Distribution Agreement

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

Signature:

_____ _____

Steve Knutson Date

A Chemical Biology Toolbox for Studying A-to-I RNA Editing

By

Steve D. Knutson
Doctor of Philosophy

Chemistry

Prof. Jennifer M. Heemstra
Advisor

Prof. Vincent P. Conticello
Committee Member

Prof. Nathan T. Jui
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

A Chemical Biology Toolbox for Studying A-to-I RNA Editing

By

Steve D. Knutson
B.S., University of Illinois at Urbana Champaign, 2010
M.S., University of Illinois at Chicago, 2015

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
2021

Abstract

A Chemical Biology Toolbox for Studying A-to-I RNA Editing

By Steve D. Knutson

RNA carries and transports information within cells to significantly influence their function and behavior. After transcription, RNA is also extensively modified by a number of enzymes to further alter this information. Adenosine-to-inosine (A-to-I) editing is one of the most widespread and impactful RNA modifications, and is catalyzed by adenosine deaminases acting on RNA (ADARs). Editing is essential for a number of biological processes, and dysfunctional editing is directly linked to several disease pathologies. Despite this major importance, our overall understanding of A-to-I editing is quite limited. In particular, existing technical challenges obscure the true prevalence and landscape of A-to-I editing in the human transcriptome, and it is unclear why certain sites are edited over others and what precise function they each serve. While it is known that dysregulated editing is linked with numerous diseases, the exact molecular and cellular mechanisms responsible for this relationship are also poorly understood. We also have indirect evidence that different cell types and tissues display vastly different editing patterns, but there are no approaches to measure these differences. As a result, we have little understanding of how this heterogeneity affects overall tissue and organ function which might contribute to disease. Given these limitations, the primary goal of my research has been to develop improved tools to better probe A-to-I RNA editing and characterize its biological functions. This thesis aims to highlight these efforts, and in Chapter 1 I will first provide an overview of the biological roles of A-to-I editing and summarize current methodology for studying and harnessing this modification. In Chapters 2 and 3, I then describe explorations of inosine labeling using acrylamide derivatives to enable affinity enrichment and chemical profiling of A-to-I edited transcripts. In Chapters 4 and 5, I develop a protein-based platform for binding and detecting inosine in RNA using Endonuclease V (EndoV), and highlight the powerful applications that have been leveraged from these investigations. In Chapter 6, I describe our use of the covalent denaturant glyoxal as a synthetic nucleic-acid modification which enables thermoreversible control over the structure and activity of a variety of biomolecular constructs. Finally, in Chapter 7 I summarize the implications of these present studies, discuss opportunities for advancing each technology platform, and describe future perspectives and challenges for studying A-to-I RNA editing.

A Chemical Biology Toolbox for Studying A-to-I RNA Editing

By

Steve D. Knutson
B.S., University of Illinois at Urbana Champaign, 2010
M.S., University of Illinois at Chicago, 2015

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
2021

Acknowledgements

Somehow, I am about to receive my doctorate in chemistry. While I have worked very hard to get to this point, I certainly could not have done this alone. Additionally, the experience of graduate school has been uniquely challenging and allowed me to grow in many ways, and I wish to individually thank the people who supported me and fostered this development.

I first want to thank Gary and Susan Knutson for being exceptional parents. You both made immense sacrifices to provide for me when I needed you most, and we had many fun memories as a family. I was also quite rebellious as a teenager and engaged in several bouts of shenanigans. Thank you for never giving up on me and always remaining patient, dedicated, and loving. You always taught me that life can sometimes be cruel and unfair, few things come for free, and that if I worked hard enough and set my mind to it, I could achieve anything. This instilled in me a singular resolve and work ethic that has propelled my career to this day. To Kyle, thank you for being an outstanding example of what an older brother can and should be. I still look up to you in many ways, and you've had a huge impact on my life. Even as young kids, whatever you were doing, I wanted to be doing it too, and you've always driven me to be best version of myself. When we were younger our parents would often say, (especially when we were fighting and otherwise very annoyed by each other) "you know you're lucky to even have a brother, right?" As an adult, I can say with utmost clarity that I completely understand this now, and I am so grateful to have you in my life. To Lyndsay, I feel equally lucky to have you as a new sister, and it's been amazing to have you in our family now. You both were also very helpful in my aspirations for even attending graduate school, and while I was not accepted at Stanford as planned, you were both instrumental in crafting personal statements and forming my career goals.

I am also grateful for the support of Frank, Diana, Katy and Kevin Laux. I love spending holidays with all of you, and your family has been a wonderful complement to ours. I can't wait until we can all break out the Trivial Pursuit board once more. Katy and Kevin were both also

excellent role models for Kyle and myself growing up, and Kevin was especially helpful in illustrating how to pursue a PhD and what it looks like to be successful.

I next want to thank Dr. Allyson Jane Boyington, who is a master chef, tailor, and a molecular designer extraordinaire in her own right, for being the perfect girlfriend and the most special person in my life. Thank you for being patient and persistent in your pursuit of dating me, and I apologize for being focused on research in much of our initial interactions. While “better late than never” remains a true and heartwarming thing to say, my one major regret is not acting on these mutual affections earlier. You’ve always supported my dreams, and have been patient and kind in teaching me organic chemistry concepts and photoredox catalysis. You continually inspire me to be a better man and scientist, and I am so excited to share life with you. I can (somewhat) draw electron pushing diagrams, my figures have never looked better, my writing has never been stronger, and I’ve never been happier in my life. I love you and can’t wait to see where we go together.

From a professional and scientific training perspective, I first want to thank Drs. Christopher Schoenherr and Neal Kitchen at the University of Illinois at Urbana-Champaign for offering me my first research opportunity as an undergraduate. This was a truly formative experience and provided me with a solid foundation of laboratory techniques and helped shape my current attitudes toward research. Neal was also instrumental in helping me secure my first job after graduating, and I would not be here without his mentorship. I next want to thank Thermo Fisher Scientific and all of the amazing people that I worked with and met during my time there. In particular, I first want to thank Dr. Liam F. Garrity as the manager of the Analytical Services Laboratory (ASL) at Thermo Fisher in Rockford, IL. Liam offered me my first job as a Quality Chemist, and he taught me how to think critically about analytical biochemistry. Liam was also integral in shaping my professional demeanor, and he importantly guided my political tact, networking style, and overall sense of humor that I currently use when interacting with senior personnel. He was always supportive of my every career aspiration, and truly wanted me to

succeed in whatever path I chose. He remains one of the best supervisors and mentors I've ever had. I also want to thank Gary Judd, as he was my primary technical mentor upon joining ASL. Gary was an incredible guide for me on both personal and professional aspects, and he personally taught me the majority of the laboratory techniques that I still employ today, especially in mammalian cell and tissue culture. It is often your (constructively critical) voice that I still hear in my head when I design, execute, and analyze experiments. ASL as a whole was also a near-perfect setting for me to development and train as early-career scientist. The breadth of scientific techniques employed in this group is impressively wide, and I credit nearly all of my own laboratory expertise and dexterity to my time training with these talented people. In particular, I thank Howard Coyer, Mike Stanaitis, Brandon Smith, Tim Ryan, Barb Olson, Zlatana (Z) Kuljanin, Scott Starwalt, and Brian Wolf, whose combined 100+ years of expertise greatly shaped my skillset.

I would also like to thank the department of Research and Development (R&D) at Thermo Fisher in Rockford, IL, with whom I worked with for 2 years before attending graduate school. In particular, Dr. Surbhi Desai was an especially important mentor to me, and not only helped me obtain a position as a research scientist, but was also a member of my Master's thesis committee. She still writes recommendation letters for me to this day, and her mentorship continues to make a significant positive impact on my career. I would also like to thank the immunoassay development group, including Kara Machleidt, Kari Severson, Matt Baker, Doug Hughes, Deven Etnyre and Carrie McCray. This group encouraged my abilities as an independent scientist and greatly increased my confidence. Additionally, the critical skills I developed in validating antibodies and publishing experiments using these materials has shaped my research style and influenced my projects significantly. I would also like to thank the many talented individuals in this department, including Chris Etienne, Greg Kilmer, Eric Hommema, Suzanne Smith, Sergei Snovida, Joanna Geddes, Kate Herting, Korin Pathammavong, Ryan Bomgarden, Penny Jensen, Erum Raja, Marie Nlend, and Nikki Jarrett. This group provided great companionship, daily doses of laughter, rounds of golf, and many drinks after work that were essential for my time at Thermo

Fisher. Christopher and Thomas Wojewodzki also deserve very special thanks, as their friendship was very significant to me. I did not know anyone when I moved to Rockford IL, and they both welcomed me into their lives like I had known them for years, and I will never forget their kindness.

I would also like to express my gratitude toward the University of Illinois Chicago College of Medicine at Rockford, which accepted me into their Master's program in Medical Biotechnology (MBT). I did not have excellent grades from my undergraduate studies that are typically required for graduate programs, and Dr. Khalifah Sidik saw potential in me and helped facilitated my acceptance. My time in the MBT program allowed me to take more advanced courses and eventually prove that I was capable of high level graduate study. Additionally, the MBT Master's thesis was my first major opportunity to independently plan and execute a research project of my choosing, and the experience was transformative for my confidence and career aspirations. I especially thank my co-authors of this study (Drs. Erum Raja, Ryan Bomgarden, Marie Nlend, Aoshuang Chen, Ramaswamy Kalyanasundaram, and Surbhi Desai) as well as those that made this research possible (Howard Coyer, Liam Garrity, Gary Judd, Eugene Cichon, Andrew Canciamille, Khalifah Sidik, and Neelu Puri). My thesis committee deserves special acknowledge, which included Drs. Surbhi Desai, Ramaswamy Kalyanasundaram Ryan Bomgarden and Aoshuang Chen. Dr. Ramaswamy was particularly helpful to me during my graduate studies, and he significantly helped me shape my research project. I also looked up to Ryan immensely, and we would talk at length about industrial and academic research and what it meant to execute a project well. I am very fortunate to have this group of exceptional scientists as a Master's thesis committee, and they pushed me to become a better scientist, writer, and educator. They also collectively wrote countless letters of recommendations for subsequent PhD and fellowship applications, and I am very grateful for their efforts.

After applying to 15 PhD programs, I was rejected by all of them except for one – The University of Utah Biosciences Program. I cannot begin to express my gratitude and thanks for this decision. Simply put, I would not be receiving a PhD at all were it not for this acceptance. I

was certainly a non-traditional candidate, and Utah saw enough potential in me to give me a chance to prove myself. While I do not know exactly who to thank for this decision, based on visitation observations and anecdotal evidence I can reasonably guess that Drs. Shawn Owen and Michael S. Kay were instrumental to this decision. Regardless of their responsibility, I also thank both of them for their support and mentorship in my early academic career.

Moving to Salt Lake City, UT to begin my PhD, I was very lucky to reconnect with family living there. In particular, the entire Barrell family made my stay in Utah an amazing experience. I was not only able to find excellent housing in an apartment attached to Danny and Kelsey Barrell's home in Sugarhouse, but I was immediately welcomed into the tight-knit and vibrant social life that they all enjoy. From the bottom of my heart, I thank Eleanor, Dan Sr, Arleen, Jeff, Denise, Danny, Kelsey, Scott, Keith, Carly, Andy, Natalie, Nick, and Tyler. We climbed mountains, backpacked, skied, cooked amazing meals together, and shared countless laughs. I was truly lucky to have such amazing family support when moving to a new city, and I would not be the person I am today without each and every one of you. Keith, Scott, and Natalie deserve special thanks, in that we spent a great deal of time together when I first arrived in Utah. They brought me into their own wide social circles, and introduced me to a host of outdoor activities and adventures in the area. In my first semester of graduate school, I also unfortunately developed a severe case of appendicitis and was hospitalized for a week. Dan Sr. and Arleen both visited me almost daily in the hospital, and Arleen was a key companion during my recovery. I am so grateful that I was surrounded by family during one of the most physically challenging periods of my life. I would also like to give very special thanks to Danny and Kelsey. Not only they did they welcome me into their home as a roommate-by-extension, but we became close friends and shared many adventures together. We talked at length about a wide-range of subjects, including but not limited to current political events, climate change, fine cuisine, and outdoor recreation. I truly cherished our time together and I was very sad to leave when our lab eventually moved to Emory. Overall, my time in Utah with the Barrells will always hold a special place in my heart, and I can say that I

grew more as a person during that one year than perhaps of any of the previous. All of you taught me much about what it means to be a good person and how to push my physical and mental limits for adventure, and in the process provided me with a confidence to tackle almost any challenge.

I would next like to thank Emory University for welcoming our lab after Jen accepted a faculty position in the Chemistry department. Packing up and moving across the country to a new University is a daunting task for any researcher, and Emory made this a smooth, straightforward, and painless process. I would like to especially thank Todd Polley, Kira Walsh, and Ana Velez for logistically arranging our transfers to make sure we had everything we needed. My PhD thesis committee members, comprised of Profs. Vincent P. Conticello, and Nathan T. Jui, have also provided invaluable support to me during my research. I especially thank Vince for writing several letters of recommendations for various fellowship applications throughout my graduate work. I also thank Nate for sharing his graduate research experiences and providing advice about choosing a lab for postdoctoral research – this was integral to my current career trajectory.

In the second year of my PhD, I was also accepted into the Atlanta chapter of the Achievement Rewards for College Scientists (ARCS) program, and I graciously thank this organization for their support and for providing me access to the valuable network of past and present ARCS scholars. I would also like to thank Drs. Cora MacBeth and Tim Raines for accepting me as a fellow in the Scholarly Inquiry and Research Experience (SIRE) program. This was a tremendously valuable experience for me to grow as an educator and to help facilitate undergraduate research throughout different disciplines at Emory.

The Heemstra laboratory has been an amazing place to work, and I want to thank all of the incredible graduate students and post-doctoral associates that I've met and interacted with over the years. In particular, I'd like to first thank Alex Rangel, Zhesen Tan, Hershel Lackey, Brad Green, Erin Price, Amber Peterson, Zhesen Tan, Satish Ellipilli, and Miguel Pereira for being so welcoming when I was rotating and first joined the lab. I'd also like to thank current and previous members Colin Swenson, Tewoderos (T) Ayele, Meredith Henry, Travis Loya, Jackie Gayton,

Misael Romero, Aimee Sanford, Rachel Bender, Hector Argueta-Gonzalez, Brea Manuel, Diane Karloff, Alex Quillin, Ansley Felix, Kristen Patterson, Trent Stubbs, and Sierra Starling. Research collaborations with Colin and T were especially fruitful during my graduate career, and Aimee in particular has been a significant research partner. It has been incredibly fun and rewarding to work closely with them. I also want to recognize the talented undergraduates that I have had the pleasure to mentor and work with at Emory, especially Megan Korn, Deanna Dailey, Ryan Johnson, and Michael Hwang. Seeing all of you grow as independent scientists has been very rewarding for me, and has helped me develop as a mentor and teacher significantly. I have also been lucky to work with several new and talented graduate students during the last year of my PhD, and I want to thank Ansley Felix, Alex Quillin, and Trent Stubbs. Each of you three are now carrying on a part of my research, and I've been so impressed at your ability to think critically and execute projects. The future of Heemstra lab is especially bright because of your talent, and I expect great things to come.

I especially want to recognize Colin Swenson and Tewoderos (T) Ayele, who were also my roommates for 3+ years when we moved to Atlanta. We spent countless days, nights, and weekends together, and I'm proud to call you both my close friends. We also collaborated on several papers, and it was especially rewarding to be able to share my personal and professional life with each of you. T and I also attended several conferences together and were able to visit Italy as part of a GRC, and I will always remember that week as an incredibly fun adventure with him. Colin and I also worked together extensively during his last few months in the lab, where he taught me many organic synthesis techniques. This period was especially fun and critical for my future post-doctoral work. The house has not been the same without you both, and I miss our time together. However, I also want to thank Mark Maust, who moved in with us after T left. Mark perfectly fit into our house "culture" and I am proud to call him my close friend now. He is a great person, an excellent roommate, and I love (when I actually do) beat him in Settlers of Catan. I

wish him great luck in his own graduate career at Emory, and have no doubts that he will be incredibly successful.

Lastly, I want to thank my advisor, Prof. Jennifer M Heemstra. Simply put, she has been an exceptional mentor and I owe the vast majority of my success to her guidance and support. I began speaking with Jen by phone in early 2015 before applying to graduate school, and it was immediately apparent that she was not only a brilliantly creative scientist but also a caring mentor for her students. This was absolutely confirmed throughout my rotation in her lab and first year of my PhD, and I had zero doubt that this group was a perfect fit. Jen has always treated me with a high amount of respect, and I always felt like her equal when discussing science, planning experiments, writing papers, or submitting grants. She also gave me near-complete intellectual freedom on almost any project of my choosing, and allowed me to explore a variety of new techniques and disciplines. This gave me extensive practice in project planning and design, and significantly improved my skillset in these areas for the future. Jen is also an excellent science communicator, and I have learned a wealth of writing and presentation skills under her guidance. These experiences have greatly enhanced my confidence as an independent researcher, and her encouragement has shaped my ambition to seek out innovative solutions to highly difficult problems. This professional relationship has not only been extremely productive (6 research articles, 1 review, 2 protocol chapters, 1 patent submission, 13 posters, 3 research grants totaling ~$1.5M, 5 semesters teaching, and 4 new courses developed), but I can also easily say that Jen has become a close friend and will remain a significant personal role model. I've gained extensive scientific training from her, but more importantly I've also learned how to be a highly effective leader of a research group and how to mentor other early-career scientists. I would not be the person and scientist I am today without her. I am excited to watch her inevitable success as she moves forward with her own career, and I look forward to a lifetime of friendship and guidance from her. I will miss her greatly as I move on to the next phase of my career, and I will be forever proud to call myself one of her students.

This dissertation and the body of work contained within it is a testament to all of the love and guidance given to me throughout my career, and is dedicated to everyone who has supported me.

Thank you all for the wings.

Table of Contents

List of Tables and Figures:

Chapter 1

Introduction

1.1 RNA Modifications

RNA is a key information-carrying biomolecule that controls cellular function in all living organisms. After transcription, RNA molecules undergo a host of processing and modification steps to yield mature transcripts. In addition to capping, polyadenylation, and splicing steps that yield mature mRNA,[1] individual nucleobases within transcripts can also be modified by a number of enzymes. These processes are now described as 'RNA editing' or 'epitranscriptomic modifications' (**Figure 1.1**)[2-5] and they affect nearly all types of RNA and encompasses a significant and increasing number of known modifications. **Figure 1.1b** displays a select group of these modified bases, and while there are many (>100) additional modifications known and more continue to be discovered, these highlighted alterations are the most prevalent in the human transcriptome and significantly influence RNA coding and function.[2-4] Adenosines in particular are frequently edited in humans, and different modifications are installed by several enzyme classes. Adenosine-to-inosine (A-to-I) conversions are catalyzed by adenosine deaminases acting on RNA (ADARs) (**Figure 1.1b**), which in turn alter the base pairing properties of the nucleobase. Similar to epigenetic modifications in DNA, methylation of certain nucleotides in RNA is emerging as a widespread cellular event for tuning RNA function and regulating gene expression.[6] In particular, N^6-methyladenosine (m^6A, **Figure 1.1b**) has been identified as the most frequent RNA editing event in humans[7-9] and appears to both enhance protein translation[10] and promote phase-separation of different transcripts into cytosolic granules.[11-12] Additional methylation of m^6A at the 2′ OH position (m^6Am) is also prevalent, and is enriched at the 5′ termini of mRNAs.[7] Uridines can also be converted into pseudouridine (Ψ), and while originally discovered as a common modification in both transfer RNA (tRNA) and ribosomal rRNA (rRNA),[13] these edits have now

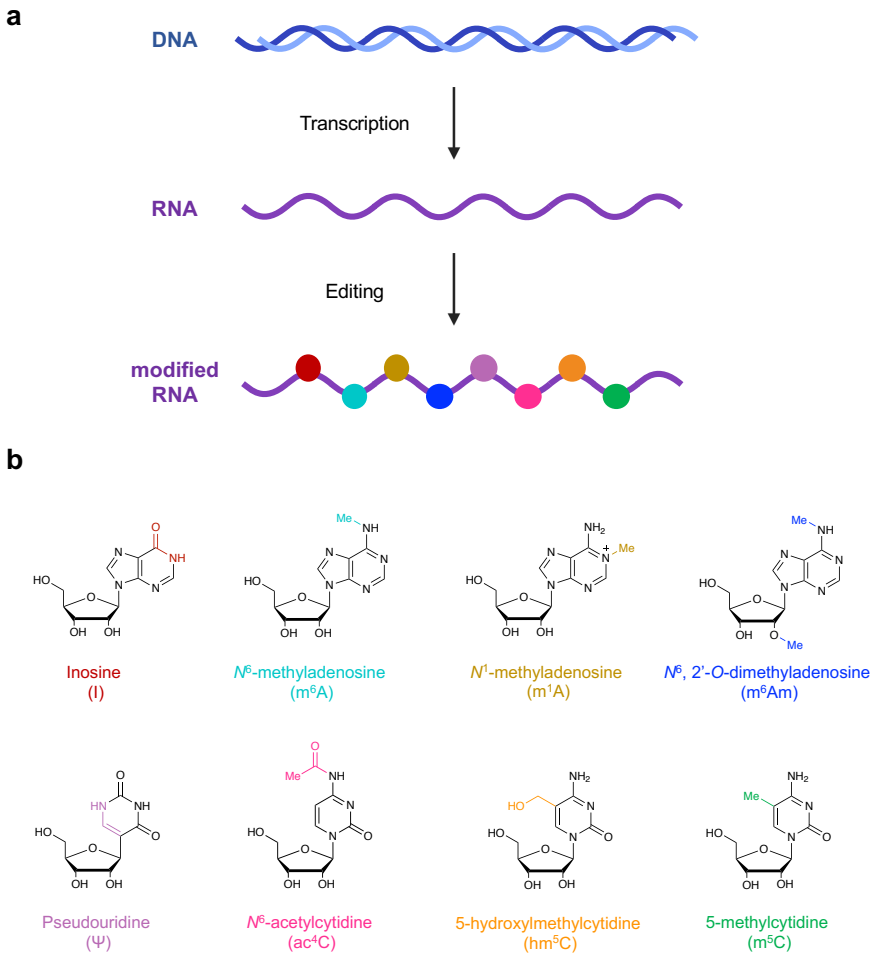


Figure 1.1 RNA editing introduces chemical and structural alterations to different nucleobases. a) General schematic of RNA editing as a post-transcriptional processing step. b) Selected chemical structures of several known modified ribonucleosides, with key molecular changes highlighted in color. Reproduced with permission.[14]

been detected in thousands of human mRNAs.[15-16] Similarly, cytidines can be converted to N^4-acetylcytidine (ac^4C),[17-18] 5-hydroxylmethylcytidine (hm^5C),[19] and 5-methylcytidine (m^5C),[20] with each of these modifications playing key roles in translational quality control and RNA stability. The discovery of these modifications represents an exciting new frontier in RNA biology, and determining their precise cellular functions will likely yield significant insight into basic cellular physiology. However, elucidating these roles has presented a significant challenge for the field and required creative strategies to detect and map these events. A-to-I editing especially

exemplifies this challenge and promise, and subsequent sections of this chapter describe this modification in greater detail.

1.2 A-to-I RNA Editing and ADAR Enzymes

Adenosine-to-inosine (A-to-I) editing is one of the most widespread and impactful RNA modifications, and is catalyzed by adenosine deaminases acting on RNA (ADARs).[21] A-to-I editing is also one of the most highly conserved modifications, occurring in many eukaryotic species and is especially enriched in metazoans (animals).[22] Deamination changes the structure and hydrogen bonding pattern of the nucleobase, and resulting inosines instead hybridize with cytosine to effectively recode these sites as guanine (**Figure 1.2**). RNA editing is widespread in both coding and non-coding transcripts, producing dramatic changes in overall cellular function. Editing of mRNAs can directly alter protein sequences through modification of codons, or by altering splice sites and regulatory elements in untranslated regions. A-to-I editing events in non-coding RNAs, particularly microRNA (miRNA) and small-interfering RNA (siRNA) precursors, can also produce significant changes in their biosynthesis, trafficking, specificity, and gene regulation properties, in turn dramatically affecting overall cellular behavior.[23-24] Although modified by a separate but similar class of enzymes, certain transfer RNAs (tRNAs) also undergo A-to-I editing in the N34 wobble position of anticodons, allowing these tRNAs to recognize multiple codons during translation.[25-26] While it is unknown why RNA editing first evolved, its pervasiveness across eukaryotes and participation in critical biological processes underscores its broad importance.

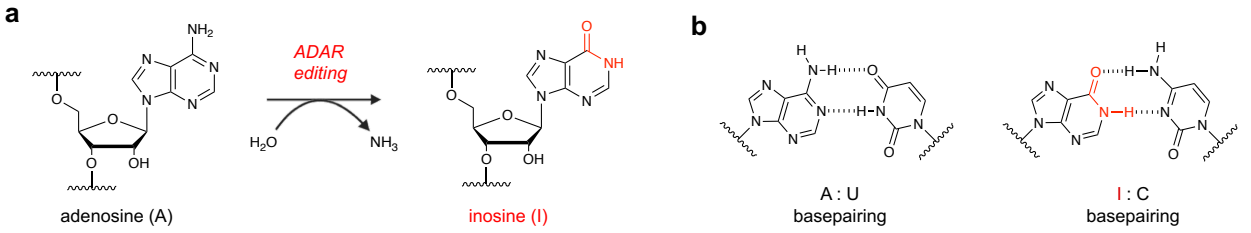


Figure 1.2 Structural alterations resulting from A-to-I RNA editing. a) Adenosine deamination is catalyzed by ADAR enzymes, incorporating oxygen from water and releasing ammonia (NH_3). b) Editing introduces a basepairing change, with inosine (I) preferentially hybridizing with cytidine (C).

ADAR1 and ADAR2 are the primary enzymes responsible for catalyzing A-to-I editing in humans.[21, 23] ADARs have strong affinity for double-stranded RNA (**Figure 1.3**),[21] and widespread *Alu* elements throughout the transcriptome form long inverted repeats which are extensively edited.[27-28] Interestingly, while ADARs can edit adenosines in fully complementary RNA duplexes, they appear to demonstrate greater activity toward mismatched base pairs, particularly in structural bulges when adenosine is opposite cytidine.[29-30] Conversely, editing activity is decreased on adenosines in very short (>15 bp) or only partly duplexed neighboring regions, strongly suggesting that RNA secondary structure plays a critical role in editing site selectivity.[31] Nearest neighbor motifs in dsRNAs and other local regulatory sequence elements are also beginning to emerge as integral to ADAR targeting.[30, 32]

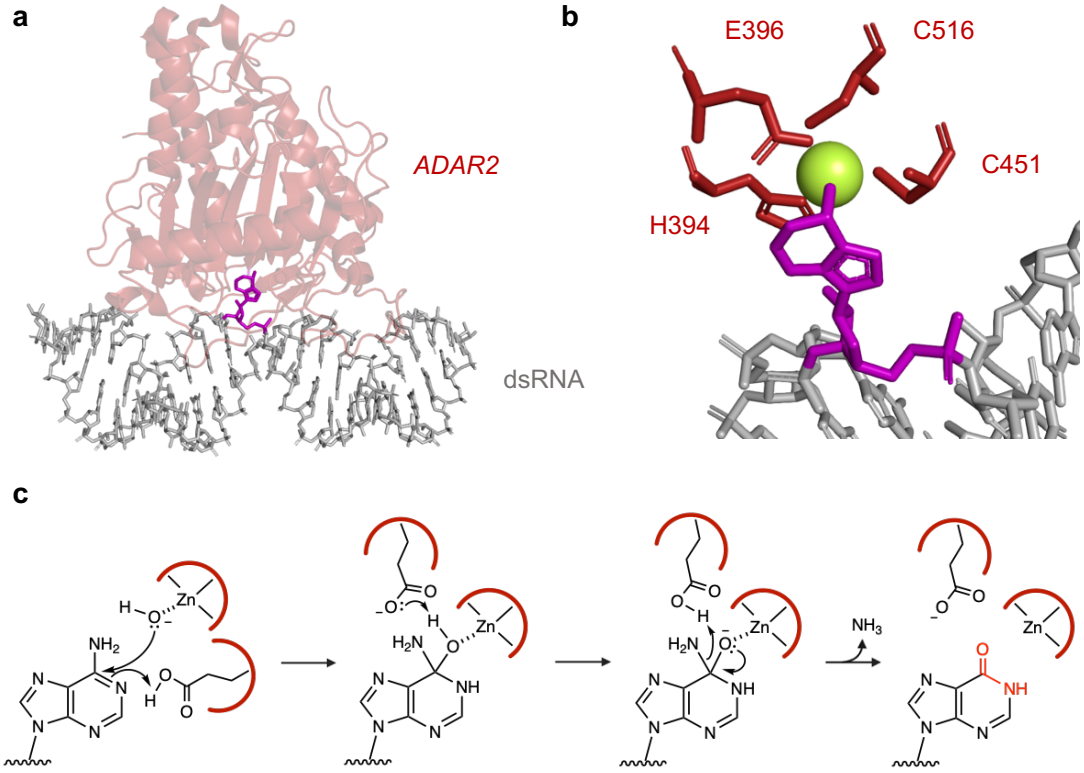


Figure 1.3. ADAR substrate recognition and catalytic mechanism. a) Crystal structure (PDB 5HP3) of the deaminase domain from human ADAR2 (red) bound to a dsRNA substrate (gray) with target adenosine (magenta). b) Key active site residues surround the adenosine substrate and facilitate deamination. Zinc co-factor shown as a green sphere. c) ADAR catalytic mechanism. A coordinated water molecule is activated by a glutamic acid in the active site, facilitating C^6 hydroxide attack. Glutamate then participates in a proton shuttling sequence to form a C^6 carbonyl and release ammonia (NH_3).

Mechanistically, ADARs convert adenosine to inosine through an exocyclic deamination as shown in **Figure 1.3**. The target adenosine is first "flipped out" (**Figure 1.3a**) to position the nucleobase into the enzyme active site. Zinc is a vital cofactor in ADAR catalysis, with a conserved H394, C451, C516 triad stabilizing the cation and facilitating coordination with a water molecule (**Figure 1.3b,c**).[33] A key proton-shuttling glutamate residue (E396) first deprotonates this water molecule to generate a reactive zinc-hydroxide that attacks C^6 of adenosine to form a hydrated tetrahedral intermediate. E396 then abstracts a proton from the newly formed O^6 hydroxyl, protonating the N^6 amino group and enabling the release of ammonia (**Figure 1.3c**).[30, 32]

Structurally, ADARs are highly conserved and modular, containing discrete domains for dsRNA binding, nuclear trafficking, and catalytic deamination (**Figure 1.4**).[23, 34] ADARs 1, 2 and 3 are the predominant isoforms in animals, with ADARs 1 and 2 responsible for the majority of editing activities. ADAR1 is the primary enzyme for catalyzing A-to-I editing in non-coding and repetitive RNA regions, while ADAR1 appears to be implicated in protein recoding activities in key mRNA editing events. ADAR1 is further divided into two variations; a larger ~150 kDa enzyme (ADAR1p150) and a smaller ~110 kDa (ADAR1p110) isoform, and the expression of these variants appears to be regulated on independent promoters.[35-36] While ADAR1p110 is constitutively expressed, ADARp150 biosynthesis is upregulated in response to certain external stimuli, including interferon.[37] ADAR1p110 also importantly lacks a nuclear export signal (NES), requiring mediation by Ran-GTPases for nucleocytoplasmic shuttling.[38] ADAR1 also contains a

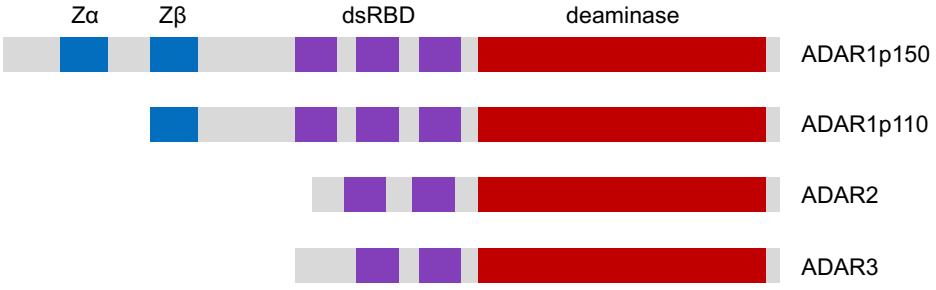


Figure 1.4 Structural domains of ADAR enzymes. These include Z-DNA binding domains (Zα and Zβ, blue), double-stranded RNA binding domains (dsRBD,purple), and deaminase domains (red). Boxes indicate relative position and size of each domain. Figure reproduced with permission.[39]

Z-DNA binding domain, and though the precise function of this domain remains unclear, some studies have pointed towards a role in the formation and processing of cytoplasmic stress granules.[40-42] ADAR1 and 2 can generally be found across all cell and tissue types, but ADAR3 is distinct in that is expressed exclusively in the brain, and uniquely contains an arginine rich motif for binding ssRNAs.[43] ADAR3 also mysteriously displays little enzymatic editing activity despite containing a functional deaminase domain. While its definitive roles remain elusive, some investigations have suggested a downregulating function for ADAR3 based upon its observed competition with ADARs1 and 2 for RNA substrates, decreasing overall editing frequencies.[44] Homodimerization of ADARs also appears to be a requirement for efficient deamination. Although this activity is mediated by dsRNA binding domains, it also occurs independently of actual dsRNA binding.[45-46] ADAR3 does not display this dimerization ability, possibly due to a lack of one or more of these domains, which may also explain its non-editing behavior. Together, it is increasingly apparent that ADAR regulation is complex, and the molecular principles by which ADARs recognize and target specific transcripts and particular bases are still not fully understood.

