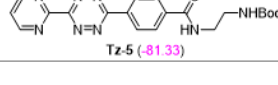
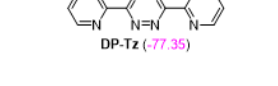
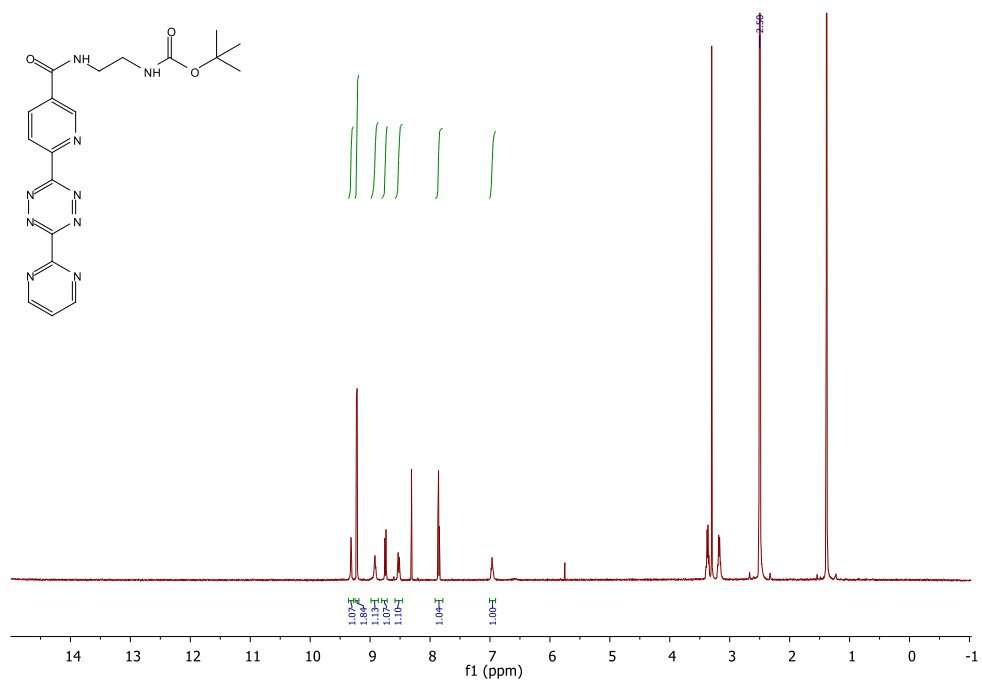
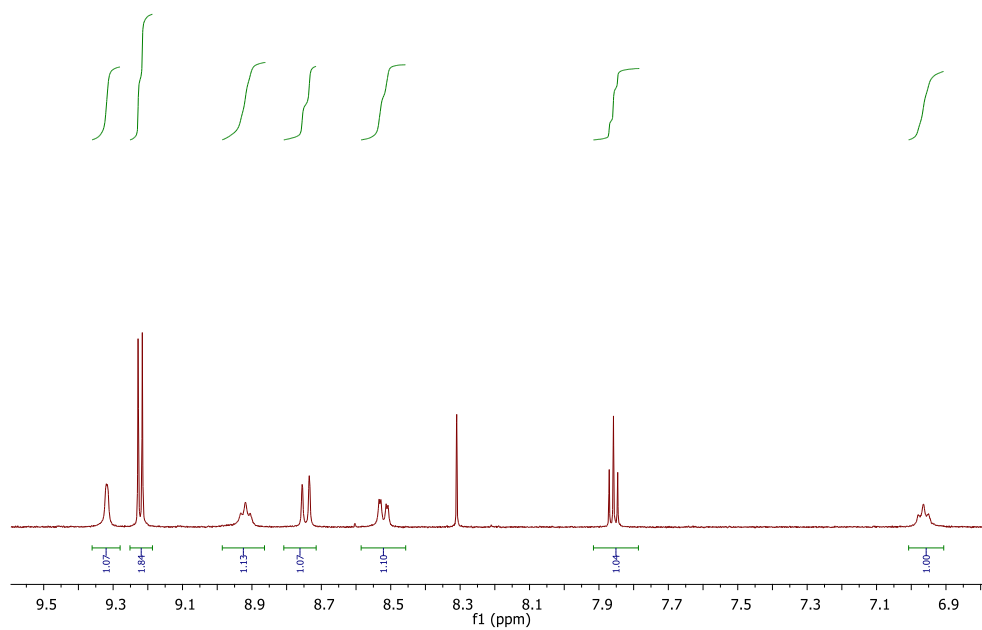
HOMO-LUMO Gap (kcal/mol) HOMO & LUMO+1 values calculated by TPSS, D3, def2-SVP (kcal/mol) k in 10% AcOH in DMSO ($M^{-1}s^{-1}$) k in DMSO ($M^{-1}s^{-1}$) k in 1:3 H₂O:DMSO ($M^{-1}s^{-1}$) * Tz-1, Tz-2 undergo hydrolysis	 5-VUb (-124.41)	 5-VU (-124.48)	 5-VC (-121.99)	 7-dVA (-114.43)	 7-dVAb (-111.87)	 2-VA (-124.53)
 Tz-1 (-83.66)	40.75 $2.49 \pm 0.19 \times 10^{-2}$ $1.49 \pm 0.12 \times 10^{-2}$	40.82 $3.95 \pm 0.24 \times 10^{-3}$ $4.42 \pm 0.03 \times 10^{-3}$	38.33 $5.72 \pm 0.44 \times 10^{-3}$ $7.03 \pm 0.25 \times 10^{-3}$	30.76 $1.67 \pm 0.68 \times 10^{-2}$ $9.79 \pm 0.31 \times 10^{-3}$	28.20 $5.35 \pm 0.54 \times 10^{-2}$ $4.33 \pm 0.08 \times 10^{-3}$	40.87 $1.89 \pm 0.14 \times 10^{-3}$ $1.55 \pm 0.11 \times 10^{-3}$
 Tz-2 (-83.87)	40.54 $2.44 \pm 0.20 \times 10^{-3}$	40.61 $1.38 \pm 0.08 \times 10^{-3}$	38.12 $2.05 \pm 0.19 \times 10^{-3}$	30.56 $7.54 \pm 0.59 \times 10^{-3}$	27.99 $1.59 \pm 0.15 \times 10^{-2}$	40.66 $6.37 \pm 0.29 \times 10^{-4}$
 Tz-3 (-73.89)	50.53 $1.46 \pm 0.13 \times 10^{-3}$ $3.80 \pm 0.33 \times 10^{-3}$	50.59 $5.06 \pm 0.44 \times 10^{-4}$ $1.08 \pm 0.07 \times 10^{-3}$	48.10 $6.13 \pm 0.43 \times 10^{-4}$ $1.37 \pm 0.03 \times 10^{-3}$	40.54 $2.05 \pm 0.04 \times 10^{-3}$ $1.48 \pm 0.14 \times 10^{-3}$	37.98 $6.66 \pm 0.66 \times 10^{-3}$ $2.17 \pm 0.15 \times 10^{-2}$	50.64 $2.88 \pm 0.28 \times 10^{-4}$ $7.44 \pm 0.13 \times 10^{-4}$
 Tz-4 (-74.30)	50.11 $8.64 \pm 0.73 \times 10^{-4}$ $2.70 \pm 0.15 \times 10^{-3}$	50.18 $1.75 \pm 0.12 \times 10^{-4}$ $6.77 \pm 0.35 \times 10^{-4}$	47.69 $5.53 \pm 0.45 \times 10^{-4}$ $7.23 \pm 0.36 \times 10^{-4}$	40.13 $1.88 \pm 0.09 \times 10^{-3}$ $7.47 \pm 0.69 \times 10^{-3}$	37.57 $6.50 \pm 0.22 \times 10^{-3}$ $1.92 \pm 0.08 \times 10^{-2}$	50.23 $1.26 \pm 0.11 \times 10^{-4}$ $4.28 \pm 0.26 \times 10^{-4}$
 Tz-5 (-81.33)	43.08 $6.15 \pm 0.62 \times 10^{-3}$	43.15 $1.17 \pm 0.10 \times 10^{-3}$	40.66 $2.70 \pm 0.27 \times 10^{-3}$	33.09 $1.88 \pm 0.18 \times 10^{-3}$	30.53 $6.73 \pm 0.54 \times 10^{-3}$	43.19 $3.69 \pm 0.33 \times 10^{-3}$
 DP-Tz (-77.35)	47.07 $8.19 \pm 0.24 \times 10^{-3}$	47.14 $1.72 \pm 0.09 \times 10^{-3}$	44.64 $4.55 \pm 0.23 \times 10^{-3}$	37.08 $3.43 \pm 0.17 \times 10^{-3}$	34.52 $1.13 \pm 0.08 \times 10^{-2}$	47.18 $6.41 \pm 0.12 \times 10^{-4}$

Figure 4-S2. HOMO-LUMO energy calculation and reaction kinetics. Calculated HOMO and LUMO+1 orbital energy in kcal/mol displayed next to compounds. HOMO-LUMO+1 energy in kcal/mol displayed in green. k ($M^{-1}s^{-1}$) in DMSO displayed in black. k ($M^{-1}s^{-1}$) in 1:3 H₂O:DMSO displayed in red. k ($M^{-1}s^{-1}$) in DMSO with 10% AcOH displayed in orange.

A**B**

C

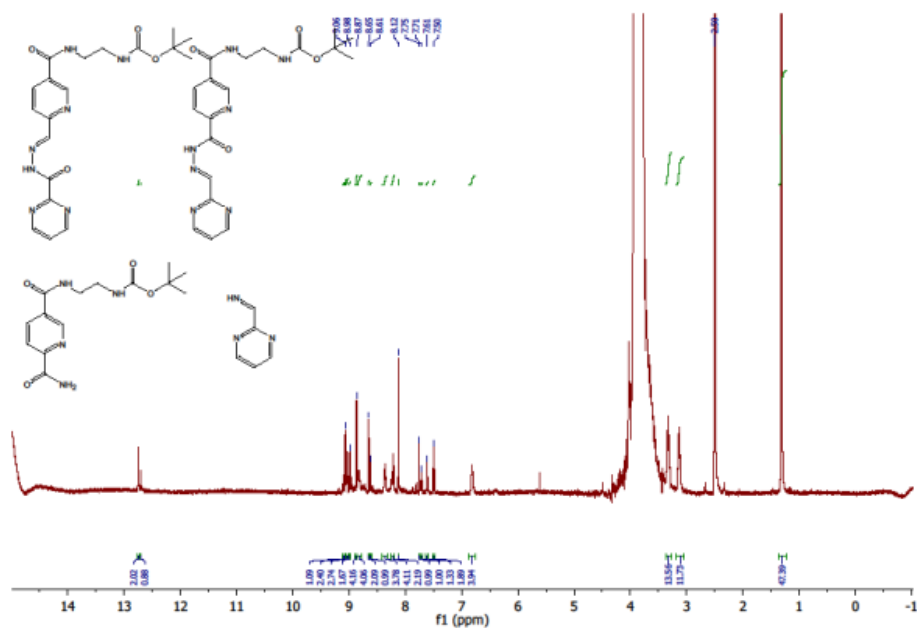


Figure 4-S3. ^1H NMR analysis of Tz-1 hydrolysis. A. initial ^1H NMR of Tz-1. B. Initial ^1H NMR of Tz-1 with zoom on aromatic region. C. ^1H NMR of Tz-1 after water addition showing hydrazide regioisomers.

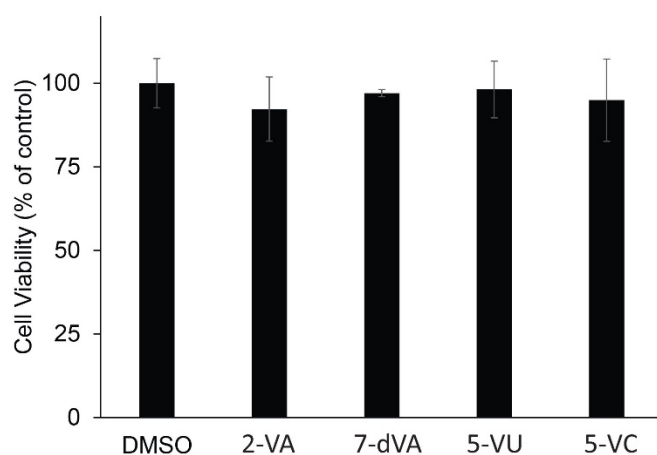


Figure 4-S4. MTT assay of vinyl-nucleosides. MTT viability assay measured proliferation of cells through reducing tetrazolium dye MTT to its insoluble formazan. Cells were treated for 12 h with 1 mM of each analog, and viability was normalized to the DMSO treated control.

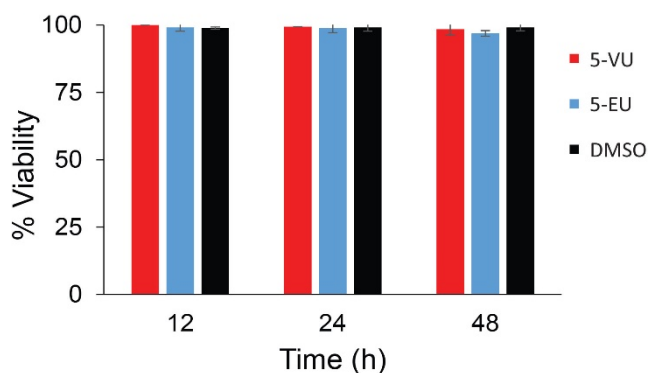
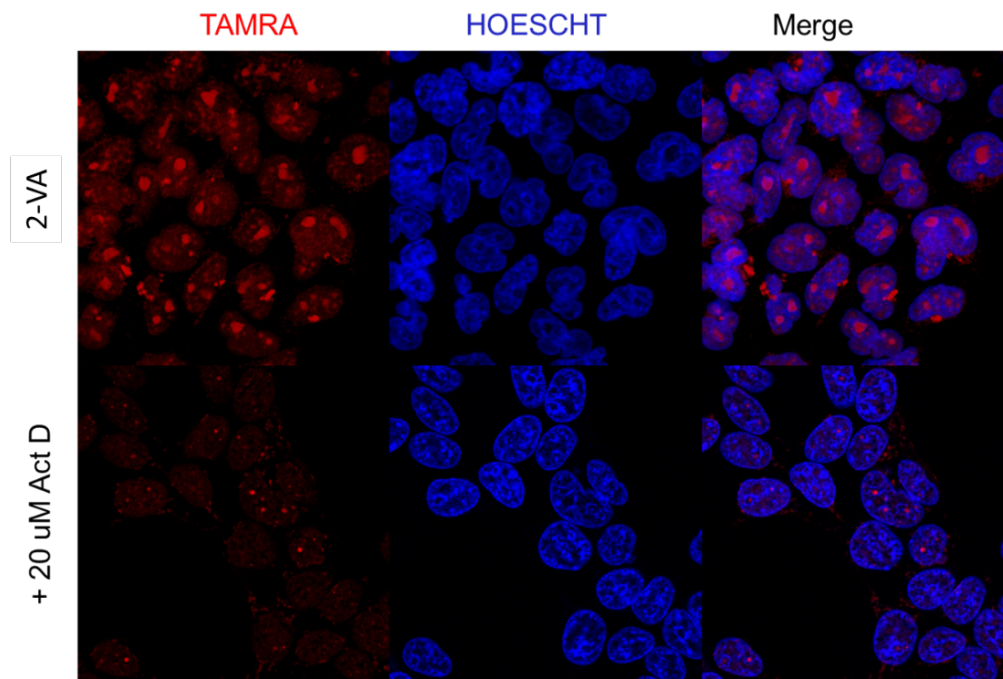
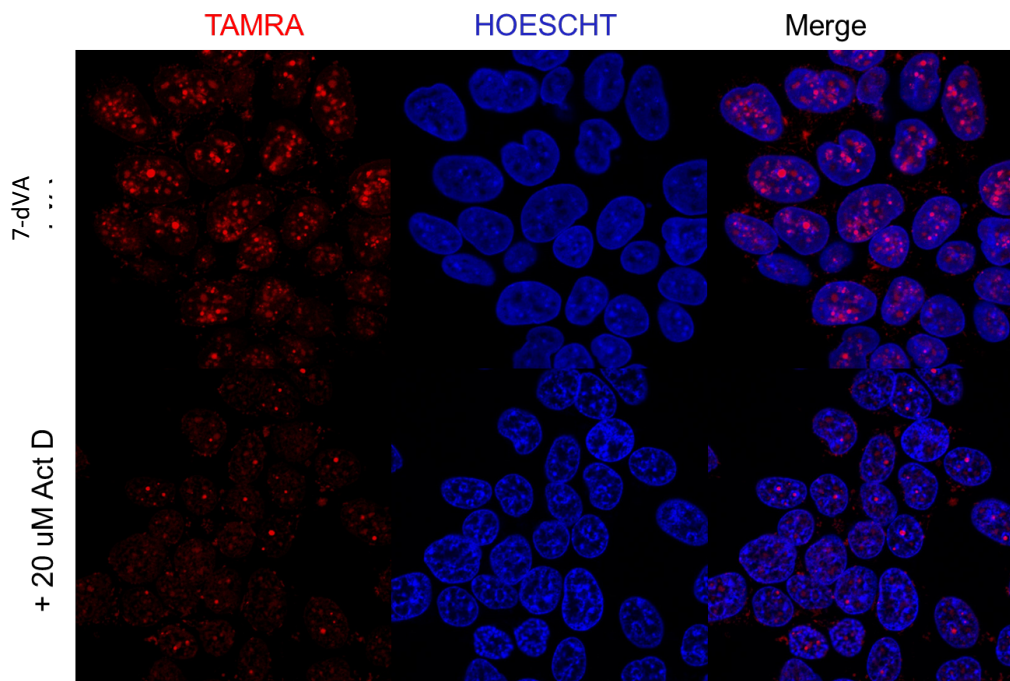


Figure 4-S5. Trypan Blue exclusion assay of 5-EU and 5-VU. Trypan Blue Exclusion assay cells treated with 5-VU and 5-EU for 12, 24, and 48 h.

A



B



C

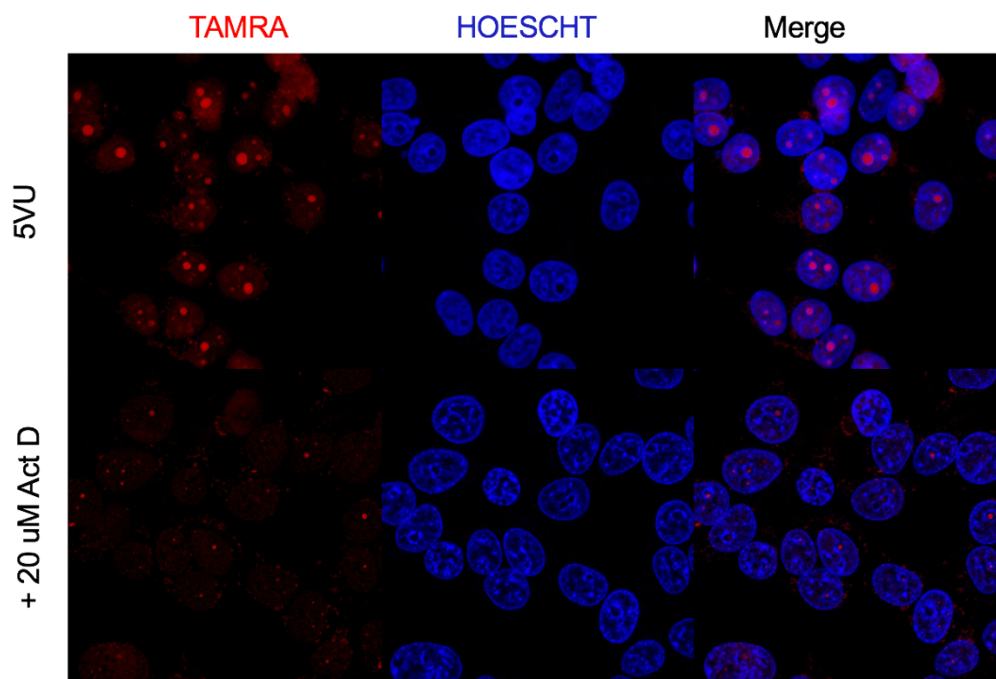


Figure 4-S6. Cell imaging with 2-VA and 7-dVA. Fluorescent imaging of **A. 2-VA** and **B. 7-dVA** and **C. 5-VU** in HEK293T cells. Cells were treated with 1 mM analog and allowed to incubate for 5 hours. After fixation and permeabilization, **TAMRA-Tz-3** was reacted with vinyl incorporated RNA for 2 h at 37 °C. Nucleus was stained with HOESCHT. Concurrent treatment with 20 μM actinomycin D and analog revealed loss of fluorescent signal revealing vinyl-nucleosides are incorporated through polymerase I/II.

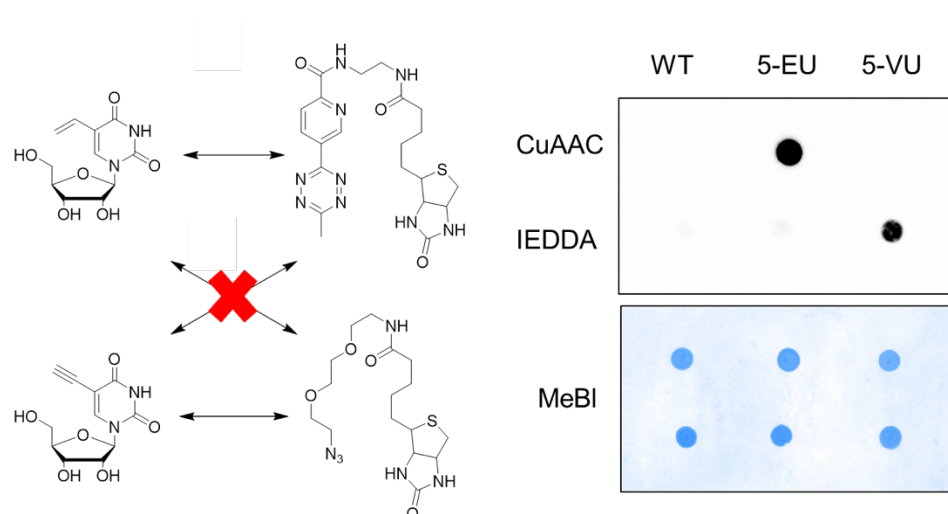


Figure 4-S7. Mutually orthogonal bioorthogonal chemistries. HEK239T cells were incubated with **5-EU** or **5-VU** and RNA was isolated. The RNA was then reacted with biotin-azide and **Biotin-Tz-4** resulting in specific reactions between CuAAC and IEDDA chemical partners as shown by dot blot analysis.

Table 4-S1. LC-MS/MS Triple Quadrupole detection parameters.

Nucleoside	Parent ion (m/z)	Daughter ion (m/z)	Dwell time (s)	Cone (V)	Collision (V)	Retention time (min)
Adenosine	268.03	136.06	0.049	30	10	0.49
Uridine	245.03	113.02	0.049	20	20	0.48
Guanosine	284.04	152.46	0.049	20	20	0.48
Cytidine	244.03	111.70	0.049	20	10	0.46
8-vinyladenosine	294.08	161.23	0.044	20	20	0.61
7-deazavinyladenosine	293.10	161.26	0.044	30	20	0.54
2-vinyladenosine	294.10	162.26	0.044	20	20	0.68
5-vinyluridine	271.09	139.20	0.044	10	10	0.65
8-vinylguanosine	310.10	178.24	0.044	20	10	0.56
5-vinylcytidine	270.10	138.23	0.044	20	10	0.46
5-iodocytidine	369.93	238.16	0.044	20	20	0.53

Ions were detected in positive ion mode.

Column conditions: water (+0.2% acetic acid) : acetonitrile

Table 4-S2. Lowest energy reaction barrier ($\Delta G_{\text{barrier}}$ in kcal/mol) computed for each tetrazine-vinylnucleotide pair, where:

$$\Delta G_{\text{barrier}} = \Delta G_{\text{TS}} - \Delta G_{\text{complex}}$$

where ΔG_{TS} and $\Delta G_{\text{complex}}$ are the free energies of the transition state and initial complex, respectively. In contrast to the driving forces derived from the HOMO-LUMO+1 gap, the computed barriers reproduce all observed trends in reactivity, i.e., **Tz-1** > **Tz-2**. Furthermore, the estimated relative rates of, for example, the **Tz-1-5-VUb** reaction compared to the **Tz-2-5-VUb** reaction, agree well with experiment: **Tz-1** was estimated to be faster than **Tz-2** by factors of 10.1 and 3.7 for the reactions with **5-VUb** and **7-dVAb**, respectively, compared to factors of 10 and 2.2 measured experimentally. This stands in stark contrast to the results from HOMO-LUMO+1 driving forces, in which essentially no difference in reactivity between **Tz-1** and **Tz-2** is predicted.

	5-VUb	7-dVAb	2-VAb
Tz-1	9.14	7.82	10.45
Tz-2	10.51	8.61	11.68

4.7 References

1. Morris, K. V.; Mattick, J. S. *Nat. Rev. Genet.* **2014**, *15*, 423-437.
2. Novikova, I. V.; Hennelly, S. P.; Tung, C.-S.; Sanbonmatsu, K. Y. *J. Mol. Biol.* **2013**, *425*, 3731-3746.
3. Prescher, J. A.; Bertozzi, C. R. *Nat. Chem. Biol.* **2005**, *1*, 13-21.
4. Miller, M. R.; Robinson, K. J.; Cleary, M. D.; Doe, C. Q. *Nat. Methods* **2009**, *6*, 439-U457.
5. Jao, C. Y.; Salic, A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 15779-15784.
6. Curanovic, D.; Cohen, M.; Singh, I.; Slagle, C. E.; Leslie, C. S.; Jaffrey, S. R. *Nat. Chem. Biol.* **2013**, *9*, 671-673.
7. Zheng, Y.; Beal, P. A. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 1799-1802.
8. Nainar, S.; Beasley, S.; Fazio, M.; Kubota, M.; Dai, N.; Correa, I. R.; Spitale, R. C. *ChemBioChem* **2016**, *17*, 2149-2152.
9. Nainar, S.; Kubota, M.; McNitt, C.; Tran, C.; Popik, V. V.; Spitale, R. C. *J. Am. Chem. Soc.* **2017**, *139*, 8090-8093.
10. Duffy, Erin E.; Rutenberg-Schoenberg, M.; Stark, Catherine D.; Kitchen, Robert R.; Gerstein, Mark B.; Simon, Matthew D. *Mol. Cell* **2015**, *59*, 858-866.
11. Paredes, E.; Das, S. R. *ChemBioChem* **2011**, *12*, 125-131.
12. Winz, M. L.; Samanta, A.; Benzinger, D.; Jaschke, A. *Nucleic Acids Res.* **2012**, *40*, e78.
13. Maeda, T.; Date, A.; Watanabe, M.; Hidaka, Y.; Iwatani, Y.; Takano, T. *Ann. Clin. Lab Sci.* **2016**, *46*, 571-577.
14. Nilsson, H.; Krawczyk, K. M.; Johansson, M. E. *J. Biotechnol.* **2014**, *192 Pt A*, 62-65.
15. 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*, 2148-2151.

16. Wu, H.; Devaraj, N. K. *Acc. Chem. Res.* **2018**, *51*, 1249-1259.
17. Etienne, B.; Philippe, B. *Curr. Org. Chem.* **2016**, *20*, 2136-2160.
18. Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 10317-10320.
19. Patterson, D. M.; Nazarova, L. A.; Xie, B.; Kamber, D. N.; Prescher, J. A. *J. Am. Chem. Soc.* **2012**, *134*, 18638-18643.
20. Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. *Nature Chem.* **2012**, *4*, 298.
21. Rossin, R.; van den Bosch, S. M.; ten Hoeve, W.; Carvelli, M.; Versteegen, R. M.; Lub, J.; Robillard, M. S. *Bioconjugate Chem.* **2013**, *24*, 1210-1217.
22. George, J. T.; Srivatsan, S. G. *Bioconjugate Chem.* **2017**, *28*, 1529-1536.
23. Bußkamp, H.; Batroff, E.; Niederwieser, A.; Abdel-Rahman, O. S.; Winter, R. F.; Wittmann, V.; Marx, A. *Chem. Commun. (Cambridge, U. K.)* **2014**, *50*, 10827-10829.
24. Grammel, M.; Hang, H. C. *Nat. Chem. Biol.* **2013**, *9*, 475.
25. Lutz, S.; Liu, L. F.; Liu, Y. C. *Chimia* **2009**, *63*, 737-744.
26. Mathews, I. I.; Erion, M. D.; Ealick, S. E. *Biochemistry* **1998**, *37*, 15607-15620.
27. Suzuki, N. N.; Koizumi, K.; Fukushima, M.; Matsuda, A.; Inagaki, F. *Structure* **2004**, *12*, 751-764.
28. Qu, D.; Zhou, L.; Wang, W.; Wang, Z.; Wang, G.; Chi, W.; Zhang, B. *Anal. Biochem.* **2013**, *434*, 128-135.
29. Grammel, M.; Hang, H.; Conrad, N. K. *ChemBioChem* **2012**, *13*, 1112-1115.
30. Hottin, A.; Betz, K.; Diederichs, K.; Marx, A. *Chem. Eur. J.* **2017**, *23*, 2109-2118.
31. Dudycz, L.; Stolarski, R.; Pless, R.; Shugar, D. *Z Naturforsch C* **1979**, *34*, 359-373.

32. Rosemeyer, H.; Toth, G.; Golankiewicz, B.; Kazimierczuk, Z.; Bourgeois, W.; Kretschmer, U.; Muth, H. P.; Seela, F. *J. Org. Chem.* **1990**, *55*, 5784-5790.
33. Yang, J.; Karver, M. R.; Li, W.; Sahu, S.; Devaraj, N. K. *Angew. Chem., Int. Ed.* **2012**, *51*, 5222-5225.
34. Li, Z.; Wang, D.; Li, L.; Pan, S.; Na, Z.; Tan, C. Y. J.; Yao, S. Q. *J. Am. Chem. Soc.* **2014**, *136*, 9990-9998.
35. Hörner, S.; Uth, C.; Avrutina, O.; Frauendorf, H.; Wiessler, M.; Kolmar, H. *Chem. Commun. (Cambridge, U. K.)* **2015**, *51*, 11130-11133.
36. Knall, A.-C.; Hollauf, M.; Slugovc, C. *Tetrahedron Lett.* **2014**, *55*, 4763-4766.
37. Lang, K.; Chin, J. W. *ACS Chem. Biol.* **2014**, *9*, 16-20.
38. Blizzard, R. J.; Backus, D. R.; Brown, W.; Bazewicz, C. G.; Li, Y.; Mehl, R. A. *J. Am. Chem. Soc.* **2015**, *137*, 10044-10047.
39. An, R.; Jia, Y.; Wan, B. H.; Zhang, Y. F.; Dong, P.; Li, J.; Liang, X. G. *Plos One* **2014**, *9*.
40. Burger, K.; Mühl, B.; Kellner, M.; Rohrmoser, M.; Gruber-Eber, A.; Windhager, L.; Friedel, C. C.; Dölken, L.; Eick, D. *RNA Biol.* **2013**, *10*, 1623-1630.
41. Triemer, T.; Messikommer, A.; Glasauer, S. M. K.; Alzeer, J.; Paulisch, M. H.; Luedtke, N. W. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E1366-e1373.
42. Farina, V.; Krishnan, B. *J. Am. Chem. Soc.* **1991**, *113*, 9585-9595.
43. Zilbershtein, L.; Silberman, A.; Fischer, B. *Org. Biomol. Chem.* **2011**, *9*, 7763-7773.

Chapter 5: An Optimized Chemical-Genetic Method for Cell-Specific Metabolic Labeling of RNA

5.1 Introduction

Tissues and organs are organized into complex structures that are comprised of many different cell types. Each cell type can have unique gene expression programs that are optimized to maintain cellular fitness. Because of high cellular diversity and tissue heterogeneity, characterizing gene expression profiles within select cell types is an incredibly challenging task ¹.

While several methods to isolate, tag and profile RNA expression in a cell-specific manner have been reported, they have varying sets of limitations. As a first concern, when a cell is removed from its niche, unique patterns of gene expression may be altered, potentially resulting in a transcriptional profile that differs from that of the cell *in vivo*. This is especially true for cells whose main task is changing gene expression in response to environmental cues.² Cell sorting through flow cytometry is commonly used to isolate specific cell types in these cases. However, it is known that cytometry protocols can introduce substantial false-positive RNA signatures due to the stress of sorting or compromised integrity of RNA during processing, in addition to the normal challenges of efficient sorting^{3,4}. Purification of ribosomes (Ribo-tag or TRAP) using cell-specifically expressed affinity tags has been used to overcome the limitations of sorting, but this procedure only isolates RNAs thought to be associated with ribosomes.^{5,6} It is therefore implicit in this approach that typically only mRNAs would be profiled, leaving out many other classes of regulatory and non-coding RNAs that serve critical roles in cellular function

and cellular identity. Interestingly, it has been shown that post-lysis associations can occur between ribosomes and mRNAs, which leads to false-positive producing interactions that were not actually present within the intact cells or tissues^{7,8}. These procedures also rely on antibody-based enrichment, which further introduces the possibility of non-specific binding of mRNAs to antibodies and/or antibody-binding to bacterial proteins used to purify antibodies. These methods, while useful, have underscored the areas of cell-specific expression profiling that require optimization. In this regard, it has created opportunities to develop more stringent methods for RNA tracking using chemical approaches.

RNA metabolic labeling utilizes exogenous nucleobase or nucleoside analogs, which are introduced into nascent transcripts through endogenous cellular machinery⁹. These metabolic reporters are endowed with biorthogonal functional groups that enable isolation of RNA through selective chemistries. For example, nucleoside analogs with alkynyl-, azido-, and thio- modifications have been successfully employed to track RNA expression, catalog RNA-protein interactions, and track RNA localization in cells with high resolution¹⁰⁻¹³. Further, such analogs are soluble and can be systemically perfused into animals, enabling RNA labeling and tracking *in vivo*⁴. However, extending these approaches for cell-specific RNA labeling has been challenging, although this is not well reported. This is because extensive characterization of the endogenous enzymes and metabolic pathways that control analog flux into RNA have yet to be fully characterized.

