Distribution Agreement

Signature:

______________________________ _______________
Diane B. Karloff Date

Small-Molecule Tools to Probe and Modulate RNA Function

By

Diane B. Karloff
Doctor of Philosophy

Chemistry

Prof. Jennifer M. Heemstra
Advisor

Prof. Monika Raj
Committee Member

Prof. Khalid Salaita
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D., MPH
Dean of the James T. Laney School of Graduate Studies

Date

Small-Molecule Tools to Probe and Modulate RNA Function

By

Diane B. Karloff
B.S., Duke University, 2018
M.Sc., Emory University, 2023

Advisor: Jennifer M. Heemstra, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2023

Abstract

Small-Molecule Tools to Probe and Modulate RNA Function

By Diane B. Karloff

Once considered a mere messenger, RNA is now known to perform diverse regulatory roles in the cell, from maintaining genome stability and modulating gene expression to scaffolding protein:protein interactions and enabling cell polarization. Aberrant RNA processing is linked to a wide array of neurodegenerative diseases and cancers, and RNA also serves as the genetic material for dozens of devastating human viruses. As small molecules are well-established tools to interrogate and disrupt biological systems, the elucidation of principles that drive RNA:small molecule binding interactions facilitates basic science discovery and therapeutic drug development. This dissertation highlights opportunities that small molecules present toward exploring and perturbing cellular RNA function. In Chapter 1, we discuss four key themes relating to the clinical importance of the transcriptome and its amenability to small molecule-based probing. In Chapter 2, we describe a photoaffinity labeling method to improve the robustness of cellular RNA imaging. Visualization of RNA dynamics is critical to unraveling effects on cell signaling. Further, reducing the molecular weight of the cargo required for RNA imaging reduces impacts on native behavior. Chapter 3 explores photoaffinity labeling as a means to evaluate RNA:drug target engagement. We find that 10 chemical probes from diverse structural classes display unique binding profiles with breast cancer-associated messenger RNAs. Chapter 4 investigates the small molecule glyoxal as a thermoreversible regulator of nucleoside antiviral bioactivity. This strategy has the potential to improve patient adherence during antiretroviral therapy and enhance the poor pharmacokinetic profile associated with this compound class. Beyond the scope of RNA:small molecule interactions, Chapter 5 elaborates on the vital role that researchers play in shaping scientific literacy and accessibility in society. This role can be realized within departments, via social media, or through engagement in global organizations. In Chapter 6, we summarize the implications of the work described herein and present exciting future opportunities at the interface of RNA:ligand binding.

Small-Molecule Tools to Probe and Modulate RNA Function

By

Diane B. Karloff
B.S., Duke University, 2018
M.Sc., Emory University, 2023

Advisor: Jennifer M. Heemstra, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry
2023

Acknowledgments

It takes a village to train a scientist. Thank you to the countless people who shaped my path since the beginning of my education. It feels surreal to have made it to this point. Thanks, me, for the decade-plus of hard work and daily investment in myself—it compounds. (Chemistry pun?)

First, I would like to acknowledge my parents, Dr. Patricia Carneval and Art Karloff. Mom, thank you for passing me your warmth, conscientious regard for others, alt rock music taste, and secretly twisted sense of humor. You have added immeasurable beautiful experiences to my life, and I'm so lucky to be your daughter. Also, I would have never fallen in love with Molecular Biology without your prodding. Dad, the analytical, mathematical, and curious mindset I learned from you is *the* reason I have been successful in my STEM career. You taught me to always wonder and probe. I am also blessed with our fun times together, from rock climbing, pool races, volleyball coaching, and tennis to beach trips and corn mazes with friends. You've both always been there for me, hoping to help me every step of the way. I love you!

Thank you to the rest of my family for the constant love and security you bring to my life. I can't wait to come back up and hang out with the Erie Carnevals. Now seems like an appropriate time to quote Nana's needlepoint: "As soon as the rush is over, I'm going to have a nervous ~~BREAKDOWN~~ I worked for it. I owe it to myself & nobody is going to deprive me of it." I would like to specially acknowledge my grandfather, Dr. Daniel Carneval, whose work ethic, mastery, and devotion to the *WSJ* I hope to emulate throughout my life. Brian Foley, thank you for always providing me a delicious meal, a luxurious view, a backhanded golf compliment, and a lungful of beach air! I am so grateful to the Karloff/Gomberg/Fox/Karlofsky/Meadow/Natkin/Shahar/Page/Majola families for your unending love and support. Grandpa and Edna, I think of you at the piano.

Without my friends, everything would be pointless. The first shout-out goes to my OG homie Katie. I love you, bro. Can't wait to visit after wrapping up. Sending love to the Kanterbraus as well. Collin Hong, Nick Cosaboom, and Amy Liu, thank you for, amazingly, staying in close touch after all these years—our long friendships mean so much to me. Anna, my forever roommate and comedic soul mate—cheers to all the shenanigans that were, that are, and some that have not yet come to pass. Reenium (alt: Reenio), thank you for inspiring the best chemistry nickname and for the steady, warm, witty, and wise counsel you sprinkle into my life. @ Joe and Hansung (+ Yuming), brooo <33 To my ATL and STL roommates, Patrick, Ani, Gerry, and Benoît, I cherished all the late-night talks, debates, dinners, and laughs. It felt like home, and y'all kept me sane (well, mostly). Likewise for Smriti Rathi, Paulina Miranda et al., my Emory tennis crew, and Kornelia + Ian (a.k.a. potassium iodide)!

My outstanding teachers, professors, and mentors at Princeton High and Duke inspired me with their work ethic, passion, and curiosity. Mrs. Giammanco, you are still one of the liveliest chemists I've ever met—thank you for your infectious enthusiasm to start my journey! Dr. Qiu Wang, your Organic I and Chemical Biology classes were my favorites at Duke. Thank you for my graduate school recommendation letter and for being an incredible scientific role model. I was so lucky to be able to work at PTC Therapeutics during summer 2015 thanks to Mark Boulding. I loved learning organic chemistry, especially under the wonderful mentorship of Dr. Christie Morrill, and that experience had a huge impact on my career (before I even knew about phenotypic screens or alternative splicing!).

I am grateful to Dr. Terry Lechler for my first cell biology research opportunity—I still think of you sometimes in the cell culture hood—and to Dr. Kwaby Badu-Nkansah for your avid mentorship and love of science, which were important examples for me early in my research career. Dr. Amanda Hargrove, the way you see chemical biology and RNA:small molecule interactions has always captivated me and highlighted an exciting research path to pursue—you definitely get credit for my dissertation topic. Thank you also for the incredible, two-year scientific training opportunity in your lab. I am further grateful to the Duke Chemistry department for being my first chemistry home. Finally, Dr. Brittany Morgan, we both know I wouldn't be exactly where I am right now without you. I can't express how much you have helped me every time I have been standing at a professional/personal crossroads. You were always there with me, patiently telling me what I needed to hear. The research world is so lucky to have you, and I can't wait to see what's up next.

Now that the pre-graduate school folks have been covered, I would like to thank my research advisor, Dr. Jennifer Heemstra. You have always kept the lab well-funded so that I could focus on my work. You are an incredible writer and captivating speaker. I gravitated toward your lab because you are one of the hardest-working, most efficient, and highest-achieving people I know. Your hands-off mentoring approach proved more challenging to navigate than I anticipated, but I’ve had the space to grow into the researcher that I wanted to be, and I know the struggle made me stronger. Thank you in addition for introducing me to life-altering concepts like the growth mindset and crucial conversations. You also have a knack for attracting brilliant researchers. I am grateful for all of the lab members I have gotten to meet and work with over the last five years—everyone taught me something, and the work environment can really make or break your day as a scientist. It was a pleasure mentoring Eddy Pineda—I can’t wait to see what’s in store for you, boss! Big W. I would like to specially acknowledge my colleagues Dr. Tewoderos Ayele, Dr. Steve Knutson, Dr. Aimee Sanford, Dr. Brea Manuel, Dr. Travis Loya, and (future Dr.) Alexandria Quillin for the immense contributions you each made to my personal and professional development during graduate school.

I feel blessed to have been a part of the Emory Chemistry community. During hard times, it always made me feel better to know the awesome people in my cohort were on the same journey with me. @ Andrew, 🪲. Participating in NOBCChE @ Emory was the single most influential career experience I had in graduate school. Drs. Ordy Gnewou and Tamra Blue, you knocked your NOBCChE presidencies out of the park. It was an honor working with you and watching you do your thing—best of luck in your next steps! Much love to my peers on the 2021–22 SAFE Committee. Thank you, Dr. Khalid Salaita and Dr. Monika Raj, for your research mentorship and the many ways you inspire me as scientists. I am grateful for the passionate and talented Emory Chemistry professors I met during my Ph.D. Thank you to the department administrators and support staff, with special shout-outs to Drs. Kira Walsh, Antonio Brathwaite, Tracy McGill, Todd Polley, and Frank McDonald for your outstanding work to make academia more inclusive and engage in dialogue with students.

I am also grateful for my research collaborators inside and outside the lab. It was a privilege to work closely with Drs. Jay Schneekloth, Curran Rhodes, and Phil Homan at the NCI. Thank you, Dr. Debra Dantzler, for your optimistic outlook and profound insights. Lastly, thank you to the WUSTL Department of Chemistry for welcoming and hosting me during the final year of my Ph.D.

Table of Contents

List of Tables and Figures

Chapter 1: Introduction

1.1 RNA Function in Cells

RNA is canonically known as a cellular messenger, but in recent decades it has emerged as a structurally and functionally diverse biomolecule that plays a critical regulatory role in the body. Moreover, under aberrant conditions, RNA drives disease. Indeed, RNA is implicated in neurodegenerative disorders such as Alzheimer's, Huntington's disease, and amyotrophic lateral sclerosis,[1] as well as numerous cancers.[2-3] RNA also serves as the genetic material for various viruses, including HIV-1, hepatitis C, and SARS-CoV-2.[4-5] A flurry of research activity revolves not only around characterizing native RNA function, but leveraging RNA as a therapeutic target. However, there remains a pressing need to develop guiding principles for the design of RNA-targeted small molecules. Substantially more is known about the structure and druggability of the proteins that are encoded by these cellular transcripts, and targeting RNA is truly a paradigm shift in therapeutic development.[6-8]

Messenger RNA (mRNA) is among the earliest and best-studied RNA classes to be identified. Human mRNA, typically 3–5 kb long but potentially much larger,[9-10] is generated via DNA-templated synthesis by RNA polymerase II before translocation of the transcript from the cellular nucleus to the cytoplasm for ribosome-mediated translation into protein.[11-12] Importantly, mRNA can undergo alternative splicing to generate diverse protein isoforms, and dysregulation of this process can lead to the accumulation of pathogenic protein products.[13] For example, alternative splicing of the *survival of motor neuron 2* (SMN2) gene is an important therapeutic gateway to treat spinal muscular atrophy (SMA).[14] As less than 20% of the proteins comprising the human proteome are thought to contain "druggable pockets" that are amenable to binding by small-molecule therapeutics, the mRNAs upstream of the proteome represent a critical secondary opportunity to alter cellular function (Figure 1.1).[7, 15]

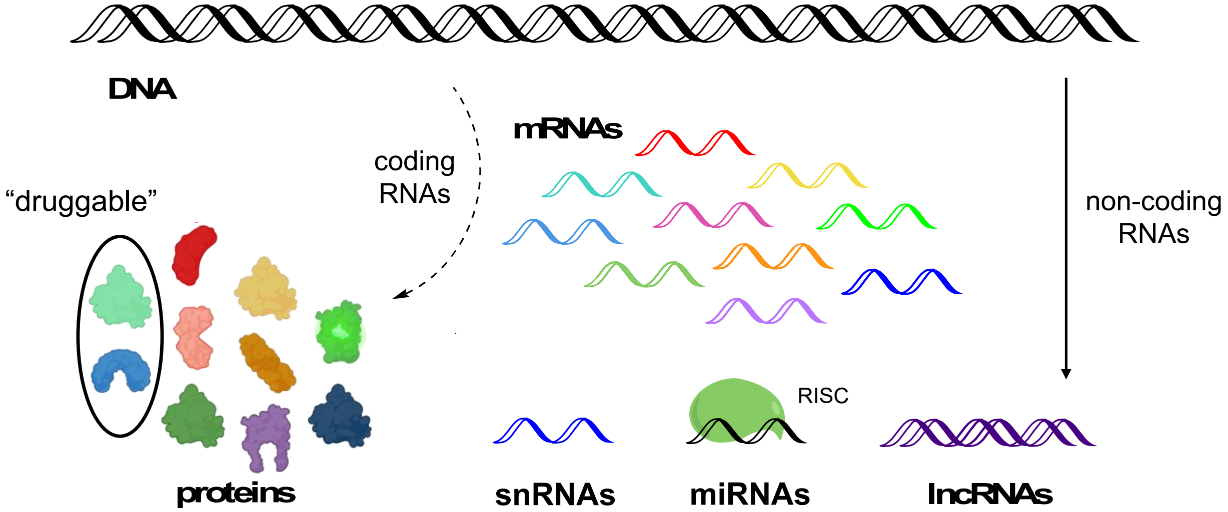


Figure 1.1: RNAs expand the scope of druggable cellular targets. mRNAs upstream of undruggable proteins present a second opportunity for therapeutic treatment. Non-coding RNAs are critical cellular regulators that can drive disease.

Fascinating work in the last few decades has revealed a plethora of non-coding RNA classes (Figure 1.1). MicroRNAs (miRNAs), small interfering RNAs (siRNAs), and piwi-interacting RNAs are among the shortest species at <50 nt in length and contain minimal structural complexity.[16] Mid-size non-coding RNAs including small nuclear RNAs (snRNAs), transfer RNAs, vault RNAs, Y RNAs, and more are 50–200 nt long. These transcripts play critical regulatory roles in the cell and are involved in processes ranging from translation to catalysis.[2, 10] For example, snRNAs direct intron removal from pre-mRNAs in nuclear Cajal bodies.[17-18] Meanwhile, long noncoding RNAs (lncRNAs) can range from 200 to >100,000 nucleotides in length and serve more diverse roles on a transcript-by-transcript basis.[3, 19] Whereas the lncRNA MALAT1 facilitates RNA processing and is upregulated in metastatic tumors,[20] HOTAIR scaffolds protein:protein interactions to support ubiquitination and senescence.[21]

A final class of biomedically relevant RNAs is found in RNA viruses and retroviruses. These pathogens contain RNA-based genomes that are translated by the infected host cell to produce viral proteins for viral replication or, in the latter case, first converted into cDNA by viral reverse-transcriptase (RT) enzymes and incorporated into the host genome for hijacking of host transcriptional and translational machinery.[4, 22-23]

RNA possesses exquisite capacity for molecular recognition. The above RNA classes adopt an astonishing array of secondary structures, such as stems, hairpins, bulges, internal loops, and G-quadruplexes.[24-25] These structures are mediated through hydrogen bonding between the nucleobases of two antiparallel regions of RNA (Watson-Crick-Franklin base-pairing), Hoogsteen base-pairing, and/or interactions with the phosphoribose backbone.[8, 26] The potential for base stacking can also promote the assembly of RNA into larger tertiary structural motifs such as triple helices, kissing loops, and pseudoknots.[27-28] miRNAs and siRNAs function primarily through complementarity with their mRNA targets.[16, 29] Meanwhile, RNA structure also strongly regulates RNA:protein interactions. The negatively-charged RNA backbone is a key feature for specific recognition by the thousands of known RNA-binding proteins.[30] As mentioned above, the striking flexibility of this biomolecule enables it to adopt complex structures for simultaneous binding of two or more proteins and scaffolding of key protein:protein interactions.[21]

Asymmetric RNA localization is responsible for protein accumulation that either forms a gradient across the cell or is highly locally concentrated.[31] For example, asymmetric distribution of β-actin is critical for neuron polarity and function in the brain.[31-33] The various RNA classes described above are differentially dispersed throughout the nucleus and the cytoplasm, and some transcripts are even anchored to the cell membrane.[34] A newly emerging field seeks to characterize the behavior of RNA in membraneless, liquid-liquid phase-separated (LLPS) biomolecular condensates such as stress granules, Cajal bodies, and P-bodies. RNA is thought to be sequestered in these condensates in response to environmental cues or the need for regulation of protein expression.[35-36] Another active area of research focuses on how to effectively visualize RNA distribution and dynamics, as probing the unique localization patterns of the thousands of different transcripts can help elucidate their function.[12] Early efforts to visualize RNA localization ingeniously leveraged native RNA:protein interactions that rely upon the molecular recognition properties described above.[37]

1.2 RNA:Small Molecule Interactions

Specific RNA molecular recognition also extends to small molecules. Bacteria possess a remarkable means of regulating intracellular metabolite levels through components known as riboswitches. These typically short (50–250 nt), highly conserved RNA sequences are often found in the 5' UTR of mRNAs that encode enzymes involved in metabolite production.[38-39] Often, when a high concentration of the metabolite is present, the riboswitch binds the molecule and undergoes a conformational change that inhibits transcription, splicing, or downstream translation, thereby maintaining metabolic homeostasis. Breaker and others have identified and characterized >55 distinct riboswitches classes, including binders of diverse structures such as the amino acid lysine, the cofactor S-adenosylmethionine, and the ion Mn^{2+}.[40]

Of potentially greater therapeutic interest is the opportunity to target pathogenic RNAs with exogenous small molecules. In this way, the function of an RNA that is overexpressed or goes awry to drive disease can be disrupted. Indeed, ligand binding could induce destabilization of toxic RNA or stabilization of beneficial RNA.[41] The first notable achievement in this area was the development of linezolid, an oxazolidinone-based antimicrobial that binds the bacterial ribosome, a ribonucleoprotein.[7, 42] Compounds that specifically target disease-associated human RNAs are perhaps even more highly coveted. Risdiplam, for example, is a drug that was approved by the FDA in 2017 to promote the inclusion of exon 7 for the production of full-length SMN2 in SMN1-deficient SMA patients.[14, 43] Meanwhile, nucleoside reverse transcriptase inhibitors are a leading class of viral RT inhibitors that indirectly target pathogenic RNA by disrupting retroviral cDNA synthesis.[44] However, the number of FDA-approved drugs that bind proteins dwarfs those that bind RNA.[7, 43]

Some groups postulate that because ribonucleotide properties diverge so widely from amino acid properties, rational drug design for each type of biomolecule should adapt accordingly. For example, one analysis was conducted by Morgan et al. compared the physicochemical, structural, and spatial properties of bioactive RNA- versus protein-targeted ligands. Interestingly, the RNA ligands were found to contain statistically significantly more nitrogen atoms and aromatic rings and fewer oxygen atoms (Figure 1.2a).[6] This knowledge could be used to identify an "RNA-privileged chemical space" that contains molecular structures which are more likely to bind RNA with high affinity and/or selectivity.[6, 45] Likewise, small-molecule screening libraries can be designed around "RNA-privileged" chemical scaffolds such as *bis*-benzimidazole, diphenylfuran, oxazolidinone, and others to increase hit rates against an RNA of interest.[41, 46-47] Finally, Weeks and coworkers posit that ligand identification is more likely to succeed with information-rich and structurally complex RNAs with protein-like small-molecule binding pockets (Figure 1.2b).[7] The existence of such pockets was confirmed through an analysis of RNA crystal structures in the Protein Data Bank.[48]

Various small-molecule tools for probing RNA structure and function have been reported to date. One such class of compounds is the electrophilic reagents that enable selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE). SHAPE reagents like 1-methyl-7-nitroisatoic anhydride and 2-methylnicotinic acid imidazolide preferentially react with the 2'-hydroxy group of the sugars in the RNA backbone that are more sterically accessible. Chemical probing thens enables classification of RNA regions as significantly or minimally involved in base-pairing.[49-50] Disney and coworkers also described a ribonuclease-targeting chimera (RIBOTAC) that recruits RNase L to an RNA of interest for *in cellulo* degradation and targeted analysis of

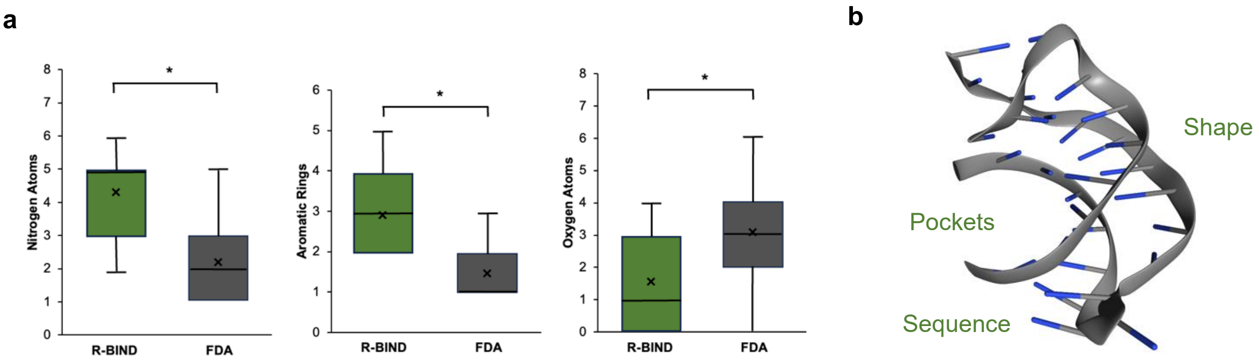


Figure 1.2: Developing guiding principles to drug RNAs. a RNA-targeted probes (R-BIND] differ structurally from protein-targeted probes (FDA). The boxes represent the middle 50% of the data, the whiskers represent the 10th and 90th percentiles, and the lines and crosses represent the medians and means, respectively. $^*p < 0.001$, Mann-Whitney *U* test. **b** RNA features that enhance ligandability are underexplored.

RNA function.[51] Thus, the amenability of RNAs to small-molecule binding fortuitously enables chemical biology studies on RNA behavior in a cellular context.

1.3 Photoaffinity Labeling

In the development of new small-molecule therapeutics, it is critical to validate that the drug engages the intended RNA target within a cellular context. To this end, compounds that react with an RNA of interest covalently would provide a more robust linkage than those that bind via a canonical, non-covalent mechanism.[52] This opens the door for more efficient target labeling in RNA imaging and RNA pulldown applications. While methods to address these goals in the context of protein:small molecule interactions are widely available and well-characterized,[53] comparable technology to assess RNA:small molecule target engagement and increase labeling efficiency remains in its infancy.

When seeking to identify the cellular targets of protein-binding small molecules, the incorporation of photoactivable crosslinker (PC) moieties into small-molecule probes has proven to be an invaluable strategy (Figure 1.3a).[54] The four major classes of PCs – benzophenones, psoralens, aryl azides, and diazirines (Figure 1.3c) – form high-energy intermediates upon irradiation with UV light, and in the process of relaxing back to the ground state, these intermediates covalently modify neighboring molecules.[53] Advantageously, PCs exhibit reactivity only during UV treatment, and their labeling results in an irreversible covalent bond that enables robust capture of transient interactions that could otherwise elude detection during pulldown and analysis. PCs have been used extensively in the proteomics field, where significant effort has been focused on genetically encoding noncanonical photocrosslinking amino acids,[55] as well as decorating protein-targeted small molecule probes with PC functionality for activity-based protein profiling.[53]

Covalent labeling strategies have also been employed to study carbohydrate[56-57] and lipid biology,[58] and early endeavors are underway to develop covalent, RNA-targeted chemical probes

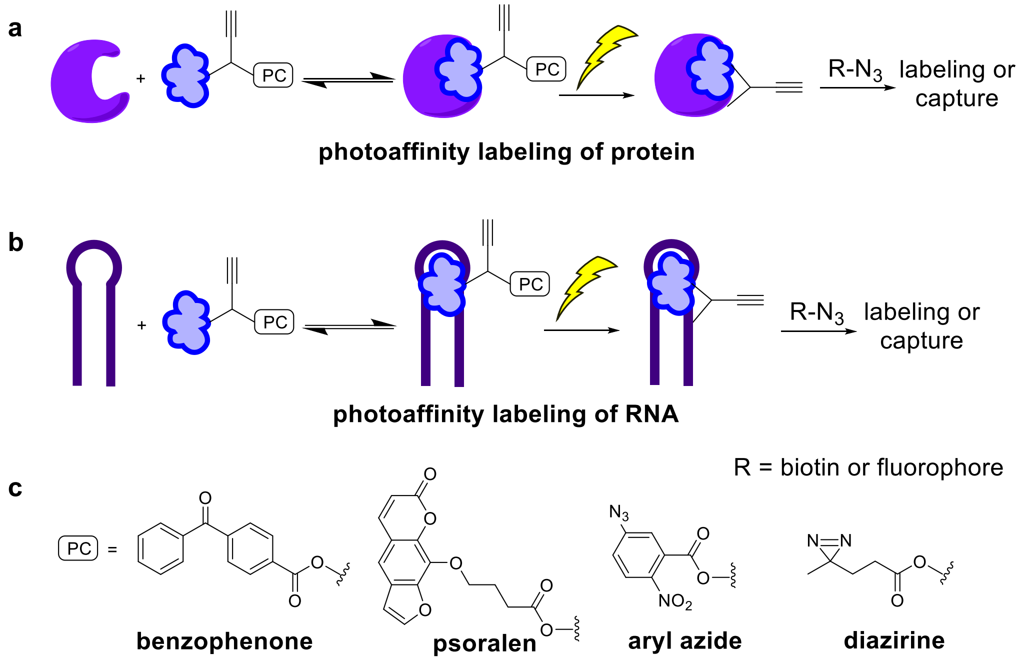


Figure 1.3: Photoaffinity labeling enables *in cellulo* capture of small-molecule interactions. a Scheme for protein:small molecule labeling. **b** Scheme for RNA:small molecule labeling. **c** Structures of four photoactivable crosslinker groups.

(Figure 1.3b). Disney and coworkers first reported Chem-CLIP, which utilizes a primary alkyl chloride electrophile and a biotin affinity handle to decorate a small molecule scaffold for target enrichment and analysis.[59] This covalent approach offers the advantage of irreversibly labeling cellular RNAs of interest for pulldown and identification.[60] However, the high background reactivity of the alkyl chloride with both nucleic acids and other cellular nucleophiles poses a significant challenge for this approach, and unlike PC probes, the reactivity of the alkyl chloride cannot be temporally controlled. Averting these limitations, Schultz and coworkers utilized a diazirine PC probe to successfully identify RNA binding site of Risdiplam and elucidate the mechanism of action.[61] While this demonstrates the promise of this approach, it explored only a single probe structure and PC class. Most recently, Disney and coworkers also explored diazirine-modified RNA ligands and chemical fragments for target engagement and discovery studies.[62-63]

It is important to note that the four PC classes (and the variants within these classes) possess distinct properties—including size, activation energy, transition state lifetime, and labeling range.[53] Given these differences in reactivity profiles, it is likely that the four PC classes react differentially with RNA versus proteins. Further, parameters such as the length and composition of the linker between an RNA probe and a PC group likely affect RNA labeling efficiency. In sum, deployment of RNA-targeted PC probes is an underutilized strategy for achieving robust RNA labeling and assessing RNA target engagement *in cellulo*. This approach can be explored more deeply to leverage the advantages of PC probes, namely that: (1) covalent trapping enables capture of transient or low-affinity interactions, and (2) light activation provides temporal control to allow capture under specific cellular conditions or in specific subcellular compartments.

1.4 RNA SELEX

The RNA:small molecule interactions described thus far include several that occur naturally, several that were identified through high-throughput screening of a synthetic compound library against an endogenous transcript of interest, and several that leverage electrophilic and/or photoreactive moieties to enhance ligand reactivity toward RNA. Another important tool in the RNA chemical biology arsenal is a method called systematic evolution of ligands by exponential enrichment (SELEX). This method enables the researcher to isolate a non-naturally occurring RNA sequence that ideally binds with high affinity and selectivity to nearly any small molecule of choice.[64] Typically, the *in vitro* evolution begins by incubating the ligand of interest with a library of $10^{13–15}$ unique RNA strands that share constant primer sequences. The RNAs that bind the target, otherwise known as aptamers, are enriched through a ligand pulldown step, converted back into DNA, amplified, and re-transcribed to generate a new RNA library with a higher abundance of the ligand-binding sequences (Figure 1.4).[64-65] The cycle is repeated until RNAs with the desired affinity and selectivity toward the target molecule are isolated. While many variations of this general protocol exist, all are geared toward engineering non-native RNA binding interactions for any number of applications.

This powerful method has been harnessed to sense a wide variety of species. Once the RNA sequence has been identified, it can be used *ex vivo* to analyze levels of therapeutics such as theophylline in samples extracted from patients.[66] Alternately, the aptamer can be expressed in bacterial or mammalian cells on its own or as a fusion with a gene of interest for *in cellulo* detection of metabolites such as fructose-6-bisphosphate, with readout coupled to fluorescence or other signal types.[64] SELEX has also enabled great strides in the field of RNA imaging. Unlike approaches that append protein-binding aptamers to an RNA of interest, small molecule-binding aptamers have the potential to reduce the attached molecular weight cargo. Further, if the small molecule has fluorogenic properties, it will emit minimal signal only when bound to the aptamer-tagged RNA, unlike constitutively fluorescent GFP.

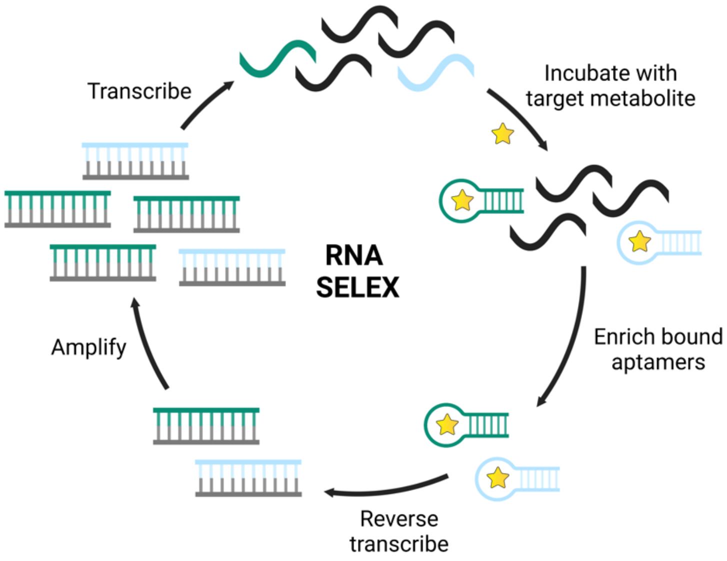


Figure 1.4: Isolation of small molecule-binding RNA aptamers. RNA aptamers for a molecule of interest are iteratively enriched and amplified from a random RNA sequence library.

A plethora of fluorogenic aptamers for RNA imaging have been reported over the years.[67] The first innovation in the field was the isolation of an RNA aptamer called Spinach that binds DMHBI, a small-molecule mimic of the GFP fluorophore.[68] In the last decade, researchers have developed aptamers that emit light across the visible spectrum.[69] Through continuous iteration, the most recent innovations using highly optimized dyes and carefully selected RNA binders are opening the door for single-molecule and super-resolution RNA imaging.[70-71]

1.5 Summary and Conclusions of this Dissertation

This dissertation aims to broadcast the utility of small molecules as powerful tools to explore native RNA properties, characterize drug target engagement, and disrupt transcript activity in living cells. The four key themes expounded upon above—RNA function in cells, RNA:small molecule interactions, photoaffinity labeling, and RNA SELEX—recur throughout the following chapters. Excitingly, the collective conclusions we are able to draw from each body of work not only expand our understanding of how to rationally design and deploy RNA:small molecule interactions, but simultaneously reinforce similar lessons and highlight the ample remaining opportunities to leverage these interactions toward endless chemical biology applications.

In Chapter 2, we describe the first photoaffinity-based method to enhance the robustness of fluorogenic RNA imaging. The method takes advantage of the malachite green aptamer (MGA), an RNA sequence identified via SELEX which binds the far-red dye malachite green (MG). Specifically, we append MGA to the U6 snRNA and visualize this construct in live cells using diazirine-functionalized MG. We demonstrate enhanced colocalization between the tagged RNA of interest and a nuclear Cajal body marker in a UV-dependent manner, showcasing the utility of photoaffinity labeling to achieve covalent capture of the RNA:dye interaction. Similarly, in Chapter 3 we decorate a panel of structurally diverse chemical fragments and biologically relevant compounds with a diazirine handle to explore *in cellulo* mRNA target engagement. In combination with RNA sequencing, this approach reveals distinct RNA binding patterns that are specific to each chemical entity. Such studies are useful for the elaboration of RNA as a therapeutic target and the discovery of guiding principles to design RNA-targeted drugs.

Expanding on prior work in our lab to modulate nucleic acid function, Chapter 4 takes advantage of the thermoreversible reaction between RNA nucleobases and the small molecule glyoxal to tune the bioactivity of nucleoside antivirals. Namely, we find that glyoxal caging impacts the potency of the pharmaceutical acyclovir toward HIV-1 *in vitro* and HSV-1 *in cellulo*. This innovation provides access to an extended-release mechanism that could improve patient adherence to antiviral treatment while addressing preexisting pharmacokinetic challenges with this class of compounds. Glyoxal thus presents a unique opportunity to target biomedically viruses, particularly retroviruses. Outside of the direct scope of RNA:small molecule interactions, Chapter 5 discusses steps that individual researchers and scientific advocacy groups can take to bolster scientific literacy in the general public. Indeed, the COVID-19 pandemic unveiled the widely varying levels of familiarity that members of our society have surrounding the principles that chemical biologists rely on on a day-to-day basis, from statistics and the Central Dogma to the Scientific Method. We as researchers arguably have a responsibility to promote understanding of these key themes among non-scientists. Finally, Chapter 6 synthesizes the work presented herein and proposes enticing future directions.

1.6 References

1. Storkebaum, E.; Rosenblum, K.; Sonenberg, N., Messenger RNA translation defects in neurodegenerative diseases. *N. Engl. J. Med.* **2023,** *388* (11), 1015-1030.

2. Cooper, T. A.; Wan, L.; Dreyfuss, G., RNA and disease. *Cell* **2009,** *136* (4), 777-793.

3. Qian, Y.; Shi, L.; Luo, Z., Long non-coding RNAs in cancer: Implications for diagnosis, prognosis, and therapy. *Front. Med. (Lausanne)* **2020,** *7*, 612393.

4. Poltronieria, P.; Sunb, B.; Mallardo, M., RNA viruses: RNA roles in pathogenesis, coreplication and viral load. *Curr. Genom.* **2015,** *16*, 327-335.

5. Khan, J.; Asoom, L. I. A.; Khan, M.; Chakrabartty, I.; Dandoti, S.; Rudrapal, M.; Zothantluanga, J. H., Evolution of RNA viruses from SARS to SARS-CoV-2 and diagnostic techniques for COVID-19: a review. *Beni-Suef Univ. J. Basic Appl. Sci.* **2021,** *10* (1), 60.

6. Morgan, B. S.; Forte, J. E.; Culver, R. N.; Zhang, Y.; Hargrove, A. E., Discovery of key physicochemical, structural, and spatial properties of RNA-targeted bioactive ligands. *Angew. Chem. Int. Ed. Engl.* **2017,** *56* (43), 13498-13502.

7. Warner, K. D.; Hajdin, C. E.; Weeks, K. M., Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **2018,** *17* (8), 547-558.

8. Vicens, Q.; Kieft, J. S., Thoughts on how to think (and talk) about RNA structure. *Proc. Natl. Acad. Sci. U. S. A.* **2022,** *119* (17), e2112677119.

9. Zhang, Z. Y.; Yang, Y. H.; Ding, H.; Wang, D.; Chen, W.; Lin, H., Design powerful predictor for mRNA subcellular location prediction in Homo sapiens. *Brief. Bioinform.* **2021,** *22* (1), 526-535.

10. Boivin, V.; Faucher-Giguere, L.; Scott, M.; Abou-Elela, S., The cellular landscape of mid-size noncoding RNA. *Wiley Interdiscip. Rev. RNA* **2019,** *10* (4), e1530.

11. Sentenac, A., Eukaryotic RNA polymerases. *CRC Crit. Rev. Biochem.* **1985,** *18* (1), 31-90.

12. Ben-Ari, Y.; Brody, Y.; Kinor, N.; Mor, A.; Tsukamoto, T.; Spector, D. L.; Singer, R. H.; Shav-Tal, Y., The life of an mRNA in space and time. *J. Cell Sci.* **2010,** *123* (Pt 10), 1761-74.

13. Scotti, M. M.; Swanson, M. S., RNA mis-splicing in disease. *Nat. Rev. Genet.* **2016,** *17* (1), 19-32.

14. Markati, T.; Fisher, G.; Ramdas, S.; Servais, L., Risdiplam: An investigational survival motor neuron 2 (SMN2) splicing modifier for spinal muscular atrophy (SMA). *Expert Opin. Investig. Drugs* **2022,** *31* (5), 451-461.

15. Dixon, S. J.; Stockwell, B. R., Identifying druggable disease-modifying gene products. *Curr. Opin. Chem. Biol.* **2009,** *13* (5-6), 549-55.

16. Kurreck, J., siRNA efficiency: Structure or sequence—that is the question. *J. Biomed. Biotechnol.* **2006,** *2006* (4), 83757.

17. Guthrie, C.; Patterson, B., Spliceosomal snRNAs. *Annu. Rev. Genet.* **1988,** *22*, 387-419.

18. Stanek, D., Cajal bodies and snRNPs - friends with benefits. *RNA Biol.* **2017,** *14* (6), 671-679.

19. Novikova, I. V.; Hennelly, S. P.; Sanbonmatsu, K. Y., Sizing up long non-coding RNAs. *BioArchitecture* **2012,** *2* (6), 189-199.

20. Arun, G.; Aggarwal, D.; Spector, D. L., MALAT1 long non-coding RNA: Functional

implications. *Noncoding RNA* **2020,** *6* (2).

21. Yoon, J. H.; Abdelmohsen, K.; Kim, J.; Yang, X.; Martindale, J. L.; Tominaga-Yamanaka, K.; White, E. J.; Orjalo, A. V.; Rinn, J. L.; Kreft, S. G.; Wilson, G. M.; Gorospe, M., Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. *Nat. Commun.* **2013,** *4*, 2939.

22. Payne, S., Introduction to RNA Viruses. In *Viruses*, 2017; pp 97-105.

23. Luganini, A.; Gribaudo, G., Retroviruses of the human virobiota: The recycling of viral genes and the resulting advantages for human hosts during evolution. *Front. Microbiol.* **2020,** *11*, 1140.

24. Eubanks, C. S.; Forte, J. E.; Kapral, G. J.; Hargrove, A. E., Small molecule-based pattern recognition to classify RNA structure. *J. Am. Chem. Soc.* **2017,** *139* (1), 409-416.

25. Fay, M. M.; Lyons, S. M.; Ivanov, P., RNA G-quadruplexes in biology: Principles and molecular mechanisms. *J. Mol. Biol.* **2017,** *429* (14), 2127-2147.

26. Takahashi, S.; Sugimoto, N., Watson-Crick versus Hoogsteen base pairs: Chemical strategy to encode and express genetic information in life. *Acc. Chem. Res.* **2021,** *54* (9), 2110-2120.

27. Higgs, P. G., RNA secondary structure: Physical and computational aspects. *Q. Rev. Biophys.* **2000,** *33* (3), 199-253.

28. Butcher, S. E.; Pyle, A.-M., The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* **2011,** *44* (12), 1302-1311.

29. Lam, J. K.; Chow, M. Y.; Zhang, Y.; Leung, S. W., siRNA versus miRNA as therapeutics for gene silencing. *Mol. Ther. Nucleic Acids* **2015,** *4* (9), e252.

30. Corley, M.; Burns, M. C.; Yeo, G. W., How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell* **2020,** *78* (1), 9-29.

31. Buxbaum, A. R.; Haimovich, G.; Singer, R. H., In the right place at the right time: Visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **2015,** *16* (2), 95-109.

32. Bassell, G. J.; Zhang, H.; Byrd, A. L.; Femino, A. M.; Singer, R. H.; Taneja, K. L.; Lifshitz,

L. M.; Herman, I. M.; Kosik, K. S., Sorting of β-actin mRNA and protein to neurites and growth cones in culture. *J. Neurosci.* **1998,** *18* (1), 251-265.

33. Mofatteh, M., mRNA localization and local translation in neurons. *AIMS Neurosci.* **2020,** *7* (3), 299-310.

34. Cohen, R. S., The role of membranes and membrane trafficking in RNA localization. *Biol. Cell* **2005,** *97* (1), 5-18.

35. Shin, Y.; Brangwynne, C. P., Liquid phase condensation in cell physiology and disease. *Science* **2017,** *357* (6357).

36. Anderson, P.; Kedersha, N., RNA granules: Post-transcriptional and epigenetic modulators of gene expression. *Nat. Rev. Mol. Cell Biol.* **2009,** *10* (6), 430-6.

37. George, L.; Indig, F. E.; Abdelmohsen, K.; Gorospe, M., Intracellular RNA-tracking methods. *Open Biol.* **2018,** *8* (10).

38. Mandal, M.; Breaker, R. R., Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **2004,** *5* (6), 451-63.

39. Serganov, A., The long and the short of riboswitches. *Curr. Opin. Struct. Biol.* **2009,** *19* (3), 251-9.

40. Kavita, K.; Breaker, R. R., Discovering riboswitches: The past and the future. *Trends Biochem. Sci.* **2023,** *48* (2), 119-141.

41. Childs-Disney, J. L.; Hoskins, J.; Rzuczek, S. G.; Thornton, C. A.; Disney, M. D., Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. *ACS Chem. Biol.* **2012,** *7* (5), 856-62.

42. Hashemian, S. M. R.; Farhadi, T.; Ganjparvar, M., Linezolid: A review of its properties, function, and use in critical care. *Drug Des. Devel. Ther.* **2018,** *12*, 1759-1767.

43. Sheridan, C., First small-molecule drug targeting RNA gains momentum. *Nat. Biotechnol.* **2021,** *39* (1), 6-8.

44. Seley-Radtke, K. L.; Yates, M. K., The evolution of nucleoside analogue antivirals: A

review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold. *Antiviral Res.* **2018,** *154*, 66-86.

45. Morgan, B. S.; Sanaba, B. G.; Donlic, A.; Karloff, D. B.; Forte, J. E.; Zhang, Y.; Hargrove, A. E., R-BIND: An interactive database for exploring and developing RNA-targeted chemical probes. *ACS Chem. Biol.* **2019,** *14* (12), 2691-2700.

46. Donlic, A.; Morgan, B. S.; Xu, J. L.; Liu, A.; Roble, C., Jr.; Hargrove, A. E., Discovery of small molecule ligands for MALAT1 by tuning an RNA-binding scaffold. *Angew. Chem. Int. Ed. Engl.* **2018,** *57* (40), 13242-13247.

47. Patwardhan, N. N.; Ganser, L. R.; Kapral, G. J.; Eubanks, C. S.; Lee, J.; Sathyamoorthy, B.; Al-Hashimi, H. M.; Hargrove, A. E., Amiloride as a new RNA-binding scaffold with activity against HIV-1 TAR. *Medchemcomm* **2017,** *8* (5), 1022-1036.

48. Hewitt, W. M.; Calabrese, D. R.; Schneekloth, J. S., Jr., Evidence for ligandable sites in structured RNA throughout the Protein Data Bank. *Bioorg. Med. Chem.* **2019,** *27* (11), 2253-2260.

49. Wilkinson, K. A.; Merino, E. J.; Weeks, K. M., Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **2006,** *1* (3), 1610-6.

50. Busan, S.; Weidmann, C. A.; Sengupta, A.; Weeks, K. M., Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. *Biochemistry* **2019,** *58* (23), 2655-2664.

51. Costales, M. G.; Suresh, B.; Vishnu, K.; Disney, M. D., Targeted degradation of a hypoxia-associated non-coding RNA enhances the selectivity of a small molecule interacting with RNA. *Cell Chem. Biol.* **2019,** *26* (8), 1180-1186 e5.

52. Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; Gonzalez-Paez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J.; Wolan, D. W.; Cravatt, B. F., Proteome-wide covalent ligand discovery in native biological systems. *Nature* **2016,** *534* (7608), 570-4.

53. Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J. S., Photo-affinity labeling (PAL) in

chemical proteomics: A handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci.* **2016,** *15*, 14.

54. Hatanaka, Y.; Sadakane, Y., Photoaffinity labeling in drug discovery and developments: chemical gateway for entering proteomic frontier. *Curr. Topic. Med. Chem.* **2002,** *2* (3), 271-288.

55. Young, D. D.; Schultz, P. G., Playing with the molecules of life. *ACS Chem. Biol.* **2018,** *13* (4), 854-870.

56. Tanaka, Y.; Bond, M. R.; Kohler, J. J., Photocrosslinkers illuminate interactions in living cells. *Mol. Biosyst.* **2008,** *4* (6), 473-80.

57. Tanaka, Y.; Kohler, J. J., Photoactivatable crosslinking sugars for capturing glycoprotein interactions. *J. Am. Chem. Soc.* **2008,** *130*, 3278-3279.

58. Peng, T.; Yuan, X.; Hang, H. C., Turning the spotlight on protein-lipid interactions in cells. *Curr. Opin. Chem. Biol.* **2014,** *21*, 144-53.

59. Rzuczek, S. G.; Colgan, L. A.; Nakai, Y.; Cameron, M. D.; Furling, D.; Yasuda, R.; Disney, M. D., Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat. Chem. Biol.* **2017,** *13* (2), 188-193.

60. Disney, M. D., Targeting RNA with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine. *J. Am. Chem. Soc.* **2019,** *141* (17), 6776-6790.

61. Wang, J.; Schultz, P. G.; Johnson, K. A., Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc. Natl. Acad. Sci. U.S.A.* **2018,** *115* (20), E4604-E4612.

62. Suresh, B. M.; Li, W.; Zhang, P.; Wang, K. W.; Yildirim, I.; Parker, C. G.; Disney, M. D., A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2020,** *117* (52), 33197-33203.

63. Tong, Y.; Gibaut, Q. M. R.; Rouse, W.; Childs-Disney, J. L.; Suresh, B. M.; Abegg, D.; Choudhary, S.; Akahori, Y.; Adibekian, A.; Moss, W. N.; Disney, M. D., Transcriptome-wide mapping of small-molecule RNA-binding sites in cells informs an isoform-specific degrader of QSOX1 mRNA. *J. Am. Chem. Soc.* **2022,** *144* (26), 11620-11625.

64. Karloff, D. B.; Heemstra, J. M., Sweet sensation: Developing a single-cell fluorescent reporter of glycolytic heterogeneity. *Cell Chem. Biol.* **2021,** *28* (11), 1539-1541.

65. Zhuo, Z.; Yu, Y.; Wang, M.; Li, J.; Zhang, Z.; Liu, J.; Wu, X.; Lu, A.; Zhang, G.; Zhang, B., Recent advances in SELEX technology and aptamer applications in biomedicine. *Int. J. Mol. Sci.* **2017,** *18* (10).

66. Ferapontova, E. E.; Olsen, E. M.; Gothelf, K. V., An RNA aptamer-based electrochemical biosensor for detection of theophylline in serum. *J. Am. Chem. Soc.* **2008,** *130* (13), 4256-4258.

67. Chen, W.; Zhao, X.; Yang, N.; Li, X., Single mRNA imaging with fluorogenic RNA aptamers and small-molecule fluorophores. *Angew. Chem. Int. Ed. Engl.* **2023,** *62* (7), e202209813.

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

69. Chen, X.; Zhang, D.; Su, N.; Bao, B.; Xie, X.; Zuo, F.; Yang, L.; Wang, H.; Jiang, L.; Lin, Q.; Fang, M.; Li, N.; Hua, X.; Chen, Z.; Bao, C.; Xu, J.; Du, W.; Zhang, L.; Zhao, Y.; Zhu, L.; Loscalzo, J.; Yang, Y., Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nat. Biotechnol.* **2019,** *37*, 1287-1293.

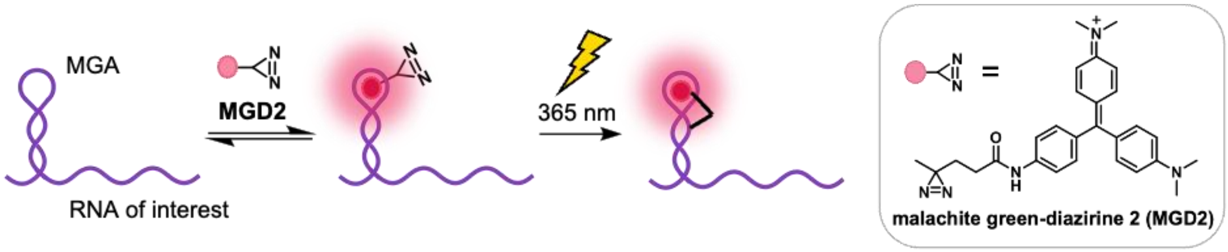
70. Cawte, A. D.; Unrau, P. J.; Rueda, D. S., Live cell imaging of single RNA molecules with fluorogenic Mango II arrays. *Nat. Commun.* **2020,** *11* (1), 1283.