1.3 Biological Functions and Disease Mechanisms

A significant number of A-to-I sites have been catalogued in the human transcriptome (> 5 million),[22, 47] and several apparent functional roles have emerged. Because inosine base pairs with cytosine and thus results in a functional A-to-G transition (**Figure 1.1b**), these alterations can directly modify codon sequences within mRNAs and subsequently impart amino acid substitutions in proteins (**Figure 1.5a**). These recoding events generate critical complexity and diversity in the cellular proteome, and serve several key functions. mRNA editing is especially prevalent in the brain,[48-51] where different ion channel and neurotransmitter receptor mRNAs undergo precise A-to-I modification to create structural changes in the resulting proteins and adjust ion permeability and membrane potential.[52-53] In particular, mRNA encoding the ionotropic glutamate receptor *GRIA2* undergoes precise A-to-I editing to induce a Gln-Arg recoding event, which importantly

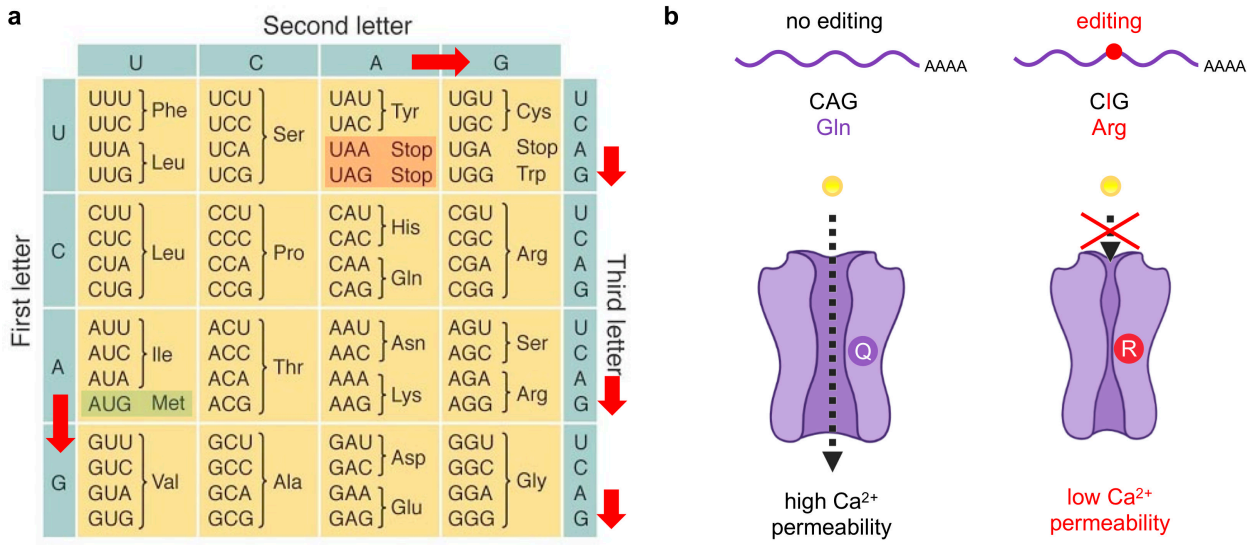


Figure 1.5 A-to-I editing in mRNA recodes amino acids. a) Universal genetic code table depicting codon sequences for each amino acid. Start (AUG) and stop codons (UAA and UAG) are highlighted, and red arrows depict possible transitions imparted by A-to-I editing. Table adapted from Griffiths et al. 2004. b) RNA editing of the GRIA2 ionotropic Ca^{2+} glutamate receptor channel alters physiological function. A-to-I editing of a glutamine (Gln, Q) codon recodes this to arginine (Arg, R), significantly perturbing ion permeability of the channel.

alters channel permeability to Ca^{2+} (**Figure 1.5b**).[52, 54] Several classes of both neurotransmitter receptors[53, 55-57] and ion channels[58] require this type of mRNA recoding for proper neuronal activity, and dysregulation of this process is directly responsible for brain development disorders,[51, 59] glioblastoma,[44] and neurodegeneration.[60-61] Further, alterations in neurotransmitter editing has also been linked to depression, schizophrenia, metabolic dysregulation, and obesity.[62-64] Numerous key proteins involved in cytoskeleton organization,[65] intracellular transport,[66] and metabolic signaling[67] also undergo A-to-I recoding, and many cancer types have been directly linked to aberrations in these events.[68-74] Abnormal mRNA editing is also generally prevalent in cancer throughout the body, and off-target A-to-I transitions in many oncogene and tumor suppressor transcripts provide a direct molecular link between RNA editing and cancer pathology.[72-73, 75] Interestingly, overall editing activity also varies between different tissue and cell types, underscoring the vital role of A-to-I editing in creating intercellular diversity and specialized tissue function.[22, 28] Moreover, recoding patterns also change significantly during tissue development and are often required for stem cell differentiation.[22, 76-80] Small regulatory RNAs are

also extensively edited, in turn modulating the specificities and biogenesis of small-interfering RNAs (siRNAs) and microRNAs (miRNAs) and affecting global gene expression patterns.[24]

Despite the significant functional importance of these sites, the vast majority of editing actually occurs in repetitive, non-coding RNA sequences.[22, 47] In particular, these sites are found in *Alu* elements embedded throughout the transcriptome and are common in non-coding regions of mRNA.[27, 81] Although unclear, *Alu* elements are thought to have originated from spontaneous duplication of the signal recognition particle RNA gene (7SL) sometime during rodent and primate evolutionary divergence, and now represent a major type of transposable element in humans and other primates. [82] Additionally, because of their high frequency (>10.7% of the human genome),[83] these are also classified as short interspersed nuclear elements (SINEs) in DNA. RNA transcribed from these *Alu* elements are especially important in the context of ADAR-mediated A-to-I editing, in that they are comprised of long (~300 bp) inverted repeats that adopt extensive dsRNA structures, which are efficiently recognized and edited by ADAR1 (**Figure 1.6**).[27, 81] *Alu* editing locations account for millions of A-to-I sites in humans,[22, 47] and may explain why editing activity is generally much higher in humans and primates compared to other mammals.[84]

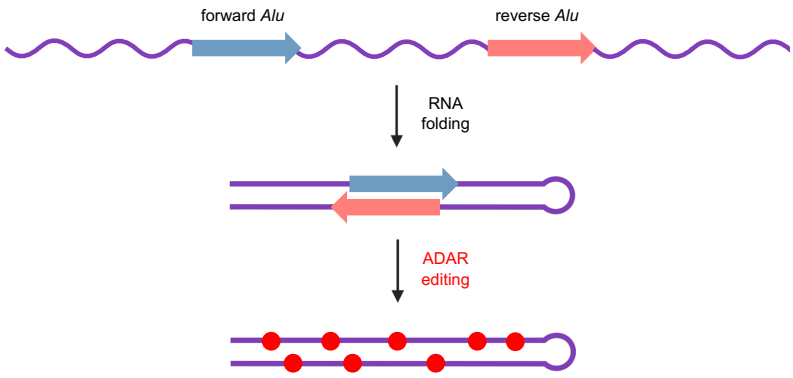


Figure 1.6 A-to-I editing of *Alu* element repeats. *Alu* elements are typically comprised of "forward" and "reverse" segments, which form inverted repeats in RNA transcripts and associate into long dsRNA structures. These are then recognized and edited extensively by ADAR enzymes.

Interestingly, widespread *Alu* RNA editing also appears to be a play a critical cellular function in humans. Unmodified dsRNA within the cytosol is typically indicative of viral infection and is quickly recognized by a suite of sensor proteins that activate the innate immune system.[85] *Alu* editing sites embedded throughout the transcriptome are now thought to help mark host cellular RNA as "self" and regulate these inflammatory pathways. [21, 23, 86-88] Unsurprisingly, malfunctions in this type of A-to-I editing are directly linked to several types of disease. ADAR1 editing of these *Alu* elements appears to be a critical molecular event during embryonic development, and genetic deletion of ADAR1 in mice is typically a lethal phenotype before birth. [80, 89-91] This is thought to occur because of uncontrolled innate immune activation in certain developing tissues, which in turn activates cytotoxic interferon responses and results in large-scale tissue death and deformation. [86, 91-92] Similar dysregulation of this process in humans is also implicated in autoimmune disease, including both systemic lupus and Aicardi-Goutieres syndrome.[93-95] Several types of cancer also exhibit ADAR1 overexpression and hyperactive *Alu* editing, which is theorized as a potential mechanism of evading normal immune system checkpoints and surveillance.[72, 96]

Despite its importance, our overall understanding of A-to-I editing regulation, function, and dynamics is significantly limited. In particular, while many sites have been identified (> 5 million),[97-98] the true prevalence and landscape of A-to-I editing in the human transcriptome is largely unknown, and it is unclear why certain sites are edited over others and what precise biological function they each serve. These knowledge gaps are especially problematic for therapeutic site-directed RNA editing strategies, which rely on the substrate preferences and inherent efficiencies of the natural ADAR machinery.[99] While global dysregulation of editing is also linked with numerous diseases, the exact molecular and cellular mechanisms responsible for this relationship are poorly understood. Similarly, because we lack tools to rapidly measure differences in ADAR activity between different cell and tissue types, we have little understanding of how this heterogeneity might contribute to disease. Gaining a deeper understanding of these and other

RNA modification patterns would provide invaluable information regarding disease progression. Moreover, addressing these knowledge gaps would lead to the identification of novel druggable protein and RNA targets and advance the development of new diagnostic tools.

1.4 Detecting A-to-I RNA Editing

Because of its importance, there is great interest in developing technologies to robustly map A-to-I sites and determine their biological significance. Early efforts were devoted to isolating and characterizing different ADAR isoforms, and yielded strong insight into editing mechanisms and substrate specificities.[29-30, 100] More recently, a powerful workflow developed to map ADAR binding sites in RNA with both high resolution and throughput.[101] This method, termed irCLASH (infrared crosslinking, ligation, and sequencing of hybrids), identifies ADAR binding sites by crosslinking the enzymes to RNA and then isolating and sequencing these regions. Interestingly, this revealed that ADARs bind ~50 bp footprints in tandem pairs on long RNA duplexes, corroborating earlier notions that ADAR dimerization is critical for activity.[45, 102] Because irCLASH can isolate specific ADAR isoforms, this method also elucidated key binding differences for both ADAR1 and ADAR2, revealing unique signatures for both protein-recoding (predominantly ADAR2) and Alu-type (ADAR1) editing sites. This study was also notable for mapping ADAR3 binding sites, which have been notoriously difficult to characterize because the enzyme does not introduce detectable editing sites in RNA. Interestingly, irCLASH datasets demonstrated that ADAR3 binding sites significantly overlapped with ADAR2 regions, suggesting it may compete with ADAR2 for these editing substrates to negatively regulate overall editing levels.[101] Although biochemical characterization of ADAR enzymes has provided some insight into editing mechanisms,[29-30, 100] these approaches indirectly infer editing site locations, and it still remains unclear why some A's are edited over others or how editing rates at individual sites are regulated. Moreover, several fundamental questions remain as to how editing is differentially regulated between different cell types and at varying stages of development.

Accurately identifying A-to-I RNA editing sites in the transcriptome is key for elucidating these roles, and robust correlation between editing activity and disease pathology will enable full leveraging of this molecular signal as a diagnostic biomarker. One of the first analytical methods developed for inosine detection relied on RNase T2 degradation of RNA followed by phosphorylation of each ribonucleotide with P^{32}. Using thin-layer chromatography (TLC), labeled ribonucleotides could then be resolved and imaged using autoradiography, and when compared to an internal inosine standard, A-to-I content in different samples could be estimated. This study was significant in that it provided the first estimate of tissue-level A-to-I editing activity in rat mRNA, and this technique has also been used to estimate ADAR editing kinetics in chemically synthesized RNA substrates.[54, 100] However, this method is also technically challenging, suffers from low precision and utilizes hazardous material, and has thus not been widely adopted outside of these experiments.

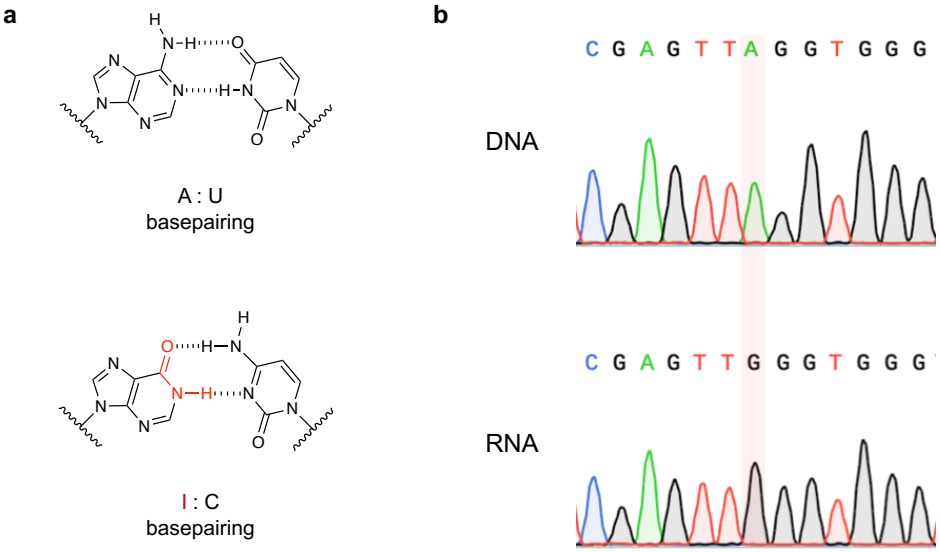


Figure 1.7 A-to-I editing sites can be identified with Sanger sequencing. a) Structural alterations introduced by A-to-I editing, with inosine being functionally recoded as guanosine (G). b) This base pairing change can be identified in Sanger sequencing traces. Sample traces from genomic DNA and reverse-transcribed RNA are aligned to illustrate this change. An A-G transition is highlighted, indicating the presence of an editing site with 100% conversion.

Because inosine is decoded as guanine by polymerases, these sites can also be reliably detected by identifying A-G conversions between DNA and RNA samples. In particular, target RNA can be reverse transcribed and analyzed using Sanger sequencing, and A-to-I sites can be reliably inferred by comparing matched genomic DNA to reverse-transcribed RNA samples (**Figure 1.7**). While this is effective for sites that are robustly converted (100% editing frequency is shown in **Figure 1.7b**), most A-to-I sites undergo very low or variable editing frequencies,[22] and these sites can be highly challenging to detect using Sanger sequencing. In particular, editing rates below ~15-20% are nearly impossible to distinguish from typical Sanger baselines. Additionally, Sanger is a low-throughput method and further requires prior knowledge of candidate sites to design specific primers, and is thus significantly limited in its ability for large-scale mapping of A-to-I editing sites in the human transcriptome.

To circumvent these limitations, newer methods have incorporated chemical labeling strategies to identify RNA modifications. In particular, acrylonitrile labeling has been used to modify and detect inosines,[49, 103] which reacts through a type of Michael addition to form N^1-cyanoethylinosine (ce^1I) (**Figure 1.8**). While this chemistry has been known for some time, it was only recently discovered that ce^1I also arrests reverse transcription. This finding was then leveraged into a new epitranscriptomic sequencing tool known as “inosine chemical erasing sequencing” (ICE-seq).[103] In this method, acrylonitrile-labeled RNAs produce truncated cDNA transcripts, eliminating their amplification by PCR for sequencing (**Figure 1.8b**). Thus, when comparing the sequencing data of treated and untreated RNA, regions that contained inosines are “erased” from sequencing chromatograms (**Figure 1.8b**). This can also be coupled with high-throughput analysis algorithms for identifying new sites, and represented a significant improvement in detection workflows for A-to-I site discovery. However, this method is also still limited in sensitivity as it still requires identification of A-G transitions in “untreated” samples, and hence suffers from the same limitations as other Sanger-based methods.

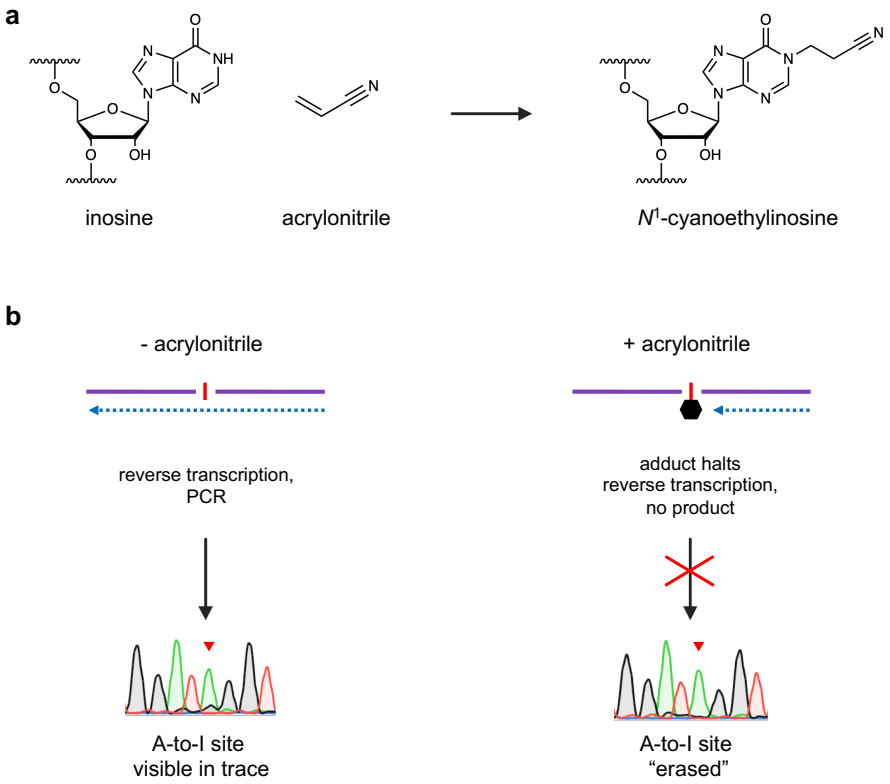


Figure 1.8 Acrylonitrile reacts with inosine and chemically alters Sanger sequencing results. a) Inosine undergoes Michael addition with acrylonitrile to form N^1-cyanoethylinosine. B) Because these adducts are on the Watson-Crick-Franklin face of the nucleobase, they arrest reverse-transcription and "erase" A-G transitions.

The development of high-throughput RNA sequencing (RNA-seq) has improved this sensitivity problem and enables detection of editing events in single RNA molecules, and is now the most widely adopted method to identify editing sites.[104] While this still relies on the sequence change introduced by A-to-I editing and enables high-throughput identification of inosine in transcripts, editing sites are still quite rare in the context of total RNA (~0.1 – <0.01% of all nucleotides).[22, 105] Additionally, key edited transcripts can be low in copy number and actual editing rates at individual sites are often highly variable.[22] RNA-seq is also limited to capturing small "windows" of RNA, making this technique inconsistent and prone to random sampling errors. Because of these challenges, RNA-seq requires excessive amounts of RNA input material, specialized instrumentation, and very high numbers of sequencing reads to achieve sufficient depth and coverage for accurate A-to-I calling.[106-107] Overexpression of ADARs is also linked with

many types of cancer,[69, 71-73] and both under-[59, 93, 108] and over-editing[73, 94, 109] phenotypes have been observed in other disorders. While global editing signatures could serve as a vital diagnostic biomarker in these examples, current methods to profile inosine abundance rely on the same RNA-seq technology described above and thus suffer from the same limitations that necessitate large investments in material and time, and require lengthy computational pipelines to generate a single metric.[28] At present, there is significant demand for technologies that both sensitively identify and map individual editing sites as well as rapidly profile the overall abundance of inosine in cellular RNA samples.

1.5 Summary and Conclusions of this Dissertation

Despite the critical importance of this process in cellular function, our overall understanding of A-to-I editing is significantly limited, and this incomplete knowledge is the direct result of technical challenges associated with current methods for detecting, studying, and controlling this modification. Given these limitations, the primary goal of my research has been to develop improved tools to better probe A-to-I RNA editing and characterize its biological functions. In this thesis, we describe in detail our investigations of chemical and biological strategies for engineering these new methods.

Building off observations that inosine can be directly labeled with Michael acceptors,[49, 103] we first construct acrylamide derivatives that can selectively react with inosine in RNA to yield a detectable signal. As an initial demonstration, we first synthesized acrylamidofluorescein to enable fluorescent detection of inosine in RNA transcripts (Chapter 2). We then expanded this chemistry into a generalizable phenylacrylamide scaffold that allows “clickable” installation of any fluorophore or small molecule payload to A-to-I sites, enabling rapid characterization of editing activity in small RNA substrates (Chapter 3). While these reagents were feasible for labeling small RNA substrates, acrylamide and acrylonitrile derivatives also display off-target labeling with other nucleobases (particularly those containing nucleophilic nitrogens) and were hence less-suited for

use with complex cellular RNA samples. Looking to Nature for a solution to this molecular recognition challenge, we then describe how we harnessed the exquisite ability of Endonuclease V (EndoV) to recognize and bind inosine-containing RNAs. In particular, we first developed EndoVIPER (Endonuclease V immunoprecipitation enrichment) as a method to selectively enrich A-to-I edited transcripts prior to RNA-seq, roughly doubling the amount of newly identified editing sites compared to standard workflow (Chapter 4). Building off of these results, we then leverage EndoV's highly specific binding affinity for inosine in RNA to construct a chemiluminescent immunoassay for direct detection of global A-to-I editing signatures in cellular RNA (Chapter 5). We find that this method is comparable to existing RNA-seq computation workflows for detecting inosine content with a dramatic reduction in both lead-time and cost. Lastly, we utilize the covalent denaturant glyoxal, which was a key step in our EndoV-based workflows, as a synthetic nucleic acid modification. Glyoxalation of different substrates affords thermoreversible control over virtually any nucleic acid, positioning this a vital platform for constructing stimuli-responsive materials medicine, biocomputing, and nanotechnology. Additionally, our demonstration expands our notions of how nucleic acid modifications may be leveraged in different synthetic applications. I then finally summarize current perspectives gained from these developments, and discuss future initiatives for optimizing and deploying these new technologies in a variety of biological contexts (Chapter 7).

1.6 References:

1. Licatalosi, D. D.; Darnell, R. B., RNA processing and its regulation: global insights into biological networks. *Nature Reviews Genetics* **2010,** *11* (1), 75-87.

2. Shafik, A.; Schumann, U.; Evers, M.; Sibbritt, T.; Preiss, T., The emerging epitranscriptomics of long noncoding RNAs. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **2016,** *1859* (1), 59-70.

3. Peer, E.; Rechavi, G.; Dominissini, D., Epitranscriptomics: regulation of mRNA metabolism through modifications. *Current opinion in chemical biology* **2017,** *41*, 93-98.

4. Grozhik, A. V.; Jaffrey, S. R., Epitranscriptomics: Shrinking maps of RNA modifications. *Nature* **2017,** *551* (7679), 174-177.

5. Li, X.; Xiong, X.; Yi, C., Epitranscriptome sequencing technologies: decoding RNA modifications. *Nature methods* **2017,** *14* (1), 23-31.

6. Shi, H.; Wei, J.; He, C., Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Molecular cell* **2019,** *74* (4), 640-650.

7. Li, K.; Cai, J.; Zhang, M.; Zhang, X.; Xiong, X.; Meng, H.; Xu, X.; Huang, Z.; Peng, J.; Fan, J., Landscape and regulation of m6A and m6Am methylome across human and mouse tissues. *Molecular cell* **2020,** *77* (2), 426-440. e6.

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

9. Meyer, K. D.; Saletore, Y.; Zumbo, P.; Elemento, O.; Mason, C. E.; Jaffrey, S. R., Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **2012,** *149* (7), 1635-1646.

10. Mao, Y.; Dong, L.; Liu, X.-M.; Guo, J.; Ma, H.; Shen, B.; Qian, S.-B., m 6 A in mRNA coding regions promotes translation via the RNA helicase-containing YTHDC2. *Nature communications* **2019,** *10* (1), 1-11.

11. Arguello, A. E.; DeLiberto, A. N.; Kleiner, R. E., RNA chemical proteomics reveals the N6-methyladenosine (m6A)-regulated protein–RNA interactome. *Journal of the American Chemical Society* **2017,** *139* (48), 17249-17252.

12. Fu, Y.; Zhuang, X., m6A-binding YTHDF proteins promote stress granule formation by modulating phase separation of stress granule proteins. *bioRxiv* **2019**, 694455.

13. Li, X.; Ma, S.; Yi, C., Pseudouridine: the fifth RNA nucleotide with renewed interests. *Current Opinion in Chemical Biology* **2016,** *33*, 108-116.

14. Knutson, S. D.; Heemstra, J. M., Protein-based molecular recognition tools for detecting and profiling RNA modifications. *Current Opinion in Structural Biology* **2021,** *69*, 1-10.

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

16. Carlile, T. M.; Martinez, N. M.; Schaening, C.; Su, A.; Bell, T. A.; Zinshteyn, B.; Gilbert, W. V., mRNA structure determines modification by pseudouridine synthase 1. *Nature chemical biology* **2019,** *15* (10), 966-974.

17. Jin, G.; Xu, M.; Zou, M.; Duan, S., The Processing, Gene Regulation, Biological Functions, and Clinical Relevance of N4-Acetylcytidine on RNA: A Systematic Review. *Molecular Therapy-Nucleic Acids* **2020,** *20*, 13-24.

18. Zhao, W.; Zhou, Y.; Cui, Q.; Zhou, Y., PACES: prediction of N4-acetylcytidine (ac4C) modification sites in mRNA. *Scientific reports* **2019,** *9* (1), 1-7.

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

20. Squires, J. E.; Preiss, T., Function and detection of 5-methylcytosine in eukaryotic RNA. *Epigenomics* **2010,** *2* (5), 709-715.

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

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

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

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

25. Gerber, A. P.; Keller, W., An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **1999,** *286* (5442), 1146-1149.

26. Rubio, M. A. T.; Pastar, I.; Gaston, K. W.; Ragone, F. L.; Janzen, C. J.; Cross, G. A.; Papavasiliou, F. N.; Alfonzo, J. D., An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. *Proceedings of the National Academy of Sciences* **2007,** *104* (19), 7821-7826.

27. Kim, D. D.; Kim, T. T.; Walsh, T.; Kobayashi, Y.; Matise, T. C.; Buyske, S.; Gabriel, A., Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome research* **2004,** *14* (9), 1719-1725.

28. Roth, S. H.; Levanon, E. Y.; Eisenberg, E., Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nature methods* **2019**, 1-8.

29. Wong, S. K.; Sato, S.; Lazinski, D. W., Substrate recognition by ADAR1 and ADAR2. *Rna* **2001,** *7* (6), 846-858.

30. Kuttan, A.; Bass, B. L., Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci U S A* **2012,** *109* (48), E3295-304.

31. Nishikura, K.; Yoo, C.; Kim, U.; Murray, J.; Estes, P.; Cash, F.; Liebhaber, S., Substrate specificity of the dsRNA unwinding/modifying activity. *The EMBO journal* **1991,** *10* (11), 3523.

32. Goodman, R. A.; Macbeth, M. R.; Beal, P. A., ADAR proteins: structure and catalytic mechanism. *Curr Top Microbiol Immunol* **2012,** *353*, 1-33.

33. Macbeth, M. R.; Schubert, H. L.; VanDemark, A. P.; Lingam, A. T.; Hill, C. P.; Bass, B. L., Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **2005,** *309* (5740), 1534-1539.

34. Nishikura, K., Functions and regulation of RNA editing by ADAR deaminases. *Annual review of biochemistry* **2010,** *79*, 321-349.

35. George, C. X.; Samuel, C. E., Human RNA-specific adenosine deaminase ADAR1 transcripts possess alternative exon 1 structures that initiate from different promoters, one constitutively active and the other interferon inducible. *Proceedings of the National Academy of Sciences* **1999,** *96* (8), 4621-4626.

36. Patterson, J. B.; Samuel, C. E., Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Molecular and cellular biology* **1995,** *15* (10), 5376-5388.

37. Hartwig, D.; Schütte, C.; Warnecke, J.; Dorn, I.; Hennig, H.; Kirchner, H.; Schlenke, P., The large form of ADAR 1 is responsible for enhanced hepatitis delta virus RNA editing in interferon-α-stimulated host cells. *Journal of viral hepatitis* **2006,** *13* (3), 150-157.

38. Strehblow, A.; Hallegger, M.; Jantsch, M. F., Nucleocytoplasmic distribution of human RNA-editing enzyme ADAR1 is modulated by double-stranded RNA-binding domains, a leucine-rich export signal, and a putative dimerization domain. *Molecular biology of the cell* **2002,** *13* (11), 3822-3835.

39. Nishikura, K., A-to-I editing of coding and non-coding RNAs by ADARs. *Nature reviews Molecular cell biology* **2016,** *17* (2), 83.

40. Herbert, A.; Alfken, J.; Kim, Y.-G.; Mian, I. S.; Nishikura, K.; Rich, A., A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proceedings of the National Academy of Sciences* **1997,** *94* (16), 8421-8426.

41. Herbert, A.; Rich, A., The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1. *Proceedings of the National Academy of Sciences* **2001,** *98* (21), 12132-12137.

42. Ng, S. K.; Weissbach, R.; Ronson, G. E.; Scadden, A., Proteins that contain a functional Z-DNA-binding domain localize to cytoplasmic stress granules. *Nucleic acids research* **2013,** *41* (21), 9786-9799.

43. Chen, C.-X.; CHO, D.-S. C.; WANG, Q.; LAI, F.; CARTER, K. C.; NISHIKURA, K., A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA* **2000,** *6* (5), 755-767.

44. Oakes, E.; Anderson, A.; Cohen-Gadol, A.; Hundley, H. A., Adenosine Deaminase That Acts on RNA 3 (ADAR3) Binding to Glutamate Receptor Subunit B Pre-mRNA Inhibits RNA Editing in Glioblastoma. *Journal of Biological Chemistry* **2017,** *292* (10), 4326-4335.

45. Cho, D.-S. C.; Yang, W.; Lee, J. T.; Shiekhattar, R.; Murray, J. M.; Nishikura, K., Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *Journal of Biological Chemistry* **2003,** *278* (19), 17093-17102.

46. Valente, L.; Nishikura, K., RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. *Journal of Biological Chemistry* **2007,** *282* (22), 16054-16061.

47. Mansi, L.; Tangaro, M. A.; Lo Giudice, C.; Flati, T.; Kopel, E.; Schaffer, A. A.; Castrignanò, T.; Chillemi, G.; Pesole, G.; Picardi, E., REDIportal: millions of novel A-to-I RNA editing events from thousands of RNAseq experiments. *Nucleic Acids Research* **2020,** *in press*.

48. Li, J. B.; Church, G. M., Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. *Nature neuroscience* **2013,** *16* (11), 1518.

49. Sakurai, M.; Ueda, H.; Yano, T.; Okada, S.; Terajima, H.; Mitsuyama, T.; Toyoda, A.; Fujiyama, A.; Kawabata, H.; Suzuki, T., A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome research* **2014,** *24* (3), 522-534.

50. Behm, M.; Öhman, M., RNA editing: a contributor to neuronal dynamics in the mammalian brain. *Trends in Genetics* **2016,** *32* (3), 165-175.

51. Hwang, T.; Park, C.-K.; Leung, A. K.; Gao, Y.; Hyde, T. M.; Kleinman, J. E.; Rajpurohit, A.; Tao, R.; Shin, J. H.; Weinberger, D. R., Dynamic regulation of RNA editing in human brain development and disease. *Nature neuroscience* **2016,** *19* (8), 1093.

52. Wright, A. L.; Vissel, B., The essential role of AMPA receptor GluR2 subunit RNA editing in the normal and diseased brain. *Frontiers in molecular neuroscience* **2012,** *5*, 34.

53. Sommer, B.; Köhler, M.; Sprengel, R.; Seeburg, P. H., RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **1991,** *67* (1), 11-19.

54. Stephens, O. M.; Yi-Brunozzi, H. Y.; Beal, P. A., Analysis of the RNA-editing reaction of ADAR2 with structural and fluorescent analogues of the GluR-B R/G editing site. *Biochemistry* **2000,** *39* (40), 12243-12251.

55. Lomeli, H.; Mosbacher, J.; Melcher, T.; Hoger, T.; Kuner, T.; Monyer, H.; Higuchi, M.; Bach, A.; Seeburg, P. H., Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* **1994,** *266* (5191), 1709-1713.

56. Burns, C. M.; Chu, H.; Rueter, S. M.; Hutchinson, L. K.; Canton, H.; Sanders-Bush, E.; Emeson, R. B., Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **1997,** *387* (6630), 303-308.

57. Daniel, C.; Wahlstedt, H.; Ohlson, J.; Björk, P.; Öhman, M., Adenosine-to-inosine RNA editing affects trafficking of the γ-aminobutyric acid type A (GABAA) receptor. *Journal of Biological Chemistry* **2011,** *286* (3), 2031-2040.

58. Bhalla, T.; Rosenthal, J. J.; Holmgren, M.; Reenan, R., Control of human potassium channel inactivation by editing of a small mRNA hairpin. *Nature structural & molecular biology* **2004,** *11* (10), 950-956.

59. Tran, S. S.; Jun, H.-I.; Bahn, J. H.; Azghadi, A.; Ramaswami, G.; Van Nostrand, E. L.; Nguyen, T. B.; Hsiao, Y.-H. E.; Lee, C.; Pratt, G. A., Widespread RNA editing dysregulation in brains from autistic individuals. *Nature neuroscience* **2019,** *22* (1), 25.

60. Khermesh, K.; D'Erchia, A. M.; Barak, M.; Annese, A.; Wachtel, C.; Levanon, E. Y.; Picardi, E.; Eisenberg, E., Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *Rna* **2016,** *22* (2), 290-302.

61. Kawahara, Y.; Ito, K.; Sun, H.; Aizawa, H.; Kanazawa, I.; Kwak, S., Glutamate receptors: RNA editing and death of motor neurons. *Nature* **2004,** *427* (6977), 801-802.

62. Akbarian, S.; Smith, M. A.; Jones, E. G., Editing for an AMPA receptor subunit RNA in prefrontal cortex and striatum in Alzheimer's disease, Huntington's disease and schizophrenia. *Brain research* **1995,** *699* (2), 297-304.

63. Singh, M.; Kesterson, R. A.; Jacobs, M. M.; Joers, J. M.; Gore, J. C.; Emeson, R. B., Hyperphagia-mediated obesity in transgenic mice misexpressing the RNA-editing enzyme ADAR2. *Journal of Biological Chemistry* **2007,** *282* (31), 22448-22459.

64. Gurevich, I.; Tamir, H.; Arango, V.; Dwork, A. J.; Mann, J. J.; Schmauss, C., Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims. *Neuron* **2002,** *34* (3), 349-356.