The most widely used approach for cell-specific metabolic labeling relies on expression of a unique, exogenous enzyme in the cell-type of interest.⁴ The enzyme is cell-specifically expressed, which can then convert an inert metabolic analog into an active metabolic analog. The analog can subsequently be introduced into nascent RNA using

endogenous metabolic enzymes (**Figure 5-1**). The gold standard of this approach utilizes *Toxoplasma gondii* uracil phosphoribosyl transferase (*TgUPRT*) to convert the nucleobase analog 4-thiouracil (4SU) into 4-thiouridine monophosphate. The analog is then phosphorylated twice more to produce the triphosphate for incorporation during transcription¹⁴. This approach has been utilized for nearly 20 years and has been employed in living animals to purify RNA from specific cell types.¹⁵⁻¹⁷ Notably, the implicit selectivity of the *TgUPRT* system relies on the inability of higher order eukaryotes to salvage uracil or uracil analogs. However, it has recently been demonstrated that the *Drosophila melanogaster* UPRT homolog, *krishah* (*kri*), has salvage activity thereby limiting the selectivity of the 4SU/*TgUPRT* system¹⁸. This is concerning, as the active site of UPRT is largely conserved throughout the animal kingdom, suggesting there may be other active homologs in cells and organisms that have been profiled using this method. There may also be other endogenous pathways for uracil salvage that have yet to be identified, which would further limit the utility of such systems mentioned above. Consequently, this renews the challenge of designing highly-stringent methods for cell-specific metabolic labeling of RNA.

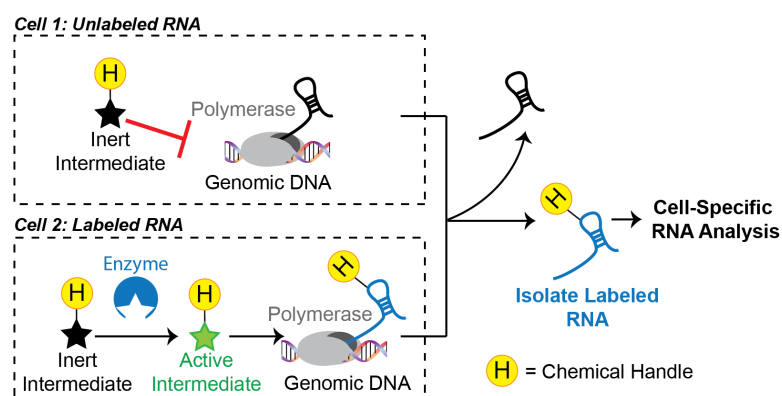


Figure 5-1. Schematic of cell-specific metabolic labeling of RNA. For cell-specific metabolic labeling of RNA, an “inert intermediate” (nucleobase or nucleoside), which

cannot be processed by wildtype cells, is introduced. Only cells ectopically expressing an enzyme, which can convert an “inert intermediate” to an “active intermediate,” can facilitate metabolic labeling of RNA. H = chemical handle for functionalization.

Herein, we evaluate and discover that mammalian cells are indeed capable of salvaging exogenous uracil analogs. We demonstrate that the human homolog of UPRT (*HsUPRT*) as well as uridine monophosphate synthase UMPS (*HsUMPS*), a bifunctional enzyme that catalyzes the final two steps of *de novo* pyrimidine biosynthesis, are capable of this salvage activity. These initial results reveal novel enzyme functionality, which in turn confirms the limitations of cell-specific labeling using 4SU/*TgUPRT*. With this knowledge, we focused on developing a novel nucleoside-enzyme pair for more stringent cell-specific metabolic labeling.

We systematically evaluate uridine analogs paired with engineered or wild type uridine-cytidine kinase 2 (UCK2), a rare kinase isoform. We report the successful pairing 2'-azidouridine (2'AzUd) and wild type UCK2, which was shown to have undetectable background labeling as well as no cytotoxic effects. To more fully understand the interactions and mechanism for analog recognition and catalysis, we performed molecular dynamics simulations wherein we predicted a novel, non-canonical binding mode. To confirm the validity of the models, we obtained the pre- and post-catalytic co-crystal structures of 2'AzUd bound to UCK2. Excitingly, we found similar hydrogen bonding profiles and experimentally confirmed the same binding modes. We note that these analyses are the first of its kind, wherein we have provided structural analysis of a biorthogonal metabolic probe co-crystallized with its enzyme. Lastly, using a litany of experiments we demonstrate

true cell selectivity in a variety of cells. Overall, our novel 2'AzUd/UCK2 pair is highly suitable for cell-specific metabolic labeling of RNA. The results presented herein not only reveal novel aspects of mammalian salvage pathways, but also serve as a new benchmark for designing, characterizing and evaluating methodologies for cell-specific labeling of biomolecules.

5.2 Results and Discussion

Endogenous enzymes can salvage uracil analogs for metabolic labeling of RNA

Understanding how cells metabolize metabolic intermediates is key to developing highly stringent cell-specific methods for metabolic labeling. We and others have recently observed that incubation of moderate to high concentrations of uracil analogs, such as 5-ethynyluracil (5EU) and 4-thiouracil (4SU), into wildtype cells results in metabolic incorporation¹⁹⁻²¹. To further explore background incorporation, we performed a series of experiments exposing many different mammalian cells to 5EU and evaluating incorporation via click-chemistry mediated biotinylation and streptavidin-HRP dot blot detection (**Figure 5-2, a & b**). 5-ethynyluridine (5EUd), which does not rely on uracil salvage, was used as a labelling control (**Figure 5-2, b**). **Figure 5-2, c** demonstrates there is background incorporation in all cell lines from both human and mouse tissues, including dissected primary cortical neurons, a cell type that is non-proliferative. Imaging analysis also revealed that 5EU incorporation mirrors 5EUd, with strongest signal originating in the nucleolus, the location of ribosomal RNA synthesis (**Figure 5-S1**). We also observe a similar pattern of background incorporation for 4SU, the most widely used analog for such studies (**Figure 5-S2**)²². Notably, our background analysis mimics conditions used for *in vivo*

studies with the UPRT/4SU system: 1 mM or above, and over the course of several days^{16,17,23}. These data demonstrate that mammalian cells can indeed process uracil analogs and that cell-specific metabolic labeling experiments utilizing *TgUPRT* may suffer from critical false positives due to this background activity.

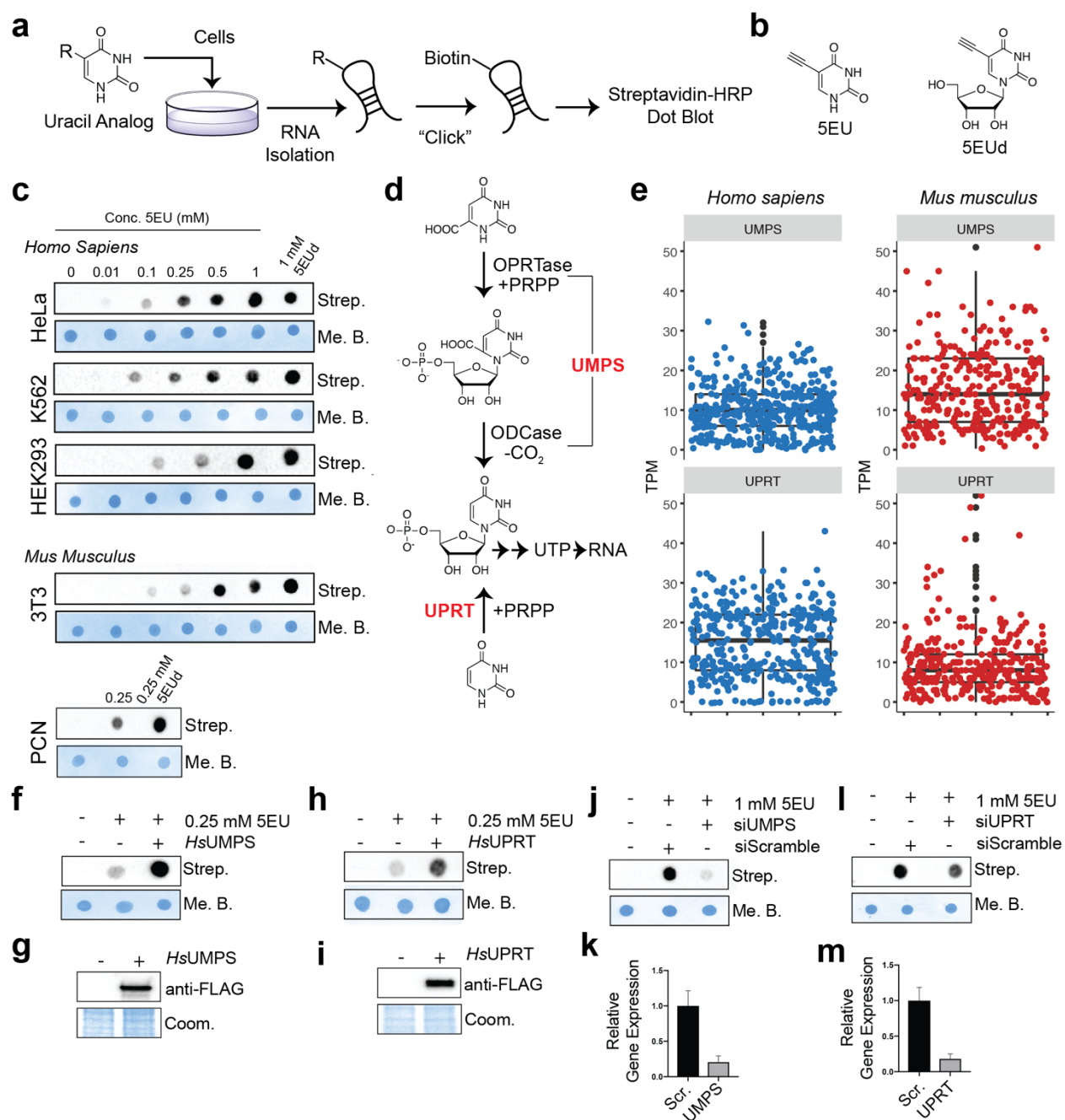


Figure 5-2. Mammalian cells are capable of salvaging exogenous uracil analogs for eventual incorporation into cellular RNA. **a.** Schematic screening experiments for assaying incorporation of exogenous uracil analog into cellular RNA. **b.** Structure of 5-ethynyluracil (for screening; 5EU) and 5-ethynyluridine (analog that is incorporated by wildtype cells; 5EUd) **c.** Time course dot-blot results. Cells were cultured with 5EU for 5 h at indicated concentrations. Total RNA was isolated and biotinylated via azide-alkyne click chemistry and analyzed by streptavidin-HRP dot blot. Strep. = streptavidin-HRP. Me.B. = methylene blue. PCN = primary cortical neurons. **d.** Known enzymatic pathways that could result in salvage of uracil or uracil analogs for eventual RNA metabolic labeling. **e.** RNA sequencing results from mouse and human body map tissue analysis demonstrate that both UPRT and UMPS are highly expressed in a variety of tissues. **f.** Dot blot demonstrating that in the presence of overexpressed *HsUMPS*-FLAG construct, there is more incorporation of 5EU into HEK293T cellular RNA. *Hs* = *Homo sapiens* **g.** Western blot demonstrating specific overexpression of *HsUMPS*-FLAG in HEK293T cells. Coom. = coomassie stain. **h.** Dot blot demonstrating that in the presence of overexpressed *HsUPRT*-FLAG construct, there is more incorporation of 5EU into HEK293T cellular RNA. **i.** Western blot demonstrating specific overexpression of *HsUPRT*-FLAG in HEK293T cells. **j.** Dot blot demonstrating treatment with 25 nM siRNA targeting endogenous *HsUMPS* results in decreased incorporation of 5EU into cellular RNA. Middle sample represents 25 nM scrambled siRNA treatment. **k.** RT-qPCR analysis confirming successful siRNA knockdown of endogenous *HsUMPS* mRNA, normalized to GAPDH control (biological duplicates). **l.** Dot blot demonstrating treatment with 25 nM siRNA targeting endogenous *HsUPRT* results in no significant change to incorporation of 5EU into cellular RNA. **m.** RT-qPCR analysis

confirming successful siRNA knockdown of endogenous *HsUPRT* mRNA, normalized to GAPDH control (biological duplicates).

We next sought to determine the source of this observed uracil salvage. Looking into the literature, we identified two possible candidate enzymes that could utilize uracil or uracil analogs: either the human homolog of uracil phosphoribosyl transferase²⁴ (*HsUPRT*) or *HsUMPS* (**Figure 5-2, d**)²⁵. UMPS functions first through converting orotic acid to orotidine-5'-monophosphate via the N-terminal enzyme orotate phosphoribosyl transferase (OPRTase). The second reaction is carried out by the C-terminal enzyme, orotidine-5'-phosphate decarboxylase (ODCase), which converts orotidine-5'-monophosphate to uridine monophosphate. It is this first reaction that we hypothesize is also able to use uracil and/or uracil analogs as substrates²⁶. In contrast to *HsUMPS*, which is a highly active metabolic enzyme, it is reported that recombinant *HsUPRT* is not catalytically active.¹⁸ However, it has been hypothesized this enzyme may be part of larger complexes or have additional cofactors in cells that facilitate catalytic activity, which were absent from kinetic studies of purified *HsUPRT*. Inspection of mammalian tissue body map data demonstrates expression of both UPRT and UMPS in many different tissue types, which would contribute to background 5EU labeling (**Figure 5-2, e**)²⁷.

To directly test if *HsUPRT* and *HsUMPS* can use 5EU as a substrate, we overexpressed FLAG-tagged clones of each enzyme in HEK293T cells and assessed the effects on labeling. In both cases, we saw a significant difference in the biotinylation signal, relative to the wild-type treated with 5EU (**Figure 5-2, f & g**). This evidence strongly supports that both enzymes are possible contributors to the background we observe in

wildtype cells lines. We note that the signal from *HsUMPS* is much stronger than *HsUPRT*, indicating that 5EU is likely a better substrate and/or UMPS is a more efficient enzyme than *HsUPRT*. Inspection of the *HsUMPS* crystal structure from human cells further supports this notion, where a spacious pocket near the 5-position may tolerate analogs such as 5EU (**Figure 5-S3, a**). We also compared these two treatments to that of 5EU-treated cells expressing the *Toxoplasma gondii* homolog of UPRT (**Figure 5-S3, b**). From this we see similar biotinylation signals from both UPRT homologs, possibly indicating that they have similar catalytic activities. Moreover, because we have seen similar levels of background labeling in mouse cells, we also hypothesize that the mouse homologs of UPRT and UMPS are similarly active. Sequence alignments of both enzymes revealed extensive conservation across a multitude of model organisms (**Figure 5-S4**), which supports the extent of background labeling we observe. Overall, these results suggest that both *HsUMPS* and *HsUPRT* are capable of facilitating 5EU incorporation into RNA.

To understand the relative contributions from each enzyme, we performed siRNA knockdown and assessed the effects on 5EU labelling. Successful knockdown of endogenous *HsUMPS* revealed a prominent reduction of 5EU incorporation into RNA (**Figure 5-2, j & k**). Conversely, while we were able to successfully knock down the *HsUPRT* mRNA, we observed limited reduction in the labelling signal (**Figure 5-2, h & i**). Based on this result, we demonstrate that the background signal in HEK293T cells is largely due to the endogenous expression of UMPS. These results reveal a novel and surprising role for *HsUMPS*, where it is not solely a *de novo* enzyme, but it also has salvage capabilities that facilitate uracil incorporation into RNA. We anticipate that the balance of contributions from each enzyme will be cell-type dependent, as their relative expression

levels can vary across different cell lines (**Figure 5-2, e**). Furthermore, due to the confirmed expression of UMPS in most mammalian tissues, endogenous machinery may severely limit the stringency of the *TgUPRT* system for cell-specific metabolic labeling of RNA.

Identification of the UCK2/2'-azidouridine pair for robust and non-toxic metabolic labeling of RNA

The observations above prompted us to work toward identifying a novel small molecule-enzyme pair that would enable cell-specific metabolic labeling with higher stringency than existing approaches. Ideally, such a pair would intercept pyrimidine triphosphate biosynthesis, as uridine and cytidine are evenly distributed throughout the transcriptome.

The first step in production of pyrimidine triphosphates is carried out by the enzyme uridine/cytidine kinase (UCK), of which there are two isoforms (UCK1 and UCK2; **Figure 5-3, a**)²⁸. UCK1 is a ubiquitously expressed kinase. Conversely, UCK2 is reported to have restricted mRNA expression in placenta tissue and protein expression is undetectable or unconfirmed in tissues *in vivo*²⁹. The lack of UCK2 expression in nearly all but a few select tissues has been highlighted as a possible target for suicide gene therapy³⁰. Human UCK1 and UCK2 show a sequence similarity of 67%, however reports indicate that UCK1 has low substrate tolerance for non-canonical substrates. Whereas, the active site of UCK2 is purportedly more flexible, endowing greater substrate promiscuity than UCK1^{31,32}. We therefore envisioned that UCK2 paired with a uridine or cytidine analog may afford a novel route for cell-specific metabolic labeling.

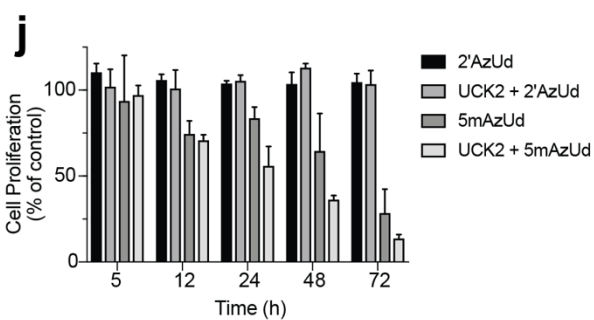
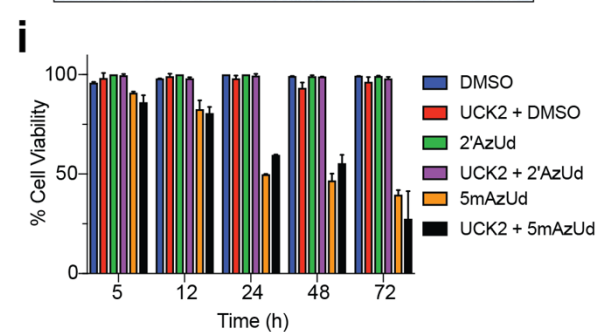
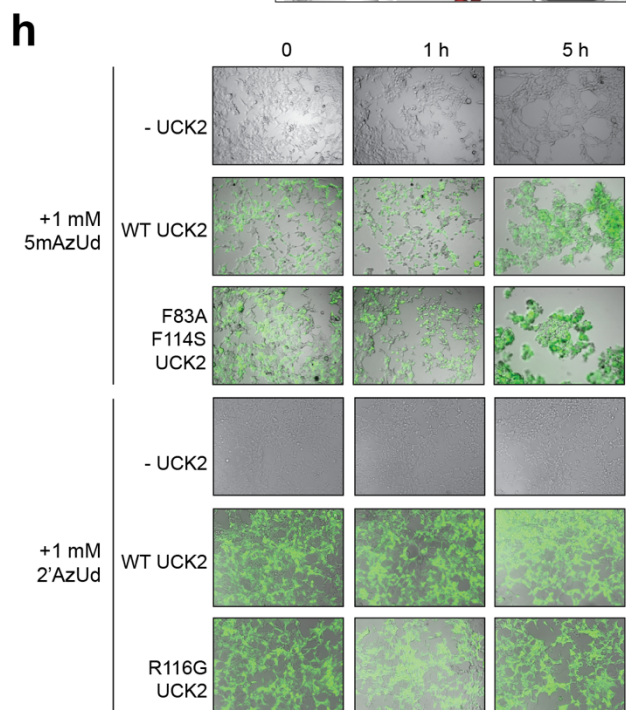
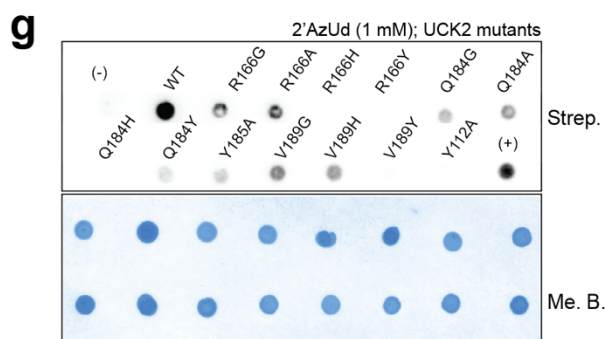
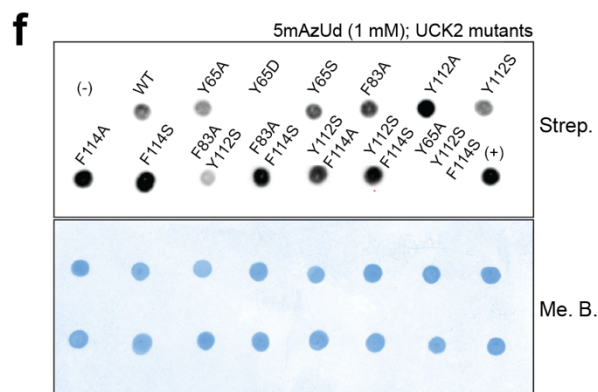
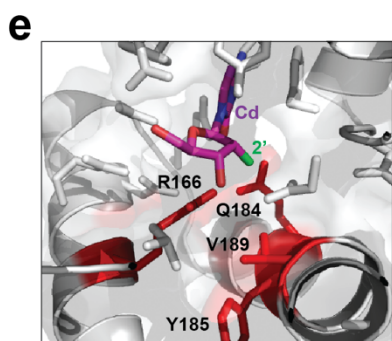
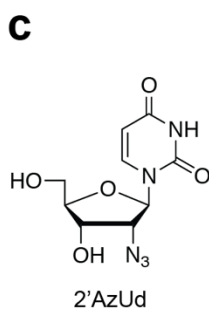
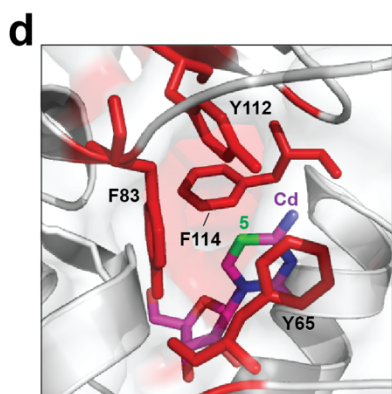
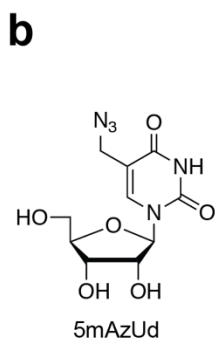
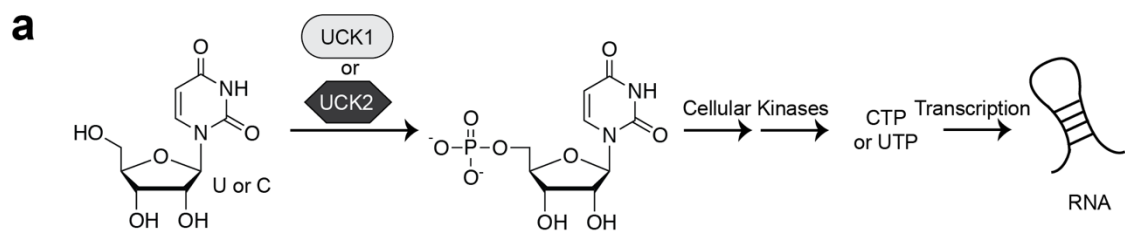


Figure 5-3. Cellular screening reveals a UCK2/2'AzUd pair amenable for metabolic labeling of RNA. **a.** Schematic of pyrimidine salvage pathway through uridine-cytidine kinase and eventual RNA biosynthesis. **b.** Structure of 5-azidomethyluridine (5mAzUd). **c.** Structure of 2'-azidouridine (2'AzUd). **d.** Crystal structure of UCK2 bound to cytidine (PDB: 1UEJ) with residues selected for mutation to accommodate 5mAzUd highlighted in red (Y65, F83, Y112, F114). C5-position is highlighted in green. **e.** Crystal structure of UCK2 bound to cytidine (PDB: 1UEJ) with residues selected for mutation to accommodate 2'AzUd highlighted in red (R166, Q184, Y185, V189). 2'-position is highlighted in green. **f.** Dot blot results from cell-screening of UCK2 mutants with 5mAzUd. HEK293T cells were transfected with a vector encoding WT or mutant UCK2-tGFP and incubated for 48 h. (-) is the negative control with no UCK2-tGFP transfection. (+) is the positive control with 5-ethynyluridine incubation, followed by click with azide biotin. Cells were then treated with 1 mM analog for 5 h. RNA was isolated and biotinylated using azide-alkyne click chemistry, followed by streptavidin-HRP dot blot detection of biotinylated RNA **g.** Dot blot results from screening UCK2 mutants with 2'-azidouridine **h.** HEK293T cell imaging highlighting changes to cell morphology, and potential toxicity of 5-mAzUd-treated cells (1 mM) expressing WT or UCK2 mutants (top). No discernable toxicity observed in 2'AzUd-treated cells (1 mM) expressing WT or mutant UCK2. **i.** Trypan blue assay of HEK293T cell permeability demonstrates that 5mAzUd has an overall greater effect on cell viability than 2'AzUd, relative to DMSO controls. WT or stable-expression UCK2 cells were treated with 1 mM analog for indicated times. Biological duplicates. **j.** Alamar blue assay demonstrates changes to HEK293T cell proliferation in 5mAzUd-treated WT or UCK2-stable expression

cells. Percent proliferation was normalized to DMSO control-treated cells (0.5% final) for each respective time point. Replicates of four.

We began by working to identify and understand the limitations of uridine and cytidine analogs incorporation into cellular RNA, with the goal of identifying analogs with no measurable background. Consistent with our previous findings, 5-azidomethyluridine (5mAzUd) (**Figure 5-3, b**) is not incorporated into cellular RNA¹¹. We also discovered herein that 2'-azidocytidine (2'AzCd) is incorporated into the RNA of WT cells while 2'-azidouridine (2'AzUd) is not (**Figure 5-3, c; Figure 5-S5**). Importantly, this lack of incorporation was assessed at 1 mM concentrations, those which yield high levels of background for uracil analogs as previously discussed (**Figure 5-2**). As such, we hypothesized that WT or engineered UCK2 could be paired with either 5mAzUd or 2'AzUd to label RNA cell-type specifically.

Inspection of the crystal structure of UCK2³¹ bound with cytidine revealed that the 5-position of the nucleobase is crowded by several bulky aromatic groups TYR65, PHE83, TYR112, and PHE114, leaving little space around the base (**Figure 5-3, d**). We hypothesized that single and combinatorial mutation of these residues to smaller functional groups (alanine and serine) would accommodate the azidomethyl- group, thereby enabling the analog to bind and be phosphorylated. We applied the same logic to making UCK2 variants that could accommodate a modification at the 2' position, where we see that ARG166, GLN184, TYR185 and VAL189 could possibly be crowding the space below the sugar (**Figure 5-3, e**). However, we do acknowledge that the area around this region is more cavernous than that of the 5-position. Furthermore, while there were other

candidate residues for substitution, many of the functional groups in this region of the binding pocket are purportedly critical for binding of nucleoside sugar moiety, which ultimately causes a conformational change allowing base binding and catalysis³¹. ARG166 and ASP84 are responsible for this recognition of the sugar, and thus we chose only to substitute ARG166. We thought that the “arginine zipper” along the backside of the binding pocket may be able to compensate for the removal of this residue. Beyond reducing the size of the functional groups, we also sought to make substitutions to aromatic groups, which could possibly trap the azide in the proper conformation, due to π -stacking interactions.

After performing site directed mutagenesis on a vector encoding UCK2 fused to TurboGFP, we had two sets of 13 mutants for screening against 5mAzUd and 2'AzUd. We utilized in-cell screening to assay the activities of each enzyme with the analogs, wherein we transiently expressed each vector in cells and monitored expression via the GFP reporter (**Figure 5-S6**). We note that 24 of the 26 mutants were robustly expressed in cells, with the exception of R166H which significantly affected cell proliferation and Y65D, which appeared to cause severe inclusion body formation. Cells expressing either WT or UCK2 variants were then treated with 1 mM analog for 5 h, followed by total RNA isolation, click chemistry-mediated biotinylation, and dot blot detection. As shown in **Figure 5-3, f**, we observed that many of the mutants as well as WT UCK2 yielded biotinylation signals after treatment with 5mAzUd. Strikingly, we also observed that all of these pairs resulted in complete cell-detachment and the appearances of cytotoxicity (explained below). While removing residues near the 5-position of 5mAzUd allowed it to be phosphorylated and incorporated, the same was not true for the mutations to accommodate 2'AzUd. In-cell screening revealed that WT had the best activity with this analog, resulting in the greatest

level of incorporation (**Figure 5-3, g**). As previously mentioned, the residues around the 2' position are important for sugar recognition, so this region could be refractory to mutations or perhaps the azide is oriented differently than we expect in the binding pocket.

In all cell types treated with 5mAzUd and expressing WT or mutant variants of UCK2, we noticed significant cell detachment and visible cytotoxicity upon incubation. To investigate this further, we monitored 5mAzUd-treated cells (100 μ M and 1 mM) expressing WT or mutant UCK2 that yielded the most robust biotinylation signal from the RNA labeling screen. Surprisingly, after only 1 hour changes in cell morphology could be observed and after 5 hours many UCK2-variant expressing cells had become partially or completely detached, compared to the DMSO control (**Figure 5-3, h; Figure 5-S7 – 12**). Furthermore, even WT UCK2- expressing cells appeared to be susceptible to 5mAzUd cytotoxicity (**Figure 5-S7**). We hypothesize that the observed toxicity is kinase-dependent, where it takes upwards of 12 hours for un-transfected cells to show signs of toxicity. At 24 hours post incubation, we saw complete detachment of cells at both concentration treatments, eliminating the need to image at longer time points. Additionally, we note that several mutants (Y112A and F83A; F114S) appear to have modest effects on cell growth and proliferation (**Figure 5-S8, 5-S10**).

In light of these findings, we demonstrate that 5-mAzUd will not be viable for profiling nascent transcription in a cell-selective manner for several reasons. First, analog-mediated toxicity leaves a small window for expression profiling, and furthermore will not be amenable to *in vivo* applications. Second, cytotoxicity will cause unintended changes in gene expression, confounding biologically meaningful results.

For 2'AzUd, the top performing enzymes were WT UCK2 and the R166G and R166A

variants. To further explore the effects of these systems on cell growth and proliferation, we performed a longer time course of imaging in each of these treatments at 100 μ M and 1 mM (**Figure 5-3, h; Figure 5-S13 – S16**). We found that the UCK2 2'AzUd pairing, in general, was innocuous to cell viability, where cell morphology was consistent with DMSO-treated controls, in contrast to 5mAzUd. However, over time (48 - 72 h post transfection) we note that the ectopic expression of R166A followed by the R166G variants appear to impede cell growth at extended time points, regardless of DMSO or analog treatment (**Figure 5-S14, 5-S15**). We speculate that the effects of these point mutations could in the long term affect the ability of the protein to properly fold and form quaternary structures, in addition to changing the binding and catalysis of canonical substrates.³¹ Based on these findings, we demonstrate that the WT UCK2 enzyme will best be paired with 2'AzUd for cell-type specific labeling.

To further ensure that ectopic expression of UCK2 will not have larger impacts on the cells over time, we created a stable expression HEK293T cell line (**Figure 5-S17**). We sought to gain further insight into the extent of analog cytotoxicity, where we performed two basic cell viability assays for quantitative comparisons at extended time points. Trypan blue staining was used to determine whether the cell membrane is intact once cells have detached, and an alamarBlue assay was used to monitor cell proliferation over time with analog treatment relative to the DMSO control. There were no significant changes in cell permeability or any detachment in 2'AzUd treated cells, WT or UCK2-expressing (**Figure 5-3, h & i**) Conversely, 5mAzUd treatment consistently resulted in complete cell detachment, which over time lead to loss of membrane integrity and reduced proliferation (**Figure 5-3, i & j**). We see similar patterns of 5mAzUd-mediated cytostasis in the cell proliferation data,

however it is more apparent at earlier time points because detectable changes to the reducing environment occurred sooner in the cells (**Figure 5-3, j; Figure 5-S18, a & b**). Furthermore, these effects were accelerated in cells expressing UCK2, where there is a significant difference between proliferation of WT and UCK2-expressing cells treated with 5mAzUd. As expected, we did not note any significant changes to proliferation of either WT or UCK2-expressing cells when treated with 2'AzUd (**Figure 5-3, j; Figure 5-S18, c & d**).

A major concern when using RNA metabolic reporters is the potential for cellular toxicity. However, extensive analysis of toxicity is rarely investigated, and only 4-thiouridine (4SUd) has been reported to be toxic³³. These observations, coupled with the substantial cytotoxic effects of 5mAzUd, prompted us to ask whether other non-cell specific nucleosides may have similar effects. To gain a more comprehensive understanding, we performed the same viability assays for cells treated with commercially available analogs 4SUd, 5EUd and the analog 2'AzCd reported herein. We found that 4SUd was indeed the most toxic of these analogs, with cells showing visible signs of stress and detachment around 12 hours post treatment (**Figure 5-S19**). 5EUd caused similar effects but at delayed time points relative to 4SUd. While Trypan staining showed no obvious changes to membrane integrity relative to the DMSO control, the proliferation of cells treated with 4SUd followed by 5EUd was significantly reduced over the 72-hour time course (**Figure 5-S20**). Changes to cell proliferation from 5EUd also mirrored results from a recent report where other C5 modified nucleosides cause cytotoxicity in cells³⁴. The analog 2'AzCd had the smallest effect on proliferation. Generally, we see that all three of the analogs displaying concerning cytotoxic effects possess modifications on the base. In addition to the type of chemical functionality installed, we postulate that functionalization at sites of the

nucleoside that frequently contact other moieties may disrupt cellular function. Traditionally, modifications to the Watson-Crick face of the base have been avoided for this reason, however this logic may be extended to the base in general. Conversely, we see that 2' modifications are by-and-large the most innocuous to cell function, which further underscores their utility for RNA labeling. Overall, our results demonstrate that in contrast to the other uridine analogs explored above, 2'AzUd is not incorporated into WT cells and the WT UCK2/2'AzUd pair is non-toxic, yielding robust metabolic labeling of nascent RNA.