71. Englert, D.; Burger, E. M.; Grun, F.; Verma, M. S.; Lackner, J.; Lampe, M.; Buhler, B.; Schokolowski, J.; Nienhaus, G. U.; Jaschke, A.; Sunbul, M., Fast-exchanging spirocyclic rhodamine probes for aptamer-based super-resolution RNA imaging. *Nat. Commun.* **2023,** *14* (1), 3879.

Chapter 2: Imaging and Tracking RNA in Live Mammalian Cells via Fluorogenic Photoaffinity Labeling

2.1 Abstract

Cellular RNA labeling using light-up aptamers that bind to and activate fluorogenic molecules has gained interest in recent years as an alternative to protein-based RNA labeling approaches. Aptamer-based systems are genetically encodable and cover the entire visible spectrum. However, the inherently temporary nature of the non-covalent aptamer-fluorogen interaction limits the utility of these systems in that multiple copies of the aptamer are often required, and imaging does not withstand dye washout. We propose that these limitations can be averted through covalent RNA labeling. Here we describe a photoaffinity approach in which the aptamer ligand is functionalized with a photoactivatable diazirine reactive group such that irradiation with UV light results in covalent attachment to the RNA of interest. In addition to the robustness of the covalent linkage, this approach benefits from the ability to achieve spatiotemporal control over RNA labeling. To demonstrate this approach, we incorporated a photoaffinity linker into malachite green and fused a single copy of the malachite green aptamer to a Cajal body-associated small nuclear RNA of interest. We observed improved sensitivity for live cell imaging of the target RNA upon UV irradiation and demonstrate visualization of RNA dynamics over a timescale of minutes.



2.2 Introduction

Trafficking of RNA to subcellular compartments plays an essential role in RNA homeostasis and cellular function. This spatiotemporal control of RNA localization is a common characteristic for a significant fraction of transcripts,[72-74] and in recent years, fluorescence microscopy has dramatically increased our understanding of the heterogeneity of transcript regulation and the complex subcellular interactions of RNAs and proteins. However, accurately mapping RNA localization relies on the ability to fluorescently label cellular RNAs without significantly perturbing their structure or localization. The earliest approaches to fluorescently tagging cellular RNAs utilized probes capable of binding to the RNA of interest (ROI) through Watson-Crick-Franklin base pairing, including fluorescence *in situ* hybridization (FISH) probes and molecular beacons. While these methods yielded much of the current day knowledge on RNA localization, they can require cell fixation or other approaches to disrupt native RNA folding and thus are limited for studying the trafficking and dynamics of cellular RNAs.[75-77] Currently, the most widely used approach for visualizing RNA uses GFP-fused RNA-binding proteins such as MS2, λ_N, PCP, or Cas proteins.[78-81] These fluorescently-tagged proteins recognize a specific RNA sequence that is incorporated multiple times into an ROI. Optimizing signal to noise in the fluorescent protein system usually requires adding a nuclear localization signal in order to reduce non-specific signal in the cytoplasm. As a trade-off, imaging nuclear RNAs is difficult due to high levels of nuclear-localized fluorescent protein, unless further optimizations are introduced.[82] Another approach to improving protein-based RNA imaging methods is to incorporate multiple copies of the binding site into the target RNA sequence. The additional binding sites, as well as the heavy load of the associated proteins (>1.3 MDa), can alter the native localization and functional properties of the RNA.[83-84]

Building off of Grate and Wilson's development of the malachite green aptamer (MGA) for chromophore-assisted laser inactivation in 1999,[85] Tsien and coworkers proposed that the small molecule recognition capabilities of RNA could potentially be used for RNA labeling, and they reported *in vitro* fluorescence enhancement of MG and sulforhodamine B in the presence of their respective aptamers.[86] Over the past decade, a number of RNA aptamers binding to fluorogenic dyes beyond malachite green (MG) have been reported, including the Spinach,[87-88] Broccoli,[89] and Mango[90] aptamers. These aptamers have been fused to RNAs and expressed in cells to enable RNA visualization. However, these approaches typically require fusion of multiple copies of the aptamer to enable fluorescence imaging, sometimes within a folding scaffold that can alter target RNA localization or stability, and even then struggle to detect low-abundance RNAs.[89, 91-92] Recently, multiple groups have reported significant advances in aptamer labeling. Cawte et al. developed an updated Mango aptamer-based system, improving signal to noise, and Yang and coworkers reported the Peppers aptamer system, which has a high signal-to-background ratio and was used to image nuclear foci in mammalian cells.[69-70] Notably, rhodamine dyes have especially progressed the RNA labeling field by allowing single molecule imaging, single-molecule localization microscopy (SMLM), and stimulated emission depletion (STED) microscopy.[71, 93-96] While the stability and cellular brightness of these systems offered a notable improvement compared to the previously reported aptamer-based systems, this overall approach still suffers from inherent limitations as a result of its non-covalent nature. For example, the reversibility of the fluorescent probe interaction with the ROI makes Peppers and other non-covalent systems unsuitable for applications that require media exchange, as labeling does not withstand washout steps. Additionally, these systems are not suitable for time-resolved investigations such as pulse-chase and tracking experiments, as the dye molecules are likely transferred between different RNA molecules over the course of the experiment, compromising the reliability of the data.

We hypothesized that these limitations could be overcome through covalent attachment of the fluorescent molecule to the target RNA. Specifically, covalent attachment would be expected to provide increased robustness of RNA labeling to maximize signal-to-background ratio with fewer aptamer copies and would allow the labeling event to withstand media exchange or washing steps. The ability to wash out unbound dyes offers a significant advantage, as many of these dyes have off-target effects and exhibit cellular toxicity, and thus minimizing cellular exposure decreases the risk of unwanted side effects that could perturb cellular function. Additionally, we envisioned that using a photoaffinity labeling group would provide spatiotemporal control over the labeling process, which is not possible using existing methods. Jäschke and coworkers recently achieved covalent RNA labeling based on the proximity-induced interaction between a dienophile-modified RNA aptamer and the corresponding tetrazine-modified ligand,[97] although this method is not genetically encodable due to the requirement for chemical modification of the aptamer. To achieve our goal, we utilized the malachite green aptamer, as binding to the malachite green dye results in significant red-shifted fluorescence enhancement.[85, 98] The excitation and emission maxima for the MG fluorogen are also located in the far-red region of the UV spectrum, averting the inherent challenges associated with cellular autofluorescence and making this aptamer-ligand pair well-suited for live cell imaging. One the first aptamers selected for a fluorogenic dye, MGA has been used for a wide range of applications outside of RNA imaging.[99-102] Despite having among the highest reported *in vitro* fluorescence enhancement of the fluorogenic dyes in the presence of its cognate aptamer, the potential of MG has not been fully realized as an RNA imaging tool due to reported cytotoxicity,[103] and MG studies were instead mostly performed in bacterial cells.[104-105] However, we were encouraged by reports that MG derivatization could mitigate cytotoxicity without significantly compromising fluorescence enhancement,[106-107] and that with our covalent labeling approach that enables fluorophore washout, cell exposure to the dye could be minimized.

To covalently label the RNA of interest, we envisioned that the aptamer could be fused to the ROI, expressed in cells, and the cells incubated with a derivative of the MG ligand having a photoaffinity handle. UV irradiation would convert the non-covalent RNA:small molecule binding interaction into a covalent linkage, resulting in robust and spatiotemporally controlled labeling of the ROI (Figure 2.1a). To create the photoreactive fluorogen, we appended a diazirine linker to

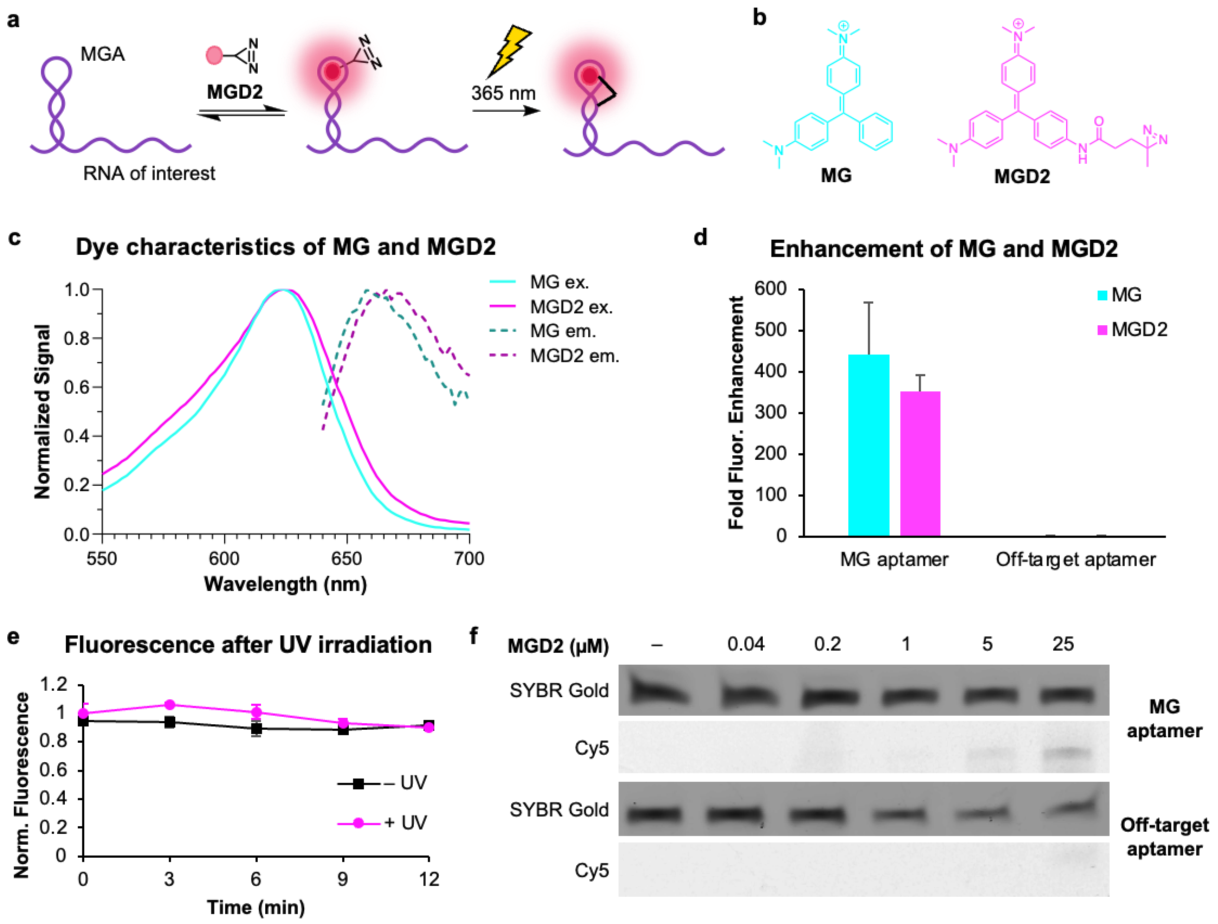


Figure 2.1: *In vitro* characterization of the MGA/MGD2 interaction. a Schematic representation of fluorogenic photoaffinity labeling with MGA-functionalized RNA. The RNA binds to the fluorogenic dye and induces fluorescence enhancement. UV irradiation results in covalent attachment of the dye to the ROI. **b** Structures of MG and MGD2. The canonical MG molecule was functionalized with a diazirine linker to enable photoaffinity labeling of MGA. **c** Representative normalized excitation (solid) and emission (dashed) spectra of MG (cyan) or MGD2 (magenta) bound to MGA. **d** Fluorescence enhancement of MG and MGD2 in the presence of MGA or an off-target aptamer in 1X PBS. Bars and error bars represent the mean and standard deviation from n = 3 independent samples. **e** Normalized, time-dependent fluorescence enhancement of MGD2 with MGA upon irradiation at 365 nm in 1X PBS. Points and error bars represent the mean and standard deviation from n = 3 independent samples. **f** Denaturing PAGE analysis of UV-dependent covalent labeling of MGA or an off-target aptamer (20 nM) with increasing concentrations of MGD2.

the dye (Figure 2.1b, Schemes A1–A2) and termed this new, minimal MG derivative malachite green-diazirine 2 (MGD2). We previously used a similar approach to design a derivative of MG termed MGD for photoaffinity labeling of fluorogen-activating proteins in live cells.[108] However, this strategy had yet to be applied to cellular RNA imaging. Upon UV-A irradiation at 365 nm, the diazirine linker is activated and produces a carbene moiety that readily reacts with nearby C-H and heteroatom-H bonds.[53] UV-C (254 nm) irradiation has been used for crosslinking RNA-protein interactions by taking advantage of the photoresponsiveness of natural amino acids such as Cys, Lys, Phe, Trp, and Tyr.[109-110] Here, the longer wavelength of 365 nm used to activate MGD2 ensures that our design does not result in unwanted crosslinking of RNA with cellular proteins. Moreover, diazirine and other photocrosslinking moieties are routinely used for *in cellulo* characterization of biomolecular interactions,[61, 111-114] and the UV irradiation required for crosslinking is known to be compatible with live cells.

Using our photoaffinity approach, we demonstrate that an RNA of interest can be labeled and imaged in live cells using a single 38 nt MGA fusion, thereby minimizing perturbation of RNA structure and localization. We show that covalent labeling enables RNA visualization under conditions where the non-covalent system performs far worse, and we utilize our approach to monitor the dynamics of RNAs in Cajal bodies. Our findings reveal that covalent attachment greatly improves RNA tracking by increasing the number of detectable foci that can be tracked for longer durations. Together, this research provides the first covalent method for cellular RNA labeling and introduces an easy-to-use strategy for the RNA community. The added robustness and temporal control achieved using a photoaffinity-based approach is anticipated to advance RNA imaging capabilities, providing important new insights into the role of RNA trafficking in biological processes such as development and disease.

2.3 Results and Discussion

2.3.1 In vitro *characterization*

To establish the feasibility of our photoaffinity labeling approach, we carried out *in vitro* experiments to investigate the reactivity and selectivity of the aptamer with the MGD2 derivative. Both MG and MGD2 displayed well-defined absorbance and fluorescence profiles with excitation maxima at 625 nm and emission maxima at 660 and 666 nm, respectively, in the presence of

MGA (Figure 2.1c). Prior to UV irradiation, fluorescence measurements revealed a 440-fold fluorescence enhancement for MG bound to MGA and a 350-fold fluorescence enhancement for MGD2 bound to MGA (Figure 2.1d). Neither probe showed measurable fluorescence enhancement in the presence of an off-target aptamer, indicating that the fluorophore recognition and fluorescence increase are specific to the MG aptamer sequence. We next investigated the effect of UV irradiation on fluorescence enhancement. For this experiment, we measured the fluorescence signal of a solution of MGD2 and MGA that was exposed to no light or up to 15 minutes of UV irradiation (Figure 2.1e). Fortuitously, no UV-dependent reduction in fluorescence was observed, indicating that crosslinking does not impede MG binding and that irradiation at 365 nm does not result in appreciable photobleaching of the fluorophore.

We also investigated the specificity of labeling using denaturing PAGE analysis. MGA and an off-target aptamer (20 nM) were incubated with increasing concentrations of MGD2 (40 nM–25 µM) and irradiated for 15 minutes. We observed concentration-dependent labeling of MGA but not the negative control aptamer (Figure 2.1f, Figure A1a), and MGA labeling increased with UV irradiation time (Figure A1b). We were further encouraged upon testing the selective reaction of MGD2 (2.5 or 25 µM) with MGA in the presence of total cellular RNA (Figure A2). After 15 minutes of UV irradiation, we observed bands in the MG channel corresponding to the target RNA with excellent specificity over other cellular RNAs. Analysis of MGD2-labeled MGA by HPLC indicated an approximate RNA labeling efficiency of 1.7% (Figure A3). While this is not an atypical yield for photocrosslinking, it also suggests that future optimization efforts focused on increasing RNA labeling yield could further improve our method.[86, 113] The collective observations from these experiments served as encouragement that our photoaffinity approach could provide selective labeling of the ROI in cells and offered guidelines for irradiation time and dye concentrations to be used in subsequent experiments.

To assess the compatibility of our approach with live cells, mouse neuroblastoma Neuro-2a (N2a) cells were treated with 200 nM MGD2 and/or irradiated with UV light under the same conditions used in later imaging experiments, then monitored for viability over 8 hours using the CellTox (Promega) assay per the manufacturer's kinetic protocol. There was no observable increase in MGD2- or UV-dependent cell death compared to the "no treatment" negative control, nor did treatment with both UV and MGD2 induce cell death up to 8 hours post-treatment (Figure A4). These results offered evidence that neither MGD2 nor UV exposure would lead to observable cytotoxicity at the levels used for live-cell imaging in this study.

2.3.2 Live-cell imaging of RNA

Having demonstrated our photoaffinity labeling approach *in vitro*, we next turned to live cell experiments to validate labeling of a target RNA. As a biological context for testing our labeling method, we chose the Cajal body. The Cajal body is a nuclear organelle found in many proliferative and/or metabolically active eukaryotic cells. Characteristically defined by the presence of coilin protein, the Cajal body accumulates proteins and RNAs involved in several aspects of RNA metabolism.[18] Small nuclear ribonucleoprotein (snRNP) metabolism, telomere maintenance, and histone mRNA processing are all associated with Cajal bodies.[115] Most importantly for our experiments, uridine-rich small nuclear RNAs (U snRNAs) preassemble within Cajal bodies before associating with the spliceosome to direct precursor messenger RNA splicing and intron removal.[116] snRNAs are challenging to visualize in cells using canonical RNA imaging methods because of their small size and nuclear localization. Specifically, the signal of snRNA-targeted FISH probes is limited due to the short length across which to tile fluorescent, complementary oligonucleotides, and the requirement for cell fixation means that FISH cannot capture the dynamic nature of snRNA localization. Meanwhile, the MS2-MCP system adds molecular weight far outweighing that of the snRNA itself, which could perturb snRNA function and localization, and the system must be further optimized to achieve suitable signal-to-noise ratio for imaging RNA in the nucleus.[82-83] We do note that several groups have successfully leveraged RNA:small molecule interactions to visualize U snRNAs,[95, 117-118] and here we aim to demonstrate the added benefits that can be achieved by covalent dye attachment.

To test our photoaffinity labeling method in live cells, we inserted a single copy of MGA into the 5’ UTR of U6 snRNA and also generated a control plasmid having U6 snRNA with no aptamer fusion. We hypothesized that colocalization of MGD2-labeled U6 snRNA with the Cajal body marker coilin-GFP would increase in an MGA- and UV-dependent manner (Figure 2.2a). To probe this question, N2a cells were exposed to one of four treatment conditions in which the cells were mock transfected or co-transfected with coilin-GFP plasmid and either 0X MGA- or 1X MGA-U6 snRNA plasmid. After 72 h, all wells were incubated with 100 nM MGD2 in fresh media for 15 min. Cell plates were then stored in the dark or irradiated using 365 nm UV light for 10 minutes to allow for covalent labeling of the aptamer (Figure 2.2a). After exchanging the media to wash away unbound dye, the cells were imaged using confocal microscopy (representative Figure 2.2b, Figures A5–A6).

Successful transfection with coilin-GFP was clearly observed by the appearance of green fluorescent puncta in conditions 2–4 as compared to condition 1 (untransfected). Relative to condition 1, MGD2 signal likewise appeared higher in conditions 2, 3, and 4, in which dye treatment was combined with UV irradiation, transfection with 1X MGA-U6 snRNA plasmid, or both, respectively. Given the relatively low signal-to-background ratios of traditional aptamer-based RNA imaging strategies, it was unsurprising that differences in red signal between images of conditions 2–4 were not easily discernible by eye.[95] Therefore, line scans were drawn in Fiji across images of nuclei from the four treatment conditions, revealing that MGD2 and coilin-GFP signals colocalized with the highest incidence in condition 4, in which the aptamer is present and UV light is used to stimulate covalent labeling (representative Figure 2.2c, Figure A7). To further quantify colocalization between MGD2 signal and coilin-GFP signal, we drew 15–30 nuclear line scans for each experimental condition. We then analyzed the red and green curves and the areas underneath them, using coilin-GFP peaks as “ground truth” to define Cajal body localization.

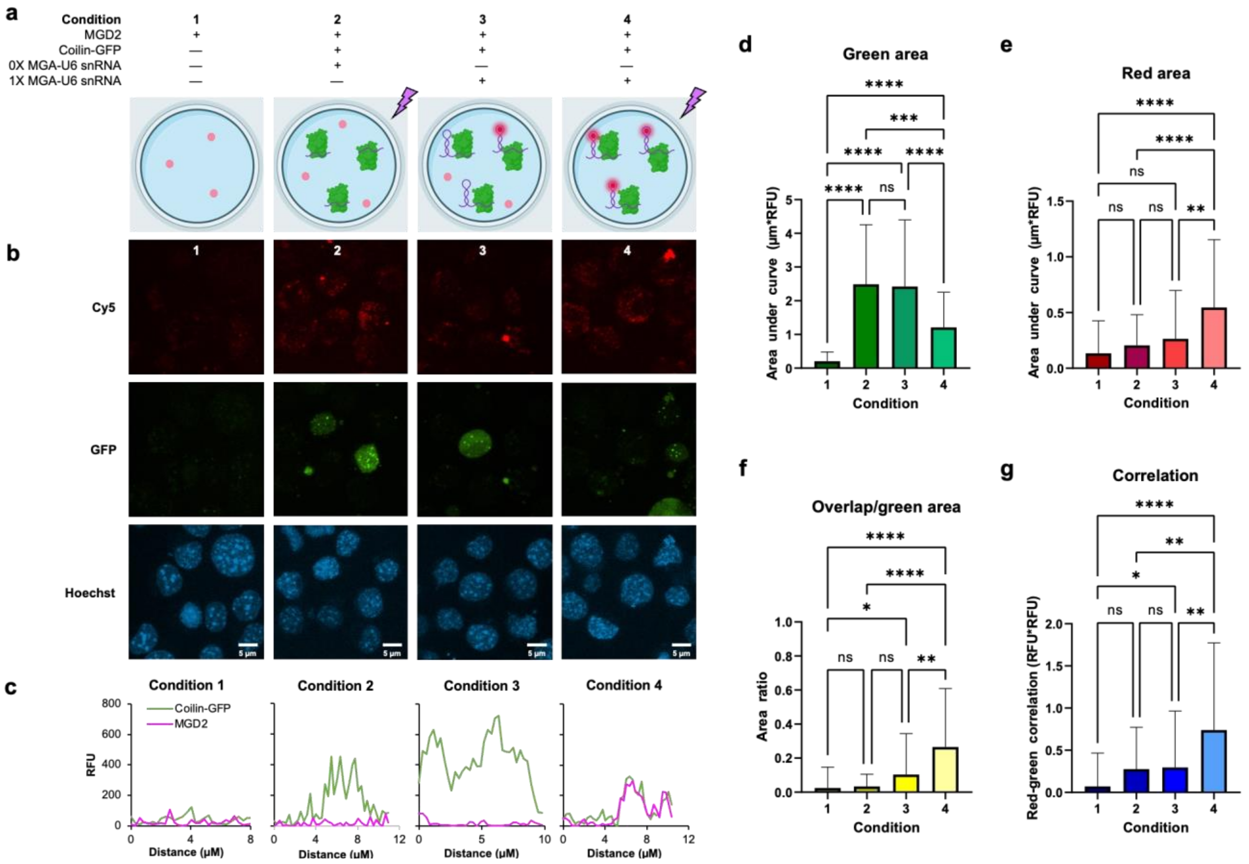


Figure 2.2: Colocalization of MGA-tagged U6 snRNA and coilin-GFP in live mammalian cells. **a** Schematic representation of U6 snRNA labeling with MGD2 in an MGA- and UV-dependent manner. MGD2 dye is expected to localize to the target RNA only when tagged with MGA, and UV irradiation covalently traps the dye-aptamer interaction to enhance RNA signal-to-noise ratio. Successful U6 snRNA labeling is visualized by co-expression with coilin-GFP to mark Cajal bodies, in which U6 snRNAs are processed. **b** Representative images of untransfected mouse Neuro-2a cells or N2a cells co-transfected with coilin-GFP and either 0X MGA- or 1X MGA-U6 snRNA. Cells were treated with 100 nM MGD2 and, in conditions 2 and 4, exposed to 10 min UV irradiation at 365 nm. MG fluorescence is monitored using the Cy5 channel. **c** Representative line scans drawn through cell nuclei using Fiji show that colocalization between MGD2 signal (red) and coilin-GFP signal (green) increases in an MGA- and UV-dependent manner (condition 4). **d–g** Global analysis of normalized MGD2 and coilin-GFP signals measured from 15–30 line scans in cell treatment conditions 1–4 over two biological replicates. For each condition, bar graphs compare **d** the mean area under the green curves, **e** the mean area under the red curves, **f** the mean area in which the red curves overlap the green curves, and **g** the correlation between the data points along the red and green curves. One-way ANOVA (95% confidence limit); post hoc test (Games-Howell), $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$, $****p \leq 0.0001$. Error bars indicate mean and standard deviation. Created with BioRender.com.

Across two biological replicates, the line scans for conditions 2–4 had a statistically higher area under the green curve than condition 1, confirming successful transfection of coilin-GFP (Figure

2.2d). Promisingly, the mean area under the red curves was significantly higher in condition 4 compared to conditions 1–3 lacking aptamer or photocrosslinking (Figure 2.2e). This suggests that the presence of the MGA sequence and subsequent irradiation with UV light covalently trap MGD2 on the target RNA, resulting in increased labeling efficiency. Most encouragingly, overlap between the MG and GFP signals was highest in condition 4 and showed a statistically significant difference from condition 3 that lacked UV photocrosslinking (Figure 2.2f). An analysis of correlation between the points on the two curves also revealed the highest association for photoaffinity labeling in condition 4, whereas non-covalent aptamer labeling (condition 3) was statistically indistinguishable from the no-aptamer control (condition 2) (Figure 2.2g). Together, these data demonstrate that we are able to label and image RNA in living cells using a single copy of the 38 nt aptamer and support the hypothesis that our photoaffinity labeling strategy provides superior robustness and capabilities in RNA imaging relative to a comparable non-covalent approach.

2.3.3 Real-time RNA tracking

Having established the applicability and robustness of our approach for live cell RNA imaging to study localization, we next investigated the RNA tracking capabilities of the MGA/MGD2 system to observe RNA dynamics. We hypothesized that by covalently linking MGD2 to MGA, we could improve both the continuity of RNA tracking throughout the acquisition time and the quality of the resulting traces. Distinct advantages relative to traditional non-covalent aptamer methods include a higher degree of confidence in tracking the same RNA throughout the entire image acquisition, as covalent labeling prevents migration of the dye between RNA molecules, as well as longer tracking periods before dissociation of the dye from the RNA. To assess the capabilities of our approach, N2a cells were co-transfected with expression plasmids for 1X MGA-U6 snRNA and coilin-GFP, treated with MGD2, and photocrosslinked via UV

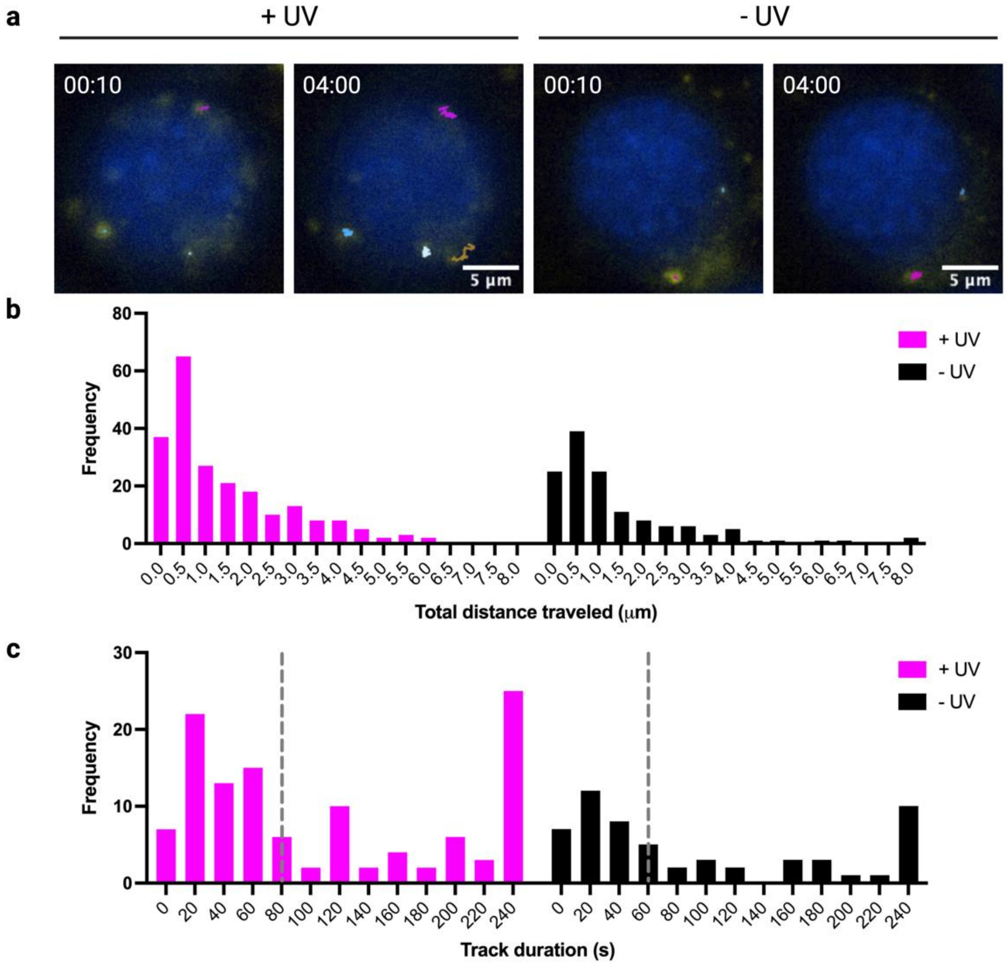


Figure 2.3: Live cell tracking of U6 snRNA foci. a Real-time widefield microscopy imaging of live Neuro-2a cells expressing 1X MGA-U6 snRNA and treated with MGD2 (100 nM) for both covalent (+ UV) and non-covalent systems (– UV) at time points 10 s and 240 s. Traces of nuclear foci were generated by tracking U6 snRNA. **b,c** Distribution plots generated from the analysis of the 1X MGA-U6 snRNA foci in 120 cells from **a** (N = 219 tracks, + UV and N = 134 tracks, – UV) with respect to their **b** total distance traveled and **c** track duration. Gray dashed lines represent the median of each data set. Data are representative of two independent experiments.

irradiation (covalent) or left in the dark (non-covalent) as previously described. Live-cell widefield imaging of both the covalent and non-covalent systems was completed over the course of four minutes to capture the spatiotemporal features of the U6 snRNA in nuclear Cajal bodies. Foci were readily detected and tracked using the TrackMate ImageJ plugin to ultimately create traces of the foci over the course of the time lapse acquisition.[119] Traces of foci are represented in varying colors shown at time points 10 seconds and 240 seconds (Figure 2.3a). Analysis of 120 cells across two replicate experiments from both conditions revealed that covalent attachment improved the total number of traces generated by 1.6-fold in comparison to the non-covalent approach. We then quantified and graphed the total distance traveled and the duration of each trace for both treatments. While the distance traveled by the 1X MGA-U6 snRNA in both covalent and non-covalent conditions was comparable, photocrosslinking yielded more traces for nearly every distance (Figure 2.3b). This trend was also observed when evaluating the track duration, such that covalent attachment enabled tracking of 2.5 times more foci throughout the entire four-minute image acquisition in comparison to the non-covalent system (Figure 2.3c). Further, the median track duration improved by 20 seconds with photocrosslinking, representing a significant increase in the amount of data that could be gathered regarding cellular RNA dynamics.

2.4 Conclusion

Fluorescent labeling and imaging of RNA is key to understanding its roles in cellular function and disease processes. While a number of methods have been reported for labeling RNA, these suffer from limitations related to either the requirement for cell fixation, the need for a large fusion added to the RNA, or the lack of robustness due to the transient nature of non-covalent interactions. We hypothesized that these challenges could be overcome through development of a covalent RNA labeling approach, akin to the SNAP-tag and Halo-tag systems that have proven to be powerful for protein labeling. Specifically, we envisioned a photoaffinity approach in which molecular recognition guides a fluorogenic small molecule to a target RNA and then brief irradiation with UV light promotes reaction between a photoreactive diazirine group and the RNA to result in covalent attachment of the fluorophore to the target RNA. We envisioned that the added robustness of the covalent linkage would provide increased signal-to-background ratio to enable imaging with a smaller RNA fusion and that the permanent attachment of the fluorophore to the RNA would generate novel capabilities for time-resolved RNA imaging applications.

To this end, we modified the malachite green fluorogen to incorporate a photoreactive diazirine linker, which allowed for covalent labeling of the cognate RNA aptamer upon irradiation with UV light. *In vitro* studies revealed that the MG derivative (MGD2) exhibited similar fluorescence enhancement properties to MG when incubated with MGA RNA. Covalent attachment proceeded selectively with MGA over an off-target aptamer and in the presence of total cellular RNA. As expected, labeling efficiency increased with probe concentration and UV irradiation time. This proof-of-principle covalent RNA aptamer tagging system achieved ~2% labeling efficiency by HPLC using the diazirine photocrosslinker. While this yield is lower than would be considered ideal, it is in line with that observed for other photoaffinity labeling methods. Moreover, as greater understanding of the relationships between photoreactive group identity, linker structure, and biomolecular targets emerges, we are optimistic that further tuning of photocrosslinker identity and spacing can increase labeling yield.

After establishing the feasibility of our approach *in vitro*, we moved to live cell studies. While cytotoxity deriving from MG and/or UV treatment represented potential concerns with this approach, we found that neither had an appreciable effect on N2a cell viability over the concentrations and treatment times used in this study. The amenability of the covalent approach to dye washout after labeling also serves to reduce undesirable off-target effects by limiting the concentration and amount of time that cells are exposed to the fluorophore. Using a fusion having a single copy of the aptamer inserted into the 5' UTR of the U6 snRNA, we demonstrated target-specific labeling and imaging of the ROI. Cells that were transfected with 1X MGA-U6 snRNA, treated with MGD2, and then UV irradiated showed the strongest overlap and correlation between MGD2 and signals for the Cajal body marker coilin-GFP, as compared to unirradiated cells or cells transfected with untagged U6 snRNA. Excitingly, we observed notable increases in overlap and correlation between U6 snRNA and coilin signals for our photoaffinity labeling approach compared to the analogous non-covalent aptamer approach, demonstrating the benefits of covalent dye attachment. Future efforts exploring installation of the aptamer in the 3' UTR of the target transcript could yield RNA species that are even less vulnerable to exonuclease degradation or labeled RNA populations that do not include potentially truncated transcripts.

Furthermore, we investigated application of our photocrosslinking approach to study cellular RNA dynamics, aiming to enhance the continuity and duration of RNA tracking throughout image acquisition. Live cell imaging of U6 snRNA in nuclear Cajal bodies revealed that covalent attachment of MGD2 to MGA improved the number of foci that could be tracked by 2.5-fold and prolonged the tracking duration over the entire acquisition compared to the traditional non-covalent system. Notably, our approach enhanced the performance of MGA in live mammalian cells, an application for this fluorescent aptamer that has otherwise remained limited due to cytotoxicity concerns at higher MG concentrations than those used here. Given its current photophysical properties, the potential of the malachite green fluorescent aptamer for single-molecule RNA imaging is limited. However, with improved affinity and fluorescence enhancement comparable to other fluorescent aptamers that have already demonstrated single-molecule RNA imaging, the malachite green fluorescent aptamer could emerge as a promising contender.[95, 120] Further, applying our covalent labeling approach to systems that have successfully imaged single RNA molecules (Broccoli, Mango II, Riboglow, RhoBAST, BiRhoBAST, and biSiRA) holds potential for further enhancing their robust imaging capabilities by providing more reliable and comprehensive tracking.[70, 94-95, 118, 121-122]

The novel strategy presented herein provides several advantages for RNA imaging applications. First, the far red-shifted fluorescence emission wavelength and the fluorogenic nature of the MGD2 dye minimize background signal generated from cellular autofluorescence and unbound small molecules, respectively. Second, the temporally controlled covalent attachment of the fluorogen to its cognate aptamer enables the labeling to withstand washout steps and allows tracking of RNAs labeled during a specific time window, a feature which is necessary for time-resolved studies such as pulse-chase experiments yet inaccessible using non-covalent aptamer labeling approaches. Third, imaging of the target RNA was achieved using a single aptamer fusion of 38 nt, an improvement over comparable methods that require multiple copies of the respective tag and/or fluorescent binding protein, producing a fusion that can add thousands of kDa to the transcript of interest. This limitation is particularly problematic for visualizing the many classes of RNA with transcripts <100–200 nt in length, including the U snRNAs. Finally, the covalent labeling strategy described here is anticipated to be highly generalizable, enabling the development of additional RNA aptamer-photoaffinity probe combinations for multiplexed and multicolor imaging. Likewise, incorporation of alternate photoreactive groups having distinct photophysical features from alkyl diazirine could serve to reduce UV irradiation time or harness alternate regions of the light spectrum for orthogonal labeling with multiple dyes. Covalent attachment additionally may allow for techniques to enrich specific transcripts from total cellular RNA as a method for studying RNA:protein interactions. Together, the research reported here advances RNA labeling technology by introducing a novel strategy that increases robustness and opens the door to new time-resolved applications. We expect that this will advance the study of basic mechanisms that underlie localization and dynamics of diverse types of RNAs and how these mechanisms go awry in disease and other important biological contexts.

2.5 Materials and Methods

2.5.1 Reagents

Chemicals and materials were purchased from Thermo Fisher Scientific unless otherwise stated. Malachite green (MG) was purchased from Chem-Impex International and used without further purification. Malachite green-diazirine 2 (MGD2) was synthesized as described in the Supplementary Information. Malachite green aptamer (MGA) RNA was generated via *in vitro* transcription of the dsDNA sequence 5’-*TAA TAC GAC TCA CTA TAG* GGA CAC AAT GGA CGG GAT CCC GAC TGG CGA GAG CCA GGT AAC GAA TGG ATC CTA ACG GCC GAC ATG AGA G-3’ ordered from Integrated DNA Technologies (IDT), with the T7 RNA Polymerase promoter sequence italicized and the MGA sequence underlined.[85] Briefly, ~1 µg of MGA dsDNA was combined with 10 µL of NTP buffer mix and 2 µL of T7 RNA polymerase (New England Biolabs, NEB) and filled to 30 µL final volume with nuclease-free water. The reaction was incubated at 37 ºC for 8 h and then incubated for an additional 15 min at 37 ºC with 10 µL of 10X DNase I reaction buffer, 2 µL of DNase I (NEB), and 60 µL of nuclease-free water. The RNA was purified using the Monarch RNA Cleanup Kit (50 µg, NEB) following the manufacturer protocol and eluting in 30 µL of nuclease-free water. The "off-target aptamer," PreQ1 riboswitch (5'-AGA GGU UCU AGC UAC ACC CUC UAU AAA AAA CUA A-3'), was purchased from IDT and used without further purification.[123]

2.5.2 Synthesis of MGD2

***N*-(4-(bis(4-(dimethylamino)phenyl)methyl)phenyl)-3-(3-methyl-3*H*-diazirin-3-yl)propenamide (2).** The synthesis of *p*-amino-leucomalachite green (2) was performed using the protocol described by Deng and coworkers.[124] In an oven dried 5 mL flask, *p*-amino-leucomalachite green (25.0 mg, 72.7 µmol) was dissolved in 1 mL of dry pyridine under inert gas. To this solution, 1.2 eq. of NHS-diazirine (19.6 mg, 86.83 µmol) was added and the solution was stirred at room temperature overnight. The reaction was then concentrated under reduced pressure to obtain dark green oil. The crude oil was dissolved in minimal amount of methanol and loaded on a preparative TLC with 30% EtOAc/MeOH as a mobile phase. The product band was scraped off from the preparative TLC and the product was filtered from the silica using MeOH and dried under reduced pressure. Yield = 11.6 mg, 35%. ^{1}H NMR (500 MHz, DMSO) δ 0.99 (s, 3H), 1.64 (t, 2H, *J* = 7.57 Hz), 2.20 (t, 2H, *J* = 7.3 Hz), 3.04 (s, 12H), 5.63 (s, 1H), 7.03 (d, 2H, *J* = 8.3 Hz), 7.24 (d, 4H, *J* = 7.3), 7.52-7.69 (m, 6H), 10.15 (s, 1H). ^{13}C NMR (500 MHz, DMSO) δ 19.36, 25.76, 29.53, 30.69, 44.52, 53.93, 119.21, 129.06, 130.14, 137.68, 169.75. LRMS (ESI-TOP) *m/z* Calcd for $C_{28}H_{33}N_5O$ $[M + H]^+$ 456.2758; Found 456.2746

***N*-(4-((4-(dimethylamino)phenyl)(4-(3-(3-methyl-3*H*-diazirin-3-yl)propanamido)phenyl) methylene)cyclohexa-2,5-dien-1-ylidene)-*N*-methylmethanaminium (3, "MGD2")**. 11.64 grams (25.55 µmol) of **2** was dissolved in 350 µL of 20% MeOH/EtOAc. To this solution, 1.2 eq. of chloranil (7.54 mg, 30.66 µmol) was added and the solution was stirred at room temperature for 3 h. The dark green solution was concentrated under reduced pressure and flash column purified using EtOAc to remove excess chloranil. Then mobile phase was then switched to 50% MeOH/DCM to isolate the crude product. The collected crude product was concentrated under reduced pressure and further purified by preparative TLC using 20% MeOH/DCM as mobile phase. Yield = 10.3 mg, 89%. ^{1}H NMR (500 MHz, DMSO) δ 1.03 (s, 3H), 1.69 (t, 2H, *J* = 7.1 Hz), 2.37 (t, 2H, *J* = 7.3 Hz), 3.14 (s, 6H), 3.27 (s, 6H), 7.07 (d, 4H, *J* = 8.8 Hz), 7.31 (t, 6H, *J* = 9.3 Hz), 7.94 (d, 2H, *J* = 8.3 Hz). ^{13}C NMR (500 MHz, DMSO) δ 19.35, 25.70, 28.93, 29.37, 40.39, 48.50, 113.63, 118.52, 126.24, 133.20, 136.25, 140.06, 156.23. LRMS (ESI-TOP) *m/z* Calcd. for $C_{28}H_{32}N_5O^+$ $[M]^+$ 454.2601; Found 454.2595

2.5.3 Absorbance/emission spectra

In triplicate, 20 µL of MGA RNA (50 µM) was combined with 20 µL of MG (300 µM in 8% DMF), MGD2 (300 µM in 8% DMF), or 8% DMF in water and brought up to 200 µL final volume with 20 µL of 10X PBS (1.37 M NaCl, 27 mM KCl, 80 mM Na_2HPO_4, 20 mM KH_2PO_4, pH 7.4) and 140 µL of nuclease-free water (0.8% DMF final). Reactions were incubated in the dark at rt for 20 min in an untreated 96-well, Clear, Flat-Bottom Polystyrene Microplate (Corning). Absorbance spectra were scanned from 300–700 nm with 2 nm step increments on a Cytation 5 Cell Imaging Multi-Mode Reader (BioTek) with the following instrument settings: Detection Method = Absorbance; Read Type = Spectral Scanning; Optics Type = Monochromators. Reactions were transferred to an untreated 96-well, Black, Flat-Bottom Polystyrene Assay Plate (Corning), and fluorescence spectra were scanned from 640 ± 10 nm to 700 nm with 2 nm step increments using Detection Method = Fluorescence Intensity and the same instrument settings otherwise.

Readings were normalized to the maximum absorbance or fluorescence intensity for each condition.

2.5.4 In vitro *fluorescence enhancement and UV dependence*

As above, MGA RNA or PreQ1 RNA (50 μM) was combined with MG or MGD2 (300 μM in 8% DMF) in 1X PBS to 20 μL final volume in triplicate. Dye-only negative control solutions were also prepared. After incubation, duplicate 2 μL fluorescence measurements were recorded and averaged for each sample on a Take3 Micro-Volume Plate (BioTek) using λ_{ex} = 624 ± 10 nm and λ_{em} = 662 ± 10 nm. Fluorescence enhancement for each replicate was calculated as the ratio of (RNA + dye) fluorescence over fluorescence of dye alone. For UV-dependent fluorescence studies, 120 μL solutions of MGA (5 μM) and MGD2 (30 μM) were incubated in 1X PBS for 20 min at rt in triplicate and divided in half. For 12 min, samples were incubated in the dark or UV irradiated at 365 nm from 12 cm in a 96-well, clear microplate under an ultraviolet, high intensity UVP lamp (Model B-100A, Analytik Jena). Aliquots were removed from each condition at 3 min intervals and the fluorescence measured as above. Values were normalized to the maximum fluorescence reading for the corresponding (± UV) 0 min timepoint.

2.5.5 Gel analysis of covalently labeled RNA

A 100 μL solution of MGA (100 nM) or PreQ1 aptamer (100 nM) was incubated with 50 μL of 10X PBS and 125 μL of 2.7% DMF or 160 nM–100 μM MGD2 in 2.7% DMF and brought up to 500 μL final volume with RNase-free water (0.7% DMF final). Reactions were incubated and UV irradiated for 15 min as above, and the RNA was ethanol precipitated via addition of 3 μL RNA-grade Glycogen, 0.1 volumes of 3 M sodium acetate (pH 5), and 2.5 volumes of ice-cold 100% ethanol. Pellets were washed with cold 70% ethanol and resuspended in 20 μL of 48% formamide/0.01% sodium dodecyl sulfate gel loading solution. RNA samples (~4 pmol) were run on 20% denaturing PAGE and imaged on an Amersham Typhoon^{TM} (Cytiva Life Sciences) in the

SYBR Gold and Cy5 channels. To probe the UV dependence of MGA labeling, MGA (20 nM) was incubated with MGD2 (25 µM) in 1X PBS, and the resulting solution was UV irradiated at 3 min intervals for up to 15 min. The RNA was purified by ethanol precipitation and analyzed by 20% denaturing PAGE. Relative band intensities were quantified by Fiji.

2.5.6 HPLC quantification of MGA labeling efficiency

Unpurified MGA (21 nmol, 4 µM final) was incubated with MG or MGD2 (100 µM final) in 1X PBS, followed by 15 min UV irradiation. The reactions were ethanol precipitated and further purified using the Monarch® RNA Cleanup Kit (New England BioLabs), eluting in 60 µL RNase-free water. MG- or MGD2-labeled RNA samples (9.75 nmol in 60 µL HPLC-grade water) were injected onto a 3.5 µm, 4.6 x 150 mm Eclipse Plus C18 Column (Agilent) and analyzed on a 1260 Infinity LC System (Agilent) using the following instrument settings: MWD1 (RNA) = 260 ± 4 nm (reference = 360 ± 20 nm); MWD2 (MGD2) = 615 ± 20 nm (no reference); flow rate = 1 mL/min. The samples were eluted over 5 minutes with isocratic 97:3 50 mM triethylammonium acetate buffer (pH 7.5):acetonitrile, followed by a 15-minute gradient mobile phase to 50:50 buffer:acetonitrile. Labeling efficiency was calculated according to Equation 1,[125] where ε_{260}(MGA) = 374,900 $M^{-1}cm^{-1}$ (calculated using IDT OligoAnalyzer Tool) and ε_{620}(MGD2) = 32,400 $M^{-1}cm^{-1}$ (data not shown).

$$\%\,\text{labeling} = \frac{A_{615} \cdot \varepsilon_{260}}{A_{260} \cdot \varepsilon_{615}} \qquad (1)$$

2.5.7 Determination of MGD2 selectivity in total cellular RNA

To determine the selectivity of MGD2 labeling, we spiked MGA (1, 5, or 25 pmol) into 5 µg of total cellular RNA that was extracted from 90% confluent N2a cells using the Monarch Total RNA Miniprep Kit (New England BioLabs). To this solution in 1X binding buffer (50 mM Tris-HCl, 100 mM KCl, 2 mM $MgCl_2$, pH 7.5), MGD2 (2.5 or 25 µM final concentration) was added and the solution was incubated for 20 min in the dark at room temperature. After 15 min of UV irradiation, the RNA samples were ethanol precipitated and 1 µg total RNA was analyzed by 12% denaturing PAGE as per the Materials and Methods section, “Gel analysis of covalently labeled RNA.” To visualize transcripts of higher molecular weight, the 2.5 µM MGD2-treated RNA samples were further purified using the Monarch® RNA Cleanup Kit (New England BioLabs), eluting in 30 µL RNase-free water. Total RNA (~400 ng) from each sample was analyzed on a 1% agarose-formaldehyde gel[126] containing 1X SYBR™ Safe DNA Gel Stain, along with 2 µL of ssRNA Ladder (“L1,” New England BioLabs), 2 µL of Low Range ssRNA Ladder (“L2,” New England BioLabs), and 0.5 or 5 nmol of free MGD2.