65. Jain, M.; Mann, T. D.; Stulić, M.; Rao, S. P.; Kirsch, A.; Pullirsch, D.; Strobl, X.; Rath, C.; Reissig, L.; Moreth, K., RNA editing of Filamin A pre-mRNA regulates vascular contraction and diastolic blood pressure. *The EMBO journal* **2018,** *37* (19).

66. Wang, X.; Wu, X.; Zhu, Z.; Li, H.; Li, T.; Li, Q.; Zhang, P.; Li, L.; Che, D.; Xiao, X., Landscape of RNA editing reveals new insights into the dynamic gene regulation of spermatogenesis. *Cell Cycle* **2019,** *18* (23), 3351-3364.

67. Chen, L.; Li, Y.; Lin, C. H.; Chan, T. H. M.; Chow, R. K. K.; Song, Y.; Liu, M.; Yuan, Y.-F.; Fu, L.; Kong, K. L., Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nature medicine* **2013,** *19* (2), 209-216.

68. Peng, X.; Xu, X.; Wang, Y.; Hawke, D. H.; Yu, S.; Han, L.; Zhou, Z.; Mojumdar, K.; Jeong, K. J.; Labrie, M., A-to-I RNA editing contributes to proteomic diversity in cancer. *Cancer cell* **2018,** *33* (5), 817-828. e7.

69. Paz, N.; Levanon, E. Y.; Amariglio, N.; Heimberger, A. B.; Ram, Z.; Constantini, S.; Barbash, Z. S.; Adamsky, K.; Safran, M.; Hirschberg, A.; Krupsky, M.; Ben-Dov, I.; Cazacu, S.; Mikkelsen, T.; Brodie, C.; Eisenberg, E.; Rechavi, G., Altered adenosine-to-inosine RNA editing in human cancer. *Genome Res* **2007,** *17* (11), 1586-95.

70. Martinez, H. D.; Jasavala, R. J.; Hinkson, I.; Fitzgerald, L. D.; Trimmer, J. S.; Kung, H.-J.; Wright, M. E., RNA editing of androgen receptor gene transcripts in prostate cancer cells. *Journal of Biological Chemistry* **2008,** *283* (44), 29938-29949.

71. Chan, T. H.; Lin, C. H.; Qi, L.; Fei, J.; Li, Y.; Yong, K. J.; Liu, M.; Song, Y.; Chow, R. K.; Ng, V. H.; Yuan, Y. F.; Tenen, D. G.; Guan, X. Y.; Chen, L., A disrupted RNA editing balance mediated by ADARs (Adenosine DeAminases that act on RNA) in human hepatocellular carcinoma. *Gut* **2014,** *63* (5), 832-43.

72. Han, L.; Diao, L.; Yu, S.; Xu, X.; Li, J.; Zhang, R.; Yang, Y.; Werner, H. M. J.; Eterovic, A. K.; Yuan, Y.; Li, J.; Nair, N.; Minelli, R.; Tsang, Y. H.; Cheung, L. W. T.; Jeong, K. J.; Roszik, J.; Ju, Z.; Woodman, S. E.; Lu, Y.; Scott, K. L.; Li, J. B.; Mills, G. B.; Liang, H., The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer Cell* **2015,** *28* (4), 515-528.

73. Paz-Yaacov, N.; Bazak, L.; Buchumenski, I.; Porath, H. T.; Danan-Gotthold, M.; Knisbacher, B. A.; Eisenberg, E.; Levanon, E. Y., Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors. *Cell Rep* **2015,** *13* (2), 267-76.

74. Hu, X.; Wan, S.; Ou, Y.; Zhou, B.; Zhu, J.; Yi, X.; Guan, Y.; Jia, W.; Liu, X.; Wang, Q.; Qi, Y.; Yuan, Q.; Huang, W.; Liao, W.; Wang, Y.; Zhang, Q.; Xiao, H.; Chen, X.; Huang, J., RNA over-editing of BLCAP contributes to hepatocarcinogenesis identified by whole-genome and transcriptome sequencing. *Cancer Lett* **2015,** *357* (2), 510-9.

75. Jiang, Q.; Crews, L. A.; Holm, F.; Jamieson, C. H. M., RNA editing-dependent epitranscriptome diversity in cancer stem cells. *Nat Rev Cancer* **2017,** *17* (6), 381-392.

76. Wahlstedt, H.; Daniel, C.; Ensterö, M.; Öhman, M., Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. *Genome research* **2009,** *19* (6), 978-986.

77. Shtrichman, R.; Germanguz, I.; Mandel, R.; Ziskind, A.; Nahor, I.; Safran, M.; Osenberg, S.; Sherf, O.; Rechavi, G.; Itskovitz-Eldor, J., Altered A-to-I RNA editing in human embryogenesis. *PLoS One* **2012,** *7* (7).

78. Goldstein, B.; Agranat-Tamir, L.; Light, D.; Zgayer, O. B.-N.; Fishman, A.; Lamm, A. T., A-to-I RNA editing promotes developmental stage–specific gene and lncRNA expression. *Genome research* **2017,** *27* (3), 462-470.

79. Wang, Q.; Khillan, J.; Gadue, P.; Nishikura, K., Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. *Science* **2000,** *290* (5497), 1765-1768.

80. Liddicoat, B. J.; Hartner, J. C.; Piskol, R.; Ramaswami, G.; Chalk, A. M.; Kingsley, P. D.; Sankaran, V. G.; Wall, M.; Purton, L. E.; Seeburg, P. H., Adenosine-to-inosine RNA editing by ADAR1 is essential for normal murine erythropoiesis. *Experimental hematology* **2016,** *44* (10), 947-963.

81. Bazak, L.; Levanon, E. Y.; Eisenberg, E., Genome-wide analysis of Alu editability. *Nucleic acids research* **2014,** *42* (11), 6876-6884.

82. Arcot, S. S.; Wang, Z.; Weber, J. L.; Deininger, P. L.; Batzer, M. A., Alu repeats: a source for the genesis of primate microsatellites. *Genomics* **1995,** *29* (1), 136-144.

83. Roy-Engel, A. M.; Carroll, M. L.; Vogel, E.; Garber, R. K.; Nguyen, S. V.; Salem, A.-H.; Batzer, M. A.; Deininger, P. L., Alu insertion polymorphisms for the study of human genomic diversity. *Genetics* **2001,** *159* (1), 279-290.

84. Paz-Yaacov, N.; Levanon, E. Y.; Nevo, E.; Kinar, Y.; Harmelin, A.; Jacob-Hirsch, J.; Amariglio, N.; Eisenberg, E.; Rechavi, G., Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. *Proceedings of the National Academy of Sciences* **2010,** *107* (27), 12174-12179.

85. Ruan, J.; Cao, Y.; Ling, T.; Li, P.; Wu, S.; Peng, D.; Wang, Y.; Jia, X.; Chen, S.; Xu, A., DDX23, an evolutionary conserved dsRNA sensor, participates in innate antiviral responses by pairing with TRIF or MAVS. *Frontiers in immunology* **2019,** *10*, 2202.

86. Mannion, N. M.; Greenwood, S. M.; Young, R.; Cox, S.; Brindle, J.; Read, D.; Nellåker, C.; Vesely, C.; Ponting, C. P.; McLaughlin, P. J., The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell reports* **2014,** *9* (4), 1482-1494.

87. Ishizuka, J. J.; Manguso, R. T.; Cheruiyot, C. K.; Bi, K.; Panda, A.; Iracheta-Vellve, A.; Miller, B. C.; Du, P. P.; Yates, K. B.; Dubrot, J., Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. *Nature* **2018,** *565* (7737), 43-48.

88. Chung, H.; Calis, J. J.; Wu, X.; Sun, T.; Yu, Y.; Sarbanes, S. L.; Thi, V. L. D.; Shilvock, A. R.; Hoffmann, H.-H.; Rosenberg, B. R., Human ADAR1 prevents endogenous RNA from triggering translational shutdown. *Cell* **2018,** *172* (4), 811-824. e14.

89. Wang, Q.; Miyakoda, M.; Yang, W.; Khillan, J.; Stachura, D. L.; Weiss, M. J.; Nishikura, K., Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene. *Journal of Biological Chemistry* **2004,** *279* (6), 4952-4961.

90. Hartner, J. C.; Schmittwolf, C.; Kispert, A.; Müller, A. M.; Higuchi, M.; Seeburg, P. H., Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. *Journal of Biological Chemistry* **2004,** *279* (6), 4894-4902.

91. Liddicoat, B. J.; Piskol, R.; Chalk, A. M.; Ramaswami, G.; Higuchi, M.; Hartner, J. C.; Li, J. B.; Seeburg, P. H.; Walkley, C. R., RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **2015,** *349* (6252), 1115-1120.

92. Pestal, K.; Funk, C. C.; Snyder, J. M.; Price, N. D.; Treuting, P. M.; Stetson, D. B., Isoforms of RNA-editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. *Immunity* **2015,** *43* (5), 933-944.

93. Shallev, L.; Kopel, E.; Feiglin, A.; Leichner, G. S.; Avni, D.; Sidi, Y.; Eisenberg, E.; Barzilai, A.; Levanon, E. Y.; Greenberger, S., Decreased A-to-I RNA editing as a source of keratinocytes' dsRNA in psoriasis. *RNA* **2018,** *24* (6), 828-840.

94. Roth, S. H.; Danan-Gotthold, M.; Ben-Izhak, M.; Rechavi, G.; Cohen, C. J.; Louzoun, Y.; Levanon, E. Y., Increased RNA editing may provide a source for autoantigens in systemic lupus erythematosus. *Cell reports* **2018,** *23* (1), 50-57.

95. Rice, G. I.; Kasher, P. R.; Forte, G. M.; Mannion, N. M.; Greenwood, S. M.; Szynkiewicz, M.; Dickerson, J. E.; Bhaskar, S. S.; Zampini, M.; Briggs, T. A., Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. *Nature genetics* **2012,** *44* (11), 1243-1248.

96. Maas, S.; Kawahara, Y.; Tamburro, K. M.; Nishikura, K., A-to-I RNA editing and human disease. *RNA biology* **2006,** *3* (1), 1-9.

97. Ramaswami, G.; Li, J. B., RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic acids research* **2013,** *42* (D1), D109-D113.

98. Picardi, E.; D'Erchia, A. M.; Lo Giudice, C.; Pesole, G., REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic acids research* **2016,** *45* (D1), D750-D757.

99. Montiel-Gonzalez, M. F.; Quiroz, J. F. D.; Rosenthal, J. J., Current strategies for site-directed RNA editing using ADARs. *Methods* **2019,** *156*, 16-24.

100. Wang, Y.; Park, S.; Beal, P. A., Selective recognition of RNA substrates by ADAR deaminase domains. *Biochemistry* **2018,** *57* (10), 1640-1651.

101. Song, Y.; Yang, W.; Fu, Q.; Wu, L.; Zhao, X.; Zhang, Y.; Zhang, R., irCLASH reveals RNA substrates recognized by human ADARs. *Nature Structural & Molecular Biology* **2020,** *27* (4), 351-362.

102. Thuy-Boun, A. S.; Thomas, J. M.; Grajo, H. L.; Palumbo, C. M.; Park, S.; Nguyen, L. T.; Fisher, A. J.; Beal, P. A., Asymmetric dimerization of adenosine deaminase acting on RNA facilitates substrate recognition. *Nucleic acids research* **2020,** *48* (14), 7958-7972.

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

104. Oakes, E.; Vadlamani, P.; Hundley, H. A., Methods for the Detection of Adenosine-to-Inosine Editing Events in Cellular RNA. In *mRNA Processing*, Springer: 2017; pp 103-127.

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

106. Zhang, R.; Li, X.; Ramaswami, G.; Smith, K. S.; Turecki, G.; Montgomery, S. B.; Li, J. B., Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nature methods* **2014,** *11* (1), 51.

107. Giudice, C. L.; Silvestris, D. A.; Roth, S. H.; Eisenberg, E.; Pesole, G.; Gallo, A.; Picardi, E., Quantifying RNA Editing in Deep Transcriptome Datasets. *Frontiers in Genetics* **2020,** *11*.

108. Hideyama, T.; Yamashita, T.; Aizawa, H.; Tsuji, S.; Kakita, A.; Takahashi, H.; Kwak, S., Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. *Neurobiology of disease* **2012,** *45* (3), 1121-1128.

109. Vlachogiannis, N. I.; Gatsiou, A.; Silvestris, D. A.; Stamatelopoulos, K.; Tektonidou, M. G.; Gallo, A.; Sfikakis, P. P.; Stellos, K., Increased adenosine-to-inosine RNA editing in rheumatoid arthritis. *Journal of autoimmunity* **2020,** *106*, 102329.

Chapter 2

Chemical Labeling and Affinity Capture of Inosine-Containing RNAs Using Acrylamidofluorescein[1*]

2.1 Abstract:

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

---

[*]Adapted from Ref. 1 with permission from Knutson, S. D.; Ayele, T. M..; Heemstra, J. M. *Bioconjugate Chemistry,* **2018**, *29*(9), 2899-2903. Copyright 2018 American Chemical Society.

2.2 Introduction:

RNA is extensively edited after transcription. Adenosine-to-inosine (A-to-I) conversion is of one of the most common and impactful forms of editing and is catalyzed by adenosine deaminases acting on RNA (ADARs).[2] Resulting inosines base pair with cytidine and are effectively decoded as guanosine by cellular machinery. A-to-I editing occurs in both coding and non-coding RNA transcripts, eliciting dramatic changes in overall cellular function and behavior. Editing of mRNA can alter protein sequence through direct modification of codons or by altering splice sites and regulatory elements in untranslated regions. A-to-I editing events are also extensive in non-coding RNAs, including microRNA and small-interfering RNA precursors, significantly altering their biosynthesis, trafficking, specificity, and gene regulation properties.[3-5] Accurately identifying A-to-I RNA editing sites in the transcriptome is necessary to improve our under-standing of these modifications and their biological functions. A recently developed method to map A-to-I editing locations employs chemical modification of inosines with acrylonitrile to form N^1-cyanoethylinosine (**Figure 2.1a**).[6-7] Termed inosine chemical erasing sequencing (ICE-seq), this technique leverages the observation that inosine cyanoethylation inhibits Watson-Crick base pairing and effectively arrests reverse transcription at A-to-I editing sites. Resulting truncated cDNAs fail to undergo PCR amplification and are “erased” from RNA sequencing chromatograms, allowing bioinformatic detection of editing sites. Although ICE-seq has improved the accuracy and scalability of mapping and discovering A-to-I RNA editing sites, this method is also limited in sensitivity, as labeled inosine-containing transcripts cannot be enriched. Additionally, while millions of A-to-I sites have been identified across the human transcriptome, actual editing rates at these sites are highly variable and dependent on cellular and environmental cues, rendering them difficult to detect, characterize, and measure with these techniques. This is particularly true in coding RNAs, where I/A ratios can range anywhere from <0.001-5% depending on tissue type or external stimuli.[8-10] Together, these challenges mask the overall prevalence and true landscape of A-to-I RNA editing across the transcriptome.

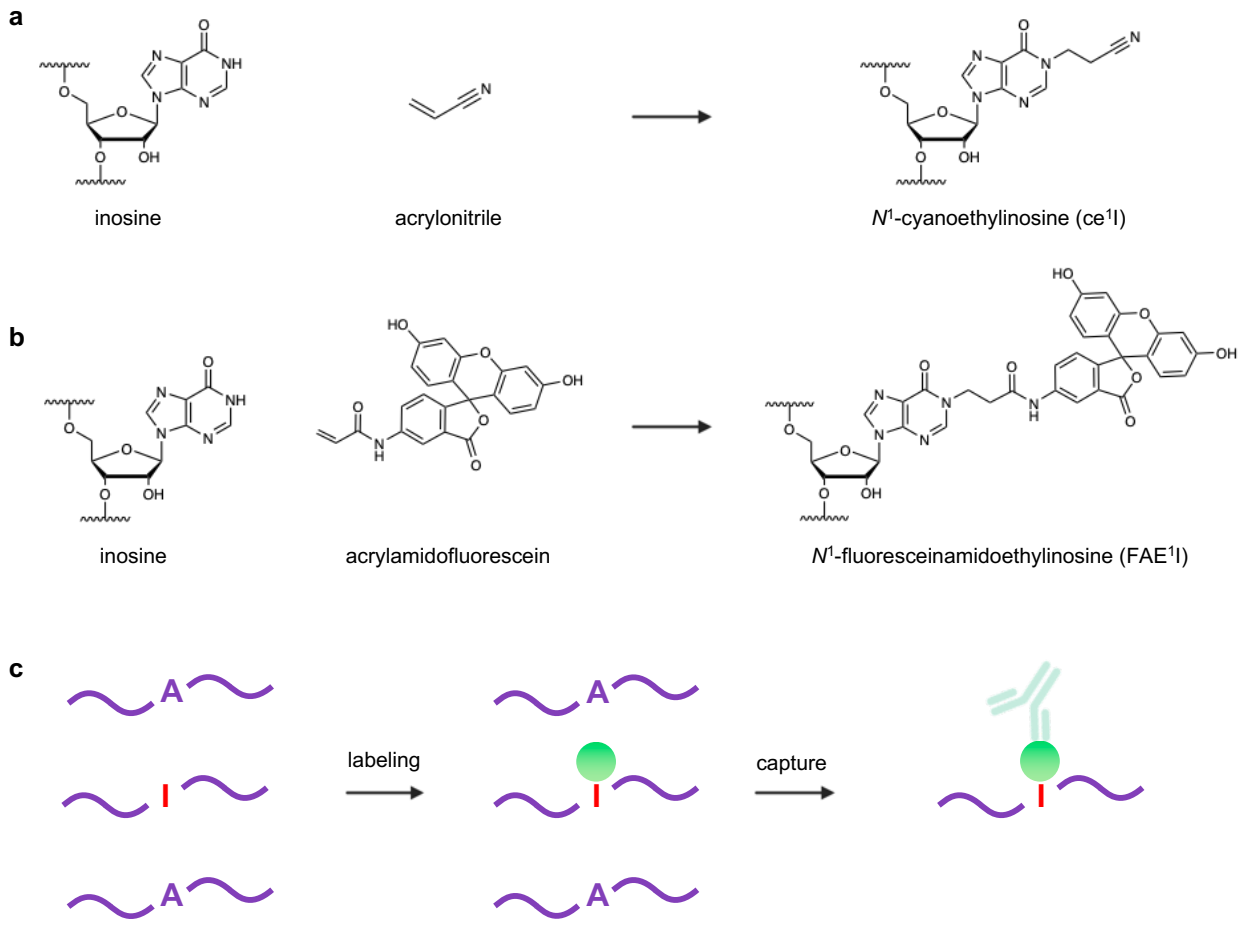


Figure 2.1 Chemical labeling of inosine. a) Acrylonitrile and b) acrylamidofluorescein produce N^1 addition products. c) Acrylamidofluorescein labeling enables affinity capture of transcripts containing inosine.

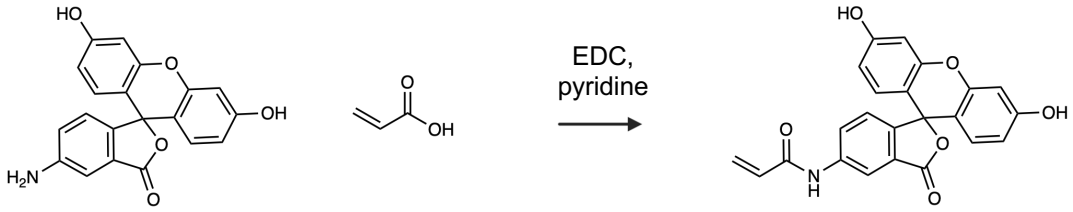
The ability to enrich A-to-I edited transcripts from more complex total RNA samples would largely address this limitation and allow for deeper interrogation and characterization of the epitranscriptome. Approaches using chemical labeling and/or antibody immunoprecipitation to capture edited transcripts have enabled significant advances in identifying and cataloging a number of other RNA modifications, including N^1- and N^6-methyladenosine, 5-methylcytidine, 5-hydroxymethylcytidine, and pseudouridine (Ψ).[11-19] While a previous study reported the production of antibodies targeting inosine for the enrichment of tRNAs, this method also displayed adsorptivity to other nucleobases and has not been further demonstrated in any other contexts.[20] Thus, no generally applicable methods currently exist for the derivatization and/or enrichment of

inosines in RNA, significantly limiting both depth and sensitivity in identifying and studying A-to-I RNA editing dynamics across the transcriptome.

In the design of a reagent for affinity capture of inosine-containing RNAs, we hypothesized that an acrylamide electrophile would provide similar reactivity towards inosine as acrylonitrile, while offering the structural flexibility to install an affinity handle for enrichment. This reagent would provide both fluorescent labeling of inosines and the ability to perform affinity capture of A-to-I edited RNA transcripts using a commercially available anti-fluorescein antibody (**Figure 2.1b,c**). Here, we develop and evaluate this strategy and demonstrate the first generalizable inosine-reactive electrophile for fluorescent detection and affinity enrichment of inosine-containing transcripts.

2.3 Results and Discussion:

To test our hypothesis, we first carried out a facile synthesis to generate acrylamidofluorescein (**Figure 2.2**). After designing and synthesizing the acrylamidofluorescein reagent, we assessed initial labeling performance by reacting acrylamidofluorescein and acrylonitrile with each of the major ribonucleosides: inosine (I), pseudouridine (Ψ), uridine (U), guanosine (G), adenosine (A), and cytidine (C). Closely mimicking the ICE reaction conditions, a mixture comprising 50 mM ribonucleoside and 250 mM of either acrylonitrile or acrylamidofluorescein was prepared in 50:50 triethylammonium acetate:ethanol at pH 8.6.



5-aminofluorescein acrylic acid acrylamidofluorescein

Figure 2.2 Synthesis of acrylamidofluorescein.

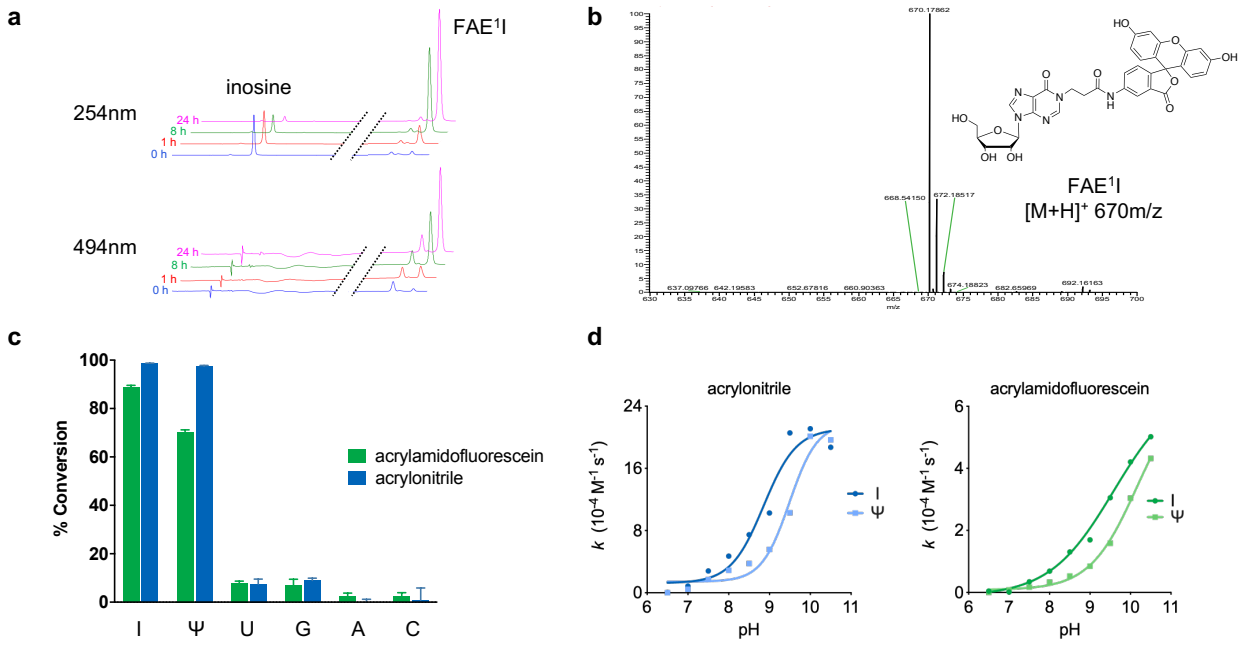


Figure 2.3 Reactivity characteristics of acrylamidofluorescein. a) Representative HPLC traces depicting the reaction between inosine and acrylamidofluorescein over 24 hours. Disappearance of inosine (I) correlates with the appearance of a new putative N1-fluoresceinacrylamidoethylinosine (FAE1I) product peak. b) ESI-MS analysis confirming mass identity of FAE1I product. c) Reactivity panel of acrylonitrile and acrylamidofluorescein with ribonucleosides after 24 hours. d) Dependence of reaction rate constants on pH for the major reacting nucleosides inosine (I) and pseudouridine (Ψ).

The solutions were incubated at 70 °C and the reaction was monitored by HPLC over 24 hours. As illustrated in **Figure 2.3a**, disappearance of inosine peaks is clearly shown along with the formation of a new product peak in both 254 nm and 494 nm chromatograms. This product peak was isolated and analyzed using ESI-MS and MS/MS analysis, confirming the identity of the predicted N^1-fluoresceinamidoethylinosine (FAE^{1}I) product (**Figure 2.3b**). Using ribonucleoside peak areas in the chromatograms, we determined the ratio of reacted vs unreacted ribonucleoside to calculate average conversion percentages for each base at various time points over 24 hours (**Figures 2.3c, A6**).

While acrylamidofluorescein and acrylonitrile exhibit similar reactivity trends, it is clear from the data that acrylonitrile has higher reaction efficiency (**Figures 2.3d**, **A6**). This is likely due to the difference in electron withdrawing properties between the two reagents, which contributes significantly to the kinetics of addition reactions.[21-22] Given that the amide group is less

withdrawing than the nitrile moiety, these results are then unsurprising. Regardless, acrylamidofluorescein and acrylonitrile display similar overall labeling selectivity, exhibiting major product formation with I and Ψ, minimal reactivity with U and G, and virtually no reactivity with A and C throughout extended reaction times. While both reagents display reactivity with Ψ, these observations are consistent with previous studies using acrylonitrile and serve to demonstrate the similar reactivity profiles of both acrylonitrile and acrylamidofluorescein. Indeed, the first reports of acrylonitrile-nucleoside labeling demonstrated its robust reactivity with N^1 on both inosine and Ψ.[21-22] To further validate addition of acrylamidofluorescein at N^1 of inosine, we assessed the effect of pH on reaction rates. Early characterizations of acrylonitrile reactivity with inosine showed that cyanoethylation is strongly pH dependent, suggesting N^1 deprotonation is required for reactivity. Similarly, the data in **Figure 2.3d** illustrate the direct correlation between reaction rate and pH and highlight the preferred reactivity with inosine at ~pH 8.5-8.6, consistent with the known p$\underline{K}_a$ values of N^1 for inosine (8.7)[23] and pseudouridine (9.5).[24] Taken together with the MS spectra, these results strongly support the predicted N^1 addition to inosine and further suggest a similar labeling mechanism of acrylamidofluorescein compared with the well characterized chemistry of acrylonitrile.

Given the promising results of our reagent with ribonucleosides, we next sought to demonstrate acrylamidofluorescein labeling of inosine in RNA oligoribonucleotides. As a test system for these studies, we chemically synthesized two short RNAs containing a 5' Cy5 fluorescent label and an adenosine (RNA-A-Cy5) or inosine (RNA-I-Cy5) at a defined position. We subjected each of these RNAs to acrylamidofluorescein labeling and denaturing PAGE analysis. As shown in **Figure 2.4a**, fluorescein labeling is clearly observed in RNA-I-Cy5 with increasing reaction times, and the labeled product exhibits a slight decrease in migration rate. In comparison, only a faint signal is observed for RNA-A-Cy5, even after a 48 hour reaction time. Given that the presence of inosine is the only molecular difference between these two RNA strands, these data are indicative of selective fluorescein addition at this nucleotide position.

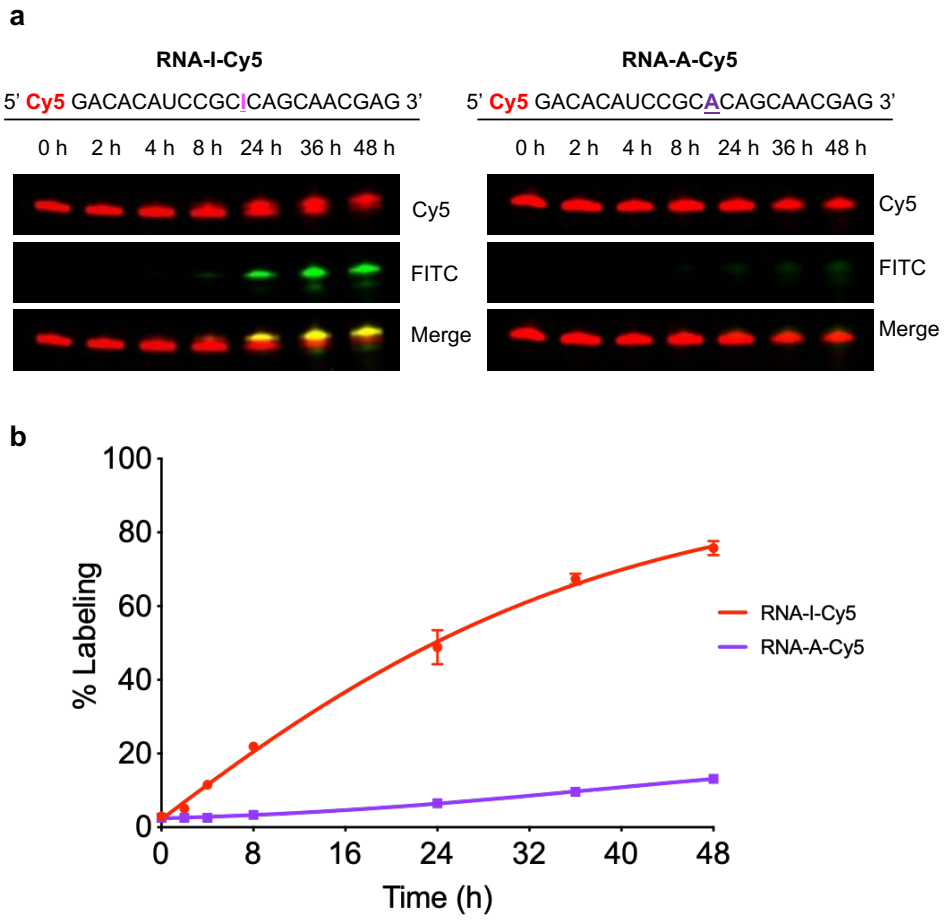


Figure 2.4. RNA labeling with acrylamidofluorescein. a) Denaturing PAGE analysis of synthetic oligoribonucleotides labeled with acrylamidofluorescein. b) Densitometric quantification of oligoribonucleotide labeling.

Densitometric analysis was performed on the labeled RNA bands in the gels and normalized to standard amounts of fluorescein and Cy5-labeled control oligo nucleotides. These data were then used to calculate labeling yield as a function of reaction time (**Figure 2.4b**), which illustrates good selectivity for labeling of RNA-I-Cy5 compared to RNA-A-Cy5. This experiment also highlights the importance of reaction time in maximizing inosine labeling efficiency while maintaining selectivity, as we observe optimal RNA-I:RNA-A labeling ratios at approximately 24 hours. While longer RNA transcripts can undergo hydrolysis in mild alkaline conditions at elevated temperatures, these data demonstrate the stability of shorter RNA segments under our reaction conditions. We envision the use of this labeling method with high-throughput RNA-seq workflows, which require

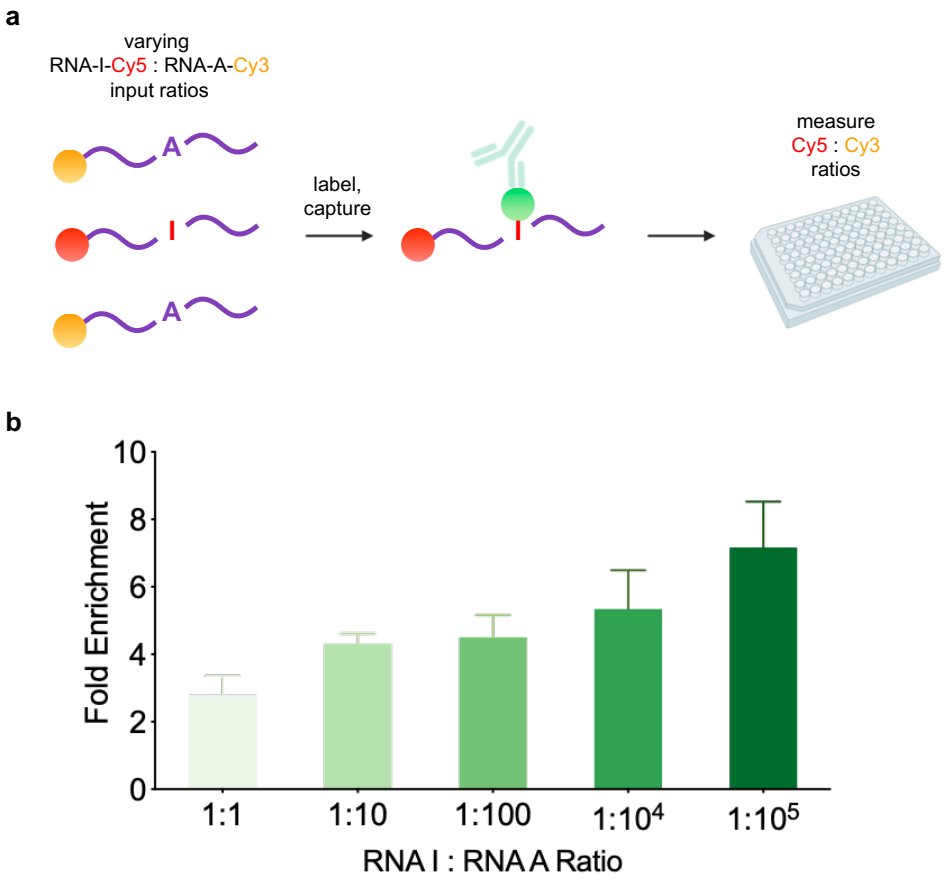


Figure 2.5. Labeling and affinity enrichment of inosine-containing RNAs. a) Workflow for quantifying pulldown efficiency with acrylamidofluorescein labeling and immunoprecipitation. b) Fold enrichment of inosine-containing oligoribonucleotides from varying mixtures.

fragmentation of longer RNAs prior to library preparation and amplification. This fragmentation step is employed upstream of chemical labeling and pulldown in the analogous strategies described above for mapping other RNA modifications,[11-12, 16-18, 25] and thus our results indicate compatibility with these platforms.