Binding and Kinetic Assays to Compare UCK1/2 Isoforms

Methods for cell-specific metabolic labeling rely on the successful identification of high fidelity enzyme-small molecule pairs, yet these efforts often end at cell-based screening. As previously mentioned, in the case of cytidine and uridine salvage, we are aware that there are two functional isoforms of this kinase, albeit with divergent expression levels. To more closely characterize the selectivity observed in cell culture, we performed biochemical assays on both UCK1 and UCK2 using our 2'-modified analogs as well as the canonical substrates. We hoped that such studies would offer greater insight into selectivity, as well as provide additional context to any limitations of this system that may arise from UCK1-dependent activity.

We expressed and purified UCK1 and UCK2, and subjected them to binding affinity assays. This was done through taking advantage of the highly aromatic binding pocket of UCK, where TYR68/65, PHE86/83 and PHE190/188 of UCK1/2 base stack, creating a signature fluorescence signal in the absence of bound substrate (**Figure 5-S21**). Upon substrate binding, this fluorescence becomes quenched, enabling the dissociation constants

to be determined as a result of this change.

Table 5-1. Binding and kinetic analysis of UCK1/2 with canonical and 2'-azido analogs.

Substrate	K_D (nM)		K_m (mM)		V_{max} ($\mu\text{mol/mg/min}$)		k_{cat} (s^{-1})		k_{cat}/K_m ($\text{s}^{-1} \text{M}^{-1}$)	
	UCK1	UCK2	UCK1	UCK2	UCK1	UCK2	UCK1	UCK2	UCK1	UCK2
Uridine	56.4 ± 11.9	5.8 ± 3.2	4.06 ± 0.82	0.15 ± 0.03	11.56 ± 1.42	78.77 ± 6.31	5.44 ± 0.67	37.40 ± 0.46	$1.3 \pm 0.3 \times 10^3$	$2.4 \pm 0.5 \times 10^5$
Cytidine	15.9 ± 1.1	3.7 ± 0.5	1.92 ± 0.32	0.13 ± 0.01	13.9 ± 0.77	70.61 ± 2.49	6.54 ± 0.36	33.52 ± 1.18	$3.4 \pm 0.6 \times 10^3$	$2.6 \pm 0.2 \times 10^5$
2'azidouridine	180.8 ± 17.1	13.4 ± 4.8	n.d.	240 ± 57	n.d.	2.16 ± 0.46	n.d.	1.03 ± 0.22	n.d.	4.3 ± 1.2
2'azidocytidine	65.7 ± 10.6	17.8 ± 6.3	118 ± 62	205 ± 43	0.23 ± 0.10	3.64 ± 0.69	0.11 ± 0.05	1.73 ± 0.33	0.9 ± 0.6	8.4 ± 1.8

We found that UCK1 binds cytidine with the highest affinity (K_d of 15.9 ± 1.1 nM), and the introduction of a 2'-azido group results in a four-fold decrease in affinity (65.7 ± 10.6 nM) (**Table 5-1**). UCK1 affinity for uridine is lower than that of cytidine (56.4 ± 11.9 nM), and UCK1 binds 2'AzUd with the lowest affinity (180.8 ± 17.1 nM). We demonstrate that UCK2 binds both canonical nucleosides (cytidine and uridine; 3.7 ± 0.5 nM and 5.8 ± 3.2 nM, respectively) with very high affinity, while 2'AzCd and 2'AzUd and have lower affinities (13.4 ± 4.8 mM 17.8 ± 6.3 nM, respectively). These studies revealed two key trends, the first being that UCK2 has higher affinity for the canonical or 2'azido substrates compared to UCK1. These results suggest that UCK1 is more discriminating in ligand binding while UCK2 is more promiscuous.³² Furthermore, for either isoform, the affinity is always higher for cytidine or 2'AzCd than it is for uridine or 2'AzUd. This is supported by published structural studies reporting additional hydrogen bonding contacts for the base of cytidine.³¹ Importantly, we observe that UCK1 binds 2'AzUd with the lowest affinity of all

the nucleosides, indicating that this pairing is the least likely to support catalysis.

Despite the similarities in binding affinity, there is a clear disparity in RNA labeling with 2'AzCd and 2'AzUd. To augment these studies, we further assessed the kinetic properties of UCK1 and UCK2 using an NADH-coupled ATPase assay (**Figure 5-S22**).³⁵ We determined the Michaelis constants of uridine and cytidine phosphorylation, which were in agreement with reported literature values for UCK1 and UCK2 (**Table 5-1**).³⁶ Furthermore, the K_m values of UCK1 were 27-fold and 15-fold higher than UCK2 for uridine and cytidine, respectively. The K_m values were significantly higher for 2'AzCd (118 mM and 205 mM for UCK1 and UCK2, respectively), while the catalytic efficiency (K_{cat}/K_m) of 2'AzCd and UCK2 was 9-fold higher than UCK1. Importantly, there was no measurable activity for 2'AzUd with UCK1, while the K_m of 2'AzUd with UCK2 was 240 mM and the catalytic efficiency was half that of 2'AzCd with UCK2. In summary, we found that despite having reasonable agreement in binding affinities, there is a striking preference for catalytic activity of UCK2 and 2'AzUd alone.

We wanted to further understand the propensity of UCK2 and not UCK1 to phosphorylate 2'AzUd, so we aligned their primary sequences and structures (PDB: 2JEO, 1UEJ) (**Figure 5-S23, a & b**). There is nearly 67% sequence identity with an additional 45 residues that are similar in terms of functional group. However, there is a distinguishable difference in one helix of the lid domain, which was characterized by Suzuki et al.³¹ In UCK2, a structured helix of the lid is critical for base recognition, and undergoes a large induced fit where ARG176 and HIS117 close down on the base, hydrogen bond and mediate catalysis. In this way, mobility of the UCK2 lid domain allows for tighter contacts to be made with its substrate. As shown in **Figure 5-S23, c**, the helix of UCK1 is observably

shorter and this region of the lid domain is less structured, possibly reducing the ability of the enzyme to tightly contact its substrate. Furthermore, while the sidechain of ARG176 in UCK2 is facing the binding pocket, the sidechain of ARG175 in UCK1 is oriented to the exterior, and would be less likely to form hydrogen bonds with a substrate. Taken together, we speculate that UCK1 is more rigid and has an overall reduced ability to change conformations and tightly bind a substrate. Given its already lowered affinity for uridine over cytidine, we expect that the azide modification even further reduces the ability of 2'AzUd to bind, preventing measurable catalysis.

Structural analysis of the 2'AzUd/UCK2 pair

To better understand the interactions of UCK2 with the 2'-azido analogs, we performed molecular dynamics (MD) simulations using the UCK2-cytidine complex (PDB: 1UEJ) as our starting model (**Supplementary Note 1 – Molecular Dynamics Studies**). We docked both 2'AzUd and 2'AzCd into the UCK2 active site, ran short (40 ps) MD simulations and found two general binding modes. The first is canonical in nature, as it mimics the orientation of the sugar captured in the reported structure (PDB: 1UEJ), whereby the ribose ring oxygen is pointed outward (**Figure 5-4, a; Figure 5-S24, a**). However, we also observed a novel binding mode that is unique, wherein the sugar has flipped around the N-glycosidic bond, orienting the oxygen inward (**Figure 5-4 b; Figure 5-S24, b**). To distinguish between the binding mode of each ligand, we refer to the canonical binding mode as 2'AzUd_{canc} / 2'AzCd_{canc} and the novel, flipped binding mode as 2'AzUd_{flip} / 2'AzCd_{flip}. For each binding mode, we ran five separate 100 ns MD simulations for a total simulation time of 500 ns. The results of our MD simulations were surprising, as they suggest that the

azide significantly alters the way in which the analog can engage with UCK2. As such, we wanted to further confirm these models through obtaining the X-ray crystal structures of UCK2 with 2'AzUd.

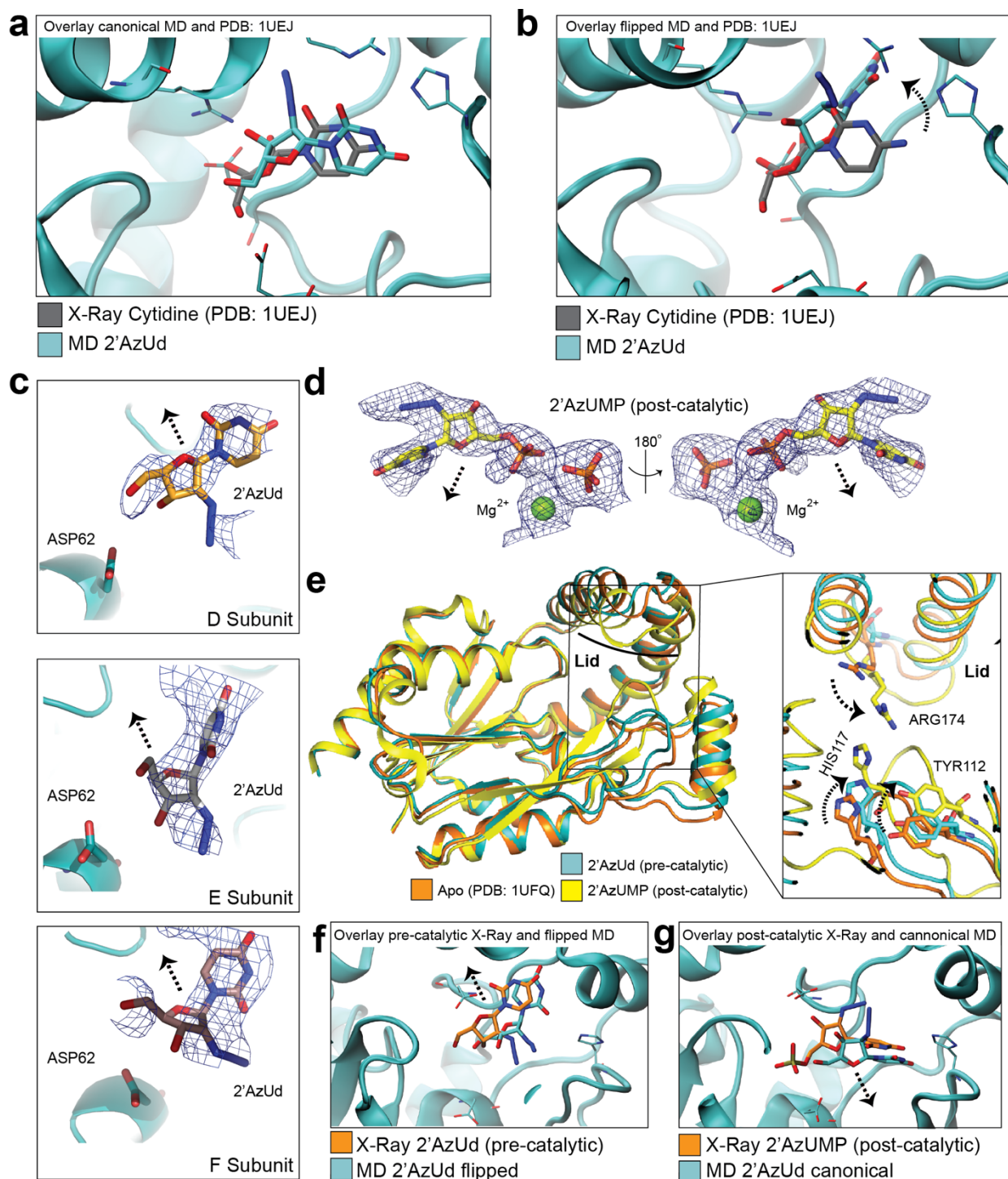


Figure 5-4. Structural analysis of the UCK2/2'AzUd pair reveals the mechanism for selectivity. **a.** Overlay of simulated 2'AzUd in the canonical binding mode (2'AzUd_{can}) and cytidine from PDB: 1EUJ. The ribose oxygen is oriented away from the active site in both the predicted MD conformation and 1EUJ. **b.** Overlay of simulated 2'AzUd in the flipped binding mode (2'AzUd_{flip}) and cytidine from PDB: 1EUJ. 2'AzUd_{flip} orientation of the ribose in the active site pocket is opposite to that of the canonical binding mode (denoted with a dashed arrow). **c.** Electron density of 2'AzUd in the pre-catalytic states in subunits D, E and F. Each of the orientations for 2'AzUd are from the same view as in a. and b. **d.** Electron density of 2'AzUMP in the post-catalytic state, where the substrate is flipped in an orientation opposite that of the pre-catalytic structures. Specifically, the ribose is flipped and now facing down and out of the active site. **e.** Structural superimposition of 2'AzUMP, 2'AzUd and apo-UCK2 demonstrates conformational changes in the lid domain and opposing loop containing key residues in the pre- and post-catalytic states, that are unique to UCK2. **f.** Overlay of 2'AzUd_{flip} MD and 2'AzUd structure determined herein and. **g.** Overlay of 2'AzUd_{can} MD and post-catalytic 2'AzUMP structure determined herein.

To validate our MD predictions, we determined the structures of UCK2 complexed with 2'AzUd in the presence and absence of ATP, and also with 2'AzCd and ATP (**Supplementary Note 2 – Crystallographic Studies**). UCK2 complexed with 2'AzUd (UCK2•2'AzUd) crystallized with eight protein subunits in the asymmetric unit (ASU). Of the eight protein subunits, only three (D, E, and F) show reasonable electron density for a bound 2'AzUd (**Figure 5-4, c**). We obtained a post-catalytic structure of UCK2 complexed with 2'AzUd and ATP, wherein a phosphorylation reaction occurred *in situ* to yield

2'azidouridine mono-phosphate (UCK2•2'AzUMP) (**Figure 5-4, d**). Similar to UCK2•2'AzUMP, UCK2 also complexed with 2'AzCd, ATP and MgCl₂ crystallized with two protein subunits in the ASU, both of which contain the product of the phosphorylation reaction, 2'-azidocytidine monophosphate (2'AzCMP) and a Mg²⁺ ion (**Figure 5-S25, a**). The structures of 2'AzCMP present good electron density, and also superimpose well with CMP from the reported structure of UCK2•CMP (PDB: 1UJ2)³¹, as the cytidine bases in both subunits are similarly oriented (**Figure 5-S25, b**).

Superimposition of pre- and post-catalytic UCK2 structures revealed structural transitions in the lid domain and a loop directly opposite. In the post-catalytic structure, several key residues have collapsed in toward the binding pocket making contacts with the substrate, providing evidence for UCK2-specific induced fit binding to promote catalysis (**Figure 5-4, e; Supplementary Note 2 – Crystallographic Studies**).

Figure 5-S26, shows our reported structures of UCK2•2'AzUd from subunits D, E, and F. Herein, the ribose of 2'AzUd is consistently oriented opposite to that of cytidine in PDB: 1UEJ, wherein the oxygen of the ring is facing inward. Strikingly, all three subunits present in what we have designated as the flipped conformation. We also see that this binding mode of 2'AzUd is moderately variable amongst each subunit, most notably in the rotation of the base (**Figure 5-4; Figure 5-S26 b, c**). This suggests that there may be some flexibility in how the base moiety binds in the active site. We hypothesize that this may be due to compensatory networks of hydrogen bonding.

For a direct comparison, **Figure 5-4, f** shows the MD model simulating 2'AzUd_{flip} overlaid with the 2'AzUd crystal structure (subunit D). The minimum computed root mean square deviation (RMSD) between them was 2.88 Å, indicating similar orientations within

the binding pocket. While the MD model has the base rotated (resembling the orientation found in subunit F), the remainder of the two structures are very similarly oriented. Most importantly, we see that the sugar orientation in both structures is indeed flipped around the N-glycosidic bond compared to PDB: 1UEJ (**Figure 5-4, b & f**).

Simulated hydrogen bond (H-bond) frequencies of one 2'AzUd_{flip} simulation reveals contact for key residues ASP62 (~7%), ASP84 (~30%), and ARG174 (lid region; ~60%), which are likewise found in our UCK2•2'AzUd structures (**Figure 5-S27; Table 5-S1**). While there was contact with catalytic residue ASP62, in one simulation, the remainder of the simulations remain in the starting configuration. Overall, these residues likely play a significant role in forming the flipped binding conformation.

Figure 5-S24, b shows the MD model of 2'AzCd in the flipped binding mode, wherein the ligand makes contact with the catalytic residue. Here, we observe that the 3' hydroxyl of 2'AzCd coordinates with the backbone of UCK2 near ASP62, which then brings the 5' hydroxyl within 3.0 Å of the catalytic residue. While the MD simulation does not show bond breaking or forming, we expect that this distance is close enough to facilitate catalysis. Compared to our simulations of 2'AzUd in the flipped binding mode, we find that the 2'AzCd ligand forms more frequent contacts with the catalytic ASP62 (**Figure 5-S28**) for roughly over 50% of the simulation time. This is consistent with the observations from the biochemical analyses, where the kinetics of 2'AzCd and UCK2 are faster than 2'AzUd. Other key residues involved in binding of 2'AzCd_{flip} appear to be ARG169 and ARG174. If given more simulation time for 2'AzUd in the flipped binding mode, we believe we would see a similar increase in contact with ASP62 as was seen for 2'AzCd.

Overall, this data supports that the MD prediction describing the existence of a flipped binding mode is confirmed by the UCK2•2'AzUd x-ray structures. Taken together, they the computational and experimental structures corroborate that the flipped binding mode is the pre-catalytic conformation, which can form contacts that are required to facilitate catalysis.

In **Figure 5-S29**, we compare our post-catalytic structure (orange) of UCK2•2'AzUMP (subunit B) to the reported structure from PDB: 1UEJ (gray) containing a bound cytidine molecule. We see that the structures very closely overlay, illustrating that the binding mode we have designated as canonical is representative of the post-catalytic state. **Figure 5-4, d** shows the conformational differences between 2'AzUMP of subunit A and subunit B, where the rotation of the base has changed nearly 180 degrees. This reasserts our findings from our flipped model simulations and structures, wherein there is considerable flexibility in the way the base moiety can orient itself within the binding site.

After simulating 2'AzUd starting from the canonical binding mode, we overlaid this with our reported structure of UCK2•2'AzUMP (subunit B) (**Figure 5-4, g**). We observed that the post-catalytic structure very closely overlays with the predicted binding mode from our MD simulations of 2'AzUd_{can} (RMSD 2.44 Å), wherein the ribose is oriented outward. When we analyzed the H-bond contact frequencies for 2'AzUd_{can} across all five simulations, we saw virtually no contact with the catalytic residue ASP62 (**Figure 5-S30**). These results support that the canonical binding mode is indeed representative of the post-catalytic state, as interactions with catalytic residue ASP62 are very infrequent, which could be expected post-catalysis.

Throughout the simulations of 2'AzUd_{can}, we also see more frequent contacts with HIS117 (~40%), TYR112 (~80%), and ARG166 (~60%) than in the 2'AzUd_{flip} simulations (~10%, ~1%, and ~10% respectively) (**Figure 5-S27, 5-S30**). This supports the observed structural changes in the UCK2•2'AzUMP structure that shows these residues shifted in toward the binding pocket (yellow, **Figure 5-4, e**). Among the H-bonds in our UCK2•2'AzUMP structures, HIS117 and ARG166 contacts concur with the MD model, and are similarly present only in this post-catalytic state, (**Figure 5-S30; Table 5-S1**).

In **Figure 5-S31**, we illustrate that our simulated model of 2'AzCd in the canonical binding mode (cyan) almost directly overlays with both our structure of UCK2•2'AzCMP (orange) and the reported structure of bound cytidine from PDB 1UEJ. Like our crystallized structure of 2'AzUMP, we similarly find that the post-catalytic 2'AzCMP is oriented in the canonical binding mode. Excitingly, the computed RMSD of our MD model to the UCK2•2'AzCMP structure is merely 0.72 Å, further providing evidence that the canonical binding mode represents the post-catalytic state. We speculate that the improved RMSD of 2'AzCMP compared to 2'AzUMP is the result of the base being similarly oriented. As previously mentioned, cytidine has more H-bonding capabilities, which could enhance the base stabilization. From analysis of hydrogen bond contact frequencies, we observe that 2'AzCd_{can} forms contacts mostly with ASP84, PHE114, HIS117, ARG166, and ARG176 (**Figure 5-S32**). Notably, there is virtually no contact with the catalytic residue ASP62, again in agreement with our findings from 2'AzUMP.

Collectively, our observations suggest that both 2'AzUd and 2'AzCd are forming unique contacts and have a novel, flipped binding mode for catalysis. Interestingly, in a simulation of flipped 2'AzUd, we observe that the ligand starts off in the flipped binding

mode, and samples different conformations where 2'AzUd is able to form contact between the 5' hydroxyl and ASP62. The 2'AzUd molecule then undergoes a conformational change to the canonical binding mode, loses contacts with key residues, and promptly exits the binding site. This simulation appears to illustrate both binding modes, wherein each is shown to play a role in the phosphorylation and release of 2'AzUd from the binding pocket. Given these observations as well as the agreement with our experimental structures, we believe we have strong evidence to support that the flipped binding mode resembles the pre-catalytic state and the canonical binding mode represents the post-catalytic state.

Within the azido-analog bound structures, we have captured a series of conformations between the *apo* and fully closed conformation. Among the UCK2•2'AzUd structures, none of the subunits were in the closed conformation, reflecting the mobility of bound 2'AzUd (**Figure 5-4, e; Figure 5-S33**). Indeed, while one molecule occupies a position in or near canonical binding, the other two appear to be in a state of entry or exit.

In the structure of UCK2•2'AzUMP, both subunits are oriented closer to the closed conformation, wherein there is a noticeable shift in HIS117, TYR112 and ARG174 toward the binding pocket (**Figure 5-4, e**). This is reflected in an increased commonality between the residues coordinating 2'AzUMP to residues previously observed coordinating CMP and UTP (PDB: 1UJ2, 1UEI) (**Table 5-S1**).³¹ Finally, the structure of the UCK2•2'AzCMP complex best approximates the closed conformation (**Figure 5-S33**).

Clearly, UCK2 is able to achieve a catalytically-active conformation with azide-modified substrates, as we see bound product and can monitor the reaction experimentally. However, these new structures of UCK2 provide more detailed models of substrate binding, from which we can elucidate the conformational changes involved in substrate binding and

recognition. Moreover, the observed structures in the X-ray data suggest a highly dynamic binding site that promotes substrate promiscuity, further corroborating previous reports that UCK2 is the less selective isoform. UCK2, unlike UCK1, is likely able to achieve a catalytically-active conformation with 2'AzUd due to its induced fit mechanism, which is facilitated by the flipped binding mode. Overall, these analyses demonstrate the structural intricacies of ligand binding and further reveal the molecular interactions that enable engagement with UCK2 for efficient kinase activity. Excitingly, we were able to discover a new binding mode through MD simulations, which we then experimentally corroborated through crystallographic studies. Notably, these results are the first-of-its-kind, wherein structural analysis of a novel metabolic reporter has provided greater insights into activity and selectivity with its metabolic enzyme.

Demonstration of stringent cell-specific metabolic labeling

Our observations thus far have suggested that the UCK2/2'AzUd pair would be a stringent system for cell-specific metabolic labeling, with the caveats that UCK1 is similarly refractory to 2'AzUd phosphorylation in cells, and the endogenous expression of UCK2 is low or undetectable in the targeted cell type. To more thoroughly investigate this, we pooled RNA expression data of UCK1 and UCK2 across a myriad of tissues from the human body map, as well as HEK293T and HeLa cells. **Figure 5-5, a** demonstrates that immortalized cell lines (293T & HeLa) possess overall higher RNA expression of UCK1 and UCK2 than any tissues³⁷. As a result, we used these cell lines as benchmarks to assess any off-target labeling, with the expectation that they possess the upper levels of UCK1 and UCK2 expression *in vivo*.

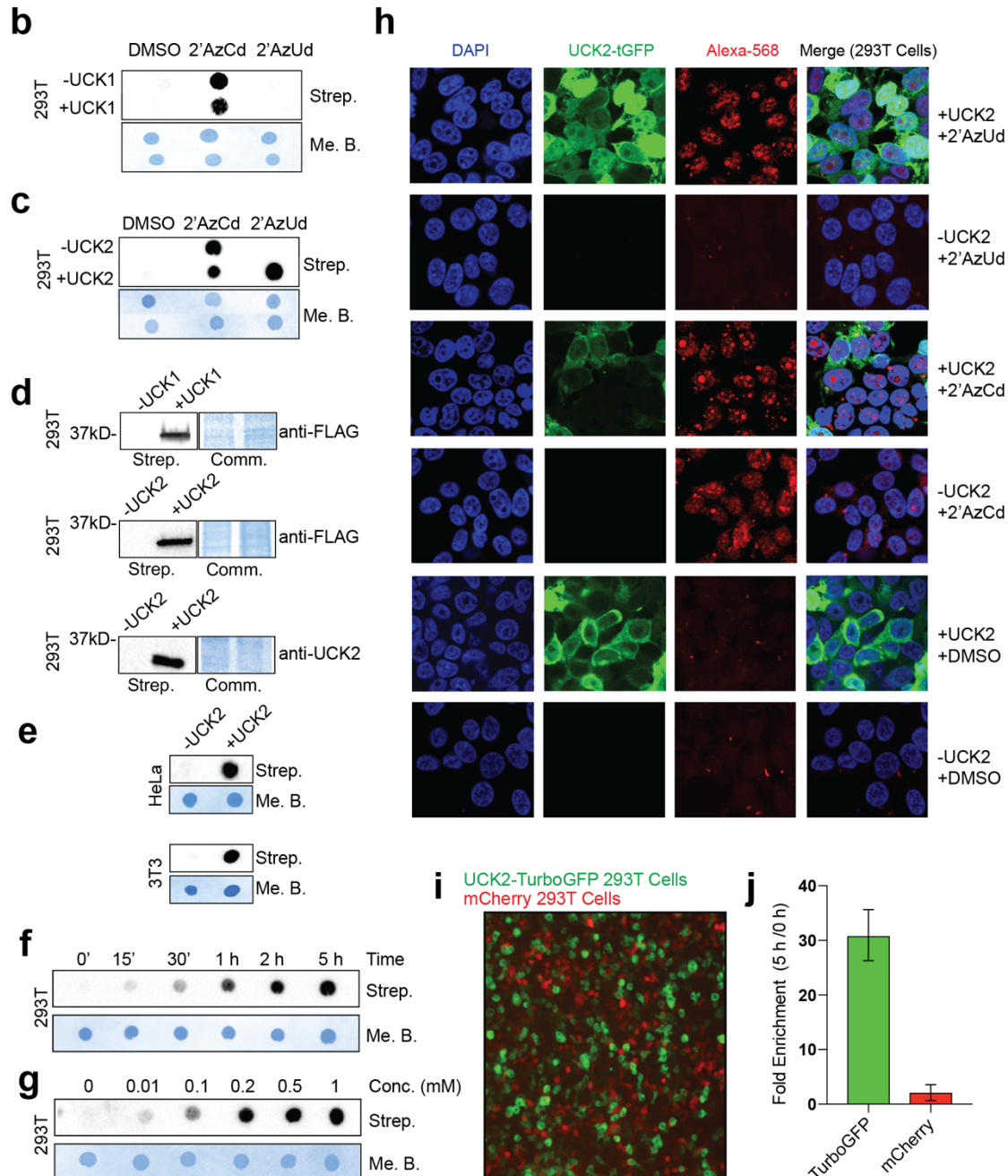
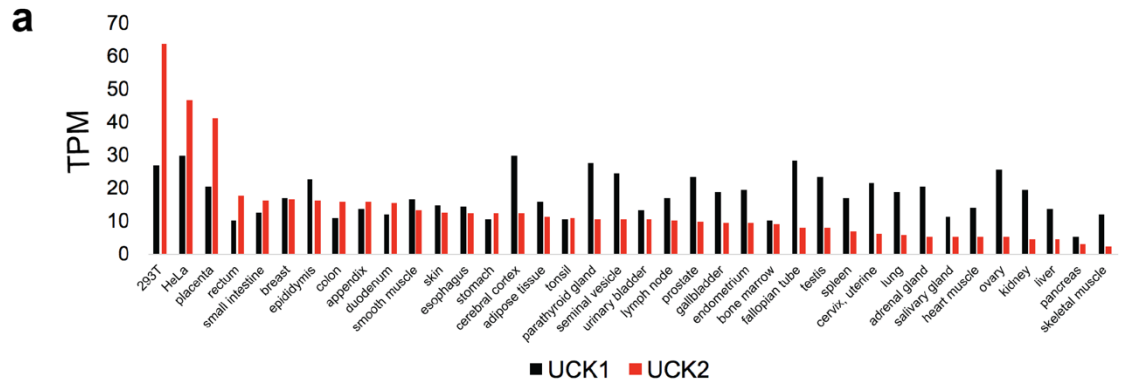


Figure 5-5. Demonstration of cell-specific metabolic labeling with UCK2/2'AzUd pair.

a. TPM (transcripts per million count) expression of UCK1 and UCK2 from tissues and two immortalized cell lines to be profiled. **b.** Dot blot demonstrating no incorporation of 2'AzUd in the presence of UCK1-FLAG overexpression. **c.** Dot blot demonstrating incorporation of 2'AzUd only in the presence of UCK2-FLAG overexpression. **d.** Western blots detecting overexpression of UCK1-FLAG and UCK2-FLAG constructs. Overexpression was confirmed with anti-FLAG or additionally UCK2 was also detected via anti-UCK2. **e.** Additional mammalian cells not expressing UCK2 are unable to introduce 2'AzUd into their RNA unless UCK2 is overexpressed **f.** Time course analysis of 2'AzUd incorporation into cellular RNA from cells expressing UCK2-FLAG and treated with 1 mM analog. **g.** Concentration titration of 2'AzUd incorporation into cellular RNA in cells expressing UCK2-FLAG. Cells were treated at indicated concentrations for 5 h **h.** Cellular imaging demonstrates 2'AzUd is incorporated into cellular RNA of only cells expressing UCK2-TurboGFP (green). Cells transfected with UCK2-TurboGFP (or mock) were incubated with 1 mM 2'AzUd for 5 h. Cells were fixed, permeabilized and clicked with Alexa-568 alkyne followed by imaging of labeled RNA (red). **i.** Co-culture cell image. HEK293T cells separately transfected with UCK2-TurboGFP fusion or mCherry were re-plated in co-culture. 2'AzUd was then added for 0 or 5 h. **j.** UCK2-tGFP/mCherry co-culture total RNA was isolated, biotinylated, reverse transcribed and cDNA:RNA hybrids were enriched using magnetic streptavidin beads. Eluted cDNA was subjected to qPCR analysis to assess enrichment of TurboGFP or mCherry. Analysis was performed in biological duplicates and technical triplicates.

While our kinetics analysis supports that there is no reactivity of UCK1 with 2'AzUd, we similarly wanted to demonstrate that this is true in cell culture. We therefore assessed 2'AzUd labeling in the presence of overexpressed UCK1 or UCK2, with expression confirmed through Western blot (**Figure 5-5, b – d**). This demonstrates that 2'AzUd was only incorporated into cellular RNA when UCK2 was overexpressed. These results are consistent with the *in vitro* data, wherein 2'AzUd is demonstrated to not be a viable substrate for UCK1. Furthermore, we also confirmed overexpression through direct detection of the UCK2 enzyme; in contrast, we were unable to detect expression of UCK2 protein in WT cells (**Figure 5-5, d**), despite having relatively higher reported mRNA expression (**Figure 5-5, a**).

We also co-expressed UCK1 and UCK2, and remarkably saw that concurrent expression actually reduced the activity of UCK2 with 2'AzUd (**Figure 5-S34, a**). Imaging of the individual and co-expressed isoforms demonstrated that UCK2 localization switched from the cytosol to the nucleus, where UCK1 is typically located (**Figure 5-S34, b**). We postulate that UCK1-UCK2 heteromers cause conformational changes that lower the binding affinity or activity with 2'AzUd. These results further support the selectivity between the ubiquitously expressed UCK1 and the rare isoform UCK2, and that this system will be selective.

We tested if there was spurious incorporation of 2'AzUd in other immortalized cell lines and found no detectable 2'AzUd background signal, in stark contrast to uracil analogs (**Figure 5-5, e; Figure 5-2**). This was true for both human as well as mouse (3T3) cultured cell lines, demonstrating that the selectivity we have observed can be applied across different mammalian organisms. This provides us with evidence that the known limited

UCK2 expression will not produce appreciable background labeling and will provide the desired selectivity for experiments in primary cells and tissues.

Next, we sought to determine the concentration and time needed for appreciable metabolic labeling. From a time course of 2'AzUd, we were able to observe signal from metabolic incorporation within 15 minutes (**Figure 5-5, f**), making this system ideal for providing a snapshot of transient gene expression. From a concentration titration, we observed labeling at concentrations as low as 0.01 mM (**Figure 5-5, g**). This was encouraging, because despite the relatively slow kinetics of UCK2 with the 2'AzUd analog, exogenous enzyme expression and its multiple-turnover capability facilitates detectable incorporation at very low concentrations.

We further demonstrated the specificity of metabolic labeling by imaging nascent RNA, where we expressed UCK2-TurboGFP and treated cells with 2'AzUd. Imaging of nascent RNA revealed the strongest signal originating from the nucleolus, the site of ribosomal RNA synthesis (**Figure 5-5, h**). We also see labeling in all cells treated with the nucleoside 2'AzCd, regardless of UCK2 expression. Lastly, we envisioned utilizing this system to enrich for tagged RNAs from a complex cellular mixture. As a proof of concept, we separately transfected cells with UCK2-TurboGFP or mCherry alone, followed by replating as a heterogeneous population (**Figure 5-5, i**). We then treated cells as normal with 2'AzUd, and isolated the RNA. Following click chemistry-mediated biotinylation, reverse transcription, and streptavidin enrichment, the extent of TurboGFP and mCherry cDNA enrichment was quantified utilizing qPCR. As predicted, the TurboGFP cDNA was considerably enriched at 5 h compared to the untreated 0 h control (**Figure 5-5, j**). Conversely, mCherry was not significantly enriched above background. Owing to the

sensitivity of qPCR amplification, we speculate that this level of selectivity will provide the desired stringency in complex cultures and tissues from which lowly-amplified sequencing libraries will be made. In summary, these results demonstrate that the 2'AzUd/UCK2 pair has the required stringency and resolution for cell-specific metabolic labeling of RNA.