2.5.8 Cell culture and transfection

Neuro-2a cells (N2a, ATCC) were used for all cellular experiments. The cells were cultured in DMEM (high glucose, pyruvate, Gibco) supplemented with 10% FBS (Gibco) and 1% Penicillin-Streptomycin (Gibco) at 37 ºC in a humidified atmosphere of 5% CO_2 and 95% air. The cells were passaged once they reached 70% confluency. All cellular imaging was done on Cellview cell culture slides (Greiner Bio-One, cat. no. 543079). Transfections were performed using FuGENE HD Transfection Reagent (Promega) following the manufacturer recommended protocol and preparing transfection mixtures in Opti-MEM I (Gibco). The media from the cell culture wells was removed and replaced with 190 mL of fresh media, and 10 µL of the transfection mixture containing 100 ng of the corresponding plasmid(s) was added directly to each chamber containing 60–80% confluent cells. Transfection was allowed to proceed for 72 h with daily media exchange before conducting further experimentation.

2.5.9 Cytotoxicity

N2a cells were grown to 90% confluence in Poly-D-Lysine-Coated 96-Well Black/Clear Bottom Plates (Corning) and exposed to MGD2 (200 nM) or no dye in fresh media at 37 ºC for 15 min. Cells were either not irradiated or irradiated at 365 nm for 10 min. from 12 cm with an ultraviolet, high intensity UVP lamp (Model B-100A, Analytik Jena). The cells were then washed 3X with 100 µL media and incubated at 37ºC for up to 8 h with fresh media containing CellTox^{TM} Green Dye (diluted 1:500). Samples were prepared in quadruplicate, and duplicate positive control wells also contained Lysis Solution (diluted 1:25). Fluorescence measurements were recorded on a Cytation 5 Imaging Reader (BioTek) at λ_{ex} = 495 nm and λ_{em} = 525 nm. Error bars represent mean and standard deviation.

2.5.10 Confocal microscopy

Live images were taken using a Nikon A1R HD25 point scanning confocal microscope equipped with an Apo/DIC 60X/1.4 oil objective and two GaAsP PMT detectors. The cells were imaged at 37 ºC in a humidified atmosphere of 5% CO_2 and 95% air. Laser excitation at 405 nm was used to image Hoechst 33342; excitation at 488 nm was used to image coilin-GFP; and excitation at 640 nm was used to image MGD2. Cellular images were obtained by taking Z-stack images with instrument-optimized framerate as well as instrument-optimized step size, with enough steps to cover the depth of each cell. Gain for each channel was optimized to minimize oversaturation while obtaining clear, fluorescent foci signal above background. An 8 kHz resonant scanner was employed to reduce photobleaching and phototoxicity.

2.5.11 Live cell imaging

Cells were mock transfected or co-transfected with 100 ng of coilin-GFP and 100 ng of either 0X MGA- or 1X MGA-U6 snRNA plasmid. After 72 h transfection, the cell media was replaced with 200 µL of media containing 100 nM MGD2 and 2 µM Hoechst 33342, and the cells were incubated in the dark for 15 min at 37 ºC. Cell plates were then wrapped in tin foil and kept at rt for 10 min or irradiated with 365 nm light for 10 min at a distance of ~20 cm with the cell plate lid removed using an ultraviolet, high intensity UVP lamp (Model B-100A, Analytik Jena). All wells were then washed 3X with 200 µL of DMEM + 10% FBS + 1% Pen-Strep with 5 min incubations at 37 ºC. Phenol-free DMEM + 10% FBS + 1% Pen-Strep replaced the final wash prior to imaging. Two or three images were collected per well.

2.5.12 Live cell image analysis

A Z-stack maximum intensity projection of the Hoechst channel was generated in ImageJ image processing software and a line drawn through a representative nucleus. The line was transferred to the GFP and MGD2 channels, and background subtraction was performed by adjusting rolling ball radius to reduce the baseline fluorescence signals to zero. Next, ~50 lines were drawn through random regions of cell nuclei in the Hoechst channel and transferred to the GFP channel. A plot profile was generated for each line across the entire GFP Z-stack to quantify fluorescence signal. If a Z slice showed one or more clearly defined GFP peaks, indicating the presence of a Cajal body, the line was saved and transferred to the MGD2 channel to generate a plot profile for the MGD2 signal as well. Co-peaks indicated colocalization between coilin-GFP and 1XMGA-U6 snRNA. This process was repeated until 15–30 line scans showing coilin-GFP signal had been collected for each treatment condition (Figure A7).

2.5.13 Statistical analysis of GFP/MGD2 plots

For each treatment condition, the line scan data containing GFP and MGD2 signal versus distance was imported as a series of text files into a custom pipeline that is available for reference on GitHub at https://github.com/gchenfc/RNA_labeling_fluorogenic_photoaffinity/tree/main. First, to remove background noise in the data, a minimum threshold of 100 RFU was imposed for both curves. Next, to account for the differing brightness of the GFP and MGD2 fluorophores, data points on each curve were normalized to their respective maxima. The area under each curve was then calculated using the composite trapezoidal rule. A third curve representing the overlapping region between the GFP and MGD2 curves was defined by the minimum value between either curve at each point along the line scan. For each line scan, the area of the overlap curve was divided by the area of the GFP curve as a metric for the proportion of GFP signal in which MGD2 colocalization occurred, with values approaching unity representing 1:1 mapping. Finally, the similarity or correlation between the two curves was defined as the dot product between the normalized GFP and MGD2 fluorescence intensity values. Graphs represent combined means and standard deviations from 15–30 line scans for each of the four cell treatment conditions across two biological replicates. One-way ANOVA using the Brown-Forsythe and Welch ANOVA tests was performed in GraphPad Prism 9 to compare the means of GFP area, MGD2 area, overlap/GFP area ratio, and correlation across the treatment conditions. The Games-Howell test was used to correct for multiple comparisons, and P values were also adjusted accordingly. The family-wise alpha threshold and confidence level were set at 0.05 (95% confidence interval).

2.5.14 Live cell RNA tracking

Cells were seeded, transfected, and treated as previously described for live cell imaging. Time-lapse videos were taken using a Crest Optics X-Light V3 Spinning Disk Confocal System on a Nikon Ti2-E inverted microscope equipped with an Apo/DIC Lambda D 60X/1.4 oil immersion objective (1.5 zoom enabled, resulting in a magnification of 90X) and an ORCA-FUSION GEN-III SCMOS CAMERA C-MT. Cells were maintained at 37 ºC and 5% CO_2 using a Tokai Hit STXG Stage Top Incubator during image acquisition for widefield microscopy. Laser excitation at 405 nm was used to image Hoechst 33342; excitation at 488 nm was used to image coilin-GFP; and excitation at 640 nm was used to image MGD2. Gain and exposure for each channel were optimized accordingly to mitigate oversaturation while achieving sufficient fluorescence. Time-lapse acquisition was completed by imaging for a duration of four minutes, acquiring images every ten seconds to mitigate photobleaching, yielding a frame rate of 0.1 fps. Three time-lapse acquisitions were obtained per well.

2.5.15 Live cell RNA tracking analysis

Live-cell single particle tracking was performed using the TrackMate plugin in ImageJ.[119] The Difference of Gaussians (DoG) detector, known for its effectiveness in detecting small spot sizes, was utilized and a quality threshold was applied to capture the maximum number of particles while eliminating those originating from background. Additionally, a filter was set to include only spots with a minimum fluorescence intensity, and this threshold was determined using previously obtained data. For trace construction, the LAP tracker algorithm was employed. To construct traces, a max distance of 1.0 micron was used for frame-to-frame linking and 1.0 micron was used for gap-closing with a maximum frame gap of 2 frames. The number, total distance traveled, and duration of the resulting traces from both treatments (+/– UV) were analyzed and graphed using Prism.

2.5.16 Plasmid construction

pCMV-5U-Venus-PSD-95-3U (Addgene plasmid # 102949 ; http://n2t.net/addgene:102949 ; RRID:Addgene_102949) was PCR-amplified using 5'-ATATGAATTCGGTTTAGTGAACCGTCAGATCC-3' and 5'-ATATGATATCGGTGGATCTTCCGGATGG-3'. The PCR product was digested with *EcoRI* and *EcoRV* and ligated into similarly digested pcDNA3.1(+) (Invitrogen) to create pcDNA3.1-5U-Venus-PSD-95-3U. This plasmid was digested with *NheI* and *NcoI* and religated to create pcDNA3.1-5U-PSD-95-3U. pcDNA3.1-5U-PSD95-3U was then digested with *AgeI* and *XbaI*. U6 snRNA[127] was PCR-amplified from pUC-U6snRNA (custom ordered from Genewiz) using 5'-ATATATACCGGTGTGCTCGCTTCGGC-3' and 5'-ATATATTCTAGACAAAATATGGAACGCTTCACG-3'. The PCR product was digested with *AgeI* and *XbaI* and ligated into the digested pcDNA3.1-5U-PSD95-3U plasmid to create the original pcDNA3.1-V0_U6snRNA plasmid. pcDNA3.1-V0_U6snRNA was digested with *NheI* and *AgeI* to create the vector backbone. The malachite green aptamer[85] was inserted via the annealed oligo pair 5'-CTAGCGGATCCCGACTGGCGAGAGCCAGGTAACGAATGGATCCA-3' and 5'-CCGGTGGATCCATTCGTTACCTGG CTCTCGCCAGTCGGGATCCG-3' with compatible overhangs to create pcDNA3.1-1XMGA-U6snRNA. The final negative control plasmid omitting the aptamer was cloned by first completing a double digest on the pcDNA3.1-1XMGA-U6snRNA with NheI-HF and AgeI-HF to remove the 1X MGA sequence, followed by a blunting reaction with mung bean endonuclease, and finally ligation to yield pcDNA3.1-0XMGA-U6snRNA. pEGFP-Coilin was a gift from Greg Matera (Addgene plasmid # 36906 ; http://n2t.net/addgene:36906 ; RRID:Addgene_36906).

2.6 References

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

2. Pessa, H. K.; Will, C. L.; Meng, X.; Schneider, C.; Watkins, N. J.; Perala, N.; Nymark, M.; Turunen, J. J.; Luhrmann, R.; Frilander, M. J., Minor spliceosome components are predominantly localized in the nucleus. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (25), 8655-60.

3. Cabili, M. N.; Dunagin, M. C.; McClanahan, P. D.; Biaesch, A.; Padovan-Merhar, O.; Regev, A.; Rinn, J. L.; Raj, A., Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **2015,** *16*, 20.

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

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

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

7. Zhang, X.; Zhou, Y.; Chen, S.; Li, W.; Chen, W.; Gu, W., LncRNA MACC1-AS1 sponges multiple miRNAs and RNA-binding protein PTBP1. *Oncogenesis* **2019,** *8* (12), 73.

8. Daigle, N.; Ellenberg, J., λN-GFP: an RNA reporter system for live-cell imaging. *Nat. Methods* **2007,** *4* (8), 633-6.

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

10. Nelles, D. A.; Fang, M. Y.; O'Connell, M. R.; Xu, J. L.; Markmiller, S. J.; Doudna, J. A.; Yeo, G. W., Programmable RNA tracking in live cells with CRISPR/Cas9. *Cell* **2016,** *165* (2), 488-96.

11. Huang, N. C.; Luo, K. R.; Yu, T. S., Development of a split fluorescent protein-based RNA live-cell imaging system to visualize mRNA distribution in plants. *Plant Methods* **2022,** *18* (1), 15.

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

13. Tutucci, E.; Vera, M.; Biswas, J.; Garcia, J.; Parker, R.; Singer, R. H., An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat. Methods* **2018,** *15*, 81-89.

14. Grate, D.; Wilson, C., Laser-mediated, site-specific inactivation of RNA transcripts. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96*, 6131-6136.

15. Babendure, J. R.; Adams, S. R.; Tsien, R. Y., Aptamers switch on fluorescence of triphenylmethane dyes. *J. Am. Chem. Soc.* **2003,** *125* (48), 14716-14717.

16. Huang, H.; Suslov, N. B.; Li, N. S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A., A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* **2014,** *10* (8), 686-91.

17. Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferre-D'Amare, A. R., Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.* **2014,** *21* (8), 658-63.

18. Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R., Broccoli: rapid selection of an RNA

mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* **2014,** *136* (46), 16299-308.

19. Trachman III, R. J.; Demeshkina, N. A.; Lau, M. W. L.; Panchapakesan, S. S. S.; Jeng, S. C. Y.; Unrau, P. J.; Ferre-D'Amare, A. R., Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Nat. Chem. Biol.* **2017,** *13* (7), 807-813.

20. Strack, R. L.; Jaffrey, S. R., Live-cell imaging of mammalian RNAs with Spinach2. *Methods Enzymol.* **2015,** *550*, 129-46.

21. Rink, M. R.; Baptista, M. A. P.; Flomm, F. J.; Hennig, T.; Whisnant, A. W.; Wolf, N.; Seibel, J.; Dolken, L.; Bosse, J. B., Concatemeric Broccoli reduces mRNA stability and induces aggregates. *PLoS One* **2021,** *16* (8), e0244166.

22. Chen, X.; Zhang, D.; Su, N.; Bao, B.; Xie, X.; Zuo, F.; Yang, L.; Wang, H.; Jiang, L.; Lin, Q.; Fang, M.; Li, N.; Hua, X.; Chen, Z.; Bao, C.; Xu, J.; Du, W.; Zhang, L.; Zhao, Y.; Zhu, L.; Loscalzo, J.; Yang, Y., Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nat. Biotechnol.* **2019,** *37*, 1287-1293.

23. Cawte, A. D.; Unrau, P. J.; Rueda, D. S., Live cell imaging of single RNA molecules with fluorogenic Mango II arrays. *Nat. Commun.* **2020,** *11* (1), 1283.

24. Wirth, R.; Gao, P.; Nienhaus, G. U.; Sunbul, M.; Jaschke, A., SiRA: a silicon rhodamine-binding aptamer for live-cell super-resolution RNA imaging. *J. Am. Chem. Soc.* **2019,** *141* (18), 7562-7571.

25. Sunbul, M.; Lackner, J.; Martin, A.; Englert, D.; Hacene, B.; Grun, F.; Nienhaus, K.; Nienhaus, G. U.; Jaschke, A., Super-resolution RNA imaging using a rhodamine-binding aptamer with fast exchange kinetics. *Nat. Biotechnol.* **2021,** *39* (6), 686-690.

26. Bühler, B.; Schokolowski, J.; Benderoth, A.; Englert, D.; Grün, F.; Jäschke, A.; Sunbul, M., Avidity-based bright and photostable light-up aptamers for single-molecule mRNA imaging. *Nat. Chem. Biol.* **2023,** *19*, 478-487.

27. Englert, D.; Burger, E. M.; Grun, F.; Verma, M. S.; Lackner, J.; Lampe, M.; Buhler, B.;

Schokolowski, J.; Nienhaus, G. U.; Jaschke, A.; Sunbul, M., Fast-exchanging spirocyclic rhodamine probes for aptamer-based super-resolution RNA imaging. *Nat. Commun.* **2023,** *14* (1), 3879.

28. Arnould, B.; Quillin, A. L.; Heemstra, J. M., Tracking the message: applying single molecule localization microscopy to cellular RNA imaging. *Chembiochem* **2023,** *24* (10), e202300049.

29. Englert, D.; Matveeva, R.; Sunbul, M.; Wombacher, R.; Jaschke, A., Aptamer-based proximity labeling guides covalent RNA modification. *Chem. Commun.* **2021,** *57* (28), 3480-3483.

30. Babendure, J. R.; Adams, S. R.; Tsien, R. Y., Aptamers switch on fluorescence of triphenylmethane dyes. **2003,** *125* (48), 14716-14717.

31. Grate, D. A.; Wilson, C., Inducible regulation of the S. cerevisiae cell cycle mediated by an RNA aptamer–ligand complex. *Bioorg. Med. Chem.* **2001,** *9*, 2565-2750.

32. Kolpashchikov, D. M., Binary malachite green aptamer for fluorescent detection of nucleic acids. *J. Am. Chem. Soc.* **2005,** *127*, 12442-12443.

33. Afonin, K. A.; Danilov, E. O.; Novikova, I. V.; Leontis, N. B., TokenRNA: a new type of sequence-specific, label-free fluorescent biosensor for folded RNA molecules. *ChemBioChem* **2008,** *9* (12), 1902-5.

34. Scharf, N. T.; Molodtsov, V.; Kontos, A.; Murakami, K. S.; Garcia, G. A., Novel chemical scaffolds for inhibition of rifamycin-resistant RNA polymerase discovered from high-throughput screening. *SLAS Discov.* **2017,** *22* (3), 287-297.

35. Zhou, H.; Zhang, S., Recent development of fluorescent light-up RNA aptamers. *Crit. Rev. Anal. Chem.* **2022,** *52* (7), 1644-1661.

36. Zhang, X.; Potty, A. S.; Jackson, G. W.; Stepanov, V.; Tang, A.; Liu, Y.; Kourentzi, K.; Strych, U.; Fox, G. E.; Willson, R. C., Engineered 5S ribosomal RNAs displaying aptamers recognizing vascular endothelial growth factor and malachite green. *J. Mol. Recognit.* **2009,** *22* (2), 154-61.

37. Yerramilli, V. S.; Kim, K. H., Labeling RNAs in live cells using malachite green aptamer scaffolds as fluorescent probes. *ACS Synth. Biol.* **2018,** *7* (3), 758-766.

38. Kraus, G. A.; Jeon, I.; Nilsen-Hamilton, M.; Awad, A. M.; Banerjee, J.; Parvin, B., Fluorinated analogs of malachite green: synthesis and toxicity. *Molecules* **2008,** *13* (4), 986-994.

39. Lux, J.; Pena, E. J.; Bolze, F.; Heinlein, M.; Nicoud, J. F., Malachite green derivatives for two-photon RNA detection. *Chembiochem* **2012,** *13* (8), 1206-13.

40. Ayele, T. M.; Knutson, S. D.; Ellipilli, S.; Hwang, H.; Heemstra, J. M., Fluorogenic photoaffinity labeling of proteins in living cells. *Bioconjug. Chem.* **2019,** *30* (5), 1309-1313.

41. Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J. S., Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci.* **2016,** *15*, 14.

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

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

44. Nakamoto, K.; Ueno, Y., Diazirine-containing RNA photo-cross-linking probes for capturing microRNA targets. *J. Org. Chem.* **2014,** *79* (6), 2463-72.

45. Wang, J.; Schultz, P. G.; Johnson, K. A., Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc. Natl. Acad. Sci. U.S.A.* **2018,** *115* (20), E4604-E4612.

46. Suresh, B. M.; Li, W.; Zhang, P.; Wang, K. W.; Yildirim, I.; Parker, C. G.; Disney, M. D., A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117* (52), 33197-33203.

47. Balaratnam, S.; Rhodes, C.; Bume, D. D.; Connelly, C.; Lai, C. C.; Kelley, J. A.; Yazdani, K.; Homan, P. J.; Incarnato, D.; Numata, T.; Schneekloth, J. S., Jr., A chemical probe based on the PreQ(1) metabolite enables transcriptome-wide mapping of binding sites. *Nat. Commun.* **2021,** *12* (1), 5856.

48. Bae, J. W.; Kim, S.; Kim, V. N.; Kim, J. S., Photoactivatable ribonucleosides mark base-specific RNA-binding sites. *Nat. Commun.* **2021,** *12* (1), 6026.

49. Stanek, D., Cajal bodies and snRNPs - friends with benefits. *RNA Biol.* **2017,** *14* (6), 671-679.

50. Machyna, M.; Heyn, P.; Neugebauer, K. M., Cajal bodies: where form meets function. *Wiley Interdiscip. Rev. RNA* **2013,** *4* (1), 17-34.

51. Didychuk, A. L.; Butcher, S. E.; Brow, D. A., The life of U6 small nuclear RNA, from cradle to grave. *RNA* **2018,** *24*, 437-460.

52. Autour, A.; S, C. Y. J.; A, D. C.; Abdolahzadeh, A.; Galli, A.; Panchapakesan, S. S. S.; Rueda, D.; Ryckelynck, M.; Unrau, P. J., Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. *Nat. Commun.* **2018,** *9* (1), 656.

53. Braselmann, E.; Wierzba, A. J.; Polaski, J. T.; Chrominski, M.; Holmes, Z. E.; Hung, S. T.; Batan, D.; Wheeler, J. R.; Parker, R.; Jimenez, R.; Gryko, D.; Batey, R. T.; Palmer, A. E., A multicolor riboswitch-based platform for imaging of RNA in live mammalian cells. *Nat. Chem. Biol.* **2018,** *14* (10), 964-971.

54. Tinevez, J. Y.; Perry, N.; Schindelin, J.; Hoopes, G. M.; Reynolds, G. D.; Laplantine, E.; Bednarek, S. Y.; Shorte, S. L.; Eliceiri, K. W., TrackMate: An open and extensible platform for single-particle tracking. *Methods* **2017,** *115*, 80-90.

55. Trachman III, R. J.; Ferre-D'Amare, A. R., Tracking RNA with light: selection, structure, and design of fluorescence turn-on RNA aptamers. *Q. Rev. Biophys.* **2019,** *52*, e8.

56. Li, X.; Kim, H.; Litke, J. L.; Wu, J.; Jaffrey, S. R., Fluorophore-Promoted RNA Folding and Photostability Enables Imaging of Single Broccoli-Tagged mRNAs in Live Mammalian Cells. *Angew. Chem. Int. Ed. Engl.* **2020,** *59* (11), 4511-4518.

57. Braselmann, E.; Stasevich, T. J.; Lyon, K.; Batey, R. T.; Palmer, A. E., Detection and quantification of single mRNA dynamics with the Riboglow fluorescent RNA tag. *Preprint at bioRxiv* **2019**.

58. Roth, A.; Winkler, W. C.; Regulski, E. E.; Lee, B. W.; Lim, J.; Jona, I.; Barrick, J. E.; Ritwik, A.; Kim, J. N.; Welz, R.; Iwata-Reuyl, D.; Breaker, R. R., A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.* **2007,** *14* (4), 308-17.

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

60. Sharma, A. K.; Plant, J. J.; Rangel, A. E.; Meek, K. N.; Anamisis, A. J.; Hollien, J.; Heemstra, J. M., Fluorescent RNA labeling using self-alkylating ribozymes. *ACS Chem. Biol.* **2014,** *9* (8), 1680-4.

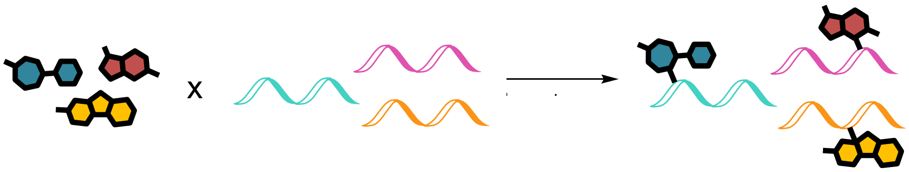
61. Rio, D. C., Denaturation and electrophoresis of RNA with formaldehyde. *Cold Spring Harb. Protoc.* **2015,** *2015* (2), 219-22.

62. Wolff, T.; Bindereif, A., Conformational changes of U6 RNA during the spliceosome cycle: an intramolecular helix is essential both for initiating the U4-U6 interaction and for the first step of splicing. *Genes & Dev.* **1993,** *7*, 1377-1389.

Chapter 3: Profiling the Ligandability of the Human mRNAome with Small-Molecule PAL Probes

3.1 Abstract

Detailed annotation of the human genome within the last decade has revealed the previously overlooked potential of RNA as a therapeutic target. Notwithstanding messenger RNAs (mRNAs) upstream of proteins categorized as undruggable, many additional mRNAs, long noncoding RNAs, microRNAs, and newly emerging RNA classes are now implicated in human disease. Small molecules represent convenient tools to modulate RNA function and activity, given their great success in treating protein-based pathologies. However, the dearth of RNA-targeted small molecule probes that have passed clinical trials highlights the need to develop guiding principles for rationally designing novel, bioactive compounds. Here, we develop a photoaffinity labeling (PAL) method to probe RNA:small molecule target engagement *in vitro* and in living cells. This covalent strategy is necessary to explore interactions that may otherwise remain undetected due to their transient and/or low-affinity nature. We validate the method using the PreQ1 riboswitch and its cognate ligand as a model system. Subsequent efforts focus on analyzing the differential labeling of mRNAs from MCF-7 breast cancer cells by a panel of 10 structurally diverse PAL ligands that include RNA- and protein-privileged scaffolds, metabolites, and FDA-approved drugs. Finally, we extensively characterize the RNA hits of the probe derived from anticancer drug palbociclib, as well as the reproducibility of these hits across *in vitro* and *in cellulo* experiments. This work provides a more robust approach toward the characterization of RNA target engagement and lends insight into the development of RNA-targeted drugs.



3.2 Introduction

RNA canonically serves the role of cellular messenger, transferring information encoded within genetic material into proteins that carry out molecular tasks. However, more recent exploration of the human genome through the Encyclopedia of DNA Elements (ENCODE) Project revealed that while over 60% of DNA is transcribed into RNA, less than 5% of that DNA codes for protein.[128] The RNA Revolution has witnessed the identification of thousands of non-coding RNAs such as microRNAs (miRNAs), small nuclear RNAs (snRNAs), and long noncoding RNAs (lncRNAs) that execute critical cellular functions.[2, 10] Meanwhile, an estimated 85–90% of proteins are deemed “undruggable” due to the absence of discrete binding pockets,[15] and the upstream messenger RNAs (mRNAs) therefore offer a second opportunity for therapeutic intervention.[7] These transcripts collectively represent valuable drug targets, as aberrant RNA activity is implicated in many disease pathologies.[129]

Substantial efforts have focused on modulating RNA function with chemical biology tools such as small molecule probes, antisense oligonucleotides (ASOs), and CRISPR-based editing technologies.[130] Small molecules are attractive candidates due to their desirable pharmacokinetic properties and widespread use as exogenous regulators of protein function.[43] Meanwhile, ASOs have met with some success in clinical trials—for example, in the treatment of GGGGCC hexanucleotide repeat-containing mRNAs in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)[131]—but are significantly limited in their oral bioavailability and blood-brain barrier (BBB) permeability. Likewise, due to inherent issues with immunogenicity, delivery, and off-target activity,[132] strategies to target pathogenic RNAs through CRISPR in a widespread fashion are likely years out from practical realization.

Several online resources such as the Nucleic Acid Ligand Database (NALDB),[133] the Small Molecule Modulators of RNA (SMMRNA) database,[134] the Inforna RNA database,[135] and the RNA-targeted Bioactive LigaNd Database (R-BIND)[45] highlight successful examples of bioactive, RNA-binding small molecules. In particular, the curation of R-BIND was prompted through a wider effort to explore the physicochemical and spatial properties of small molecules that promote interaction with RNAs. Cheminformatic analysis determined that, compared to protein-targeted probes, bioactive RNA-targeted probes are enriched in nitrogen and aromatic rings with reduced oxygen content.[6] Such studies facilitate efforts in small molecule-based RNA targeting by establishing guiding principles that may be exploited in the design and development of "RNA-privileged" screening libraries and hit compounds.

Despite these advances in the scientific community's understanding of RNA:small molecule targeting, as well as the identification of numerous bioactive chemical probes, few drugs are FDA-approved for targeting RNA. Namely, the linezolid antibiotics were developed to target the bacterial ribosome and, recently, Risdiplam was identified as a modulator of human SMN2 mRNA splicing.[7, 14, 42] Complementary to the above question of whether certain small-molecule features promote interaction with RNA, it is possible that certain RNA features promote interaction with small molecules. Weeks and coworkers argue that researchers should not only identify disease-related RNAs and screen appropriate small molecule libraries for activity against them, but also place greater emphasis on ensuring that the structures of the selected RNA targets themselves are likely to yield hits.[7] In particular, information-rich and structurally complex RNAs with deep binding pockets reminiscent of those found in proteins may be more ligandable than simple RNA structures. Indeed, a computational analysis of liganded RNA and protein pockets revealed substantial overlap in properties such as buriedness, hydrophobicity, and pocket volume, suggesting that RNA pockets may be as amenable to the binding of synthetic probes as protein targets.[48] Further exploration of the sequence and structural dependence of RNA on ligandability will expedite efforts in the development of RNA-targeted therapeutics.

One attractive strategy to probe RNA target engagement is inspired by a report from Parker et al. which applied photoaffinity labeling (PAL) with a panel of diazirine-containing small-molecule fragments to profile low-affinity interactions with the human proteome.[136] In general, photocrosslinking functional groups installed within chemical probes are a powerful means to irreversibly and temporally label proximal biomolecules. Upon UV irradiation, photoreactive crosslinking groups such as diazirine are excited to a high-energy state, and during relaxation, covalently insert themselves into neighboring C–H and heteroatom–H bonds as can be found within RNA.[137] The diazirine moiety in particular has the advantage of small size and thus lower likelihood to perturb native RNA:small molecule interactions that exist in cells. Through additional decoration with a minimal alkyne handle, fragments derivatives can be used for enrichment, sequencing, and bioinformatic analysis of target RNAs.

It is well established that photoactivatable diazirines can form stable crosslinks not only with protein but with RNA bases.[111] By appending this reactive group to a small-molecule RNA binder, Disney and coworkers were able to profile individual RNA:small molecule interactions in cells.[112] Additionally, work from Mukherjee et al. showed that PAL can be applied to profile transcriptome-wide.[138] In a related approach, Tong et al. applied chemical cross-linking and isolation by pull-down to map small molecule:RNA binding sites (Chem-CLIP-Map-Seq) using a panel of small-molecule fragments, one of which was shown to have functional consequences in cells based on RNA target engagement.[63, 139]

Here we build on this foundation by applying a modified strategy to profile mRNA:small molecule interactions *in vitro* and in live cells. We focus on mRNA as a target toward ultimately controlling/disrupting downstream protein translation.[7] To assess the generalizability of this approach, we synthesized a diverse panel of 10 PAL compounds that includes reported RNA- and protein-binding fragments, cellular metabolites, and FDA-approved drugs. Chem-CLIP-Map-Seq was performed *in vitro* by treating mRNA derived from MCF-7 breast cancer cells with the compounds, applying UV irradiation to induce photocrosslinking, pulling down labeled transcripts, and sequencing to assess differential probe engagement (Fig 1A).[139] More focused application of competitive Chem-CLIP-Map-Seq (Comp-Chem-CLIP-Map-Seq) with compound **9**, a derivative of an FDA-approved drug for breast cancer called palbociclib, revealed that many of these small-molecule interactions are specific and reproducible.[140-141] To assess whether the *in vitro* interactions between compound **9** and the observed RNA targets are preserved in a cellular context and have biological relevance, we also repeated the Chem-CLIP-Map-Seq experiments for compound **9** in live cells. We found hits consistent across both experiments, revealing potential RNA targets of the drug. Validation of these hits will establish this photocrosslinking method as a powerful strategy for researchers to develop therapeutically active, RNA-targeted small molecules.

3.3 Results and Discussion

3.3.1 Non-covalent PreQ1 RNA capture

The initial approach we designed to assess the binding preferences of cellular RNAs for drug fragments was to immobilize the fragments on beads, incubate with total cellular RNA, and characterize enrichment of the non-covalently bound RNAs (Figure 3.1a). To validate this strategy, we first performed proof-of-principle experiments using a known RNA:ligand pair, the bacterial prequeuosine 1 (PreQ1) riboswitch, whose interaction with PreQ1 is well-characterized (Figure 3.1b).[123, 142]

Unlike mammals, bacteria employ special functional RNA elements known as riboswitches to sense the intracellular concentrations of critical metabolites.[38, 40] For example, PreQ1 is a precursor of the hypermodified guanine nucleotide queuosine (Q), which is inserted into the wobble position of several bacterial and eukaryotic tRNAs to regulate translational fidelity.[142] Whereas eukaryotes cannot synthesize PreQ1 and must obtain it from their diets, bacteria regulate the expression of PreQ1 biosynthesis genes at the transcriptional level with the installation of the Class I PreQ1 aptamer at the 5’ untranslated region (UTR) of associated mRNAs. Specifically, PreQ1 binding induces a conformational change in the riboswitch that inhibits downstream transcription (Figure 3.1b). This interaction is well studied with crystal structures reported.[143] To leverage this interaction for RNA pulldown applications, we prepared a PreQ1 derivative, (polyethylene glycol)$_2$-PreQ1 (PEG$_2$-PreQ1, Figure 3.1c). Fluorescence

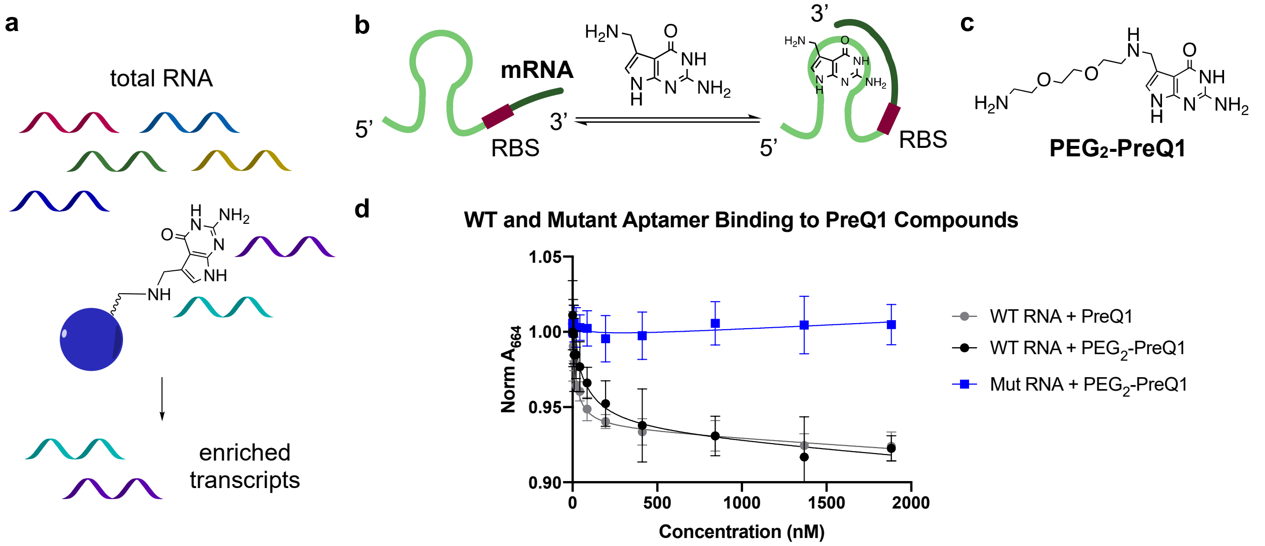


Figure 3.1: Proof-of-principle PreQ1-mediated RNA enrichment. a Scheme to isolate fragment-binding RNA transcripts. **b** PreQ1 riboswitch-mediated translational downregulation of PreQ1 biosynthesis. **c** PEG_2-PreQ1 derivative for bead coupling. **d** Fluorescence titration of PreQ1 and derivative binding to the WT aptamer. Mut = mutant; RBS = ribosomal binding site; WT = wild-type.

titration studies indicated that PEG_2-PreQ1 bound to the 5'-Cy5-wild-type (WT) riboswitch with comparable affinity to the native ligand, whereas PEG_2-PreQ1 binding to a riboswitch double mutant rationally designed on the basis of the crystal structure was not observed (Figure 3.1d).[143]

Incorporation of this primary amine linker onto PreQ1 enabled conjugation of the ligand to NHS ester-functionalized beads. However, negligible WT riboswitch RNA in a 4 µM solution was recovered from the beads compared to unfunctionalized control beads (Figure B1a). These results indicated either that the probe had not successfully attached to the beads or that, despite attachment, the RNA could not interact favorably with the probe, potentially due to steric interference. To investigate these alternatives, we optimized a test coupling reaction with PEG_2-PreQ1 and a biotin NHS ester (Figure B1b). The reaction proceeded in >95% yield by high-performance liquid chromatography-mass spectrometry (HPLC-MS, Figure B1c), thereby establishing the reactivity of PEG_2-PreQ1. We therefore hypothesized that the longer linker on the biotin-PEG_6-PreQ1 (BP_6P) probe would further increase the accessibility of the PreQ1 moiety for interaction with the aptamer *in vitro*. Biotin also enabled pulldown experiments to be performed on streptavidin beads, as opposed to NHS ester beads, as an orthogonal testing system.

Concerningly, significant enrichment of the WT versus the mutant RNA was not observed on the streptavidin beads whether the probe was preincubated with the RNA or with the beads before RNA pulldown (Figure B1d,e). Nor was nonspecific mutant RNA binding reduced in the presence of Tween or excess sodium chloride.

We further characterized the interaction of the WT aptamer with both PreQ1 derivatives by microscale thermophoresis (MST) to better understand the basis for the lack of enrichment. RNA (1 nM) was incubated with two-fold serially diluted probe from 0.002–55 μM concentration in pH 7 buffer containing 50 mM Tris, 100 mM KCl, and 1 mM $MgCl_2$. Substantially ablated RNA binding was observed for BP_6P compared to PEG_2-PreQ1, demonstrating the intolerance of the riboswitch for significant ligand modification (Figure B2). Given the exquisite molecular discrimination that riboswitches must exhibit to differentiate structurally similar cellular analogs, these results were not unsurprising, but they led us to conclude that non-covalent RNA:ligand binding interactions may be too weak to enrich target RNAs on fragment-immobilized beads. Therefore, we sought a more robust approach to isolate cellular RNAs interacting with a small molecule probe.

3.3.2 Validation of PreQ1 RNA photoaffinity labeling

Inspired by recently published work from Schultz and coworkers which utilized a diazirine-modified small molecule to identify cellular RNA targets of the probe,[61] as well as recent success in our own lab with diazirine-based cellular RNA imaging,[144] we synthesized several diazirine-modified PreQ1 derivatives which could be used for selective labeling of the cognate aptamer (Figure 3.2a). Although these modifications could potentially perturb RNA:ligand binding in a similar manner to BP_6P, specific interactions occurring even transiently could be captured irreversibly via covalent probe attachment into the RNA. Preliminary data showed the *in vitro* formation of RNA:probe photocrosslinking adducts by MALDI-TOF (Figure 3.2b) and denaturing polyacrylamide gel electrophoresis (PAGE, Figure 3.2c), with up to an estimated 50% labeling

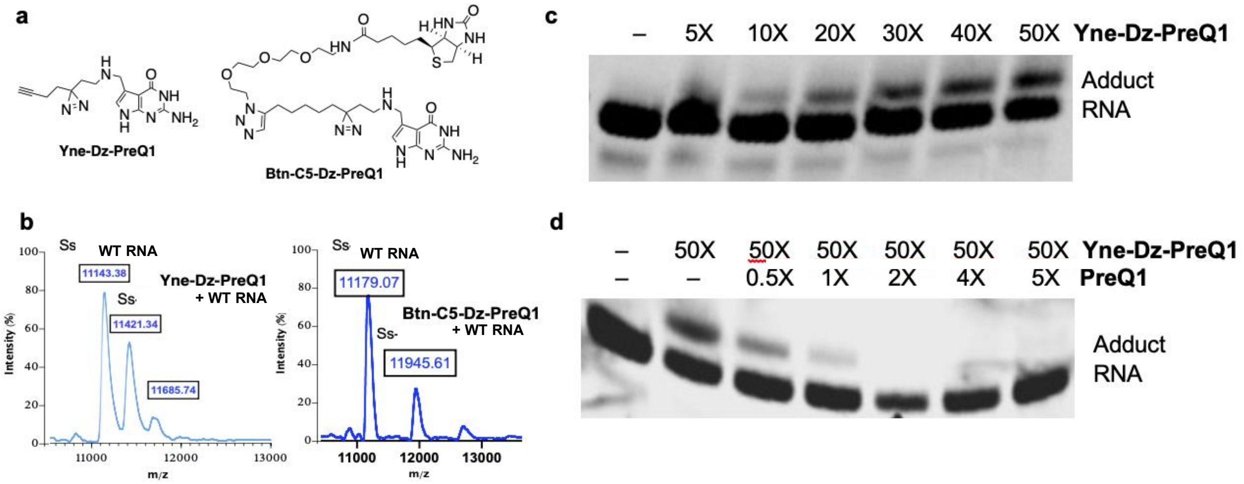


Figure 3.2: Preliminary studies on PreQ1 aptamer photocrosslinking. **a** Structures of two photocrosslinking PreQ1 probes. **b,c** Formation of covalent RNA:probe adducts upon UV irradiation by **b** matrix-assisted laser desorption/ ionization–time-of-flight mass spectrometry (MALDI-TOF) and **c** denaturing PAGE. **d** Ablated RNA crosslinking in the presence of unlabeled PreQ1 competitor.

efficiency by densitometry analysis. Further, covalent labeling was diminished in the presence of PreQ1 competitor (Figure 3.2d), indicating a specific interaction. Notably, several of the bands on the PAGE gels appeared oversaturated, which could lead to an artificially high estimate of labeling efficiency.

To validate these results, we evaluated the photocrosslinking and RNA pulldown efficiency of the biotinylated probe, which could be isolated and analyzed directly on streptavidin beads. Briefly, 5’-Cy5-labeled WT or mutant PreQ1 riboswitch (10 µM) was annealed and then incubated with 50-fold excess probe for 30 minutes in the dark before 20 minutes of UV irradiation under 360 nm light. Finally, labeled RNAs were enriched on streptavidin beads (Figure B3a**)**. We observed moderate recovery of the WT RNA compared to a no-probe control, but the mutant RNA was also recovered to a comparable extent, indicating nonspecific crosslinking (Figure B3b). We attributed these results to two possible factors: i) unfavorable RNA:probe interaction due to the relatively large size of the biotin probe and ii) background PreQ1 binding by the mutant aptamer.

We addressed these concerns by exploring the reactivity of the more minimal yne-dz-PreQ1 probe (Figure 3.2a) and by replacing our mutant aptamer negative control with the Sadenosyl-methionine (SAM)-II aptamer.[145] We then designed a novel, fluorescence-based assay to quantify the presence of alkynyl-modified RNAs after UV labeling with the probe. After performing photocrosslinking as described above, RNAs are ethanol precipitated to remove unreacted probe and then further functionalized with an azide-conjugated dye through copper-assisted azide-alkyne cycloaddition (CuAAC). The appearance of the dye signal, as well as a shift in oligonucleotide size, can be visualized by PAGE to identify labeled RNAs (Figure 3.3a). Validation of this experimental setup was achieved using an alkyne-modified, FAM-labeled DNA control strand (Figure 3.3b).

Using the protocol described above and an Alexa Fluor 488 (AF488)-azide, we observed less than 5% crosslinking efficiency between the Cy5-WT PreQ1 aptamer and the minimal yne-dz-PreQ1 probe (Figure 3.3c). This value was much lower than the prior results (Figure 3.2c). Furthermore, we observed comparable labeling between the Cy5-PreQ1 RNA and the Cy5-SAM-

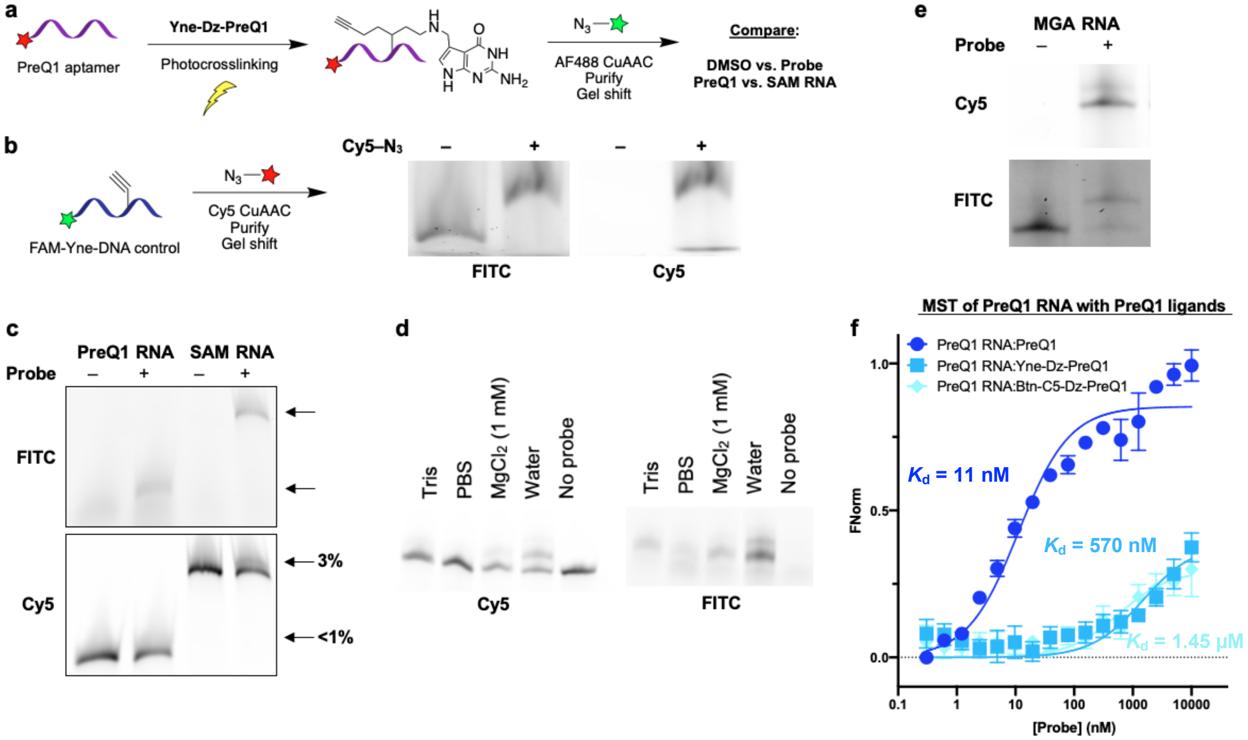


Figure 3.3: PreQ1 RNA labeling with a minimal alkyne PAL. a CuAAC-mediated PreQ1 aptamer labeling scheme. **b** Labeling validation with Cy5-azide and an alkynyl DNA control strand. **c** Minimal Cy5-PreQ1 and Cy5-SAM-II riboswitch labeling (arrows) by the probe in Tris buffer. **d** Increased Cy5-PreQ1 riboswitch labeling in reduced salt content buffers. **e** Off-target probe labeling of the malachite green aptamer (MGA), visualized by Cy5-N_3 CuAAC. **f** PreQ1 RNA binding to the native ligand and two PAL derivatives.

II aptamer negative control RNA (Figure 3.3c). Neither specificity nor labeling efficiency were improved with reduced DMSO content. Further, crosslinking increased in buffers with reduced salt content, reflecting nonspecific reactivity with unfolded RNA (Figure 3.3d). The probe also showed significant background reactivity with an additional off-target RNA, the malachite green aptamer (Figure 3.3e).[85] It is likely that the presence of the diazirine motif on the PreQ1 molecule inhibits RNA:ligand binding more severely than does the PEG chain on PEG$_2$-PreQ1 (Figure 3.1d). This conclusion is supported by MST binding studies, in which the WT aptamer binds much more tightly to its native ligand than to either diazirinyl derivative (Figure 3.3f).

These results raised serious concerns that the PreQ1 RNA:ligand pair does not bind robustly when perturbed from its native state. More generally, the experiments indicate that fragments modified with the diazirine-alkyne handle may not interact with the same cellular substrates as would the fragments alone. However, we also explored the crosslinking efficiency of the yne-dz-PreQ1 probe with a larger 70 nt PreQ1 aptamer construct that contained within it the 34 nt sequence used in the above studies. These experiments were performed in the presence of total RNA to examine the specificity of the probe for its target RNA. Promisingly, the studies demonstrated more convincing evidence of specific aptamer labeling, with roughly 40% labeling efficiency for 1 µM aptamer doped into 5 µg MCF-7 total RNA in the presence of 250 µM probe (Figure 3.4). Further, competition by native PreQ1 again suggests that this interaction is specific (lanes 4–6), and importantly, aptamer labeling is dependent on UV irradiation (lane 7). The findings suggest

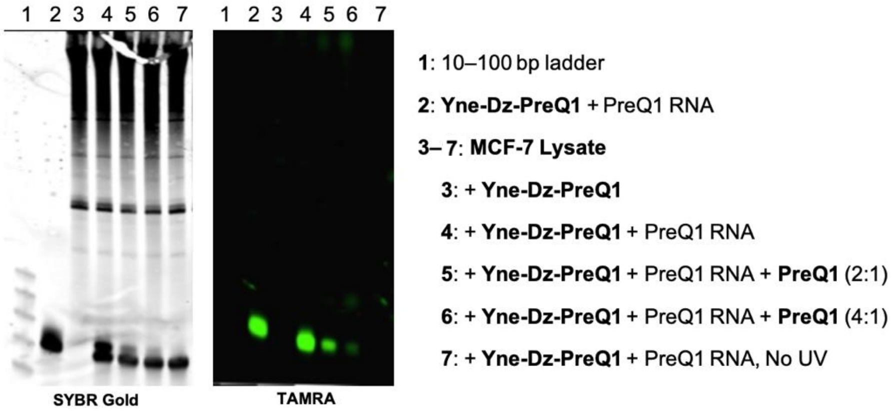


Figure 3.4: Selective, UV-dependent PreQ1 RNA labeling in total RNA. The 70 nt aptamer (1 µM) was labeled with yne-dz-PreQ1 (250 µM) in the presence of 5 µg MCF-7 total cellular RNA. Labeling was visualized by CuAAC with TAMRA-N$_3$.

that the larger PreQ1 aptamer construct binds PreQ1 and the alkyne-dz-PreQ1 derivative more robustly for specific labeling in the presence of cellular RNA. With proof-of-principle labeling successfully achieved for a known RNA:small molecule pair, subsequent efforts could focus on expanding this approach to a library of diazirine-functionalized fragments.