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

were combined in varying ratios, subjected to acrylamidofluorescein labeling, and then affinity captured using an anti-fluorescein monoclonal antibody and protein A/G magnetic beads. After extensive washing, bound oligoribonucleotides were eluted and quantified using a fluorescence plate reader (**Figure 2.5a**). Final concentrations of RNA-A-Cy3 and RNA-I-Cy5 after pull-down were compared to input ratios to calculate fold-enrichment. As shown in **Figure 2.5b**, acrylamidofluorescein labeling coupled with IP enables upwards of 7-fold enrichment of inosine-containing oligoribonucleotides, with the highest enrichment factors achieved for samples containing the lowest ratios of the inosine-containing RNA.

2.4 Conclusion:

Chemical modification strategies coupled with affinity capture have significantly improved the sensitivity and accuracy in sequencing, mapping, and characterizing several modified RNA bases.[11-19] However, there are no extant methods for enriching A-to-I edited transcripts, greatly limiting our ability to understand the true scale and impact of A-to-I modifications on cell and tissue function. Here we address this challenge through the synthesis of a novel acrylamidofluorescein reagent that chemically labels inosine and enables the enrichment of A-to-I edited transcripts. While the observed reactivity between acrylamidofluorescein and Ψ may seem problematic for the effective isolation and enrichment of inosine-containing transcripts from biological samples, Ψ is found predominantly in ribosomal RNAs and tRNAs, and thus effective fractionation of total RNA samples can remove significant quantities of this modified base. In coding RNAs, I also vastly outnumbers Ψ, with current estimates of ~500:1 I:Ψ.[8, 18, 26] Additionally, methods have now been developed to selectively label and/or deplete Ψ from total RNA pools using biotinylated carbodiimide reagents.[18] We envision that acrylamidofluorescein could be coupled with carbodiimide labeling to achieve simultaneous selective modification and separate enrichment of transcripts containing I and Ψ, respectively. We also recognize the potential to improve enrichment by reducing reactivity with the natural ribonucleosides U and G, and efforts are

underway to explore alternative acrylamide structures toward this goal. Regardless, given the present lack of methods for isolating inosine-containing RNAs, the research presented here represents a critical first step toward integrating chemical labeling and enrichment methods for this important application.

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

2.5 Materials and Methods:

Synthesis of Acrylamidofluorescein

To a solution of 5-aminofluorescein (1.00 g, 2.88 mmol) in pyridine (8.00 ml, 98.9 mmol), *N*-(3- dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (828 mg, 4.32 mmol) and acrylic acid (0.390 ml, 5.76 mmol) were added and left to stir at room temperature overnight. Once the 5- aminofluorescein was consumed, as determined by TLC, the reaction was dried under reduced pressure to form a crude oil. The crude oil was added into 20 mL of 10% sodium hydroxide and extracted using dichloromethane. The organic layer was collected and acidified by adding concentrated hydrochloric acid until orange precipitates formed. The product was then vacuum filtered and dried to yield 0.930 g (80.7%) of orange powder. ^{1}H NMR (400 MHz, DMSO-d6) δ 10.93 (s, 1H), 8.47 (s, 1H), 7.96 (d, *J* = 8.6, 1H), 7.22 (d, *J* = 8.2, 1H), 6.77 (s, 2H), 6.53-6.68 (m, 5H), 6.35 (d, *J* = 16, 1H), 5.84 (d, *J* = 11.3, 1H). ^{13}C NMR (400 MHz, DMSO-d6) δ 168.9, 164.2, 160.9, 152.8, 141.2, 132.0, 130.0, 128.2, 127.7, 126.7, 125.4, 114.8, 113.7, 110.6, 102.7. HRMS m/z (ESI) calcd for $C_{23}H_{15}NO_6$ (M+H)+ 402.09776, found 402.09658.

Ribonucleoside Labeling and HPLC Analysis

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

Acrylamidofluorescein reactions were analyzed using a linear mobile phase gradient from 5% to 45% acetonitrile in water over 25 minutes. All mobile phases contained 0.1% trifluoracetic acid.

RNA Oligoribonucleotides

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

RNA-I-Cy5	5' Cy5/GACACAUCCGC**I**CAGCAACGAG 3'
RNA-A-Cy3	5' Cy3/GACACAUCCGC**A**CAGCAACGAG 3'
RNA-A-Cy5	5' Cy5/GACACAUCCGC**A**CAGCAACGAG 3'

Oligoribonucleotide Labeling and PAGE Analysis

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

Oligoribonucleotide Labeling and Immunoprecipitation Pulldown

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

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

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

2.6 References:

1. Knutson, S. D.; Ayele, T. M.; Heemstra, J. M., Chemical Labeling and Affinity Capture of Inosine-Containing RNAs Using Acrylamidofluorescein. *Bioconjugate chemistry* **2018,** *29* (9), 2899-2903.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

22. Ofengand, J., The Function of Pseudouridylic Acid in Transfer Ribonucleic Acid I. THE SPECIFIC CYANOETHYLATION OF PSEUDOURIDINE, INOSINE, AND 4-THIOURIDINE BY ACRYLONITRILE. *Journal of Biological Chemistry* **1967,** *242* (21), 5034-5045.

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

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

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

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

Chapter 3

Chemical Profiling of A-to-I RNA Editing Using a Click-Compatible Phenylacrylamide[1*]

3.1 Abstract:

Straightforward methods for detecting adenosine-to-inosine (A-to-I) RNA editing are key to better understanding its regulation, function, and connection with disease. We address this need by developing a novel reagent, *N*-(4-ethynylphenyl)acrylamide (EPhAA), and illustrating its ability to selectively label inosine in RNA. EPhAA is synthesized in a single step, reacts rapidly with inosine, and is "click"-compatible, enabling flexible attachment of fluorescent probes at editing sites. We first validate EPhAA reactivity and selectivity for inosine in both ribonucleosides and RNA substrates, and then apply our approach to directly monitor *in vitro* A-to-I RNA editing activity using recombinant ADAR enzymes. This method improves upon existing inosine chemical labeling techniques and provides a cost-effective, rapid, and non-radioactive approach for detecting inosine formation in RNA. We envision this method will improve study of A-to-I editing and enable better characterization of RNA modification patterns in different settings.

[*]Adapted from Ref. 1 with permission from Knutson, S. D.; Korn, M. M.; Johnson, R. P.; Monteleone, L. R.; Dailey, D. M.; Swenson, C. S.; Heemstra, J. M. *Chemistry–A European Journal,* **2020***, 26*(44), 9874-9878. Copyright 2020 Chemistry Europe.

3.2 Introduction:

RNA is chemically modified by a number of enzymes after transcription, in turn influencing RNA stability, localization and activity within the cell. Adenosine-to-inosine (A-to-I) RNA editing is one of the most widespread modifications, and is performed by adenosine deaminases acting on RNA (ADARs) (**Figure 3.1a**).[2] Adenosine deamination changes the molecular structure and hydrogen bonding pattern of the nucleobase, and resulting inosines instead base pair with cytidine to effectively recode these sites as guanosine. Editing sites within protein-coding mRNAs directly alter amino acid sequences and produce different protein isoforms. Non-coding RNAs also undergo extensive editing, including microRNAs and small-interfering RNAs, significantly altering their biosynthesis, localization, and gene regulation properties.[3-4] A-to-I editing is essential for a number of biological processes including tissue development,[5-6] neurological function,[7] and immune system activation.[8] Dysfunctional editing is also directly linked with autoimmune diseases,[9-10] neurological disorders,[11] and several types of cancer.[12-13]

Despite this importance, our overall understanding of A-to-I editing regulation is limited. In particular, while many sites have been identified (> 5 million),[14-15] it is unclear why certain sites are edited at higher frequency than others and what precise function they each serve.[16] Efforts to map A-to-I locations and ADAR binding sites have revealed that editing patterns are highly complex and variable in humans,[8, 17-19] and the precise mechanisms by which ADAR enzymes bind to and edit specific RNA sequences remain unclear. This gap is also significant for therapeutic site-directed RNA editing strategies,[20] as both the design and precise implementation of this machinery is reliant on a thorough understanding of ADAR regulation.

Detecting inosine formation in RNA is of central importance for characterizing editing mechanisms. While high-throughput RNA sequencing (RNA-seq) is commonly employed for large scale detection and mapping of A-to-I sites,[21] this method is also costly, prone to random sampling errors, and requires complex bioinformatic analyses.[22-23] Alternatively, model reactions using ADAR enzymes with small RNA substrates (~20-50 nt) have yielded substantial insights into how

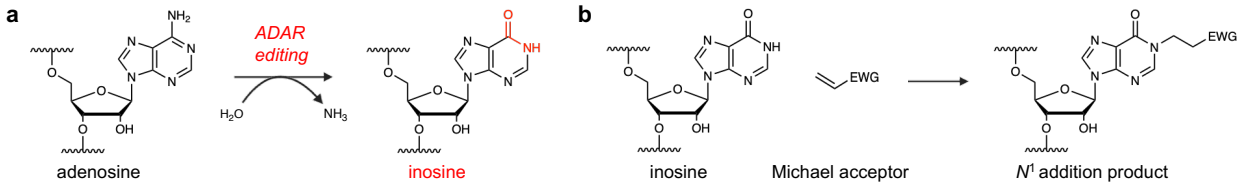


Figure 3.1. Formation and detection of inosine in RNA. a) A-to-I editing is catalyzed by adenosine deaminases acting on RNA (ADAR, red). b) Inosine nucleobases can be detected by reacting with Michael acceptors to yield N^1 addition products.

certain RNA sequences and structural motifs are recognized and edited.[18, 24-27] Although A-to-I sites are "visible" as A-G transitions in Sanger sequencing,[28-29] these methods require relatively large RNA substrates (>300-400 nt), and are incompatible with the small RNA strands that are ideal for these experiments. To detect inosine in smaller substrates, adenosines within chimeric RNA strands are often internally radiolabeled with P^{32}. After ADAR editing, RNA substrates are then digested with nuclease P1 and A-to-I nucleotide changes are detected with autoradiographic thin layer chromatography.[24, 26-27] While this method is effective, it is also time-consuming to construct each RNA substrate, and assays using these radioactive materials require specialized training, instrumentation, and waste disposal protocols. Alternatively, deamination can be detected by incorporating alkyne-modified or fluorogenic thiolated-adenosine analogues into RNA substrates,[30-31] but these approaches can introduce structural alterations into RNA targets that impact ADAR targeting, and they require lengthy phosphoramidite monomer synthesis.

Direct chemical detection of A-to-I editing would circumvent these assay limitations, and inosine has been shown to react with Michael acceptors to yield N^1 addition products (**Figure 3.1b**).[32] In particular, a recent approach utilized acrylonitrile to alkylate inosines for reverse-transcription termination sequencing. Termed "inosine chemical erasing sequencing" (ICE-seq), this technique improved the accuracy of detecting A-to-I sites using Sanger sequencing, but also suffered from significant limitations in sensitivity, and requires matched DNA and RNA samples for each assay.[33] Subsequent work derivatized acrylonitrile for use in "clickable"-biotinylation and enrichment of A-to-I edited transcripts.[34] While acrylonitrile is a promising scaffold for chemical

detection of inosine, derivatizing these reagents is difficult and requires several synthetic and purification steps. Alternatively, we recently reported an acrylamidofluorescein reagent that enables fluorescent detection and enrichment of inosine in RNA.[35] While our initial study demonstrated feasibility, acrylamidofluorescein also displayed poor solubility and was restricted to fluorescein addition. However, acrylamide scaffolds are simple to modify, and we were interested in elaborating upon this architecture to develop an improved and more generalizable inosine probe.

3.3 Results and Discussion:

Toward this goal, we first screened potential acrylamide scaffolds for inosine reactivity using our previously established reaction conditions (50:50 EtOH:1M triethylammonium acetate pH 8.6, 70 ºC) (**Figure 3.2b**). Acrylamide and *N*-phenylacrylamide were both highly reactive toward inosine, whereas alkylacrylamide scaffolds (mPEG acrylamide and *N*-hydroxyethylacrylamide) gave little to no product formation (**Figures 3.2b, B1**). Interestingly, *N*-phenylacrylamide is structurally similar to acrylamidofluorescein (**Figure 3.2b**), and it is likely this moiety exhibits sufficient electron-withdrawing properties consistent in other Michael acceptors.

We next sought to derivatize *N*-phenylacrylamide to enable secondary functionalization with fluorescent probes using copper-catalyzed azide-alkyne cycloaddition (CuAAC), or “click” chemistry. We quickly identified 4-ethynylaniline as a commercially available, alkyne-functionalized intermediate, enabling us to employ a one-step coupling with acrylic acid to yield *N*-(4-ethynylphenyl)acrylamide (EPhAA, **Figure 3.3**). After verifying product identity (**Figures B2-B4**), we tested EPhAA for reactivity with inosine and confirmed appearance of the expected addition product N^1-ethynylphenylamidoethylinosine ($EPhAE^1I$) by HPLC and ESI-MS analysis (**Figures 3.4.a-b, B5, B7**). Deprotonation of N^1 on inosine is known to mechanistically drive this reaction,[32] and so to further confirm that EpHAA undergoes addition at this position, we tested our labeling reaction across different pH values (**Figure 3.4c**).

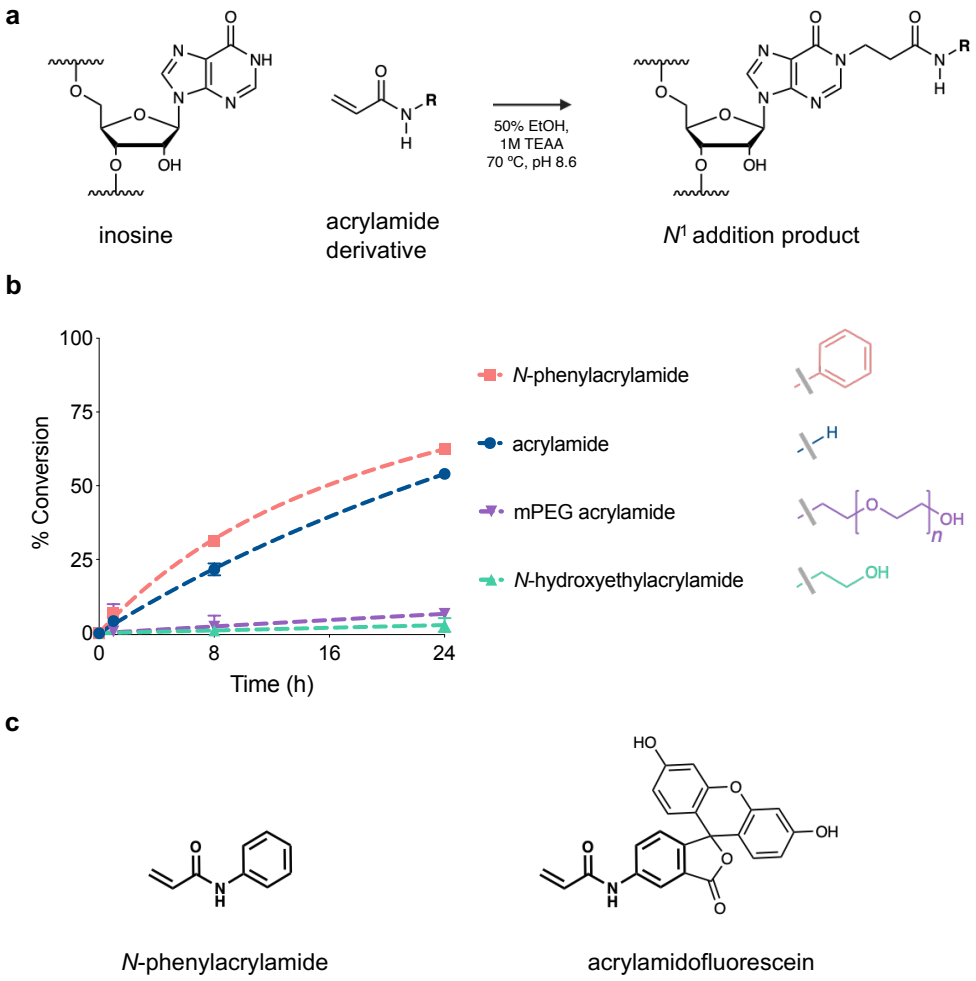


Figure 3.2. Evaluating inosine reactivity of acrylamide scaffolds. a) Different acrylamide derivatives were evaluated for reactivity with inosine by b) monitoring product formation via HPLC. Values represent mean with S.D. error bars (n = 3). c) Structural similarities (bold) between *N*-phenylacrylamide and acrylamidofluorescein.

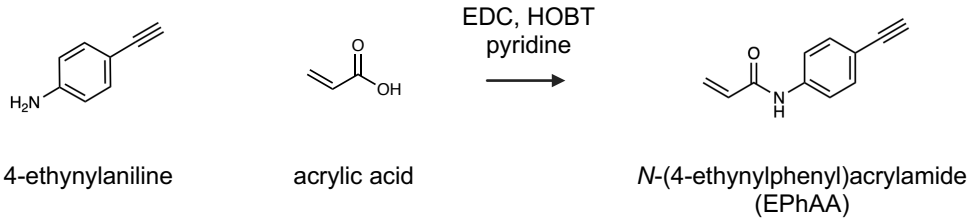


Figure 3.3. One-step synthesis of *N*-(4-ethynylphenyl)acrylamide (EPhAA).

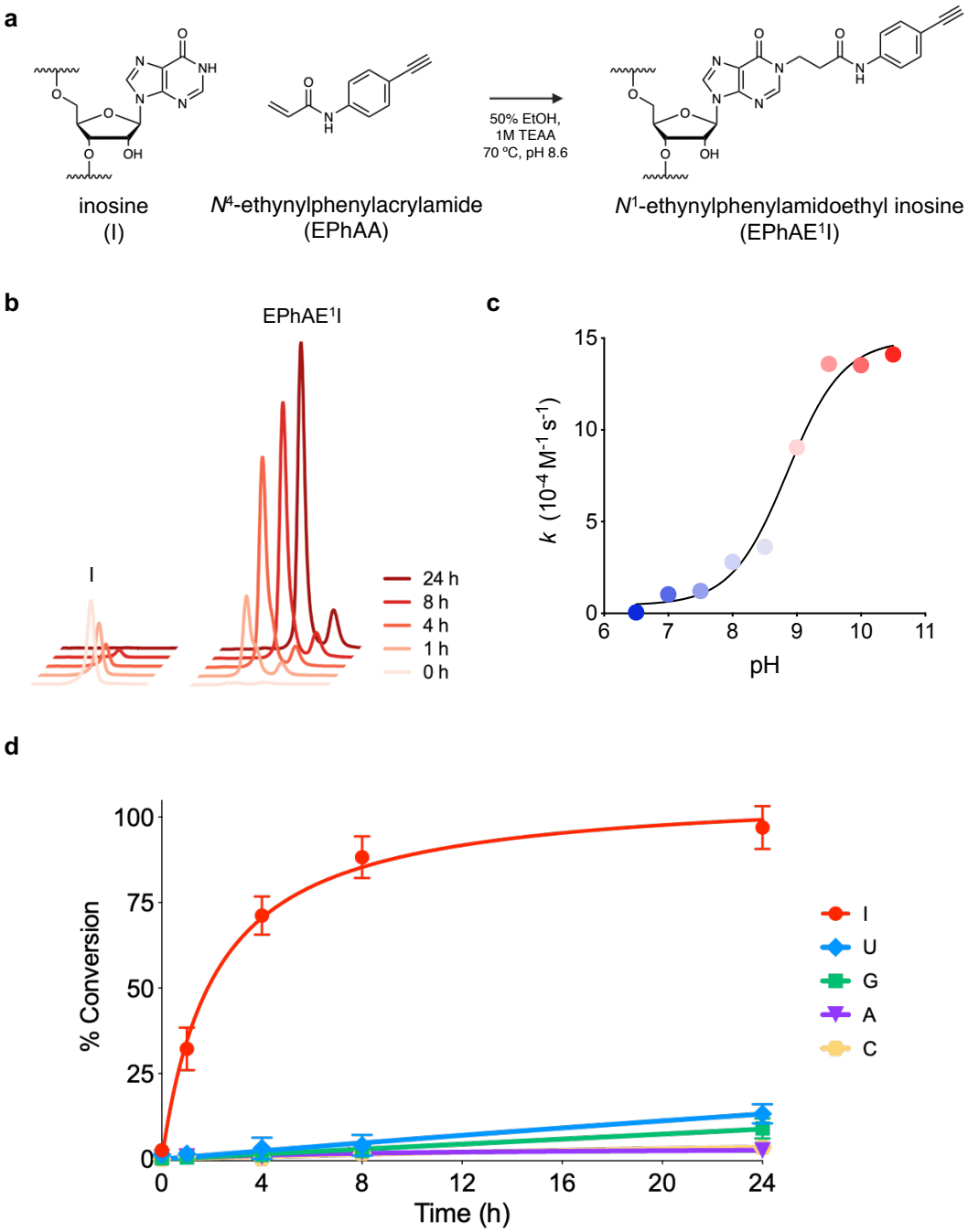


Figure 3.4. Reactivity assessment of EPhAA with ribonucleosides. a) Reaction scheme of EPhAA with inosine. b) Representative HPLC traces depicting formation of EPhAE^{1}I over 24 hours. c) Dependence of pH on reaction rate constants for EPhAA addition on inosine. d) EPhAA reactivity with each of the major ribonucleosides over 24 hours. Values represent mean with S.D. error bars (n = 3).

As expected, we observed a steep increase in reaction rates consistent with the known pK_a value of inosine N^1 (~8.7).[36] In assembling reaction mixtures, we also noted improved solubility of EPhAA compared to acrylamidofluorescein, and we were able to double our normal working concentrations to ~500 mM. As expected, this resulted in more rapid overall reaction kinetics, and when compared to our previous reagent, we observed a ~2-3-fold increase in conversion percentages at similar reaction times (**Figure 3.4d**).[35] Next, we incubated EPhAA with the remaining ribonucleosides uridine (U), guanosine (G), adenosine (A), and cytidine (C) (**Figure 3.4d, B5**). In this test, we observed robust selectivity towards I, and while U and G have similar acidic nitrogens that can be labeled with Michael acceptors,[37-38] N^1 on inosine displays much higher nucleophilicity and reactivity with these reagents,[32] and off-target labeling was only observed at extended reaction times. Acrylonitrile and acrylamide reagents are also known to react with pseudouridine (Ψ),[32] and we determined that our reagent exhibited similar reactivity characteristics, as we observed the expected N^1 addition product (**Figure B5f, B6, B8**). While this off-target reactivity may seem problematic, the primary application we envision for our EPhAA reagent is detecting inosine in model RNA strands to monitor ADAR activity, and Ψ can be omitted from these substrates. Additionally, if assays necessitate the use of cellular RNA or require Ψ content, existing carbodiimide reagents can be employed to selectively block and deplete Ψ sites.[39-40] Lastly, we assessed general EPhAA stability by incubating the reagent in the absence of ribonucleoside, and while we observed some degradation of the reagent in our labeling conditions, this effect was minimal and only seen in extended reaction times (**Figure B5g**).

With our validated reagent in hand and given our ultimate goal of detecting ADAR-mediated A-to-I editing, we next sought to label inosine in RNA oligonucleotides. ADARs commonly deaminate double-stranded RNA substrates at A:C mismatches, “flipping out” the target A into the active site to leave behind an “orphan C” base.[2-3, 41] To mimic this, we synthesized a target strand inspired by an mRNA hairpin (HER1) that undergoes editing by human ADAR1 (hADAR1) at a defined site (**Figure 3.5a**).[24] To detect inosine, we planned to use CuAAC to install

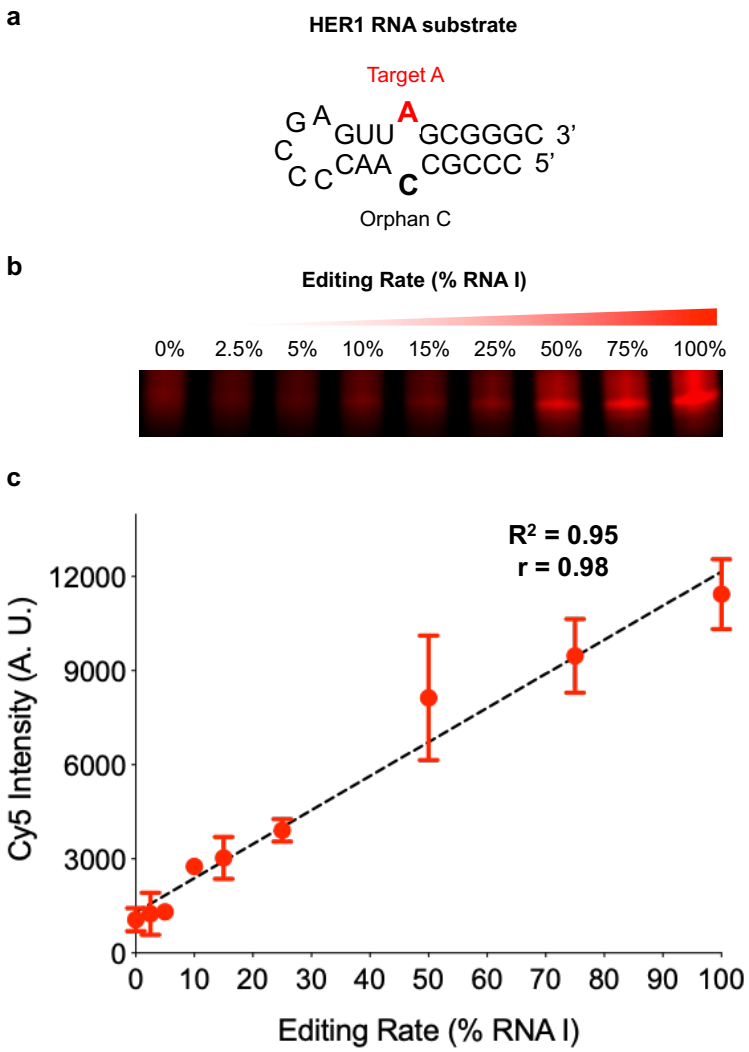


Figure 3.5. Validating chemical detection of inosine in RNA. a) Sequence and structure of the model HER1 RNA hairpin substrate with target A (red) and the orphan C base (black). b) EPhAA labelling and PAGE analysis of different simulated RNA editing rates and c) densitometric quantification of signal (arbitrary units, A.U.) across reactions. Values represent mean with S.D. error bars (n = 2).

a fluorophore after EPhAA labeling. We first wanted to verify that our click-labeling conditions were optimal, so we subjected an alkyne-functionalized DNA oligonucleotide to a standard CuAAC protocol with a picolyl azide-functionalized Cyanine5 fluorophore (Cy5-N_3).[42] As shown in **Figure B9**, we observed complete labeling of the alkyne-modified strand, indicating these conditions would be compatible with our workflow. Given our previous data showing background labeling of U and G nucleotides at very long reaction times (**Figure. 3.4d**), we were next interested in optimizing EPhAA labeling time to minimize off-target attachment. To test this, we reacted

HER1 RNA A and I substrates with EPhAA for increasing amounts of time in independent duplicate trials followed by CuAAC labeling (**Figure B10**). In RNA I samples, we expectedly saw a rapid appearance of a major product band, indicative of inosine fluorescent chemical labeling. RNA A did not produce significant signal, but did exhibit a "smear" in longer reaction times, likely indicating a mixture of products resulting from off-target labeling at U and G residues. This distribution was also observed in RNA I reactions with extended labeling times, further corroborating this hypothesis. We also stained all RNA species in gels using SYBR gold to assess overall labeling efficiency (**Figure B10**). Although we did not achieve full conversion of the RNA I strand, we identified 6 hours as an optimal EPhAA reaction time to achieve robust selectivity (~60-fold I vs A labeling, **Figure B10b**). In particular, when compared to our previous acrylamidofluorescein reagent, we achieved significantly better selectivity (~60-fold vs ~8-fold) and with much shorter reaction times (6 h vs 24 h).[35] Lastly, given our ultimate goal of detecting RNA editing, we were also interested in assessing the linearity of our method for measuring different A-to-I editing "rates." To test this, we performed a series of duplicate labeling reactions using varying ratios of A and I substrate while keeping the total amount of RNA constant. As shown in **Figures. 3.5b-c**, inosine content was highly proportional to fluorescent intensity and we observed linearity between these variables (R^2 = 0.95, r = 0.98), providing additional confidence that our method could accurately measure A-to-I editing activity.

Finally, we wanted to directly illustrate the utility of our method for detecting ADAR-mediated A-to-I editing. Given that HER1 is selectively recognized and edited by hADAR1, we first expressed and purified recombinant deaminase domains from this enzyme. Additionally, we prepared a mutant hADAR1 enzyme (E1008Q) which displays increased catalytic activity and speed, likely by providing enhanced stability of the orphan C nucleobase (**Figure 3.6a**).[24, 41] We envisioned that these enzyme variants would be a suitable test of our labeling method and further validate this approach for detecting catalytic deamination differences arising from biochemical variations in ADAR enzymes. As shown in Figures. **3.6b-d**, we performed *in vitro* deamination

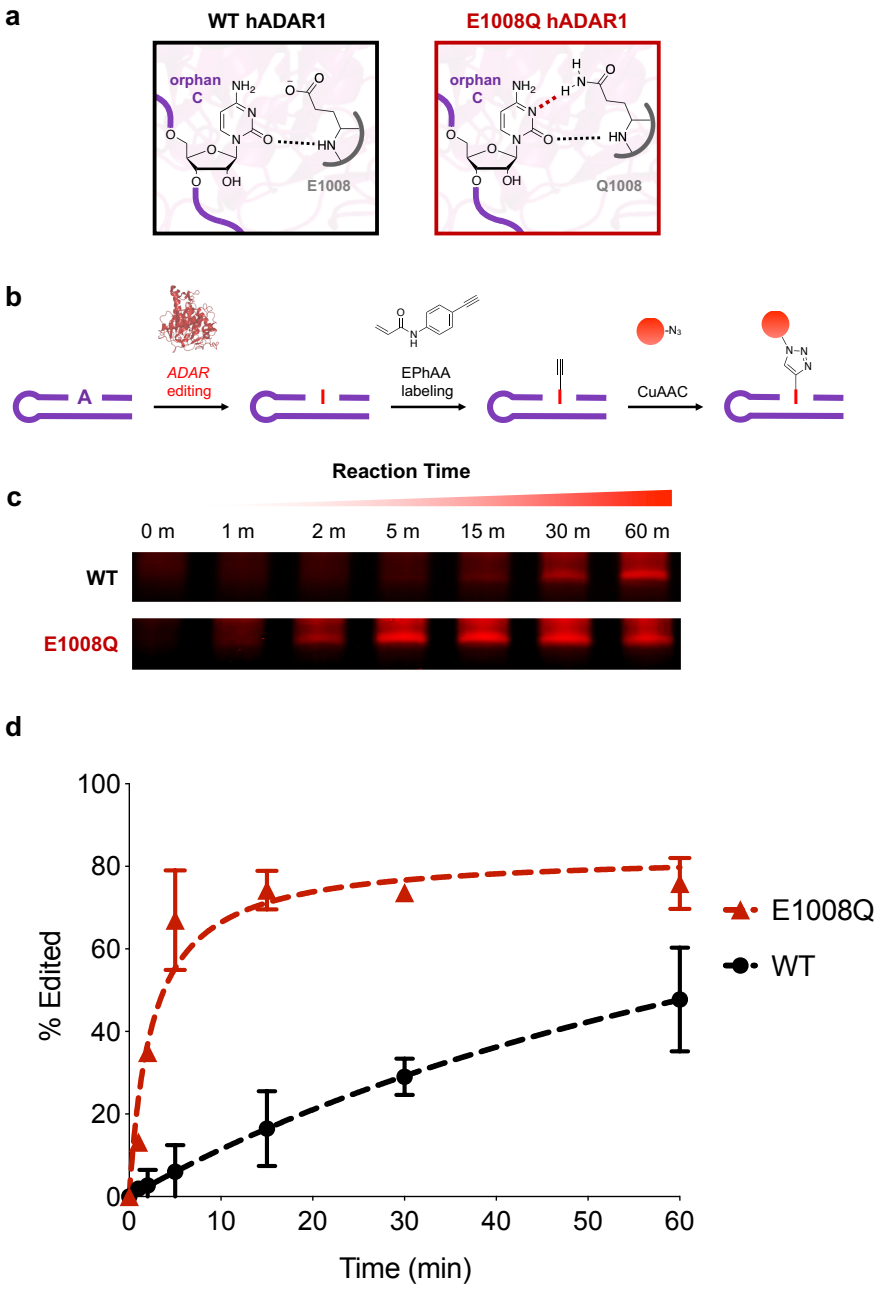


Figure 3.6. Chemical detection of ADAR1-mediated A-to-I RNA editing. a) ADAR1 amino acids interact with the orphan C base, with the E1008Q point mutation providing increased stability and overall catalytic efficiency. b) Overall workflow for detecting A-to-I editing with EPhAA labeling and CuAAC. c) EPhAA labelling and PAGE analysis of *in vitro* A-to-I RNA editing reactions. d) Densitometric quantification of signal across different RNA editing reactions using wild type (WT, black) and E1008Q mutant (red) ADAR enzymes. Values represent mean with S.D. error bars (n = 2).

experiments on our HER1 RNA A substrate with both enzymes, and we were able to fluorescently detect A-to-I conversion and robustly distinguish activity between wild type ADAR and the hyperactive E1008Q mutant. In addition to plotting overall editing activity, we also estimated initial velocities (v_i) for both enzymes and observed ~14-fold increase in turnover speed for the E1008Q mutant (**Figure B11**), which is in close agreement with previous activity comparisons of these hADAR1 isoforms.[24]

3.4 Conclusions:

A-to-I RNA editing is a widespread post-transcriptional modification that is essential for a variety of cellular processes, and aberrant RNA editing is directly linked to a number of diseases. Despite progress in characterizing A-to-I editing regulation and dynamics, significant gaps remain in our understanding of why certain sites are edited more than others, and what functional roles these editing events play. Simple and straightforward methods for detecting inosine formation in RNA and measuring ADAR activity are integral to addressing these knowledge gaps. In this work, we show the development and validation of a novel reagent, *N*-(4-ethynylphenyl)acrylamide (EPhAA), as an economical and rapid chemical labeling method for assaying A-to-I RNA editing *in vitro*. This reagent is simple to synthesize, improves upon existing labeling approaches, and robustly detects inosine in RNA. We envision this method will be a valuable tool to complement existing techniques for characterizing ADAR mechanisms and deciphering A-to-I RNA editing signatures in a variety of contexts. In particular, we view EPhAA labeling as a cost-effective and rapid method to elucidate the effects of RNA sequence and structure on ADAR editing activity *in vitro*, better assess the biochemical impact of disease-relevant ADAR mutations on pathological A-to-I editing, and accurately measure the activity of engineered recombinant enzymes for site-directed RNA editing.