5.3 Conclusion

Analysis of biomolecules that reside within specific cell types continues to be a major challenge in method development. In particular, methods to analyze RNA expression profiles or dynamics of RNA expression within unique cell types can be especially challenging, as they frequently have limitations that can result in false-positives in the data. These shortcomings present unique opportunities for the development of more stringent and novel approaches. Key to this endeavor are careful design aspects and thorough testing of the system, which will ensure the method is reliably stringent in different cell types as well as species.

An underappreciated facet of developing cell-specific metabolic methods for analysis of biomolecules is truly understanding the cellular pathways responsible for installing the biorthogonal reporter molecules. Herein, we have demonstrated that there must be additional considerations made when using *TgUPRT* and uracil analogs. We have made the seminal discovery that mammalian cells can salvage uracil analogs for their eventual incorporation into RNA. Notably, we have specifically identified that the enzymes *HsUMPS* and *HsUPRT* are capable of contributing to metabolic flux of exogenous uracil analogs into RNA. This surprising result not only expands our understanding of mammalian

salvage capabilities, but also cautions against indiscriminate use of uracil analogs in cell types that will have high levels of background activity.

Because endogenous expression of enzymes that will salvage uracil analogs seems largely unavoidable, we set out to intercept other steps of nucleoside salvage pathways. We focused on the step of uridine-cytidine mono-phosphorylation. As a result, we have discovered that pairing uridine-cytidine kinase 2 (UCK2) with biorthogonal metabolic probes can be a highly-stringent approach for cell-specific metabolic labeling of RNA. First, we used basic protein engineering and in-cell screening to identify which UCK2/analog pairs enable robust metabolic labeling without background (-UCK2) labeling. We successfully found several UCK2 variants worked with 5mAzUd and WT UCK2 paired with 2'AzUd. However, we observed significant toxicity associated with the former, and thus we performed more extensive analysis of the cytotoxic effects of each system. Following this investigation, we demonstrated no measurable toxicity of our selected analog, 2'AzUd, in stark contrast to all 5mAzUd pairs. This investigation was critical, as the implication of such methodologies is their eventual use in living systems, which cannot sustain reduced viability in addition to changed gene expression in response to toxicity. We further investigated the cytotoxic effects of other non-specific metabolic probes, such as 4SUd, 5EUd and 2'AzCd. We note a general pattern of toxicity wherein base modifications produced the most significant effects on cell viability, compared to sugar modification. As a result, we selected the UCK2/2'AzUd pair for further characterization.

Biochemical analyses of purified enzyme to determine the affinity of 2'AzUd and 2'AzCd with UCK1 and UCK2, revealed similar patterns of binding affinity. However, kinetic analysis revealed clear disparities in rates of catalysis, wherein only UCK2 could

phosphorylate 2'AzUd. We performed MD simulations to understand the interactions of 2'AzUd in the binding pocket of UCK2, from which we predicted a novel, flipped binding mode. We further validated our models through crystal structures of UCK2 complexed with 2'AzUd, wherein the pre-catalytic structures assumed the flipped binding mode, while the post-catalytic structures represent the canonical binding mode.

Interestingly, our initial attempts to mutate UCK2 to better accommodate 2'AzUd did not yield better labeling than the WT enzyme. In light of our MD and crystallographic findings, the evidence of an alternative binding mode could possibly explain this failure to identify a better UCK2 variant. The residues and contacts we hypothesized were important were supported by analysis of a reported structure, wherein the ligand was captured in the canonical binding mode. Overall, our structural analysis revealed several key findings including the necessary interactions for binding as well as the binding modes pre- and post-catalysis, which in the future, we envision will guide further analog design and UCK2 engineering. To our knowledge, this is the first comprehensive biochemical and structural analysis of a metabolic enzyme with a biorthogonal metabolic reporter.

Lastly, we have demonstrated stringent cell-specificity in a series of cellular experiments. Consistent with *in vitro* experiments, overexpression of UCK1 does not result in background labeling. This is in stark contrast to the uracil (UMPS and UPRT) data presented in **Figure 5-2**. Systematic testing in other cell lines also showed clear dependency on overexpression of UCK2. Analysis of publically available gene expression datasets also demonstrates that UCK2 is more highly expressed in the tested cell lines than *in vivo*. As such, our system should give clear cell-specificity and non-toxicity when used in these settings. We also demonstrated that RNA can be imaged and purified in a cell-specific

manner, further supporting the notion that our UCK2/2'AzUd novel enzyme/small-molecule should serve as a highly-stringent pairing for cell-specific RNA analysis.

In summary, we have developed and thoroughly characterized a novel approach for cell-specific RNA tagging. Through our investigation, we have uncovered a novel aspect of mammalian cell metabolism and biology: uracil salvage. This inspired us to identify a new, disparate approach for high-resolution labeling of RNA. We anticipate that the advancements described herein will serve as a new paradigm for developing such methods, and will supply the field with a much-needed, stringent system for cell-specific RNA tracking and analysis. Finally, we note that both 2'AzUd and UCK2 expression plasmids are commercially available (**5.5 Methods**), and as such this system can be easily adopted throughout the scientific community.

5.4 Acknowledgements

The 2'-modified analogs used in preliminary studies were synthesized by Miles Kubota. Whitney England performed analysis of transcriptomic data. Bonnie Cuthbert, along with the help of Robert Quechol, performed binding studies and obtained the structures of UCK2. Nathan Lim performed molecular dynamics simulations of the analogs docked into the binding pocket of UCK2.

5.5 Methods

Vectors for transient transfection:

pCMV6-UCK2-GFP (Origene #RG202406)

pCMV6-UCK2-Myc-DDK (Origene #RC202406)

pCMV6-UCK1-Myc-DDK (Origine #RC220876)

UCK1-GFP (UCK1 ORF cloned into pCMV6-GFP vector through restrictions sites Xho1 and SacII)

pCMV6-UPRT-Myc-DDK (Origene #RC205030)

pLJC2-UPMS-3xFLAG (a gift from David Sabatini, Addgene plasmid # 87975)

<http://n2t.net/addgene:87975> ; RRID:Addgene_87975)

pcDNA3.3-mini-INT-HA-TgUPRT (a gift from Mike Cleary Lab)

pMSCV-IRES-mCherry (a gift from Ellen Rothenberg , Addgene plasmid # 80139)

<http://n2t.net/addgene:80139> ; RRID:Addgene_80139)

Vector for stable line generation:

Lentivirus EF1a-UCK2-P2A-EGFP; PGK-Puro

List of qPCR primers used:

Target	Forward	Reverse
GAPDH	TGTCAAGCTCATTTCTGGTAT	CTCTCTTCCTCTTGTGCTCTT
UPRT	ATGGTGACCACTCCAACAGG	TTGCCTCACCGCTTCTCATT
UMPS	CGGCTCTCACGCAGAGTATTA	CAAAGGCAGGCTTGACGAAC
TurboGFP	CTACCACTTCGGCACCTACC	GTACTTCTCGATGCGGGTGT
mCherry	GATAACATGGCCATCATCAAGGA	CGTGGCCGTTACGGAG

All primers were empirically determined to have efficiencies between 95 – 103%.

Site-directed mutagenesis of UCK2-GFP vector.

Amino acid substitutions in the open reading frame of UCK2 were made using the Q5 Mutagenesis kit (New England Biolabs). Back to back primers containing mutations were designed for the targeted region, allowing for exponential application of the plasmid. Mutants with more than one substitution were made from template containing one previously made mutation. Following kinase, ligation and Dpn1 digestion of the amplification reaction, the mutant plasmids were transformed into chemically competent DH5alpha *E. coli*, selected via ampicillin resistance. Following plasmid isolation and sequence validation, 1-2 µg of each mutant was transfected into HEK293T cells, and analyzed for kinase activity via streptavidin-HRP dot blot, described below.

Mutation	Forward	Reverse
Y65A	GGATAGCTTCgcccGTGTCCTTACCTC	TGGCTCAGGATGACCACCTGCT TC
Y65S	GGATAGCTTCagcCGTGTCTTACCTC	TGGCTCAGGATGACCACCTGCT TC
Y65D	GGATAGCTTCgacCGTGTCTTACCTC	TGGCTCAGGATGACCACCTGCT TC
F83A	CCAGTTCAACgccGACCACCCGGATGCC	CCCTTCAGGGCTTTGGCCTTCT
Y112A	GATCCCCGTGgccGACTTTGTCTC	TGGACTGTTTTCCCTTCAGTGA T
Y112S	GATCCCCGTGagcGACTTTGTCTC	TGGACTGTTTTCCCTTCAGTGA T

F114A	CGTGTATGACgccGTCTCCCATTCC	GGGATCTGGACTGTTTTCCCTT C
F114S	CGTGTATGACcagcGTCTCCCATTCC	GGGATCTGGACTGTTTTCCCTT C
F83A; F114S	CCAGTTCAACgccGACCACCCGGATGCC	CCCTTCAGGGCTTTGGCCTTCT
F83A; Y112S	CCAGTTCAACgccGACCACCCGGATGCC	CCCTTCAGGGCTTTGGCCTTCT
Y112S; F114A	cgccGTCTCCCATTCCCGGAAG	tcgctCACGGGGATCTGGACTGT
Y112S; F114S	cagcGTCTCCCATTCCCGGAAG	tcgctCACGGGGATCTGGACTGT
Y65A; Y112S; F114S	GGATAGCTTCgccCGTGTCTTACCTC	TGGCTCAGGATGACCACCTGCT TC
R166G	GCTCTCACGcgccGTATTAAGGGAC	CGGGTGTCCGCATCTGTATCCA C
R166A	GCTCTCACGcgccGTATTAAGGGACATCAGC	CGGGTGTCCGCATCTGTATCCA C
R166H	GCTCTCACGCcacGTATTAAGGGACATCAGC	CGGGTGTCCGCATCTGTATCCA C
R166Y	GCTCTCACGctacGTATTAAGGGACATCAGC	CGGGTGTCCGCATCTGTATCCA

		C
Q184G	GATTTTATCTggcTACATTACGTTCGTCAAGC CTGC	TGCTCAAGATCCCTGCCTCTCT C
Q184A	GATTTTATCTggcTACATTACGTTCGTCAAGC CTGC	TGCTCAAGATCCCTGCCTCTCT C
Q184H	GATTTTATCTcacTACATTACGTTCGTCAAGC	TGCTCAAGATCCCTGCCTCTCT C
Q184Y	GATTTTATCTtacTACATTACGTTCGTCAAGC CTG	TGCTCAAGATCCCTGCCTCTCT C
Y185A	TTTATCTCAGgccATTACGTTCGTCAAG	ATCTGCTCAAGATCCCTGCCT
V189G	CATTACGTTCggcAAGCCTGCCT	TACTGAGATAAAATCTGCTCAA GATC
V189A	CATTACGTTCgccAAGCCTGCCT	TACTGAGATAAAATCTGCTCAA GATC
V189H	CATTACGTTCcacAAGCCTGCCTTTG	TACTGAGATAAAATCTGCTCAA G
V189Y	CATTACGTTCtacAAGCCTGCCTTTG	TACTGAGATAAAATCTGCTCAA G

Cell lines and culture conditions.

HEK293T, HeLa, NIH-3T3, (ATCC) were cultured in DMEM containing glucose and L-glutamine (Corning) supplemented with 10% FBS, 1% penicillin and streptomycin. K562

(ATCC) cells were cultured in RPMI with L-glutamine (Life Technologies) supplemented with 10% FBS, 1% penicillin and streptomycin. Primary cortical neurons (PCN) were isolated as previously described²⁰ and maintained in Neuroal Basal media supplemented with 1x B27 supplements, 1% penicillin and streptomycin and 1% glutamax. All cell lines were grown at 37°C, in an atmosphere of 5% CO₂.

Cell transfection and stable line generation.

Cells were transiently transfected with plasmid 1-2 µg of plasmid using jetPRIME DNA transfection reagent (Polyplus Transfection) and were incubated for 24 – 48 h to allow for expression. A UCK2 stable expression cell line was generated in HEK293T cells following transduction with lentivirus encapsulated EF1a-UCK2-P2A-EGFP-PGK-Puro vector. Cells were puromycin selected to obtain UCK2 positive cells, which were then clonally isolated through limiting dilution.

Nucleoside analog treatment in cultured cells.

5-ethynyl uridine (Click Chemistry Tools), 4-thiouracil (Sigma), 4-thiouridine (Cayman Chemical). 5-ethynyluracil was synthesized as previously described¹⁹ 2'azidocytidine and 2'aziduridine were purchased from Carbosynth. Analogs were added at various concentrations and times to complete culture medium from 200 mM DMSO stocks (final concentration of <1% DMSO).

Alamar Blue Assay.

HEK293T cells were seeded at densities of 10,000 cells into a 96-well plate and grown for 1 day in supplemented DMEM. Cells were then incubated for various lengths of time with each analog at titrated concentrations or DMSO, with 0.5% DMSO final. Following incubation, 10 μ L of alamarBlue reagent (Thermo Fisher Scientific) was added to each well and incubated for 4 hours to assess the cells' ability to irreversibly reduce resazurin to resorufin, resulting in a color change from blue to red. Absorbance readings were taken at 570 nm and 600 nm as a reference, and were used to calculate cell proliferation relative to the DMSO-treated control in replicates of five to six.

Trypan Blue Assay.

HEK293T cells were seeded at densities of $1.5 - 2.5 \times 10^5$ into 6 well plates and grown for 18 - 24 hours. WT or UCK2-expressing cells were then incubated for 12, 24 and 48 h with analogs 4SUd, 5EUd, 2'AzCd, and 2'AzUd at 1 mM final concentration. Following incubation, cells were trypsinized, removed from the plate and spun down. The supernatant was aspirated and the cells were re-suspended in 1 mL of culture medium, after which 10 μ L was aliquoted for Trypan staining. To each aliquot, 10 μ L of 0.4% Trypan Blue solution (Gibco) was added and 10 μ L of this solution was applied to a hemocytometer for assessment of cell viability. Total and dead cell counts was recorded for each of the four corners of the grid, and repeated for biological duplicates of two.

RNA isolation and biotinylation via azide-alkyne click chemistry.

After labeling with azide or alkyne modified nucleosides/nucleobase, total cellular RNA was harvested using Trizol Reagent (Invitrogen) following the manufacturer's instructions, with two final washes with 70% ethanol, followed by resuspension in 30 – 50 μ L of nuclease free water. CuAAC or SPAAC reactions were prepared using 5-10 μ g of total RNA in solutions containing either: 500 μ M CuSO_4 , 2.5 mM THPTA (Sigma), 2.5 mM sodium ascorbate, and 1 mM alkyne or azide-biotin (Sigma) with a final DMSO concentration of 10% or 10 mM Tris-HCL pH 7.4, 10 mM EDTA, 1 mM DBCO-biotin (Sigma). CuAAC reactions were incubated with at room temperature for 30 min at 400 RPM and SPAAC reactions were incubated at 37 °C for 1 h. Biotinylation reactions were purified using Zymo RNA clean and concentrator-5 spin column according to the manufacturer instructions, with RNA eluted in 10 μ L of nuclease free water.

4SU and 4Sud RNA biotinylation.

Cells were treated as previously described above, and RNA was Trizol extracted. Biotinylation was carried out in a reaction containing 100 μ M MTSEA-biotin (TRC chemicals), 1 mM EDTA, 20 mM HEPES pH 7.4 and 5-10 μ g of total RNA. The biotinylation reaction was allowed to proceed for 1 h at room temperature, after which it was column purified using Zymo clean & concentrator-5.

HRP-streptavidin Dot blotting.

Equal amounts of column-purified, biotinylated RNA were spotted onto 2x SSC equilibrated Hybond-N+ membrane (GE Healthcare), allowing spots to dry in between. The membrane

was crosslinked using the auto crosslink function at 254 nm on a UV Stratalinker (Stratagene). The membrane was then blocked in a 10% SDS blocking solution followed by incubation with 1:10,000 dilution of high sensitivity streptavidin-HRP (Thermo Fisher Scientific) in blocking solution. The membrane was washed x2 in a 1:10 solution of blocking solution and x2 in Tris-saline buffer. It was then incubated in SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) according to manufacturer's instructions and imaged on a ChemiDoc MP imaging system (Bio-Rad) under the chemiluminescence channel. Following imaging, the blot was stained in a 0.04% methylene blue, 0.3 M sodium acetate solution to visualize all RNA as a loading control. The blot was imaged again using the colorimetric channel.

siRNA knockdown of nucleoside metabolism enzymes in mammalian cells.

siRNA pools targeting human uridine monophosphate synthase (UMPS) and uracil phosphoribosyltransferase (UPRT) (SMARTpool: ON-Targetplus siRNA, Dharmacon) were re-suspended to working concentrations of 20 μ M. Cells were transfected with 25 nM of siRNA using INTERFERin transfection reagent, (PolyPlus Jet Fusion) and allowed to incubate for 72 h. Cells were then treated with analog as previously described, and were subsequently collected for extraction and dot blotting or RT-qPCR, or lysate was collected for Western blotting.

RT-qPCR analysis of knockdown.

Cells treated with siRNA were isolated and total RNA was extracted. Following RNA precipitation, 1 – 2 μ g of RNA was subjected to DNase Turbo (Thermo Fisher Scientific)

treatment according to the manufacturer's instructions, followed by Zymo column purification. RNA was then subjected to reverse transcription using the SuperScript IV Reverse Transcriptase kit (Thermo Fisher) with oligo(dT) priming. Following first strand synthesis, the reactions were inactivated by heating to 85 °C for 10 min, after which cDNA was directly subjected to qPCR analysis. qPCR was carried out on a CFX Connect Real-Time PCR Detection System using the iQ Supermix (Bio-Rad). Relative gene expression of UPRT or UMPS to the housekeeping gene GAPDH was determined through the delta-delta C_q expression method. Knockdowns were performed in biological duplicates and technical triplicates.

Western blotting detection of protein ectopic expression and knockdown.

Cells were seeded at a density of 2.5×10^5 into 6 well plates and were grown to 60-70% confluency. Cells were transiently transfected with plasmid using jetPRIME transfection reagent (Polyplus transfection) followed by incubation for 24 – 72 h. After allowing for protein to express, the media was decanted and 300 uL of lysis buffer (150 mM NaCl, 5 mM EDTA pH 8.0, 50 mM Tris, pH 8.0, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) spiked with DNase Turbo and RNaseOut (Thermo Fisher Scientific) was added to the cells. Cells were scraped and transferred to tubes, which were incubated on ice for 15 min to allow for lysis to occur. Lysates were spun down at 14,000 RPM, 4 °C for 25 min, and the soluble fraction was collected for protein quantification using the Qubit Protein Assay kit and Qubit fluorimeter (Thermo Fisher Scientific). Equal amounts of protein (10-20 ug) lysate were mixed with 2x Laemelli sample buffer (Bio-Rad) containing β -mercaptoethanol, and boiled at 95 °C for 5 min. Samples were cooled to room temperature, loaded onto a pre-

cast 4-20% TGX gel (Bio-Rad) and resolved. Proteins were transferred onto nitrocellulose in a midi-transfer stack using the Trans-Blot Turbo transfer apparatus (Bio-Rad). The membrane was blocked in 5% milk, TBST for 1 h followed by hybridization with primary antibody at room temperature for 1 h. UCK2 over-expression was first detected through 1:500 of anti-UCK2 human (Abcam, Ab167683). UCK2, UMPS, UCK1 and UPRT were detected through 1:1000 dilution of anti-FLAG (Cell Signaling technologies #14793). UMPS was detected at 1:500 dilution of anti-UMPS (Abcam, Ab155763), and UPRT was detected at 1:500 dilution of anti-UPRT (Abcam, Ab111701). The membrane was washed 3 times for 5 min using TBST, followed by incubation with secondary antibody, either 1:2000 anti-mouse HRP or anti-rabbit HRP (Cell Signaling Technologies, #7076, #7074). The membrane was then incubated with SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) according to the manufacturer's instructions and imaged on a ChemiDoc MP imaging system (Bio-Rad) under the chemiluminescence channel. The membrane was subsequently washed and stained with PageBlue Protein Staining solution (Thermo Fisher Scientific) for protein loading controls. Following de-staining overnight, the membrane was imaged under the colorimetric channel.

UCK1 and UCK2 expression and purification.

A plasmid encoding an N-terminally His-tagged UCK1 was purchased from addgene (plasmid number 42374). UCK1 was expressed in *E. coli* BL21 (DE3) Gold cells (Agilent). Cells were grown aerobically at 37 °C in LB medium containing 50 µg/mL ampicillin. UCK2 expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactoside at an OD₆₀₀ ≈ 0.6-0.8 and expressed overnight at 18 °C. Cells were collected by centrifugation at

5,500 x g for 10 min and then resuspended in resuspension buffer (50 mM Tris/HCl, pH 7.5, 0.5 M NaCl and 1 mM BME). Cells were lysed in resuspension buffer containing 10 mg/mL lysozyme and 1 mM phenylmethylsulfonyl fluoride (PMSF), and centrifuged for an hour at 4 °C at 14000 rpm to remove debris. Clarified lysate was loaded onto a column of Ni-NTA beads, Ni-NTA resin was washed with resuspension buffer supplemented with 20 mM imidazole. UCK1 was eluted with stepwise increases in imidazole concentration (50 mM - 2 M) in resuspension buffer. Eluents containing UCK1 were combined, dialyzed overnight into resuspension buffer and concentrated to a ~0.5 mg/mL (~18 µM) using a 30-kDa centrifugal concentrator (Centricon; Millipore).

A plasmid encoding an N-terminal GST fusion UCK2 (residues 1-250) construct as described by Suzuki *et al.* was generated by Genscript in the vector pGEX-6p (Genscript) (Suzuki et al. 2004; Suzuki et al. 2003). UCK2-GST was expressed in BL21 (DE3) Gold cells and lysed as described for UCK1, however UCK2-GST was solubilized in 0.1 M sodium phosphate pH 7.4 and 0.5 M NaCl. Clarified lysate was loaded onto a GSTrap FF column (GE Healthcare), washed with PBS buffer and eluted with a gradient of 50 mM Tris/HCl pH 8, 200 mM NaCl and 10 mM reduced glutathione. Eluted protein was dialyzed into 50 mM Tris/HCl pH 7, 150 mM NaCl, 1 mM EDTA and 1 mM DTT before the GST-tag was removed by digestion with Pierce HRV 3C Protease (Thermo Scientific) at 4 °C overnight. Remaining GST-tagged UCK2 was separated using glutathione sepharose 4B resin (GE Healthcare,). The untagged UCK2 protein was further purified by gel filtration on a Superdex 200 column (GE Healthcare) in 20 mM Tris/HCl pH 7.5, 150 mM NaCl.

UCK1 and UCK2 fluorescence-based nucleoside binding experiments.

The affinities of UCK1 and UCK2 for uridine, cytidine, and their 2'-azido analogs were determined using fluorescence microscopy. Titrations of nucleoside into UCK1/UCK2 were monitored by fluorescence, observing the change of tyrosine fluorescence intensity with increasing concentration of ligand. UCK1 and UCK2 (100 nM) and the titratable ligand (at 1.25, 2.5 and 5 μ M) were prepared in 20 mM Tris/HCl pH 7.5, 150 mM NaCl. Ligands were titrated into protein solutions in 5-500 nM increments with 3 minute incubations between measurements. The tyrosine fluorescence emission was acquired at 303 nm through excitation at 274 nm using a Hitachi F-4500 fluorescence spectrophotometer at 25 °C. The fluorescence emission at 303 nm when plotted against increasing concentrations of ligand provides the dissociation constant (K_d) at half maximal binding. The K_d was determined by the non-linear least squares regression equation for bimolecular model binding: $Y = \frac{(F_{max}-F_{min})x}{K_d+x} + F_{min}$, where x is the ligand concentration, F_{max} refers to the maximal observed fluorescence and F_{min} to the minimal observed fluorescence. Experiments were performed in triplicate.

NADH-coupled activity assay.

Purified enzyme was subjected to kinetic analysis using an ATPase-coupled assay that monitors the reduction of NADH to NAD⁺ via decrease in absorbance at 340 nm.^{35,38,39} Each reaction contained 20 mM HEPES (pH 7.2), 100 mM KCl, 2 mM MgCl₂, 50-300 μ M cytidine or uridine for UCK2, 50-8000 μ M cytidine or uridine for UCK1, 100 μ M – 10 mM 2'-AzCd or 2'-AzUd for UCK2, 1-40 mM 2'-AzCd or 2'-AzUd, 1 mM ATP, 250 μ M NADH, 250 μ M phosphoenolpyruvate, 0.5 U lactate dehydrogenase from rabbit muscle (Sigma), 0.5 U

pyruvate kinase from rabbit muscle (Sigma), and 100 nM of purified UCK1 or UCK2 enzyme. Continuous measurements were taken at 340 nm in 150 μ L fill volumes in 96-well UV-transparent plates (Corning), using a Spark Multimode microplate reader (Tecan), with samples run in triplicate. Kinetic constants were obtained through a nonlinear fit of the data using the Michaelis-Menten equation with Prism 8 GraphPad software.

UCK2 crystallization experiments.

UCK2 crystals were grown by hanging drop vapor-diffusion. UCK2 (15 mg/mL) was screened for crystallization with 100 mM 2'-azidouridine (2'AzUd); or 50 mM 2'-AzUd, 5 mM ATP, 10 mM MgCl₂ or 50 mM 2'-azidocytidine (2'AzCd), 5 mM ATP and 10 mM MgCl₂. All resulting crystals were mounted and collected under cryoconditions at 70K with the addition of 10-25% (vol/vol) glycerol as a cryoprotectant to the reservoir condition.

Datasets were indexed, integrated and scaled in HKL2000 (Otwinowski and Minor 1997), solved by molecular replacement using Phaser and refined using Phenix.refine in the Phenix suite. Molecular replacement searches were performed with UCK2 structures PDB: 1UJ2 (UCK2•CMP) and 1UEI (UCK2•UTP)³¹.

Structure of UCK2 complexed with 2'AzUd.

UCK2 with 2'AzUd crystallized in 50 mM BIS-TRIS pH 6.5, 25 mM ammonium sulfate, and 32% pentaerythritol ethoxylate and were cryo-cooled after 12 days. The dataset was collected at ALS 8.2.1, resulting in a 100% complete dataset up to 3.1 Å (**Table 5-S1**). A molecular replacement solution of the processed dataset was found with a search for eight molecules with PDB: 1UJ2. Refinement was carried out in subsequent rounds using

Phenix.refine (Adams et al. 2010) and Coot (Emsley and Cowtan 2004) to a final R-work/R-free of 21.1/25.8. Data collection and refinement statistics are presented in **Supplementary Note 2 – Crystallographic Studies Table 5-1.**

Structures of UCK2 complexed with ATP, MgCl₂ and 2'AzCd or 2'AzUd.

UCK2 was screened for crystallization with 2'AzCd, MgCl₂ and ATP. The best diffracting crystals were found in the PACT suite condition 58: 0.2 M sodium/potassium phosphate, 20% PEG 3350. Crystals were optimized and grown in this condition for 10 days before cryo-cooling for data collection. To obtain a structure of UCK2 with ATP and 2'AzUd, UCK2 was screened for crystallization with 2'AzUd, ATP, and MgCl₂. However, the best diffracting crystals of UCK2 with ATP and 2'AzUd were observed in the UCK2•2'AzCMP crystallization condition. Following crystallization optimization an UCK2•2'AzUMP dataset was collected from crystals grown for 3 days in 0.2 M sodium/potassium phosphate, 22% PEG 3350.

The UCK2•2'AzCMP dataset was collected at ALS 8.2.1 with diffraction to 2.4 Å and was processed as described above (Supplementary Note 2 – Table 5-1). Molecular replacement was carried out using a search for two molecules of PDB: 1UEI, and the solution was refined to a final R-work/R-free of 18.9/22.7, as described above. The UCK2•2'AzUMP dataset was collected at SSRL 9-2 with diffraction to 2.7 Å. Data processing and molecular replacement were performed as per the UCK2•2'AzCMP dataset. Refinement resulted in final R-work/R-free of 19.6/24.4.

RNA fluorescence imaging via CuAAC.

Coverslips in 6-well tissue culture plates were treated with 1x poly-D lysine (Corning) solution for 3 h at 37 °C. HEK293T cells were seeded at densities of 2.5×10^5 and grown to ~50% confluency on glass cover slips. Cells with or without ectopic expression of UCK2 were treated with nucleosides to final concentrations of 1 mM and incubated for 5 hours. After labeling, cells were washed twice with DPBS and fixed for 10 min at room temperature with 3.7% paraformaldehyde. Cells were quenched with 50 mM glycine, DPBS for 5 min, and then washed twice more. Cells were permeabilized in 0.1% Triton-X DPBS for 15 min, and then washed x2 with DPBS. Cells were then treated with solutions containing 500 μ M CuSO_4 , 2.5 mM THPTA (Sigma), 2.5 mM sodium ascorbate, and 15 μ M Alexa-Fluor 568 (Click Chemistry Tools) and incubated at room temperature for 1 h in the dark. Cells were washed x2 with 0.1% Triton-X DPSB for 2 min, and then x5 with DPBS for 5 min each on an orbital shaker. Cells were stained with a solution of 1:3000 Hoechst 333242 for 10 min (Thermo Fisher). Coverslips were briefly washed and then mounted using VectaShield mounting medium (Vector Labs). Slides were imaged via fluorescence confocal microscopy using a 63x oil immersion objective on a Leica 700 Carl Zeiss microscope.

Co-culture enrichment of biotinylated transcripts.

HEK293T cells were separately transfected with mCherry or UCK2-GFP plasmid for 24 hours, following by 1:1 re-plating together to generate a mixed culture. Cell were allowed to grow and express for another 16 h, followed by treatment with 1 mM 2'AzUd for 0 or 5 h. Following RNA extraction and SPAAC biotinylation, RNA was DNase treated and 1 μ g was

subjected to reverse transcription with oligo(dT) priming. Following cDNA synthesis, 1:10 of the reaction was set aside and diluted 1:100 to use as an input control. The remainder of the reaction was subjected to streptavidin bead enrichment using MyOne Streptavidin C1 beads (Thermo Fisher Scientific) with 10 μ L of beads per 1 μ g of input RNA. Prior to hybridization, beads were washed x4 in binding buffer (1 M NaCl, 100 mM Tris-HCl pH 7.0, 10 mM EDTA, 0.2% (v/v) Tween-20) before blocking in RNase-free glycogen (Life Technologies) for 1 h at room temperature. After blocking, beads were washed twice more and re-suspended in the original volume using binding buffer. cDNA reactions were added to beads and the reactions were incubated for 20 min at room temperature with rotation to ensure that the beads were mixing. Following hybridization, beads were sequestered on magnetic rack and the supernatant was removed. Beads bound by cDNA:RNA hybrids were transferred to new tubes and were washed x4 with high salt wash buffer (4 M NaCl, 100 mM Tris-HCl pH 7.0, 10 mM EDTA, 0.2% (v/v) Tween-20) followed by x2 washes with nuclease free water. Beads were then spun down and re-suspended in 50 μ L of cDNA elution buffer (10 U RNase H, 1x RNase H buffer (New England Biolabs), 4 μ g RNase A/T1 mix (Thermo Fisher Scientific), 12.5 mM D-biotin) and incubated at 37 °C for 30 min at 750 RPM. Following incubation 1 μ L of DMSO was added and samples were heated to 95 °C for 8 min. The eluted cDNA was purified using the Zymo DNA Clean and Concentrator-5 kit with cDNA eluted in 20 μ L of nuclease free water.

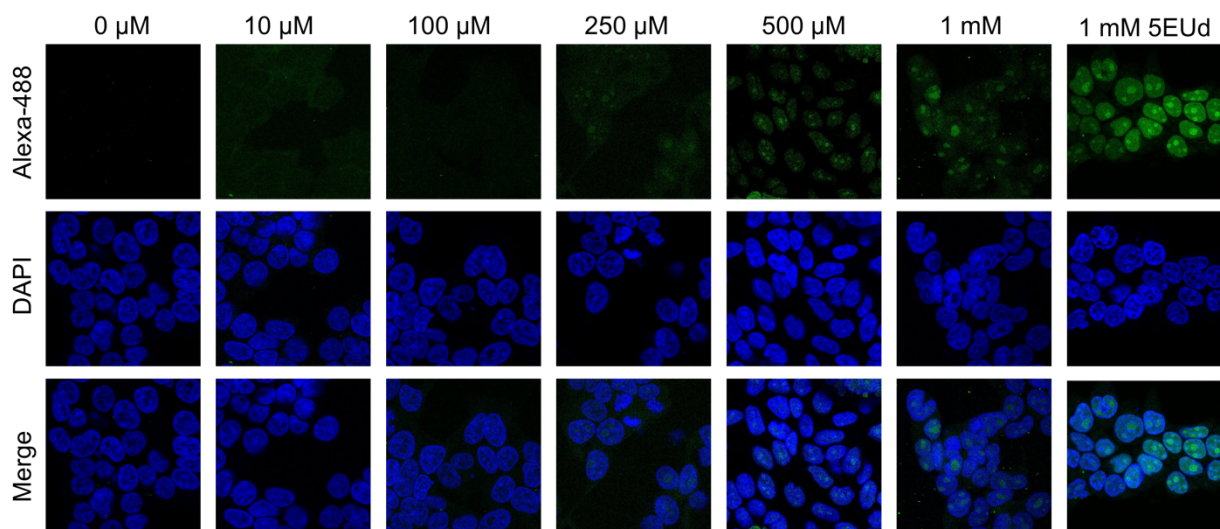
qPCR analysis of enriched turboGFP from co-culture.