3.3.3 Preparation of a 10-compound PAL panel

The above experiments exploring proof-of-principle PreQ1 RNA photocrosslinking with a diazirine-modified cognate ligand highlight the sensitivity of aptamer binding to ligand derivatization. These findings raise the concern that attempts to derivatize fragments of interest with a diazirine moiety could disrupt contacts that the fragment would otherwise make with cellular RNAs. Nevertheless, the demonstration that a PreQ1 aptamer construct is selectively labeled in the presence of off-target RNAs encourages additional investigation of this fragment-based approach to study RNA:small molecule recognition. We were also optimistic that the incorporation of UV-dependent crosslinking functionality would impart novel advantages over non-covalent or electrophilic probes in the quest to elucidate critical principles governing the design of RNA-targeted probes.

Based on recent work that explored fragment-based screening of cellular protein binding events, we designed an analogous shotgun approach for cellular RNA labeling. A panel of 10 small-molecule PAL probes was prepared by appending each compound to a linker that contained both a diazirine photoreactive group for covalent labeling of interacting RNAs and an alkyne handle for downstream RNA pulldown and sequencing (Figure 3.5a). Compound **1** was prepared to serve as a binding control, as it contains minimal structure for selective and high-affinity RNA interaction. Compounds **2** and **3** are from the R-BIND database of RNA-interacting small molecules and occupy “RNA-privileged chemical space” on the basis of 20 cheminformatic parameters.[6] Compound **4** was chosen due to its benzodiazepine core, as benzodiazepine is a known protein-interacting scaffold.[146-147] Compound **5** was selected due to the arrangement of the

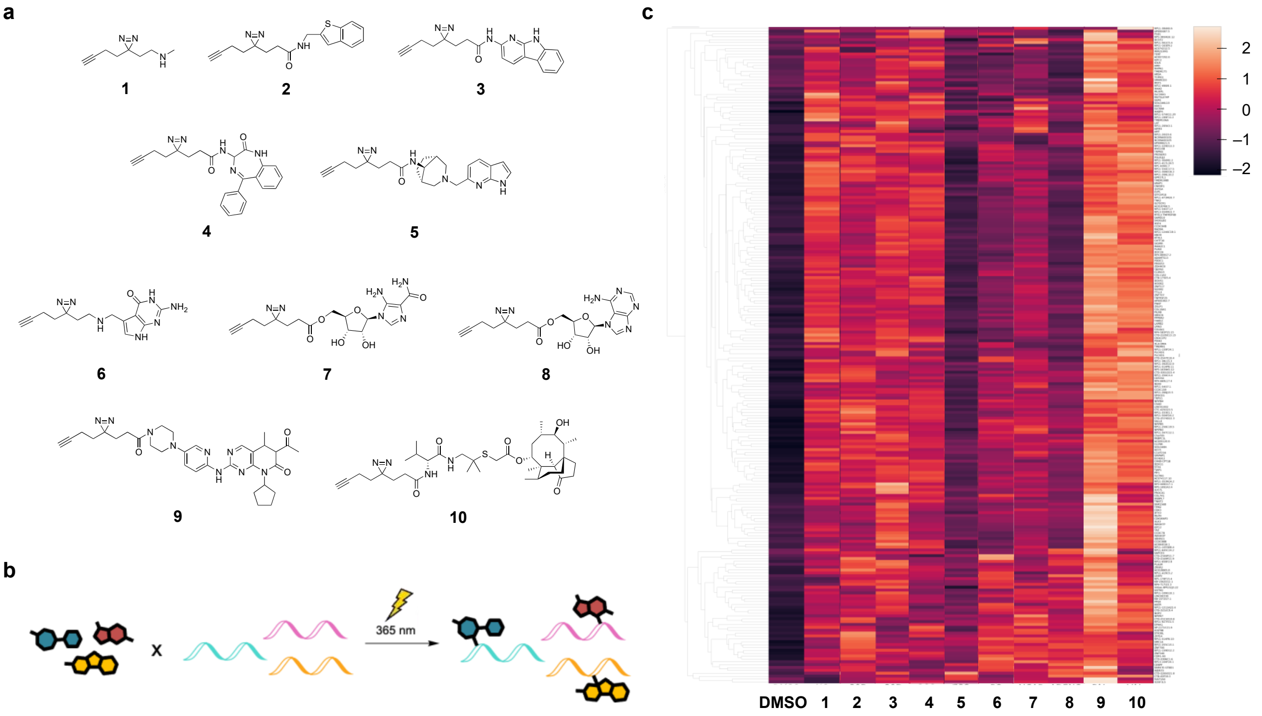


Figure 3.5: Analysis of *in vitro* RNA target engagement using PAL. a Structures of 10 photocrosslinking probes representing negative control (**1**), RNA-privileged (**2**–**3**), protein-privileged (**4**–**5**), metabolite-like (**6**–**8**), and FDA-approved (**9**–**10**) compounds. **b** General scheme to assess probe labeling of poly-A-enriched mRNA from MCF-7 breast cancer cells. **c** Heatmap showing compound-specific interactions with the enriched and sequenced mRNA pools. The heatmap includes select mRNAs that were statistically significantly enriched versus the DMSO control sample (FC > 2, FDR < 0.01) for at least one of the test compounds. The scale on the top right shows the fold-change values for the mRNA levels in each sample, with 0 representing the average TPM-normalized sequencing counts across the pooled data.

hydrogen-bonding groups in the 7-aza-indole scaffold, which is another motif that is commonly featured in protein-binding small molecules.[148] These RNA- and protein-binding fragments were selected to investigate whether this technology would be useful in fragment-based screens or scaffold hopping applications. Notably, while simple chemical fragments are less likely to bind targets with high affinity, lead compounds can be derivatized into potent bioactives.[136]

Three compounds were selected to investigate potential metabolite interactions in the cell. Compound **6** is recycled from the PreQ1 validation experiments above (**Fig. 8a**). Compound **7** is also based on a riboswitch binding ligand, 5′-aminoimidazole-4-carboxamide (AICA). AICA binds to the aptamer of the ZTP riboswitch, albeit with lower affinity than the cognate ligand, ZTP.[149-150] Compound **8** is based on adenosine. Functional adenine riboswitches have also been annotated in bacteria.[151] We envisioned that by investigating the RNA targets of these metabolite-like aptamer-binding ligands, we could potentially discover aptamers-like RNAs in the human mRNAome. In addition, this set of compounds enabled us to investigate the role of the nucleobase sugar in these binding events. It is also worth highlighting the structural similarity of compounds **7** and **8**, which enabled us to assess how highly similar compounds perform.

Compounds **9** and **10** are derived from two FDA-approved drugs, palbociclib and valnemulin. Palbociclib is a CDK4/6 inhibitor that is clinically approved for HR-positive and HER2-negative breast cancers.[140] Valnemulin is a pleuromutilin antibiotic used to treat bacterial infections in pigs.[152] These compounds were both selected for this study due to previous findings suggesting that they may bind to structured RNAs with high affinity. In particular, palbociclib, the core structure of which is a Parke-Davis RNA-binding motif,[153] has recently been shown to bind to RNA targets, including the HIV-1 TAR hairpin.[154-155] Interestingly, both palbociclib and valnemulin were identified as potential inhibitors of –1 programmed ribosomal frameshifting from a dual luciferase-based screen of FDA-approved small molecules in SARS-CoV-2 and related coronaviruses.[156] In addition, inclusion of these compounds allowed us to evaluate whether this strategy could be used to validate off-target mRNA:drug interactions, which may shed light on differential patient outcomes or reveal other pharmacological insights. For example, Kool and coworkers recently demonstrated that a related approach termed reactivity-based RNA profiling, which relies upon the ability of acylimidazole linkers to react with the 2'-OH of RNA, can be used to identify off-target RNA interactions of FDA-approved drugs.[157]

3.3.4 In vitro *Chem-CLIP-Map-Seq profiling of compounds 1–10*

By including all 10 of these diverse ligands in the same large-scale sequencing experiment, we were able to dive more deeply into specific comparisons between compound pairs, which should shed light on the chemical determinants that drive RNA binding. We chose MCF-7 breast cancer cells as a model system to explore RNAs that are required for normal functioning and RNAs overexpressed in cancer.[158] Cellular RNA from MCF-7 cells was extracted and mRNA-enriched using poly-A selection. The enriched pool of mRNA was incubated with 200 µM of each compound briefly at room temperature before UV crosslinking with 365 nm light (Figure 3.5b). Inclusion of the alkyne tag enabled us to install a biotin handle via CuAAC for isolation of the crosslinked mRNAs from the bulk solution. The DMSO control sample was not subjected to the streptavidin pull-down and therefore represents the RNA background that is present in the original population. After elution of the RNA from the beads, sequencing libraries were prepared and the samples submitted for NGS. It is important to note that while poly-A selection was performed, this is not expected to completely eliminate non-coding RNAs such as rRNAs or tRNAs from the starting RNA pool. Therefore, sequencing datasets were pre-filtered to remove noncoding RNAs from future analysis in an effort to focus on the mRNAome.

Despite recent reports indicating that chemical crosslinks on RNA cause reverse transcription (RT) stops or mutations, it remains unclear whether all such modifications elicit this effect and, if so, to what extent.[138] To circumvent this issue, we chose not to perform RNA fragmentation before pull-down with streptavidin. In this way, the workflow isolates the intact mRNA, and even if a crosslinking event causes RT stops or mutations, high sequence coverage should be retained in other regions of the transcript, resulting in a net enrichment of counts for the gene after alignment. With this analysis strategy, RNA targets were identified by comparing transcripts-per-million (TPM)-normalized count values across sample replicates to the DMSO sample and searching for statistically significant enrichments. Alternative count normalization strategies, including fragments per kilobase million and counts per million, were also explored using the raw counts files. The identity of hits for individual compounds did not change significantly with the normalization method employed. The inclusion of compound **1**, the negative control diazirine, allowed us to remove any false positive RNA hits that could be interacting with the diazirine-alkyne linker or undergoing enrichment due to an experimental artifact.

As anticipated, the various ligands employed in this experiment showed notable differences in both the number and identity of enriched RNA targets (Figure 3.5c). The gene set for this heatmap was created by filtering the TPM-normalized counts matrix to include only genes that were statistically significant hits for each individual compound when compared to the DMSO control [fold-change (FC) > 2, false discovery rate (FDR) < 0.01, n = 429]. The average TPM counts values for each of these hit genes was then used to generate the heatmap. Thus, the heatmap highlights the selectivity of the observed interactions, with each compound associating with distinct RNAs that were not pulled down by the other compounds in the same experiment. Interestingly, hits for compounds **7** and **8** clustered together in the heatmap, suggesting that the binding events are indeed being driven primarily by chemical structure, as these two probes share the ribose sugar ring and other molecular determinants. Compounds **9** and **10**, both of which are based on FDA-approved drugs, had many diverse hits with FC > 2 (87 and 92, respectively). In contrast, compounds **4** and **5**, both of which feature previously annotated protein-binding scaffolds, had the fewest interactions (18 and 0, respectively).

This analysis also allowed us to identify cellular mRNAs in MCF-7 cells that may be highly structured. For example, POLR2J2 mRNA encodes the protein RNA polymerase II subunit J2 required for mRNA synthesis.[159] This transcript was a hit for 6/10 compounds in our panel (Figure 3.5c), suggesting that it may have a highly stable structure or a population of structures that renders it particularly ligandable with small molecules. Alternatively, many mRNAs were not enriched by any compound, suggesting that these may be more difficult to target with small molecules, at least in the absence of proteins and other cellular machinery. Importantly, no correlation was observed between the abundance of an RNA in the sequencing pool and the likelihood of enrichment in this experiment.

A more in-depth analysis of individual hits for compound **9** revealed excellent agreement between individual sample replicates. A volcano plot between compound **9** and the DMSO-treated samples showed the anticipated net enrichment of most transcripts by the PAL ligand (Figure 3.6a). This net enrichment was also observed with all other ligands assayed and is likely a consequence of non-specific interactions that occur due to the high concentration of PAL ligand. However, with the exception of compound **5** which generated no significant hits, individual hits such as CDK5RAP3 mRNA for compound **9** separate from the bulk and are shifted into the significantly enriched quadrant of the volcano plot (Figure 3.6a). Indeed, closer inspection of the average normalized count values for CDK5RAP3 across the DMSO-, compound **1**- and compound **9**-treated samples confirms specific enrichment of this RNA target (Figure 3.6b). CDK5 regulatory subunit-associated protein 3 (CDK5RAP3) is involved in transcriptional regulation, cell cycle progression, and potentially tumorigenesis.[160]

As indicated earlier, compounds **7** and **8** were employed here to probe metabolite:RNA interactions and also investigate how two structurally related compounds perform in this workflow. To assess this question, the Tanimoto similarity scores were calculated for compounds **7**, **8**, and **9**.[161] As expected, compounds **7** and **8** are more structurally similar (Tanimoto score = 0.68) than either compound is to compound **9** (Tanimoto scores = 0.48 in both cases). Thus, if compound structure is the primary factor in determining hit identity, then we expected to observe a higher degree of correlation in hits between these structurally similar compounds. Indeed, a positive correlation was observed between the fold-change of enrichment for compounds **7** and **8** (R^2 = 0.74). Importantly, when the same Pearson correlation analysis was performed with compound **9** and either **7** or **8**, the correlation is significantly weaker (R^2 = 0.24 and 0.19, respectively). Despite

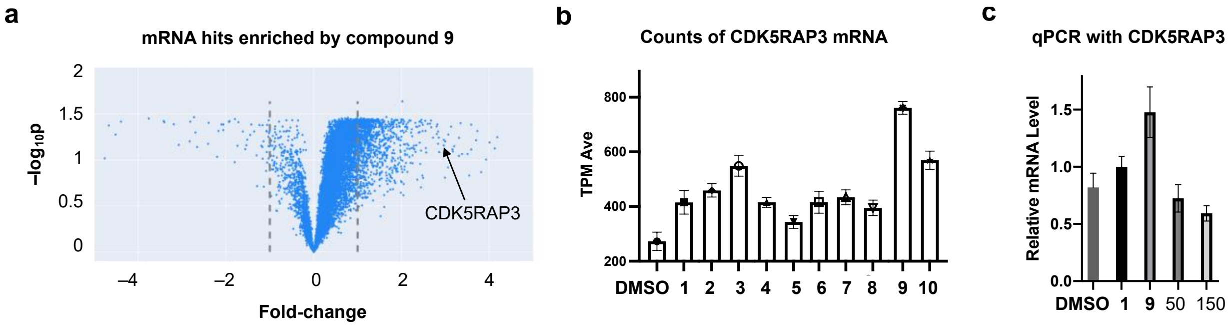


Figure 3.6: CDK5RAP3 mRNA is selectively enriched by compound 9. a Volcano plot depicting mRNAs enriched by compound **9** by *in vitro* Chem-CLIP-Map-Seq. **b** CDK5RAP3 mRNA sequencing counts are compared across the 11 RNA treatment conditions. **c** CDK5RAP3 mRNA is enriched from a pool of MCF-7 total RNA upon treatment with compound **9** by RT-qPCR. This enrichment is reversed in the presence of 50 or 150 μM palbociclib competitor.

the high correlation and structural similarity, only 10 hits were shared between compounds **7** and **8** in this experiment. Most of the hits for both **7** and **8** were unique to each compound (74% and 88%, respectively). Together, this analysis suggests that while similar compounds interact with similar RNA targets, unique interactions can still be discovered with this platform.

3.3.5 In vitro *RT-qPCR validation of compound 9*

Compound **9** was selected for further *in vitro* validation studies for several reasons. Firstly, the RNA pool was derived from MCF-7 breast cancer cells, and compound **9** is based on palbociclib, a clinically prescribed CDK4/6 inhibitor that is approved for the treatment of breast cancer.[140] Secondly, previous small-molecule microarray (SMM) screening efforts and recently published literature have indicated that this compound binds to various RNA structures.[154-155] Finally, in the initial 10-compound screen, this probe yielded a large pool of mRNA hits with diverse functions. *In vitro* RT-qPCR experiments with compound **9** in extracted in poly-A-enriched RNA confirmed the enrichment of several mRNA targets, including CDK5RAP3, METTL3, SULT1A3, SULT1A4, and MLPH, over both the DMSO- and negative control-treated samples (Figure B4). Additionally, these interactions appeared to be competable, as co-incubation of **9** with 50 or 150 µM palbociclib competitor reduced enrichment of the target RNAs in a concentration-dependent manner (Figure B4). This RT-qPCR experiment was also performed in non-poly-A-selected, MCF7 total RNA to analyze CDK5RAP3 mRNA enrichment upon treatment with compound **1**, compound **9**, or compound **9** + free palbociclib competitor (Figure 3.6c). Excitingly, the results were reproducible in the more complex RNA mixture. The additional proteins encoded by the additional mRNA hits explored above, methyltransferase 3 (METTL3), sulfotransferase family 1A members 3 and 4 (SULT1A3/4), and melanophilin (MLPH), are involved in N6-adenosine mRNA methylation,[162] phenol sulfotransfer to signaling and drug molecules,[163-164] and pigmentation,[165] respectively.

The mRNA hits highlighted thus far underscore several striking features. First, the four encoded proteins listed above have been linked to cancer.[165-168] Thus, our photocrosslinking strategy enriched pharmacologically relevant RNA targets, and the strategic selection of MCF-7 breast cancer cells as a model system may have been conducive to this end. Second, POLR2J2, CDK5RAP3, and METTL3 are involved in transcriptional regulation. It is possible that an RNA-focused enrichment method such as ours preferentially yields transcripts that are both involved in RNA processing and relevant for therapeutic modulation of RNA pathways. Finally, this method detected two homologous genes, SULT1A3 and SULT1A4, lending confidence to the reproducibility and specificity of the identified RNA:small molecule interactions.

3.3.6 In vitro *competitive Chem-CLIP-Map-Seq validation of compound 9*

CDK5RAP3 was most significantly enriched by compound **9** in the 10-compound sequencing experiment (Figure 3.6b). While compounds **3** and **10** also displayed a moderate increase in TPM counts for CDK5RAP3 when compared to the DMSO control, only compound **9**-treated samples were significantly higher than the negative control (**1**)-treated samples, highlighting the potential selectivity of the interaction between **9** and CDK5RAP3 (Figure 3.6b). To assess the robustness of this photoaffinity-based RNA enrichment strategy, the Chem-CLIP-Map-Seq experimental workflow was repeated, focusing only on compound **9**. This time, samples were also prepared in which compound **9** was co-incubated with 55, 167, or 500 µM palbociclib competitor (Figure 3.7a). If the identified RNA:small molecule interactions are specific, then free palbociclib should decrease TPM counts in the sequencing data. The same analysis pipeline was used to process the data and identify hits.

In the absence of competitor, the most significantly enriched RNA target of compound **9** was again CDK5RAP3. Correlative analysis between the fold-change of individual RNAs in the two Chem-CLIP-Map-Seq experiments revealed a strong correlation ($R^2 = 0.76$) for compound **9**-treated samples. This is especially notable as the starting pools for the two experiments were not identical, with the first Chem-CLIP-Map-Seq experiment being performed on poly-A-selected mRNA and the second being performed on total RNA. In contrast, when the fold-change of all mRNAs in the competitive Chem-CLIP-Map-Seq experiment with compound **9** was compared to the fold-change of all mRNAs in the first Chem-CLIP experiment with other compounds such as compound **8**, significantly lower correlation was observed ($R^2 = 0.30$). Additionally, the observed correlation between hits for compound **9**-treated samples in the two Chem-CLIP experiments decreased in the presence of 55 µM free palbociclib competitor ($R^2 = 0.56$).

These conclusions are recapitulated in an analysis of unique versus shared hits between different compound treatment conditions (Figure 3.7b). Namely, a large proportion of hits are shared between the two Chem-CLIP-Map-Seq experiments for the samples that were treated with compound **9**. This proportion is drastically lower when comparing compound **9** treatment in the second sequencing experiment with compound **8** treatment in the first. Finally, an intermediate number of hits are shared between **9** + 55 µM competitor treatment in experiment 2 and **9** treatment in experiment 1 (Figure 3.7b).

This Comp-Chem-CLIP-Map-Seq experiment also enabled us to examine the selectivity of the profiled interactions. As anticipated, the number of enriched transcripts decreased with increasing concentrations of competitor, as is clearly depicted in the volcano plots of the data (Figure 3.7c). At the highest concentration of competitor (500 µM, 2:1 competitor:PAL ligand), the sample much more closely resembled the DMSO control, with very few genes showing significant changes. K-means clustering was used to identify clusters of genes that respond similarly to the treatment. In total, four clusters were identified (Figure 3.7d). The genes in cluster A show the expected pattern of enrichment over the DMSO control, with competition occurring in a concentration-dependent manner.

Importantly, CDK5RAP3 is a member of cluster A, and it showed the expected pattern of enrichment over DMSO, followed by concentration-dependent competition with free palbociclib (Figure 3.7e). The other validated genes including METTL3, MLPH, NOP2 and SULT1A3/A4 were

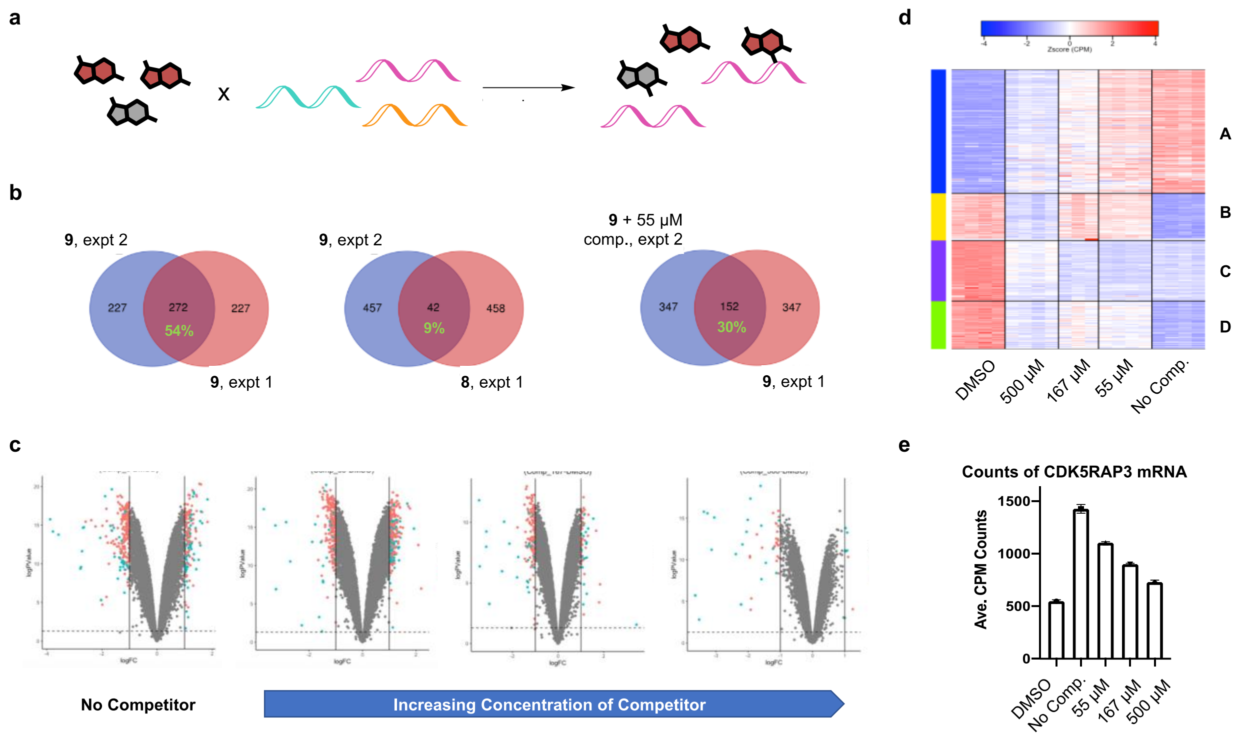


Figure 3.7: Competitive Chem-CLIP-Map-Seq with compound 9. a General scheme to assess 250 μM compound **9** labeling of total RNA from MCF-7 breast cancer cells in the presence or absence of 55, 167, or 500 μM palbociclib competitor. **B** Quantification of statistically significant mRNA enriched in Chem-CLIP-Map-Seq experiments 1 and 2 for the delineated treatment conditions. **c** Volcano plots demonstrate decreases in the number of transcripts enriched by compound **9** during Comp-Chem-CLIP-Map-Seq in a concentration-dependent manner. **d** K-means clustering of the transcripts that were differentially enriched in the five treatment conditions revealed four clusters. The species in cluster A were depleted with increasing concentrations of palbociclib. **e** CDK5RAP3 mRNA was again enriched by compound **9**, but palbociclib reversed this effect.

not members of cluster A. However, inspection of the raw TPM counts for these genes revealed that METTL3, MLPH and NOP2 all show the expected pattern of enrichment with **9** only in the absence of the competitor. These genes simply did not show a concentration-dependent decrease, with maximal competition observed even at the lowest concentration (55 μM). In contrast, SULT1A3/4 did not seem to be competable, at least at the concentrations of free palbociclib used here. Like the enriched hits highlighted earlier, nucleolar protein 2 (NOP2) is also involved in both RNA processing and cancer, catalyzing 5-methylcytosine installation in rRNA toward ribosome biogenesis and promoting cell proliferation and tumorigenesis.[169-171]

Finally, we note that the enrichment patterns and concentration-dependent competition described above were not observed for GAPDH or other housekeeping genes such as NRAS or

MYC, nor were they hits for compound **9** in the first Chem-CLIP-Map-Seq experiment. Altogether, these analyses suggests that the Chem-CLIP experiment is reproducible and that the compound structure of the PAL ligand is the determining factor in the resulting hits. The data also support palbociclib as a potentially specific binder of CDK5RAP3 mRNA.

3.3.7 In cellulo *Chem-CLIP-Map-Seq of compound 9*

Additional Chem-CLIP-Map-Seq experiments were conducted to profile the interactions of compound **9** with RNA targets in live MCF-7 cells (Figure 3.8a). We were curious to investigate how the presence of proteins and other cellular factors affecting localization would impact the RNA:small molecule interactions observed *in vitro*. In contrast to the *in vitro* experiments on isolated cellular RNA, signaling pathways are intact in cells, and the treatment of MCF-7 cells with palbociclib has been shown to result in dramatic transcriptome-wide changes in RNA levels that could impact our results.[172-173] This could be particularly problematic at the high concentrations of **9** necessary to achieve efficient PAL crosslinking in the *in vitro* experiments. To avoid these effects, we replaced the DMSO control with an "RNA input" control sample that is collected after UV crosslinking but before streptavidin pull-down. By using the input sample as a control, any transcriptome-wide changes in the RNA population that might result from compound **9** binding to cellular targets such as CDK4/6 could be removed from our final analysis. This strategy is thus expected to provide a much clearer picture of direct binding interactions between the tested compound and RNAs in the complex cellular environment.

Principal component analysis (PCA) revealed a moderate level of clustering between individual replicates, with **9**-treated input and output samples separating along the first principal component, which accounts for 41% of the observed variation (Figure 3.8b). In the case of the cell-based experiment, analysis with TPM-normalized counts resulted in an unexpected skew in the dataset that separated transcripts into coding and noncoding populations. However, with FPKM or CPM counts this effect was not observed, suggesting that it is a unique result of normalizing the counts with TPM. Additionally, this effect was not observed in either *in vitro* Chem-CLIP-Map-Seq experiment when the results were analyzed with TPM. Based on these observations, we speculate that this discrepancy is likely caused by differences in the cellular localization of these transcripts or a result of utilizing the input control in place of DMSO.

Interestingly, there was significantly less correlation between the results of the *in vitro* data set and the Chem-CLIP experiment in live cells. The lack of observed correlation between the *in vitro* and *in cellulo* experiments suggests that the overall targetability of the mRNAs has changed, likely due to the aforementioned interactions with proteins and/or other cellular machinery. Despite this lack of correlation, 22 of the validated *in vitro* targets of compound **9**, including CDK5RAP3, were also hits in the Cell-Chem-CLIP-Map-Seq experiment, with excellent agreement between individual replicates (Figure 3.8c). Encouragingly, CDK5RAP3 mRNA was enriched by compound **9** in the cell-based experiments (Figure 3.8d).

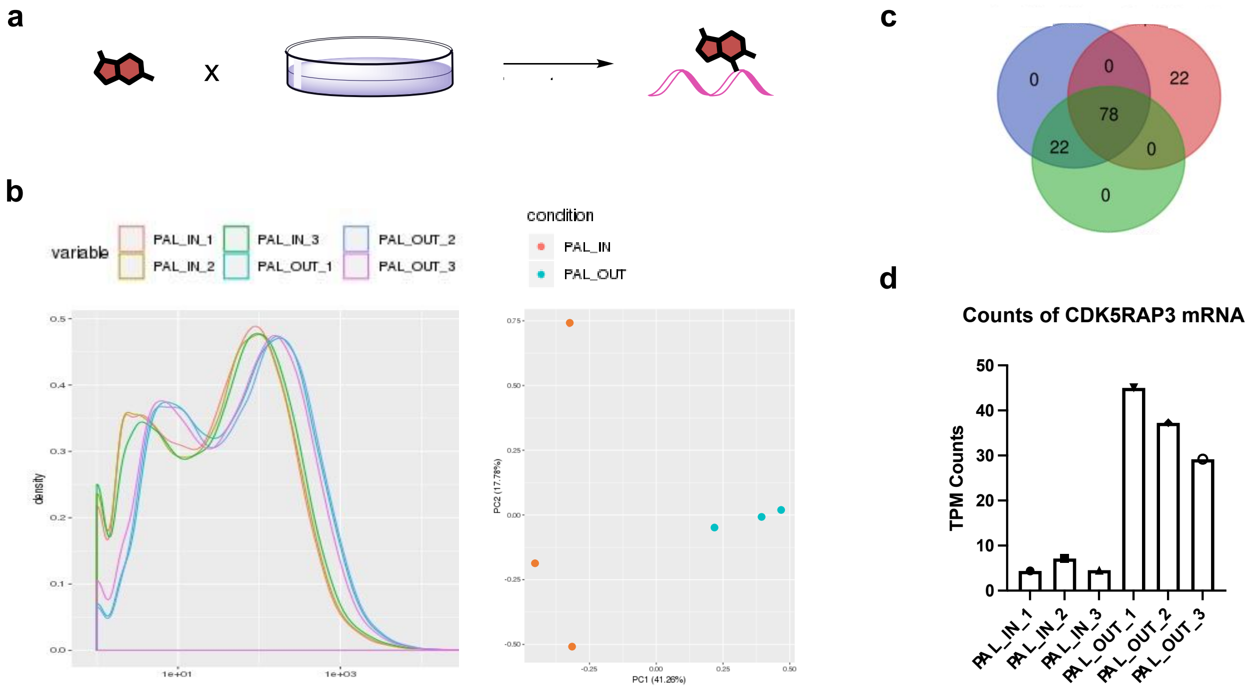


Figure 3.8: *In cellulo* Chem-CLIP-Map-Seq with compound 9. a General scheme to assess probe labeling of native mRNA in MCF-7 breast cancer cells. **b** Density distribution plot and 2D PCA showing clustering of cell-treated compound **9** replicates and the separation of replicates into two distinct groups. The x-axis of the PCA plot has 42% of the variance. **c** The three treatment replicates share 78 enriched mRNA targets, demonstrating reproducibility. **c** CDK5RAP3 mRNA is enriched in a cellular environment.

3.4 Conclusion

The limited druggability of various oncoproteins has provided a significant roadblock to researchers working to provide treatments for cancer patients. One way to circumvent the lack of effective compounds targeting these critical pathways at the protein level is to move upstream and target the mRNA. By targeting mRNA splicing, translation, stability, or various other regulatory mechanisms, the effective pool of an oncoprotein can be depleted.[174-177]

While RNA is an exciting therapeutic target, the rules that govern RNA ligandability have only sparsely been explored. Gaining a better understanding will accelerate the discovery of novel therapeutics. Recent developments such as the clinical success and FDA approval of Risdiplam have precipitated a huge increase in interest in the development of small-molecule probes and inhibitors of complex RNA structures.[14] Despite these efforts, the discovery of selective RNA targeting ligands with cellular activity has proven challenging. There are numerous reasons for this roadblock, including the cost and instability of *in vitro*-synthesized RNAs, a lack of high-resolution RNA structures, and a lack of methods to screen small molecules against RNA targets either *in vitro* or *in cellulo*. Additionally, the few ligands that have been described, such as naturally occurring aminoglycosides, suffer from poor target selectivity and non-ideal pharmacokinetic properties that restrict their applications.[178] There is a strong need for novel and orthogonal methods that enable researchers to assess the broad targetability of the human mRNAome.

Our first approach in this work to profile transcriptome-wide RNA:small molecule interactions involved immobilization of a test ligand, PreQ1, on beads to validate non-covalent pulldown of the cognate PreQ1 aptamer. Perhaps unsurprisingly, we found by MST that perturbation of the native interaction led to reduced binding affinity and unreliable RNA pulldown. This led us to pivot to covalent RNA labeling strategies via the incorporation of a diazirine-alkyne pulldown handle into the ligand. We found that yne-dz-PreQ (compound **6**) enabled irreversible PreQ1 RNA upon irradiation with UV light. Further, this interaction was competable with excess PreQ1 ligand in the presence of MCF-7 total RNA.

Using a structurally diverse, 10-compound ligand panel, we were then able to show that, similarly to what is observed with proteins, human cellular RNAs form discrete interactions and exhibit unique reactivity profiles. Some compounds were more reactive than others, with the more elaborated FDA-approved drugs and metabolites included in the panel yielding more hits than the R-BIND or protein-binding fragments selected. This finding could reveal the need for structurally complex RNA-targeted ligands in order to interact with RNA binding pockets in a specific manner. The lack of activity for compounds **4** and **5** is particularly notable and perhaps to be expected, as these are both described in the literature as privileged protein-binding scaffolds.[146-148] Meanwhile, the moderate number of targets pulled down with compounds **2** and **3** could indicate that these are privileged RNA-interacting scaffolds. Note that an ideal RNA-targeting scaffold would selectively interact with only one or a few RNA targets. Interestingly, only five targets (out of 80 and 21 hits, respectively) were overlapping between compounds **2** and **3**, suggesting that these two R-RBIND ligands have a non-overlapping set of mRNA targets.

Compounds **6**, **7** and **8**, all of which were chosen due to their annotated interactions with RNA aptamers in bacteria, allowed us to investigate the impact of closely related structures in RNA:small molecule recognition. The high affinity and selectivity of these bacterial riboswitches highlight the ability of RNA to assume stable 3D structures that can discriminate between highly similar chemical matter. Compound **6** had been assayed previously to develop covalent inhibitors of the PreQ1 riboswitch.[113] This compound is structurally similar to guanosine and is therefore related to compounds **7** and **8**, which are respectively related to a purine nucleobase precursor (AICA) and adenine. As expected, a high level of correlation was observed for compound **6** to both **7** and **8** ($R^2 = 0.79$ and 0.85 respectively). Interestingly, there were more overlapping hits between **6** and **8** than between any other two compounds included in this study. This was somewhat unexpected, as we anticipated that the shared ribose moiety in compounds **7** and **8** would play a greater role in dictating RNA targets. Compounds **6**, **7**, and **8** possessed seven RNA targets in common. It is possible that one or more of these RNAs is involved in a previously unannotated regulatory pathway that relies upon nucleobase/metabolite recognition, although further work is needed to validate these findings.

Compounds **9** and **10** were included with the aim of profiling potential off-target interactions of these important FDA-approved small molecules. As indicated previously, both ligands have been shown to interact with multiple RNA structural motifs in SMM screening datasets and in the literature.[154-156] We are interested in exploring in the future how the RNA targets of palbociclib, the parent of compound **9**, vary from those of other FDA-approved CDK4/6 inhibitors such as abemaciclib or ribociclib.[179-180] Despite having the same primary cellular target, these compounds show highly different potency and side effects in clinical trials, some of which may be mitigated by functional RNA off-target interactions. Indeed, Kool and coworkers recently showed that a related covalent approach could be used to annotate off-target RNA-drug interactions for levofloxacin and dasatinib.[157] Application of the orthogonal approach described here could be used to validate and better understand the mechanisms of these drugs.

While one goal of this work was to evaluate broad chemical space and investigate which ligands preferentially recognize RNAs, this method also allows us to evaluate the ligandability of the human mRNAome. Interestingly, we saw no correlations between the abundance of a transcript in the RNA pools tested and the fold-change of enrichment. High concentrations of PAL ligand were used to promote covalent capture of weak or transient interactions, particularly for the fragment-based compounds. When comparing transcripts with similar abundance levels, clear preferences emerged across all compounds. For example, POLR2J2 mRNA was a hit for 6/10 compounds assayed, while similarly abundant mRNAs such as MYC and NRAS never emerged as hits. This finding could indicate that the POLR2J2 RNA folds into a complex, stable structure that can accommodate many of the ligands tested here. Alternatively, it is possible that these six ligands share a chemical determinant that promotes the interaction. Interestingly, POLR2J2 was also a hit in the Cell-Chem-CLIP-Map-Seq experiment that was conducted with compound **9**, suggesting that this mRNA may retain said structure even in the cellular context. This approach adds to the arsenal of tools available to the community of researchers working on identifying new RNA targets with high-throughput screening efforts. Many of the mRNA hits in these studies encoded proteins affiliated with RNA processing and cancer.

After validating that the photocrosslinking method could be used to distinguish between *in vitro* binding events even for highly similar molecules and discover reproducible, competable interactions, we extended this method to cells. Despite a high level of correlation for compound **9**-treated samples in the two *in vitro* Chem-CLIP-Map-Seq experiments, this correlation was mostly abolished in the cellular context. This is consistent with what has been reported in the literature and suggests that many of our *in vitro*-profiled interactions are unlikely to be preserved in cells. However, the majority of the significantly enriched *in vitro* hits were observed in the cells, and 22 transcripts were compound **9** hits in all three experiments. CDK5RAP3 mRNA is a notable example among these, although it was not the most significantly enriched target *in cellulo*. These experiments highlight the critical importance of context in RNA structure and suggest that future screens should take place in cells to discover RNA:small molecule interactions with functional consequences. Additionally, it may be necessary to combine this or a related strategy to traditional protein-based PAL to truly uncouple the complex regulatory mechanisms underlying the observed interactions.

Together, these results underscore the challenges associated with exploring and developing RNA-targeted chemical probes. While non-covalent interactions may be sufficient to elicit biological responses *in cellulo* or *in vivo*, they are not necessarily strong enough to withstand the harsher conditions necessary for RNA pulldown and bioinformatic analysis. Promisingly, covalent photocrosslinking reagents can be utilized to irreversibly capture RNA:small molecule binding events that occur in cells. However, decorating probes of interest with such photocrosslinking moieties can alter their native RNA contacts. The use of unmodified competitor molecules is a promising strategy for exploring the specificity and selectivity of fragment binding to cellular RNAs toward both better understanding the features of ligandable RNAs and identifying new lead molecules for therapeutic purposes. Future efforts will focus on validating the RNA hits identified through these studies.

3.5 Materials and Methods

3.5.1 General information

All reagents were used as purchased from their chemical manufacturer. Reagents were purchased from MilliporeSigma unless otherwise specified. pH values were measured using a Mettler Toledo SevenCompact S220 pH Meter. MCF-7 cells were maintained in DMEM (High Glucose, Pyruvate, Gibco) supplemented with 10% Fetal Bovine Serum (Gibco) at 37 ºC in a humidified atmosphere of 5% CO_2 and 95% air. Cells were passaged at 80–90% confluence.

3.5.2 Preparation of PAL probe library

The substrate 3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanoic acid (10 mg, 1.0 eq.) was preactivated by stirring in the presence of O-(1H-6-chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (4.0 eq.) and diisopropylethylamine (8.0 eq.) in anhydrous dimethylformamide (0.5 mL). After 1–2 min, the amine- or alcohol-containing small-molecule substrate (1.2 eq.) was added to the solution, and the mixture was stirred at rt for 2 h. Then, the reaction mixture was concentrated under vacuum and purified via column chromatography on a CombiFlash Rf system (0–40% gradient, MeOH:DCM) or via preparative HPLC (10–90% gradient of H_2O:MeCN) as necessary, to isolate the purified product.

3.5.3 Microscale thermophoresis

RNAs were Cy5-labeled using the *Label* IT Nucleic Acid Labeling Kit, Cy5 (MirusBio) and following the manufacturer protocol. In brief, the *Label* IT Reagent and 10X Labeling Buffer A were warmed to room temperature before use, and 3.5 µL of each was combined with 3.5 µg RNA and the reaction filled to 35 µL final volume with RNase-free water, with the *Label* IT Reagent being added to the reaction mixture last. The reaction was incubated at 37°C in the dark for 1 h, with brief centrifugation after 0.5 h. The reaction volume was brought up to 100 µL with water and the reaction ethanol-precipitated to isolate the labeled RNA and remove unreacted Cy5 reagent.

The RNA was diluted in riboswitch buffer (50 mM Tris-HCl, 100 mM KCl, 2 mM $MgCl_2$, pH 7.5) to ~200 nM and analyzed on a Monolith 2020 (TNG) (MM-167) MST instrument in Nano - RED and binary binding modes. All samples were measured in duplicate using standard capillaries. The RNA was diluted until the signal read ~1000 counts, and that RNA concentration was used for all subsequent binding experiments. Meanwhile, the small-molecule probe was prepared as a 20 mM stock solution in DMSO and further diluted to 400 µM in riboswitch buffer. Then, 16 serial dilutions were prepared with 10 µL of riboswitch buffer and 10 µL of the 400 µM probe solution to obtain concentrations ranging from ~6 nM–200 µM. To each tube was added 10 µL RNA. The samples were incubated at 60 ºC for 10 min and then at rt for 5 min to promote RNA folding. MST binding curves were collected for each RNA:small molecule pair in triplicate.

3.5.4 RT-qPCR

Total RNA was purified from MCF-7 cells using the Monarch Total RNA Miniprep Kit (NEB) with DNase I treatment performed to remove unwanted nucleic acid. The RNA was diluted in riboswitch buffer + 0.05% Tween-20 to ~530 ng/µL and annealed by heating at 90 ºC for 4 mins, followed by snap cooling on ice for 5 min. To 95 µL aliquots were added 5 µL DMSO or 5 mM compound **1** or compound **9** ± the desired concentration of free palbociclib in DMSO. The reaction was mixed and incubated at 37 ºC for 1 h to allow probe:RNA binding. Then, the 100 µL solutions were pipetted into 96-well clear plate and UV irradiated at 365 nm under an ultraviolet, high intensity UVP lamp (Model B-100A, Analytik Jena) for 15 minutes at rt. Each solution was then transferred an Eppendorf tube.

Next, the alkyne-labeled RNAs were biotinylated by CuAAC. First, 90 µL of THPTA (200 mM) was combined with 180 µL $CuSO_4$ (50 mM), vortexed, and incubated at rt for 15 min to form the copper pre-complex. Meanwhile, 10 µL of biotin-N_3 (5 mM in DMSO) was added to the RNA sample. After addition of 10 µL of freshly prepared ascorbic acid (200 mM) and vortexing, 30 µL of THPTA:$CuSO_4$ mixture was added and further vortexed. The reaction was incubated at rt for 1 h, vortexing every 15–20 min, followed by EtOH precipitation and resuspension of the RNA pellet in 50 µL riboswitch buffer + 0.05% Tween.

The RNA was incubated with 50 µL of Streptavidin Magnetic Beads (4 mg/mL, NEB) at rt for 30 min with constant rotation, as well as vortexing every 10 min. After application of the magnet, the supernatant containing the unbound RNAs was discarded and the beads washed 3X with 200 µL riboswitch buffer + 0.01% Tween-20 to remove nonspecifically-bound RNAs. The supernatant was again removed, and the bound RNA was eluted in 50 µL of 95% formamide + 10 mM EDTA (pH 8.2) for 2 min at 92 ºC. The Monarch RNA Cleanup Kit (NEB) was used to isolate biotin-mRNA, eluting in 25 µL water and quantifying recovery by NanoDrop.

cDNA was prepared using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher) following the manufacturer protocol. A 2X RT master mix was prepared by combining the following: 10X RT Buffer (21 µL), 25X dNTP Mix (100 mM, 8.4 µL), 10X RT Random Primers (21 µL), ddH2O (44.1 µL), and MultiScribe RTase (10.5 µL) to 105 µL final volume. Careful pipetting, 10 µL 2X master mix solution was combined with 10 µL RNA samples. The reactions were incubate in a thermocycler as follows: 25 ºC, 10 minutes; 37 ºC, 2 h; 85 ºC, 5 min. Each sample was diluted with 20 µL water and stored temporarily at –20 ºC.

Finally, iTaq Universal SYBR Green Supermix (Bio-Rad) was thawed to rt, vortexed, centrifuged, and stored on ice in the dark. The following were then combined: iTaq Supermix (2X, 105 µL), gene-specific forward primer (100 µM, 2.1 µL, 1 µM final), gene-specific reverse primer (100 µM, 2.1 µL, 1 µM final), and ddH2O (48.6 µL) to 157.5 µL volume total. Aliquots (15 µL) of each gene master mix were transferred to a qPCR plate and combined with 5 µL of the appropriate triplicate cDNA samples. The plate was centrifuged for 20 s to remove air bubbles. Finally, the reactions were heated in the thermocycler as follows: 94 ºC for 3 min; 45 cycles of (94 ºC for 15 s, 60 ºC for 30 s, 68 ºC for 30 s); 68 ºC for 5 min. Data analysis was performed by comparing enrichment of the RNAs of interest in probe- over control-treated samples, normalizing to GAPDH.

3.5.5 In cellulo *Chem-CLIP-Map-Seq*

Three plates of ~90% confluent MCF-7 cells grown on 150 mm Nunc EasYDish Dishes (Thermo Fisher) were grown for each treatment condition. The media was removed, the cells were rinses with 5 mL warm 1X PBS, and the cells were incubated with 5 mL serum-free media containing 200 µM probe in the dark for 30 min at 37 ºC. An additional 5 mL serum-free media was added to the cells before irradiation for 10 min from 12 cm distance at 365 nm under the UV lamp to label cellular RNAs. The media was then discarded and the cells rinsed with 10 mL warm PBS. Using trypsin, the cells were pelleted at 900 RPM for 4 min and the supernatant discarded. The pellets were resuspended in 1 mL RNA Lysis Buffer by pipetting gently to avoid foaming. The total RNA was isolated using the Monarch Total RNA Miniprep Kit (NEB) with DNase I treatment performed to remove unwanted nucleic acid. The RNA concentration was recorded by NanoDrop and the RNA ethanol-precipitated to further concentrate it.

Next, the alkyne-labeled RNAs were biotinylated as described above. The maximum amount of RNA available from the cell treatment and RNA isolation protocol was used. The RNA pellet was resuspended in 55.5 µL riboswitch buffer. An aliquot of 10% of the RNA sample was removed and stored at –80 ºC to preserve the "input"/non-enriched RNA sequences. The remaining biotinylated RNA was enriched on streptavidin beads as described above. RNA concentration was analyzed by NanoDrop and Bioanalyzer.

3.6 References

1. Consortium, E. P., An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012,** *489* (7414), 57-74.
2. Cooper, T. A.; Wan, L.; Dreyfuss, G., RNA and disease. *Cell* **2009,** *136* (4), 777-793.

3. Boivin, V.; Faucher-Giguere, L.; Scott, M.; Abou-Elela, S., The cellular landscape of mid-size noncoding RNA. *Wiley Interdiscip. Rev. RNA* **2019,** *10* (4), e1530.

4. Dixon, S. J.; Stockwell, B. R., Identifying druggable disease-modifying gene products. *Curr. Opin. Chem. Biol.* **2009,** *13* (5-6), 549-55.

5. Warner, K. D.; Hajdin, C. E.; Weeks, K. M., Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **2018,** *17* (8), 547-558.

6. Connelly, C. M.; Moon, M. H.; Schneekloth, J. S., Jr., The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem Biol* **2016,** *23* (9), 1077-1090.

7. Zhu, Y.; Zhu, L.; Wang, X.; Jin, H., RNA-based therapeutics: An overview and prospectus. *Cell Death Dis.* **2022,** *13* (7), 644.