3.5 Materials and Methods:

Inosine labeling with acrylamide derivatives and HPLC analysis

Acrylamide, *N*-phenylacrylamide, and *N*-hydroxyethylacrylamide were purchased from Sigma Aldrich Corporation (St. Louis, MO). mPEG-acrylamide (MW 1000 g/mol) was purchased from Creative PEGWorks (Chapel Hill, NC). Inosine ribonucleoside was purchased from Sigma Aldrich Corporation (St. Louis, MO). Labeling reaction mixtures were comprised of 50 mM inosine and 250 mM of each acrylamide derivative in 50:50 EtOH:1M triethylammonium acetate (TEAA) pH 8.6. Reactions were incubated at 70 °C for the time periods indicated. Reversed-phase HPLC analysis was performed on an Agilent 1260 Infinity II system using a 4 µm, 150 x 4.6 mm Phenomenex Synergi Fusion-RP 80A C18 column. Samples were prepared in a stationary phase solution of 5% acetonitrile in PBS. Acrylamide, mPEG-acrylamide, and *N*-hydroxyethylacrylamide reactions were analyzed using an isocratic mobile phase of 5:95 acetonitrile:water. *N*-phenylacrylamide reactions were analyzed using a linear mobile phase gradient from 5% to 45% acetonitrile in water over 15 minutes. All mobile phases contained 0.1% trifluoracetic acid. Percent conversion in each reaction was defined as the inosine peak area relative to unreacted inosine at the same time point without any reagent.

Synthesis of *N*-(4-ethynylphenyl)acrylamide

Unless otherwise noted, all starting materials were obtained from Sigma Aldrich Corporation (St. Louis, MO) and were used without further purification. Flash column chromatography was carried out using silica gel 60 (230–400 mesh). ^{1}H NMR chemical shifts are expressed in parts per million (δ). Mass spectra were obtained on an Agilent 6230 TOF LC/MS. To a solution of 4-ethynylaniline (2.0 g, 17.1 mmol), 0.2 eq hydroxybenzotriazole (462.1 mg, 3.42 mmol), and 1.2 eq 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (3.934 g, 20.52 mmol) in anhydrous pyridine (30 mL, 372.4 mmol) under N_2 was added 1.2 eq of acrylic acid (1.41 mL, 20.49 mmol). The reaction was stirred at room temperature overnight. Consumption of 4-

ethynylaniline was confirmed by TLC in 1:1 hexanes:ethyl acetate. The crude reaction mixture was diluted in 50 mL ethyl acetate and washed sequentially with water and brine. The aqueous layer was back-extracted twice with ethyl acetate, and the collected organic layer was dried with $MgSO_4$ and filtered. The organic layer was concentrated under reduced pressure and purified by column chromatography (1:1 hexanes:ethyl acetate). The purified product was concentrated under reduced pressure and dried *in vacuo* to yield 1.26 g (42%) of a salmon-colored powder. ^{1}H NMR (400 MHz, DMSO-d6) δ 10.29 (s, 1H), 7.65 (d, *J* = 8.5 Hz, 2H), 7.40 (d, *J* = 8.3 Hz, 2H), 6.40 (dd, *J* = 17.6, 10.0 Hz, 1H), 6.29 – 6.19 (m, 1H), 5.78 – 5.70 (m, 1H), 4.06 (s, 1H). ^{13}C NMR (400 MHz, DMSO-d6) δ 163.73, 139.97, 133.23, 132.85, 132.78, 132.04, 127.87, 119.60, 116.81, 83.96, 80.44. HRMS m/z (ESI) calculated for $C_{11}H_{10}NO$ $(M+H)^+$ 172.0762, found 172.0637.

Ribonucleoside labeling and HPLC analysis

Ribonucleosides inosine, guanosine, adenosine, cytidine and uridine were purchased from Sigma Aldrich Corporation (St. Louis, MO). Pseudouridine (Ψ) was purchased from Santa Cruz Biotechnology (Santa Cruz, CA). Labeling reaction mixtures were comprised of 50 mM ribonucleoside and 500 mM *N*-(4-ethynylphenyl)acrylamide reagent in 50:50 EtOH:reaction buffer. Phosphate buffered saline (PBS) was used for reactions from pH 6.5-7.5 and 1M triethylammonium acetate (TEAA) for pH 8.0-10.5. Reactions were incubated at 70 °C for the time periods indicated. Reversed-phase HPLC analysis was performed on an Agilent 1260 Infinity II system using a 4 µm, 150 x 4.6 mm Phenomenex Synergi Fusion-RP 80A C18 column. Samples were diluted 1:100 in a stationary phase solution of 5% acetonitrile in PBS. Reactions (1 µL injection) were analyzed using a linear mobile phase gradient from 5% to 45% acetonitrile in water over 15 minutes. All mobile phases contained 0.1% trifluoracetic acid.

DNA oligonucleotide click labeling and PAGE analysis

A FAM labeled DNA strand with an internal alkyne modification (5-octadiynyl deoxyuridine, i5OctdU) was purchased from Integrated DNA Technologies (Skokie, IL) as shown below.

Alkyne DNA 5' FAM/AGCAGCAGGACG/**i5OctdU**/AGCAGAACAGAC 3'

100 pmol of alkyne DNA was CuAAC labeled in a total volume of 100 µL using the Click-&-Go Plus Labeling Kit (Click Chemistry Tools, Scottsdale, AZ) according the manufacturer's protocol. Reactions were given either 1 µL of DMSO (vehicle) or 1 µL of a 5 mM Cy5 picolyl azide (Click Chemistry Tools, Scottsdale, AZ) solution in DMSO. Reactions were incubated for 1 hour at room temperature, after which they were ethanol precipitated and resuspended in 20 µL nuclease-free water. 1 pmol of each sample was resolved on a 10% denaturing polyacrylamide gel and imaged with a GE Amersham Typhoon RGB scanner.

HER1 RNA labeling selectivity

Custom RNA oligonucleotides were purchased from the University of Utah DNA synthesis core facility (Salt Lake City, UT) as shown below.

HER1 RNA A 5' CCCGCCAACCCCGAGUU**A**GCGGGC 3'

HER1 RNA I 5' CCCGCCAACCCCGAGUU**I**GCGGGC 3'

In duplicate, 100 pmol of either HER 1 RNA A or HER1 RNA I was added to a 0.1 mL solution of 500 mM *N*-(4-ethynylphenyl)acrylamide in 50:50 EtOH:TEAA buffer, adjusted to pH 8.6, and incubated at 70 °C for the indicated time points. Samples were then ethanol precipitated and resuspended in 20 µL nuclease-free water. RNA was then CuAAC labeled using the Click-&-Go Plus Labeling Kit and 1µL of a 5 mM Cy5 picolyl azide solution in DMSO (Click Chemistry Tools, Scottsdale, AZ). Reactions were incubated for 1 hour at room temperature, after which they were ethanol precipitated and resuspended in 10 µL nuclease-free water. 20 pmol of each sample was resolved on a 10% denaturing polyacrylamide gel, stained with SYBR gold (Thermo Fisher

Scientific, Rockford, IL), and imaged with a GE Amersham Typhoon RGB scanner. Densitometric quantification of product bands was performed using ImageJ software. Fold selectivity was defined as the Cy5 lane intensity of RNA I divided by RNA A.

HER1 editing rate linearity

In duplicate, varying mixtures of HER1 RNA I and HER1 RNA A were prepared in a 0.1 mL solution of 500 mM *N*-(4-ethynylphenyl)acrylamide in 50:50 EtOH:TEAA buffer, adjusted to pH 8.6, and incubated at 70 °C for 6 hours. Mixtures were defined as follows:

Editing Rate (%)	pmol HER1 RNA A	pmol HER1 RNA I
100	0	100
75	25	75
50	50	50
25	75	25
15	85	15
10	90	10
5	95	5
2.5	97.5	2.5
0	100	0

After labeling, samples were then ethanol precipitated and resuspended in 20 µL nuclease-free water. RNA was then CuAAC labeled using the Click-&-Go Plus Labeling Kit and 1µL of a 5 mM Cy5 picolyl azide solution in DMSO (Click Chemistry Tools, Scottsdale, AZ). Reactions were incubated for 1 hour at room temperature, after which they were ethanol precipitated and resuspended in 10 µL nuclease-free water. 20 pmol of each sample was resolved on a 10% denaturing polyacrylamide gel and imaged with a GE Amersham Typhoon RGB scanner. Densitometric quantification of bands was performed using ImageJ software. Linear regression and Pearson r correlation analysis was performed using GraphPad Prism 8 software.

hADAR1 WT and E1008Q overexpression and purification

hADAR1 wildtype (WT) and hADAR1 E1008Q deaminase domains were generally expressed and purified as previously described (Macbeth and Bass, 2007; Matthews et al., 2016). Mutagenesis of human ADAR1 (hADAR1) deaminase domain was performed using QuickChange XL Site-Directed Mutagenesis (Agilent, Santa Clara, CA) and transformed into XL10-Gold Ultracompetent cells (Agilent, Santa Clara, CA). S. *cerevisiae* BCY123 cells were transformed with a pSc-ADAR construct encoding hADAR1 WT or hADAR1 E1008Q. Cells were streaked on yeast minimal medium minus uracil (Cm-ura) plates. A single colony was used to inoculate a 15 mL Cm-ura starter culture, which was shaken at 300 r.p.m. at 30 ˚C overnight. The starter culture was used to inoculate 1.5 L yeast growth medium. After cells reached an optical density between 1 - 2, cells were induced with 165 mL of sterile 30% galactose, and protein was expressed for 6 h. Cells were collected by centrifugation and stored at -80 ºC. Cells were lysed in 20 mM Tris-HCl, pH 8.0, 5 % glycerol, 1 mM BME, 750 mM NaCl, 30 mM imidazole, 0.05 % Triton X-100 supplemented with cOmplete EDTA-free protease inhibitor (Sigma Aldrich, St. Louis, MO). Cell lysate was clarified by centrifugation (18,000 rpm, 60 min). Lysate was passed over a 5 mL Ni-NTA column, which was then washed with 50 mL of wash I buffer (20 mM Tris-HCl, pH 8.0, 5 % glycerol, 1 mM BME, 750 mM NaCl, 30 mM imidazole) and 100 mL of wash II buffer (20 mM Tris-HCl, pH 8.0, 5 % glycerol, 1 mM BME, 350 mM NaCl, 30 mM imidazole. Protein was eluted with 20 mM Tris-HCl, pH 8.0, 5 % glycerol, 1 mM BME, 400 mM imidazole and 350 mM NaCl. Fractions containing protein were dialyzed against 50 mM Tris-HCl, pH 8.0, 10 % glycerol, 1 mM DTT, 5 mM EDTA, pH 8.0, 0.01 % NP-40 and 200 mM KCl. Protein concentration was determined through BSA standards visualized by SYPRO Orange (ThermoFisher Scientific) staining on SDS-polyacrylamide gels. Purified protein was stored at -70 ºC in 50 mM Tris-HCl, pH 8.0, 10 % glycerol, 1 mM DTT, 5 mM EDTA, pH 8.0, 0.01 % NP-40 and 200 mM KCl.

***In vitro* deamination assays**

In duplicate for each time point, 100 pmol HER1 RNA A was mixed with 20 pmol of WT or E1008Q hADAR1 enzyme in 20 µL 10 mM Tris HCl pH 7.5, 8.5 mM EDTA, 0.001% Nonidet P-40, 3% glycerol and 40.5 mM potassium glutamate. Each reaction was incubated at 37 °C for the indicated time periods, after which deaminated RNAs were immediately extracted using the Monarch RNA Cleanup Kit (New England Biolabs, Ipswitch, MA). Samples were eluted in 10 µL nuclease-free water and mixed with a 90 µL solution of 500 mM *N*-(4-ethynylphenyl)acrylamide in 50:50 EtOH:TEAA buffer. Reactions were adjusted to pH 8.6 and incubated at 70 °C for 6 hours. After labeling, samples were then ethanol precipitated and resuspended in 20 µL nuclease-free water. RNA was then CuAAC labeled using the Click-&-Go Plus Labeling Kit and 1µL of a 5 mM Cy5 picolyl azide solution in DMSO (Click Chemistry Tools, Scottsdale, AZ). Reactions were incubated for 1 hour at room temperature, after which they were ethanol precipitated and resuspended in 10 µL nuclease-free water. 20 pmol of each sample was resolved on a 10% denaturing polyacrylamide gel and imaged with a GE Amersham Typhoon RGB scanner. Densitometric quantification of bands was performed using ImageJ software. %Editing was calculated using a standard curve of labeling reactions consisting of defined mixtures of HER1 A and I RNA oligos as described earlier. For analysis of initial velocities (below), linear regression was performed using GraphPad Prism 8 software.

3.6 References:

1. Knutson, S. D.; Korn, M. M.; Johnson, R. P.; Monteleone, L. R.; Dailey, D. M.; Swenson, C. S.; Beal, P. A.; Heemstra, J. M., Chemical Profiling of A-to-I RNA Editing Using a Click-Compatible Phenylacrylamide. *Chemistry–A European Journal* **2020,** *26* (44), 9874-9878.

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

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

4. Kawahara, Y.; Zinshteyn, B.; Sethupathy, P.; Iizasa, H.; Hatzigeorgiou, A. G.; Nishikura, K., Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* **2007,** *315* (5815), 1137-1140.

5. Wahlstedt, H.; Daniel, C.; Ensterö, M.; Öhman, M., Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. *Genome research* **2009,** *19* (6), 978-986.

6. Shtrichman, R.; Germanguz, I.; Mandel, R.; Ziskind, A.; Nahor, I.; Safran, M.; Osenberg, S.; Sherf, O.; Rechavi, G.; Itskovitz-Eldor, J., Altered A-to-I RNA editing in human embryogenesis. *PLoS One* **2012,** *7* (7).

7. Hwang, T.; Park, C.-K.; Leung, A. K.; Gao, Y.; Hyde, T. M.; Kleinman, J. E.; Rajpurohit, A.; Tao, R.; Shin, J. H.; Weinberger, D. R., Dynamic regulation of RNA editing in human brain development and disease. *Nature neuroscience* **2016,** *19* (8), 1093.

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

9. Roth, S. H.; Danan-Gotthold, M.; Ben-Izhak, M.; Rechavi, G.; Cohen, C. J.; Louzoun, Y.; Levanon, E. Y., Increased RNA editing may provide a source for autoantigens in systemic lupus erythematosus. *Cell reports* **2018,** *23* (1), 50-57.

10. Vlachogiannis, N. I.; Gatsiou, A.; Silvestris, D. A.; Stamatelopoulos, K.; Tektonidou, M. G.; Gallo, A.; Sfikakis, P. P.; Stellos, K., Increased adenosine-to-inosine RNA editing in rheumatoid arthritis. *Journal of autoimmunity* **2020,** *106*, 102329.

11. Tran, S. S.; Jun, H.-I.; Bahn, J. H.; Azghadi, A.; Ramaswami, G.; Van Nostrand, E. L.; Nguyen, T. B.; Hsiao, Y.-H. E.; Lee, C.; Pratt, G. A., Widespread RNA editing dysregulation in brains from autistic individuals. *Nature neuroscience* **2019,** *22* (1), 25.

12. Maas, S.; Kawahara, Y.; Tamburro, K. M.; Nishikura, K., A-to-I RNA editing and human disease. *RNA biology* **2006,** *3* (1), 1-9.

13. Han, L.; Diao, L.; Yu, S.; Xu, X.; Li, J.; Zhang, R.; Yang, Y.; Werner, H. M. J.; Eterovic, A. K.; Yuan, Y.; Li, J.; Nair, N.; Minelli, R.; Tsang, Y. H.; Cheung, L. W. T.; Jeong, K. J.; Roszik, J.; Ju, Z.; Woodman, S. E.; Lu, Y.; Scott, K. L.; Li, J. B.; Mills, G. B.; Liang, H., The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer Cell* **2015,** *28* (4), 515-528.

14. Picardi, E.; D'Erchia, A. M.; Lo Giudice, C.; Pesole, G., REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic acids research* **2016,** *45* (D1), D750-D757.

15. Ramaswami, G.; Li, J. B., RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic acids research* **2013,** *42* (D1), D109-D113.

16. Chalk, A. M.; Taylor, S.; Heraud-Farlow, J. E.; Walkley, C. R., The majority of A-to-I RNA editing is not required for mammalian homeostasis. *Genome biology* **2019,** *20* (1), 1-14.

17. Song, Y.; Yang, W.; Fu, Q.; Wu, L.; Zhao, X.; Zhang, Y.; Zhang, R., irCLASH reveals RNA substrates recognized by human ADARs. *Nature Structural & Molecular Biology* **2020**, 1-12.

18. Wong, S. K.; Sato, S.; Lazinski, D. W., Substrate recognition by ADAR1 and ADAR2. *Rna* **2001,** *7* (6), 846-858.

19. Deffit, S. N.; Hundley, H. A., To edit or not to edit: regulation of ADAR editing specificity and efficiency. *Wiley Interdisciplinary Reviews: RNA* **2016,** *7* (1), 113-127.

20. Montiel-Gonzalez, M. F.; Quiroz, J. F. D.; Rosenthal, J. J., Current strategies for site-directed RNA editing using ADARs. *Methods* **2019,** *156*, 16-24.

21. Oakes, E.; Vadlamani, P.; Hundley, H. A., Methods for the Detection of Adenosine-to-Inosine Editing Events in Cellular RNA. In *mRNA Processing*, Springer: 2017; pp 103-127.

22. Pinto, Y.; Levanon, E. Y., Computational approaches for detection and quantification of A-to-I RNA-editing. *Methods* **2019,** *156*, 25-31.

23. Ouyang, Z.; Liu, F.; Zhao, C.; Ren, C.; An, G.; Mei, C.; Bo, X.; Shu, W., Accurate identification of RNA editing sites from primitive sequence with deep neural networks. *Scientific reports* **2018,** *8* (1), 1-12.

24. Wang, Y.; Park, S.; Beal, P. A., Selective recognition of RNA substrates by ADAR deaminase domains. *Biochemistry* **2018,** *57* (10), 1640-1651.

25. Herbert, A.; Rich, A., The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1. *Proceedings of the National Academy of Sciences* **2001,** *98* (21), 12132-12137.

26. Phelps, K. J.; Tran, K.; Eifler, T.; Erickson, A. I.; Fisher, A. J.; Beal, P. A., Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. *Nucleic acids research* **2015,** *43* (2), 1123-1132.

27. Stephens, O. M.; Yi-Brunozzi, H. Y.; Beal, P. A., Analysis of the RNA-editing reaction of ADAR2 with structural and fluorescent analogues of the GluR-B R/G editing site. *Biochemistry* **2000,** *39* (40), 12243-12251.

28. Sanger, F.; Nicklen, S.; Coulson, A. R., DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* **1977,** *74* (12), 5463-5467.

29. Burns, C. M.; Chu, H.; Rueter, S. M.; Hutchinson, L. K.; Canton, H.; Sanders-Bush, E.; Emeson, R. B., Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **1997,** *387* (6630), 303-308.

30. Mizrahi, R. A.; Shin, D.; Sinkeldam, R. W.; Phelps, K. J.; Fin, A.; Tantillo, D. J.; Tor, Y.; Beal, P. A., A Fluorescent Adenosine Analogue as a Substrate for an A-to-I RNA Editing Enzyme. *Angewandte Chemie International Edition* **2015,** *54* (30), 8713-8716.

31. Phelps, K. J.; Ibarra-Soza, J. M.; Tran, K.; Fisher, A. J.; Beal, P. A., Click modification of RNA at adenosine: structure and reactivity of 7-ethynyl-and 7-triazolyl-8-aza-7-deazaadenosine in RNA. *ACS chemical biology* **2014,** *9* (8), 1780-1787.

32. Yoshida, M.; Ukita, T., Modification of nucleosides and nucleotides: VII. Selective cyanoethylation of inosine and pseudouridine in yeast transfer ribonucleic acid. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis* **1968,** *157* (3), 455-465.

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

34. Li, Y.; Göhl, M.; Ke, K.; Vanderwal, C. D.; Spitale, R. C., Identification of Adenosine-to-Inosine RNA Editing with Acrylonitrile Reagents. *Organic letters* **2019**.

35. Knutson, S. D.; Ayele, T. M.; Heemstra, J. M., Chemical Labeling and Affinity Capture of Inosine-Containing RNAs Using Acrylamidofluorescein. *Bioconjugate chemistry* **2018,** *29* (9), 2899-2903.

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

37. Boncel, S.; Gondela, A.; Walczak, K., Michael-type addition as a convenient method for regioselective N-alkylation of ambident uracils. *Synthesis* **2010,** *2010* (10), 1573-1589.

38. Manso, J. A.; Camacho, I. F. C.; Calle, E.; Casado, J., Alkylating potential of α, β-unsaturated compounds. *Organic & biomolecular chemistry* **2011,** *9* (18), 6226-6233.

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

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

41. Kuttan, A.; Bass, B. L., Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci U S A* **2012,** *109* (48), E3295-304.

42. Uttamapinant, C.; Tangpeerachaikul, A.; Grecian, S.; Clarke, S.; Singh, U.; Slade, P.; Gee, K. R.; Ting, A. Y., Fast, cell-compatible click chemistry with copper-chelating azides for biomolecular labeling. *Angewandte Chemie International Edition* **2012,** *51* (24), 5852-5856.

Chapter 4

Selective Enrichment of A-to-I Edited Transcripts from Cellular RNA Using Endonuclease V[1*]

4.1 Abstract:

Creating accurate maps of A-to-I RNA editing activity is vital to improving our understanding of the biological role of this process and harnessing it as a signal for disease diagnosis. Current RNA sequencing techniques are susceptible to random sampling limitations due to the complexity of the transcriptome, and require large amounts of RNA material, specialized instrumentation, and high read counts to accurately interrogate A-to-I editing sites. To address these challenges, we show that *Escherichia coli* Endonuclease V (eEndoV), an inosine-cleaving enzyme, can be repurposed to bind and isolate A-to-I edited transcripts from cellular RNA. While Mg^{2+} enables eEndoV to catalyze RNA cleavage, we show that similar levels of Ca^{2+} instead promote binding of inosine without cleavage and thus enable high affinity capture of inosine in RNA. We leverage this capability to demonstrate EndoVIPER-seq (Endonuclease V inosine precipitation enrichment sequencing) as a facile and effective method to enrich A-to-I edited transcripts prior to RNA-seq, producing significant increases in the coverage and detection of identified editing sites. We envision the use of this approach as a straightforward and cost-effective strategy to improve the epitranscriptomic informational density of RNA samples, facilitating a deeper understanding of the functional roles of A-to-I editing.

[*]Adapted from Ref. 1 with permission from Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M. *Journal of the American Chemical Society*. **2020**, *142*(11), 5241–5251. Copyright 2020 American Chemical Society.

4.2 Introduction:

Adenosine-to-inosine (A-to-I) RNA editing is an abundant post-transcriptional modification found in animals. Catalyzed by adenosine deaminases acting on RNAs (ADARs), this reaction alters both the chemical structure and hydrogen bonding patterns of the nucleobase.[2] Inosines preferentially base pair with cytidine, effectively recoding these sites as guanosine. A-to-I editing is widespread across the transcriptome and present in most types of RNA. In mRNA, these sites are primarily found in repetitive and untranslated regions, affecting transcript stability, localization, and interactions with cellular pathways. mRNA editing sites can also augment transcript splicing and directly alter amino acid sequences in open reading frames.[3] Additionally, A-to-I editing modulates the target specificities and biogenesis of small-interfering RNAs (siRNAs) and microRNAs (miRNAs), in turn affecting global gene expression patterns and overall cellular behavior.[4] A-to-I editing continues to be implicated in a variety of critical biological processes including embryogenesis, stem cell differentiation, and innate cellular immunity.[3, 5] Dysfunctional A-to-I editing has also been linked with numerous disease processes such as autoimmune disorders and several types of cancer.[6-7] Recent work has also demonstrated A-to-I editing as a vital driver of human brain development and overall nervous system function, and dysregulated activity has similarly been implicated in a variety of neurological disorders including epilepsy, amyotrophic lateral sclerosis, glioblastoma, schizophrenia, autism, and Alzheimer's disease.[8-14]

Robust identification and detection of A-to-I sites is vital to understanding these broader biological roles, regulation dynamics, and relationships with disease. Because inosine is decoded as guanosine during reverse transcription, most contemporary methods utilize high-throughput RNA sequencing (RNA-seq) to identify editing sites from A-G transitions.[15] While seemingly simple, the natural complexity of cellular RNA and large dynamic ranges between individual transcripts renders RNA-seq inherently susceptible to random sampling and technical variability, making it challenging to consistently capture and detect RNA editing events, especially in light of the relative scarcity of A-to-I editing sites. Although ~5 million sites have been identified across

the transcriptome,[16-18] inosine content is low in the context of total cellular RNA, appearing in relatively few actual reads in RNA-seq datasets. This can be attributed to the fact that many key edited transcripts are expressed at low copy number. Moreover, the editing rates at individual sites can be very low or only conditionally active, and can differ significantly across cell and tissue types, individual organisms, developmental stages, and disease states.[19-21] Because of these technical challenges in RNA-seq, stringent bioinformatic analyses are also crucial for accurate detection, and extensive computational screening is needed to separate true A-to-I sites from sequencing errors, single-nucleotide polymorphisms (SNPs), somatic mutations, or spurious chemical alterations in RNA.[22]

These limitations can be overcome in part by using significant quantities of starting RNA material and/or collecting very large numbers of sequencing reads to achieve sufficient depth and coverage for accurate A-to-I calling. Alternatively, microfluidic or droplet-based PCR methods have been developed to specifically amplify regions of interest prior to RNA-seq, achieving greater sensitivity in detecting editing activity at focused A-to-I sites.[23] However, amplification-based enrichment is also significantly lower in throughput, susceptible to PCR bias, and requires both specialized instrumentation and prior knowledge of the target transcripts. While current approaches enable characterization of A-to-I editing and have yielded substantial insights into the "inosinome" in a variety of different species and tissues,[19] these methods remain impractical, expensive, and time consuming. Together, present technical limitations have made it challenging to both characterize existing A-to-I editing activity as well as increasingly difficult to discover new editing sites, restricting our overall understanding of these epitranscriptomic dynamics.

Enriching A-to-I edited transcripts prior to sequencing would largely address these challenges by depleting RNAs that otherwise lead to "wasted" sequencing reads while also helping to validate the editing sites that are observed. Despite the simplicity of this idea, effective methods to specifically target and isolate inosine in RNA have remained elusive. While a previous report detailed the generation of inosine-targeting polyclonal antibodies for isolating modified

tRNAs, these were also found to cross-react with several other nucleobases, and this research has not been reproduced.[24] We and others have also explored inosine chemical labeling strategies using acrylamide and acrylonitrile derivatives, and while these approaches are feasible for labeling and capturing inosine-containing RNAs, these reagents also irreversibly modify transcripts with adducts that inhibit reverse transcription, and inherently display off-target reactivity with pseudouridine and uridine, limiting enrichment efficiency.[25-26] Taken together, these attempts to improve A-to-I editing detection through enrichment remain limited and do not address existing technical challenges. As a result, the most widely used approach for detecting A-to-I sites remains a standard RNA-seq workflow followed by bioinformatic detection. In pursuit of alternative enrichment methods, we identified EndonucleaseV (EndoV), a conserved nucleic acid repair enzyme capable of recognizing and binding to inosine. In prokaryotes, EndoV cleaves downstream of inosine lesions resulting from oxidative damage in DNA to promote base excision repair.[27] In humans and other metazoans, EndoV has now been implicated in the metabolism of A-to-I edited RNAs.[28-29] We hypothesized that if cleavage activity could be selectively suppressed without compromising recognition and binding, then EndoV could be leveraged for enriching A-to-I edited RNAs. While human EndoV (hEndoV) appears to be a good candidate toward this goal, its biological functions and substrate preferences are still not entirely known. Recent studies have identified possible affinity toward both unedited double-stranded RNA (dsRNA) and ribosomal RNA (rRNA), properties which could be problematic for use in cellular RNA samples.[30] Interestingly, these reports also showed that *Escherichia coli* EndoV (eEndoV) was both specific and highly active toward inosine in single-stranded RNA (ssRNA) and exhibited minimal sequence bias.[28-29] These observations, as well as the commercial availability of a purified recombinant enzyme, encouraged us to explore eEndoV for the pulldown and enrichment of A-to-I edited transcripts. Herein we demonstrate EndoVIPER-seq (Endonuclease V inosine precipitation enrichment sequencing) as a novel and effective approach to bind and isolate inosine-containing

transcripts prior to RNA-seq, producing significantly improved coverage and detection of A-to-I editing sites in cellular RNA.