Quantification of input and enriched samples was assessed by qPCR analysis as previously described. Briefly, 1 μ L of input or enriched cDNA from 0 h and 5 h 2'AzUd treatments were

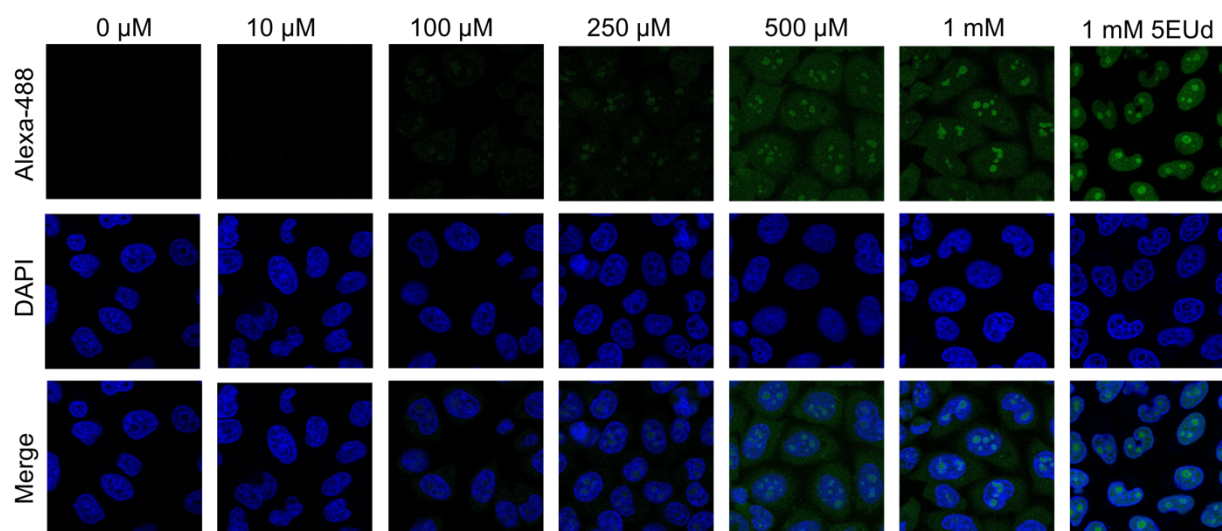
profiled for mCherry and turboGFP expression, after which the fold enrichment of each gene was calculated against the 0 h treatment, negative control: $2^{[(Cq-INPUT_{5\text{ hr}} - Cq-ENRICH_{5\text{ hr}}) - (Cq-INPUT_{0\text{ hr}} - Cq-ENRICH_{0\text{ hr}})]}$. Enrichment analysis was performed in biological duplicates and technical triplicates.⁴⁰

5.6 Supplementary Data

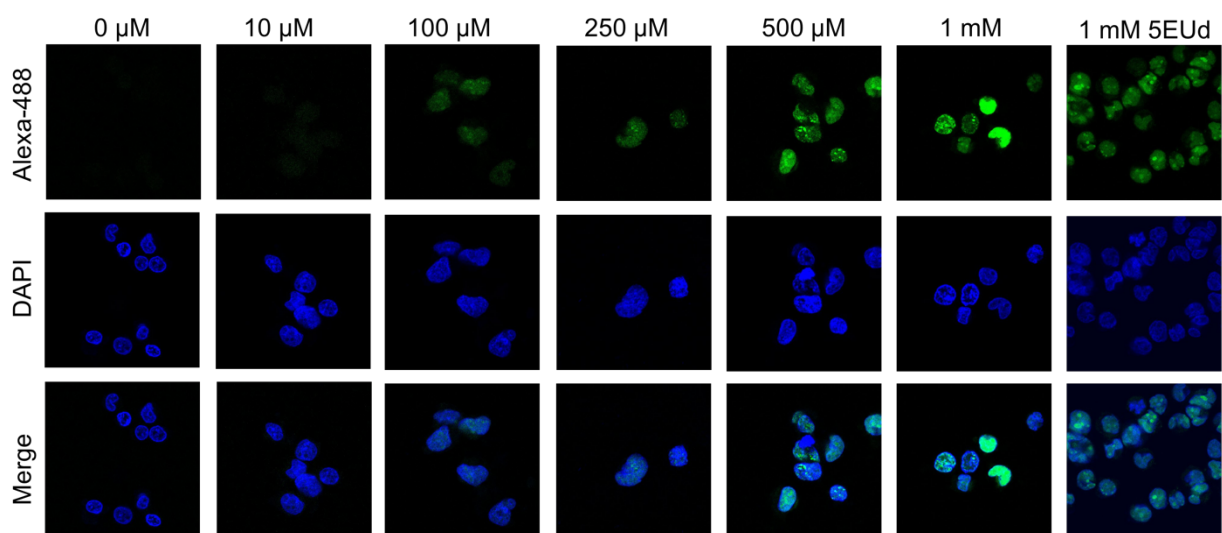
HEK293T



HeLa



K562



NIH3T3

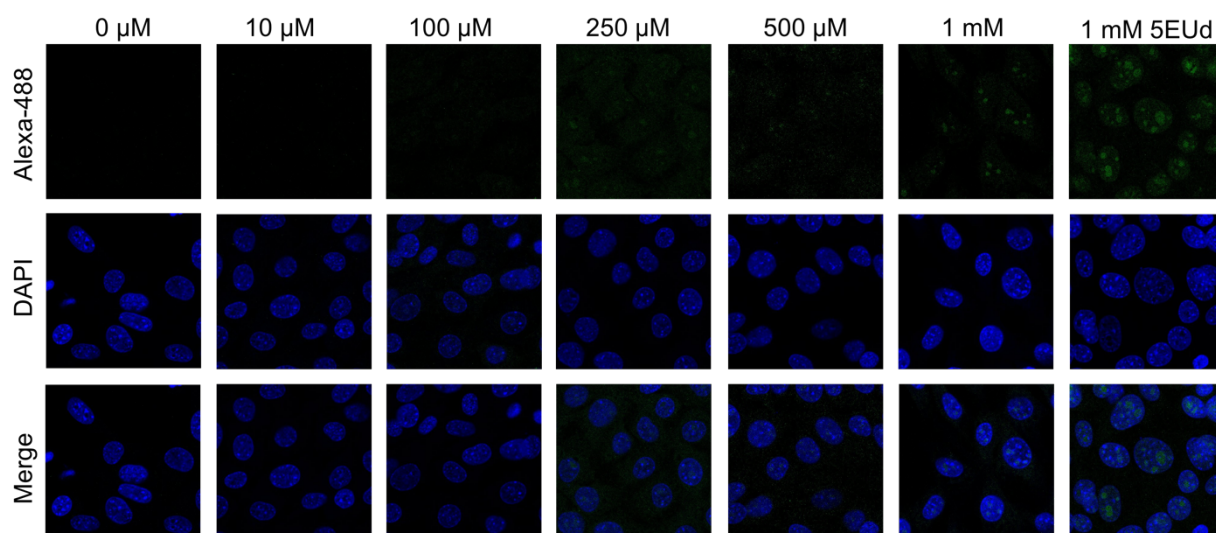


Figure 5-S1. CuAAC imaging of 5EU background incorporation in human and mouse cultured cells. Cells were incubated with titrated 5EU for 5 h, followed by fixation permeabilization and click with Alexa-488 to visualize background labeling of nascent transcription. Imaging analysis also revealed that 5EU incorporation mirrors 5EUd, with strongest signal originating in the nucleolus, the location of ribosomal RNA synthesis.

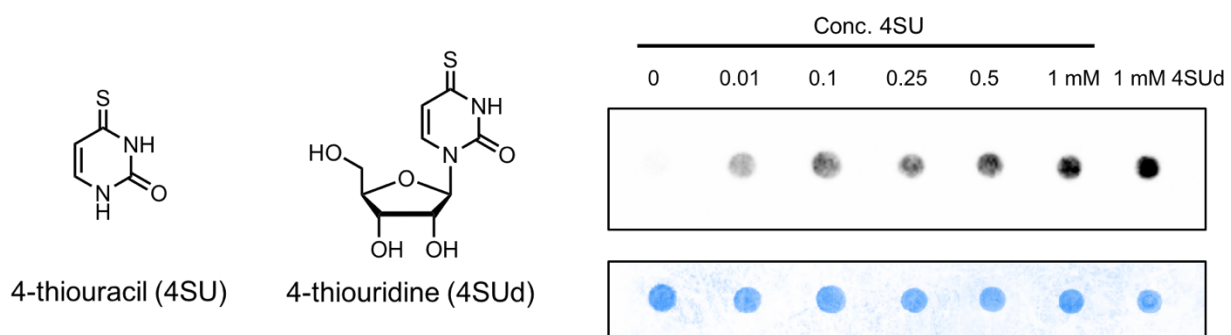


Figure 5-S2. The analog 4SU is incorporated into wildtype cells. HEK293T cells were incubated with titrated 4-thiouracil (4SU) and 4-thiouridine (4SUd) as a positive control for 5 h. Following incubation, RNA was extracted and subjected to MTSEA biotinylation,

followed by streptavidin-HRP dot blot detection. 4SU is able to be incorporated through endogenous cellular machinery by wildtype cells.

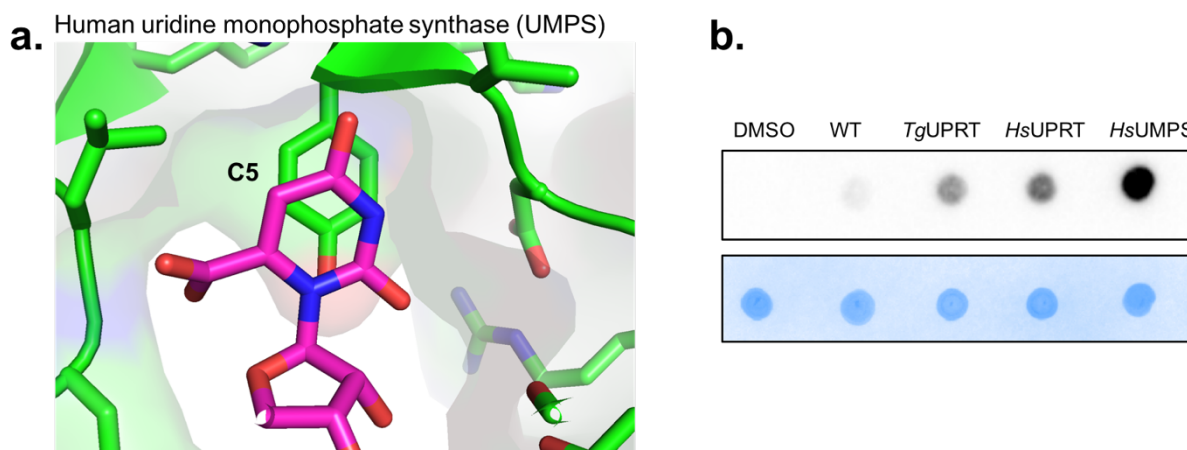


Figure 5-S3. Structure and salvage activity of human UMPS. **a.** Structure of human uridine monophosphate synthase PDB: 2WNS, demonstrates a spacious pocket around C5 of the substrate orotidine monophosphate. This region may allow for C5 functionalized pyrimidine substrates to bind. **b.** Comparison of 5EU labeling signals produced upon expression of *Toxoplasma gondii* UPRT or *Homo sapiens* UPRT homologs suggest similar activity levels. Cells were incubated with 0.25 mM 5EU, 5 h followed by RNA isolation, biotinylation and dot blot. Similar labeling signals may indicate that the *Toxoplasma gondii* and *Homo sapiens* UPRT enzymes may have similar activities.

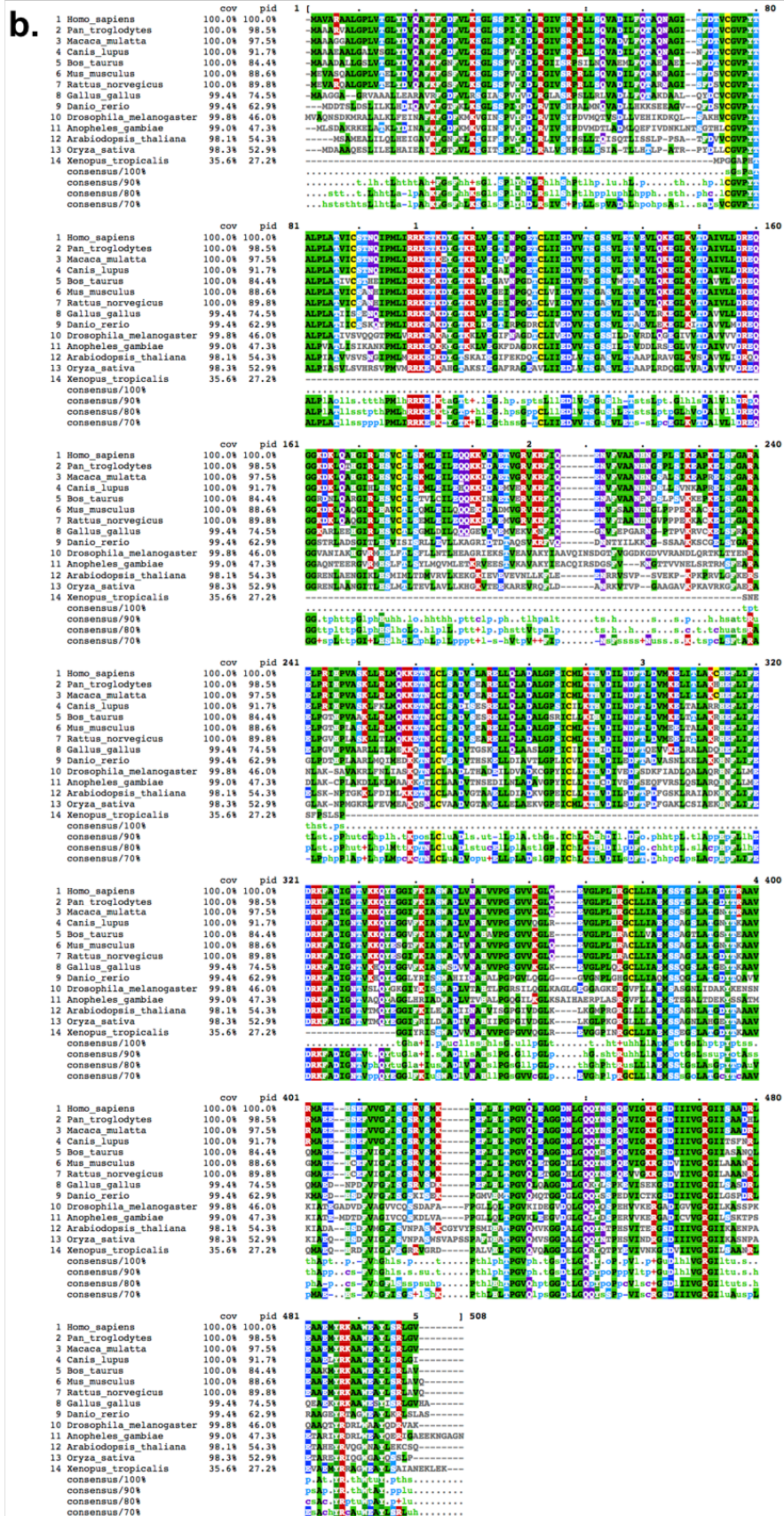


Figure 5-S4. Primary sequence alignments of UPRT and UMPS homologs. Alignments were made for these two enzymes amongst various representative organisms suggests there may be similar levels of activity.

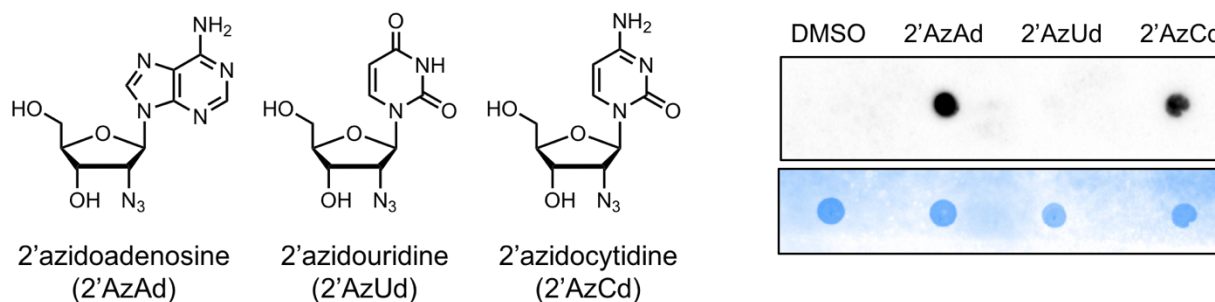
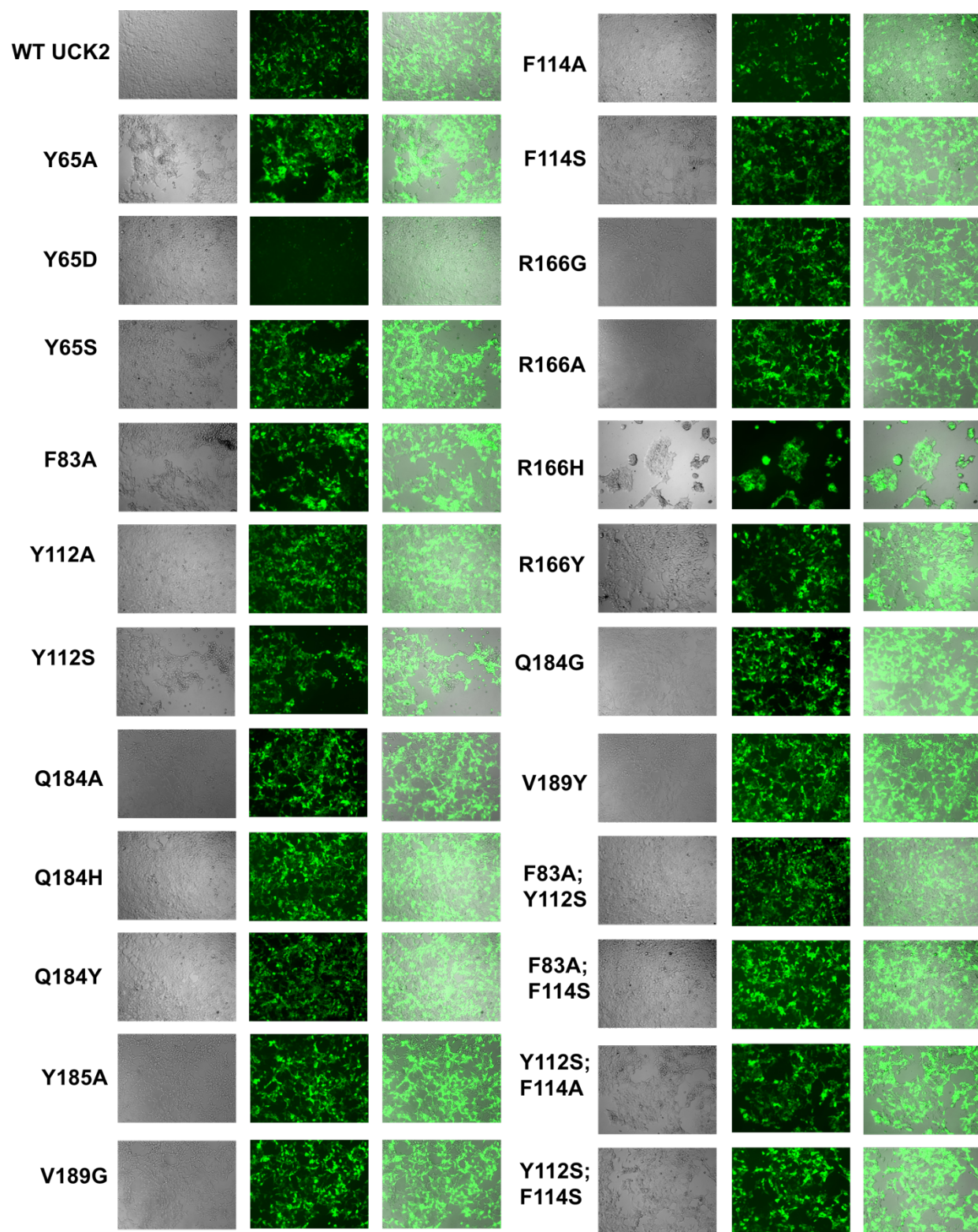


Figure 5-S5. Screening incorporation of 2'azido-nucleosides into nascent RNA. HEK293T cells were treated with 2'azidoadenosine, -uridine or -cytidine followed by RNA extraction, biotinylation and streptavidin HRP dot blot. Cells were able to incorporate 2'AzAd and 2'AzCd, while 2'AzUd was not detected.



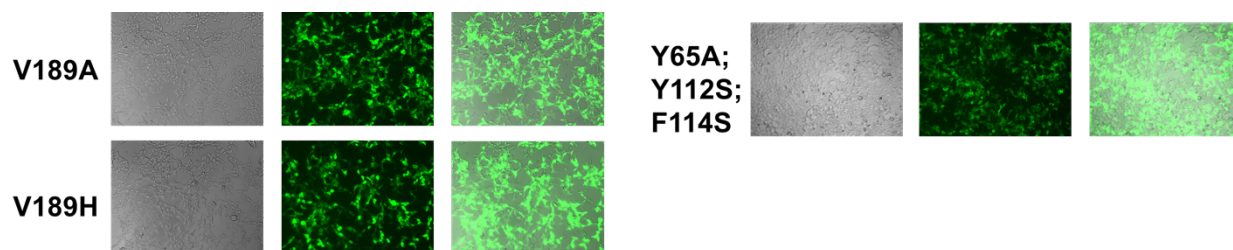


Figure 5-S6. Expression of UCK2 mutants. Site directed mutagenesis was performed on a vector encoding UCK2 fused to TurboGFP. Following cloning and sequence confirmation, we had two sets of 13 mutants for screening against 5mAzUd and 2'AzUd. Mutants were transfected into HEK293T cells and expression was monitored via the GFP reporter. Images were taken at different confluencies. We note that 24 of the 26 mutants were robustly expressed in cells, with the exception of R166H which significantly affected cell proliferation and Y65D, which appeared to cause severe inclusion body formation.

WT UCK2 + 5mAzUd

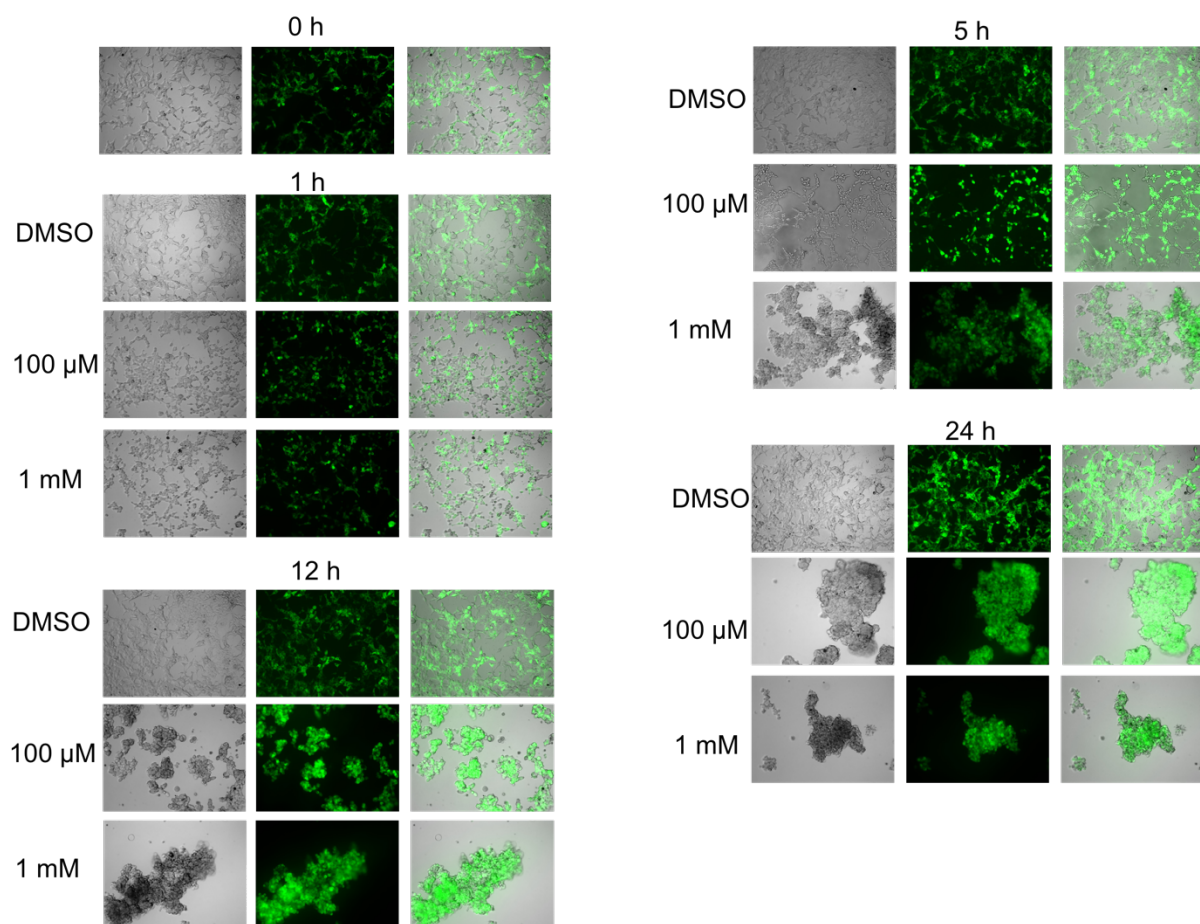


Figure 5-S7. HEK293T cells transfected with WT UCK2 followed by treatment with 5mAzUd.

Y112A + 5mAzUd

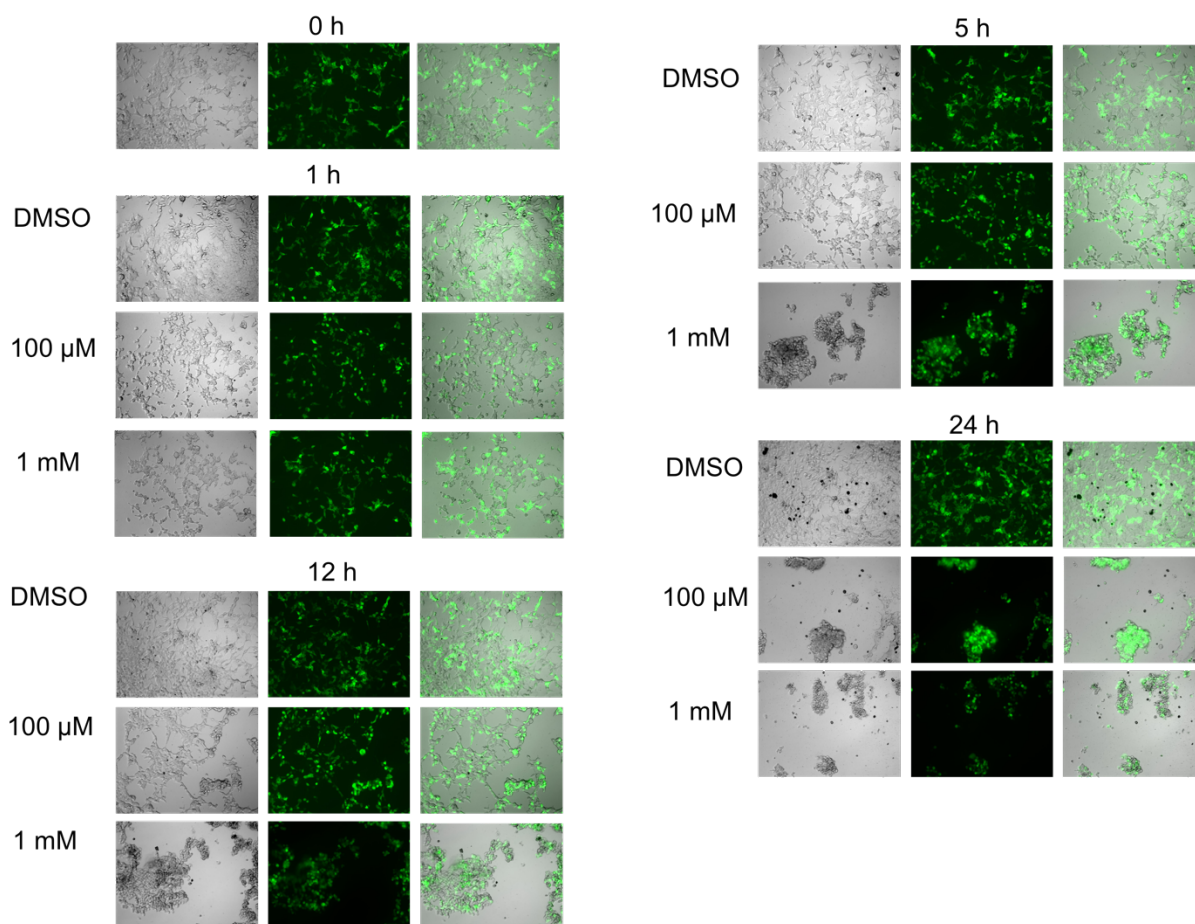


Figure 5-S8. HEK293T cells transfected with Y112A UCK2 followed by treatment with 5mAzUd.

F114S + 5mAzUd

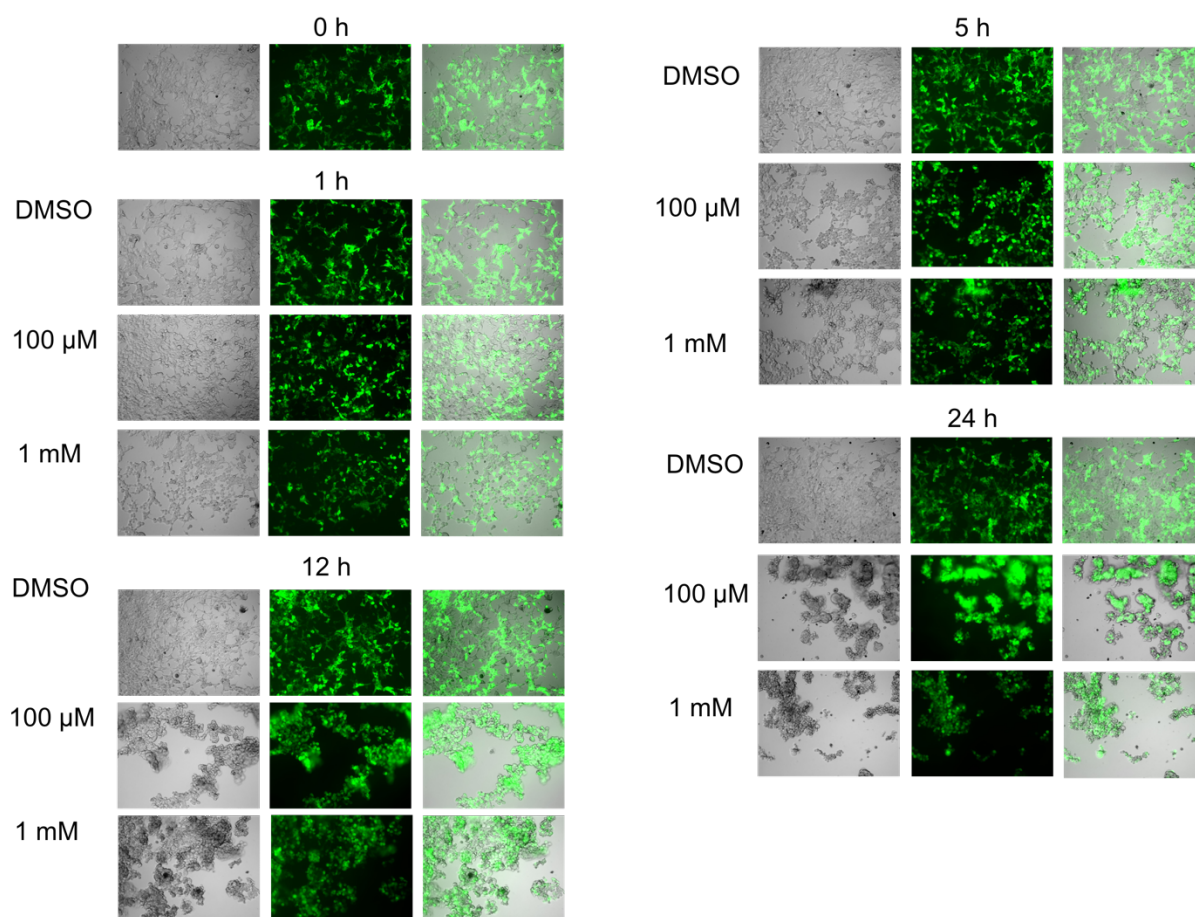


Figure 5-S9. HEK293T cells transfected with F114S UCK2 followed by treatment with 5mAzUd.