8. Sheridan, C., First small-molecule drug targeting RNA gains momentum. *Nat. Biotechnol.* **2021,** *39* (1), 6-8.

9. Ly, C. V.; Miller, T. M., Emerging antisense oligonucleotide and viral therapies for amyotrophic lateral sclerosis. *Curr Opin Neurol* **2018,** *31* (5), 648-654.

10. Pawluk, A.; Amrani, N.; Zhang, Y.; Garcia, B.; Hidalgo-Reyes, Y.; Lee, J.; Edraki, A.; Shah, M.; Sontheimer, E. J.; Maxwell, K. L.; Davidson, A. R., Naturally Occurring Off-Switches for CRISPR-Cas9. *Cell* **2016,** *167* (7), 1829-1838 e9.

11. Kumar Mishra, S.; Kumar, A., NALDB: nucleic acid ligand database for small molecules targeting nucleic acid. *Database (Oxford)* **2016,** *2016*.

12. Mehta, A.; Sonam, S.; Gouri, I.; Loharch, S.; Sharma, D. K.; Parkesh, R., SMMRNA: a database of small molecule modulators of RNA. *Nucleic Acids Res* **2014,** *42* (Database issue), D132-41.

13. Disney, M. D.; Winkelsas, A. M.; Velagapudi, S. P.; Southern, M.; Fallahi, M.; Childs-Disney, J. L., Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chem Biol* **2016,** *11* (6), 1720-8.

14. Morgan, B. S.; Sanaba, B. G.; Donlic, A.; Karloff, D. B.; Forte, J. E.; Zhang, Y.; Hargrove,

A. E., R-BIND: An interactive database for exploring and developing RNA-targeted chemical probes. *ACS Chem. Biol.* **2019,** *14* (12), 2691-2700.

15. Morgan, B. S.; Forte, J. E.; Culver, R. N.; Zhang, Y.; Hargrove, A. E., Discovery of key physicochemical, structural, and spatial properties of RNA-targeted bioactive ligands. *Angew. Chem. Int. Ed. Engl.* **2017,** *56* (43), 13498-13502.

16. Hashemian, S. M. R.; Farhadi, T.; Ganjparvar, M., Linezolid: A review of its properties, function, and use in critical care. *Drug Des. Devel. Ther.* **2018,** *12*, 1759-1767.

17. Markati, T.; Fisher, G.; Ramdas, S.; Servais, L., Risdiplam: An investigational survival motor neuron 2 (SMN2) splicing modifier for spinal muscular atrophy (SMA). *Expert Opin. Investig. Drugs* **2022,** *31* (5), 451-461.

18. Hewitt, W. M.; Calabrese, D. R.; Schneekloth, J. S., Jr., Evidence for ligandable sites in structured RNA throughout the Protein Data Bank. *Bioorg. Med. Chem.* **2019,** *27* (11), 2253-2260.

19. Parker, C. G.; Galmozzi, A.; Wang, Y.; Correia, B. E.; Sasaki, K.; Joslyn, C. M.; Kim, A. S.; Cavallaro, C. L.; Lawrence, R. M.; Johnson, S. R.; Narvaiza, I.; Saez, E.; Cravatt, B. F., Ligand and target discovery by fragment-based screening in human cells. *Cell* **2017,** *168* (3), 527-541 e29.

20. Hill, J. R.; Robertson, A. A. B., Fishing for Drug Targets: A Focus on Diazirine Photoaffinity Probe Synthesis. *J Med Chem* **2018,** *61* (16), 6945-6963.

21. Nakamoto, K.; Ueno, Y., Diazirine-containing RNA photo-cross-linking probes for capturing microRNA targets. *J. Org. Chem.* **2014,** *79* (6), 2463-72.

22. Suresh, B. M.; Li, W.; Zhang, P.; Wang, K. W.; Yildirim, I.; Parker, C. G.; Disney, M. D., A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117* (52), 33197-33203.

23. Mukherjee, H.; Blain, J. C.; Vandivier, L. E.; Chin, D. N.; Friedman, J. E.; Liu, F.; Maillet, A.; Fang, C.; Kaplan, J. B.; Li, J.; Chenoweth, D. M.; Christensen, A. B.; Petersen, L. K.; Hansen, N. J. V.; Barrera, L.; Kubica, N.; Kumaravel, G.; Petter, J. C., PEARL-seq: A

photoaffinity platform for the analysis of small molecule-RNA interactions. *ACS Chem. Biol.* **2020,** *15* (9), 2374-2381.

24. Velagapudi, S. P.; Li, Y.; Disney, M. D., A cross-linking approach to map small molecule-RNA binding sites in cells. *Bioorg. Med. Chem. Lett.* **2019,** *29* (12), 1532-1536.

25. Tong, Y.; Gibaut, Q. M. R.; Rouse, W.; Childs-Disney, J. L.; Suresh, B. M.; Abegg, D.; Choudhary, S.; Akahori, Y.; Adibekian, A.; Moss, W. N.; Disney, M. D., Transcriptome-wide mapping of small-molecule RNA-binding sites in cells informs an isoform-specific degrader of QSOX1 mRNA. *J. Am. Chem. Soc.* **2022,** *144* (26), 11620-11625.

26. Kim, E. S.; Scott, L. J., Palbociclib: A review in HR-positive, HER2-negative, advanced or metastatic breast cancer. *Target. Oncol.* **2017,** *12* (3), 373-383.

27. Serra, F.; Lapidari, P.; Quaquarini, E.; Tagliaferri, B.; Sottotetti, F.; Palumbo, R., Palbociclib in metastatic breast cancer: Current evidence and real-life data. *Drugs Context* **2019,** *8*, 212579.

28. Roth, A.; Winkler, W. C.; Regulski, E. E.; Lee, B. W.; Lim, J.; Jona, I.; Barrick, J. E.; Ritwik, A.; Kim, J. N.; Welz, R.; Iwata-Reuyl, D.; Breaker, R. R., A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.* **2007,** *14* (4), 308-17.

29. Eichhorn, C. D.; Kang, M.; Feigon, J., Structure and function of preQ1 riboswitches. *Biochim. Biophys. Acta* **2014,** *1839* (10), 939-950.

30. Mandal, M.; Breaker, R. R., Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **2004,** *5* (6), 451-63.

31. Kavita, K.; Breaker, R. R., Discovering riboswitches: The past and the future. *Trends Biochem. Sci.* **2023,** *48* (2), 119-141.

32. Connelly, C. M.; Numata, T.; Boer, R. E.; Moon, M. H.; Sinniah, R. S.; Barchi, J. J.; Ferre-D'Amare, A. R.; Schneekloth, J. S., Jr., Synthetic ligands for PreQ1 riboswitches provide structural and mechanistic insights into targeting RNA tertiary structure. *Nat. Commun.* **2019,** *10*

(1), 1501.

33. Wang, J.; Schultz, P. G.; Johnson, K. A., Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc. Natl. Acad. Sci. U.S.A.* **2018,** *115* (20), E4604-E4612.

34. Ayele, T. M.; Loya, T.; Valdez-Sinon, A. N.; Bassell, G. J.; Heemstra, J. M., Imaging and tracking mRNA in live mammalian cells via fluorogenic photoaffinity labeling. *Available on bioRxiv* **2020**, https://doi.org/10.1101/2020.02.10.942482.

35. Batey, R. T., Recognition of S-adenosylmethionine by riboswitches. *Wiley Interdiscip. Rev. RNA* **2011,** *2* (2), 299-311.

36. Grate, D.; Wilson, C., Laser-mediated, site-specific inactivation of RNA transcripts. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96*, 6131-6136.

37. Filippakopoulos, P.; Picaud, S.; Fedorov, O.; Keller, M.; Wrobel, M.; Morgenstern, O.; Bracher, F.; Knapp, S., Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family. *Bioorg. Med. Chem.* **2012,** *20* (6), 1878-86.

38. Smith, S. G.; Sanchez, R.; Zhou, M. M., Privileged diazepine compounds and their emergence as bromodomain inhibitors. *Chem. Biol.* **2014,** *21* (5), 573-83.

39. Irie, T.; Sawa, M., 7-Azaindole: A versatile scaffold for developing kinase inhibitors. *Chem. Pharm. Bull.* **2018,** *66* (1), 29-36.

40. Kim, P. B.; Nelson, J. W.; Breaker, R. R., An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol. Cell* **2015,** *57* (2), 317-28.

41. Tran, B.; Pichling, P.; Tenney, L.; Connelly, C. M.; Moon, M. H.; Ferre-D'Amare, A. R.; Schneekloth, J. S., Jr.; Jones, C. P., Parallel discovery strategies provide a basis for riboswitch ligand design. *Cell Chem. Biol.* **2020,** *27* (10), 1241-1249 e4.

42. Mandal, M.; Breaker, R. R., Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* **2004,** *11* (1), 29-35.

43. Paukner, S.; Riedl, R., Pleuromutilins: Potent drugs for resistant bugs—Mode of action and resistance. *Cold Spring Harb. Perspect. Med.* **2017,** *7* (1).

44. Garber, K., The cancer drug that almost wasn't. *Science* **2014,** *345* (6199), 865-867.

45. Pauls, E.; Badia, R.; Torres-Torronteras, J.; Ruiz, A.; Permanyer, M.; Riveira-Munoz, E.; Clotet, B.; Marti, R.; Ballana, E.; Este, J. A., Palbociclib, a selective inhibitor of cyclin-dependent kinase 4/6, blocks HIV-1 reverse transcription through the control of sterile α motif and HD domain-containing protein-1 (SAMHD1) activity. *AIDS* **2014,** *28* (15), 2213-22.

46. Shortridge, M. D.; Vidalala, V.; Varani, G., The kinase inhibitor Palbociclib is a potent and specific RNA-binding molecule. *Available on bioRxiv* **2022**, https://doi.org/10.1101/2022.01.20.477126.

47. Munshi, S.; Neupane, K.; Ileperuma, S. M.; Halma, M. T. J.; Kelly, J. A.; Halpern, C. F.; Dinman, J. D.; Loerch, S.; Woodside, M. T., Identifying inhibitors of -1 programmed ribosomal frameshifting in a broad spectrum of coronaviruses. *Viruses* **2022,** *14* (2), 177.

48. Fang, L.; Velema, W. A.; Lee, Y.; Xiao, L.; Mohsen, M. G.; Kietrys, A. M.; Kool, E. T., Pervasive transcriptome interactions of protein-targeted drugs. *Nat. Chem.* **2023,** *15* (10), 1374-1383.

49. Comşa, Ş.; Cîmpean, A. M.; Raica, M., The story of MCF-7 breast cancer cell line: 40 years of experience in research. *Anticancer Res.* **2015,** *35* (16), 3147-3154.

50. Belkevich, A. E.; Pascual, H. G.; Fakhouri, A. M.; Ball, D. G.; Knutson, B. A., Distinct interaction modes for the eukaryotic RNA polymerase alpha-like subunits. *Mol. Cell Biol.* **2023,** *43* (6), 269-282.

51. Egusquiaguirre, S. P.; Liu, S.; Tosic, I.; Jiang, K.; Walker, S. R.; Nicolais, M.; Saw, T. Y.; Xiang, M.; Bartel, K.; Nelson, E. A.; Frank, D. A., CDK5RAP3 is a co-factor for the oncogenic transcription factor STAT3. *Neoplasia* **2020,** *22* (1), 47-59.

52. Bajusz, D.; Racz, A.; Heberger, K., Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **2015,** *7*, 20.

53. Liu, S.; Zhuo, L.; Wang, J.; Zhang, Q.; Li, Q.; Li, G.; Yan, L.; Jin, T.; Pan, T.; Sui, X.; Lv, Q.; Xie, T., METTL3 plays multiple functions in biological processes. *Am. J. Cancer Res.* **2020,**

10 (6), 1631-1646.

54. Gamage, N.; Barnett, A.; Hempel, N.; Duggleby, R. G.; Windmill, K. F.; Martin, J. L.; McManus, M. E., Human sulfotransferases and their role in chemical metabolism. *Toxicol. Sci.* **2006,** *90* (1), 5-22.

55. Bairam, A. F.; Rasool, M. I.; Alherz, F. A.; Abunnaja, M. S.; El Daibani, A. A.; Gohal, S. A.; Alatwi, E. S.; Kurogi, K.; Liu, M. C., Impact of SULT1A3/SULT1A4 genetic polymorphisms on the sulfation of phenylephrine and salbutamol by human SULT1A3 allozymes. *Pharmacogenet. Genomics* **2019,** *29* (5), 99-105.

56. Hernando, B.; Pena-Chilet, M.; Ibarrola-Villava, M.; Martin-Gonzalez, M.; Gomez-Fernandez, C.; Ribas, G.; Martinez-Cadenas, C., Genetic 3'UTR variation is associated with human pigmentation characteristics and sensitivity to sunlight. *Exp. Dermatol.* **2017,** *26* (10), 896-903.

57. Zeng, C.; Huang, W.; Li, Y.; Weng, H., Roles of METTL3 in cancer: Mechanisms and therapeutic targeting. *J. Hematol. Oncol.* **2020,** *13* (1), 117.

58. Zou, J.; Li, H.; Huang, Q.; Liu, X.; Qi, X.; Wang, Y.; Lu, L.; Liu, Z., Dopamine-induced SULT1A3/4 promotes EMT and cancer stemness in hepatocellular carcinoma. *Tumor Biol.* **2017,** *39* (10).

59. Xiao, F.; Zhang, P.; Wang, Y.; Tian, Y.; James, M.; Huang, C. C.; Wang, L.; Wang, L., Single-nucleotide polymorphism rs13426236 contributes to an increased prostate cancer risk via regulating MLPH splicing variant 4. *Mol. Carcinog.* **2020,** *59* (1), 45-55.

60. Liao, H.; Gaur, A.; McConie, H.; Shekar, A.; Wang, K.; Chang, J. T.; Breton, G.; Denicourt, C., Human NOP2/NSUN1 regulates ribosome biogenesis through non-catalytic complex formation with box C/D snoRNPs. *Nucleic Acids Res.* **2022,** *50* (18), 10695-10716.

61. Cui, W.; Pizzollo, J.; Han, Z.; Marcho, C.; Zhang, K.; Mager, J., Nop2 is required for mammalian preimplantation development. *Mol. Reprod. Dev.* **2016,** *83* (2), 124-31.

62. Wang, F.; Yuan, J. H.; Wang, S. B.; Yang, F.; Yuan, S. X.; Ye, C.; Yang, N.; Zhou, W.

P.; Li, W. L.; Li, W.; Sun, S. H., Oncofetal long noncoding RNA PVT1 promotes proliferation and stem cell-like property of hepatocellular carcinoma cells by stabilizing NOP2. *Hepatology* **2014,** *60* (4), 1278-90.

63. Warth, B.; Palermo, A.; Rattray, N. J. W.; Lee, N. V.; Zhu, Z.; Hoang, L. T.; Cai, Y.; Mazurek, A.; Dann, S.; VanArsdale, T.; Fantin, V. R.; Shields, D.; Siuzdak, G.; Johnson, C. H., Palbociclib and fulvestrant act in synergy to modulate central carbon metabolism in breast cancer cells. *Metabolites* **2019,** *9* (1).

64. Lanceta, L.; O'Neill, C.; Lypova, N.; Li, X.; Rouchka, E.; Waigel, S.; Gomez-Gutierrez, J. G.; Chesney, J.; Imbert-Fernandez, Y., Transcriptomic profiling identifies differentially expressed genes in palbociclib-resistant ER+ MCF7 breast cancer cells. *Genes (Basel)* **2020,** *11* (4).

65. Hanning, J. E.; Saini, H. K.; Murray, M. J.; van Dongen, S.; Davis, M. P.; Barker, E. M.; Ward, D. M.; Scarpini, C. G.; Enright, A. J.; Pett, M. R.; Coleman, N., Lack of correlation between predicted and actual off-target effects of short-interfering RNAs targeting the human papillomavirus type 16 E7 oncogene. *Br. J. Cancer* **2013,** *108* (2), 450-60.

66. Culjkovic-Kraljacic, B.; Fernando, T. M.; Marullo, R.; Calvo-Vidal, N.; Verma, A.; Yang, S.; Tabbo, F.; Gaudiano, M.; Zahreddine, H.; Goldstein, R. L.; Patel, J.; Taldone, T.; Chiosis, G.; Ladetto, M.; Ghione, P.; Machiorlatti, R.; Elemento, O.; Inghirami, G.; Melnick, A.; Borden, K. L.; Cerchietti, L., Combinatorial targeting of nuclear export and translation of RNA inhibits aggressive B-cell lymphomas. *Blood* **2016,** *127* (7), 858-68.

67. Harden, M. E.; Prasad, N.; Griffiths, A.; Munger, K., Modulation of microRNA-mRNA target pairs by human papillomavirus 16 oncoproteins. *mBio* **2017,** *8* (1).

68. Thompson, P. A.; Eam, B.; Young, N. P.; Fish, S.; Chen, J.; Barrera, M.; Howard, H.; Sung, E.; Parra, A.; Staunton, J.; Chiang, G. G.; Gerson-Gurwitz, A.; Wegerski, C. J.; Nevarez, A.; Clarine, J.; Sperry, S.; Xiang, A.; Nilewski, C.; Packard, G. K.; Michels, T.; Tran, C.; Sprengeler, P. A.; Ernst, J. T.; Reich, S. H.; Webster, K. R., Targeting oncogene mRNA

translation in B-cell malignancies with eFT226, a potent and selective inhibitor of eIF4A. *Mol. Cancer Ther.* **2021,** *20* (1), 26-36.

69. Chittapragada, M.; Roberts, S.; Ham, Y. W., Aminoglycosides: Molecular insights on the recognition of RNA and aminoglycoside mimics. *Perspect. Medicin. Chem.* **2009,** *3*, 21-37.

70. Balaratnam, S.; Rhodes, C.; Bume, D. D.; Connelly, C.; Lai, C. C.; Kelley, J. A.; Yazdani, K.; Homan, P. J.; Incarnato, D.; Numata, T.; Schneekloth, J. S., Jr., A chemical probe based on the PreQ(1) metabolite enables transcriptome-wide mapping of binding sites. *Nat. Commun.* **2021,** *12* (1), 5856.

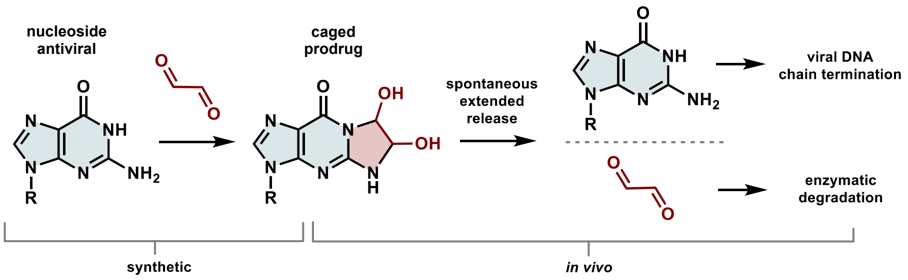
71. Piezzo, M.; Cocco, S.; Caputo, R.; Cianniello, D.; Gioia, G. D.; Lauro, V. D.; Fusco, G.; Martinelli, C.; Nuzzo, F.; Pensabene, M.; De Laurentiis, M., Targeting cell cycle in breast cancer: CDK4/6 inhibitors. *Int. J. Mol. Sci.* **2020,** *21* (18).

72. Al-Qasem, A. J.; Alves, C. L.; Ehmsen, S.; Tuttolomondo, M.; Terp, M. G.; Johansen, L. E.; Vever, H.; Hoeg, L. V. A.; Elias, D.; Bak, M.; Ditzel, H. J., Co-targeting CDK2 and CDK4/6 overcomes resistance to aromatase and CDK4/6 inhibitors in ER+ breast cancer. *NPJ Precis. Oncol.* **2022,** *6* (1), 68.

Chapter 4: Glyoxal-Caged Nucleoside Antivirals as Extended-Release Prodrugs

4.1 Abstract

Viral infections pose a significant public health threat, manifesting in acute pandemic events such as the COVID-19 outbreak and long-lasting endemic challenges such as HIV prevalence in sub-Saharan Africa. Nucleoside antivirals, including viral DNA polymerase inhibitors and nucleoside reverse transcriptase inhibitors (NRTIs), are a leading class of compounds often prescribed as a first-line treatment for viral infections. However, inherent limitations of antiviral nucleoside drugs such as low solubility and circulation lifetime sometimes necessitate multi-intraday dosing. A promising solution to address this challenge is the elaboration of currently approved therapeutics into prodrug moieties, as simple modifications of existing antiviral drugs could simultaneously enhance pharmacokinetic properties and promote an extended-release mechanism to combat poor patient adherence. Here, we deploy the 1,2-dialdehyde glyoxal to generate antiviral nucleoside prodrugs that undergo spontaneous self-activation, providing a mechanism for timed control over viral DNA polymerase and RT activity. As proof of principle, we show that the near-quantitative reaction of glyoxal with acyclovir (ACV) drastically improves ACV solubility and enables subsequent release of the drug with a half-life of 2.3 h. Further, glyoxal caging reversibly disrupts ACV activity against HIV-1 *in vitro* and HSV-1 *in cellulo*. Finally, we demonstrate the amenability of a panel of NRTIs to glyoxal caging, showcasing the potential of this highly versatile method for achieving timed-release activation of a clinically important class of antiviral drugs.



4.2 Introduction

The risk of global health threats from viral infections such as SARS-CoV-2 and HIV-1, combined with the rising prevalence of antiviral drug resistance, poses a significant biomedical challenge. Among the most effective treatments for viral infections are nucleoside analogues, which mimic naturally occurring nucleosides and thereby cause selective termination of the nascent DNA chain during viral DNA replication or reverse transcription.[1-2] Despite their efficacy in antiviral therapy, many nucleoside analogues have a short mean plasma half-life and poor solubility, leading to intravenous (IV) dosing regimens up to five times per day.[3-5] As a result, patient adherence to therapeutic requirements is often inadequate, in turn compromising patient outcomes and contributing to drug resistance.[6-8] Current approaches to overcome these challenges have primarily focused on the development of novel pharmaceuticals. However, these efforts require extensive time and resources, and discoveries are specific to each chemical entity.

Extended-release antiviral therapies offer a promising solution to combat the challenges that lead to poor drug adherence, as they can obviate the need for multi-intraday dosing.[9] Indeed, the identification of prodrug moieties acting upon the base-pairing face may provide a general approach to creating extended-release versions of a broad range of existing nucleoside-based pharmaceuticals. Prodrug moieties that are amenable to an extended-release mechanism may also impart dual functionality by improving pharmacokinetic properties such as solubility and/or circulation lifetime. This approach is particularly attractive for viral DNA polymerase inhibitors and nucleoside reverse transcriptase inhibitors (NRTIs), as the therapeutic window for many of these drugs is limited by solubility and permeability constraints, contributing to undesirable side effects including nephrotoxicity.[10]

Herein, we report a novel prodrug strategy that uses the simplest dialdehyde, glyoxal, to generate nucleoside antivirals that undergo spontaneous activation over several hours. Glyoxal is highly electron deficient and thus susceptible under mild aqueous conditions to nucleophilic addition from amines such as those found in cytosine, adenine, and especially guanine nucleobases.[11-14] Our lab recently deployed glyoxal caging to thermoreversibly tune the activity of a wide range of oligonucleotides, including aptamers, gRNAs, and antisense oligonucleotides.[15] Here, we extend this principle to nucleoside-based small-molecule therapeutics, characterizing the reactivity of glyoxal with the guanosine analogue acyclovir (ACV), as well as using glyoxal to modulate ACV bioactivity in the context of viral DNA polymerase and reverse transcriptase inhibition. Finally, we show that a wider panel of antiviral nucleoside compounds is also susceptible to glyoxal caging, underscoring the generalizability of this approach toward enhancing the pharmacokinetic properties of an important class of antiviral drugs.

4.3 Results and Discussion

4.3.1 Reaction of glyoxal with ACV

ACV is a nucleoside antiviral that serves as a structural mimetic of guanosine, undergoing intracellular phosphorylation before binding to herpes simplex virus type 1 (HSV-1) DNA polymerase and other viral DNA polymerases to induce chain termination and inhibit viral activity.[16] However, the poor solubility, permeability, and bioavailability of ACV limit its therapeutic potential.[17-18] Indeed, ACV has a maximum solubility of 2.5 mg/mL at pH 7, suffers from a short mean plasma half-life (2.5 h), and induces nephropathy in 5–10% of patients receiving IV-administered treatment due to the precipitation of needle-shaped crystals.[3, 10, 19-20] Envisioning that glyoxal caging could improve these properties, we first explored the reactivity of ACV with 10 equivalents of glyoxal in pH 6 buffer, analyzing reaction progress over 24 h by HPLC (Figure 4.1a). We found that caging proceeded rapidly, leading to the complete consumption of ACV and the formation of a new major product in <1 h (Figure 4.1b). A similar experiment conducted with 0.8 or 1 equivalents of glyoxal relative to ACV revealed near-quantitative conversion of the starting material into a product having two new signals, corresponding to hemiaminal protons of the caged ACV species, by ^{1}H NMR in D_2O (Figure C1). These peaks are both singlets, consistent with a

90° dihedral angle, in accordance with the Karplus plot.[21] The ^{13}C NMR spectrum also indicated formation of the expected glyoxal-ACV species (Figure C2). Finally, the expected product was confirmed by high-resolution ESI-MS (Figure C3).

We then proceeded to assess the decaging kinetics of the glyoxal-modified species. Upon incubation of 10 µM caged ACV in 1X PBS (pH 7.4) at 37 ºC for 1–24 h, equilibrium increasingly favored disassociation of the glyoxal cage and regeneration of the parent ACV compound (Figure 4.1c). An analogous experiment conducted over 48 h with 10 µM–1 mM caged ACV revealed that at lower concentration, the fraction of decaged product increases with a slower half-life (Figure 4.1d). These data provide encouraging evidence that in a therapeutic context, low intracellular concentrations of caged ACV would deliver the active antiviral compound over a time span of several hours, consistent with other extended-release systems.[3]

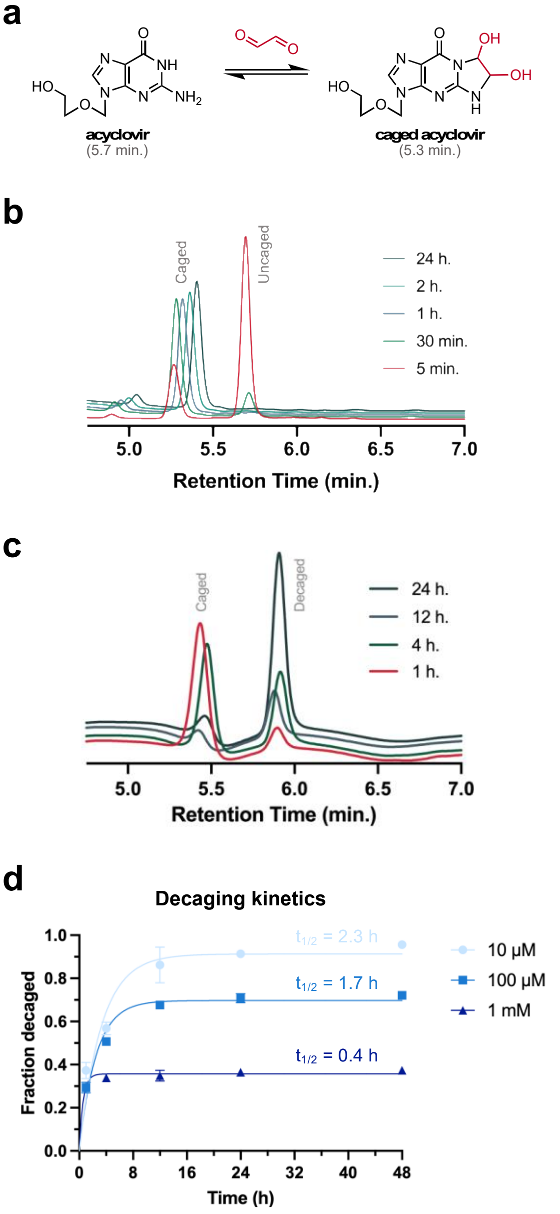
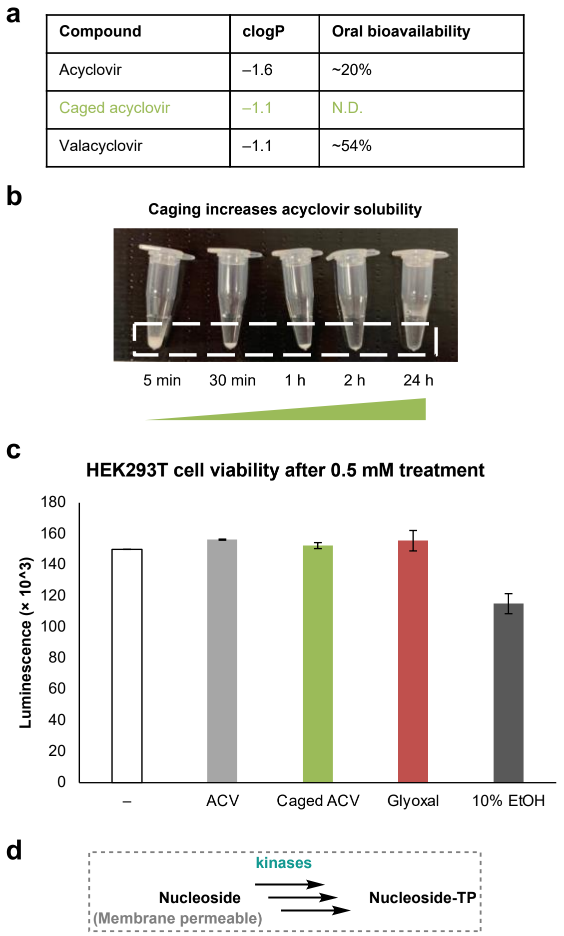


Figure 4.1: Characterization of ACV caging and decaging with glyoxal. **a** ACV caging scheme. **b** Conversion of ACV (100 mM) into caged ACV by HPLC after 5 min–24 h treatment with 10 equivalents of glyoxal in ammonium acetate buffer (10 mM, pH 6) at 50 ºC. **c** Reconversion of purified, caged ACV (10 µM) into ACV by HPLC after 1–24 h incubation in 1X PBS (pH 7.4) at 37 ºC. **d** Effects of caged ACV concentration on decaging rate and fraction converted over 48 h (n = 2). Error bars indicate mean and standard deviation.

4.3.2 Pharmacokinetic effects of ACV caging

We anticipate that glyoxal caging will provide several benefits to the pharmacokinetic properties of ACV. While the oral bioavailability of caged ACV has not yet been measured, we determined the calculated logP (clogP) values of

ACV and caged ACV to be –1.6 and –1.1, respectively. Meanwhile, the L-valyl ester of ACV, valacyclovir, also has a clogP value of –1.1 and was developed for its significantly increased oral bioavailability compared to that of ACV (Figure 4.2a).[22-23] These preliminary calculations support our hypothesis that glyoxal caging of nucleoside antivirals may increase the bioavailability of these pharmaceuticals by enhancing membrane permeability. Likewise, glyoxal caging has a striking impact on compound solubility. ACV has a maximum solubility of 2.5 mg/mL,[19] but 24 h incubation of 23 mg of ACV with a single equivalent of glyoxal in 1 mL of 1X PBS enabled complete solubilization of the precipitate, representing a near 10-fold improvement in solubility (Figure 4.2b). We further found that 4 h treatment of HEK293T cells with 0.5 mM ACV, caged ACV, or glyoxal had no measurable effect on cell viability (Figure 4.2c), nor was there an impact with up to 20 mM caged ACV (Figure C4). We envision that with putatively enhanced membrane permeability relative to ACV, caged ACV could facilitate delivery of the pharmaceutical intracellularly before subsequent phosphorylation by viral and endogenous kinases and time-controlled release to the active antiviral (Figure 4.2d).

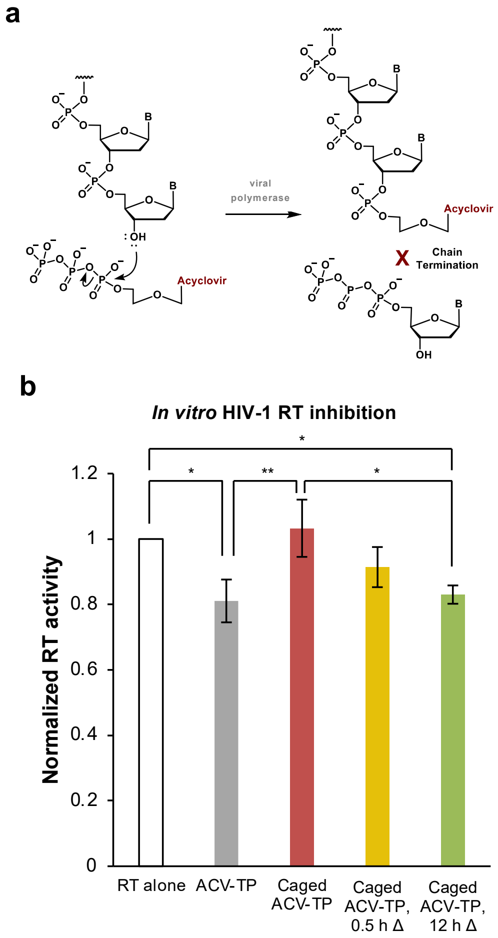


Compound	clogP	Oral bioavailability
Acyclovir	–1.6	~20%
Caged acyclovir	–1.1	N.D.
Valacyclovir	–1.1	~54%

Figure 4.2: Effects of glyoxal caging on the pharmacokinetic properties of ACV. a Calculated logP values and measured oral bioavailability of ACV, glyoxal-caged ACV, and valacyclovir. N.D., no data. **b** Ten-fold increase in solubility of ACV (100 mM) upon 5 min–24 h incubation with 1 equivalent of glyoxal in 1X PBS (pH 7.4). **c** HEK293T cell viability upon 4 h treatment with 0.5 mM ACV, 0.5 mM caged ACV, 0.5 mM glyoxal, or 10% ethanol control (n = 2). Error bars indicate mean and standard deviation. **d** Proposed scheme for caged ACV prodrug delivery to cells and intracellular phosphorylation to active antiviral compound.

4.3.3 Thermoreversible ACV bioactivity

To probe the effects of glyoxal caging on drug bioactivity, our next studies involved ACV-triphosphate (ACV-TP). In the treatment of HSV-1, therapeutically delivered ACV is thought to be specifically converted to the monophosphate by virally encoded thymidine kinase, followed by generation of ACV-TP via endogenous cellular kinases.[16, 24] ACV-TP then selectively binds the viral DNA polymerase, inducing termination of the nascent DNA chain (Figure 4.3a). Interestingly, although ACV is prescribed primarily for the treatment of herpesviruses, clinical data suggest that in co-infected patients, ACV has a beneficial therapeutic effect on HIV-1 as well.[25-27] One study investigating this phenomenon concluded that the presence of herpesvirus kinases is necessary for ACV monophosphorylation before conversion into ACV-TP, a species that directly inhibits HIV-1 RT.[28] However, separate studies found that herpesviruses may not be required for ACV phosphorylation and downstream HIV-1 suppression.[26, 29] We leveraged a commercially available HIV-1 RT assay kit to directly explore the effects of glyoxal caging on ACV-TP inhibition activity. After confirming that ACV-TP exhibits comparable caging and decaging behavior to ACV (Figure C5), we measured the *in vitro* activity of 1 ng RT alone or in the presence of 100 μM ACV-TP, caged ACV-TP, or caged



c

Compound	EC_{50} (μM)
ACV	0.97
Caged ACV	5.53
Caged ACV, 1 d Δ	1.69

Figure 4.3: *In vitro* and *in cellulo* antiviral activity of ACV after glyoxal caging and decaging. a Schematic depicting the mechanism of ACV-TP incorporation into nascent DNA strands by viral polymerases. **b** Normalized *in vitro* activity of HIV-1 RT upon 100 μM treatment with ACV-TP, caged ACV-TP, or caged ACV-TP after 0.5 h or 12 h decaging at 1 mM in 1X PBS (pH 6.8) at 37 ºC (n = 3). One-way ANOVA (95% confidence limit); post-hoc test (Tukey HSD), $^*p < 0.05$, $^{**}p < 0.01$. Error bars indicate mean and standard deviation. **c** Half-maximal effective concentration (EC50) measured for ACV, caged ACV, and caged ACV after 1 d decaging at 1 mM concentration in an HSV-1 cytopathic effects (CPE) assay. Human foreskin fibroblast cells were infected with HSV-1 and treated for 5 d with 0.048–150 μM ACV or 0.003–10 μM caged and decaged ACV species. Compound EC50 values were interpolated from experimental data and calculated relative to cells exposed to HSV-1 alone, resulting in 100% CPE.

ACV-TP that had been pre-heated in 1X PBS for 0.5 or 12 h to promote decaging (Figure 4.3b). While we observed only modest potency of ACV-TP against HIV-1 RT, caging fully disrupted this activity. More excitingly, ACV-TP activity was restored upon decaging in a time-dependent manner. In line with our hypothesis, these results indicate that glyoxal caging can be used to deactivate and spontaneously reactivate the biological effects of nucleoside antivirals while improving pharmacokinetic properties such as solubility.

We then conducted preliminary *in cellulo* and *in vivo* tests to further characterize these properties. To directly probe the relative activity of ACV vs. caged and decaged ACV against the canonical viral target, HSV-1, we performed a cytopathic effects (CPE) assay that measures viral-mediated reduction in cell growth and corresponding rescue by antiviral agents.[30] Human foreskin fibroblast cells were treated with varying concentrations of ACV, caged ACV, and caged ACV that first underwent 1 d decaging at 37 ºC in 1X PBS. The cells were then infected with HSV-1 and further incubated for 5 d until 100% CPE was observed in virus-only control wells. The half-maximal effective concentration (EC50) values at which viral activity was inhibited for the various compound treatments could then be extrapolated (Figure 4.3c, Table C1).[30] As expected, ACV demonstrated strong potency against HSV-1 with an EC50 of 0.97 µM. Meanwhile, this potency was reduced 5.7-fold for the caged compound. We speculate that the reduction in antiviral activity for caged ACV was not more pronounced due to the required length of the assay, during which time a proportion of caged ACV likely equilibrated back to the parent compound at 37 ºC. However, the EC50 value for the decaged species lay in between that of ACV and caged ACV, suggesting that the loss of antiviral activity due to glyoxal modification was indeed reversible during the 1 d decaging period. This experiment highlights the timed-release nature of caged ACV activity against HSV-1 in a biologically relevant context.

As final validation of our ACV glyoxal caging approach, we administered IV bolus dosing of 1 mg/kg ACV or caged ACV to mice (n = 3) and measured compound plasma concentration over time by LC-MS/MS. Preliminary results indicated that the initial concentration of caged ACV was much higher than that of ACV, and further that detectible caged ACV circulated in the bloodstream for 4 h, as opposed to 1 h for ACV (Figure C6). A separate study assessing compound stability in purified mouse plasma confirmed by LC-MS/MS that caged ACV indeed converts back into ACV over time (data not shown). Together, these studies showcase caged ACV as a promising, facilely synthesized ACV prodrug that retains timed-release antiviral activity in biologically relevant contexts.

4.3.4 NRTI substrate scope

Lastly, we sought to explore whether nucleoside antivirals beyond ACV could be similarly modified with glyoxal for potential therapeutic applications. Considering the prevalence of multidrug combinations in antiviral therapy and specifically the use of multiple NRTIs in combination,[31] information on the substrate scope for our glyoxalation strategy is particularly important for future therapeutic applications. We selected a panel of three NRTIs to test: entecavir, zalcitabine, and abacavir, which are guanosine, cytidine, and adenosine analogs, respectively.[32] The compounds were analyzed by HPLC before and after caging at 1 mM concentration with varying equivalents of glyoxal. Entecavir, a guanosine analog like ACV, reacted the most readily with glyoxal, requiring only 10 equivalents to achieve >99% conversion of the starting material (Figure 4.4a). Zalcitabine reacted fully with 1,000 equivalents to form a significantly more polar product (Figures 4b), and abacavir reacted 85% with 1,000 equivalents to form three new species (Figure 4.4c). At first puzzled by the multiple peaks formed from the reaction of glyoxal with entecavir and abacavir, we postulate that glyoxal is adding to either of two faces of the entecavir guanidine moiety or, likewise, two faces of the abacavir guanidine moiety as well as the amidine group (Figure C7). The formation of multiple species does not preclude the use of the caged NRTI as a therapeutic, as long as the species are isolated by HPLC and individually tested for extended-release antiviral activity in the future. High-resolution ESI-MS analysis of each caged NRTI reaction mixture confirmed the formation of products consistent with

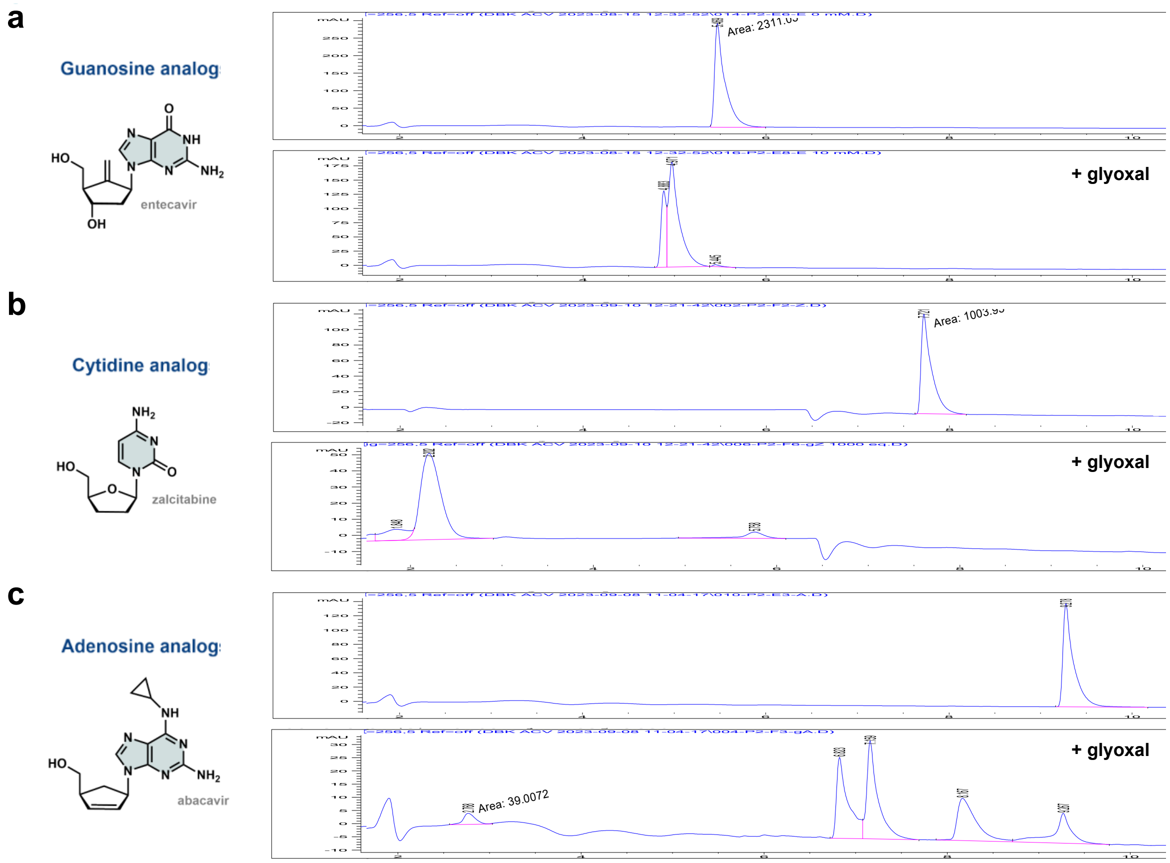


Figure 4.4: Glyoxal caging of a panel of structurally diverse NRTIs. a Entecavir (1 mM) was analyzed by HPLC before (top) and after (bottom) 1 d incubation with 10 equivalents of glyoxal in ammonium acetate buffer at 50 ºC. **b** Zalcitabine and **c** abacavir were similarly analyzed after 1 d incubation with 1000 equivalents of glyoxal.

the addition of glyoxal (Figure C8). Thus, a wide array of nucleoside-like small molecule drugs are amenable to glyoxal caging, representing an innovative and efficacious strategy to iterate off previously developed pharmaceuticals while improving pharmacokinetic parameters and imparting thermoresponsiveness.

4.4 Conclusion

The work presented here aims to combat low antiviral drug adherence through the glyoxal-based caging of antiviral therapeutics, which may provide a mechanism for extended release *in vivo* by temporarily blocking the Watson-Crick-Franklin base pairing face of the compound. This strategy is generalizable to diverse nucleoside mimetics with varying nucleobase and “sugar” structures. Additionally, we found that glyoxal caging may enhance pharmacokinetic properties, such as solubility and permeability, which are accepted as limitations within this drug class.[10] These improvements would reduce dosing frequency and thus enable superior patient outcomes. Advantageously, the reverse decaging reaction is spontaneous under physiological conditions and does not require exogenous reagents for activation.[33]

As prior work in our lab utilized glyoxal to modulate the function of diverse nucleic acid scaffolds,[15] we recognized the potential to extend this principle to nucleoside-like small molecules. We first characterized the glyoxal caging and decaging kinetics of acyclovir, a competitive inhibitor of HSV-1 DNA polymerase. We found that caging considerably enhanced ACV solubility and impacted both *in vitro* and *in cellulo* antiviral activity in a thermoreversible manner. The glyoxal caging method also demonstrated a broad substrate scope, extending to a wider panel of guanosine-, adenosine-, and cytosine-like NRTIs. Thus, this design is innovative in the use of glyoxal as a simple and generalizable approach to generate extended-release antivirals. These discoveries have implications that may greatly advance both the efficacy and levels of adherence to medicines used in the treatment of a variety of acute viral diseases.

4.5 Materials and Methods

4.5.1 General information

All reagents were used as purchased from their chemical manufacturer. Reagents were purchased from MilliporeSigma unless otherwise specified. All reactions were run under ambient atmosphere. ^{1}H NMR and ^{13}C NMR spectra were collected using Varian INOVA 400 MHz, Bruker AVANCE III HD 600 MHz, Bruker NANO HD III 400 MHz, or Agilent DD2 500 MHz spectrometers and processed by MestReNova, with chemical shifts (δ) reported in ppm. ^{1}H NMR samples were prepared by adding 100 µL of the crude reaction mixture to 600 µL of D_2O, and ^{13}C NMR samples were prepared by adding 650 µL of the crude reaction mixture to 50 µL of D_2O. Mass spectra were acquired on a Bruker MaXis Quadrupole Time-of-Flight (QTOF) mass spectrometer in positive-ion mode. pH values were measured using a Mettler Toledo SevenCompact S220 pH Meter. Computational logP calculations were performed using the ChemDraw Professional software package. Data analysis was performed on GraphPad Prism 9 software.

HPLC chromatograms were collected using an Agilent 1260 Infinity II LC System equipped with autosampler. Acyclovir (ACV) samples were prepared for HPLC by adding 10 µL of the crude reaction mixture to 90 µL of 0.1% formic acid in water. Samples were eluted over a 10-minute gradient mobile phase of 100:0 to 80:20 of 0.1% formic acid in water:acetonitrile at a flow rate of 0.6 mL/min on a 3 µm, 4.6 × 75 mm Waters Atlantis T3 analytical column heated to 30 °C with an injection volume of 2 µL (200 µL/min draw speed, 400 µL/min eject speed), visualizing by UV absorbance at 256 ± 5 nm. ACV-triphosphate (ACV-TP) and nucleoside reverse transcriptase inhibitor (NRTI) samples were prepared for HPLC by diluting ~2.5 nmol material up to 50 µL with 1X PBS (Thermo Fisher Scientific, adjusted to pH 6.8). ACV-TP samples were eluted over 3 minutes with isocratic 20 mM potassium phosphate buffer (pH 7) at a flow rate of 0.8 mL/min on a 2.7 µm, 4.6 × 150 mm Agilent Poroshell 120 EC-C18 analytical column run at room temperature with an injection volume of 45 µL and otherwise identical instrument settings to ACV samples. Entecavir and abacavir samples were eluted over a 10-minute gradient mobile phase of 98:2 to 70:30 of potassium phosphate buffer:acetonitrile, with otherwise identical instrument settings to ACV-TP samples. Zalcitabine samples were eluted over 3 minutes with isocratic potassium phosphate buffer, followed by a 7-minute gradient mobile phase to 70:30 of potassium phosphate buffer:acetonitrile, with otherwise identical instrument settings to ACV-TP samples.

4.5.2 General method to cage nucleoside antivirals

Stock solutions (10 mM) of each nucleoside drug were prepared in DMSO. In a PCR tube, 2.5 µL of the 10 mM stock solution was combined with 2.8 µL of a 40% glyoxal solution (~8.8 M), 2.5 µL of 100 mM ammonium acetate buffer (pH 6), and 17.2 µL of nuclease-free water to 25 µL final volume. Reactions alternately proceeded with 2.5 µL of 10X PBS (Thermo Fisher Scientific) with pH adjusted to 6, or 6.8 in the case of ACV-TP to preserve compound stability. Reactions were heated at 50 ºC for 5 min–2 d, as delineated in figure captions, or 40 ºC in the case of ACV-TP. Final antiviral and glyoxal concentrations also varied by experiment between 1–100 mM and 0.01–1 M, respectively, as delineated in figure captions. ACV-TP was purchased from Jena Bioscience. Zalcitabine was purchased from APExBIO.