4.3 Results and Discussion:

Structural analyses have revealed that EndoV requires Mg^{2+} as a cofactor for inosine recognition and strand scission (**Figure 4.1b**).[31] Similar studies have shown that replacing Mg^{2+} with Ca^{2+} facilitates binding of EndoV to inosine-containing substrates without supporting catalysis.[32] Thus, we hypothesized that supplementing eEndoV with Ca^{2+} would enable enrichment of inosine-containing RNAs from cellular RNA. As an initial test of feasibility, we synthesized a pair of Cy5-labeled oligoribonucleotides having either A or I in a defined position and evaluated eEndoV activity in the presence of both cations. Consistent with previous reports, we observed not only specific cleavage activity towards inosine in ssRNA (RNA I) when benchmarked against a non-edited control (ssRNA A), but also an obligate Mg^{2+} requirement for cleavage (**Figure 4.1c**). After verifying that EndoV was unable to cleave target strands in the presence of increasing amounts of Ca^{2+} (**Figure C1**), we next evaluated the effect of Ca^{2+} supplementation on the ability of eEndoV to bind and isolate inosine-containing ssRNA. The recombinant enzyme is fused to a maltose-binding protein (MBP) tag, conveniently enabling us to implement a magnetic IP workflow using anti-MBP functionalized beads, which we term EndoVIPER (Endonuclease V inosine precipitation enrichment, **Figure 4.1e**). We used this method to attempt pulldown of both ssRNA A and ssRNA I in the presence of variable amounts of Ca^{2+}, while monitoring the initial, unbound (flowthrough), and elution fractions after washing (**Figure 4.1f**). Not surprisingly, omitting Ca^{2+} produced little binding of either oligonucleotide, supporting the idea that both recognition and cleavage of inosine is mediated through divalent cations. Increasing amounts of Ca^{2+} from 0-10 mM improved binding efficiency substantially, approaching ~80% recovery with excellent selectivity (~350-fold over pulldown of RNA A). Additional supplementation beyond 10 mM Ca^{2+} quickly decreased pulldown efficiency and

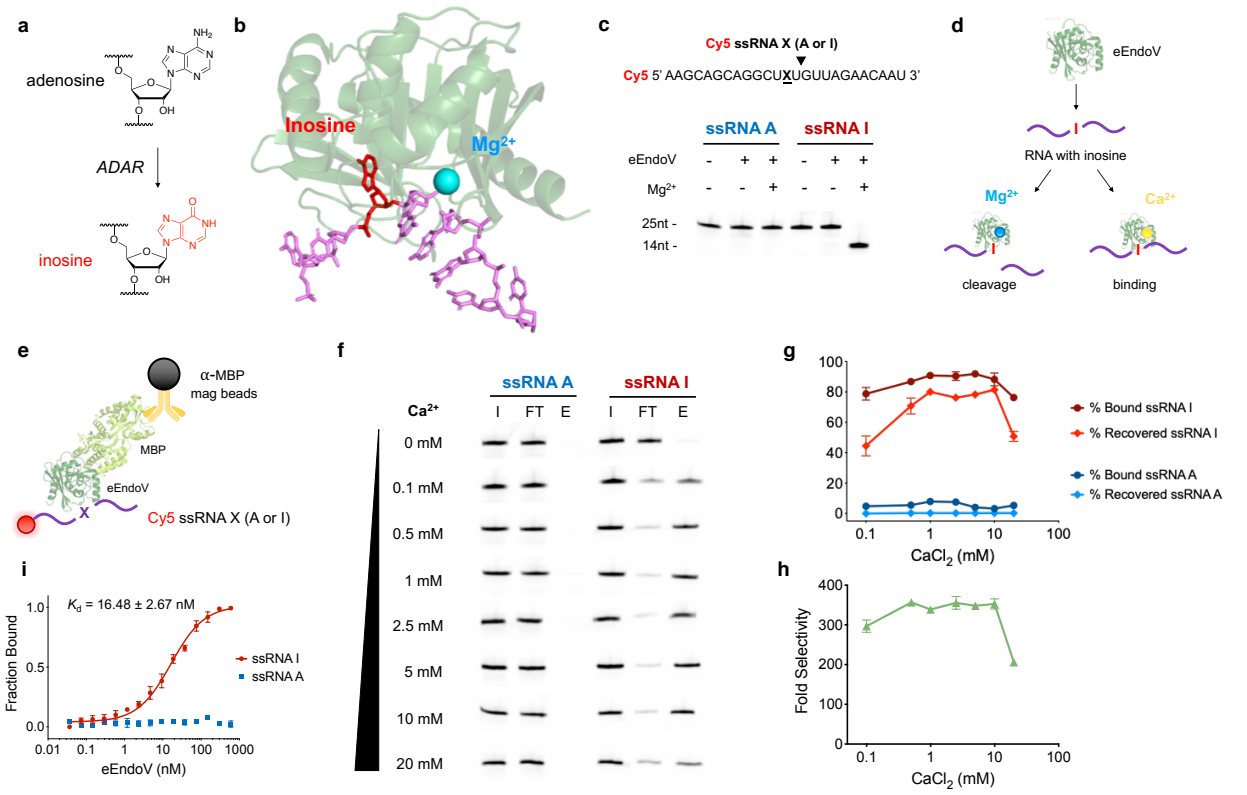


Figure 4.1. eEndoV recognizes inosine in ssRNA. Supplementation with Ca^{2+} enables high affinity binding and selective immunoprecipitation of inosine-containing ssRNAs. a) Chemical alterations of adenosine-to-inosine RNA editing catalyzed by ADAR enzymes. b) Crystal structure (PDB 2W35) of eEndoV (green) complexed with ssDNA (purple), illustrating recognition of inosine (red) in a nucleic acid substrate and Mg^{2+} (cyan) positioned adjacent to cleavage site. c) Oligoribonucleotide test sequences with putative cleavage site (arrow) and PAGE analysis of digestion reactions with eEndoV illustrating specificity toward RNA I and confirming Mg^{2+} requirement for cleavage. d) Mg^{2+} or Ca^{2+} supplementation modulates eEndoV activity towards inosine-containing RNA substrates between cleavage and binding. e) EndoVIPER schematic targeting a Cy5-labeled ssRNA using recombinant eEndoV-MBP fusion protein and anti-MBP magnetic beads. f) Representative PAGE analysis of initial (I), flowthrough (FT) and eluate (E) EndoVIPER fractions, illustrating the effects of Ca^{2+} supplementation on pulldown efficiency. g-h) Densitometric analysis of pulldown efficiency for A- and I-containing RNA. i) Quantification of eEndoV binding affinity towards ssRNA I (red) and ssRNA A (blue) using MST. Values represent mean with standard deviation, and K_d denotes mean with 95% confidence interval. (n = 3).

selectivity (**Figure 4.1g,h**), and while unconfirmed, these results likely arise from electrostatic shielding of the negative charge on the RNA phosphodiester backbone, disrupting interactions with key amino acid residues on the protein. In any case, we selected 5 mM Ca^{2+} as a suitable concentration for maximizing both recovery and selectivity. We then applied these conditions to measure the binding affinity of eEndoV for each RNA substrate using microscale thermophoresis

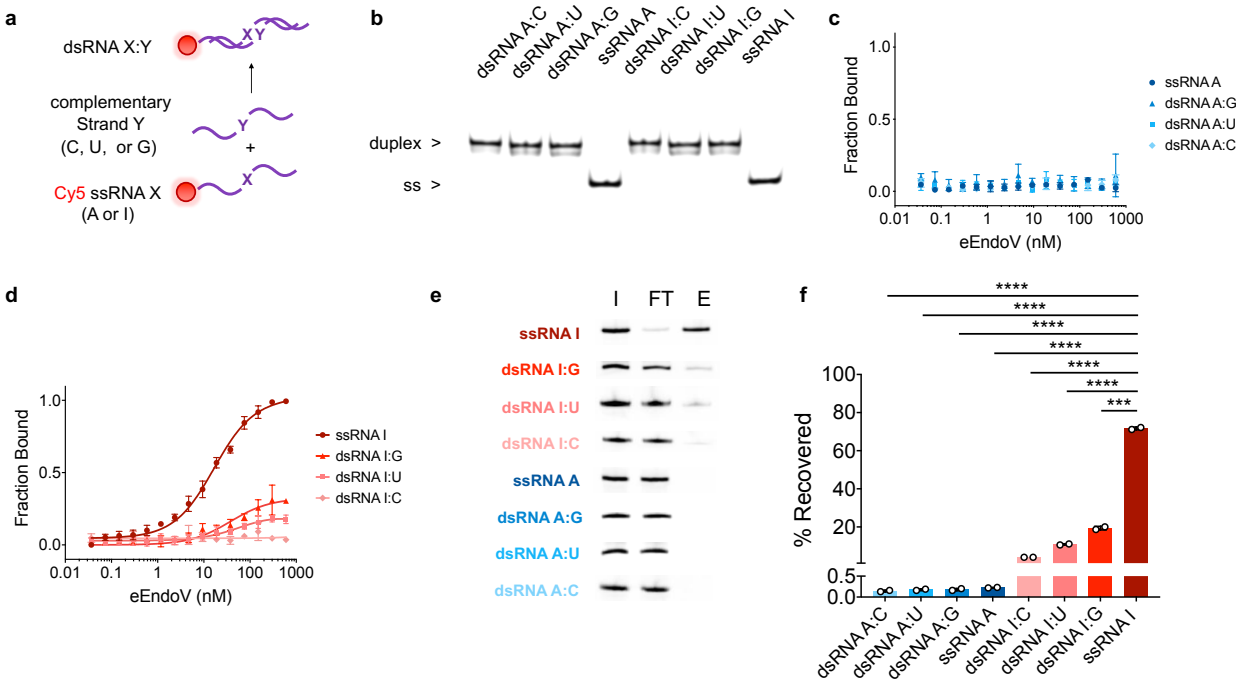


Figure 4.2. eEndoV binding favors ssRNA over dsRNA substrates. a) Schematic of dsRNA target annealing and b) duplex verification by 10% native PAGE. c) MST analysis of eEndoV binding affinity towards dsRNA A and d) dsRNA I targets using MST. Values represent mean with standard deviation. (n = 3) e) Representative PAGE analysis of initial (I), flowthrough (FT) and eluate (E) EndoVIPER fractions when tested with various dsRNA targets. f) Densitometric analysis of EndoVIPER efficiency for dsRNA targets. Values represent mean with standard deviation (n = 2). Unpaired t-test was performed for all samples against ssRNA I pulldowns (*** denotes $p = 0.0003$ and **** denotes $p < 0.0001$).

(MST) and observed low nanomolar affinity for ssRNA I and no measurable binding to the ssRNA A control (**Figure 4.1i**).

While these results were encouraging, we also recognized that ADAR primarily targets structured duplexes,[2-3] and thus the majority of inosine likely resides in the context of dsRNA. We were concerned that eEndoV may have difficulty interacting with inosine in these substrates in our native binding conditions, so we synthesized several complementary RNA strands to both ssRNA A and ssRNA I targets with differing bases opposite the A/I position. After annealing these strands together (**Figures 4.2a-b**), we assessed eEndoV affinity and EndoVIPER performance with each of the duplex constructs (**Figures. 4.2c-f**). The enzyme exhibited no detectable binding with any unedited dsRNA A substrates, yet binding affinity towards dsRNA I combinations was

highly variable and dependent on the identity of the opposing base in the complementary strand. In particular, a fully complementary duplex (dsRNA I:C) showed virtually no detectable binding by both MST and EndoVIPER (**Figures. 4.2d-f),** while mismatches ranging from I:U to I:G demonstrated increased binding in both assays. These results are also intriguing in that they are consistent with prior studies of eEndoV on DNA repair,[31] together indicating an approximate substrate preference of ssI>>> dsI:G > dsI:U > dsI:C. While interesting, these results posed a challenge to our ultimate goal of designing an unbiased approach to enriching A-to-I edited transcripts from cellular RNA, and we recognized the need to reduce or eliminate RNA secondary structure in order to mitigate the effect of these affinity biases.

Our first attempt involved reducing the ionic strength of our buffer conditions, as duplex formation is highly dependent on the presence of cations. While we initially chose 5 mM Ca^{2+} for the pulldown step, our results indicate that ~1-10 mM Ca^{2+} produce similar pulldown efficiencies (**Figure 4.1g**). These tests also employed a standard Tris-buffered saline (19 mM Tris, 137 mM NaCl, 2.7 mM KCl, pH 7.4), and we recognized that lower concentrations of monovalent cations may be tolerated. To explore these options, we assayed conditions having varying concentrations of each cation and found that removing KCl altogether and reducing $CaCl_2$ to 1 mM resulted in highly similar binding affinity and EndoVIPER performance (**Figure C2**). However, we also found that NaCl concentrations below 100 mM resulted in a significant increase in non-specific binding (**Figure C2**). Despite some promising results, both EndoVIPER and MST analyses indicated that this approach remained insufficient for opening RNA duplexes in our system, and that binding remained highly dependent on structure (**Figure C3**).

We next investigated stronger chemical methods to fully denature potential dsRNA targets. While several non-covalent denaturants, including formamide and urea, are effective in unfolding stable RNA structures, these also act on proteins, and we doubted it would be possible to denature RNA structure while maintaining native eEndoV activity. Due to these concerns, we searched for covalent methods to reversibly denature RNA prior to EndoVIPER. We required a

reagent that 1) rapidly reacts with RNA under non-degrading conditions, 2) stably maintains RNA in a single-stranded state, 3) does not interfere with eEndoV binding, and 4) can be fully removed for downstream sequencing. We were inspired by previous reports of glyoxal modification of RNA, as this reagent reacts readily with amines on the Watson-Crick-Franklin face to form stable adducts that interfere with basepairing and RNA secondary structure.[33] While glyoxal can react with A, C, and G, the N^1,N^2-dihydroxyguanosine adduct is by far the most stable (**Figure 4.3a**).[34] Importantly, glyoxal does not react with inosine, an observation that has been leveraged to study A-to-I locations through RNase T1-mediated cleavage assays.[35-36] While this appeared promising, we were uncertain if RNA glyoxalation would be compatible with eEndoV binding. To assess this, we first subjected our ssRNA I and ssRNA A oligoribonucleotides to glyoxal treatment using previously reported conditions, and observed the expected upward shift in molecular weight when analyzed via 20% PAGE (**Figure C4a**). We then analyzed binding affinity of eEndoV towards each of the treated RNAs. Surprisingly, we observed a slight improvement in affinity toward glyoxalated ssRNA I, as well as some increased non-specific response towards ssRNA A at higher concentrations of eEndoV (**Figure C5a**). We hypothesized that installation of hydrophilic groups capable of hydrogen bonding with the protein might be responsible for this non-specific activity. We also theorized that lower concentrations of eEndoV would likely confer improved specificity, so we titrated the amount of eEndoV used in the pulldown step and observed a clear optimum for both selectivity and efficiency at 100 nM enzyme (**Figures C5b-d**). Next, we repeated our full performance assay on dsRNA A and I duplex combinations. We first treated the target and complementary strands with glyoxal and unsurprisingly observed no duplex formation between glyoxalated RNAs and their complementary strands via 10% native PAGE (**Figure 4.3d**). We then tested binding affinity (**Figure 4.3e,f**) and EndoVIPER efficiency (**Figures 4.3g,h**) on the denatured RNA duplexes and observed equivalent performance across all RNA I combinations, indicating successful elimination of structural biases in eEndoV binding. While we were encouraged by these results, intermolecular duplexes are relatively easy to disrupt, so we also

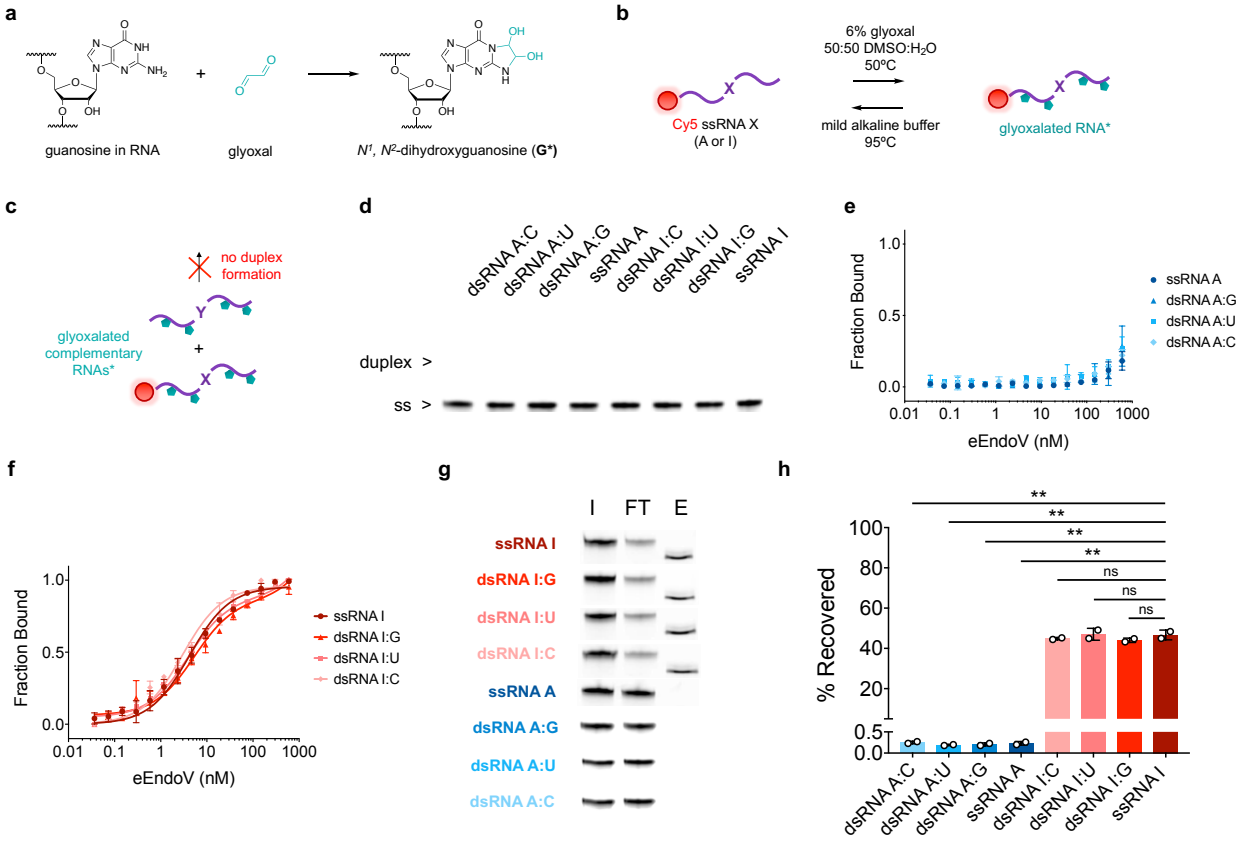


Figure 4.3. Glyoxal treatment disrupts RNA secondary structure and enables unbiased pulldown of inosine in both ssRNA and dsRNA. a) Schematic of glyoxal addition to the Watson-Crick-Franklin face on guanosine residues, forming a N^1,N^2-dihydroxyguanosine adduct. b) General reaction conditions for installation and removal of glyoxal adducts on test RNA strands. c) Disruption of dsRNA target annealing by glyoxal treatment and d) verification by 10% native PAGE. e) MST analysis of eEndoV binding affinity towards glyoxal-treated dsRNA A and f) dsRNA I targets using MST. Values represent mean with standard deviation. (n = 3) g) Representative PAGE analysis of initial (I), flowthrough (FT) and eluate (E) EndoVIPER fractions when tested with various glyoxal-treated dsRNA targets. h) Densitometric analysis of EndoVIPER efficiency for glyoxal-treated dsRNA targets. Values represent mean with standard deviation (n = 2). Unpaired t-test was performed for all samples against ssRNA I pulldowns (** denotes p = 0.0045 and "ns" indicates no significant difference).

wanted to ensure that glyoxal treatment prior to EndoVIPER was similarly robust in RNAs having a highly stable internal secondary structure. To test this, we designed a hairpin substrate representing a "worst case" RNA target due to its high melting temperature (**Figures C6a,b**). When we chemically denatured this hairpin with glyoxal, we observed almost identical EndoVIPER performance compared to previous experiments (**Figures C6d-f**). Together, these data demonstrated that we could overcome even strong secondary structure to enable pulldown with little to no effect on selectivity or enrichment of edited RNAs. However, due to the preferential

reaction of glyoxal with guanosine,[34] we were also concerned about the possibility that G bases adjacent to or near an inosine site could inhibit eEndoV binding. To address this concern, we synthesized a "G heavy" RNA strand as an additional "worst case" test substrate (**Figure C7a**), and we again observed nearly identical pulldown and binding affinity towards this substrate (**Figure C7b-d**). While there was a slight increase in overall binding affinity when measured by MST (**Figure C7d**), there was no detectable difference in pulldown performance (**Figure S7c**). Together, these experiments demonstrated that our optimized EndoVIPER protocol is robust and displays minimal bias *in vitro*, and thus we were ready to test our method in a high-throughput sequencing workflow using cellular RNA.

We selected human brain mRNA to quantify EndoVIPER-seq performance, as this tissue is known to have high A-to-I editing activity and would thus provide ample editing sites to validate our method. Additionally, nervous system tissue is a biologically interesting setting for exploring the enrichment and clinical detection of RNA editing sites crucial for neurological function or indicative of disease. To prepare for the ultimate step of high-throughput sequencing, we needed to randomly fragment our starting RNA material into smaller strand lengths. We recognized that this step would also fortuitously decrease any remaining likelihood of secondary structure formation, enhancing the resolution and performance of our pulldown. Inspired by the approaches in other RNA pulldown workflows,[37] we targeted fragment sizes of ~200-500 nt, and found that ~1 minute treatment time with Mg^{2+} at 94 ºC was sufficient to yield the desired size distribution (**Figures C8a,b**).

While glyoxal removal is well-characterized and has been used previously in both Sanger and RNA-seq applications,[35-36] we also wanted to confirm that this step was compatible with EndoVIPER performance. We first subjected fragmented mRNA to full glyoxalation and deprotection, while maintaining an identical untreated sample as a control. We then reverse transcribed cDNA for two neuronal ionotropic receptor mRNAs having known A-to-I editing sites,

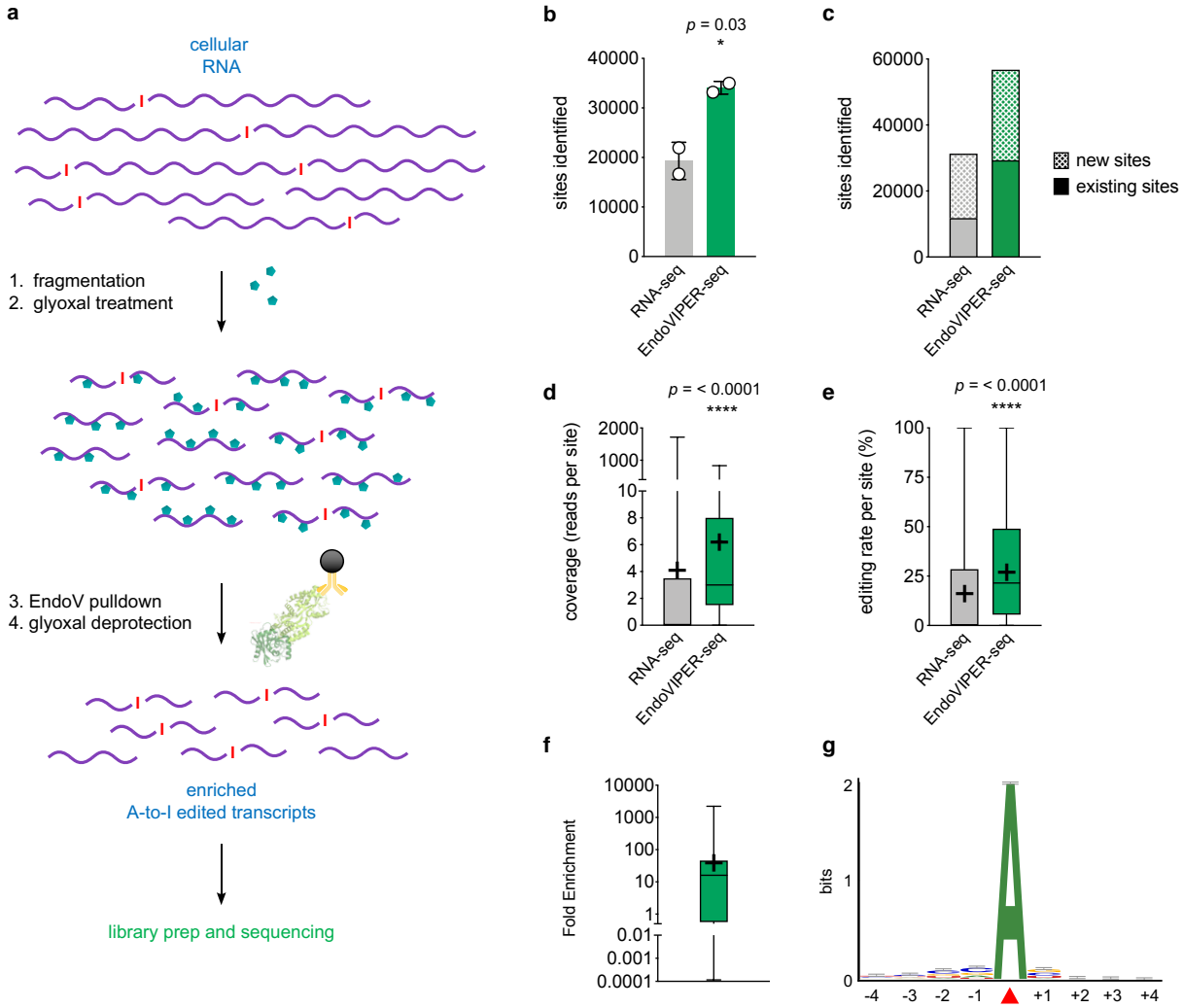


Figure 4.4. EndoVIPER-seq enables enrichment and high-throughput analysis of A-to-I RNA editing sites. a) Schematic of EndoVIPER-seq workflow. Cellular RNA is first randomly hydrolyzed into ~200-500 nt fragments, followed by glyoxal denaturation. A-to-I edited RNA is then enriched by eEndoV pulldown, followed by glyoxal removal, library preparation and high-throughput sequencing. b) Mean number of sites between duplicate RNA-seq and EndoVIPER-seq samples shows significantly increased (unpaired t-test, p = 0.03) detection of called A-to-I positions. c) Merged datasets cross-referenced against known databases show that detection of both novel and existing A-to-I sites is enhanced by EndoVIPER. Box and whisker plots show that d) read coverages and e) editing rate at all A-to-I editing sites (n = 73,578) are significantly increased by EndoVIPER (paired t-tests, p = < 0.0001). Means are denoted by black crosses. f) Box and whisker plot of calculated fold enrichment at all sites (mean, black cross = ~38-fold, n = 73,578 sites). g) Sequence motif analysis compiled from the top 200 most enriched transcripts. Red arrow denotes A/I site.

GRIA2 and *KCNA1* (**Figure C9a**).[38-39] Using real-time PCR (RT-PCR), we monitored the amplification of both transcripts and observed no kinetic difference between untreated and glyoxalated/deprotected mRNA samples, indicative of complete glyoxal removal (**Figure C9b**). We also confirmed the compatibility of glyoxalation/deprotection with Sanger sequencing, as we not only observed identical electropherogram traces between samples, but also detected the known A-to-I editing sites in both transcripts for both control and treatment samples (**Figures C10, C11**). Together with previous studies using glyoxal in sequencing workflows,[35-36] these experiments confirm that glyoxal denaturation is fully reversible and does not interfere with critical EndoVIPER library preparation steps.

With these conditions established, we next sought to directly benchmark EndoVIPER to the currently used RNA-seq methodology, and so we fragmented 2 µg of mRNA and divided this material into duplicate “RNA-seq” and “EndoVIPER-seq” groups (500 ng each), and EndoVIPER samples were subjected to the enrichment workflow (**Figure 4.4a**). After all samples underwent deprotection using heat, all samples were analyzed for size distribution and integrity, confirming that our full workflow could be completed without appreciable RNA degradation (**Figure C8c**). We then prepared libraries using ~4 ng of each respective RNA-seq and EndoVIPER-seq mRNA and proceeded to sequencing. To assess and measure A-to-I editing across samples, we employed a read aligner optimized for RNA editing (RASER[40]) as well as the specialized REDITools script package and associated filtering steps.[41] From these analyses, it was immediately apparent that the total number of identified sites was significantly higher in EndoVIPER samples (mean 34,084 sites), achieving ~1.8-fold more called A-to-I editing sites compared to RNA-seq without enrichment (mean 19,308 sites, **Figure 4.4b**). We also merged grouped data and screened these sites against the RADAR,[17] REDIPortal,[16] and DARNED[18] databases, observing a large increase in both existing and novel A-to-I locations in EndoVIPER samples (**Figure 4.4c**). Although the number of newly identified sites was larger than we expected in both sample groups (RNA-seq 19,515 novel positions out of 31,310 total called sites versus

EndoVIPER 27,429 novel positions out of 56,744 total called sites) it is worth noting that these databases catalog sites only when detected in several genome-matched donors across many RNA-seq experiments. Our experiment utilized commercially available brain mRNA (Takara Bio) isolated and pooled from a small number of donors, and larger scale verification studies are needed to further characterize these candidate sites. In any case, we applied consistent computational assessment between RNA-seq and EndoVIPER-seq samples and reliably observed a large increase in the detection of both known and novel editing sites, demonstrating the effectiveness of our method for increasing the sensitivity of detecting A-to-I editing. As a further measure to functionally validate our method, we merged and aligned all RNA-seq and EndoVIPER datasets (73,578 sites) and compared both coverage and editing rate at each detected A-to-I location. We observed a significant increase in both metrics across paired sites, indicating that EndoVIPER-seq selectively enriched A-to-I edited RNAs (**Figures 4.4d,e**). We also observed, on average, ~38-fold enrichment from read coverage values across all sites, with >75% of these sites displaying equivalent or significantly increased sequencing depth (**Figure 4.4f**).

Because of the inherent complexity of cellular RNA, we were curious whether EndoVIPER enrichment was affected by individual transcript abundance. When we plotted read coverage against enrichment scores for all sites detected in RNA-seq samples, we expectedly observed an overall decrease in fold enrichment with increased abundance (**Figure C12**). However, the overall correlation between these two variables was poor (R^2 = < 0.005), indicating that EndoVIPER is capable of enriching edited RNA transcripts across a large dynamic range of relative abundance. To ensure that eEndoV did not display a sequence context bias, we compiled the top 200 most enriched A-to-I sites and performed a sequence motif analysis. We observed no discernable consensus surrounding the editing site in highly enriched transcripts, suggesting minimal EndoVIPER sequence bias (**Figure 4.4g**). While certainly desirable towards our goal, this activity is also somewhat expected given the canonical role of eEndoV in agnostic, genome-wide surveillance and repair of oxidative lesions in DNA.[27, 31]

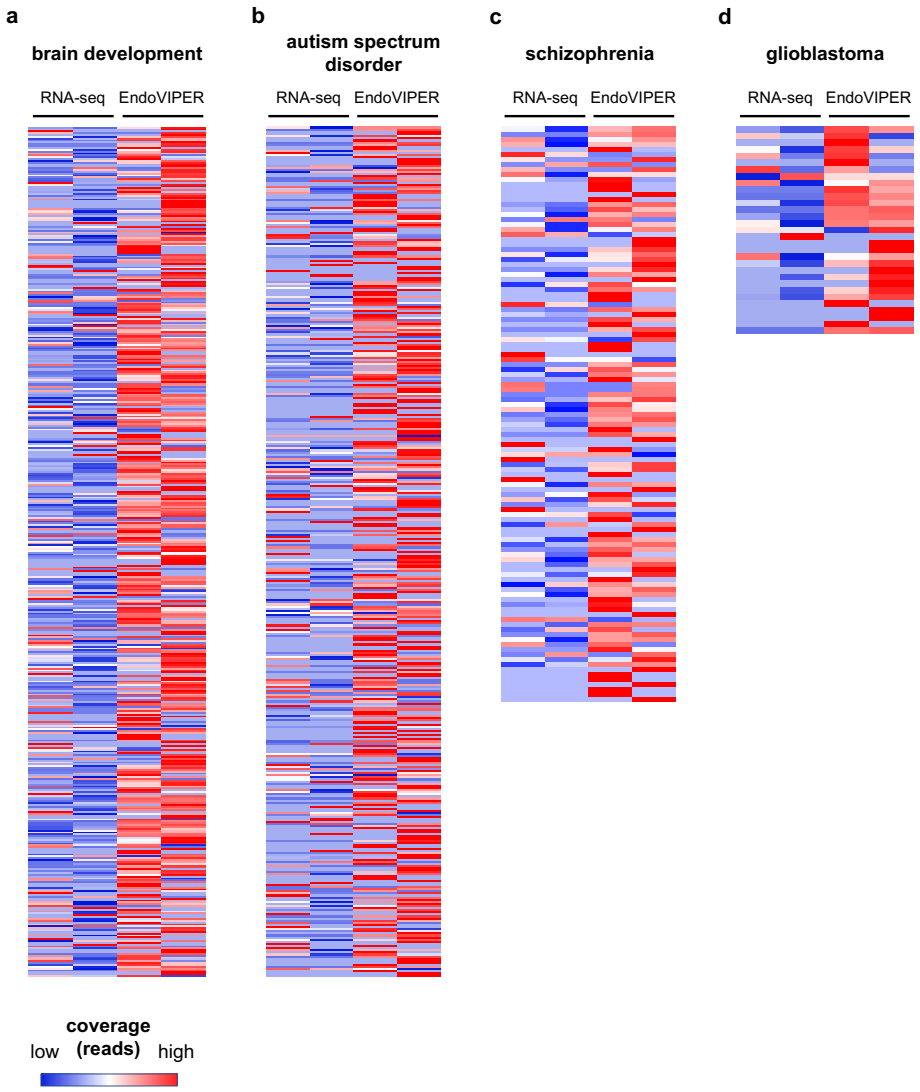


Figure 4.5. EndoVIPER-seq enhances detection of clinically relevant A-to-I editing sites. Read coverage heatmaps in upregulated RNA editing sites of interest, in a) brain development (462 sites), b) autism spectrum disorder (403 sites), c) schizophrenia (115 sites) and d) protein recoding events in glioblastoma (31 sites), demonstrating increased coverage of important editing sites in EndoVIPER treated samples. Heatmap columns display both replicate RNA-seq datasets for RNA-seq and EndoVIPER-seq samples, and each row denotes an individual editing site scaled to illustrate low (blue) and high (red) read coverage between groups.

A-to-I editing is critical for normal brain development and function, and editing activity has now been identified as a reliable, differential biomarker in a number of neurological disorders. Detection of these pathological editing events is likely to be a vital component of future RNA-based diagnostic applications, and thus we sought to employ EndoVIPER-seq for monitoring specific editing sites of interest to demonstrate its utility for improving such epitranscriptomic

characterization. In particular, we applied RNA-seq and EndoVIPER datasets toward four specific editing site panels, assessing read coverage at 462 editing sites upregulated in postnatal brain development,[9] 403 increased editing events found in autism spectrum disorder,[10] 115 sites with increased editing activity in schizophrenic patients,[13] and 31 hyperedited protein recoding events implicated in glioblastoma carcinogenesis.[7, 14] We directly compared read coverage at these sites in both RNA-seq and EndoVIPER-seq samples, and saw a consistent overall increase in total read coverage at these positions (**Figures 4.5a-d**). We also expressed these data as the number of "edited reads" containing inosine by multiplying coverage with respective calculated editing rate at each site, and this trend was expectedly similar (**Figure C13**). Together, these data indicate that EndoVIPER-seq both increased coverage at sites of interest as well as improved specific detection of pathological, edited transcript isoforms, positioning this method as a valuable tool for future clinical epitranscriptomics applications.

4.4 Conclusions:

As a scientific community, we now understand that A-to-I RNA editing is a vital post-transcriptional change affecting a variety of essential cellular pathways. Additionally, dysregulated editing underlies the molecular pathogenesis of many diseases, and is particularly important in the human nervous system. However, the true landscape and prevalence of A-to-I editing in the transcriptome remains unknown, and the total amount of A-to-I sites in the transcriptome is estimated to be significantly greater than those currently found in existing databases. Mapping and determining the precise function of individual editing sites is difficult due to current technical limitations in RNA-seq experiments, which display high sampling variability and remain challenging to specifically apply toward A-to-I editing. These studies also require specialized instrumentation for focused library preparation on a small number of sites, or use costly "brute-force" approaches that require large amounts of input RNA material and very high sequencing

depth. Together, these challenges make it increasingly difficult to access the epitranscriptomic data landscape in RNA-seq experiments.

Herein we present EndoVIPER as a new method for the affinity pulldown of inosine-containing transcripts from cellular RNA, overcoming many of the limitations outlined above and significantly improving detection and characterization of A-to-I RNA editing in complex samples. We first assess and verify that eEndoV displays high affinity and selectivity for inosine in RNA, and we chemically optimize our conditions to eliminate bias arising from different structural motifs. We then validate and demonstrate EndoVIPER-seq with brain mRNA, and show a significant increase in the ability to detect and discover A-to-I sites. In addition, we show the utility of this method for focused characterization in four panels of biologically relevant A-to-I sites, illustrating the power of our method in detecting critical RNA editing events in brain development, autism, schizophrenia, and glioblastoma. EndoVIPER is simple, straightforward, and flexible, and is easily implemented in standard library preparation workflows for RNA-seq experiments in different biological contexts. Additionally, our approach utilizes low-cost, commercially available reagents with little to no modification, enabling researchers to obtain significantly more epitranscriptomic data with smaller amounts of input RNA material. There remains opportunity for further improvement regarding EndoVIPER efficiency and selectivity, and we look forward to evaluating the performance of different EndoV orthologs, as well as exploring directed evolution strategies to further enhance binding affinity and selectivity. We also plan to extensively apply our overall strategy toward much larger scale studies of A-to-I editing across multiple individuals, tissues, and disease states, in turn providing a more detailed understanding of the overall inosine landscape in humans. To our knowledge, EndoVIPER is also the first demonstrated repurposing of an enzyme “reader” toward binding and enriching edited RNA transcripts, and our results provide strong evidence for the versatility of this strategy for isolating other epitranscriptomic or epigenetic modifications using their cognate readers. Extending this approach to other modified

nucleotides would generate a new toolbox for characterizing the corresponding transcriptional changes, and we plan to explore this in the immediate future.

Together, this report details a simple yet powerful new tool to complement existing epitranscriptomic sequencing technologies. The overall ease of use and accessibility of this method create potential for broad utility in many research disciplines. We anticipate that this will significantly improve our understanding of the dynamics and global regulation of A-to-I RNA editing across a multitude of biological contexts, further probing the potential of this epitranscriptomic mark to reveal critical information about biological function and disease progression.