F83A; F114S + 5mAzUd

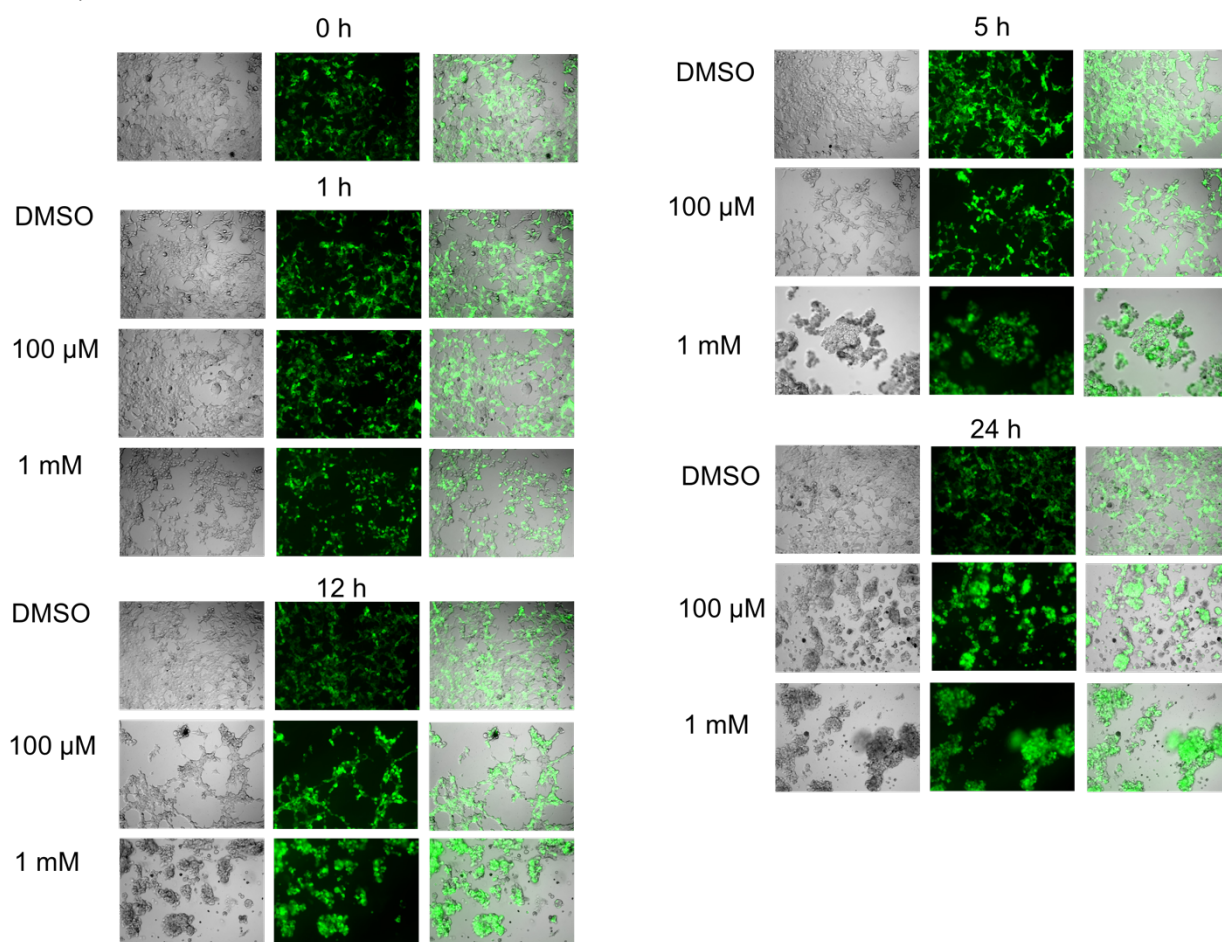


Figure 5-S10. HEK293T cells transfected with F83A; F114S UCK2 followed by treatment with 5mAzUd.

Y112S; F114S + 5mAzUd

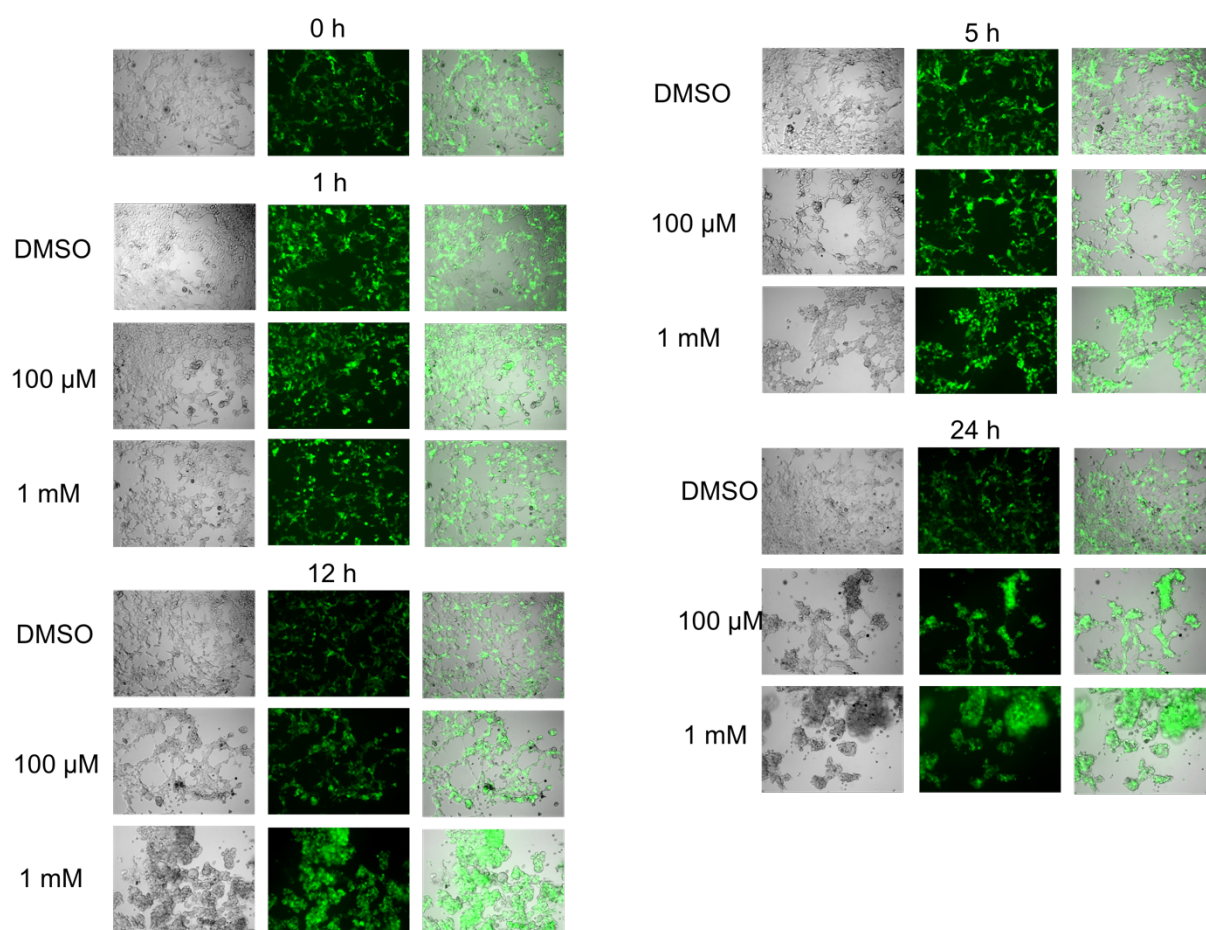


Figure 5-S11. HEK293T cells transfected with Y112S; F114S UCK2 followed by treatment with 5mAzUd.

WT -UCK2 + 5mAzUd

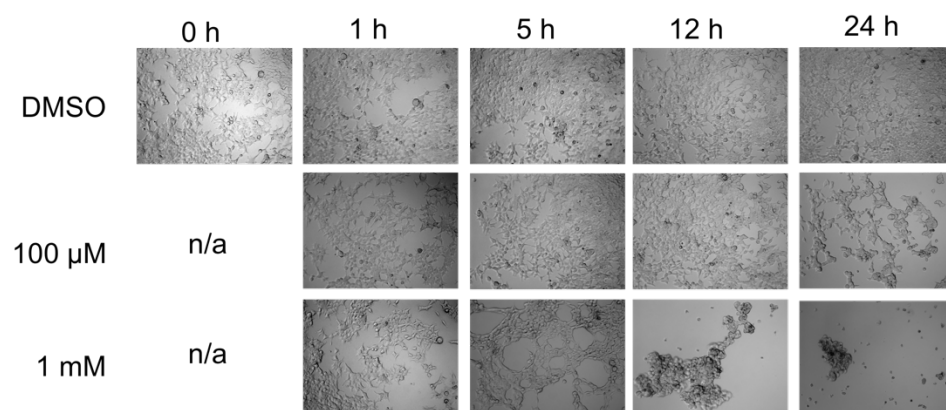


Figure 5-S12. WT HEK293T cells treated with 5mAzUd followed by treatment with 5mAzUd.

WT UCK2 + 2'AzUd

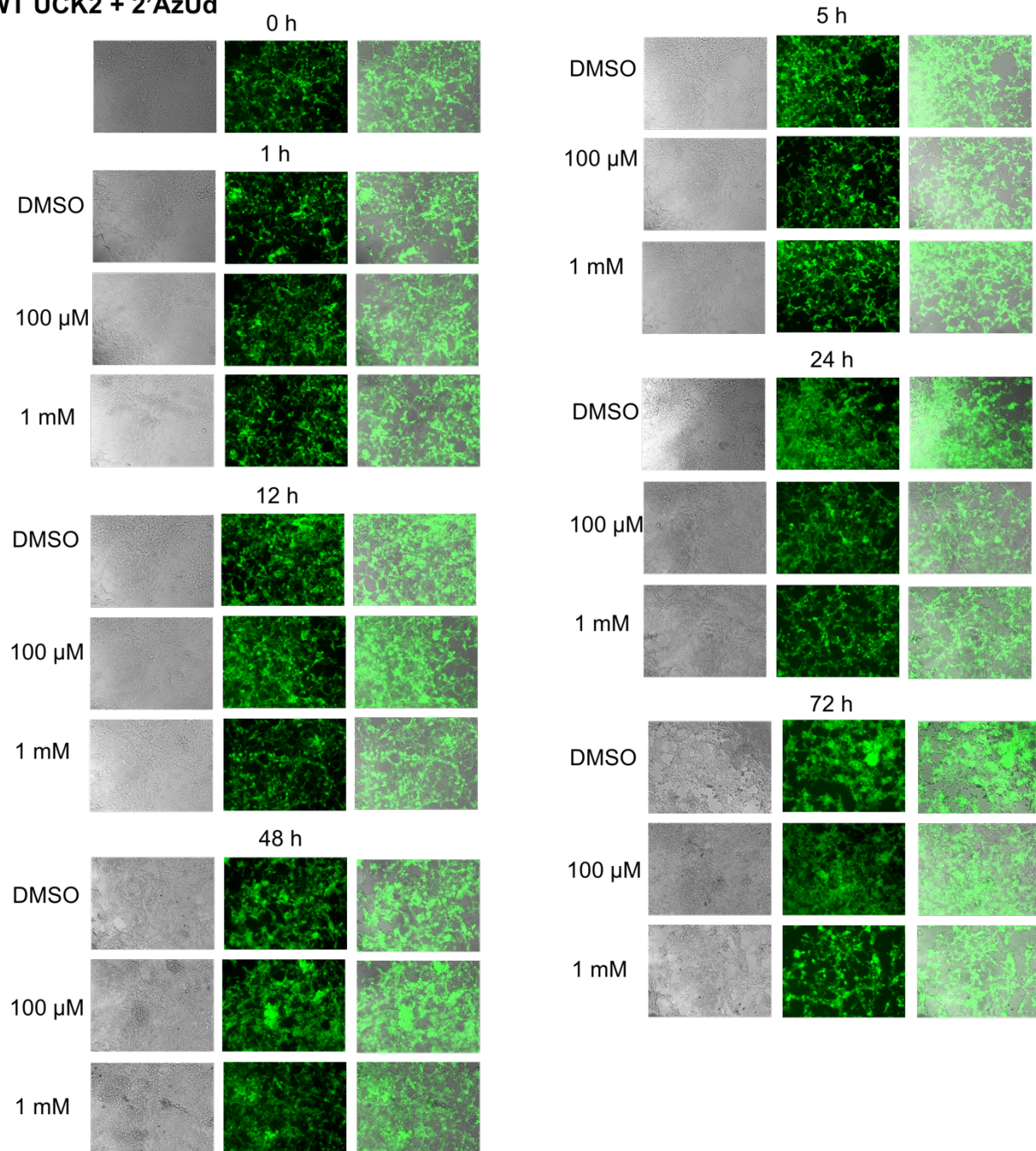


Figure 5-S13. HEK293T cells transfected with WT UCK2 followed by treatment with 2'AzUd.

R166G + 2'AzUd

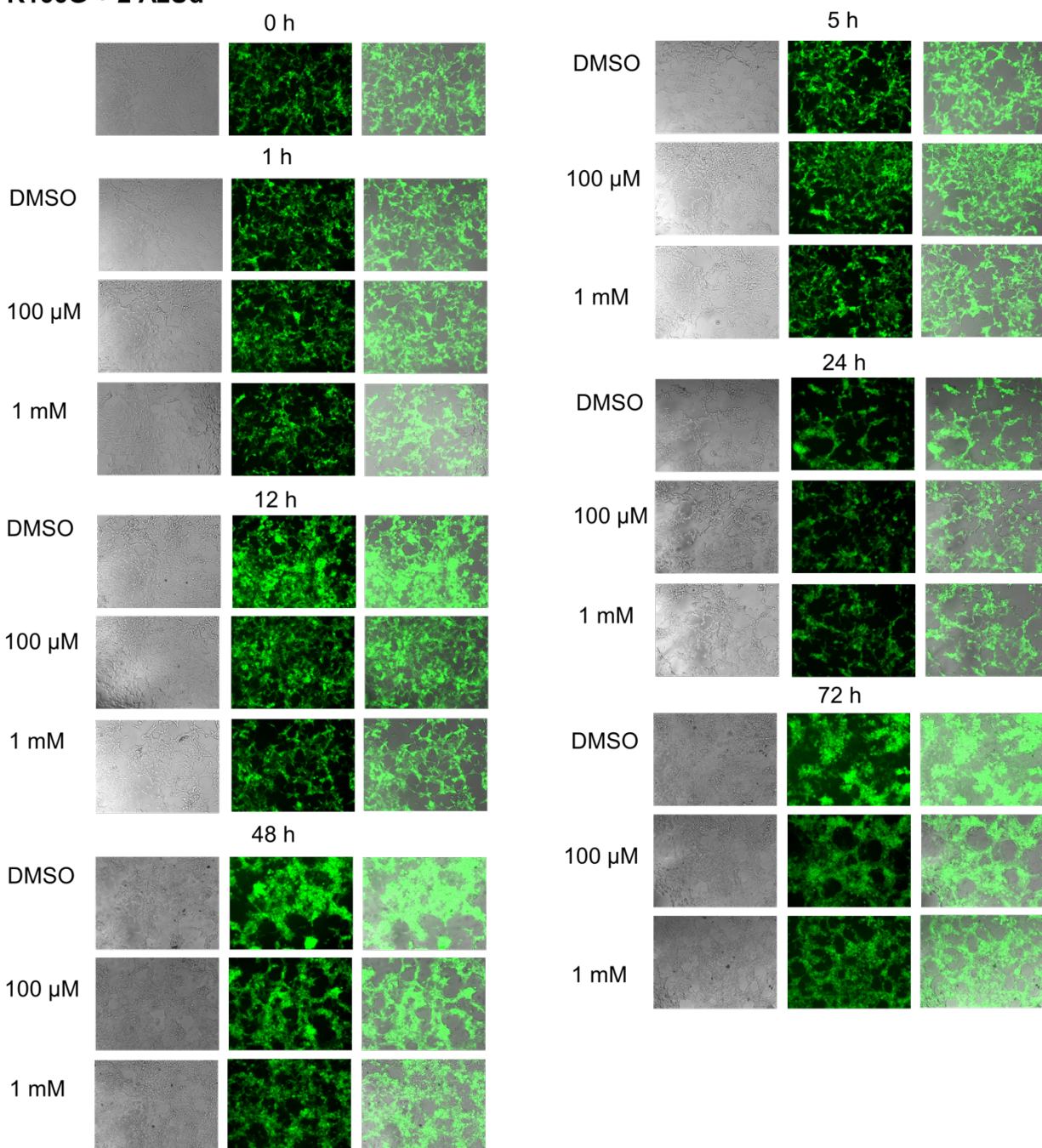


Figure 5-S14. HEK293T cells transfected with R166G UCK2 followed by treatment with 2'AzUd.

R166A + 2'AzUd

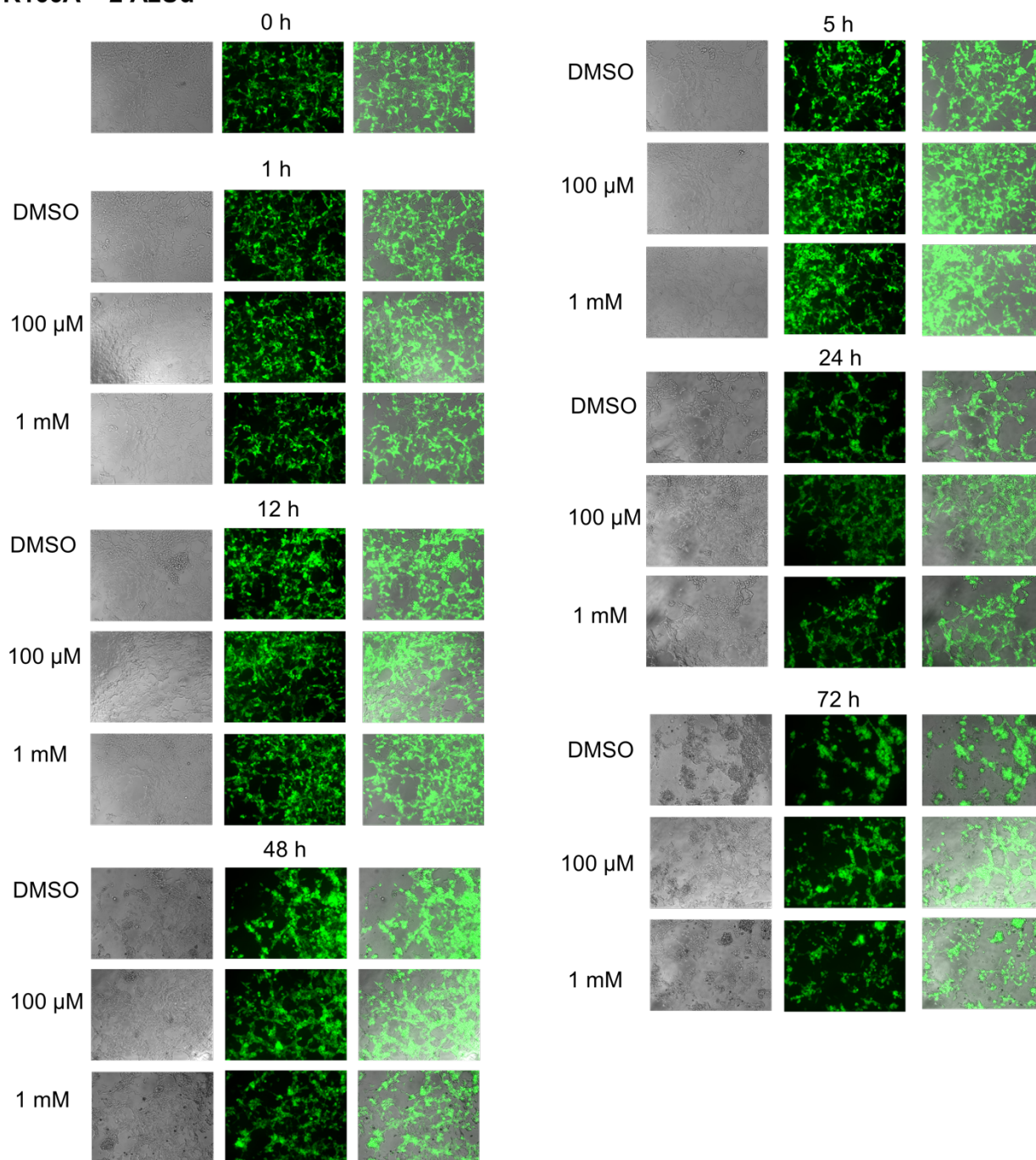


Figure 5-S15. HEK293T cells transfected with R166A UCK2 followed by treatment with 2'AzUd.

WT -UCK2 + 2'AzUd

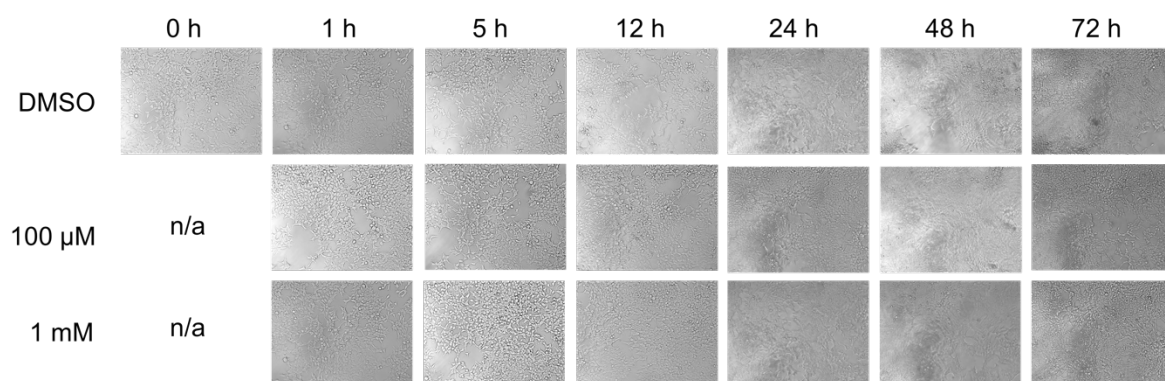
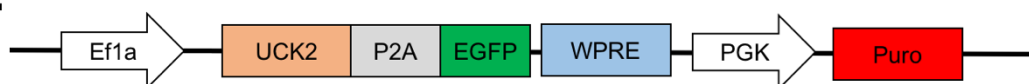
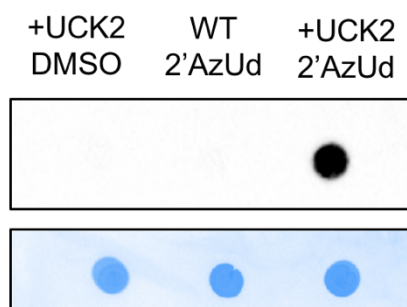


Figure 5-S16. WT HEK293T cells following treatment with 2'AzUd.

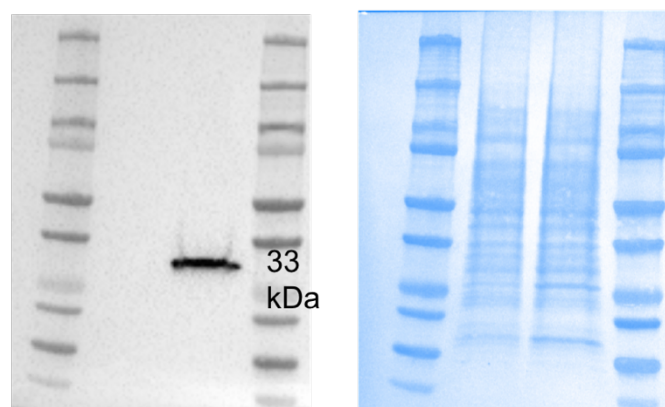
a.



b.



c.



d.

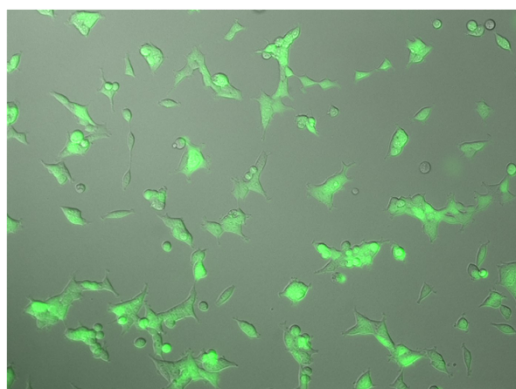


Figure 5-S17. Generation of UCK2 stable expression cell line. HEK293T cells were transduced with lentivirus to create a stably expressing UCK2 cell line. **a.** Design of lentivirus vector encoding UCK2 upstream of a read-through linker attached to an EGFP reporter. **b.** WT or UCK2-expressing cells were treated with 1 mM 2'AzUd for 5 h, followed by streptavidin dot blot to demonstrate robust UCK2-dependent labeling. **c.** Western blot analysis of UCK2 expression for lentivirus encoded UCK2. Anti-UCK2 probing of lysate extracted from WT cells (left) or cells expressing UCK2 (right) demonstrates presence of UCK2 protein only in stable cell line. **d.** Imaging EGFP reporter to identify UCK2 positive cells.

Figure 5-S18. Cell proliferation of 5mAzUd and 2'AzUd-treated cells. Cell proliferation measured via the ability of metabolically active cells to reduce resazurin to resorufin. WT

or HEK293T cells stably expressing UCK2 were grown in standard media and treated with titrated **a. & b.** 5mAzUd or **c. & d.** 2'AzUd for 5 – 72 h. Percent cell proliferation was normalized to cells treated with the DMSO control only for each respective time point (0.5% final).

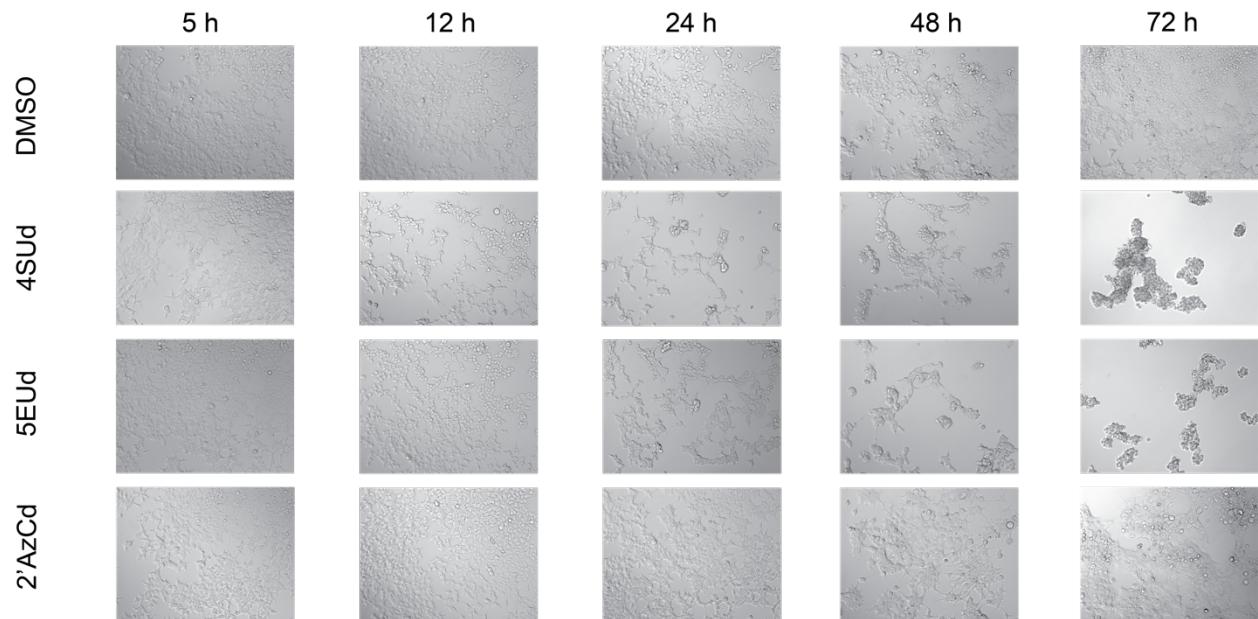


Figure 5-S19. Effects of other RNA metabolic reporters on cell growth and morphology. Imaging of HEK293T cells following incubation with 1 mM 4SUd, 5EUd, 2'AzCd or DMSO control for 5 – 72 h. Cells treated with 4SUd followed by 5EUd showed the greatest changes in cell morphology and detachment.

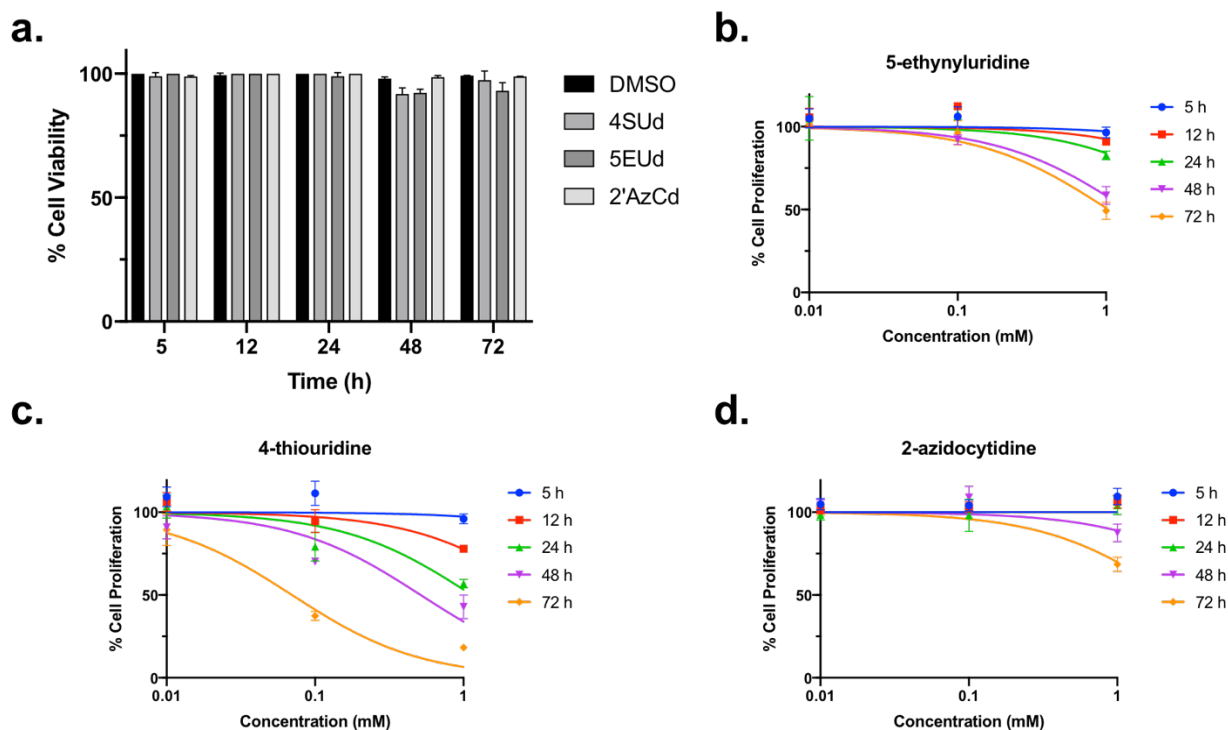


Figure 5-S20. Assessment of the effects of other RNA metabolic reporters on cell viability via trypan blue exclusion and Alamar Blue analysis. **a.** Cells treated with 1 mM 4SUd, 5EUd, 2'AzCd or DMSO for 5 – 72 h were assessed for their ability to exclude Trypan blue, indicating sustained membrane integrity. Alamar blue analysis measured the ability of metabolically active cells to reduce resazurin to resorufin. HEK293T cells were grown in standard media and treated with **b.** 5EUd, **c.** 4SUd, or **d.** 2'AzCd for 5 – 72 h. Percent cell proliferation was normalized to cells treated with the DMSO control only for each respective time point (0.5% final).

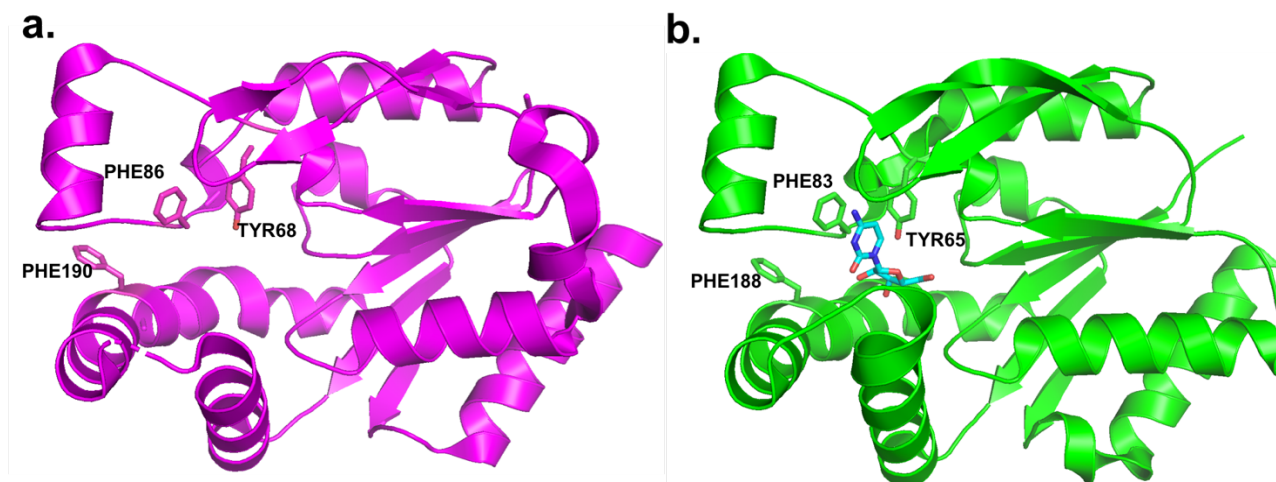


Figure 5-S21. Residues monitored in binding assays. Aromatic residues in **a.** UCK1 (PDB: 2JEO) and **b.** UCK2 (PDB: 1UEJ) used in fluorescence quenching experiments to determine binding affinity of canonical and 2'azido analogs.

Figure 5-S22. Kinetic curves of uridine, cytidine, 2'AzUd and 2'AzCd phosphorylation by purified UCK1 and UCK2. The data was fitted to the Michaelis-Menton equation using a non-linear fit. Each data point is the mean of three experiments.