4.5.3 General method to decage nucleoside antivirals

Antiviral that had been reacted with the minimal equivalent of glyoxal to achieve full caging by HPLC was diluted to 0.01–1 mM in 1X PBS (pH 7.4, or 6.8 in the case of ACV-TP). Decaging was performed at 37 ºC for 0.5 h–2 d, as delineated in figure captions.

4.5.4 Cell viability assays

All cell viability experiments were performed using HEK293T cells. Cells were maintained in DMEM (High Glucose, Pyruvate, Gibco) supplemented with 10% Fetal Bovine Serum (Gibco) at 37 ºC in a humidified atmosphere of 5% CO_2 and 95% air. Cells were passaged at 80% confluence. For viability assays, cells were seeded in 150 µL of media at 10,000 cells/well in a BioCoat™ Poly-D-Lysine, 96-well, White/Clear, TC-treated Microplate (Corning), with triplicate wells for each treatment condition. The cells were allowed to grow for 1–2 d until reaching 50–80% confluence. The effects of 0.5 mM ACV, caged ACV, and glyoxal on viability were then assessed using the CellTiter-Glo® 2.0 Cell Viability Assay kit (Promega) following the manufacturer protocol. In brief, 60 µL of media was removed from each well and replaced with 10 µL of water, 5 mM ACV, 5 mM caged ACV, 5 mM glyoxal, or 100% ethanol to 100 µL final volume. The plate was incubated at 37 ºC for 3.5 h, followed by 0.5 h incubation at room temperature. Meanwhile, CellTiter-Glo® 2.0 Reagent that had been thawed from –80 ºC to 4 ºC overnight and then equilibrated to room temperature for 1 h was gently inverted to mix and added in 100 µL volume to each well, avoiding light exposure and ATP contamination. The plate was protected from light, mixed on an orbital for 2 min to induce cell lysis, and incubated at room temperature for an additional 10 min to allow stabilization of luminescence signal. Finally, luminescence was recorded on a BioTek Cytation 5 Cell Imaging Multi-Mode Reader with the following instrument settings: Luminescence Endpoint; Integration Time: 1 s; Filter Set 1 (Emission: Full Light; Optics: Top; Gain: 135); Read Speed: Normal; Delay: 100 msec; Extended Dynamic Range; Read Height: 1 mm. Triplicate measurements for each treatment condition were averaged for each experiment and then normalized across two independent biological replicates to account for day-to-day instrument variability and variations in cell confluence. Caged ACV dose response experiments were performed following a similar protocol, except that 120 µL of media was removed from each well and replaced with 20 µL of water, 25% ethanol, or 50, 10, or 2 mM caged ACV to 50 µL final volume, followed by addition of 50 µL CellTiter-Glo® 2.0 Reagent.

4.5.5 Quantification of HIV-1 reverse transcriptase (RT) inhibition

The inhibitory effects of ACV-TP on *in vitro* HIV-1 RT activity and the corresponding influences of glyoxal caging and decaging of ACV-TP were assessed using the HIV-1 Reverse Transcriptase Assay Kit (XpressBio) following the manufacturer protocol. Namely, ACV-TP (10 mM) was diluted in 1X PBS (pH 6.8) to 1 mM, or first caged with glyoxal (12 mM), diluted in 1X PBS to 1 mM, and then further incubated at 37 ºC for 0.5 or 12 h to promote decaging. Then, 12 µL of these solutions or 12 µL duplicates of 1X PBS controls were mixed with 14 µL each of pre-thawed Reaction Buffers 1 and 2. Meanwhile, HIV-1 RT enzyme stock was thawed and briefly vortexed with pre-warmed Lysis Buffer in a 1:399 v:v ratio to produce sufficient enzyme solution for the assay. All 40 µL compound solutions were combined with 80 µL of enzyme solution except for one PBS control sample, which was instead combined with 80 µL of Lysis Buffer as an experimental blank. The final reactions contained 1 ng HIV-1 RT and 100 µM test compound. All solutions were prepared in duplicate. The reactions were incubated at 37 ºC for 20 min in a Bio-Rad T100 Thermal Cycler, and 100 µL were then transferred to a pre-warmed, streptavidin-coated

96-well plate for further incubation at 37 ºC for 20 min. Each well was washed 5X with 230 µL of pre-warmed 1X Wash Buffer and then incubated with 100 µL of pre-warmed HRP Anti-Digoxigenin Conjugate solution at 37 ºC for 45 min. After again washing the plate, 100 µL of ABTS Substrate Solution was added to each well before heating at 37 ºC for 60 min. Absorbance at 405 nm was recorded on a BioTek Cytation 5 Cell Imaging Multi-Mode Reader with the following instrument settings: Detection Method = Absorbance; Read Type = Spectral Scanning; Optics Type = Monochromators. Absorbance values were averaged for each duplicate, with higher absorbance values corresponding to higher HIV-1 RT activity. The experiment was repeated in triplicate and the data normalized both to each other and to the "HIV-1 RT enzyme alone" condition to account for day-to-day instrument variability and directly compare native vs. inhibited enzyme activity, respectively. The five treatment conditions were analyzed by one-way ANOVA and significant differences identified via a post-hoc Tukey HSD test.

4.5.6 Cytopathic effects assay

In cellulo assessment of HSV-1 inhibition by ACV, caged ACV, and decaged ACV was conducted as previously described.[30] In brief, human foreskin fibroblast (HFF) cells derived from human foreskin tissue[34] were seeded at 5,000 cells/well in a 384-well plate supplemented with MEM with Earle's salts, 2% FBS, and standard concentrations of L-glutamine, penicillin, and gentamicin. The cells were cultured for 1 d at 37 ºC in a humidified atmosphere of 5% CO_2 and 95% air to establish a confluent monolayer. ACV was then added to the wells in 5-fold serial dilutions from 0.048–150 µM. Meanwhile, caged ACV was added in 5-fold serial dilutions from 0.0032–10 µM or first decaged at 37 ºC for 1 d at 1 mM concentration in 1X PBS (pH 7.4) before addition to the cells. All treatments were performed in duplicate wells. Cells were then infected with HSV-1 at a multiplicity of infection of ~0.005 plaque forming units (PFU) of virus per cell, as previously described,[35-37] until 100% CPE toward the HFF cells was observed in virus-only control wells after 5 d at 37 ºC. Cytopathology was quantified using CellTiter-Glo reagent (Promega).

Compound concentration resulting in half-maximal reduction of viral replication (EC50) for ACV and the caged and decaged species could then be extrapolated using standard methods in Excel, averaging across the duplicate wells. Notably, although the experiment was designed to evaluate the effect of ACV caging on HSV-1 inhibition, it is likely that during the 5 d viral infection period, a substantial portion of caged ACV equilibrated back to the parent compound, thus resulting in substantial CPE. The assay also included positive and negative control compounds as well as infected and uninfected control wells to ensure valid interpretation of the data. Liquid handling steps were performed on a BioMek 4000 Automated Liquid Handler.

4.5.7 In vivo *circulatory half-life*

The circulatory lifetime of ACV and caged ACV was quantified by injecting three fed, male CD-1 mice with 1 mg/kg compound delivered via intravenous bolus as a 0.2 mg/mL formulation in 10% DMSO, 10% PEG 400, 40% propylene glycol, and 40% water. Plasma samples were collected at 0.083, 0.25, 0.5, 1, 2, 4, 8 and 24 h. An aliquot of 20 µL sample was protein precipitated with 120 µL of 100 ng/mL gemcitabine in acetonitrile as an internal standard. The mixture was vortexed and then centrifuged at 4,000 RPM for 10 min at 4 ºC. Supernatant (100 µL) was transferred to a sample plate, mixed with 50 µL water, and the plate was shaken at 800 RPM for 10 min. Finally, 0.4–1 µL of this solution was injected with a flow rate of 0.6 mL/min onto a 1.7 µm, 2.1 × 50 mm ACQUITY UPLC BEH Amide analytical column. The sample was eluted over a 1.5-minute gradient mobile phase of 95:5 to 40:60 of 0.1% formic acid in acetonitrile:0.1% formic acid in water. ACV was confirmed by ESI-MS with (m/z) calcd for $C_8H_{11}N_5O_3$ ([M + H+]): 226.1, found: 226.1 ([M + H+]). Caged ACV was confirmed by ESI-MS with (m/z) calcd for $C_{10}H_{13}N_5O_3$ ([M + H+]): 284.1, found: 284.1 ([M + H+]). No adverse effects were recorded during the animal study.

4.6 References

1. Seley-Radtke, K. L.; Yates, M. K., The evolution of nucleoside analogue antivirals: A review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold. *Antiviral Res.* **2018,** *154*, 66-86.

2. Holec, A. D.; Mandal, S.; Prathipati, P. K.; Destache, C. J., Nucleotide reverse transcriptase inhibitors: A thorough review, present status and future perspective as HIV therapeutics. *Curr. HIV Res.* **2017,** *15* (6), 411-421.

3. Durai, R. D., Drug delivery approaches of an antiviral drug: A comprehensive review. *Asian J. Pharm.* **2015,** *9* (1), 1-12.

4. Pau, A. K.; George, J. M., Antiretroviral therapy: current drugs. *Infect. Dis. Clin. North Am.* **2014,** *28* (3), 371-402.

5. Lembo, D.; Swaminathan, S.; Donalisio, M.; Civra, A.; Pastero, L.; Aquilano, D.; Vavia, P.; Trotta, F.; Cavalli, R., Encapsulation of acyclovir in new carboxylated cyclodextrin-based nanosponges improves the agent's antiviral efficacy. *Int. J. Pharm.* **2013,** *443* (1-2), 262-72.

6. Iacob, S. A.; Iacob, D. G.; Jugulete, G., Improving the adherence to antiretroviral therapy, a difficult but essential task for a successful HIV treatment-clinical points of view and practical considerations. *Front. Pharmacol.* **2017,** *8*, 831.

7. Barnhart, M., Long-acting HIV treatment and prevention: Closer to the threshold. *Glob. Health Sci. Pract.* **2017,** *5* (2), 182-187.

8. Luccarelli, S. V.; Lucentini, S.; Martellucci, C. A.; Marelli, L.; Sacchi, M.; Nucci, P., Impact of adherence (compliance) to oral acyclovir prophylaxis in the recurrence of herpetic keratitis: Long-term results from a pediatric cohort. *Cornea* **2021,** *40* (9), 1126-1131.

9. Soriano, V.; Barreiro, P.; de Mendoza, C., Long-acting antiretroviral therapy. *Nat. Mater.* **2020,** *19* (8), 826-827.

10. Razonable, R. R., Antiviral drugs for viruses other than human immunodeficiency virus. *Mayo Clin. Proc.* **2011,** *86* (10), 1009-26.

11. Kasai, H.; Iwamoto-Tanaka, N.; Fukada, S., DNA modifications by the mutagen glyoxal: Adduction to G and C, deamination of C and GC and GA cross-linking. *Carcinogenesis* **1998,** *19* (8), 1459-1465.

12. Staehelin, M., Inactivation of virus nucleic acid with glyoxal derivatives. *Biochim. Biophys. Acta* **1959,** *31* (2), 448-54.

13. Shapiro, R.; Cohen, B. I.; Shiuey, S. J.; Maurer, H., On the reaction of guanine with glyoxal, pyruvaldehyde, and kethoxal, and the structure of the acylguanines. A new synthesis of N2-alkylguanines. *Biochemistry* **1969,** *8* (1), 238-245.

14. Nakaya, K.; Takenaka, O.; Horinishi, H.; Shibata, K., Reactions of glyoxal with nucleic acids, nucleotides and their component bases. *Biochem. Biophys. Acta* **1968,** *161* (1), 23-31.

15. Knutson, S. D.; Sanford, A. A.; Swenson, C. S.; Korn, M. M.; Manuel, B. A.; Heemstra, J. M., Thermoreversible control of nucleic acid structure and function with glyoxal caging. *J. Am. Chem. Soc.* **2020,** *142* (41), 17766-17781.

16. O'Brien, J. J.; Campoli-Richards, D. M., Acyclovir: An updated review of its antiviral activity, pharmacokinetic properties and therapeutic efficacy. *Drugs* **1989,** *37*, 233-309.

17. Wagstaff, A. J.; Faulds, D.; Goa, K. L., Aciclovir: A reappraisal of its antiviral activity, pharmacokinetic properties and therapeutic efficacy. *Drugs* **1994,** *41* (1), 153-205.

18. Weller, S.; Blum, M. R.; Doucette, M.; Burnette, T.; Cederberg, D. M.; de Miranda, P.; Smiley, M. L., Pharmacokinetics of the acyclovir pro-drug valaciclovir after escalating single- and multiple-dose administration to normal volunteers. *Clin. Pharmacol. Ther.* **1993,** *54* (6), 595-605.

19. Sawyer, M. H.; Webb, D. E.; Balow, J. E.; Straus, S. E., Acyclovir-induced renal failure. *Am. J. Med.* **1988,** *84* (6), 1067-1071.

20. Karmoker, J. R.; Hasan, I.; Ahmed, N.; Saifuddin, M.; Reza, M. S., Development and optimization of acyclovir loaded mucoadhesive microspheres by Box – Behnken design. *Dhaka Univ. J. Pharm. Sci.* **2019,** *18* (1), 1-12.

21. Karplus, M., Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys.*

1959, *30* (1), 11-15.

22. Balfour, H. H., Antiviral drugs. *N. Engl. J. Med.* **1999,** *340*, 1255-1268.

23. Beutner, K. R., Valacyclovir: a review of its antiviral activity, pharmacokinetic properties, and clinical efficacy. *Antiviral Res.* **1995,** *28* (4), 281-90.

24. Whitley, R. J.; Gnann, J. W., Acyclovir: A decade later. *N. Engl. J. Med.* **1992,** *327*, 782-789.

25. Celum, C.; Wald, A.; Lingappa, J. R.; Magaret, A. S.; Wang, R. S.; Mugo, N.; Mujugira, A.; Baeten, J. M.; Mullins, J. I.; Hughes, J. P.; Bukusi, E. A.; Cohen, C. R.; Katabira, E.; Ronald, A.; Kiarie, J.; Farquhar, C.; Stewart, G. J.; Makhema, J.; Essex, M.; Were, E.; Fife, K. H.; de Bruyn, G.; Gray, G. E.; McIntyre, J. A.; Manongi, R.; Kapiga, S.; Coetzee, D.; Allen, S.; Inambao, M.; Kayitenkore, K.; Karita, E.; Kanweka, W.; Delany, S.; Rees, H.; Vwalika, B.; Stevens, W.; Campbell, M. S.; Thomas, K. K.; Coombs, R. W.; Morrow, R.; Whittington, W. L.; McElrath, M. J.; Barnes, L.; Ridzon, R.; Corey, L., Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. *N. Engl. J. Med.* **2010,** *362* (18), 427-439.

26. Vanpouille, C.; Lisco, A.; Introini, A.; Grivel, J. C.; Munawwar, A.; Merbah, M.; Schinazi, R. F.; Derudas, M.; McGuigan, C.; Balzarini, J.; Margolis, L., Exploiting the anti-HIV-1 activity of acyclovir: Suppression of primary and drug-resistant HIV isolates and potentiation of the activity by ribavirin. *Antimicrob. Agents Chemother.* **2012,** *56* (5), 2604-2611.

27. Perti, T.; Saracino, M.; Baeten, J.; Johnston, C.; Diem, K.; Ocbamichael, N.; Huang, M.-L.; Selke, S.; Magaret, A.; Corey, L.; Wald, A., High-dose valacyclovir decreases plasma HIV-1 RNA more than standard-dose acyclovir in persons coinfected with HIV-1 and HSV-2. *J. Acquir. Immune Defic. Syndr.* **2013,** *63* (2), 201-208.

28. Lisco, A.; Vanpouille, C.; Tchesnokov, E. P.; Grivel, J. C.; Biancotto, A.; Brichacek, B.; Elliott, J.; Fromentin, E.; Shattock, R.; Anton, P.; Gorelick, R.; Balzarini, J.; McGuigan, C.; Derudas, M.; Gotte, M.; Schinazi, R. F.; Margolis, L., Acyclovir is activated into a HIV-1 reverse transcriptase inhibitor in herpesvirus-infected human tissues. *Cell Host Microbe* **2008,** *4* (3), 260-

270.

29. McMahon, M. A.; Parsons, T. L.; Shen, L.; Siliciano, J. D.; Siliciano, R. F., Consistent inhibition of HIV-1 replication in CD4+ T cells by acyclovir without detection of human herpesviruses. *J. Virol.* **2011,** *85* (9), 4618-22.

30. Hartline, C. B.; Keith, K. A.; Eagar, J.; Harden, E. A.; Bowlin, T. L.; Prichard, M. N., A standardized approach to the evaluation of antivirals against DNA viruses: Orthopox-, adeno-, and herpesviruses. *Antiviral Res.* **2018,** *159*, 104-112.

31. Moreno, S.; Perno, C. F.; Mallon, P. W.; Behrens, G.; Corbeau, P.; Routy, J. P.; Darcis, G., Two-drug vs. three-drug combinations for HIV-1: Do we have enough data to make the switch? *HIV Med* **2019,** *20* (4), 2-12.

32. Amblard, F.; Patel, D.; Michailidis, E.; Coats, S. J.; Kasthuri, M.; Biteau, N.; Tber, Z.; Ehteshami, M.; Schinazi, R. F., HIV nucleoside reverse transcriptase inhibitors. *Eur. J. Med. Chem.* **2022,** *240*, 114554.

33. Burnett, W. V., Northern blotting of RNA denatured in glyoxal without buffer recirculation. *BioTechniques* **2018,** *22* (4), 668-670.

34. Prichard, M. N.; Williams, J. D.; Komazin-Meredith, G.; Khan, A. R.; Price, N. B.; Jefferson, G. M.; Harden, E. A.; Hartline, C. B.; Peet, N. P.; Bowlin, T. L., Synthesis and antiviral activities of methylenecyclopropane analogs with 6-alkoxy and 6-alkylthio substitutions that exhibit broad-spectrum antiviral activity against human herpesviruses. *Antimicrob. Agents Chemother.* **2013,** *57* (8), 3518-27.

35. Kern, E. R.; Prichard, M. N.; Quenelle, D. C.; Keith, K. A.; Tiwari, K. N.; Maddry, J. A.; Secrist, J. A., 3rd, Activities of certain 5-substituted 4'-thiopyrimidine nucleosides against orthopoxvirus infections. *Antimicrob. Agents Chemother.* **2009,** *53* (2), 572-9.

36. Toth, K.; Spencer, J. F.; Ying, B.; Tollefson, A. E.; Hartline, C. B.; Richard, E. T.; Fan, J.; Lyu, J.; Kashemirov, B. A.; Harteg, C.; Reyna, D.; Lipka, E.; Prichard, M. N.; McKenna, C. E.; Wold, W. S. M., USC-087 protects Syrian hamsters against lethal challenge with human species

C adenoviruses. *Antiviral Res.* **2018,** *153*, 1-9.

37. Prichard, M. N., Function of human cytomegalovirus UL97 kinase in viral infection and its inhibition by maribavir. *Rev. Med. Virol.* **2009,** *19* (4), 215-29.

Chapter 5: Expanding STEM Accessibility and Inclusivity

5.1 Engagement in Departmental, Local, National, and International Initiatives Promoting Scientific Literacy

Throughout my Ph.D., I have engaged in community-building and scientific outreach efforts alongside my research projects in order to make the field of chemistry more accessible. I firmly believe science should be an open door that any interested party can walk through, as well as a tool we can all wield comfortably when making important personal and professional decisions. I joined the Emory Department of Chemistry to become an expert in chemical biology, but I am leaving with the added goal of shaping the future of STEM education and the STEM workforce through a career in science policy. I learned more than I ever anticipated while serving on the NOBCChE @ Emory executive board, acting as a student liaison to the department through the Student Advisory for Full Engagement (SAFE) Committee, and sharing key lessons from these experiences at national and international platforms.

I am grateful to those I met along the way who share the goals of improving the academic experience and spreading the wonders of scientific discovery. Thank you for the sometimes uncompensated, sometimes unappreciated work you do to further these goals. Thank you to Emory Chemistry for recognizing my work through the Quayle Citizen Scholar Award. Below, I list my contributions to Emory and the wider chemistry community that I am most proud of.

5.1.1 Community building

As a founding executive board member and three-year Outreach Chair (2018–21) on Emory’s newly reestablished chapter of the National Organization for the Professional Advancement of Black Chemists and Chemical Engineers (NOBCChE, https://www.nobcche.org/), I got to play a central role in defining our identity as an organization. I helped plan NOBCChE’s department-wide, end-of-year social events and built working relationships with other Emory Chemistry student groups. Over the first three years, my NSF Graduate Research Fellowship Program (GRFP) professional development series brought together NOBCChE general body members, the wider graduate student body, Emory Chemistry administrators, and faculty to conduct peer reviews of students' fellowship application essays. Amazingly, these discussions resulted in the conferral of six prestigious NSF GRFP awards and four Honorable Mentions to students in our department. In addition, during the founding year of the SAFE Committee, I served as the informal chair, guiding graduate and undergraduate students to work together, solicit feedback from the wider student population, and collaborate with departmental leadership. Continuing prior work, we organized student interviews with faculty candidates and reported the results to Emory Chemistry's Faculty Recruiting & Development Committee. SAFE also brainstormed with the Diversity, Equity, and Inclusion Committee on how to address unmet student needs at the interface of academics and equity.

5.1.2 Outreach

To engage with students in the Atlanta community, I organized participation in NOBCChE's very first event, Marietta High Science Night (Figure 5.1), a tradition that continued five years later. I arranged for NOBCChE's first booth at the Atlanta Science Festival, which due to the pandemic was not realized until 2022. In the interim, I helped the department plan an on-campus, drive-through science demo show to keep kids excited about science during quarantine. Beyond my work in the department, I collaborated with the scientific communication organization Color Me Ph.D. (https://www.colormephd.

Figure 5.1: NOBCChE @ Emory, 2018. Marietta High Science Night and the Emory Chemistry Carnival.

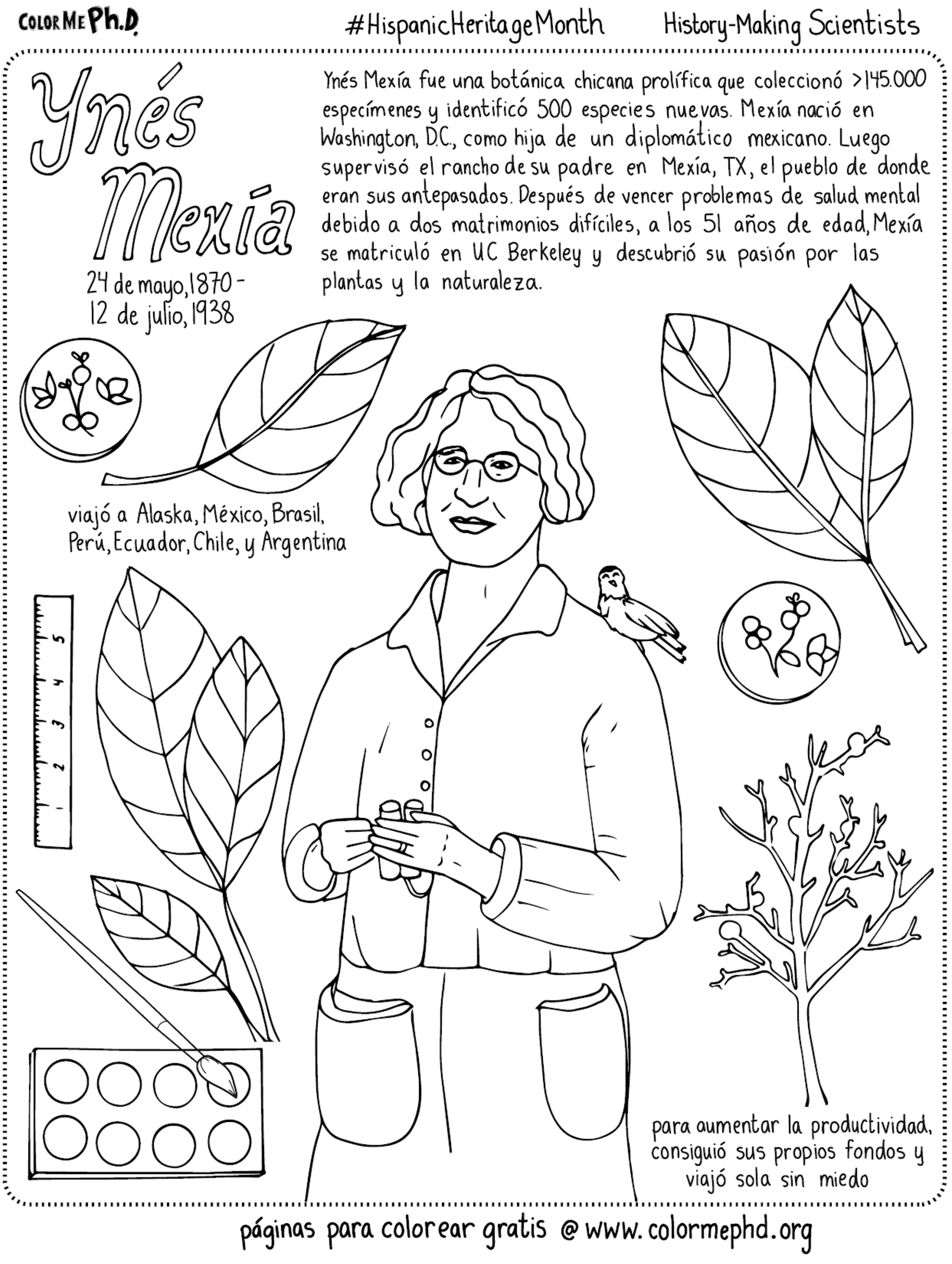
org/) to create their first coloring pages in honor of both Black History Month and Hispanic Heritage Month. I also translated the latter pages into Spanish (Figure 5.2). These pages were part of the "History-Making Scientists" series. Importantly, this work highlights historically underrecognized scientists from history while providing a free and fun educational resource to the public.

5.1.3 Broadening participation

In consultation with the Undergraduate Committee, SAFE recruited two undergraduate students to contribute meaningful student perspectives to committee discussions. Our work during the first year of SAFE laid the foundation for students to be integrated onto almost all department committees in the subsequent year, enabling students to provide input and unique insights during important decision-making moments in the development of the Emory Chemistry Department. In spring 2022, I conceptualized and executed the department's first State of the Department Address & Town Hall event, which invited members of all Emory Chemistry stakeholder groups to learn about current leadership efforts in the department, weigh in on them, and discuss future directions together. The event was held again in spring 2023.

5.1.4 Public education/scholarship

I co-authored the 2020 *Nat. Rev. Chem.* article, "Recruit and retain a diverse workforce," which challenges Emory and our peer academic institutions to take full ownership of the effort and thoughtfulness owed to the researchers from underrepresented groups in STEM whom we often try to attract to our programs, sometimes for selfish reasons. In 2022, I was invited by the Emory Chemistry Director of DEI, Dr. Tracy McGill, to co-present my work on the SAFE Committee at the ACS National Meeting, which sparked a robust discussion with students, faculty, and industry chemists from across the US on how student–faculty partnerships can advance DEI efforts. Since spring 2022, I have served as an Associate Editor for the international *Journal of Science Policy & Governance* in an effort to bolster evidence-based policymaking worldwide.



Ynés Mexía fue una botánica chicana prolífica que coleccionó >145.000 especímenes y identificó 500 especies nuevas. Mexía nació en Washington, D.C., como hija de un diplomático mexicano. Luego supervisó el rancho de su padre en Mexía, TX, el pueblo de donde eran sus antepasados. Después de vencer problemas de salud mental debido a dos matrimonios difíciles, a los 51 años de edad, Mexía se matriculó en UC Berkeley y descubrió su pasión por las plantas y la naturaleza.

viajó a Alaska, México, Brasil, Perú, Ecuador, Chile, y Argentina

para aumentar la productividad, consiguió sus propios fondos y viajó sola sin miedo

páginas para colorear gratis @ www.colormephd.org

Figure 5.2: Color Me Ph.D.'s highlight of Ynés Mexía during National Hispanic Heritage Month. One of the coloring pages we created to promote awareness of underrepresented History-Making Scientists.

5.2 Recruit and Retain a Diverse Workforce[1]

As the world erupts with demands for racial justice, the chemistry community has the obligation, opportunity, and momentum to drive for diversity and inclusion in the sciences. Efforts toward that end must begin by allocating opportunities for success on the basis of potential, not privilege, and follow through by soliciting and acting upon feedback from the scholars we have recruited.

5.2.1 Introduction

The heterogeneity of chemists across ethnic, gender, and socioeconomic spectrums has significance at every echelon, from the young students first discovering the periodic table to the entrepreneurs launching their pharmaceutical startups. Although our scientific community is increasingly engaged in honest dialogue surrounding diversity and inclusion — particularly after the killings of Ahmaud Arbery, Breonna Taylor, and George Floyd — efforts to increase diversity through admissions and hiring have historically focused on meeting racial and gender quotas rather than evaluating candidates holistically. That is, academic programs can exhibit tokenism by recruiting a few members from underrepresented groups in order to meet the lowest standard of diversity, rather than to expand the institution's analytical skillset and capacity for community engagement. Moreover, compared with the time and resources that we expend to recruit scientists from underrepresented groups in science, technology, engineering, and mathematics (STEM), far less attention is given to also creating a work environment that respects all individuals and embraces their feedback. As a result, retention of underrepresented scientists constitutes a major challenge in academia.

What will it take to make our undergraduate and graduate researchers, our postdoctoral interview candidates, our faculty, and our academic leaders reflect local and global populations — and why should we bother? As a melting pot of aspiring innovators, it is both our extraordinary privilege and our duty to answer these questions. We work in a research group whose expertise in enriching functional nucleic acids from a starting pool of trillions of random sequences has, oddly, provided us with a unique perspective on how to select, retain, and invest in the potential of would-be members of our own chemistry community. While our experiences are mostly grounded in the American academic system, here we present general insights on how to recruit diverse groups of people, provide them with the resources to flourish, and support them in dreaming big.

5.2.2 Assessing candidates

Carrying forward our research analogy, the “first law of directed evolution” coined by Frances Arnold states that “you get what you screen for.”[2] Every application process represents a screening step and, if biases (explicit or implicit) are present at each career stage, these accumulate over time to exponentially skew candidate pools. Admissions and hiring committees must therefore look beyond traditional CV metrics to consider both the context of an applicant’s achievements and the value of the knowledge and perspective that they bring to the community. Additionally, we must challenge the mentality, “Does this person have what it takes to be successful in our program?” and instead ask, “Does our program have what it takes for this person to be successful?” This approach is a first step in recognizing the inherent structural biases that make success easier for some than others. It also reframes each new student, postdoc, or faculty member as someone in whom the institution is deeply invested, as opposed to a diversity statistic. In this section, we discuss approaches for reframing admissions and hiring decisions to focus on the potential of each candidate rather than the diversity needs of the institution.

What exactly is diversity? Kenneth Gibbs Jr, a program director in the Division of Training, Workforce Development and Diversity at the National Institute of General Medical Sciences, defines diversity as, “Difference. Science workforce diversity refers to cultivating talent and promoting the full inclusion of excellence across the social spectrum.”[3] In order to cultivate talent in STEM from diverse groups, we should first recognize that excellent students come from all ethnic, religious, socioeconomic and academic backgrounds. Indeed, the ability of coworkers to think in distinct ways fosters creativity and innovation. This concept is exemplified in the acclaimed 2016 film *Hidden Figures*, which is based on the true story of Black NASA mathematicians Katherine Johnson, Mary Jackson, and Dorothy Vaughan and honors their essential roles in making John Glenn the first American to orbit the Earth. Multiple examples in the literature also support the role of diversity in positive scientific outcomes,[4,5] including in industry.

After recognizing the inherent benefits of having varied perspectives, we must next critically evaluate our motivation for soliciting diverse applicants: is it to fulfil social norms? To improve the likelihood of obtaining grant funding? Or is it because we truly believe everyone deserves equal access to educational opportunities? And because we want to invest in people who can provide new insights into our organization’s challenges? Committing to the goal of diversity for the right reasons ensures that the initiatives we pursue are consistent with the desired end goal: positive personal and professional outcomes for all parties involved.

Another major component of selecting among candidates involves evaluating their experience and past achievements. While a head-to-head comparison of publication output, conference talks, grade point average (GPA), or awards is a tempting basis for evaluating individuals, we must also consider the circumstances under which those achievements were realized. How might socioeconomic or cultural factors have affected access to education and opportunity? It is well known that standardized tests such as the Graduate Record Examination (GRE), which remain widely used in the US, show poor correlation with research productivity in graduate school while demonstrably disadvantaging non-native English speakers and students of color. Likewise, what does it take to publish at a large, fully equipped, and funded university versus a smaller institution with fewer resources? Although privilege often begets early professional success, individuals who face and overcome adversity throughout their careers develop critical skills that uniquely prepare them to face new and bigger challenges at later career stages.[6] Therefore, choosing the candidate who appears more qualified according to conventional metrics may mean losing the one who has become expert in navigating obstacles and persevering through difficulties as an underrepresented scientist.

An added benefit to the practices outlined above is that they help to prevent the tokenism that can emerge when seeking to diversify admissions or interview pools. By considering the benefits of recruiting a diverse group of people as well as each individual's merit in the context of their access to resources and opportunities, we recognize that we are recruiting people — not numbers or quotas. Funding agencies may contribute to this process by following up after initial grant disbursal not only on scientific progress but also on diversity statements. Finally, we define "potential" as the capacity to choose and then realize ambitious goals. A highly motivated applicant already possesses the fuel — employers simply impart the tools to light the fire.

5.2.3 Origins of exclusivity

Recruiting a diverse group of students, postdocs, and faculty is just the first step toward achieving equitable representation in STEM. This work is fruitless without also building an environment that offers resources and mentorship to retain and support these recruits — a particularly pressing issue in STEM, as attrition rates are higher among women[7] and people of color. Below, we describe several factors that exacerbate feelings of exclusion, as well as steps that institutions and departments may take to mitigate these effects.

Not seeing others who look like us engenders feelings of exclusion and negative thoughts such as, "Am I here by mistake? What if I don't succeed?" One way to combat these is by increasing representation at all levels. Toldson reports that historically Black colleges and universities (HBCUs), while representing only 3% of higher education institutions in the USA, graduate more Black science and engineering students than all predominantly white institutions (PWIs) combined. This is due in part to the significantly higher percentage of Black faculty at doctoral-granting HBCUs (67%) than at PWIs (4%).[8] Unsurprisingly, our own success is much easier to envision when we meet and have the opportunity to be mentored by trailblazers like us.

Furthermore, as depicted in the *Hidden Figures* film, a major problem in STEM is the lack of recognition and appreciation for the work done by women and underrepresented minorities (URMs). Of the more than 900 Nobel laureates awarded by 2018, only 14 (1.5%) were Black; moreover, only 21 women and no Black researchers have won a Nobel Prize in science. In particular, from 1901 to 1920, only four women won a Nobel Prize. Almost a century later, between 2000 and 2019, only 24 of 244 laureates were women. Likewise, the President's National Medal of Science (National Science Foundation) has been awarded to 482 scientists in STEM from 1962 to 2014 with 8% being women. Meanwhile, from 2017 to 2019, the American Chemical Society awarded over 200 chemists in various subfields. Of these, approximately 27% were women, and only about 5% were URMs. As recent findings indicate, scientific innovations by women and people of color are substantial but also less likely to be valued or pursued further.[9] Meanwhile, career trajectory is heavily influenced by conference talks,[10] citations, and the receipt of awards. Therefore, it is critical to recognize how these inequities become compounded across a scientist's career and then to actively promote inclusive practices in literature citation and the selection of speakers and award winners.

Finally, women and URMs in STEM often lack a sense of belonging. When you walk into a room full of people, what do you do? Do you look around? Do you feel like all eyes are on you? Legendary tennis player Arthur Ashe once said, "Like many other Blacks, when I find myself in a new public space situation, I will count. I always count. I count the number of Black and brown faces present."[11] Signs of exclusivity can extend beyond the demographics of students and faculty to also appear in the culture of the department. Oftentimes department-wide activities in STEM fields include trivia, dinners, and movie nights. How much might an international student know about local culture, a white student know about Black history, or a man know about women's history? There can also be a lack of inclusivity in movie choices or even the food served at these events, leaving students feeling as though they need to conform to the exclusive environment in which they find themselves.

5.2.4 Cultivating inclusion

Creating spaces for individuals from underrepresented groups in STEM to share experiences and build platforms for advocacy helps to kindle a sense of belonging. Within the Emory Department of Chemistry, we have established several graduate student extracurricular groups for female (@AWIS_Emory on Twitter), international (@IGSSemory), and other underrepresented student communities (@NOBCChE_Emory and @Emory_SACNAS) to connect and discuss challenges. These organizations regularly host events that foster social networks outside of research, advance professional development and provide opportunities to educate interested members of the majority group and potential allies about the experiences of underrepresented populations in chemistry. For example, as board members on the Emory chapter of the National Organization for the Professional Advancement of Black Chemists and Chemical Engineers (NOBCChE), the authors have coordinated events to commemorate Black History Month at Emory, as well as helping to prepare members of the general body for conferences and educational milestones. Overall, these efforts serve to empower our members as departmental leaders while also engaging the greater Atlanta community.

Finally, a crucial component of supporting diversity in the sciences is effective listening. As long as those holding positions of authority come from groups having more privilege than the students and faculty they are recruiting, our employees will have encountered considerable obstacles that our leaders have never conceived. This can be overcome by building robust infrastructure for anonymous feedback, town halls, and private meetings for concerns to be voiced and then heard. For those of us in positions of leadership, when someone is describing an experience foreign to us, that is the time to ask questions, develop a deeper understanding, and empathize — **not to occupy the s**ilence ourselves — in order to best appreciate and address the issue. Finally, if we cannot immediately remedy the problem, demonstrating to someone that they have been heard and understood is a powerful step towards comprehending the intricacies of exclusion.

5.2.5 Outlook

Diversifying science for the mere sake of fulfilling societal expectations does a huge disservice to the communities that these efforts are intended to benefit. Further, it squanders the unique perspectives and experiences that each individual can bring to our departments and institutions. Focusing our evaluation criteria on factors beyond GPA, publications, and academic pedigree is an important first step toward equitable admissions and hiring. Doing so prepares us to invest in each new student or faculty member's career goals, enabling diversification of science without exhibiting tokenism. Importantly, recruiting diverse applicants is insufficient without simultaneous efforts to foster an inclusive environment for everyone to thrive. This begins with hiring diverse faculty members and establishing support groups such as the Association for Women in Science (AWIS), NOBCChE, and the Society for Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS) in STEM departments. Moreover, while we can increase the diversity of a department by accepting diverse candidates, their efforts and accomplishments must be judged or praised by fair and equal criteria.

As with nucleic acids in our lab, the scientific community begins with a large pool of scientists. Within this large pool are groups of underrepresented individuals that possess great skill. If we can enrich these populations of scientists in every round of admissions or hiring, and then provide an environment that allows them to cultivate and utilize their talents, over time we will produce greater diversity in leadership positions. As the cycle repeats, diverse students in later generations will see more and more students, faculty members, and academic leaders who look like them. Eventually, instead of constituting a roadblock, diversity and inclusion in STEM will finally become an asset for researchers to fully leverage.

5.3 References

1. Manuel, B. A.; Karloff, D. B., Recruit and retain a diverse workforce. *Nat. Rev. Chem.* **2020**, *4*, 435-437.

2. Schmidt-Dannert, C.; Arnold, F. H., Directed evolution of industrial enzymes. *Trends Biotechnol.* **1999**, *17*, 135-136.
3. Gibbs, K. Jr., Diversity in STEM: What it is and why it matters. *Sci. Am.* **2020**, https://blogs.scientificamerican.com/voices/diversity-in-stem-what-it-is-and-why-it-matters/
4. AlShebli, B. K.; Rahwan, T.; Woon, W. L., The preeminence of ethnic diversity in scientific collaboration. *Nat. Commun.* **2018**, *9*, 5163.
5. Hong, L.; Page, S. E., Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc. Natl Acad. Sci. U.S.A.* **2004**, *101*, 16385-16389.
6. Henry, M. A.; Shorter, S.; Charkoudian, L.; Heemstra, J. M.; Corwin, L. A., FAIL is not a four-letter word: A theoretical framework for exploring undergraduate students' approaches to academic challenge and responses to failure in STEM Learning Environments. *CBE Life Sci.* **2019**, *Ed. 18*, ar11.
7. Xu, Y. J., Attrition of women in STEM. *J. Career Dev.* **2004**, *44*, 3-19.
8. Toldson, I. A., Why historically Black colleges and universities are successful with graduating Black baccalaureate students who subsequently earn doctorates in STEM. *J. Negro Educ.* **2018**, *87*, 95-98.
9. Hofstra, B.; Kulkarni, V. K.; Munoz-Najar Galvez, S.; He, B.; Jurafsky, D.; McFarland, D. A., The diversity-innovation paradox in science. *Proc. Natl Acad. Sci. U.S.A.* **2020**, *117*, 9284-9291.
10. Ford, H. L.; Brick, C.; Azmitia, M.; Blaufuss, K.; Dekens, P., Women from some minorities get too few talks. *Nature* **2019**, *576*, 32-35.
11. Steele, C. M., *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do (Issues of Our Time)* (W. W. Norton & Company, 2011).

5.4 Related Links

1. Association for Women in Science (AWIS): https://www.awis.org/

2. Carter, C. J. Diversity in STEM - important for society, good for business. *Huffington Post*: https://cra.org/cra-wp/diversity-stem-important-society-good-business/#:~:text=The%20study%20concludes%20that%20diversitynew%20value%20to%20the%20industry
3. Funk, C. & Parker, K. Diversity in the STEM workforce varies widely across jobs. *Pew Research Center*: https://www.pewsocialtrends.org/2018/01/09/diversity-in-the-stem-workforce-varies-widely-across-jobs/
4. Langin, K. A wave of graduate programs drops the GRE application requirement, *Science*: https://doi.org/10.1126/science.caredit.aay2093
5. Morgan, W. No Black scientist has ever won a Nobel – that's bad for science, and bad for society. *The Conversation:* https://theconversation.com/no-black-scientist-has-ever-won-a-nobel-thats-bad-for-science-and-bad-for-society-104456 (2018).
6. National Organization for the Professional Advancement of Black Chemists and Chemical Engineers (NOBCChE): https://www.nobcche.org/
7. Past National Award recipients (American Chemical Society): https://www.acs.org/content/acs/en/funding-and-awards/awards/national/recipients.html
8. Society for Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS): https://www.sacnas.org/
9. The President's National Medal of Science (National Science Foundation): https://www.nsf.gov/od/nms/results.jsp

Chapter 6: Conclusions and Future Perspectives

RNA:small molecule interactions are potent resources in the chemical biology toolbox. The physicochemical properties of small molecules can be tuned to impart targeted reactivity toward the varied and distinct structural attributes of RNA (Figure 6.1). For example, SHAPE reagents selectively react with sterically accessible 2' OH groups in the ribose backbone, providing insight into biomolecular secondary structure.[49-50] Likewise, glyoxal adds reversibly to the Watson-Crick-Franklin base-pairing face of guanine, adenine, and cytosine nucleobases, disrupting hybridization to complementary nucleic acid sequences.[194] RNA is also reactive to canonical protein photocrosslinkers, a useful property when probing RNA target engagement.[59, 61, 143] Finally, highly flexible and equipped with discerning molecular recognition capacity, RNA binds a broad array of molecular chemotypes in nature.[40, 217] Researchers can further leverage this remarkable property to isolate novel RNA sequences that execute a previously undescribed function (Figure 6.1). SELEX has been harnessed to develop a veritable cornucopia of fluorogenic dye-binding aptamers for RNA imaging applications—Spinach, Broccoli, Corn, Mango, and Peppers, to name a few.[68-69, 89, 117, 218] Cech and coworkers used the same technique to select a catalytically active RNA sequence, otherwise known as a ribozyme, that performs peptidyl transferase reactions.[219] Ribozymes in nature hint at the possibility that RNA was once the primary driver of life on Earth, an idea that RNA enthusiasts call the RNA World hypothesis.[220]

But many unknowns remain with regard to this molecule. The field is constantly discovering new classes and biochemical functions.[10] Intracellular RNA localization is known to drive cell polarization, but transcriptional dynamics are challenging to capture, particularly without perturbing native behavior.[31] In addition, while RNA is certainly able to bind diverse molecules, the field lacks knowledge on how to reliably and reproducibly design ligands that target pathogenic RNAs.[7] This problem could not have been underscored more starkly than with the onset of the COVID-19 pandemic, caused by the RNA virus SARS-CoV-2.[221] Such questions are the focus of

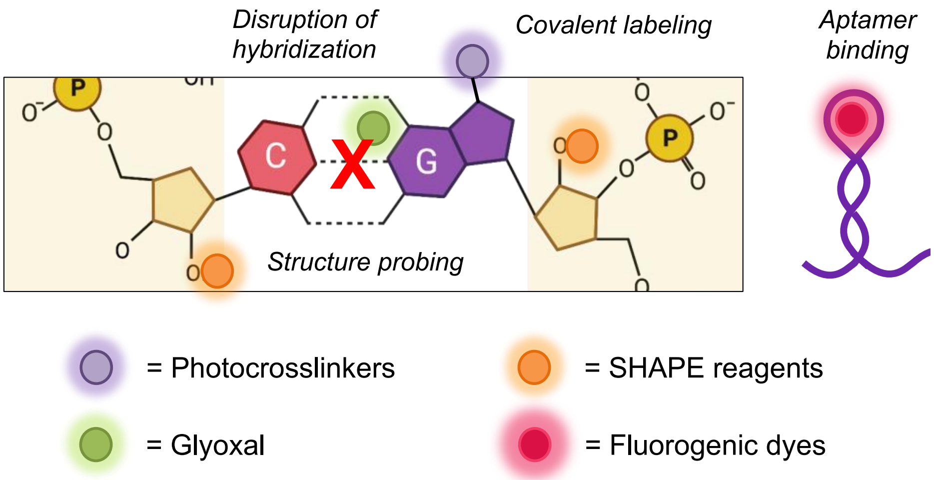


Figure 6.1: Small molecules are powerful tools to probe and disrupt RNA function. Chemical biologists leverage the reactivity of RNA nucleobases and sugars, in addition to specific binding by naturally evolved and artificially selected aptamers, to characterize biomolecular function.

this dissertation. In Chapter 2, we develop the first photoaffinity-based method to image RNA dynamics in live cells. We anticipate that this strategy will enable a reduction in the molecular weight of the cargo typically necessary for visualization of transcripts in real time. In Chapter 3, we again utilize photoaffinity labeling, in this case to explore guiding principles that drive certain mRNAs to be more ligandable than others. As less than 20% of proteins are thought to contain druggable binding pockets, the mRNAs upstream of these proteins, as well as other functional RNA classes, represent underutilized therapeutic targets.[15] In Chapter 4, we deploy glyoxal to tune the bioactivity and pharmacokinetic properties of nucleobase-like drugs toward bolstering patient adherence in the clinical treatment of DNA and RNA viruses.

Given that Chapters 2 and 3 harness RNA photoaffinity labeling, perhaps the most pressing unanswered research question emerging from this dissertation is whether diazirine is indeed the best photocrosslinker (PC) to achieve covalent RNA labeling. In Chapter 2, we observed ~2% *in vitro* labeling efficiency between our malachite green (MG)-diazirine-alkyne probe and the MG aptamer (MGA). In Chapter 3, two PAGE gels indicated a labeling efficiency of 40–50% between PreQ1-diazirine-alkyne, but a third gel reflected < 5% labeling efficiency. Higher labeling yields could drastically improve the robustness and efficacy of both methods. A related question is whether PC labeling yield is context-specific.

Certainly, alkyl diazirine is the most prevalent PC in the literature for RNA labeling.[61-63, 111] However, few groups quantify and report the labeling efficiency for multiple photoaffinity probes tested head-to-head. Mukherjee et al. report yields between 5–15% for aryl azide- and alkyl diazirine-based PAL ligands.[138] Schneekloth and coworkers measured 30–50% efficiency for alkyl diazirine probes and either ~20% efficiency or only trace labeling for select trifluoromethyl phenyl diazirine probes.[113] Meanwhile, photoaffinity labeling (PAL) has been in use in the field of proteomics for decades, with labeling characteristics well characterized.[53] Given the vast differences in composition, structure, and binding paradigms between RNA and proteins, it is highly likely that the same rules for achieving maximal labeling efficiency do not apply. Trifluoromethyl phenyl diazirines, aryl azides, benzophenones (BPs), and psoralens are highly underexplored in comparison to alkyl diazirines.[53, 222] The capacity of BP for repeated photoactivation could render it highly useful in increasing RNA labeling yield.[53] A systematic study of the effects of PC identity, linker identity, and linker length on RNA photocrosslinking efficiency would rapidly advance this burgeoning area of research.