4.5 Materials and Methods:

RNA Oligoribonucleotides

All oligonucleotides used in this study were custom designed and purchased from Integrated DNA Technologies. Complete sequences are shown below:

ssRNA I	5’ Cy5 AAGCAGCAGGCUIUGUUAGAACAAU 3’
ssRNA A	5’ Cy5 AAGCAGCAGGCUAUGUUAGAACAAU 3’
complementary RNA C	5’ AUUGUUCUAACACAGCCUGCUGCUU 3’
complementary RNA U	5’ AUUGUUCUAACAUAGCCUGCUGCUU 3’
complementary RNA G	5’ AUUGUUCUAACAGAGCCUGCUGCUU 3’
G ssRNA I	5’ Cy5 AAGCAGCAGGGGIGGUUAGAACAAU 3’

Hairpin RNA I (hRNA , below)

5’ Cy5 AAGCAGCAGGCUIUGUUAGAACAAUAAGUGUUCUAACACAGCCUGCUGCUU 3’

RNA Cleavage Assays

10 pmol of either ssRNA I or ssRNA A was incubated in the presence or absence of both Mg^{2+} at a 10 mM final concentration and/or 9 pmol recombinant eEndoV (New England Biolabs) in a total volume of 10 µL. Final buffer conditions in all reactions were 10 mM Tris, 125 mM NaCl, 15 µM EDTA, 150 µM DTT, 0.025% Triton X-100, 30 µg/ml BSA, 7% glycerol, pH 7.4. Reactions were incubated for 1 hour at 25 ºC, followed by a 10 min heat inactivation at 85 ºC. Reaction products were separated using 10% denaturing PAGE, and gels were imaged with a GE Amersham Typhoon RGB scanner using 635 nm excitation laser and the Cy5 670BP30 emission filter. To test cleavage in the presence of Ca^{2+}, 10 pmol of ssRNA I was incubated with 840 nM eEndoV with either 10 mM $MgCl_2$ or variable amounts of $CaCl_2$ (0, 0.1, 0.5, 1, 2.5, 5, 10 and 20 mM) in a total volume of 50 µL. Final buffer conditions in all reactions were 19 mM Tris, 137 mM NaCl, 3 mM KCl, 15 µM EDTA, 150 µM DTT, 0.025% Triton X-100, 30 µg/ml BSA, 7% glycerol, pH 7.4. Reactions were incubated at room temperature for 3 hours, after which a 3 µL sample was taken for 10 % denaturing PAGE analysis as described above.

EndoVIPER Magnetic IP Assays

For our initial binding tests (**Figure 4.1e**), 10 pmol of either RNA I or RNA A was combined with 840 nM eEndoV and variable amounts of $CaCl_2$ (0, 0.1, 0.5, 1, 2.5, 5, 10 and 20 mM) in a total volume of 50 µL. Final buffer conditions were 19 mM Tris, 137 mM NaCl, 3 mM KCl, 15 µM EDTA, 150 µM DTT, 0.025% Triton X-100, 30 µg/ml BSA, 7% glycerol, pH 7.4. Reactions were incubated at room temperature for 30 min, after which a 3 µL sample (initial, I) was taken and set aside for later analysis. Separately, 70 µL of anti-MBP magnetic bead slurry (New England Biolabs) was washed extensively with a buffer containing 19 mM Tris, 137 mM NaCl, 3 mM KCl, 7% glycerol, and variable amounts of $CaCl_2$ (0, 0.1, 0.5, 1, 2.5, 5, 10 and 20 mM), pH 7.4. After washing, beads were resuspended in eEndoV-RNA samples and incubated at 25 ºC for two hours with end-over-end rotation. Magnetic field was applied to the beads and a 3 µL sample (unbound,

UB) of the supernatant was saved for later analysis. Beads were washed extensively with respective buffer containing variable amounts of Ca^{2+}, and resuspended in 50 µL 19 mM Tris, 137 mM NaCl, 3 mM KCl, 47.5% formamide 0.01% SDS, pH 7.4 and heated to 95 ºC for 10 min. Magnetic field was applied and a 3 µL final sample (eluate, E) of the supernatant was taken of each reaction. Collected fractions were analyzed using 10% denaturing PAGE, and gels were imaged using a GE Amersham Typhoon RGB scanner. Densitometric quantification of bands was performed using ImageJ software. % Bound is expressed as a band intensity ratio of unbound versus initial fractions. % Recovered was defined as the intensity ratio of eluate versus initial fractions. Fold-selectivity was calculated as the ratio of ssRNA I versus ssRNA A recovery percentages. For experiments utilizing RNA duplexes (**Figure 4.2e**), stock constructs were first annealed as described in the later section and 10 pmol of this duplex was used for pulldown using the same protocol as outlined above. For buffer optimization experiments (**Figures C1,2**), this pulldown procedure was identical to our initial studies above while altering the components of the buffer as outlined in the figure. We will refer to these optimal formulations as 1X EndoVIPER (EV) binding buffer (19 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 15 µM EDTA, 150 µM DTT, 0.025% Triton X-100, 30 µg/ml BSA, 7% glycerol, pH 7.4.) and 1X EV wash buffer (19 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 7% glycerol, pH 7.4). To identify optimal eEndoV concentrations (**Figures C4b-d**), the pulldown procedure was performed by combining 10 pmol of glyoxalated ssRNA I or ssRNA A with 25 nM, 50 nM, 75 nM, 100 nM, 150 nM 200 nM, 400 nM, or 840 nM eEndoV in 1X EV binding buffer and bead-purified with 1X EV wash buffer as described above. Final elution was performed in 50 µL 0.5 M triethylammonium acetate (TEAA) pH 8.6, 47.5% formamide 0.01% SDS ("1X EV elution buffer") and heated to 95 ºC for 10 min, after which samples were analyzed and imaged using 10% denaturing PAGE as described earlier. For pulldown analysis of the hairpin RNA I substrate (hRNA I, **Figure C6d**), 10 pmol of glyoxalated and untreated RNA was incubated with 100 nM eEndoV in 1X EV binding buffer and purified, eluted and analyzed as described

earlier using 1X EV wash and EV elution buffers respectively. 10 pmol of “G heavy” RNA strand (G ss RNA I, **Figure C6b**), was tested in an identical manner using 1X EV buffers.

Microscale Thermophoresis (MST)

For our initial binding studies (**Figure 4.1h, 4.2c,d**), varying amounts of eEndoV were combined with 6 fmol of respective ssRNA or dsRNA targets in a final volume of 20 µL and allowed to incubate for 30 min at room temperature. Final buffer conditions in these samples were 19 mM Tris, 137 mM NaCl, 3 mM KCl, 5 mM $CaCl_2$, 15 µM EDTA, 150 µM DTT, 0.025% Triton X-100, 30 µg/ml BSA, 7% glycerol, pH 7.4. After incubating, samples were loaded into NT.115 standard glass capillaries. MST experiments were performed using a Nanotemper Monolith NT.115 Pico instrument. All measurements were analyzed using the Pico-RED filter with 12% LED intensity and 40% laser power. Data were fitted using GraphPad Prism 8 analysis software to determine K_d values. Binding tests were performed in triplicate in separate trials. For subsequent experiments (**Figures 4.3e,f, C4a, C5f, C6d**), RNAs were treated with glyoxal and purified as described below and incubated in 1X EV binding buffer and analyzed with the Nanotemper instrument using the same settings as our initial studies above.

RNA Duplex annealing

To assess duplex formation, 100 pmol of each RNA pair (untreated or glyoxalated) were mixed together in 19 mM Tris, 137 mM NaCl, 3 mM KCl, pH 7.4. Mixtures were heated to 95 ºC for 5 minutes and slowly cooled to room temperature over the course of approximately 1 hour. 10 pmol of annealed construct was then loaded onto a 10% native non-denaturing polyacrylamide gel and imaged with a GE Amersham Typhoon RGB scanner.

Glyoxal treatment and deprotection

For our initial tests of RNA glyoxalation (**Figure C4a**), 5 ug of ssRNA A or ssRNA I was added to 100 µL of 50% DMSO, 6% glyoxal (Sigma Aldrich) in nuclease-free water. Samples were reacted for 1 hour at 50 ºC and ethanol precipitated. 10 pmol of treated and purified RNA was then analyzed by 10% denaturing PAGE and imaged using a Typhoon RGB scanner. To remove glyoxal adducts (**Figure C3b**), 10 pmol of treated and purified RNA was added to 50 µL 0.5 M TEAA pH 8.6, 47.5% formamide, 0.01% SDS and heated to 95 ºC for 0, 0.5, 1, 2, 5, 10, 15, and 20 minutes. 5 µL of these reactions were directly analyzed by 20% denaturing page and imaged as described earlier.

mRNA glyoxal deprotection, Real-Time PCR, and Sanger sequencing

2 µg human brain mRNA (Takara bio, whole brain tissue pooled from 8 Caucasian males, ages: 43-66) was fragmented for 1 minute at 94 ºC using the NEBNext® Magnesium RNA Fragmentation Module (New England Biolabs) and ethanol precipitated. Purified pellet was then dissolved in nuclease-free water and quantified using a NanoDrop™ spectrophotometer (Thermo Fisher Scientific). 1 µg of fragmented mRNA was then reacted for 1 hour at 50 ºC in 100 µL of 50% DMSO, 6% glyoxal (Sigma Aldrich) in nuclease-free water, followed by ethanol precipitation. Purified pellet was then dissolved in 200 µL of 1X EV elution buffer and glyoxal was removed by heating to 95 ºC for 10 min. RNA was then purified with the Monarch® RNA Cleanup Kit and eluted in nuclease-free water. To ensure full removal of glyoxal adducts, RNA was then incubated at 65 ºC for 2 hours in 100 µL 50% DMSO in 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4, pH 7.4 followed by ethanol precipitation, resuspension in nuclease-free water, and quantification using a NanoDrop™ spectrophotometer. In separate tubes, 100 ng of untreated or deprotected mRNA fragments were combined with 20 pmol of gene specific reverse primer (*GRIA2* reverse primer 5’ CCACACACCTCCAACAATGCG 3’ and *KCNA1* reverse primer 5’ CTCGGTGGTAGAAATAGTTGAAATTGGACAC 3’) and heated to 70 ºC for 10 minutes and then

placed on ice. cDNA was then synthesized at 42 ºC for 1 hour using OneTaq® M-MuLV reverse transcriptase (New England Biolabs). 10 µL of each cDNA reaction was then mixed with 10 µL of 2X iTaq Universal SYBR Green Supermix (BioRad) and 10 pmols of gene specific forward and reverse primers (*GRIA2* forward 5' GAGAACTTGTATATGGGAAAGCTGATATTGC 3', *GRIA2* reverse 5' CCACACACCTCCAACAATGCG 3', *KCNA1* forward 5' GAATCTTCAAGCTCTCCCGCCAC 3', *KCNA1* reverse 5' CTCGGTGGTAGAAATAGTTGAAATTGGACAC 3'). PCR reactions were monitored in real-time using a LightCycler® 96 instrument (Roche) using the following thermal cycling program: 94 °C for 3 min, followed by 45 cycles of (94 °C for 15 s, 60 °C for 30 s, 68 °C for 30 s), 68 °C for 5 min. Amplification traces were analyzed using the LightCycler® software, and cycle thresholds (C_t) were determined using a default fluorescence value setting of 0.2 RFUs. PCR reactions were then purified using the Monarch PCR & DNA Cleanup Kit (New England Biolabs). 80 ng of each purified amplicon was then analyzed on a 1 % agarose gel and compared to a GeneRuler 50 bp DNA ladder (Thermo Fisher Scientific). 50 ng of each purified amplicon was also submitted for Sanger sequencing (Genscript). Sequencing traces were analyzed using SnapGene viewer.

EndoVIPER-seq

2 µg human brain mRNA (Takara bio, whole brain tissue pooled from 8 Caucasian males, ages: 43-66) was fragmented for 1 minute at 94 ºC using the NEBNext® Magnesium RNA Fragmentation Module (New England Biolabs) and ethanol precipitated. Fragmented mRNA was then reacted for 1 hour at 50 ºC in 100 µL of 50% DMSO, 6% glyoxal (Sigma Aldrich) in nuclease-free water, followed by ethanol precipitation. Purified pellet was then dissolved in nuclease-free water and quantified using a NanoDrop™ spectrophotometer (Thermo Fisher Scientific). 500 ng of fragmented and glyoxalated mRNA was then added to each of two tubes (duplicate "RNA-seq" samples) containing 30 µL nuclease-free water and frozen at -80 ºC for later use. For EndoVIPER samples, 500 ng of fragmented, glyoxalated mRNA was added to each of two tubes containing a

250 µL solution of 100 nM eEndoV and 120 units RNaisin Plus inhibitor (Promega) in 1X EV binding buffer, and was incubated at room temperature for 30 minutes. Separately, 300 µL anti-MBP magnetic bead slurry (New England Biolabs) was added to a new microfuge tube and washed extensively with 1X EV wash buffer. After washing, beads were resuspended in the eEndoV-mRNA samples and incubated at room temperature for two hours with end-over-end rotation. Magnetic field was applied and the supernatant was discarded. Beads were then washed three times with 500 µL 1X EV wash buffer and then resuspended in 200 µL of 1X EV elution buffer. Bound mRNA was then eluted by heating to 95 ºC for 10 min. Residual magnetic beads were removed from the collected supernatant using 0.22 µm microfuge spin filters (Corning® Costar®), and RNA was purified further with the Monarch® RNA Cleanup Kit and eluted in nuclease-free water. To ensure full removal of glyoxal adducts, RNA was incubated at 65 ºC for 2 hours in 100 µL 50% DMSO in 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4, pH 7.4 followed by ethanol precipitation and resuspension in nuclease-free water. Starting mRNA material, fragmented RNA-seq mRNA, and enriched EndoVIPER mRNA were quantified and assessed for size distribution (**Figure C6**) using an Agilent 2100 Bioanalyzer instrument and the Agilent 6000 RNA Pico kit. 8 ng of each RNA-seq and EndoVIPER-seq RNA replicate was then used to prepare sequencing libraries with the SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input kit (Takara Bio), standard 8-bp i5 and i7 Illumina index barcodes and adapters were added to each library. All libraries were then sequenced using a NextSeq 550 (Illumina) to produce paired end 150-bp reads.

Read trimming and mapping

Adapter and barcode sequences were removed using Trimmomatic[42] and processed with the following additional parameters (HEADCROP:3 LEADING:31 TRAILING:31 SLIDINGWINDOW:6:31, ILLUMINACLIP using a custom list of known Illumina adapter sequences. HEADCROP was used to remove the first three nucleotides of the second paired-end

sequencing read (read 2), which originate from the Pico v2 SMART template switching oligonucleotide adapter utilized in the SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input kit (Takara Bio). Trimmed reads were aligned to the human reference genome assembly GRCh37 (hg19) using reads aligner for SNPs and editing sites of RNA (RASER)[40] with previously optimized parameters $m = 0.05$ and $b = 0.03$, and uniquely mapped reads were retained. PCR duplicates were removed with the Genome Analysis Toolkit program MarkDuplicates (Picard).

Identification of RNA Editing Sites

To call editing sites, we used the REDItools[41] python package, filtering for sites with mapping quality score ≥ 10, variant call quality ≥ 20, minimum read coverage ≥ 3, minimum number of reads supporting variation ≥ 3, and minimum editing frequency ≥ 0.1. We also removed substitutions in homopolymeric regions of ≥ 5 nt, discarded any hits corresponding to intronic regions 4 nt next to known splice sites, and retained only AG and TC transitions. We also removed known common SNPs (dbSNP151) obtained from the UCSC genome browser.

Statistics

Total called editing sites were compiled for each sample, and averaged across groups. Means were graphed and significance (unpaired t-test) was calculated using Prism 8. Coverage and editing rate were compared by merging datasets from each group and aligning by editing site. Box and whisker plots and statistical significance (paired t-test) were graphed and determined in Prism. To enable calculation of fold-enrichment, all datasets were merged and aligned by site, and a pseudocount of 0.1 was added to all raw coverage values.[43] Box and whisker plots were graphed and determined in Prism. To analyze enrichment against transcript abundance, sites detected in RNA-seq samples were plotted against fold-enrichment scores and a semi-log regression analysis was calculated using Prism.

Motif analysis

The top 200 most enriched sites were compiled and neighboring sequence information (4 nucleotides up and downstream of editing site) was obtaining using the Integrative Genomics Viewer (Broad Institute).[44] Sequence logo graph representing consensus frequency was created using WebLogo (Berkeley, CA).[45]

Heatmaps

Heatmap images were generated from compiled read coverages and "edited" reads using the web-based Heatmapper tool.[46] "Edited" read counts were calculated by multiplying read coverage with editing frequency at each site.

4.6 References:

1. Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M., Selective enrichment of A-to-I edited transcripts from cellular RNA using Endonuclease V. *Journal of the American Chemical Society* **2020,** *142* (11), 5241-5251.

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

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

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

5. Ishizuka, J. J.; Manguso, R. T.; Cheruiyot, C. K.; Bi, K.; Panda, A.; Iracheta-Vellve, A.; Miller, B. C.; Du, P. P.; Yates, K. B.; Dubrot, J., Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. *Nature* **2018,** *565* (7737), 43-48.

6. Slotkin, W.; Nishikura, K., Adenosine-to-inosine RNA editing and human disease. *Genome medicine* **2013,** *5* (11), 105.

7. Han, L.; Diao, L.; Yu, S.; Xu, X.; Li, J.; Zhang, R.; Yang, Y.; Werner, H. M. J.; Eterovic, A. K.; Yuan, Y.; Li, J.; Nair, N.; Minelli, R.; Tsang, Y. H.; Cheung, L. W. T.; Jeong, K. J.; Roszik, J.; Ju, Z.; Woodman, S. E.; Lu, Y.; Scott, K. L.; Li, J. B.; Mills, G. B.; Liang, H., The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer Cell* **2015,** *28* (4), 515-528.

8. Mehler, M. F.; Mattick, J. S., Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiological reviews* **2007,** *87* (3), 799-823.

9. Hwang, T.; Park, C.-K.; Leung, A. K.; Gao, Y.; Hyde, T. M.; Kleinman, J. E.; Rajpurohit, A.; Tao, R.; Shin, J. H.; Weinberger, D. R., Dynamic regulation of RNA editing in human brain development and disease. *Nature neuroscience* **2016,** *19* (8), 1093.

10. Tran, S. S.; Jun, H.-I.; Bahn, J. H.; Azghadi, A.; Ramaswami, G.; Van Nostrand, E. L.; Nguyen, T. B.; Hsiao, Y.-H. E.; Lee, C.; Pratt, G. A., Widespread RNA editing dysregulation in brains from autistic individuals. *Nature neuroscience* **2019,** *22* (1), 25.

11. Khermesh, K.; D'Erchia, A. M.; Barak, M.; Annese, A.; Wachtel, C.; Levanon, E. Y.; Picardi, E.; Eisenberg, E., Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *Rna* **2016,** *22* (2), 290-302.

12. Li, J. B.; Church, G. M., Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. *Nature neuroscience* **2013,** *16* (11), 1518.

13. Breen, M. S.; Dobbyn, A.; Li, Q.; Roussos, P.; Hoffman, G. E.; Stahl, E.; Chess, A.; Sklar, P.; Li, J. B.; Devlin, B., Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia. *Nature neuroscience* **2019,** *22* (9), 1402-1412.

14. Silvestris, D. A.; Picardi, E.; Cesarini, V.; Fosso, B.; Mangraviti, N.; Massimi, L.; Martini, M.; Pesole, G.; Locatelli, F.; Gallo, A., Dynamic inosinome profiles reveal novel patient stratification and gender-specific differences in glioblastoma. *Genome biology* **2019,** *20* (1), 33.

15. Oakes, E.; Vadlamani, P.; Hundley, H. A., Methods for the Detection of Adenosine-to-Inosine Editing Events in Cellular RNA. In *mRNA Processing*, Springer: 2017; pp 103-127.

16. Picardi, E.; D'Erchia, A. M.; Lo Giudice, C.; Pesole, G., REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic acids research* **2016,** *45* (D1), D750-D757.

17. Ramaswami, G.; Li, J. B., RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic acids research* **2013,** *42* (D1), D109-D113.

18. Kiran, A.; Baranov, P. V., DARNED: a DAtabase of RNa EDiting in humans. *Bioinformatics* **2010,** *26* (14), 1772-1776.

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

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

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

22. Pinto, Y.; Levanon, E. Y., Computational approaches for detection and quantification of A-to-I RNA-editing. *Methods* **2018**.

23. Zhang, R.; Li, X.; Ramaswami, G.; Smith, K. S.; Turecki, G.; Montgomery, S. B.; Li, J. B., Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nature methods* **2014,** *11* (1), 51.

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

25. Knutson, S. D.; Ayele, T. M.; Heemstra, J. M., Chemical Labeling and Affinity Capture of Inosine-Containing RNAs Using Acrylamidofluorescein. *Bioconjugate chemistry* **2018,** *29* (9), 2899-2903.

26. Li, Y.; Göhl, M.; Ke, K.; Vanderwal, C. D.; Spitale, R. C., Identification of Adenosine-to-Inosine RNA Editing with Acrylonitrile Reagents. *Organic letters* **2019**.

27. Yao, M.; Hatahet, Z.; Melamede, R. J.; Kow, Y. W., Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3'endonuclease, from Escherichia coli. *Journal of Biological Chemistry* **1994,** *269* (23), 16260-16268.

28. Morita, Y.; Shibutani, T.; Nakanishi, N.; Nishikura, K.; Iwai, S.; Kuraoka, I., Human endonuclease V is a ribonuclease specific for inosine-containing RNA. *Nature communications* **2013,** *4*, 2273.

29. Vik, E. S.; Nawaz, M. S.; Andersen, P. S.; Fladeby, C.; Bjørås, M.; Dalhus, B.; Alseth, I., Endonuclease V cleaves at inosines in RNA. *Nature communications* **2013,** *4*, 2271.

30. Fladeby, C.; Vik, E. S.; Laerdahl, J. K.; Neurauter, C. G.; Heggelund, J. E.; Thorgaard, E.; Strøm-Andersen, P.; Bjørås, M.; Dalhus, B.; Alseth, I., The human homolog of Escherichia coli endonuclease V is a nucleolar protein with affinity for branched DNA structures. *PLoS One* **2012,** *7* (11), e47466.

31. Dalhus, B.; Arvai, A. S.; Rosnes, I.; Olsen, Ø. E.; Backe, P. H.; Alseth, I.; Gao, H.; Cao, W.; Tainer, J. A.; Bjørås, M., Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nature structural & molecular biology* **2009,** *16* (2), 138.

32. Hitchcock, T. M.; Gao, H.; Cao, W., Cleavage of deoxyoxanosine-containing oligodeoxyribonucleotides by bacterial endonuclease V. *Nucleic acids research* **2004,** *32* (13), 4071-4080.

33. Nakaya, K.; Takenaka, O.; Horinishi, H.; Shibata, K., Reactions of glyoxal with nucleic acids, nucleotides and their component bases. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis* **1968,** *161* (1), 23-31.

34. Aubert, M.; Bellemare, G.; Monier, R., Selective reaction of glyoxal with guanine residues in native and denatured Escherichia coli 5S RNA. *Biochimie* **1973,** *55* (2), 135-142.

35. Morse, D. P.; Bass, B. L., Detection of inosine in messenger RNA by inosine-specific cleavage. *Biochemistry* **1997,** *36* (28), 8429-8434.

36. Cattenoz, P. B.; Taft, R. J.; Westhof, E.; Mattick, J. S., Transcriptome-wide identification of A> I RNA editing sites by inosine specific cleavage. *Rna* **2013,** *19* (2), 257-270.

37. Lee, F. C.; Ule, J., Advances in CLIP technologies for studies of protein-RNA interactions. *Molecular cell* **2018,** *69* (3), 354-369.

38. Wright, A. L.; Vissel, B., The essential role of AMPA receptor GluR2 subunit RNA editing in the normal and diseased brain. *Frontiers in molecular neuroscience* **2012,** *5*, 34.

39. Levanon, E. Y.; Hallegger, M.; Kinar, Y.; Shemesh, R.; Djinovic-Carugo, K.; Rechavi, G.; Jantsch, M. F.; Eisenberg, E., Evolutionarily conserved human targets of adenosine to inosine RNA editing. *Nucleic acids research* **2005,** *33* (4), 1162-1168.

40. Ahn, J.; Xiao, X., RASER: reads aligner for SNPs and editing sites of RNA. *Bioinformatics* **2015,** *31* (24), 3906-3913.

41. Picardi, E.; Pesole, G., REDItools: high-throughput RNA editing detection made easy. *Bioinformatics* **2013,** *29* (14), 1813-1814.

42. Bolger, A. M.; Lohse, M.; Usadel, B., Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014,** *30* (15), 2114-2120.

43. Zhang, C.; Zhang, B.; Lin, L.-L.; Zhao, S., Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC genomics* **2017,** *18* (1), 583.

44. Robinson, J. T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E. S.; Getz, G.; Mesirov, J. P., Integrative genomics viewer. *Nature biotechnology* **2011,** *29* (1), 24.

45. Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome research* **2004,** *14* (6), 1188-1190.

46. Babicki, S.; Arndt, D.; Marcu, A.; Liang, Y.; Grant, J. R.; Maciejewski, A.; Wishart, D. S., Heatmapper: web-enabled heat mapping for all. *Nucleic acids research* **2016,** *44* (W1), W147-W153.

Chapter 5

Direct Immunodetection of Global A-to-I RNA Editing Activity with a Chemiluminescent Bioassay*

5.1 Abstract:

Adenosine-to-inosine (A-to-I) RNA editing patterns can vary significantly between different cell and tissue types, and hyperactive A-to-I signatures are indicative of several diseases, including cancer and autoimmune disorders. Because of the importance of these differences, there is significant need for efficient methods to measure overall editing levels in cellular RNA. Current standard approaches rely on RNA-seq to indirectly detect editing sites, which requires significant investments in time and material as well as extensive computational analysis. Here, we utilize Endonuclease V (EndoV), which binds specifically to inosine in RNA, to develop a protein-based chemiluminescent bioassay to directly profile A-to-I RNA editing activity. We previously showed that EndoV can bind and enrich A-to-I edited transcripts prior to RNA-seq, and we now leverage this activity to construct an EndoV-linked immunosorbency assay (EndoVLISA) as a rapid, plate-based chemiluminescent method for measuring global A-to-I editing signatures in cellular RNA. We first optimize and validate our assay, illustrating selective and sensitive detection of inosine in RNA. We then demonstrate rapid detection of inosine content in treated cell lines, demonstrating equivalent performance against RNA-seq approaches. Lastly, we deploy our EndoVLISA for profiling differential A-to-I RNA editing signatures in normal and diseased human tissue, illustrating the utility of our platform as a diagnostic bioassay. Together, the EndoVLISA method is cost-effective, straightforward, and utilizes common laboratory equipment, offering a highly accessible new approach for studying A-to-I editing. Moreover, the multi-well plate format makes this the first assay amenable for direct high-throughput quantification of A-to-I editing for applications in disease detection and drug development.

*Authors: Knutson, S. D.; Arthur, R. A.; Johnston, H.R.; Heemstra, J. M. *Submitted,* 2021.

5.2 Introduction:

Adenosine-to-inosine (A-to-I) RNA editing is catalyzed by adenosine deaminases acting on RNA (ADARs), and is a critical and widespread RNA modification in eukaryotes.[1] Deamination changes the structure and hydrogen bonding pattern of the nucleobase, and resulting inosines instead hybridize with cytosine and are decoded as guanine by cellular machinery. Due to this change, editing sites within protein-coding regions can directly alter amino acid sequences and produce different protein isoforms.[2] Small regulatory RNAs are also edited, in turn modulating the activities of small-interfering RNAs (siRNAs) and microRNAs (miRNAs) and affecting global gene expression patterns.[3]

However, the vast majority of A-to-I editing sites are found within repetitive *Alu* elements, which are embedded throughout the human transcriptome and form long (~300 bp) inverted dsRNA repeats that are recognized and edited by ADAR1.[4] Because *Alu* elements are widespread, millions of A-to-I sites have been identified,[4-8] and editing within these regions is now recognized as an essential cellular mechanism to regulate innate immune system activation and differentiate “host” RNA from pathogenic transcripts.[9] In the absence of ADAR1, unmodified dsRNA accumulates within the cell and activates downstream cytotoxic interferon responses.[10-12] Genetic knockout of ADAR1 is also typically a lethal phenotype in mice, causing developmental defects arising from uncontrolled innate immune responses in specific tissues.[10, 13-15] In humans, dysregulated A-to-I editing is also strongly linked with autoimmune disorders,[16-18] and most cancer types display ADAR1 overexpression and hyperediting signatures as a potential means of suppressing and evading the immune system.[19-20] Interestingly, deletion of ADAR1 has been shown to overcome immune checkpoint blockade resistance in tumors,[21] and there is now significant interest in developing treatment modalities to specifically inhibit ADAR enzymes and decrease global editing activity. Beyond these critical roles in immune system function and embryogenesis, A-to-I editing is also broadly implicated in stem cell differentiation and neurological activity,[2, 9, 22] and editing malfunctions in the central nervous system are implicated

in epilepsy, amyotrophic lateral sclerosis, glioblastoma, schizophrenia, autism, and Alzheimer's disease.[23-30]

Although a critically important and dynamic cellular process, our overall understanding of A-to-I editing regulation and its broader functional relationships remain limited. Similarly, despite its clear relationship with several classes of disease,[16-17, 20, 27, 31-35] both ADAR activity and global A-to-I editing signatures have not been fully leveraged as a biomarker for disease diagnosis. Both of these limitations are the direct result of technical challenges associated with current methods for detecting and measuring editing activity. One of the earliest attempts at quantifying inosine content employed enzymatic digestion of isolated mRNA followed by P^{32} radiolabeling and successive thin-layer chromatography.[36] While this study provided an initial rough estimate of inosine content in rat mRNA (~0.01 – <0.001% of all nucleotides) and suggested the first tissue-level differences in editing activity, this method is laborious, hazardous, and suffers from low precision, and thus has not been applied or adopted beyond these initial experiments. Because inosine is decoded as guanine during reverse transcription, the current standard method utilizes high-throughput RNA sequencing (RNA-seq) to identify editing sites from A-G transitions.[37] This approach offers high-resolution mapping of A-to-I sites throughout the transcriptome, and when focused solely on *Alu* elements, enables computational estimation of both ADAR1 activity and global editing levels in cellular RNA.[38] While these bioinformatic methods are effective and additionally allow post-hoc measurement of A-to-I activity from published RNA-seq experiments, generating new datasets still requires high-cost consumables, specialized instrumentation, and data turnaround times can often extend past weeks or even months. RNA-seq is also limited to capturing small "windows" of RNA, making this technique inconsistent and prone to random sampling errors. Editing sites are also quite rare in the context of total RNA[4, 36] and so RNA-seq is less suited for comprehensive measurement of editing levels from cellular samples. Because of these challenges, RNA-seq requires excessive amounts of RNA input material, very high numbers of sequencing reads, and customized computational pipelines to achieve sufficient depth

and coverage for accurate indexing of global A-to-I levels.[38-40] Modified bases can also be quantified by digesting cellular RNA and analyzing by LC/MS,[41-44] and while capable of measuring inosine content in biological samples,[45-46] these methods are also low in throughput, often require impractically high amounts of RNA input, and utilize costly and highly-specialized instrumentation. Additionally, LC/MS typically lacks the resolution needed to detect low-frequency modifications or discriminate between structurally similar nucleotides. A bioluminescent system was also recently developed, in which A-to-I editing of a UAG stop codon embedded in a luciferase mRNA reporter produces a measurable signal for inferring ADAR activity levels in immortalized cells.[47] While this method is high-throughput-compatible and enabled the first screening of a small molecule library for potential ADAR-modulating drugs, this approach relies on indirect estimation of A-to-I activity at a single putative editing site and does not directly assay actual inosine content in cellular RNA. Additionally, this approach requires genetic manipulation for establishing a stable reporter cell line, and thus cannot be extended beyond *in vitro* contexts or used as a diagnostic tool in primary cells or tissues.

Despite these advances, RNA-seq is still by far the most widely adopted and current method-of-choice for detecting A-to-I editing, and there remains a significant need for technology that can rapidly and inexpensively profile global editing signatures in a variety of biological contexts. Such an advance would not only provide faster and more accurate information regarding the biochemical regulation of global A-to-I editing activity in humans, but would also enable researchers to leverage differential A-to-I signatures as a diagnostic biomarker in several disease contexts. Additionally, an assay platform capable of detecting global changes in editing levels would be an enabling advance for designing and testing pharmacological inhibitors of ADAR activity in both immortalized cells and primary tissues.

Toward these goals, we were inspired by our previous explorations of Endonuclease V (EndoV), a conserved nucleic acid repair enzyme that can specifically recognize inosine in nucleic acids.[48-52] EndoV is present across all domains of life,[53] and naturally utilizes Mg^{2+} to cleave

inosine-containing substrates. In lower prokaryotes, this activity appears to have originally evolved for repair of inosine lesions in DNA,[50] while in humans and other higher eukaryotes, EndoV exhibits specific activity toward inosine in RNA and is speculated to have metabolic roles in degrading A-to-I edited transcripts.[51-52] Interestingly, enzyme activity can also be modulated by replacing Mg^{2+} with Ca^{2+}, enabling EndoV to bind instead of cleave inosine-containing nucleic acid substrates.[49] Inspired by this, we recently demonstrated that recombinant *E. coli* EndoV (eEndoV) has high affinity and selectivity for A-to-I edited transcripts and can be repurposed to act as an "antibody" to bind and enrich edited transcripts prior to RNA-seq.[54-55]

Here, we leverage these properties to construct an EndoV-linked immunosorbency assay (EndoVLISA) to directly measure global A-to-I RNA editing signatures using a simple chemiluminescent plate-based bioassay. We first design an assay workflow to denature, biotinylate, and immobilize RNA into streptavidin-coated plates for subsequent immunodetection and chemiluminescent measurement. We then systematically optimize key parameters in our assay, validate its accuracy and sensitivity for detecting inosine in RNA, and further benchmark its performance against LC/MS and RNA-seq methods. We then demonstrate EndoVLISA as a facile bioassay for rapidly profiling A-to-I RNA editing in several human cell and tissue samples, addressing a significant gap in RNA-based analysis and improving our ability to profile global epitranscriptomic changes in a variety of biological contexts.