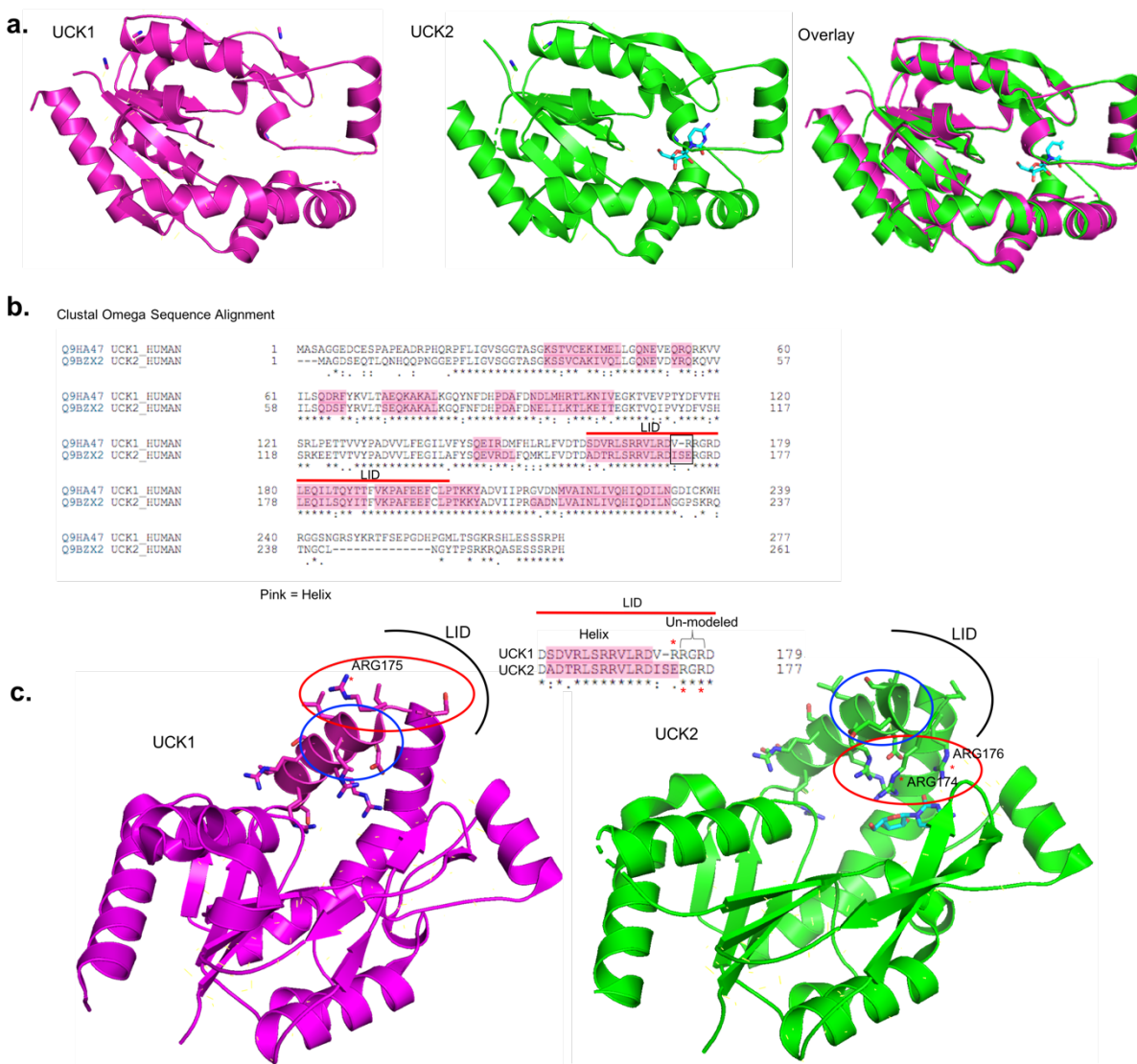


Figure 5-S23. Comparison of UCK1 and UCK2 sequence and structure. **a.** UCK1 PDB: 2JEO and UCK2 PDB: 1UEJ overlay similarly and **b.** have 67% sequence similarity. **c.** UCK1 has a shorter helix of the lid domain, and less structure (circled in blue) compared to UCK2. Arginine residues critical for induced fit substrate binding are differentially oriented. ARG175 of UCK1 faces the exterior of the enzyme, making it less likely to interact with the base. ARG174 and ARG176 face the interior, toward the binding pocket (circled in red).

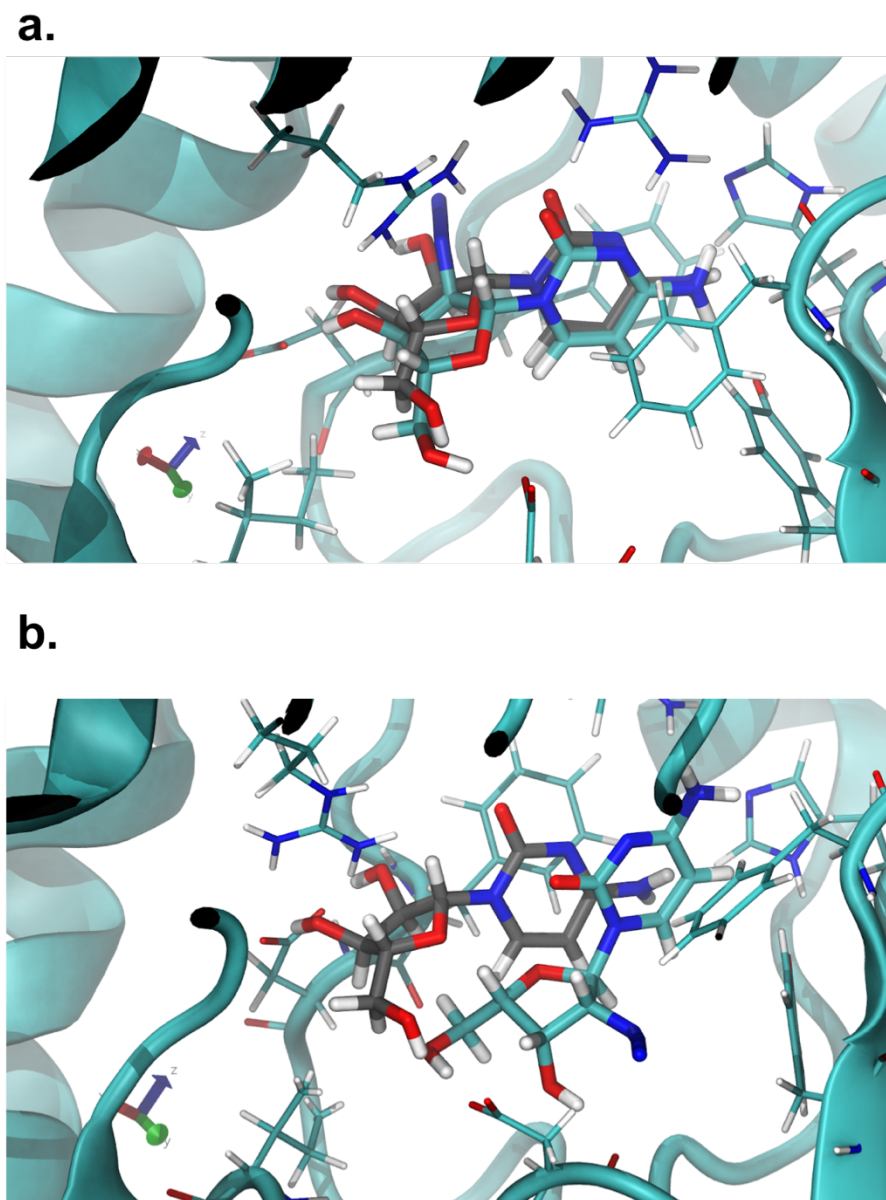


Figure 5-S24. Comparison of the canonical binding mode and the flipped binding mode simulations of 2'AzCd. The canonical binding mode is represented by the cytidine molecule (gray) from the crystallized UCK2-cytidine complex (PDB: 1UEJ), wherein **a.** one binding mode of simulated 2'AzCd (cyan) similarly overlays in the canonical binding mode (2'AzCd_{can}) and **b.** a second binding mode is flipped, wherein the sugars are oppositely oriented (2'AzCd_{flip}).

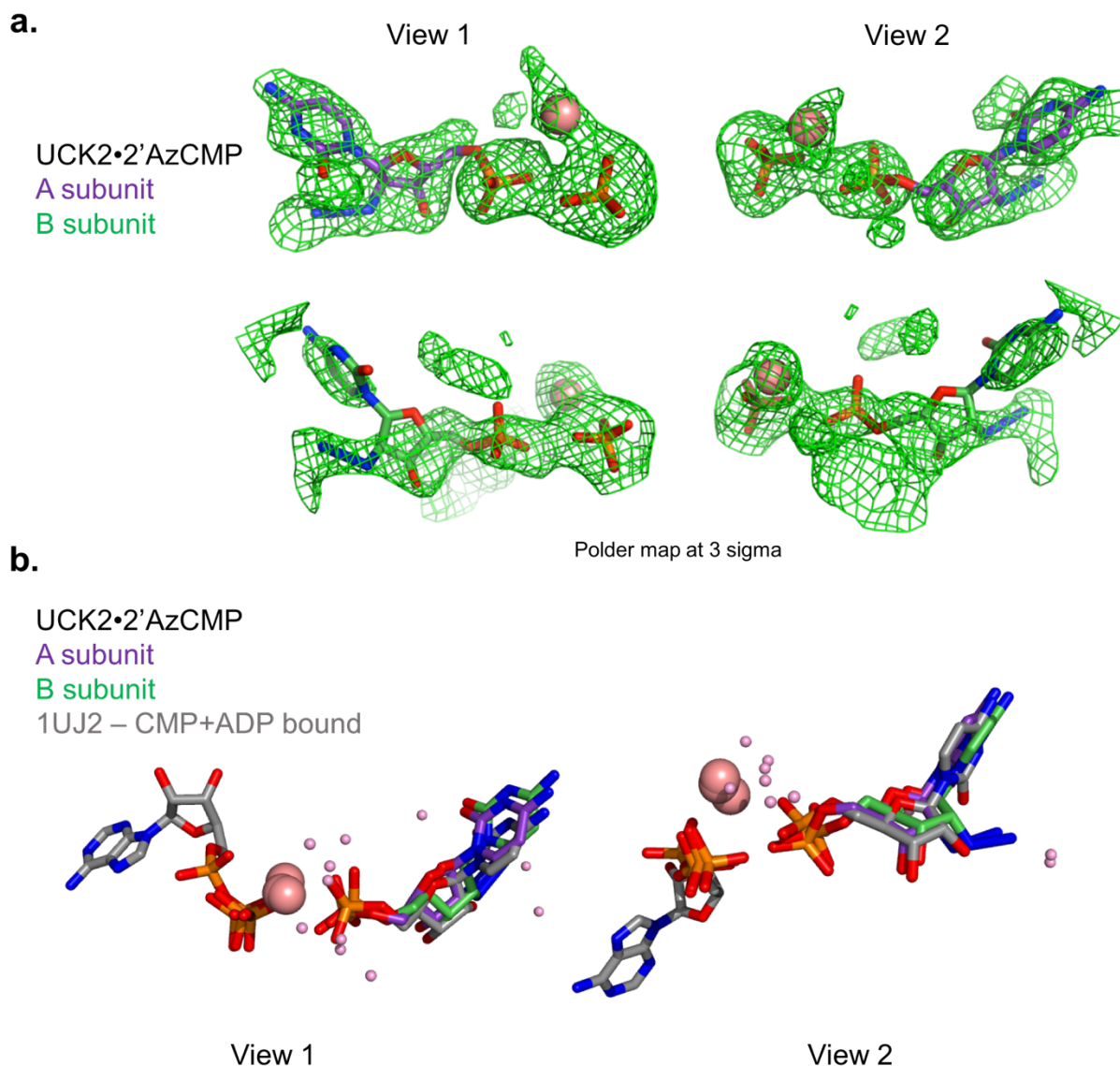


Figure 5-S25. Crystal structures of UCK2•2'AzCMP. UCK2 complexed with 2'AzCd, ATP and MgCl_2 crystallized with two protein subunits in the ASU, both of which contain the product of the phosphorylation reaction, 2'-azidocytidine monophosphate (2'AzCMP) and a Mg^{2+} ion. **a.** Density of A and B subunits showing the phosphorylation product 2'AzCMP. **b.** Superimposed structures of 2'AzCMP and CMP from PDB: 1UJ2 display agreement in binding configurations.

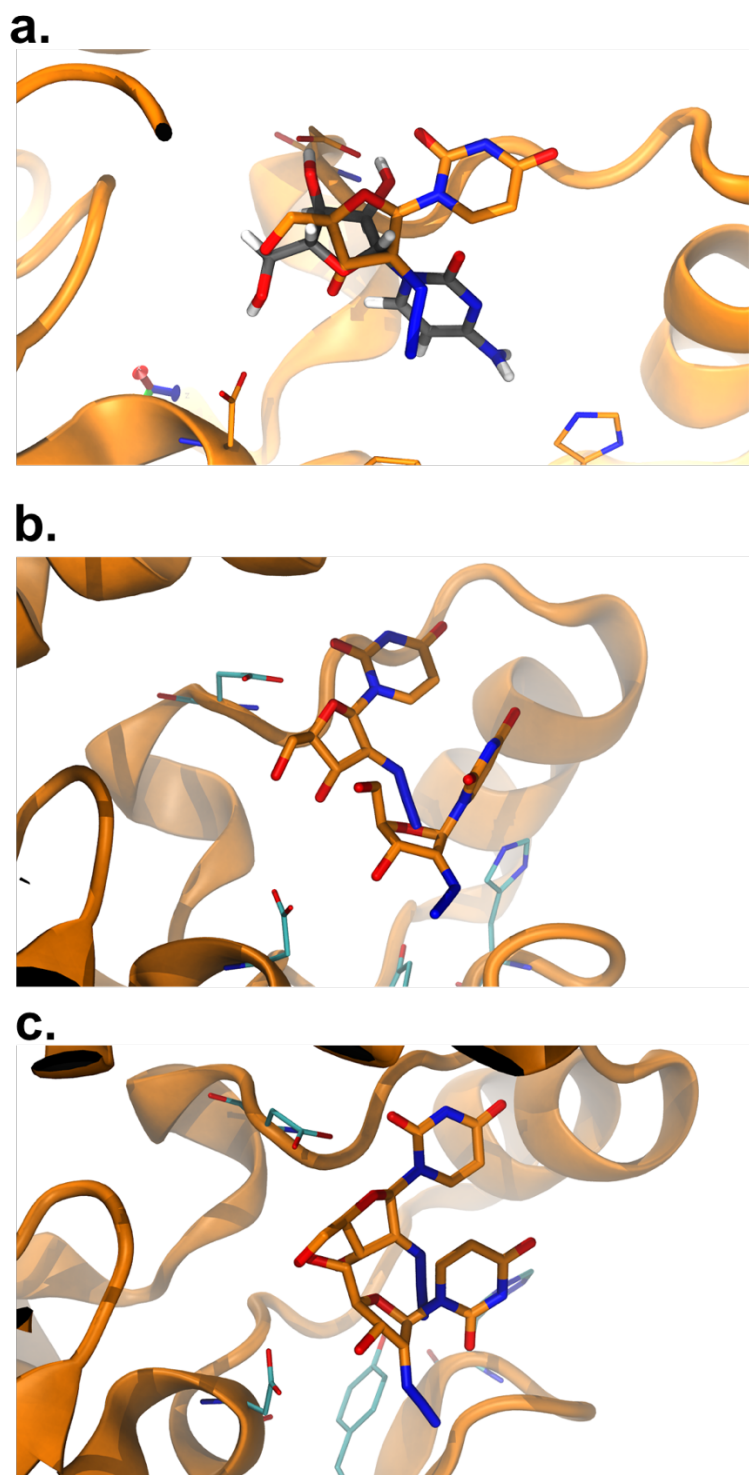


Figure 5-S26. Crystal structures of UCK2•2'AzUd represent the flipped binding mode.

a. Crystal structure of UCK2•2'AzUd (orange) taken from subunit D and the bound cytidine molecule (gray) from PDB: 1UEJ. **b.** UCK2•2'AzUd subunit D (above) and subunit E (below)

and c. 2'AzUd from subunit D (above) and subunit F (below) demonstrate a wide degree of flexibility in the base orientation.

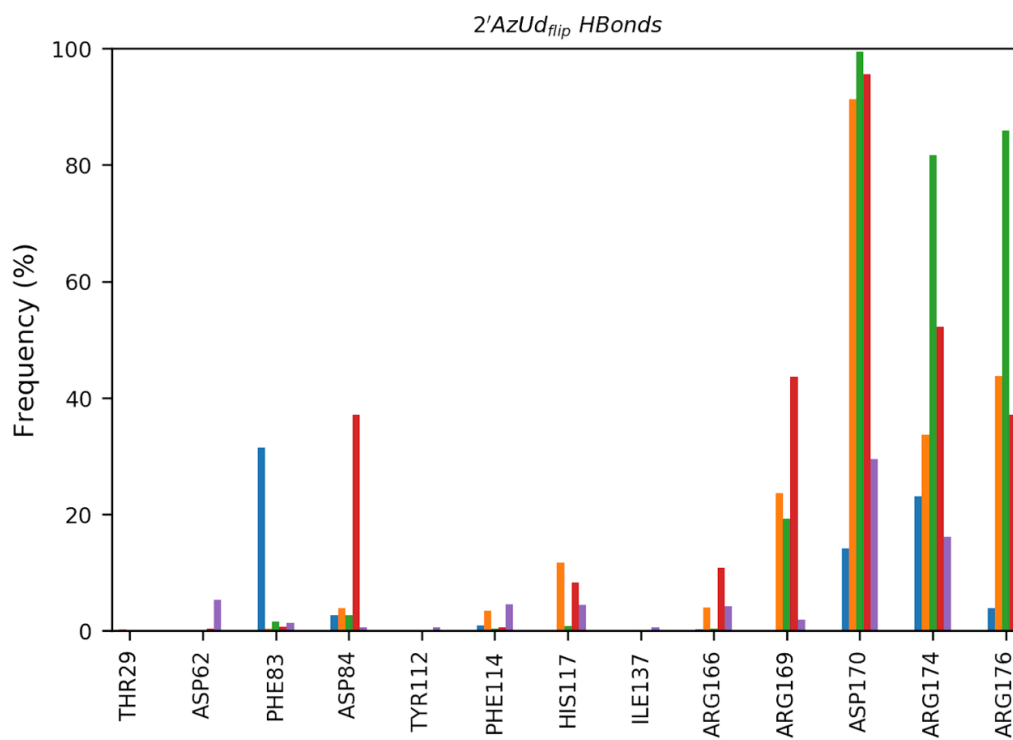


Figure 5-S27. Hydrogen bond contact frequency of 2'AzUd_{flip} Frequency of hydrogen bond contacts from the five 100 ns MD simulations of 2'AzUd in the flipped binding mode. Each colored bar represents the hydrogen bonding frequency from one MD simulation.

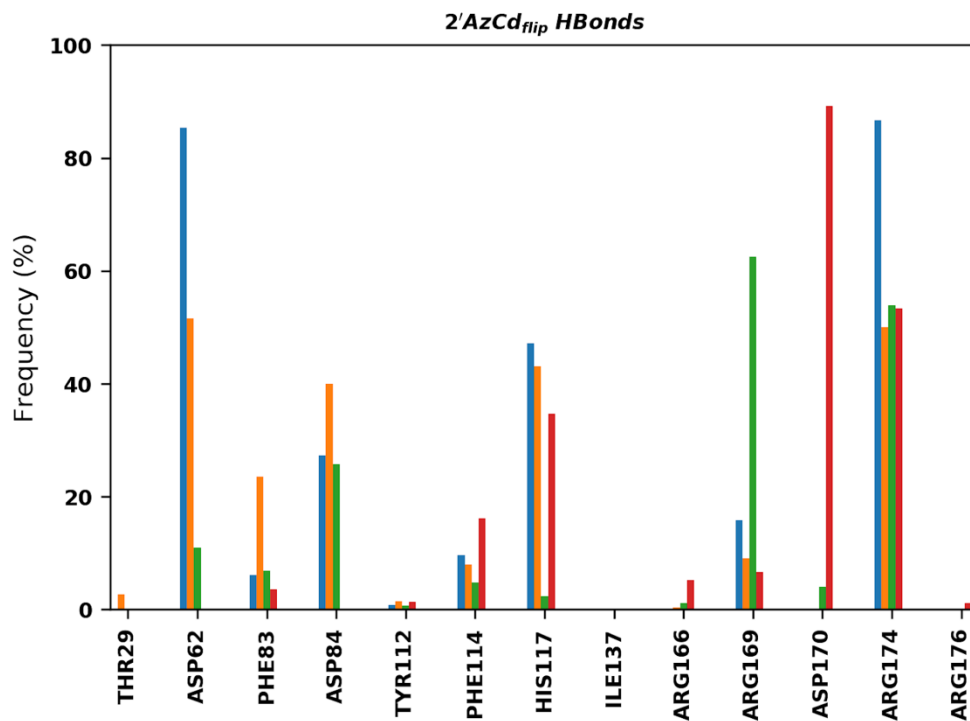
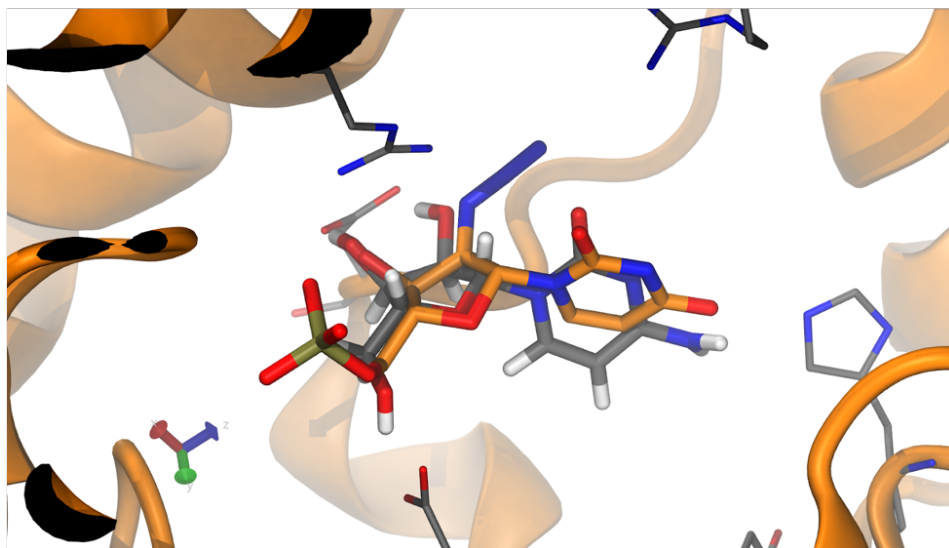


Figure 5-S28. Hydrogen bond contact frequency of 2'AzCd_{flip}. Frequency of hydrogen bond contacts from five 100 ns MD simulations of 2'AzCd in the flipped binding mode. Each colored bar represents the hydrogen bonding frequency from one MD simulation.

a.



b.

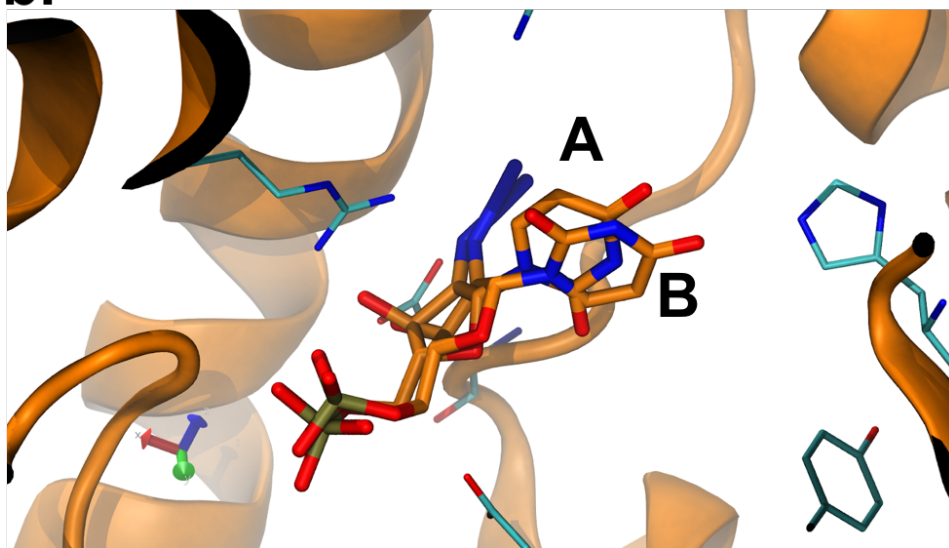


Figure 5-S29. Crystal structures of UCK2•2'AzUMP represent the canonical binding mode. a. Crystal structure of UCK2•2'AzUMP (orange) taken from subunit B and the crystallized cytidine molecule (gray) from PDB ID: 1UEJ. **b.** Overlay of crystal structures of UCK2•2'AzUMP taken from subunit A and subunit B (as labeled), representing the canonical binding mode. Subunit A is represented by the molecule in which the oxygen

atoms are oriented upwards. Subunit B is represented by the molecule in which the oxygen atoms are oriented downward.

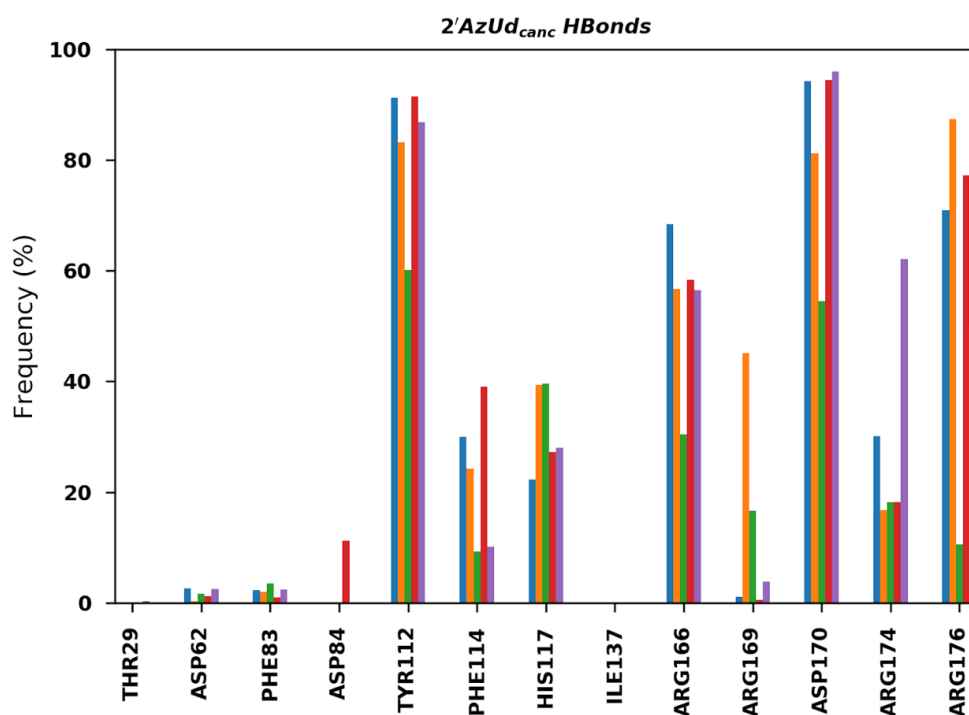


Figure 5-S30. Hydrogen bond contact frequency of 2'AzUd_{canc}. Frequency of hydrogen bond contacts from five 100 ns MD simulations of 2'AzUd in the canonical binding mode. Each colored bar represents the hydrogen bonding frequency from one MD simulation.

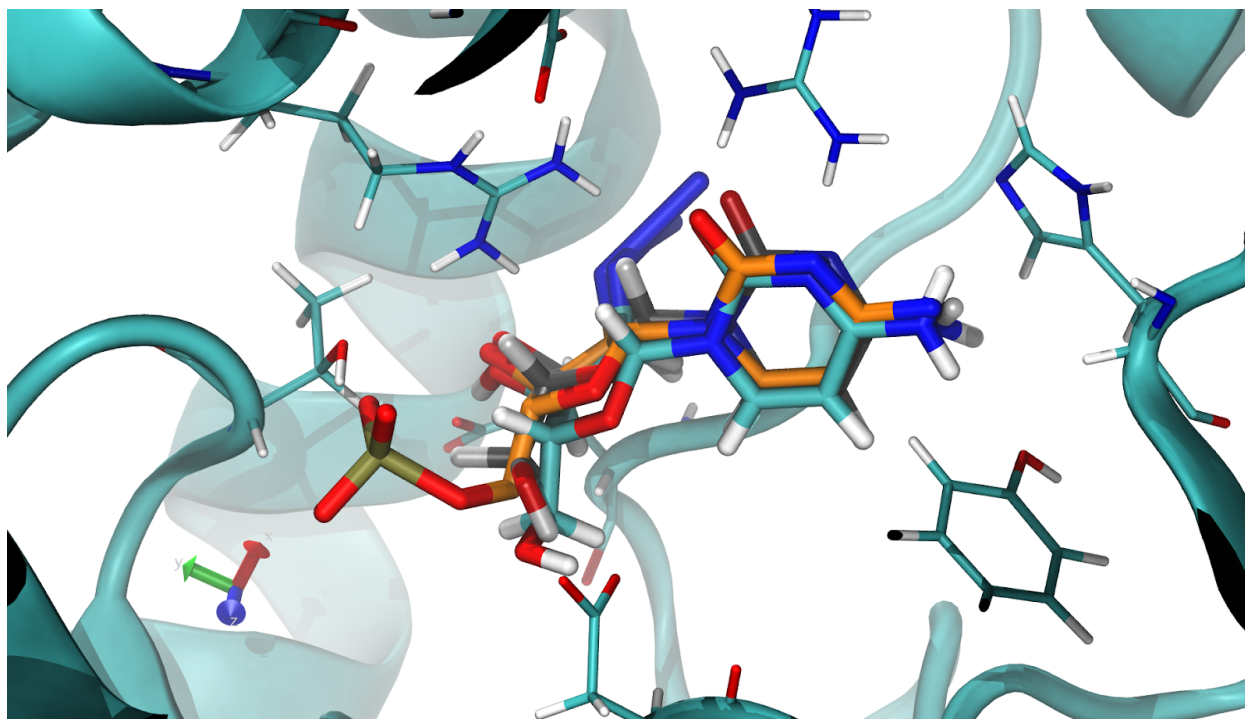


Figure 5-S31. Crystal structures of UCK2•2'AzCMP represent the canonical binding mode. Shown here are the structures of crystal structure from UCK2•2'AzCMP (subunit A- orange), MD model of 2'AzCd (cyan) in the canonical binding mode, and cytidine (PDB ID: 1UEJ - gray).

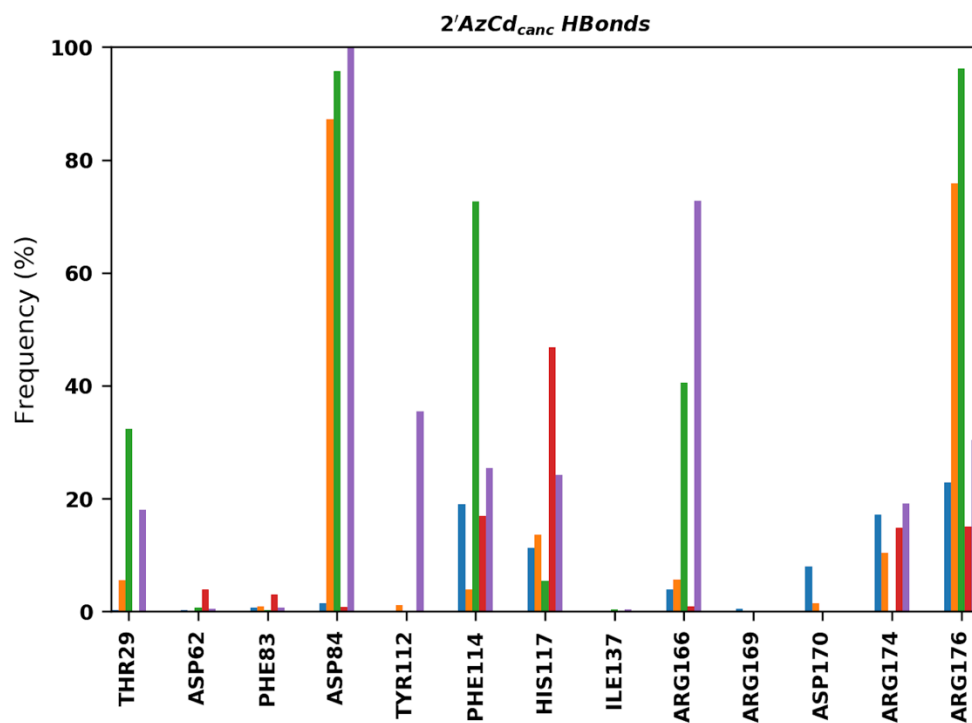


Figure 5-S32. Hydrogen bond contact frequency of 2'AzCd_{canc}. Frequency of hydrogen bond contacts from five 100 ns MD simulations of 2'AzCd in the canonical binding mode. Each colored bar represents the hydrogen bonding frequency from one MD simulation.

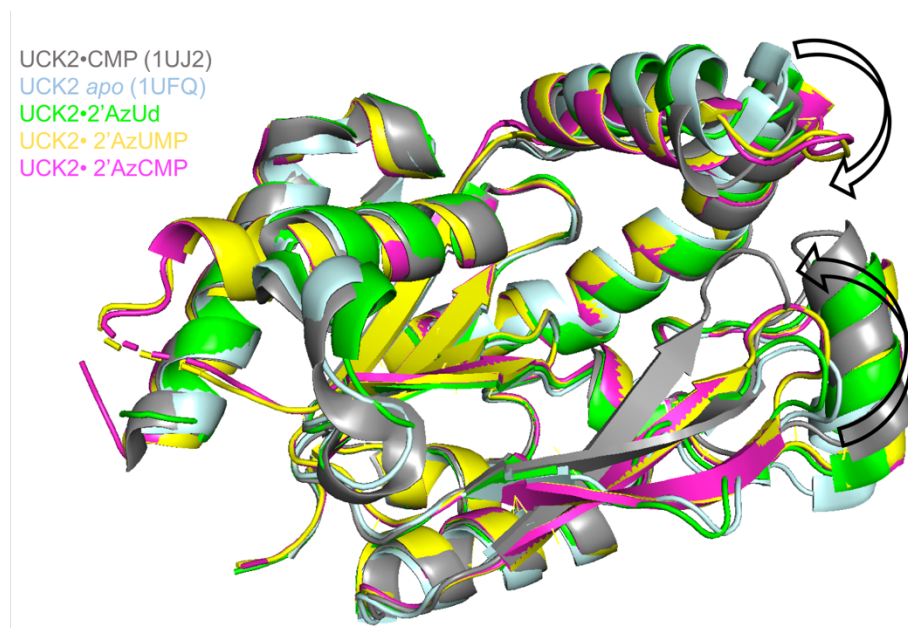


Figure 5-S33. Structural transitions of UCK2. Overlay of UCK2•CMP (PDB: 1UJ2) in gray, UCK2 *apo* (PDB: 1UFQ) in light blue, UCK2•2'AzUd in green, UCK2•2'AzUMP in yellow and UCK2•2'AzCMP in magenta illustrates the open and closed conformations found in the azido analog structures.