With higher labeling yields and better characterization of RNA:PC properties, our RNA imaging method reported in Chapter 2 could be expanded in several ways. Firstly, multiplexed, multicolor RNA imaging could be achieved by developing alternate aptamer:fluorogenic dye pairs that are amenable to PC installation. Current efforts in our lab are focused on exploring PAL strategies with the Peppers aptamer, an attractive candidate due to the wide spectral range of emission that this system offers (485–620 nm).[69] Given cutting-edge developments in aptamer-based super-resolution RNA imaging, it would be exciting to investigate potential further advantages by combining rhodamine-based systems with covalent RNA labeling.[71] Secondly, if multiple PC groups provide suitable labeling yields, the distinct photophysical properties could be used to establish imaging of multiple RNAs of interest in an a spatiotemporally orthogonal manner.[53] Thirdly, Chenoweth and coworkers recently developed a photoinducible system to control intracellular transport of proteins and even organelles. Using our RNA photocrosslinking strategy, an analogous workflow could easily be designed to access light-induced intracellular RNA transport.[223-224] Such a tool would be invaluable for studying the ties between RNA localization and function. Finally, we imaged MGA-tagged U6 snRNA in Chapter 2, but our protocol could be used to study any number of biologically important transcripts.

An important next step to establish the utility of the photoaffinity-based RNA target engagement method described in Chapter 3 is biochemical validation of the identified mRNA hits of compound 9, the palbociclib-diazirine-alkyne probe. We have already pursued two parallel strategies to this end, based on the PEARL-Seq method by Mukherjee et al. and based on an in-house designed one- or two-step "mini-SELEX" enrichment protocol (Figure 6.2).[138] Unfortunately, despite extensive optimization, neither method has enabled mRNA hit validation.

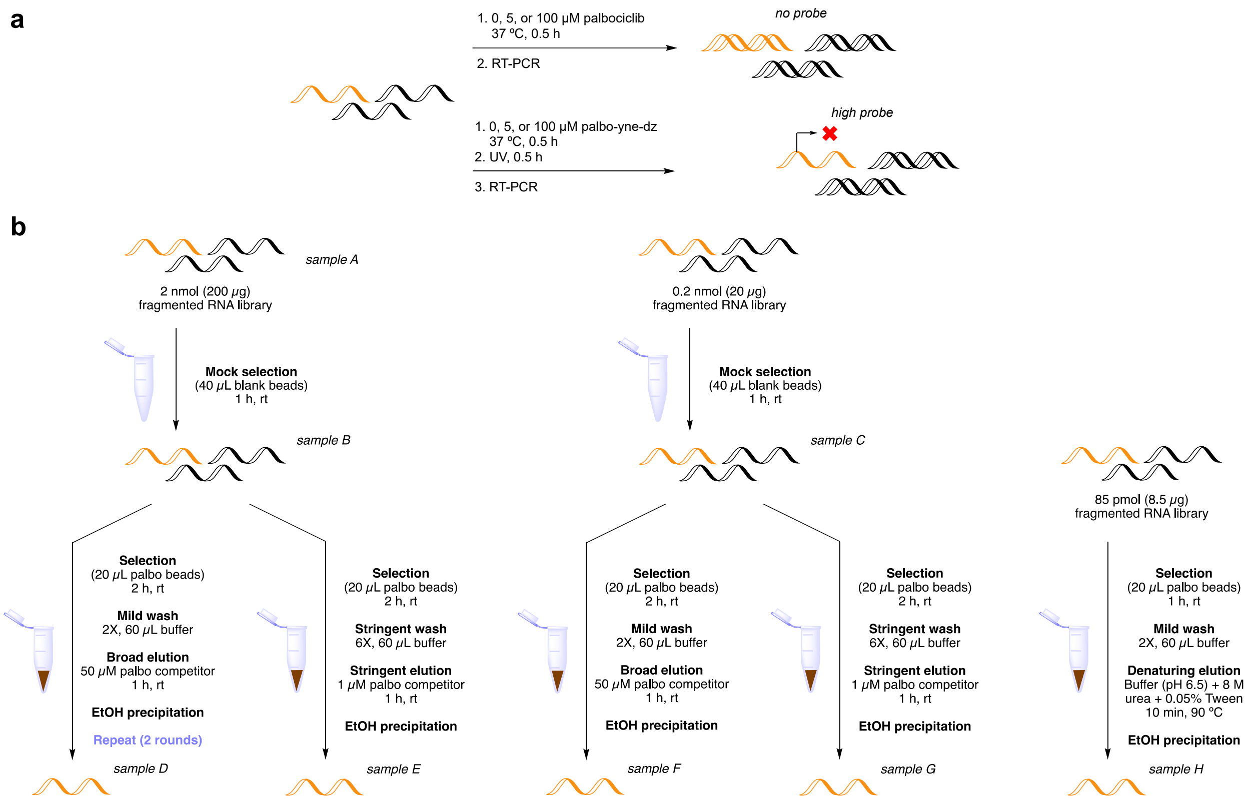


Figure 6.2: Two approaches to validate Chem-CLIP-Map-Seq mRNA hits. a PEARL-Seq-based scheme to detect palbociclib- or compound **9**-induced RT stops in *in vitro*-transcribed mRNA sequencing hits. **b** One- or two-step selection scheme to enrich mRNA sequencing hits on palbociclib-functionalized beads.

As mentioned in Chapter 3, one formidable challenge in the development of RNA-targeted probes is the cost and practicality of synthesizing lengthy, *in vitro*-transcribed RNAs for molecular biology studies. Future target validation efforts will focus on more well-established methods, such as SHAPE and/or proteomic analysis of palbociclib-treated MCF-7 cells, to further characterize transcriptome-wide effects of palbociclib treatment. Upon additional optimization, the PAL workflow could be used to study any number of RNA ligands.

Meanwhile, the preliminary *in cellulo* and *in vivo* data on glyoxal-mediated control over acyclovir (ACV) bioactivity and improvement of undesirable pharmacokinetic properties in Chapter 4 make us excited to pursue further characterization. This could include additional mouse studies focused on absorption, distribution, metabolism, and excretion, as well as further validation of timed-release conversion of glyoxal-caged ACV back into the parent compound. Additionally, while we showed proof-of-principle glyoxal caging of a three-member panel of nucleoside reverse transcriptase inhibitors (NRTIs), follow-up studies are needed to characterize the observed reaction products, glyoxal decaging kinetics vis-à-vis pharmacologically ideal half-life, and most importantly, effects of glyoxal caging on viral RT inhibition.

Another avenue we are interested in pursuing is nucleoside antiviral caging with glyoxal derivatives for improved drug circulation lifetime. Prior work in our lab and by He and coworkers has shown that nucleic acids are amenable to caging by glyoxal derivatives such as methylglyoxal and kethoxal.[225] Building upon this precedent that substituted 1,2-dicarbonyls react efficiently and reversibly with nucleobases, we will explore nucleoside antiviral caging using glyoxal derivatives that incorporate fatty acid, steroid, or PEG chains (Figure 6.3), as these are shown to be effective in extending circulation lifetime through binding to

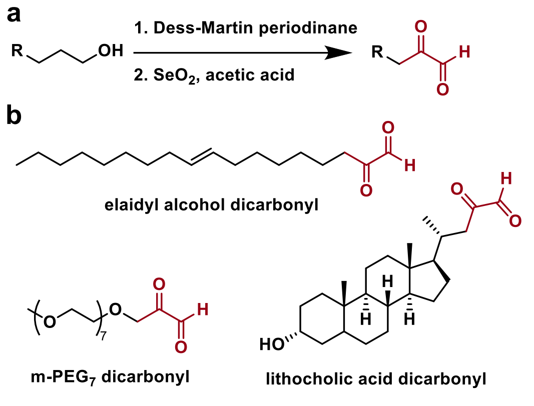


Figure 6.3: Antiviral caging with fatty acid, steroid, and PEG derivatives of glyoxal. a Reaction scheme for oxidation of primary alcohols to 1,2-dicarbonyls. **b** Representative structures of target 1,2-dicarbonyl caging groups to be synthesized and tested.

serum albumin or reduction of renal clearance.[226-228] These studies will utilize commercially available moieties such as the fatty acid elaidyl alcohol and the steroid lithocholic acid.

Beyond the scope of RNA:small molecule interactions, humans are constantly interfacing with scientific principles. Chapter 5 discusses collaborative efforts to raise scientific awareness outside of the immediate chemistry community. As technology and innovation continue to advance at an exponential pace, the potential for knowledge gaps between individuals grows as well. The recent COVID-19 pandemic underscored the importance of a scientifically literate society. Fortunately, endless avenues exist for researchers to engage in outreach opportunities that span departmental and global levels.

6.1 References

1. Wilkinson, K. A.; Merino, E. J.; Weeks, K. M., Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **2006,** *1* (3), 1610-6.

2. Busan, S.; Weidmann, C. A.; Sengupta, A.; Weeks, K. M., Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. *Biochemistry* **2019,** *58* (23), 2655-2664.

3. Knutson, S. D.; Sanford, A. A.; Swenson, C. S.; Korn, M. M.; Manuel, B. A.; Heemstra, J. M., Thermoreversible control of nucleic acid structure and function with glyoxal caging. *J. Am. Chem. Soc.* **2020,** *142* (41), 17766-17781.

4. Rzuczek, S. G.; Colgan, L. A.; Nakai, Y.; Cameron, M. D.; Furling, D.; Yasuda, R.; Disney, M. D., Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat. Chem. Biol.* **2017,** *13* (2), 188-193.

5. Wang, J.; Schultz, P. G.; Johnson, K. A., Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc. Natl. Acad. Sci. U.S.A.* **2018,** *115* (20), E4604-E4612.

6. Connelly, C. M.; Numata, T.; Boer, R. E.; Moon, M. H.; Sinniah, R. S.; Barchi, J. J.;

Ferre-D'Amare, A. R.; Schneekloth, J. S., Jr., Synthetic ligands for PreQ1 riboswitches provide structural and mechanistic insights into targeting RNA tertiary structure. *Nat. Commun.* **2019,** *10* (1), 1501.

7. Kavita, K.; Breaker, R. R., Discovering riboswitches: The past and the future. *Trends Biochem. Sci.* **2023,** *48* (2), 119-141.

8. Mattick, J. S.; Amaral, P. P.; Carninci, P.; Carpenter, S.; Chang, H. Y.; Chen, L. L.; Chen, R.; Dean, C.; Dinger, M. E.; Fitzgerald, K. A.; Gingeras, T. R.; Guttman, M.; Hirose, T.; Huarte, M.; Johnson, R.; Kanduri, C.; Kapranov, P.; Lawrence, J. B.; Lee, J. T.; Mendell, J. T.; Mercer, T. R.; Moore, K. J.; Nakagawa, S.; Rinn, J. L.; Spector, D. L.; Ulitsky, I.; Wan, Y.; Wilusz, J. E.; Wu, M., Long non-coding RNAs: Definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* **2023,** *24* (6), 430-447.

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

10. Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R., Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* **2014,** *136* (46), 16299-308.

11. Warner, K. D.; Sjeklоća, L.; Song, W.; Filonov, G. S.; Jaffrey, S. R.; Ferré-D'Amaré, A. R., A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. *Nat. Chem. Biol.* **2017,** *13*, 1195-1201.

12. Autour, A.; S, C. Y. J.; A, D. C.; Abdolahzadeh, A.; Galli, A.; Panchapakesan, S. S. S.; Rueda, D.; Ryckelynck, M.; Unrau, P. J., Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. *Nat. Commun.* **2018,** *9* (1), 656.

13. Chen, X.; Zhang, D.; Su, N.; Bao, B.; Xie, X.; Zuo, F.; Yang, L.; Wang, H.; Jiang, L.; Lin, Q.; Fang, M.; Li, N.; Hua, X.; Chen, Z.; Bao, C.; Xu, J.; Du, W.; Zhang, L.; Zhao, Y.; Zhu, L.; Loscalzo, J.; Yang, Y., Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nat. Biotechnol.* **2019,** *37*, 1287-1293.

14. Zhang, B.; Cech, T. R., Peptide bond formation by in vitro selected ribozymes. *Nature* **1997,** *390*, 96-100.

15. Higgs, P. G.; Lehman, N., The RNA World: Molecular cooperation at the origins of life. *Nat. Rev. Genet.* **2015,** *16* (1), 7-17.

16. Boivin, V.; Faucher-Giguere, L.; Scott, M.; Abou-Elela, S., The cellular landscape of mid-size noncoding RNA. *Wiley Interdiscip. Rev. RNA* **2019,** *10* (4), e1530.

17. Buxbaum, A. R.; Haimovich, G.; Singer, R. H., In the right place at the right time: Visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **2015,** *16* (2), 95-109.

18. Warner, K. D.; Hajdin, C. E.; Weeks, K. M., Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **2018,** *17* (8), 547-558.

19. Hu, B.; Guo, H.; Zhou, P.; Shi, Z. L., Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **2021,** *19* (3), 141-154.

20. Dixon, S. J.; Stockwell, B. R., Identifying druggable disease-modifying gene products. *Curr. Opin. Chem. Biol.* **2009,** *13* (5-6), 549-55.

21. Nakamoto, K.; Ueno, Y., Diazirine-containing RNA photo-cross-linking probes for capturing microRNA targets. *J. Org. Chem.* **2014,** *79* (6), 2463-72.

22. Suresh, B. M.; Li, W.; Zhang, P.; Wang, K. W.; Yildirim, I.; Parker, C. G.; Disney, M. D., A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2020,** *117* (52), 33197-33203.

23. Tong, Y.; Gibaut, Q. M. R.; Rouse, W.; Childs-Disney, J. L.; Suresh, B. M.; Abegg, D.; Choudhary, S.; Akahori, Y.; Adibekian, A.; Moss, W. N.; Disney, M. D., Transcriptome-wide mapping of small-molecule RNA-binding sites in cells informs an isoform-specific degrader of QSOX1 mRNA. *J. Am. Chem. Soc.* **2022,** *144* (26), 11620-11625.

24. Mukherjee, H.; Blain, J. C.; Vandivier, L. E.; Chin, D. N.; Friedman, J. E.; Liu, F.; Maillet, A.; Fang, C.; Kaplan, J. B.; Li, J.; Chenoweth, D. M.; Christensen, A. B.; Petersen, L. K.; Hansen, N. J. V.; Barrera, L.; Kubica, N.; Kumaravel, G.; Petter, J. C., PEARL-seq: A

photoaffinity platform for the analysis of small molecule-RNA interactions. *ACS Chem. Biol.* **2020,** *15* (9), 2374-2381.

25. Balaratnam, S.; Rhodes, C.; Bume, D. D.; Connelly, C.; Lai, C. C.; Kelley, J. A.; Yazdani, K.; Homan, P. J.; Incarnato, D.; Numata, T.; Schneekloth, J. S., Jr., A chemical probe based on the PreQ(1) metabolite enables transcriptome-wide mapping of binding sites. *Nat. Commun.* **2021,** *12* (1), 5856.

26. Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J. S., Photo-affinity labeling (PAL) in chemical proteomics: A handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci.* **2016,** *15*, 14.

27. Kurz, M.; Gu, K.; Lohse, P. A., Psoralen photo-crosslinked mRNA–puromycin conjugates: A novel template for the rapid and facile preparation of mRNA–protein fusions. *Nucleic Acids Res.* **2000,** *28* (18), e83.

28. Englert, D.; Burger, E. M.; Grun, F.; Verma, M. S.; Lackner, J.; Lampe, M.; Buhler, B.; Schokolowski, J.; Nienhaus, G. U.; Jaschke, A.; Sunbul, M., Fast-exchanging spirocyclic rhodamine probes for aptamer-based super-resolution RNA imaging. *Nat. Commun.* **2023,** *14* (1), 3879.

29. Ballister, E. R.; Ayloo, S.; Chenoweth, D. M.; Lampson, M. A.; Holzbaur, E. L. F., Optogenetic control of organelle transport using a photocaged chemical inducer of dimerization. *Curr. Biol.* **2015,** *25* (10), R407-R408.

30. Wu, D. Z.; Lampson, M. A.; Chenoweth, D. M., Photoactivatable trimethoprim-based probes for spatiotemporal control of biological processes. *Methods Enzymol.* **2020,** *638*, 273-294.

31. Wu, T.; Lyu, R.; You, Q.; He, C., Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ. *Nat. Methods* **2020,** *17* (5), 515-523.

32. van Witteloostuijn, S. B.; Pedersen, S. L.; Jensen, K. J., Half-life extension of biopharmaceuticals using chemical methods: Alternatives to PEGylation. *ChemMedChem* **2016,**

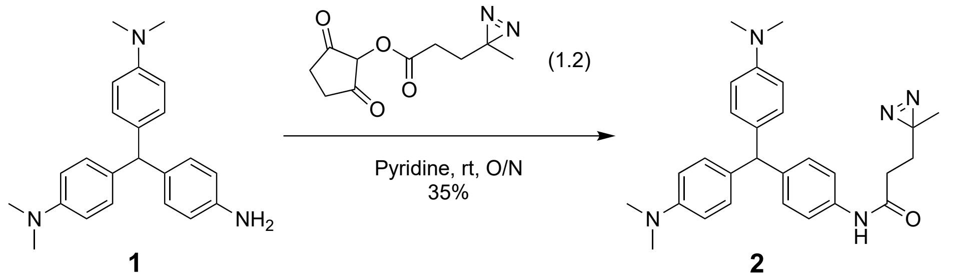
11 (22), 2474-2495.

33. Bech, E. M.; Pedersen, S. L.; Jensen, K. J., Chemical strategies for half-life extension of biopharmaceuticals: Lipidation and its alternatives. *ACS Med. Chem. Lett.* **2018,** *9* (7), 577-580.

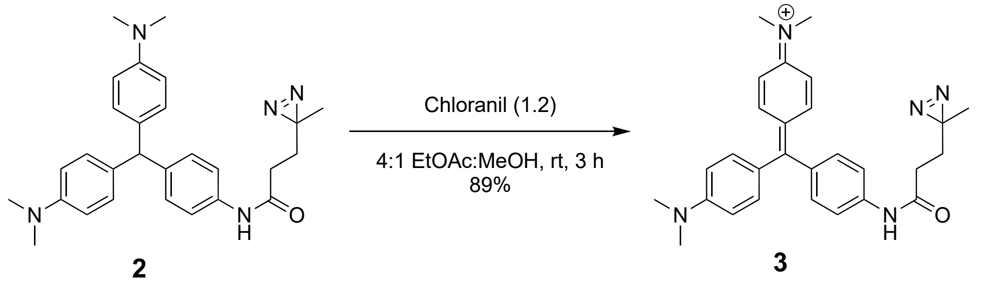
34. Wang, C.; Cheng, S.; Zhang, Y.; Ding, Y.; Chong, H.; Xing, H.; Jiang, S.; Li, X.; Ma, L., Long-acting HIV-1 fusion inhibitory peptides and their mechanisms of action. *Viruses* **2019,** *11* (9).

Appendix A: Data Omitted from Chapter 2

Scheme A1: Synthesis of reduced MGD2 precursor (2).



Scheme A2: Synthesis of MGD2 (3).



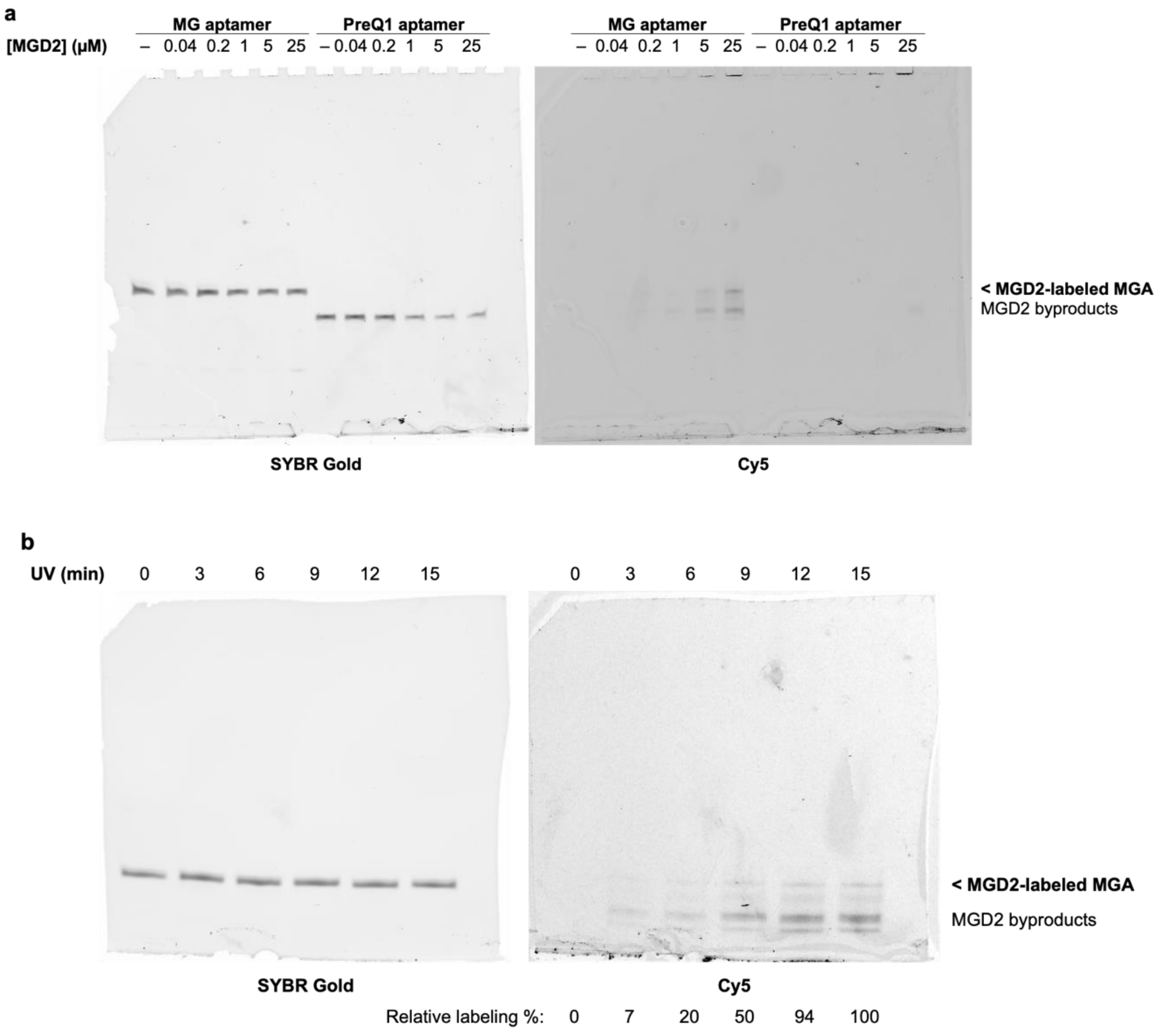


Figure A1: Concentration- and UV-dependence of RNA aptamer labeling. a MGA RNA is selectively labeled over an off-target aptamer in the presence of increasing MGD2 concentration. **b** Labeling of MGA by MGD2 increases with UV irradiation time.

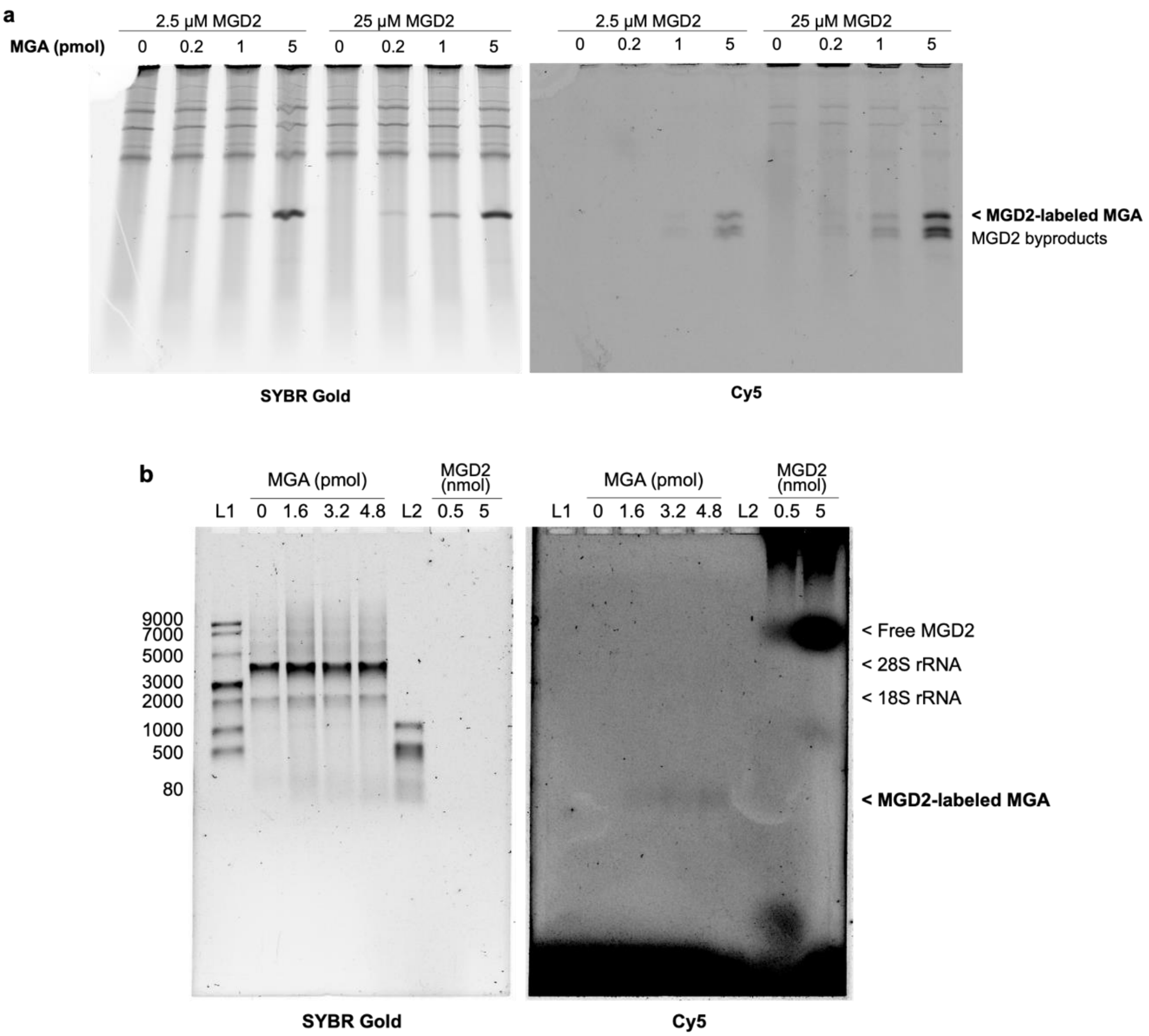
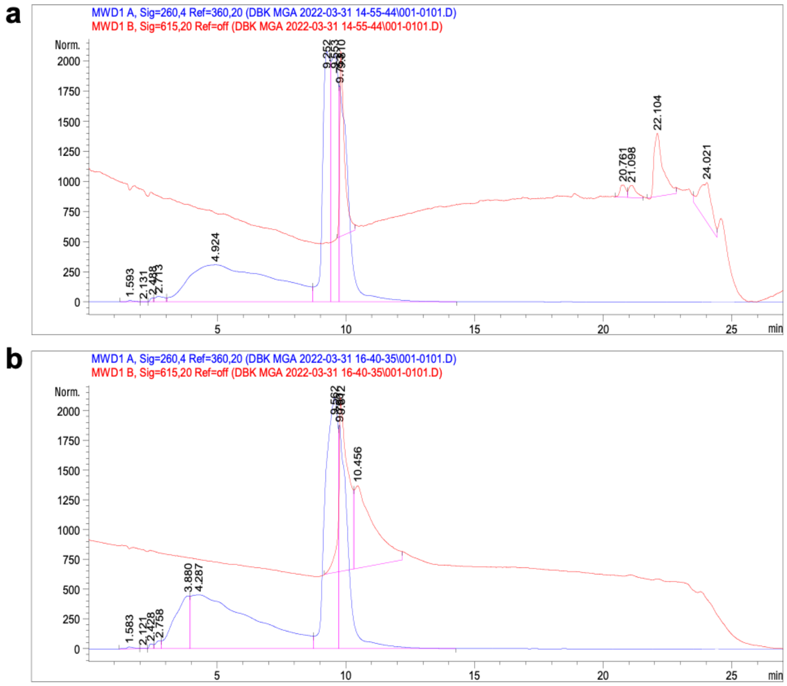


Figure A2: Selectivity of MGA labeling. MGA labeling was visualized in the presence of total cellular RNA extracted from N2a cells, as analyzed by **a** 12% denaturing PAGE or **b** 1% denaturing agarose gel.



	A260 (RNA)	A615 (MGD2)	% labeled
a. MG-labeled MGA	1.25e5	17.53	0.16%
b. MGD2-labeled MGA	1.15e5	173.58	1.7%

Figure A3: HPLC traces of probe-labeled MGA. a 9.75 nmol of MG-labeled MGA and **b** 9.75 nmol of MGD2-labeled MGA. MWD1 (RNA) and MWD2 (MGD2) are shown in blue and red, respectively. **c** Areas under blue and red curves and calculation of % labeling according to Chapter 2.5.6, Equation 1.

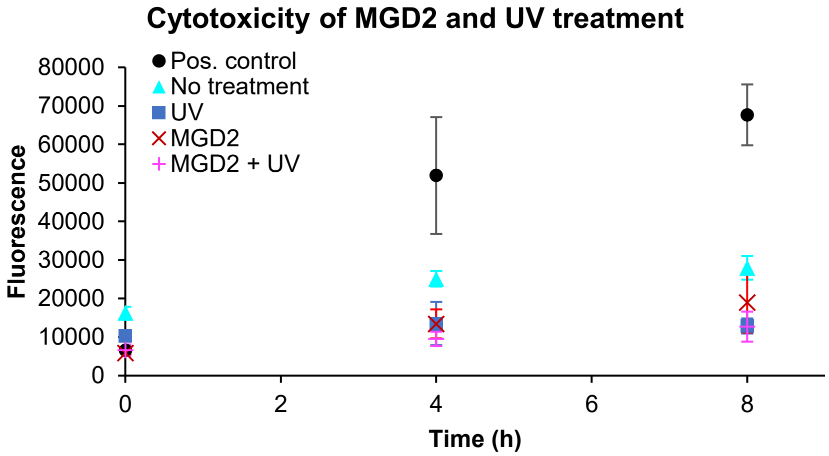
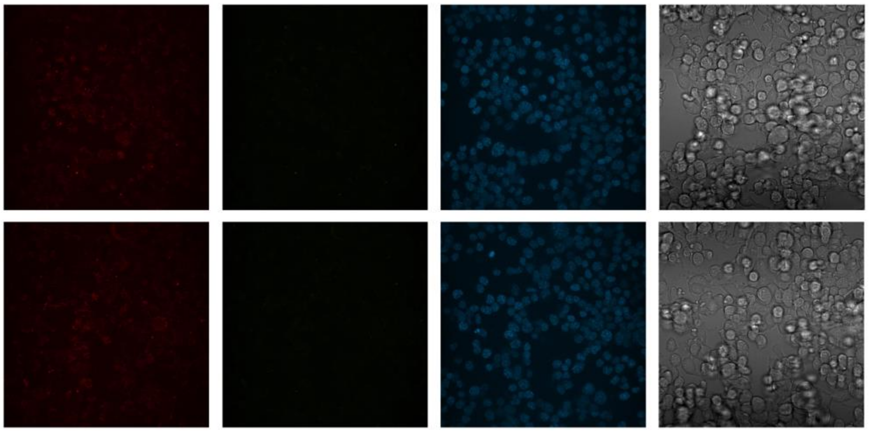


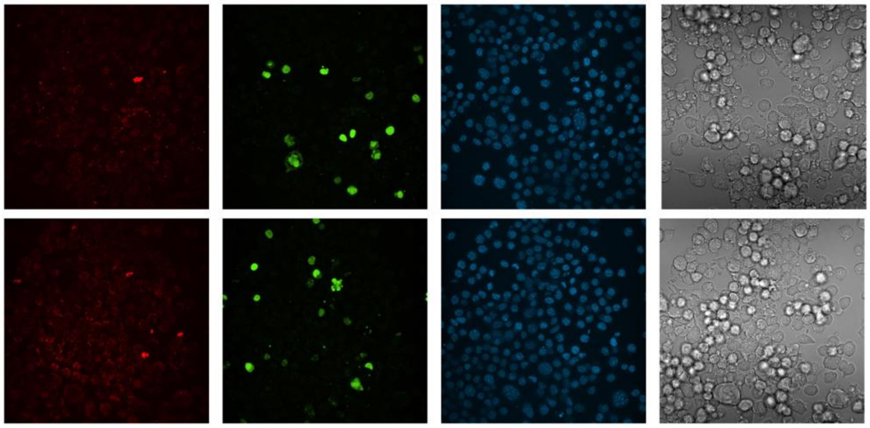
Figure A4: Assessment of UV and MGD2 cytotoxicity. Cytotoxic effects up to 8 h post-treatment of N2a cells with MGD2 (200 nM) for 15 min, UV irradiation for 10 min, or both using CellTox^{TM} Green Cytotoxicity Assay (Promega). Viability was compared to a “no treatment” negative control and a lysis buffer-treated positive control.

Replicate 1

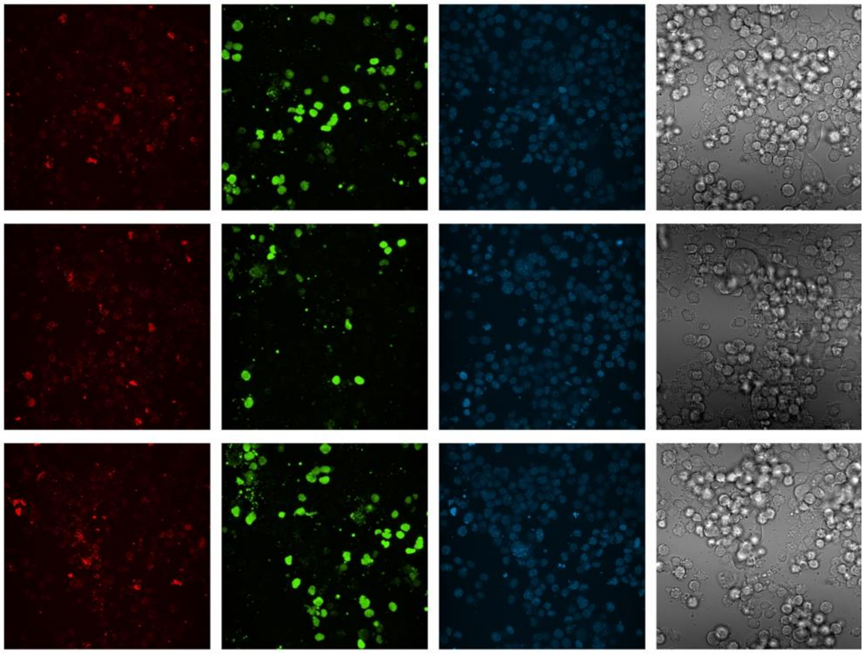
Condition 1



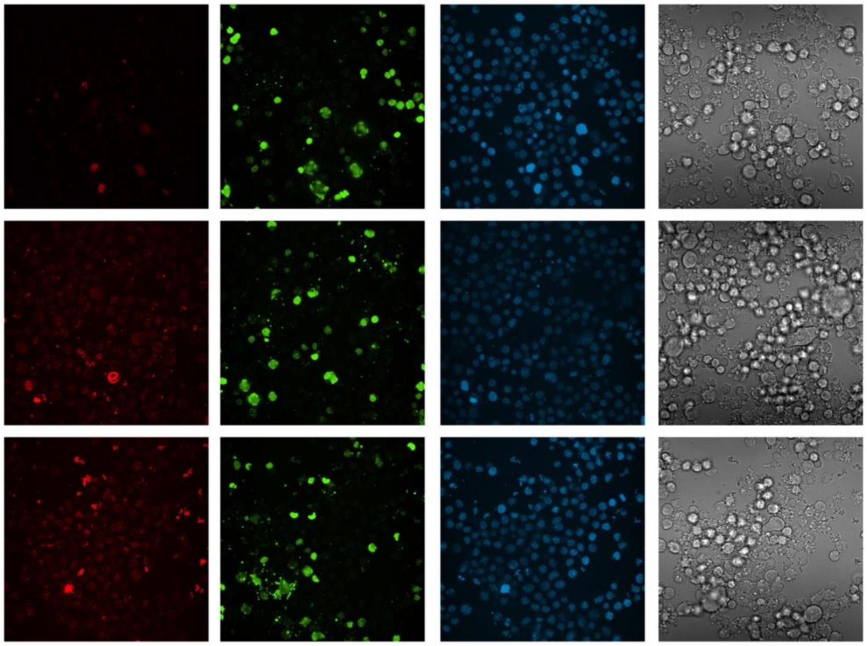
Condition 2



Condition 3

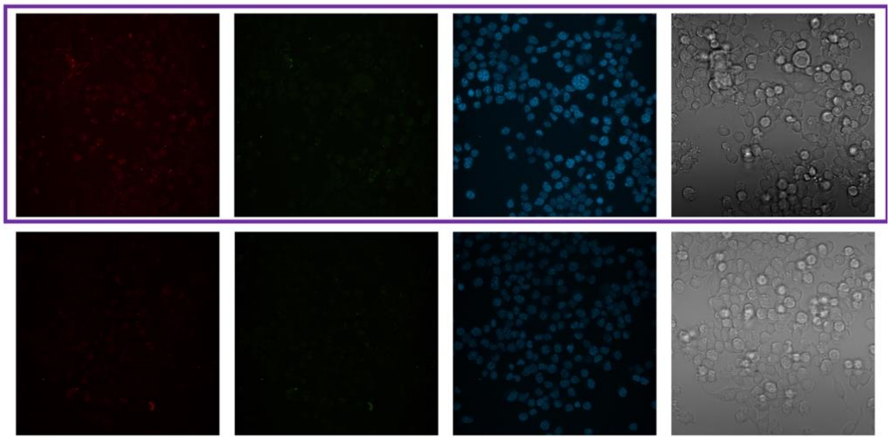


Condition 4

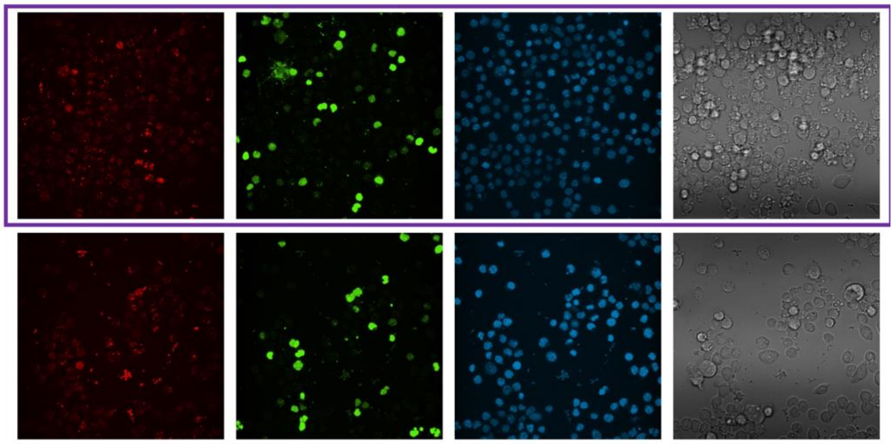


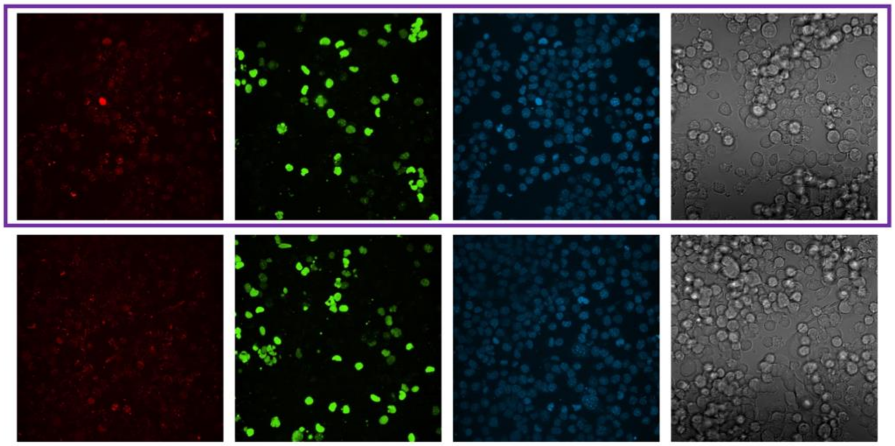
Replicate 2

Condition 1



Condition 2




Condition 3

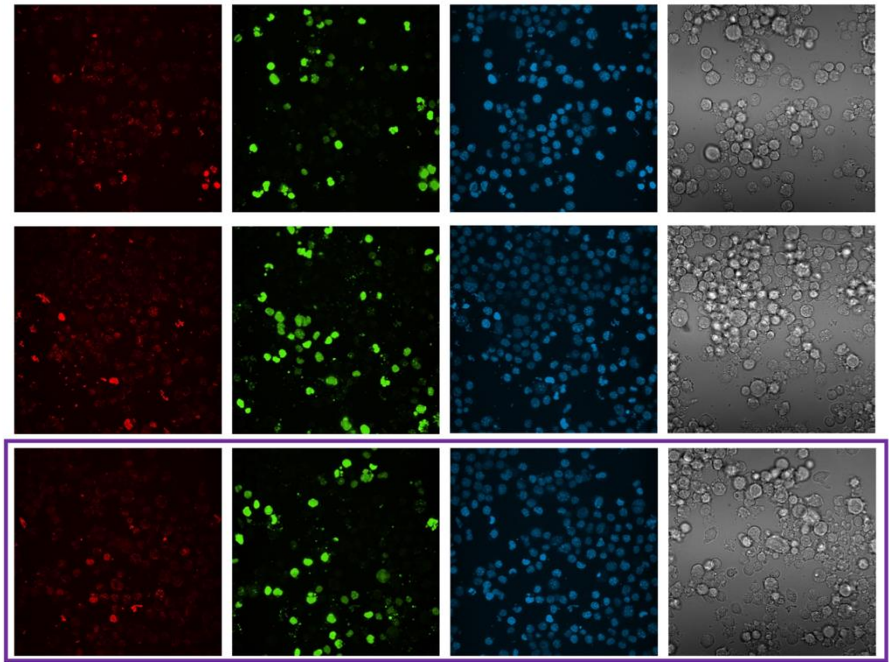

Condition 4

Figure A5: Confocal microscopy images of cell treatment conditions 1–4 across two replicates. Mouse Neuro-2a cells were treated with one of the four following conditions: 1) Mock transfection (3 d), 100 nM MGD2 (15 min); 2) Coilin-GFP and 0X MGA-U6 snRNA transfection (3 d), 100 nM MGD2 (15 min), 365 nm UV light (10 min); 3) Coilin-GFP and 1X MGA-U6 snRNA transfection (3 d), 100 nM MGD2 (15 min); 4) Coilin-GFP and 1X MGA-U6 snRNA transfection (3 d), 100 nM MGD2 (15 min), 365 nm UV light (10 min). The four channels correspond to MGD2 (red), coilin-GFP (green), Hoechst (blue), and brightfield. Boxes indicate the representative images that were zoomed in and depicted for each condition in main text.

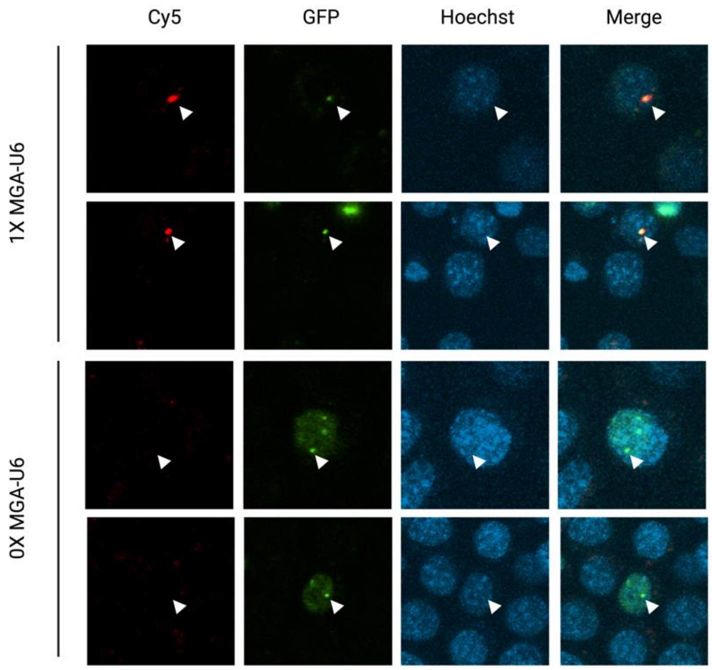
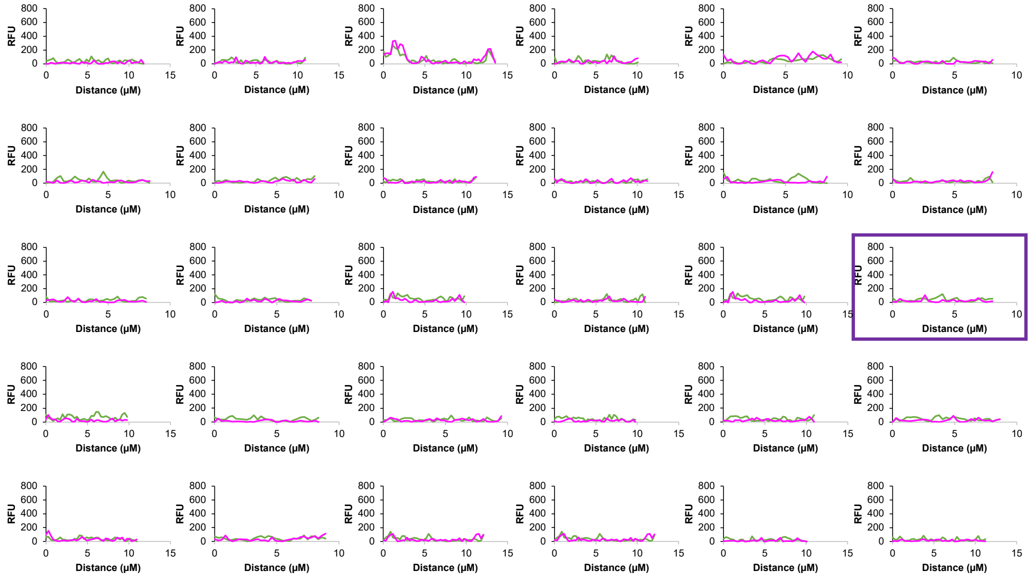


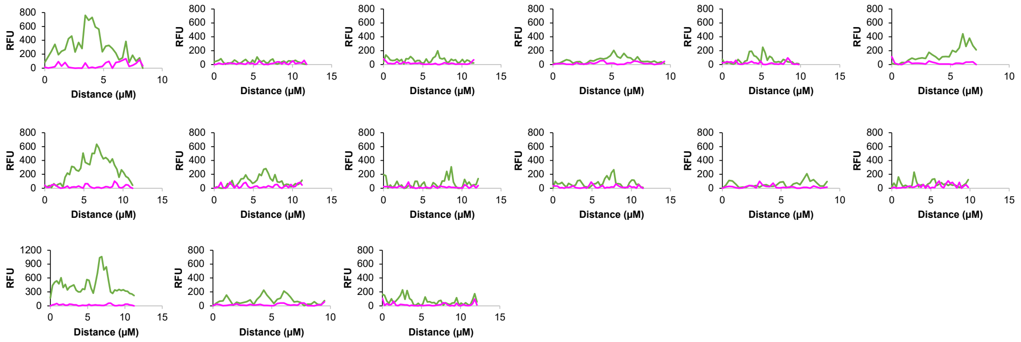
Figure A6: Colocalization of U6 snRNA and coilin-GFP in nuclear puncta. Mouse Neuro-2a cells were treated with 1) Coilin-GFP and 1X MGA-U6 snRNA transfection (3 d), 100 nM MGD2 (15 min), 365 nm UV light (10 min) or 2) Coilin-GFP and 0X MGA-U6 snRNA transfection (3 d), 100 nM MGD2 (15 min), 365 nm UV light (10 min). The four channels correspond to MGD2 (Cy5), coilin-GFP (GFP), Hoechst (blue), and brightfield. Arrows indicate nuclear puncta, and two examples are provided for both treatments.