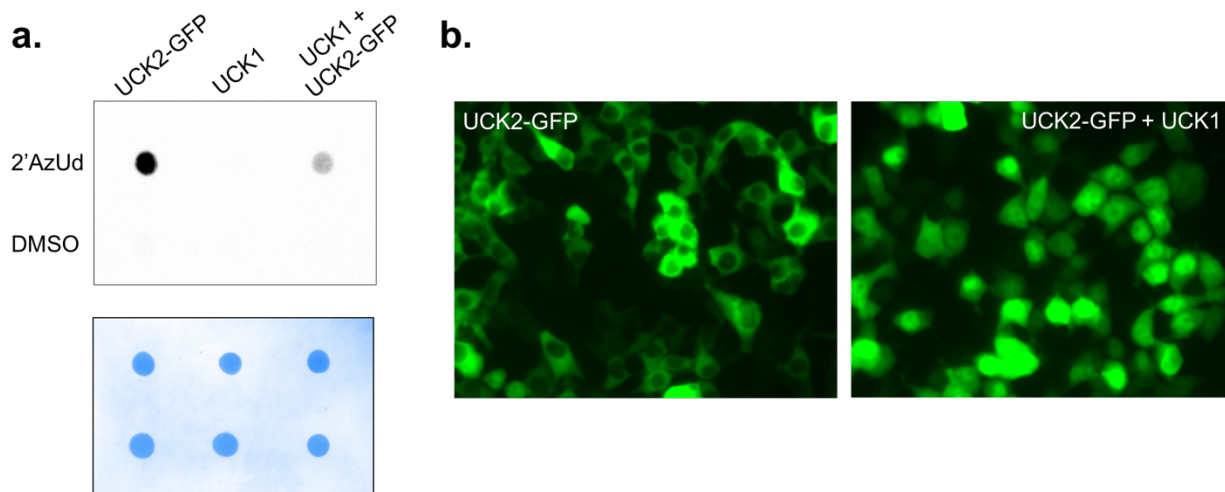


Figure 5-S34. Co-expression of UCK1 and UCK2 reduces the incorporation of 2'AzUd into RNA. **a.** HEK293T cells expressing either UCK2-GFP and/or UCK1 were treated with

2'AzUd, and RNA was isolated followed by streptavidin dot blot analysis. Co-expression of both UCK isoforms appears to reduce the labeling efficiency of UCK2. **b.** UCK2-GFP is normally localized to the cytosol, but in the presence of high levels of UCK1, localization is altered to the nucleus.

Table 5-S1. Distances between ligand and UCK2 residue atoms representing polar contacts and potential hydrogen bonds. Interaction distances for each chain are shown for UCK2•2'AzUd, UCK2•2'AzUMP and UCK2•2'AzCMP. Only one representative chain shown for canonical structures of UCK2•UTP (PDB: 1UEI) and UCK2•CMP (PDB: 1UJ2).

Residue:	Thr29	Ala30	Lys33	Asp62	Tyr65	Asp 84	Asp84	Tyr112	Phe114	His117	Glu135	Arg166	Arg166
Atom:	Oy1	N	NZ	Oδ1/Oδ2	OH	Oδ1	Oδ2	OH	O	Nδ1/Nε2	Oε1	NH1	NH2
UCK2•2'AzUD D				4.1 (O3')	2.1 (O3')	3.4 (O4')	3.5 (N1)					3.3 (O4')	
UCK2•2'AzUD E				4.2 (O3')	3.9 (O4')								
UCK2•2'AzUD F					2.8 (O4')	3.7 (O5')							
UCK2•2'AzUMP A	2.7(OP2), 3.4(O5')	3.2(OP2)	2.8(OP3)	2.8(OP1), 4.0(O2)	2.9(O4'), 3.0(O2)		2.7(N3'), 3.3(N4')			3.7(O4)	3.8(OP3)	2.4(O3')	3.0(N3'), 3.1(O3'), 3.3(N4'), 3.8(N5')
UCK2•2'AzUMP B	2.2(OP2), 3.7(OP1)	4.2(OP3)	3.5(OP3)	2.8(O5'), 4.1(O4')	3.4(O4')	2.8(N3')	2.3(O3')			2.8(O4)		3.0(N3'), 3.1(N4'), 3.6(N5')	3.0(OP2), 3.3(O3')
UCK2•2'AzCMP A	2.5(O07), 4.0(O09)	4.0(O08)	2.8(O08)	3.0(O05), 3.0(O05)	3.7(O10)	2.7(N13), 3.2(N14), 4.0(N15), 3.6(O01)	2.5(O01), 3.5(N13)	3.7(O10)	2.5(N23)	2.8(N23)		2.9(O01), 3.5(N13), 3.8(O07)	3.0(N13), 2.9(N14), 3.3(N15)
UCK2•2'AzCMP B	2.2(O09)	3.7(O09)	3.4(O07)	3.7(O08)	3.7(O10)	2.6(N13), 3.2(N14), 3.4(O01), 4.0(N15)	2.3(O01), 3.4(N13)		2.9(N23)	2.9(N23)		3.3(O01), 4.0(N13)	3.0(N13), 3.0(N14), 3.5(N15), 4.0(O01)
1UEI - UTP A	2.7(O2A)	3.1(O2A)	2.9(O1A)	3.4/3.7(O5')		2.5(O3')	2.6(O2')	2.8(O4)		2.7(O4), 3.6(N3)		2.7(O3'), 3.8(O4')	3.4(O2')
1UJ2-CMP B	2.5(O2P), 3.8(O3')	3.4(O2P)	2.8(O3P)	2.7(O5')	4.0(O5')	2.7(O2')	2.5(O3')	2.8(N4)		2.9(C5P), 3.8(N3)		2.8(O3'), 3.3(O2')	2.9(O2'), 3.2(O2)

Residue:	Arg169	Arg169	Asp170	Arg174	Arg174	Arg176	Arg176	Gln184	Gln184	Water 1	Water 2	Water 3	Water 4	Water 5
Atom:	NH1	NH2	Oδ1	NH1	NH2	NH1	NH2	Nε2	Oε1					
UCK2•2'AzUD D								4.1 (N1)	3.9 (N3), 3.3 (O4)					
UCK2•2'AzUD E				1.3 (O2)	3.1 (N3'), 2.9 (N4'), 3.1 (N5')	4.0 (O4)								
UCK2•2'AzUD F				3.8 (N1), 4.1 (N2)	3.5 (N3')					2.5 (O5')				
UCK2•2'AzUMP A	3.2 (OP2)									3.8 (O4'), 3.9 (N1)	3.0 (N5')			
UCK2•2'AzUMP B	3.1 (OP1)									3.1 (O4')	2.8 (OP3), 3.6 (O5')			
UCK2•2'AzCMP A	3.2 (O09)									3.5 (O08), 3.3 (O05)	2.6 (O08), 2.7 (O05)	O09 (2.3)	2.7 (N15)	3.3 (N15)
UCK2•2'AzCMP B	3.9 (O08)		4.1 (O22)							2.9 (O08), 3.7 (O.07)	3.1 (O07)			
1UEI - UTP A	2.9 (O2A), 3.1 (O3A)			3.1 (O4')		2.6 (O2)	2.7 (N3), 3.3 (O2), 3.7 (O4)		3.3 (O2)	2.6 (O2), 3.4 (O4'), 3.5 (N1)				
1UJ2-CMP B	3.2 (O2P), 3.5 (O1P)	3.3 (O2P)		3.9 (O4')	3.7 (O1P)	2.9 (O2), 3.7 (N3)	3.0 (N3), 3.5 (O2), 4.1 (N4)	3.3 (O2), 3.3 (O2')		2.8 (O2)	2.9 (O3P)	2.9 (O3P), 3.1 (O1P)	3.3 (O1P)	

5.7 Supplementary Note 1 – Molecular Dynamics Studies

1.0 Computational Methods

A total of 500 ns of MD simulation time was conducted for each binding mode by running 5 simulations each, where each copy of the simulation began from the same starting structure. The metastable binding modes sampled during our MD simulations were defined by constructing a Markov State Model (MSM) from our pool of MD simulation data and clustering with perron-cluster cluster analysis (PCCA). We then compared these metastable binding modes with X-ray crystal structures for 2'AzUd and 2'AzCd by computing the root-mean-square deviation (RMSD) between the ligand heavy atoms. We would like to note that the x-ray crystal structures of 2'AzUd and 2'AzCd were not used for setup or seen prior to conducting our MD simulations. Our simulation results, along with the x-ray crystal structures, support the hypothesis that the ligands must adopt the "flipped" binding mode for catalytic turnover.

1.1 Setup

1.1.1 Docking

After equilibration of the UCK2-cytidine complex, we used HYBRID docking⁴¹ from OpenEye Toolkits (v2019.4b2) to dock our nucleoside analogs (2'AzUd and 2'AzCd) into the binding site. HYBRID differs from the standard docking approach, such that the software will use the co-crystallized ligand as a reference point and attempt to fit the nucleoside analogs within the binding site by overlaying the docked ligands with the crystallographic ligand. From HYBRID docking, we generated up to 50 different poses for each nucleoside analog and then followed the same equilibration protocol described

previously. After equilibration, we found by visual inspection, that the poses began to resemble two separate groups: 1 pose which resembled the "canonical" binding mode and another which had the ribose moiety "flipped". We then selected 1 conformer from each representative group and carried these forward for our NPT 100 ns MD simulations (no restraints).

1.1.2 Protein preparation

To prepare the protein system (PDB: 1UEJ) for MD simulations, we used PDBFixer (pandegroup Github) to model in missing residues, add missing hydrogen atoms, protonate sidechains, and solvate our system. Sidechains were protonated in accordance with the pH = 8.0 environment of the enzyme assay experiments³⁸ and as described in other computational studies.⁴² Following protein preparation, with the cytidine molecule bound (using the pose from PDB: 1UEJ), we energy minimized the protein-ligand complex for a maximum of 30,000 steps and followed with an equilibration protocol as follows. The equilibration protocol occurs in four 10 picoseconds (ps) stages, whereby a progressively declining restraining force was used to help the protein-ligand system gradually relax. First, we apply a restraining force of $2.0 \text{ kcal/mol} \cdot \text{\AA}^2$ to the heavy atoms of the protein-ligand complex, simulate for 10 ps using constant volume (NVT), and follow-up with 10 ps at constant pressure (NPT). Next, we decrease the restraining force to $0.5 \text{ kcal/mol} \cdot \text{\AA}^2$ and then conduct an NPT simulation for 10 ps. Last, we use a $0.1 \text{ kcal/mol} \cdot \text{\AA}^2$ restraining force on the alpha-carbons (protein backbone) and the ligand heavy atoms and then simulate in the NPT ensemble simulate for 10 ps.

1.1.3 MD simulation parameters

All simulations in this study were conducted using OpenMM v7.1.1⁴³ at $T = 300$ K and $P = 1$ atm. Bonds with hydrogens were constrained. Here, we use time steps of 4 fs by employing the hydrogen mass repartitioning (HMR) scheme.⁴⁴ This scheme allows us to take larger time steps by slowing down the fastest motions (i.e. hydrogen bond stretching) in our MD simulations. The HMR scheme re-allocates mass from the connected heavy atom to the hydrogens. The protein-ligand systems were placed in a periodic box with explicit TIP3P water molecules using a 10 Å solvent padding distance and counter ions (Na⁺ Cl⁻) were added using a concentration of 150 mM. A 10 Å cut-off distance was used for the particle-mesh Ewald method for computing long-range (e.g. electrostatic) interactions. Protein atoms were parameterized using the ‘amber99sbildn’ forcefield⁴⁵ and the ligands were parameterized using GAFF2⁴⁶ in which atomic charges were assigned using the AM1-BCC charge model.⁴⁷ Both charges and GAFF2 parameters were assigned using openmoltools v0.8.1 and OpenEye Toolkits v2019.4b2.

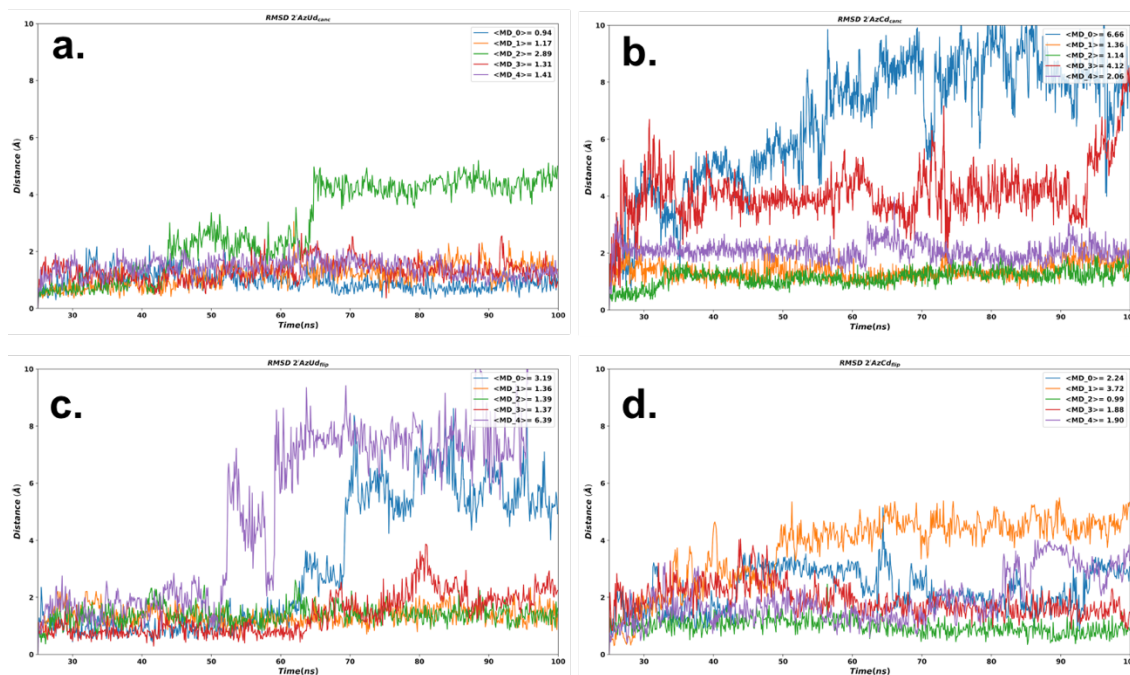
1.2 Analysis

For each nucleoside analog (2’AzCd and 2’AzUd), we ran five 100 nanoseconds (ns) MD simulations starting from each of the two binding modes (canonical and flipped). They are denoted as 2’AzCd_{can}, 2’AzCd_{flip}, 2’AzUd_{can}, and 2’AzUd_{flip}. Collectively, over the course of the 100ns of simulation time, the root-mean-square deviation (RMSD) for the ligand atoms began to stabilize after 25 ns. Thus, we discard trajectory frames from 0 – 25 ns as additional equilibration time and only perform further analysis from the 25 ns – 100 ns time frame Fig.2a. For construction of our Markov State models (MSM), we use the

PyEMMA v2.5.5⁴⁸ toolkit. To analyze hydrogen bonding contacts we use the MDTraj v1.9.1⁴⁹ and VMD v1.9.3⁵⁰ toolkits.

Supplementary Note 1 - Table S1. Calculated RMSDs ($\text{\AA}$) of the ligand heavy atoms from each MD simulation for the 25-100 ns time frame.

	RMSD	RMSD	RMSD	RMSD	RMSD	AVG
2'AzCd _{canc}	6.66	1.36	1.14	4.12	2.06	3.07
2'AzUd _{canc}	0.94	1.17	2.89	1.31	1.41	1.54
2'AzCd _{flip}	2.24	3.72	0.99	1.88	1.90	2.15
2'AzUd _{flip}	3.19	1.36	1.39	1.37	6.39	2.74



Supplementary Note 1 – Figure S1. Calculated RMSDs (Å) over time. RMSDs for the ligand heavy atoms for **a.** 2'AzUd and **b.** 2'AzCd in the canonical binding mode and **c.** 2'AzUd and **d.** 2'AzCd in the flipped binding mode from 25 – 100 ns. Each colored line represents each of the five MD simulations.

1.2.2 Comparisons to X-ray Crystal Structures

Here, we compared our metastable binding modes from our MD simulations against each subunit found in the experimental x-ray crystal structures and list the values for the randomly sampled frame against the subunit which minimizes the computed RMSD. To compare the metastable binding modes from our MD simulations with experimental x-ray crystal structures, we first align the two structures using the protein backbone. Particularly, we align by the protein backbone using residues 19 to 229 but exclude residues 48 to 52 as these were missing in the x-ray crystal structures. Once aligned by the protein backbone, we then compute the RMSD between crystallized ligand and the matching ligand heavy atoms from our defined metastable binding modes.

Supplementary Note 1 – Table S2A. Calculated RMSDs (in Angstroms) of the matching ligand heavy atoms of each metastable binding mode for 2’AzCd against the crystallized ligand (within each subunit) from our UCK2-2’AzCd-ATP structure.

	Subunit D	Subunit E	Subunit F
2’AzUd _{flip} (cluster 0)	3.17	5.02	4.93
2’AzUd _{flip} (cluster 1)	2.88	4.30	4.02
2’AzUd _{canc} (cluster 0)	5.28	4.54	4.14
2’AzUd _{canc} (cluster 1)	5.15	4.27	3.93

Supplementary Note 1 – Table S2B. Calculated RMSDs (in Angstroms) of the matching ligand heavy atoms of each metastable binding mode for 2’AzUd against the crystallized ligand (within each subunit) from our UCK2-2’AzUMP structure.

	Subunit A	Subunit B
2’AzUd _{flip} (cluster 0)	4.76	4.70
2’AzUd _{flip} (cluster 1)	4.45	4.17
2’AzUd _{canc} (cluster 0)	3.61	2.71
2’AzUd _{canc} (cluster 1)	3.34	2.44

Supplementary Note 1 – Table S2C. Calculated RMSDs (in Angstroms) of the matching ligand heavy atoms of each metastable binding mode for 2’AzUd against the crystallized ligand (within each subunit) from our UCK2-2’AzUd structure.

	Subunit A	Subunit B
2’AzCd _{flip} (cluster 0)	3.89	4.05
2’AzCd _{flip} (cluster 1)	2.95	2.90
2’AzCd _{flip} (cluster 2)	3.10	2.89
2’AzCd _{canc} (cluster 0)	0.76	1.22
2’AzCd _{canc} (cluster 1)	0.76	0.98
2’AzCd _{canc} (cluster 2)	0.84	1.17
2’AzCd _{canc} (cluster 3)	0.72	1.12

1.2.2 Hydrogen bond contacts

Using the ‘compute_contacts’ tool from MDTraj, we compute the distance between the closest heavy atoms in the ligand and nearby surrounding residues for hydrogen bonding. Here, we calculate the frequency in which the distance between the ligand and the residues are less than or equal to 3.0 Å, which we choose to define as the minimum distance needed to form a hydrogen bond. The frequency of hydrogen bond contact is calculated by taking the number of frames, in which the minimum distance is met, over the total number of

frames from each of our 100 ns MD simulations. Contact frequencies from each MD simulation is represented by each individual colored bar.

5.8 Supplementary Note 2 – Crystallographic Studies

Supplementary Note 2 – Table S1. Data collection and refinement statistics for UCK2 datasets with 2'-azidouridine and cytidine.

	2'-azidouridine	2'-azidouridine, ATP, MgCl ₂	2'-azidocytidine, ATP, MgCl ₂
Data collection			
Space group	P 2 ₁	F 2 2 2	F 2 2 2
Cell dimensions <i>a, b, c</i> (Å)	92.7, 85.8, 153.4	89.9, 142.4, 248.6	90.3, 141.3, 247.9
(°)	90, 95.4, 90	90, 90, 90	90, 90, 90
Resolution (Å)	50.0-3.1 (3.15-3.1)*	50.0-2.7 (2.75-2.7)	50.0-2.4 (2.44-2.4)
<i>R</i> _{sym}	0.089 (0.607)	0.078 (0.448)	0.097 (1.033)
<i>I</i> / <i>I</i>	22.4 (1.9)	29.3 (1.8)	31.9 (1.7)
Completeness (%)	99.8 (100.0)	98.0 (85.5)	99.9 (99.7)
Redundancy	3.7 (3.4)	8.0 (5.3)	7.0 (6.5)
Refinement			
Resolution (Å)	38.5-3.1 (3.2-3.1)	46.8-2.7 (2.8-2.7)	45.2-2.4 (2.5-2.4)
No. reflections	43704 (3766)	21895 (1867)	29822 (2616)
<i>R</i> _{work} / <i>R</i> _{free}	0.211 / 0.258	0.196 / 0.244	0.189 / 0.227
No.			
Protein atoms	12755	3256	3286
Ligand molecules: Nucleotide, Phosphate, Mg ²⁺	3, 8, NA	2, 2, 2	2, 2, 2
Solvent molecules: water, glycerol	184, 27	65, 5	145, 7
<i>B</i> -factors			
Protein	76.5	68.6	59.9
Ligands: Nucleotide, Phosphate, Mg ²⁺	103.3, 68.1, NA	100.0, 58.1, 91.0	79.2, 51.1, 79.9
Solvent: water, glycerol	62.8, 88.0	73.6, 93.2	63.0, 87.5
Ramachandran favored, outliers (%)	98.7, 0	99.8, 0	99.8, 0
R.M.S. deviations			
Bond lengths (Å)	0.002	0.002	0.003
Bond angles (°)	0.56	0.76	0.52

*Values in parentheses are for highest-resolution shell.

5.9 References

1. Jung, J.; Jung, H. *Anim. Cells Syst.* **2016**, *20*, 113-117.
2. Wu, Y. E.; Pan, L.; Zuo, Y.; Li, X.; Hong, W. *Neuron* **2017**, *96*, 313-329.e316.
3. Feng, H.; Zhang, X.; Zhang, C. *Nature Communications* **2015**, *6*, 7816.
4. Handley, A.; Schauer, T.; Ladurner, A. G.; Margulies, C. E. *Mol. Cell* **2015**, *58*, 621-631.
5. Guttman, M.; Russell, P.; Ingolia, Nicholas T.; Weissman, Jonathan S.; Lander, Eric S. *Cell* **2013**, *154*, 240-251.
6. Shigeoka, T.; Jung, J.; Holt, C. E.; Jung, H. In *RNA Detection: Methods and Protocols*; Gaspar, I., Ed.; Springer New York: New York, NY, 2018, p 85-94.
7. Riley, K. J.; Yario, T. A.; Steitz, J. A. *Rna* **2012**, *18*, 1581-1585.
8. Mili, S.; Steitz, J. A. *Rna* **2004**, *10*, 1692-1694.
9. Rabani, M.; Levin, J. Z.; Fan, L.; Adiconis, X.; Raychowdhury, R.; Garber, M.; Gnirke, A.; Nusbaum, C.; Hacohen, N.; Friedman, N.; Amit, I.; Regev, A. *Nat. Biotechnol.* **2011**, *29*, 436-442.
10. Fauster, K.; Hartl, M.; Santner, T.; Aigner, M.; Kreutz, C.; Bister, K.; Ennifar, E.; Micura, R. *ACS Chem. Biol.* **2012**, *7*, 581-589.
11. Nainar, S.; Beasley, S.; Fazio, M.; Kubota, M.; Dai, N.; Correa, I. R., Jr.; Spitale, R. C. *ChemBioChem* **2016**, *17*, 2149-2152.
12. Zheng, Y.; Beal, P. A. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 1799-1802.
13. Jao, C. Y.; Salic, A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 15779-15784.
14. Cleary, M. D.; Meiering, C. D.; Jan, E.; Guymon, R.; Boothroyd, J. C. *Nat. Biotechnol.* **2005**, *23*, 232.

15. Gay, L.; Miller, M. R.; Ventura, P. B.; Devasthali, V.; Vue, Z.; Thompson, H. L.; Temple, S.; Zong, H.; Cleary, M. D.; Stankunas, K.; Doe, C. Q. *Genes Dev.* **2013**, *27*, 98-115.
16. van Velthoven, C. T. J.; de Morree, A.; Egner, I. M.; Brett, J. O.; Rando, T. A. *Cell Reports* **2017**, *21*, 1994-2004.
17. Tomorsky, J.; DeBlander, L.; Kentros, C. G.; Doe, C. Q.; Niell, C. M. *eneuro* **2017**, *4*, ENEURO.0181-0117.2017.
18. Ghosh, A. C.; Shimell, M.; Leof, E. R.; Haley, M. J.; O'Connor, M. B. *Sci. Rep.* **2015**, *5*, 13176.
19. 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*, 2148-2151.
20. Zajackowski, E. L.; Zhao, Q.-Y.; Zhang, Z. H.; Li, X.; Wei, W.; Marshall, P. R.; Leighton, L. J.; Nainar, S.; Feng, C.; Spitale, R. C.; Bredy, T. W. *ACS Chem. Neurosci.* **2018**, *9*, 1858-1865.
21. 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*, e138-e138.
22. Miller, M. R.; Robinson, K. J.; Cleary, M. D.; Doe, C. Q. *Nat. Methods* **2009**, *6*, 439.
23. Chatzi, C.; Zhang, Y.; Shen, R.; Westbrook, G. L.; Goodman, R. H. *Eneuro* **2016**, *3*.
24. Li, J.; Huang, S.; Chen, J.; Yang, Z.; Fei, X.; Zheng, M.; Ji, C.; Xie, Y.; Mao, Y. *J. Hum. Genet.* **2007**, *52*, 415.
25. Jones, M. E. *Annu. Rev. Biochem.* **1980**, *49*, 253-279.
26. Suchi, M.; Mizuno, H.; Kawai, Y.; Tsuboi, T.; Sumi, S.; Okajima, K.; Hodgson, M. E.; Ogawa, H.; Wada, Y. *Am. J. Hum. Genet.* **1997**, *60*, 525-539.

27. Petryszak, R.; Keays, M.; Tang, Y. A.; Fonseca, N. A.; Barrera, E.; Burdett, T.; Fullgrabe, A.; Fuentes, A. M.; Jupp, S.; Koskinen, S.; Mannion, O.; Huerta, L.; Megy, K.; Snow, C.; Williams, E.; Barzine, M.; Hastings, E.; Weisser, H.; Wright, J.; Jaiswal, P.; Huber, W.; Choudhary, J.; Parkinson, H. E.; Brazma, A. *Nucleic Acids Res.* **2016**, *44*, D746-752.
28. Huang, M.; Graves, L. M. *Cell. Mol. Life Sci.* **2003**, *60*, 321-336.
29. Uhlen, M.; Fagerberg, L.; Hallstrom, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigartyo, C. A.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P. H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Ponten, F. *Science* **2015**, *347*.
30. Malami, I.; Abdul, A. B. *Biomed. Pharmacother.* **2019**, *109*, 1506-1510.
31. Suzuki, N. N.; Koizumi, K.; Fukushima, M.; Matsuda, A.; Inagaki, F. *Structure* **2004**, *12*, 751-764.
32. Van Rompay, A. R.; Norda, A.; Lindén, K.; Johansson, M.; Karlsson, A. *Mol. Pharmacol.* **2001**, *59*, 1181-1186.
33. Burger, K.; Mühl, B.; Kellner, M.; Rohrmoser, M.; Gruber-Eber, A.; Windhager, L.; Friedel, C. C.; Dölken, L.; Eick, D. *RNA Biol.* **2013**, *10*, 1623-1630.
34. Zhang, Y.; Kleiner, R. E. *J. Am. Chem. Soc.* **2019**, *141*, 3347-3351.
35. Tomoiike, F.; Nakagawa, N.; Fukui, K.; Yano, T.; Kuramitsu, S.; Masui, R. *Biochem. Biophys. Rep.* **2017**, *11*, 93-98.
36. Meinsma, R.; van Kuilenburg, A. B. *Nucleos. Nucleot. Nucl.* **2016**, *35*, 613-618.

37. Hishiki, T.; Kawamoto, S.; Morishita, S.; Okubo, K. *Nucleic Acids Res.* **2000**, *28*, 136-138.
38. Tomoike, F.; Nakagawa, N.; Kuramitsu, S.; Masui, R. *Biochemistry* **2011**, *50*, 4597-4607.
39. Agarwal, K. C.; Miech, R. P.; Parks, R. E. In *Methods Enzymol.*; Academic Press: 1978; Vol. 51, p 483-490.
40. Pfaffl, M. W. *Nucleic Acids Res.* **2001**, *29*, e45-e45.
41. McGann, M. J. *Comput.-Aided Mol. Des.* **2012**, *26*, 897-906.
42. Tanaka, W.; Shoji, M.; Tomoike, F.; Ujiie, Y.; Hanaoka, K.; Harada, R.; Kayanuma, M.; Kamiya, K.; Ishida, T.; Masui, R.; Kuramitsu, S.; Shigeta, Y. *Biophys. Physicobiol.* **2016**, *13*, 77-84.
43. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. *PLoS Comput. Biol.* **2017**, *13*, e1005659.
44. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. *J. Chem. Theory Comput.* **2015**, *11*, 1864-1874.
45. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950-1958.
46. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
47. Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623-1641.

48. Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. *J. Chem. Theory Comput.* **2015**, *11*, 5525-5542.
49. McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernandez, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S. *Biophys. J.* **2015**, *109*, 1528-1532.
50. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33-38, 27-38.

Chapter 6: Conclusions

6.1 Conclusions

The RNA transcriptome has revealed itself to be extraordinarily diverse in function, as the majority of cellular RNA is non-coding. It has further been demonstrated that RNA controls or regulates nearly every step of gene expression, from transcription to translation.¹ Unique RNA expression patterns give rise to many distinct cell types that are derived from the same genome. The RNA transcriptome is thus highly dynamic and varies amongst cell types, even within the same tissue. Despite the critical importance of RNA and its specificity of expression, there is a deficit of chemical methods to probe the transcriptomes of unique cell types residing in complex cellular environments.

Metabolic reporters, when properly identified and characterized, present a facile approach to install chemical functionality into cellular RNA. Azides and alkynes are currently the most amenable modifications that can be made to the base or sugar of nucleoside analogs.^{2,3} However, the former modification is more desirable because it eliminates the necessity of copper catalyst, which is both cytotoxic and can cause RNA degradation. The incorporation of an azide into RNA endows greater variety in the choice of conjugation, and can allow researchers to append fluorophores or other modifications in live cells or *in vivo*.^{4,5} Our lab has discovered a suite of azido-adenosine analogs that can be metabolically incorporated into cellular RNA, with the analog bearing an azide at the 2'-position of ribose being the most robustly incorporated.⁶ This 2'-modified adenosine has provided a multitude of uses to the lab, where in one particular application it was used in a chemical method for cell-specific RNA tagging through a photo-click reaction.

Moreover, the analysis of functional group tolerance and placement of modifications was expanded for work detailing vinyl-nucleoside reporters. These analyses have not only provided new tools to the community, but have also begun to bridge the gap in understanding of how these analogs are metabolized and their resulting effects on the cells. Tagging RNAs in this way presents an especially unique challenge because the functional group must be distributed throughout a transcript, only such that it may be efficiently functionalized and enriched. Over-incorporation may cause unintended secondary effects, which will mediate changes to gene expression that could modulate and skew the expected patterns of expression.

We also tested uridine and cytidine analogs with the same 2' modification. While 2'azidocytidine was robustly incorporated, 2'azidouridine was not. Although the kinase responsible for mono-phosphorylating these nucleosides is the same—uridine-cytidine kinase (UCK)—the literature reports that endogenous uridine is less favored as a substrate. Thus, further modification to uridine predictably made 2'azidouridine an even poorer substrate for phosphorylation. Interestingly, there are two isoforms of UCK, where UCK1 is the more highly expressed and substrate sensitive. UCK2, while rarely and very lowly expressed in normal tissues, has been demonstrated to be far more promiscuous and has been leveraged for phosphorylating nucleoside analogs.⁷ With this, the overexpression of UCK2 significantly improved the amount of 2'-azidouridine incorporation to levels similar to that of 2'-azidocytidine, paving the way for the creation of a new method for optimized cell-specific RNA labeling.

While up and coming, chemical approaches to profile nascent transcription are still underdeveloped in the field of chemical biology. The work presented here serves to

augment as well as characterize the various metabolic reporter approaches to install chemical functionality into cellular RNA. Importantly, all of these methods rely on the careful examination of endogenous cellular machinery that plays a role in nucleoside analog metabolism. A common thread throughout all of these projects has been to provide a deeper understanding of the limitations and stringency of these systems, and to supplant areas of significant challenges with optimized approaches. I have done this through widening the scope of chemistries that are available for RNA conjugation, as well as characterizing and developing cell-specific labeling approaches. Overall, my work can serve as a foundation for further efforts to install bioorthogonal functional groups and potentially other chemical modifications into cellular RNA, in addition to engineer orthogonal systems for cell-specific tagging in complex tissues and organisms.

6.2 References

1. Morris, K. V.; Mattick, J. S. *Nat. Rev. Genet.* **2014**, *15*, 423-437.
2. Salic, A.; Mitchison, T. J. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 2415-2420.
3. Jao, C. Y.; Salic, A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 15779-15784.
4. Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16793-16797.
5. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004**, *126*, 15046-15047.
6. Nainar, S.; Beasley, S.; Fazio, M.; Kubota, M.; Dai, N.; Correa, I. R., Jr.; Spitale, R. C. *ChemBioChem* **2016**, *17*, 2149-2152.
7. Van Rompay, A. R.; Norda, A.; Lindén, K.; Johansson, M.; Karlsson, A. *Mol. Pharmacol.* **2001**, *59*, 1181-1186.