Replicate 1

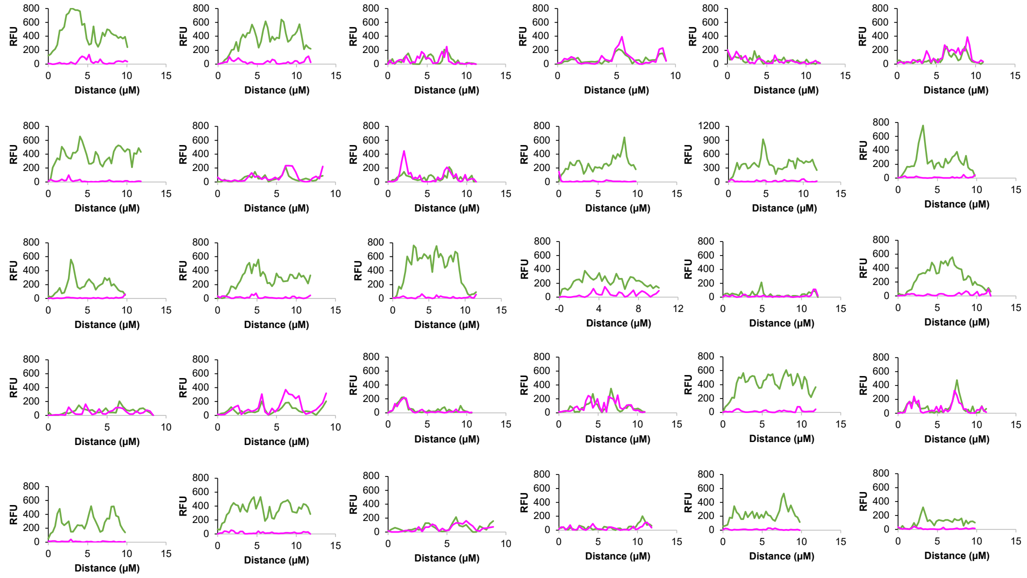
Condition 1



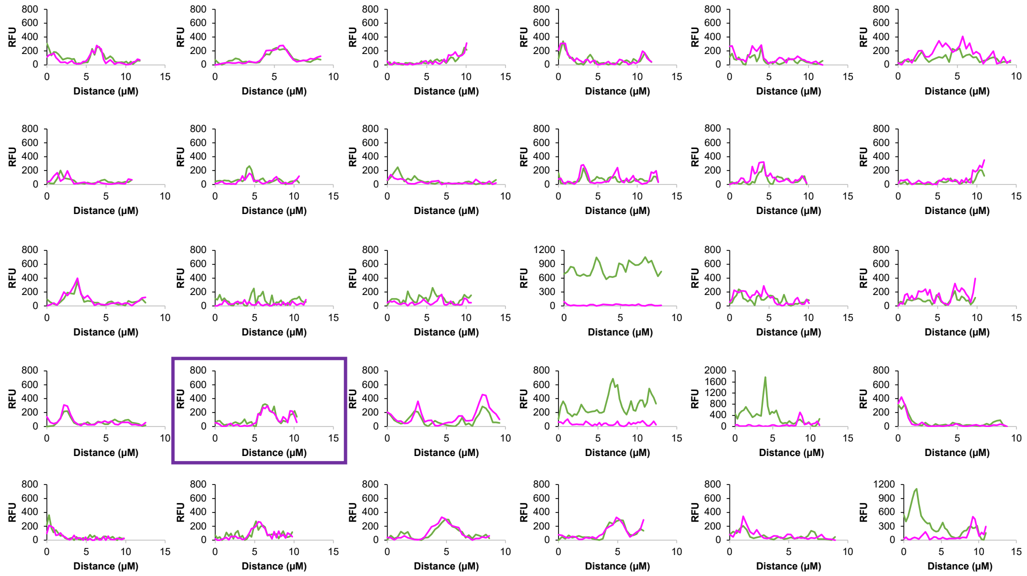
Condition 2



Condition 3

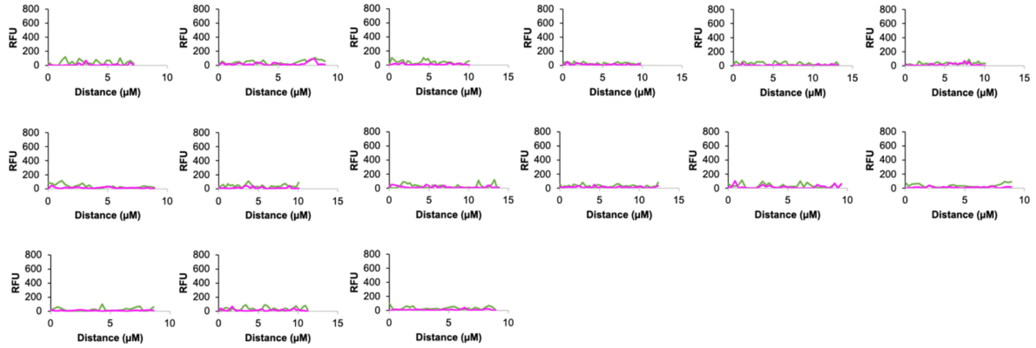


Condition 4

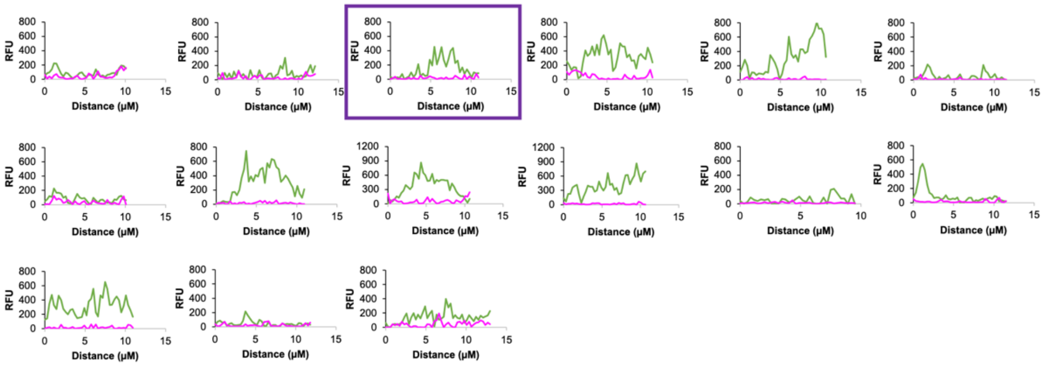


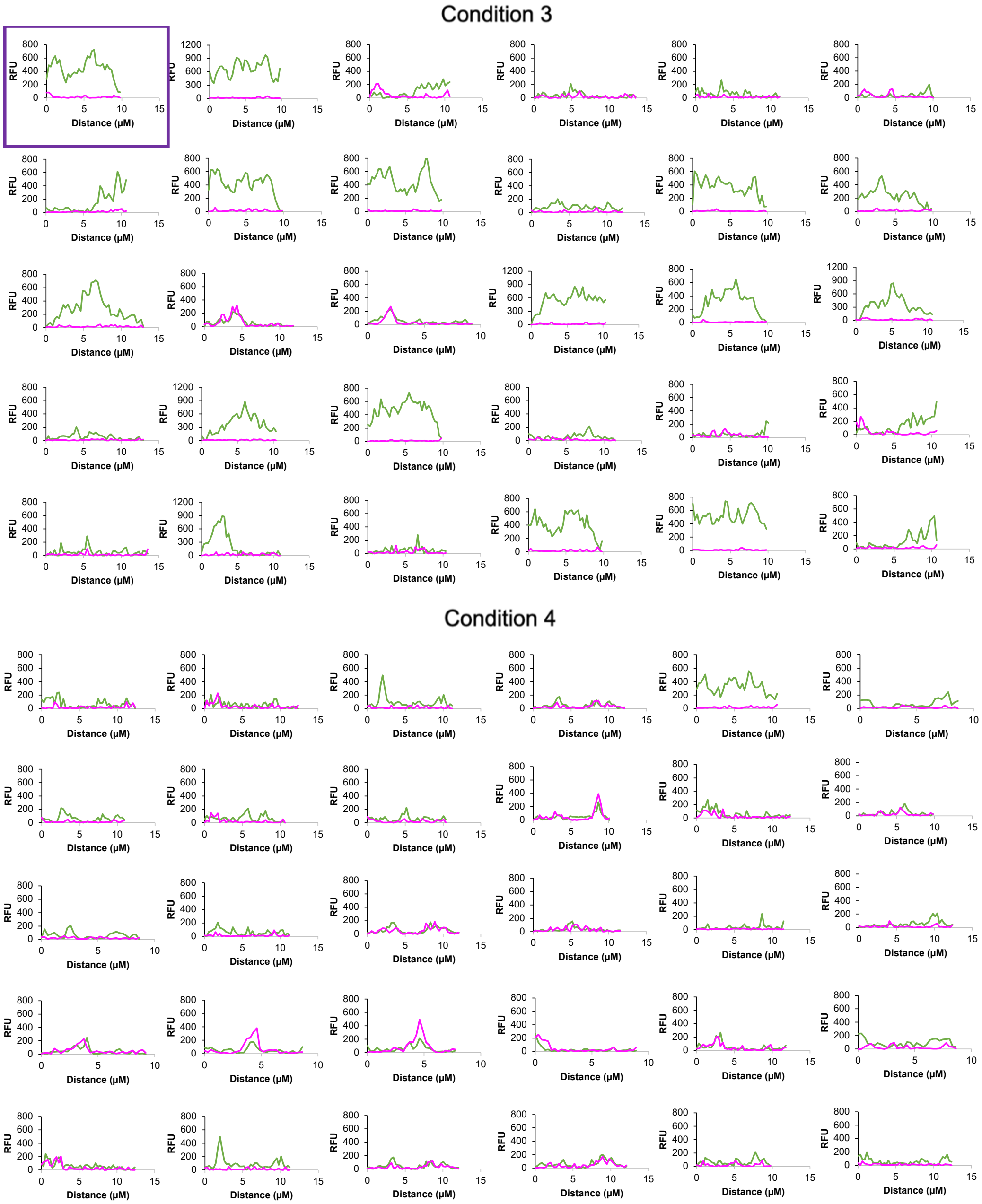
Replicate 2

Condition 1



Condition 2




Condition 3

Condition 4

Figure A7: Line scans of cell treatment conditions 1–4 across two replicates. Mouse Neuro-2a cells were treated with one of the four following conditions: 1) Mock transfection (3 d), 100 nM MGD2 (15 min); 2) Coilin-GFP and 0X MGA-U6 snRNA transfection (3 d), 100 nM MGD2 (15 min), 365 nm UV light (10 min); 3) Coilin-GFP and 1X MGA-U6 snRNA transfection (3 d), 100 nM MGD2 (15 min); 4) Coilin-GFP and 1X MGA-U6 snRNA transfection (3 d), 100 nM MGD2 (15 min), 365 nm UV light (10 min). Plot profiles depict GFP (green) and MGD2 (red) fluorescence signals across 15–30 nuclear line scans, as per the Materials and Methods section, "Image analysis." We observed that coilin-GFP signal was lowest in mock-transfected Condition 1, as expected. Further, MGD2 colocalization with coilin-GFP followed the expected trend: Condition 4 > Condition 3 > Condition 2 ≈ Condition 1. Boxes indicate the representative graph depicted in main text for each condition.

Appendix B: Data Omitted from Chapter 3

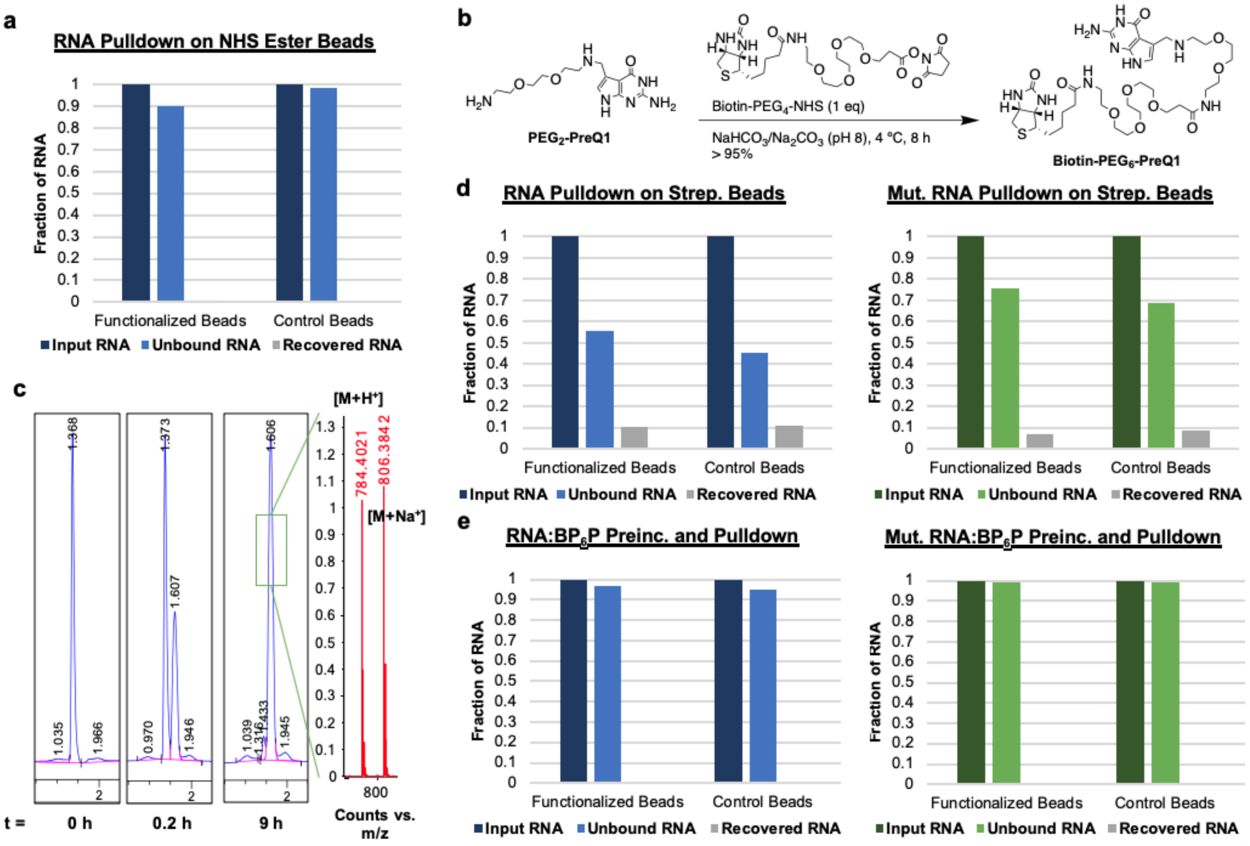


Figure B1: Non-covalent RNA pulldown on PreQ1-functionalized beads. a PreQ1 aptamer recovery from PEG$_2$-PreQ1-functionalized or unfunctionalized NHS ester beads. **b** Synthesis of Biotin-PEG$_6$-PreQ1 (BP$_6$P) as NHS ester test reaction. **c** Characterization of BP$_6$P formation by HPLC-MS. **d** Pulldown of WT and mutant aptamer onto BP$_6$P-functionalized streptavidin beads. **e** Preincubation of WT and mutant aptamer with BP$_6$P and pulldown onto streptavidin beads.

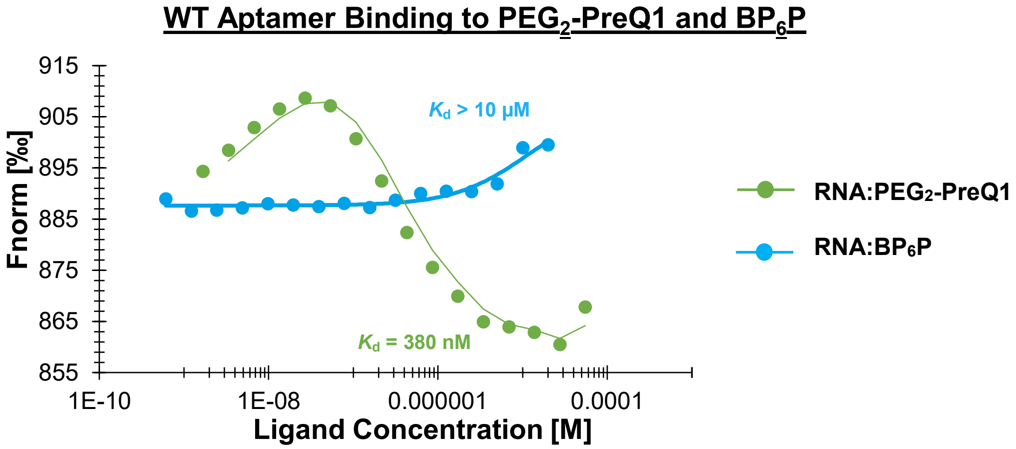


Figure B2: Excessive PreQ1 modification ablates aptamer binding. The WT PreQ1 aptamer binds tightly to PEG2-PreQ1 and poorly to BP6P by MST.

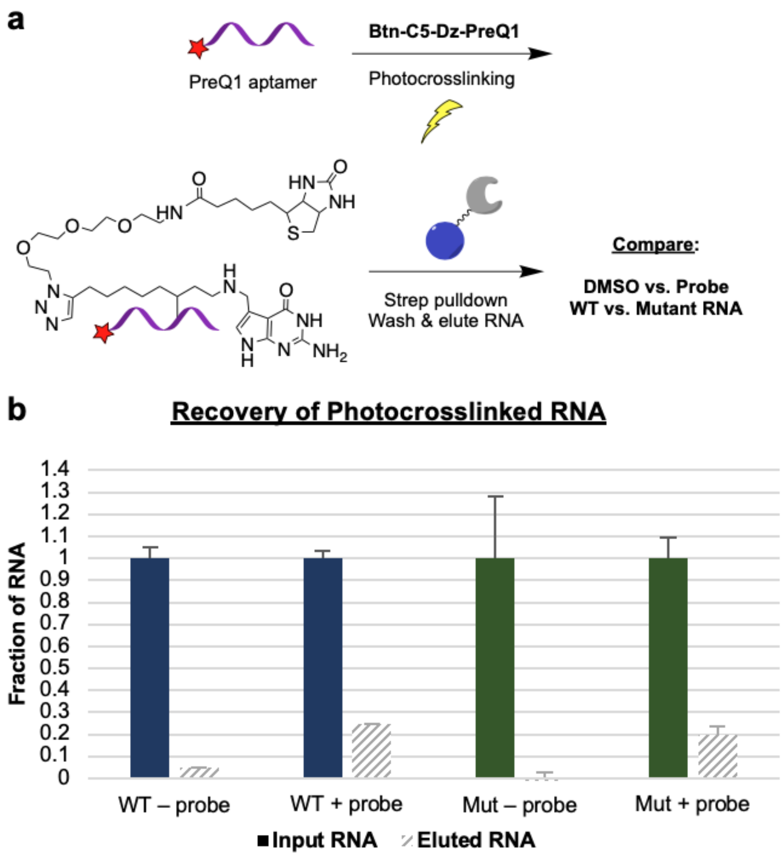


Figure B3: RNA enrichment with a biotinylated photocrosslinking probe. a PreQ1 aptamer enrichment scheme. **b** Recovery of covalently labeled WT (blue) vs. mutant (green) PreQ1 RNAs on streptavidin beads.

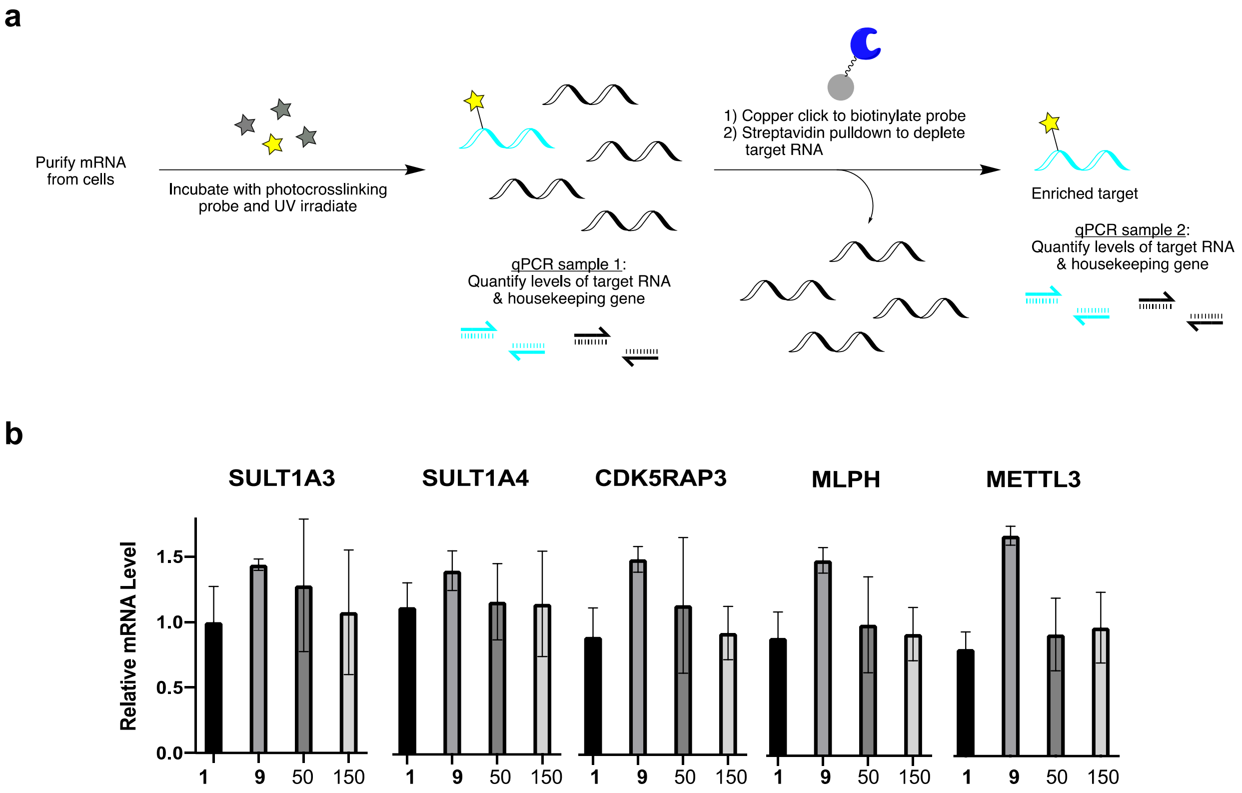


Figure B4: RT-qPCR validation of *in vitro* mRNA interactions with compound 9. a Scheme depicting the experimental workflow to validate interactions of PAL ligand **9** with mRNA hits of interest that were initially identified in the 10-compound, *in vitro* RNA target engagement screen. Poly-A-enriched RNA from MCF-7 cells was incubated with compound **1**, compound **9** (yellow star), or compound **9** + 50 or 150 μM free palbociclib (gray stars). Using gene-specific primers, levels of target mRNA were quantified in each sample before and after pulldown and enrichment. Transcript abundance was normalized to the housekeeping gene GAPDH. **b** Relative abundance of five mRNAs of interest under the four compound treatment conditions. Concentration-dependent decreases in abundance in the presence of free palbociclib indicate specific competition of PAL probe binding.

Appendix C: Data Omitted from Chapter 4

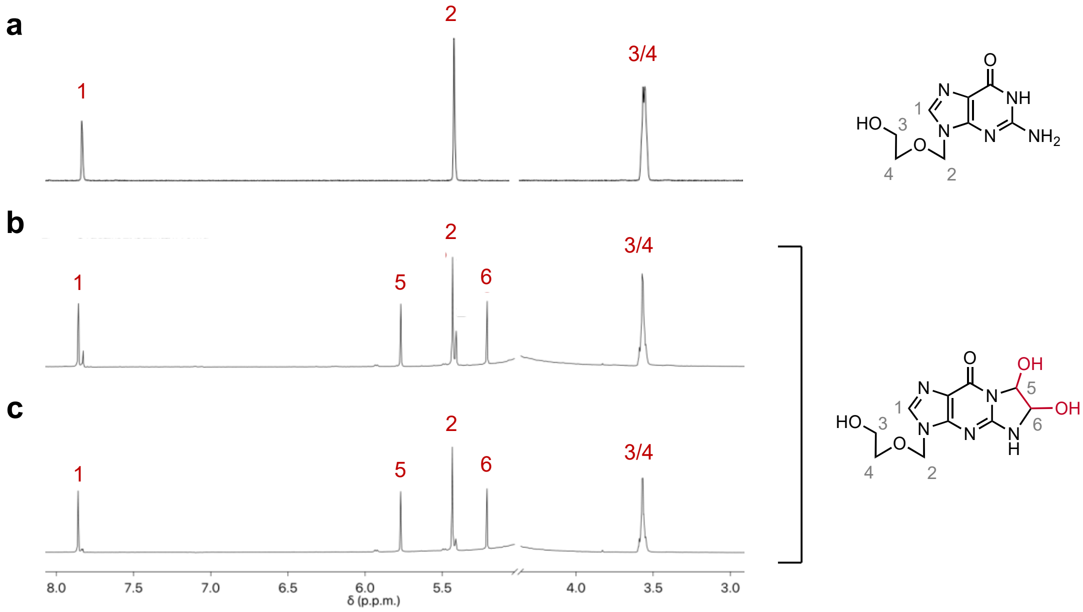


Figure C1: ^{1}H NMR of ACV before and after glyoxal caging. ^{1}H NMR analysis in D_2O of **a** ACV, or the reaction between ACV (100 mM) and **b** 0.8 equivalents of glyoxal or **c** 1.0 equivalent of glyoxal after heating in 1X PBS (pH 6) for 2 h at 50 °C.

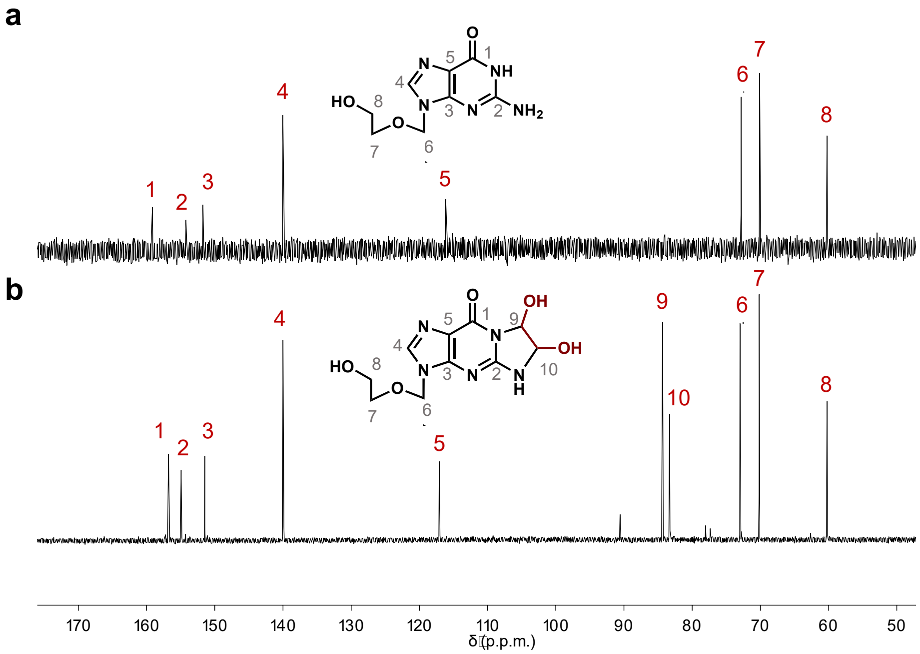


Figure C2: ^{13}C NMR of ACV before and after glyoxal caging. ^{13}C NMR analysis in D_2O of **a** ACV or **b** the reaction between ACV (100 mM) and 1.0 equivalent of glyoxal after heating in 1X PBS (pH 6) for 2 h at 50 °C.

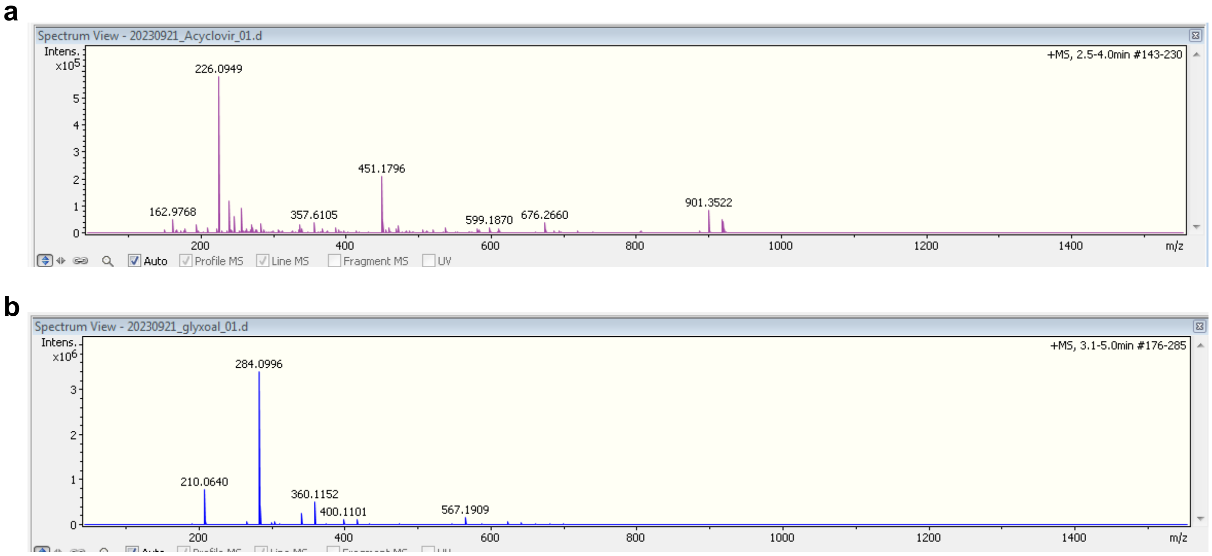


Figure C3: High-resolution ESI-MS of ACV before and after glyoxal caging. a ACV was confirmed by ESI-MS with (m/z) calcd for $C_8H_{11}N_5O_3$ ([M + H+]): 226.0935, found: 226.0949 ([M + H+]). **b** Caged ACV was confirmed by ESI-MS with (m/z) calcd for $C_{10}H_{13}N_5O_3$ ([M + H+]): 284.0989, found: 284.0996 ([M + H+]).

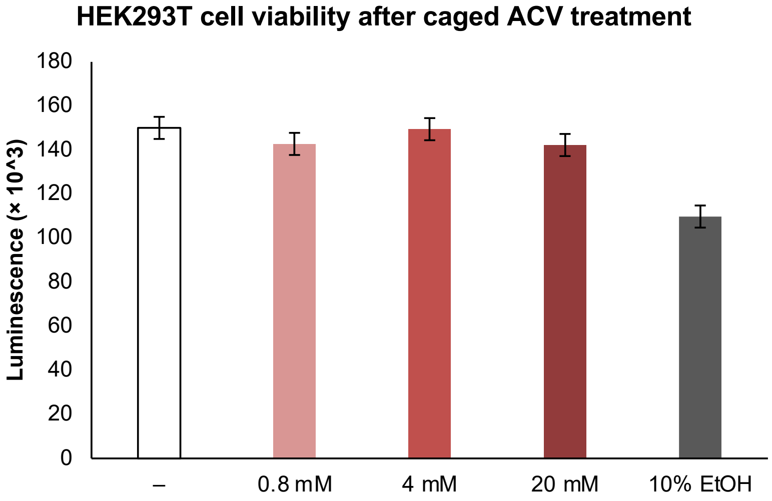


Figure C4: Cell viability dose response. HEK293T cell viability upon 4 h treatment with 0.8–20 mM caged ACV or 10% ethanol control (n = 2). Error bars indicate mean and standard deviation.

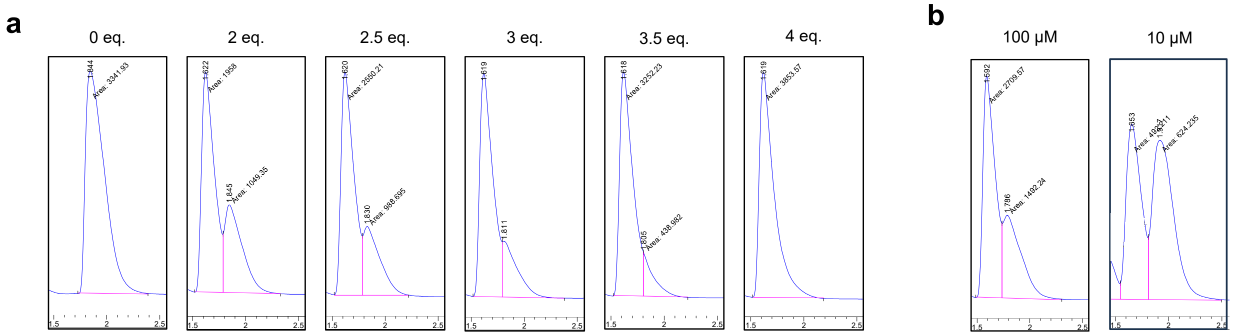


Figure C5: HPLC analysis of ACV-TP caging and decaging reactivity. a ACV-TP (10 mM) was reacted with 2–4 equivalents of glyoxal in 1X PBS (pH 6.8) at 40 ºC for 1 h to achieve complete caging. **b** ACV-TP that had undergone complete caging with 4 equivalents of glyoxal was diluted to 100 or 10 μM in 1X PBS (pH 6.8) and incubated at 37 ºC for 12 h to assess recovery of the decaged parent compound.

Table C1: Inhibition of HSV-1 cytopathic effects (CPE) in human foreskin fibroblast (HFF) cells. HFF cells were treated with five-fold serial dilutions of no compound (NC), tenofovir negative control compound (TNV), acyclovir (ACV), or caged ACV with or without prior decaging at 1 mM concentration in 1X PBS (pH 7.4) at 37 ºC for 1 d. The cells were then infected with HSV-1 and cultured over 5 d until 100% CPE was observed in NC wells. CellTiter-Glo reagent was used to measure half-maximal effective concentration (EC50) of HSV-1 CPE inhibition for the various compound treatments. All data was calculated from duplicate wells.

% CPE	NC (μM)	% CPE	TNV (μM)	% CPE	ACV (μM)	% CPE	Caged ACV (μM)	% CPE	Caged ACV, 1 d Δ (μM)
111.5	300	**49.2**	150	**34.1**	150	**35.5**	10	**37.3**	10
106.5	60	**61.0**	30	**28.7**	30	**61.4**	2	**39.6**	2
94.1	12	**82.9**	6	**28.4**	6	**98.7**	0.4	**92.3**	0.4
94.7	2	**99.0**	1.20	**42.0**	1.20	**100.0**	0.08	**91.4**	0.08
98.5	0.48	**97.7**	0.24	**74.9**	0.24	**78.0**	0.016	**92.8**	0.016
94.7	0.10	**97.8**	0.048	**70.1**	0.048	**104.0**	0.0032	**94.6**	0.0032
		EC_{50}		**EC_{50}**		**EC_{50}**		**EC_{50}**	
		142.13 μM		0.97 μM		5.53 μM		1.69 μM	

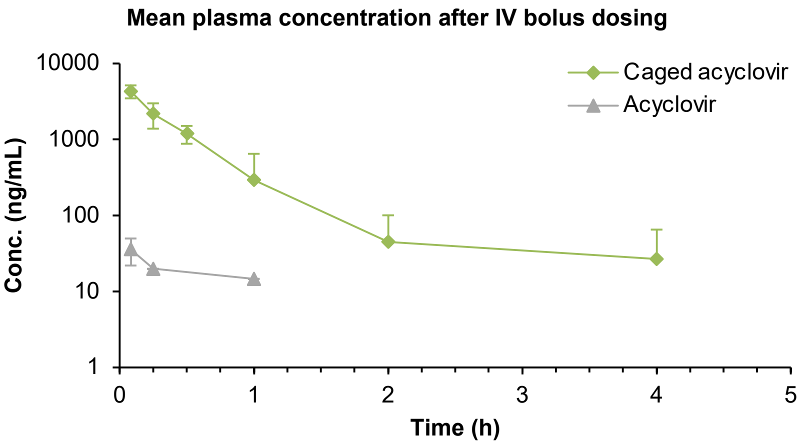


Figure C6: Circulation lifetime of ACV and caged ACV in mice. Male CD-1 mice (n = 3) were injected with an IV bolus containing 1 mg/kg of either ACV or caged ACV. Plasma samples were collected at 5 min, 15 min, 30 min, 1 h, 2 h, and 4 h post-injection, precipitated, and the supernatants analyzed for remaining drug concentration by LC-MS/MS.

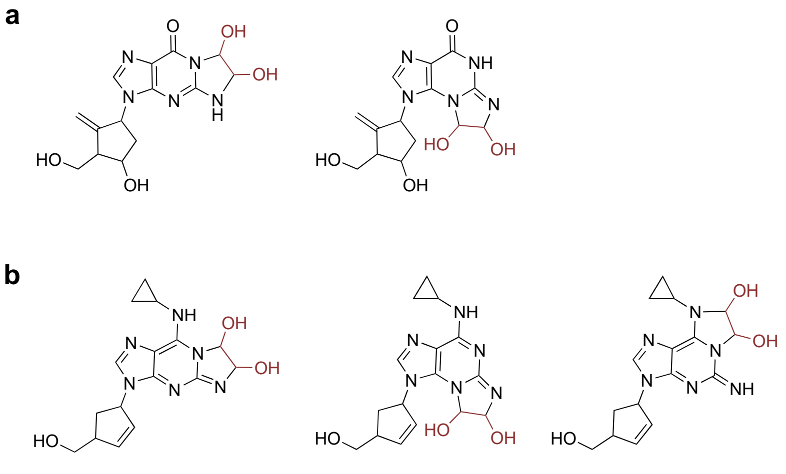
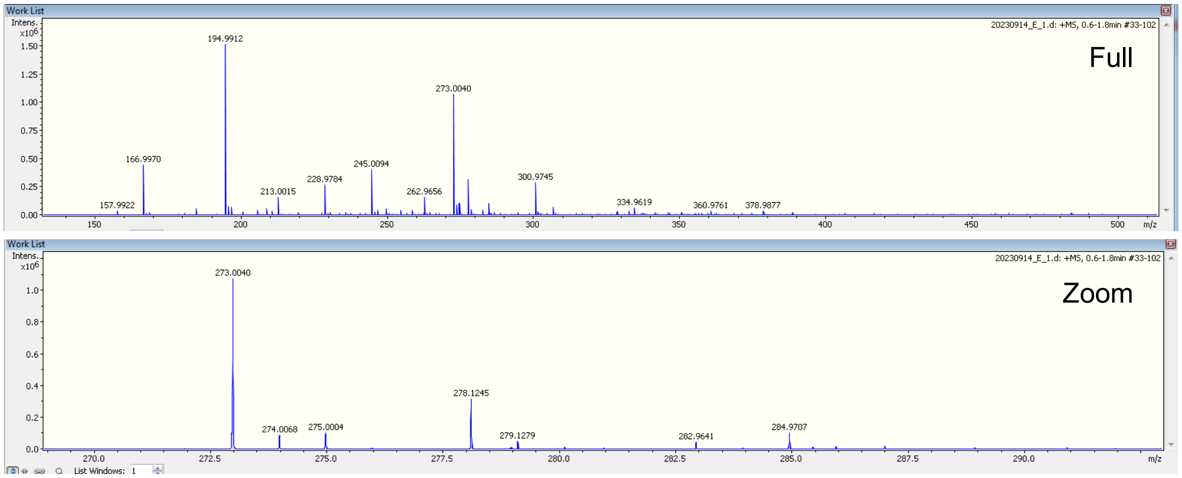
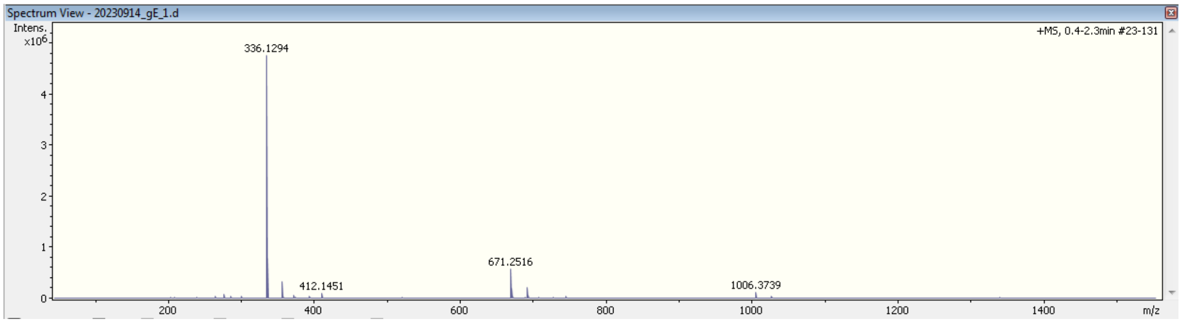


Figure C7: Predicted products of NRTI caging panel. Putative glyoxal caging products of **a** entecavir and **b** abacavir.

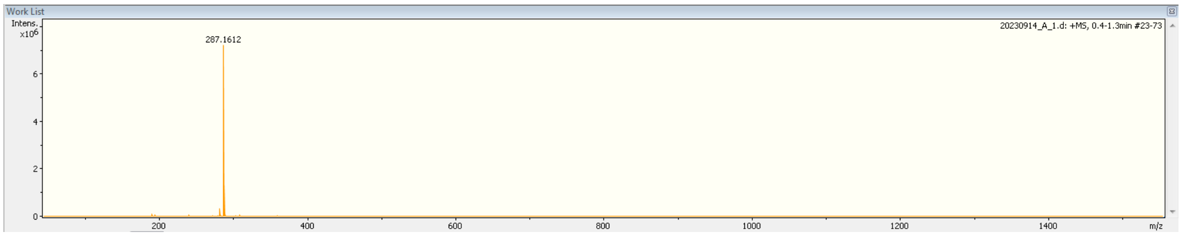
a


Work List
Intens.
x10^6
1.50
1.25
1.00
0.75
0.50
0.25
0.00
20230914_E_1.d: +MS, 0.6-1.8min #33-102
Full
194.9912
273.0040
166.9970
245.0094
228.9784
300.9745
213.0015
262.9656
157.9922
334.9619
360.9761
378.9877
150
200
250
300
350
400
450
500
m/z
Work List
Intens.
x10^6
1.0
0.8
0.6
0.4
0.2
0.0
20230914_E_1.d: +MS, 0.6-1.8min #33-102
Zoom
273.0040
278.1245
274.0068
275.0004
279.1279
282.9641
284.9707
270.0
272.5
275.0
277.5
280.0
282.5
285.0
287.5
290.0
m/z
List Windows: 1

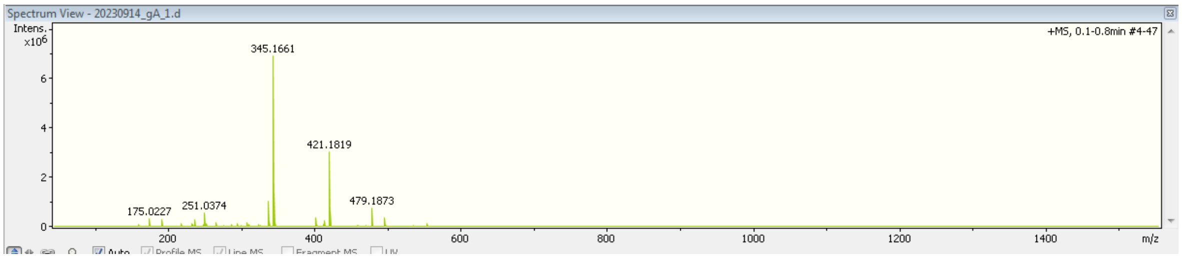
b


Spectrum View - 20230914_gE_1.d
Intens.
x10^6
+MS, 0.4-2.3min #23-131
336.1294
4
3
2
1
0
671.2516
412.1451
1006.3739
200
400
600
800
1000
1200
1400
m/z

c


Work List
Intens.
x10^6
20230914_A_1.d: +MS, 0.4-1.3min #23-73
287.1612
6
4
2
0
200
400
600
800
1000
1200
1400
m/z

d


Spectrum View - 20230914_gA_1.d
Intens.
x10^6
+MS, 0.1-0.8min #4-47
345.1661
6
4
2
0
421.1819
479.1873
175.0227
251.0374
200
400
600
800
1000
1200
1400
m/z
Auto
Profile MS
Line MS
Fragment MS
UV

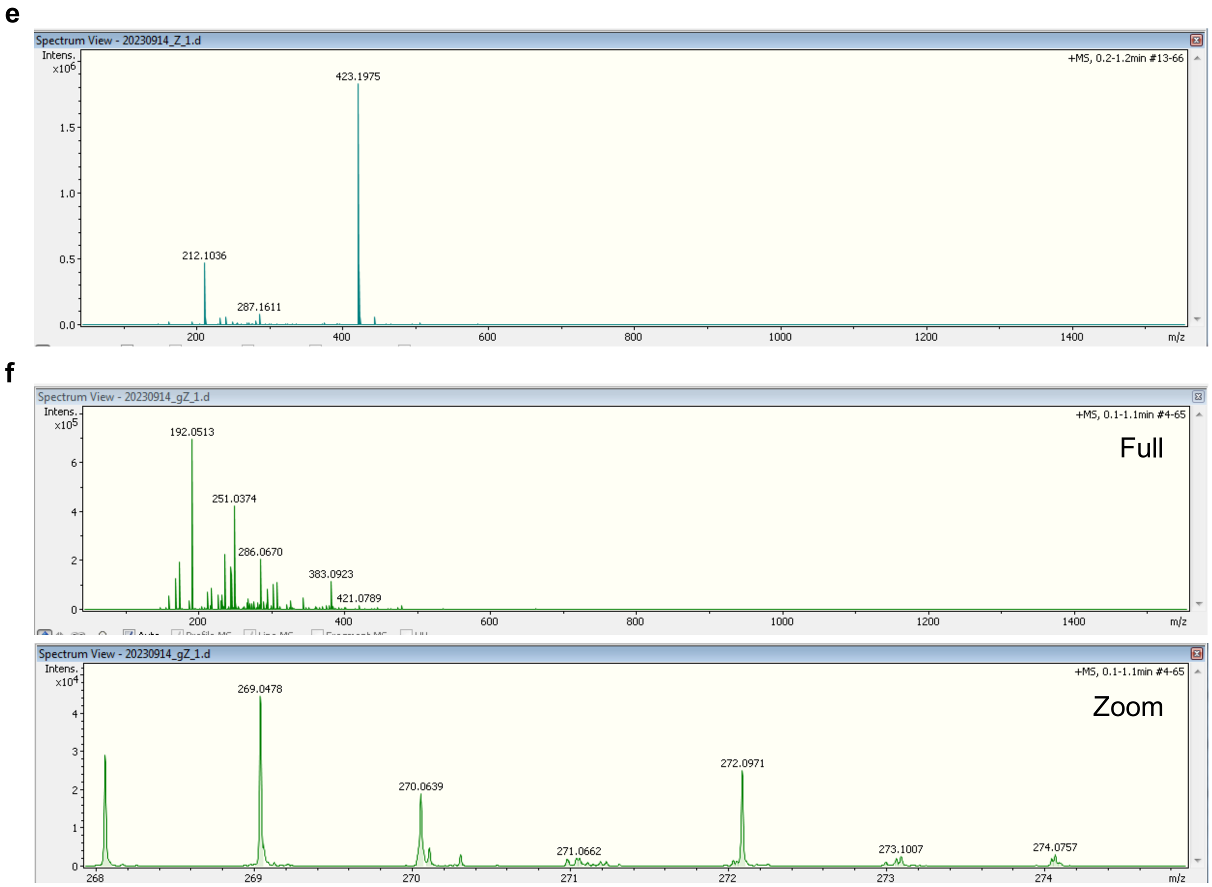


Figure C8: High-resolution ESI-MS of NRTI panel before and after glyoxal caging. a Entecavir was confirmed by ESI-MS with (m/z) calcd for $C_{12}H_{15}N_5O_3$ ([M + H+]): 278.1248, found: 278.1245 ([M + H+]). **b** Caged entecavir was confirmed by ESI-MS with (m/z) calcd for $C_{14}H_{17}N_5O_5$ ([M + H+]): 336.1302, found: 336.1294 ([M + H+]). **c** Abacavir was confirmed by ESI-MS with (m/z) calcd for $C_{14}H_{18}N_6O$ ([M + H+]): 287.1615, found: 287.1612 ([M + H+]). **d** Caged abacavir was confirmed by ESI-MS with (m/z) calcd for $C_{16}H_{20}N_6O_3$ ([M + H+]): 345.1675, found: 345.1661 ([M + H+]). **e** Zalcitabine was confirmed by ESI-MS with (m/z) calcd for $C_9H_{13}N_3O_3$ ([M + H+]): 212.1030, found: 212.1036 ([M + H+]). **f** Caged zalcitabine was confirmed by ESI-MS with (m/z) calcd for $C_{11}H_{15}N_3O_5$ ([M + H+]): 270.1084, found: 270.0639 ([M + H+]).