5.3 Results and Discussion:

In our initial demonstration using eEndoV for capturing A-to-I edited RNAs,[54] we first employed microscale thermophoresis (MST) to validate and quantify binding affinity of the protein for inosine in RNA. MST is a well-characterized and highly sensitive method that monitors solvation and mobility changes in fluorescently-tagged molecules upon ligand or receptor binding,[56-57] and we hypothesized that MST detection of EndoV binding events with fluorescently-tagged RNAs would be a feasible approach for measuring global A-to-I editing activity (**Figure**

D1a,b). We also observed that RNA secondary structure was a key factor for EndoV pulldown efficiency in edited RNAs, and found that the enzyme had a strong preference for binding inosine in unstructured, single-stranded RNA (ssRNA).[54] This was especially problematic because ADAR primarily targets structured duplexes for A-to-I editing,[1, 58] and thus inosines are highly likely to reside in double-stranded RNAs (dsRNA). To circumvent this problem, we identified glyoxal, a covalent denaturant that reacts with the Watson-Crick-Franklin face of guanosine, adenosine, and cytidine, as an effective means for disrupting RNA secondary structure (**Figure D2a**).[59] Importantly, glyoxal does not react with inosine, and we found that this treatment step was not only compatible with EndoV-RNA pulldown, but greatly enhanced the ability of EndoV to bind A-to-I edited transcripts regardless of structure (**Figure D2b**).[54]

Combining glyoxal treatment with MST analysis, we envisioned a method for rapid global editing measurement. In this approach, cellular RNA is first fragmented into smaller strands and then each molecule labeled with a Cyanine5 (Cy5) fluorophore for subsequent EndoV binding and MST measurement (**Figure D3a**). We utilized sodium metaperiodate ($NaIO_4$) to end label each strand (**Figure D3b,c**)[60] then tested the protocol on short (24 nt) RNA oligonucleotides containing A or I in a defined position (RNA A and I), and confirmed complete fluorescent labeling of glyoxal denatured strands (**Figure D3d**). With these materials in hand, we next sought to maximize sensitivity for detecting inosine in our MST workflow. Because inosine levels are typically very low in mRNA (estimated ~0.01 – <0.001% of nucleotides),[4, 36] we first identified 10 ng as the maximum amount of Cy5-labeled RNA that could be used in MST without saturating the detector (**Figure D4**). Next, we incubated 10 ng of each RNA A or I with increasing amounts of EndoV and assessed MST response (change in normalized fluorescence, ΔF_{norm}). Consistent with our previous explorations,[54] we observed high binding affinity toward the inosine-containing strand (RNA I K_d 3.39 ± 0.36 nM) and optimal EndoV selectivity for inosine (~70-fold) at ~100 nM (**Figure D5**). Additionally, off-target interactions between RNA A and EndoV were only observed at very high concentrations (> 1 μM). While promising, actual inosine content in these experiments

was still quite high (~4 %), and we unfortunately found that the signal became undetectable below ~0.5-1% inosine content (**Figure D6**). Unfortunately, this limit of detection is well above predicted A-to-I levels in cellular RNA,[4, 36] and so the use of MST for measuring global editing did not appear feasible.

In redesigning a biosensing platform for measuring global changes in A-to-I RNA editing, we were again inspired by the "antibody-like" qualities of EndoV and hypothesized that these could be applied in other established immunodetection approaches. In particular, we identified enzyme-linked immunosorbency assay (ELISA) as a potential platform due to its well-characterized performance, high sensitivity, and overall versatility for quantifying low abundance analytes in highly complex mixtures.[61-62] Because of these attractive qualities, ELISA is now a routine technique for both basic research and biomedical diagnostics, and has enabled sensitive detection of human antibody titers,[63] cytokine and small-molecule analytes in serum,[64] and viral particles in different biological fluids.[65-66] Moreover, primary antibodies identified for other rare nucleic acid modifications, including N^6-methyladenosine (m^6A),[67-68] 5-methylcytosine (5mC),[69] 5-hydroxymethylcytosine (5hmC), and 5-carboxylcytosine (5caC),[70] have now also been successfully deployed in commercial ELISA formats for sensitively detecting global editing rates in RNA and DNA. ELISA also utilizes cheap consumables and commonplace laboratory equipment, and these collective advantages uniquely position this method for potentially addressing several major technical limitations in rapidly measuring global A-to-I editing rates.

To apply this assay format for quantifying inosine, we envisioned that our original workflow could be slightly modified so that RNA would first be biotinylated and immobilized into a streptavidin-coated well of a micro-well plate to enable the binding, washing, and detection steps that are typical of other ELISA approaches (**Figure 5.1c, d**). Given our prior success in end-labeling RNA strands with Cy5 (**Figure D3**), we again utilized periodate oxidation and hydrazide coupling as a chemical strategy for 3' biotinylation of RNA transcripts (**Figure 5.1d**). Additionally, we reasoned that end-labeling and immobilizing each RNA strand from the 3' terminus would in

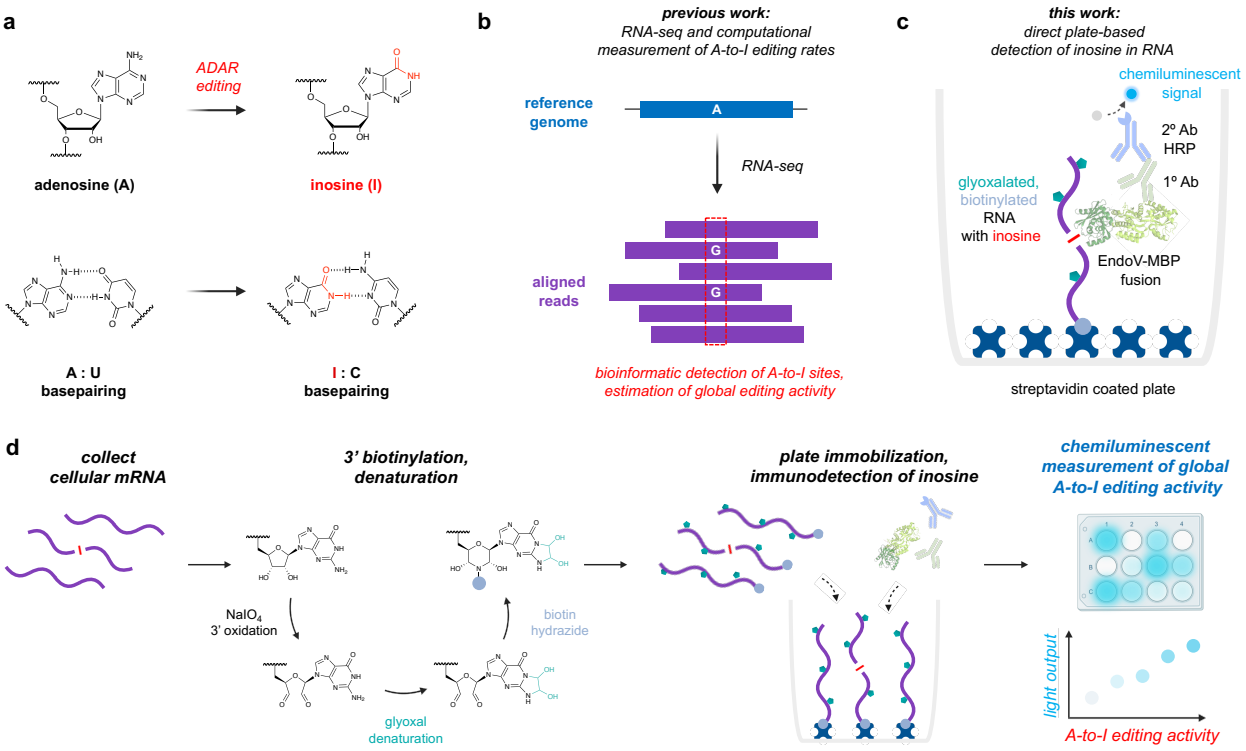


Figure 5.1. Designing a protein-based chemiluminescent bioassay for direct detection of inosine in RNA. a) Adenosines in RNA are converted to inosine by ADAR enzymes, and introduce a base pairing change. b) Detecting A-to-I sites and measuring editing activity is typically achieved with next-generation RNA sequencing (RNA-seq), where inosines are identified as A-G transitions between RNA-seq reads and a reference genome. c) In this study, we propose direct detection of A-to-I sites with an Endonuclease V – Linked Immunosorbency Assay (EndoVLISA). Cellular RNA is first glyoxal denatured, biotinylated, and immobilized in a streptavidin-coated well. Inosine is then specifically recognized by EndoV, which is fused to a maltose-binding protein (MBP) affinity tag. Wells are then probed with a mouse anti-MBP primary (1º) antibody and a goat anti-mouse secondary (2º) antibody conjugated to horse-radish peroxidase (HRP) to generate a chemiluminescent signal. d) Chemical strategy to prepare RNA for analysis. Sodium meta-periodate ($NaIO_4$) is first used to oxidize terminal 3' OH groups in each RNA strand, followed by covalent denaturation with glyoxal. Biotin-PEG_4-hydrazide is then reacted with 3' dialdehyde groups to enable immobilization of RNA mole-cules onto a streptavidin-coated surface.

theory maximize our potential sensitivity in detecting inosine by enabling multiple EndoV binding events per transcript while only occupying one streptavidin site on the two-dimensional surface of the well. As an initial feasibility test, we first sequentially treated our test RNA I strand with $NaIO_4$, glyoxal, and biotin-PEG_3-hydrazide. We then also fluorescently tagged this RNA with Cy5 and loaded increasing amounts into each well of a streptavidin-coated 96-well plate. After extensively washing each well, we then measured fluorescent signal across the plate. As shown in **Figure**

D7, signal increased proportionally from 0 to 5 pmol of the treated RNA and plateaued at higher loading amounts, suggesting a maximum capacity of ~5 pmol biotinylated RNA. Considering the median length of human mRNA (~1400 nt),[71] this capacity would conveniently enable us to measure inosine content in up to ~2-3 µg mRNA per well. Additionally, we observed virtually no binding in an RNA control strand that was not biotinylated, indicating that non-specific interactions in this step were minimal (**Figure D7**).

One of the key reasons that ELISA offers such high sensitivity is that sequential binding of the analyte of interest to both a primary (1º) and secondary (2º) antibody enables signal amplification. This is then significantly enhanced by conjugating an enzyme to the 2º antibody, such as horse-radish peroxidase (HRP), which provides catalytic generation of multiple detection signals for every initial analyte molecule bound. While powerful, this design also presents significant assay complexity, and ELISA platforms are known for requiring extensive optimization to balance detection sensitivity with background signal. Our proposed EndoVLISA system shares this potential complexity, wherein we first bind inosine in RNA with a recombinant eEndoV fused to a maltose-binding protein (MBP) affinity tag, followed by probing with a mouse anti-MBP 1º antibody and an HRP-conjugated, goat-anti-mouse 2º antibody (**Figure 5.2a**).

Each of these individual assay components are key to robustly measuring global changes in A-to-I editing, so we first sought to systematically optimize our workflow to maximize sensitivity and minimize off-target interactions. Toward this end, we also identified mRNA as an analyte of choice, as the vast majority of ADAR1 editing events occur within repetitive *Alu* elements in these transcripts and would likely improve our potential detection efficiency.[4-8] Isolating polyA+ transcripts prior to EndoVLISA detection would also eliminate RNA species that might interfere with our assay, including several transfer RNA (tRNA) species that contain inosine in the anticodon loop.[72-74] While this modification is essential for protein translation, tRNA editing is performed by a different enzyme family (ADAT) and is unrelated to A-to-I editing activity catalyzed

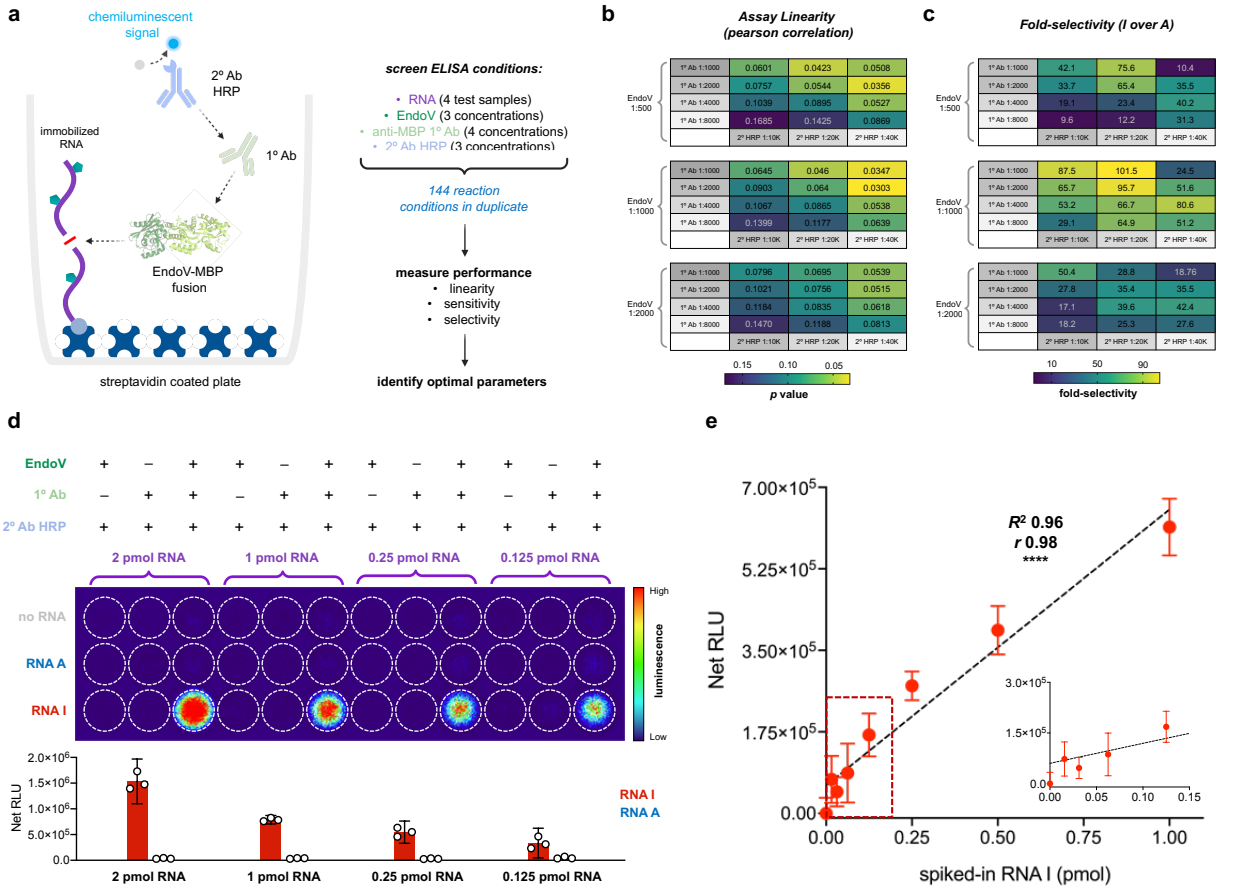


Figure 5.2. Optimizing EndoVLISA performance for sensitive and selective detection of A-to-I RNA editing. a) Schematic of initial EndoVLISA optimization and main functional components – (1) RNA of interest, (2) eEndoV-MBP fusion protein, (3) anti-MBP primary antibody, and (4) goat anti-mouse HRP-conjugated secondary antibody, and (5) chemiluminescent substrate. Several dilutions of each component 2-4 were tested in a large screen for optimal conditions. Assay linearity and selectivity were chosen as critical evaluation metrics. b,c) In duplicate, various combinations of different EndoV and antibody concentrations were tested with on- and off-target RNAs (RNA I and A), and both linearity and selectivity were measured. b) Pearson *r* p-value was computed in GraphPad Prism using on-target RNA I signals from 0, 0.1 and 2 pmol RNA signals. c) Fold-selectivity was calculated as the chemiluminescent signal of RNA I divided by RNA A (2 pmol/well). d) Representative image of chemiluminescent EndoVLISA detection of inosine. RNA was immobilized into a 96-well streptavidin coated plate, followed by immunodetection with the indicated reagent combinations and incubated with chemiluminescent substrate for 1 minute with shaking. Reactions were then transferred to a clear 96-well plate and chemiluminescent signal was captured using a Typhoon biomolecu-lar imager. White outlines indicate relative position of each well. Heat map values represent luminescent intensity generated using the acquisition software. Bar graph denotes signals generated from different RNA A and RNA I loading amounts (2, 1, 0.25, and 0.125 pmol per well). e) Linearity and sensitivity of detecting inosine in complex mixtures. Decreasing amounts of RNA I strand were spiked into 1 µg of in vitro transcribed mRNA and detected using optimized EndoVLISA workflow (EndoV 1:1000, 1º 1:1000, 2º HRP 1:40,000, and SuperSignal™ West Atto Substrate). Inset displays a zoomed-in portion of the curve (red-dashed box, < 150 fmol RNA I) illustrating a lower limit of detection ~62.5 fmol of inosine per µg RNA. Values represent mean (n = 3) and error bars denote 95% confidence intervals. Linear regression (black dashed line), R2 and pearson (r) correlation (**** denotes p < 0.0001) were computed in Prism.

by ADAR1. Using the rough estimate of rat inosine content (~50-200 fmol/μg mRNA)[36] as well as RNA-seq analysis showing that global human editing rates are ~5-fold higher,[5] we wanted to first optimize our EndoVLISA approach to target this range (~0.1-2 pmol/μg mRNA). We also designed a combinatorial screen to systematically test different conditions for each component (EndoV-MBP, 1º and 2º Ab-HRP) so that each EndoVLISA mixture could be assayed with 1) no RNA (blank), 2) "low on-target" (100 fmol/well RNA I), 3) "high on-target" (2 pmol/well RNA I) and 4) "high off-target" (2 pmol/well RNA A) (**Figure 5.2a**). Proportional on-target signal is a key metric for robust ELISA performance, and so we stringently measured assay linearity from 0, 0.1, and 2 pmol RNA I using Pearson correlation coefficients (*r*) and associated *p* values for all EndoVLISA combinations. Interestingly, while EndoV concentration had minimal effect, a clear trend emerged in that higher amounts of 1º MBP-targeting antibody (1:1000 and 1:2000 dilutions) and lower amounts of 2º antibody-HRP (1:20,000 and 1:40,000) yielded excellent linearity ($p < 0.05$, **Figure 5.2b**). Conversely, lower 1º and higher 2º antibody concentrations were less linear and displayed both higher variability and greater saturation effects in on-target binding curves (**Figures D8-D10**). This can be explained in part by the fact that we employed a monoclonal 1º antibody which binds to MBP in a 1:1 ratio, whereas multiple 2º HRP (polyclonal) antibodies can subsequently attach to this complex. Thus, coupled with catalytic chemiluminescent signal generation from HRP, certain antibody combinations (low 1º, high 2º) may result in disproportionate signal generation that does not accurately reflect inosine content per well.

Off-target background signal is also a major consideration with ELISA development, so we evaluated assay performance with high amounts of both RNA A and RNA I (2 pmol/well). While the resulting trend was less clear than in our linearity tests, better overall selectivity (I/A) was again observed with higher 1º antibody concentrations (1:1000 and 1:2000, **Figures 5.2c**, **D11-13**). Interestingly, EndoV concentration also had a more significant impact across these experiments, and there was a clear optimum in EndoVLISA performance when using 2º HRP at a dilution of 1:20,000. From this large screen, we ultimately identified several EndoVLISA reagent

combinations that produced comparable performance and excellent robustness (**Table D1**), suggesting overall assay flexibility in a variety of potential conditions and indicating that minor adjustments in these components would not significantly alter assay quality.

Using optimized conditions from this screen, (EndoV 1:1000, 1º antibody 1:1000, and 2º antibody-HRP 1:20,000), we next confirmed that our method was specific and that each assay component was necessary and sufficient for signal generation. As shown in **Figure 5.2d**, signal was only observed in wells receiving all 3 detection components. Importantly, no response was generated without EndoV, indicating very low non-specific binding of either the 1º or 2º HRP antibody. We also began testing a larger range of inosine concentrations in RNA, and we found that our EndoVLISA assay response was not only highly linear (particularly below 1 pmol) but also sensitive, with an estimated lower limit of detection of ~100 fmol (**Figure D14**).

Seeking to maximize this detection sensitivity, we lastly explored the chemiluminescent substrate used in our assay. In particular, we employed SuperSignal™ ELISA Pico substrate (Thermo Fisher Scientific) for our initial optimization assays, and next tested the SuperSignal™ West Atto variation, which is currently reported as the most sensitive commercially available substrate. Interestingly, when we incorporated this material into our optimized detection conditions, we observed erratic and overall poor signal compared to our previous EndoVLISA tests using the Pico substrate (**Figure D15**). Given that West Atto substrate is specifically formulated to detect ultra-low analyte concentrations, we wondered if the 2º HRP concentration was simply too high and that fast substrate turnover was resulting in rapid loss of light signal output. Interestingly, when diluting the 2º antibody-HRP to 1:40,000, we saw a dramatic improvement in performance, with excellent linearity below 0.5 pmol inosine and a lower limit of detection approaching ~30 fmol and representing a 2-3-fold improvement over the Pico substrate (**Figure D16**). Diluting this further (1:80,000-1:160,000) produced diminishing signal in our RNA I standard curves (**Figure D15**), so we selected 1:40,000 as an optimal 2º antibody concentration for subsequent assays.

While our results thus far were encouraging, detecting small amounts of A-to-I editing in complex samples is pivotal to our eventual goals, and so we next tested EndoVLISA performance in the presence of off-target mRNA. We first synthesized an ~800 nt mRNA by *in vitro* transcription, importantly using only the four canonical ribonucleoside triphosphates (A, U, C, G) to enable precise control over nucleobase content in each sample for subsequent inosine quantification. We first combined this mRNA (1 µg) with different ratios of our RNA I control strand, and subjected these samples to both glyoxal denaturation and 3' biotinylation prior to EndoVLISA. As shown in **Figure D17**, we did observe an overall signal decrease in the presence of off-target mRNA. These samples are undoubtedly more complex than our previous analyses using only small oligonucleotides, and it is likely that EndoV and antibody binding is less efficient in this crowded microenvironment. Despite this, EndoVLISA displayed excellent linearity from 0-1 pmol (**Figure 5.2d**). Additionally, we were still able to reliably detect ~100 fmol inosine per µg mRNA, roughly indicating that our method can sense 1 inosine molecule for every ~30,000 nucleotides (0.003%), which is encouragingly near estimated lower levels of A-to-I content (~0.01 – <0.001% of all bases).[5, 36]

Although LC/MS is not typically used to quantify inosine content in cellular RNA due to practical limitations, we were curious to benchmark our EndoVLISA assay relative to this approach. We first analyzed decreasing amounts of inosine ribonucleoside, and found that MS detection was only reliable above ~5 pmol (**Figures D18, D19**). To directly compare with our previous EndoVLISA selectivity test (**Figure 5.2d**), we also spiked inosine into 1 µg of an equimolar mixture containing each of the four canonical nucleosides (A, U, C, G) as well as the two other major modified bases found in RNA (m^6A and pseudouridine) (**Figure D20, D21**). Given the inclusion of an upstream chromatography separation step, LC/MS was predictably unaffected by this added sample complexity, and yielded similar performance whether inosine was injected alone or when mixed with other ribonucleosides (**Figure D21**). This lower limit of detection (~5-10 pmol using an accurate mass TOF-ESI instrument) was comparable to other studies employing

LC/MS for inosine quantification,[45-46] but still represented >100-fold decrease in sensitivity compared to our demonstrated EndoVLISA performance (**Figure 5.2d**). This likely explains why this method has found use in the case of more frequent RNA modifications but has not been widely adopted for measuring A-to-I editing signatures.

After thoroughly demonstrating feasibility and robustness of our EndoVLISA method for detecting inosine in RNA, we next sought to test our approach in a cellular context for measuring global changes in A-to-I editing frequency. Additionally, we were curious how the EndoVLISA would compare to the current standard method of inosine quantification utilizing RNA-seq. Because the majority of editing sites reside within repetitive *Alu* elements in mRNA,[4-8] the "*Alu* Editing Index" (AEI) was developed to computationally measure editing frequencies across these sites, in turn providing a global estimate of ADAR1 editing activity.[38, 75-76]

To compare this approach to EndoVLISA, we first selected HEK293T as a suitable cell line to induce overexpression of ADAR1. Interestingly, most immortalized cell lines also exhibit very low editing activity compared to human primary cells and tissues, which is partly explained by overall lower ADAR1 expression levels.[76] AEI profiling of different cell lines also showed that 293T cells exhibit particularly low editing levels (**Figure D22**), and this would fortunately allow us to simulate and detect increases in ADAR1 expression that are characteristic of different developmental and disease-specific changes in global editing activity. We first induced overexpression using increasing amounts of a plasmid encoding a GFP-tagged ADAR1 p110 isoform.[77] We then harvested both whole cell lysate and mRNA from treated 293T cells and analyzed these materials by western blotting, RNA-seq, and EndoVLISA (**Figure 5.3**). We confirmed very low ADAR1 levels in untreated 293T cells, and observed a clear increase in ADAR1 expression with increasing amounts of transfected plasmid by both western blot (**Figure 5.3a**,**b**) and GFP fluorescence in collected lysates (**Figure D23**).

Using isolated mRNA from these samples, we performed RNA-seq and calculated global A-to-I editing activity in treated cells using the *Alu* index and observed a proportional rise in AEI

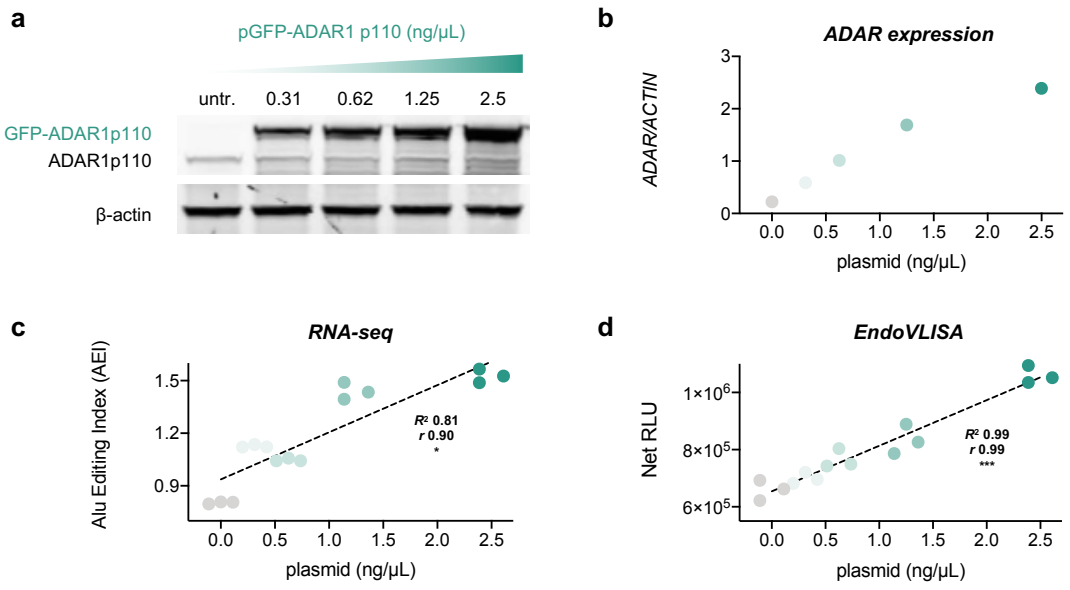


Figure 5.3. EndoVLISA detects global cellular changes in A-to-I RNA editing. HEK293T cells were transfected with increasing amounts of a plasmid encoding a GFP-tagged ADAR1 p110 isoform. a) Western blotting of lysates obtained from untreated (untr.) 293T cells and those receiving increasing amounts of transfected plasmid, illustrating higher levels of GFP-ADAR expression. ADAR1 p110 band (black) in blot signifies endogenous expression in untreated cells. GFP-ADAR (green) is larger due to addition of GFP tag. b) Densitometric estimation of ADAR1 expression in different samples as a ratio of ADAR1 band intensity compared to a control protein (β-actin). c) RNA-seq was performed on isolated mRNA from each treatment group, and A-to-I editing activity was calculated using the Alu Editing Index (AEI). Values represent individual samples for each group (n = 3). d) The same mRNA material was also biotinylated, glyoxal-denatured, and tested using EndoVLISA (1 μg mRNA per well, EndoV-MBP 1:1000, 1º antibody 1:1000, 2º antibody-HRP 1:40,000, West Atto substrate). Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Data points represent individual values from each well (n = 3 for each treatment group). Linear regression (black dashed lines), R2 and pearson (r) correlation (* denotes $p < 0.05$, *** $p < 0.001$) were computed in Prism.

values correlating with plasmid amount and overall ADAR expression (**Figure 5.3c**). In parallel, we tested this same mRNA material using our EndoVLISA approach, and were delighted to see a commensurate rise in chemiluminescent response that correlated with increasing ADAR1 plasmid transfection (**Figure 5.3d**). Using a standard curve comprised of known amounts of RNA I control strand added to 1 μg *in vitro* transcribed mRNA, we also estimated inosine concentrations in these samples and saw a comparable increase in these levels (**Figure D24**). Both RNA-seq and EndoVLISA showed statistically significant linearity (**Figures 5.3c,d**), and were also in good agreement with each other (**Figure D25**), indicating our method was reliable for detecting global

increases in A-to-I editing. Although the AEI was more sensitive in detecting smaller potential changes in ADAR1 activity, particular between untreated 293T cells and those receiving lower plasmid amounts (<1 ng/ul), this approach was less responsive in higher transfection ranges. As shown in **Figures 5.3c-d**, similar index values were observed in these samples, whereas EndoVLISA yielded more distinct and proportional responses. Overall, our results demonstrated that EndoVLISA is feasible and effective for accurately detecting global cellular changes in A-to-I RNA editing. Additionally, our method not only produced highly similar responses when compared to the current standard method using RNA-seq, but the EndoVLISA also offered significant time and cost-savings compared to RNA-seq, representing a ~200-fold reduction in analysis costs per sample (**Tables D2, D3**) and positioning this method as a highly accessible and robust tool for characterizing RNA editing signatures at medium- to high throughput.

We were next interested in testing our method in mRNA from primary human samples, as A-to-I editing activity can vary between different organs, tissues, and cell types within the body.[4, 38] These editing changes are important drivers of stem cell differentiation, embryogenesis, and immune cell activation, and large-scale fluctuations in global editing frequency can also directly cause or be indicative of a variety of developmental diseases, autoimmune illnesses, neurodegenerative disorders, and cancer types.[7, 12, 21-27]

Towards our goal of profiling these changes, we first tested relative inosine levels in a panel of normal human tissues. Although most adult tissue types exhibit similar A-to-I activity, higher editing has been observed in both brain and aorta tissues, while skeletal muscle and pancreatic mRNA display uniquely low editing rates.[4, 38] As shown in **Figure 5.4**, we were excitingly able to detect these large-scale differences using our EndoVLISA bioassay (**Figure 5.4b**). As an additional level of validation for our method, these global trends were also highly similar to results from a large-scale *Alu* indexing of >9,000 RNA-seq datasets from the Genotype-Tissue Expression (GTEx) project (**Figure 5.4a**).[38, 78] In particular, while brain and aorta mRNA displayed the highest mean response among all tissues tested in both RNA-seq and EndoVLISA,

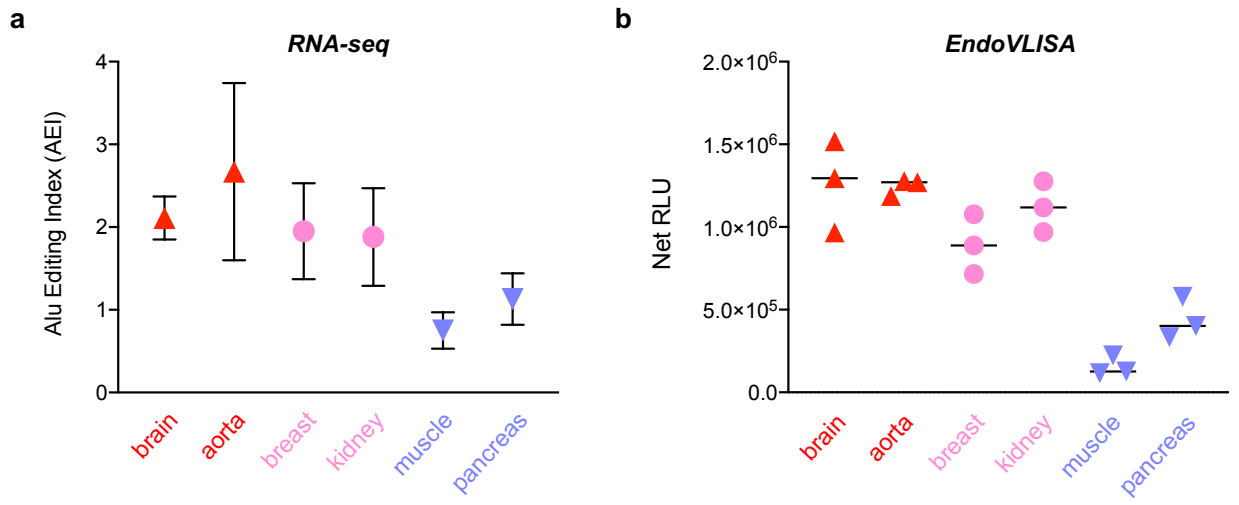


Figure 5.4. EndoVLISA detects tissue-specific A-to-I editing signatures. a) RNA-seq analysis of A-to-I editing activity in human tissues from the GTEx project. AEI was calculated from RNA-seq datasets across all indicated human tissues. Reproduced from compiled data, and values represent mean and SD (brain n = 103 individuals, aorta n = 246, breast n = 219, kidney n = 38, muscle n = 450, pancreas n = 192). b) Purified mRNA from the indicated human tissues was tested using EndoVLISA (500 ng mRNA per well, EndoV-MBP 1:1000, 1º antibody 1:1000, 2º antibody-HRP 1:40,000, West Atto substrate.) Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Data points represent individual values from each well (n = 3 for each sample).

these were also statistically similar to either breast (mammary gland) or kidney mRNA (**Figures 5.4a,b**). However, both skeletal muscle and pancreatic tissue mRNA were substantially lower than all 4 other tissues in both methods and reproduced previous observations of these differences.[4, 38] Overall signal was also much higher than in our previous experiments using HEK293T cells (**Figure 5.3**), supporting previous analyses showing lower overall editing levels in immortalized cell lines.[76] This higher signal did require us to use 500 ng instead of 1 µg mRNA per well as used in our previous EndoVLISA tests, and we were also able to estimate inosine content using an appropriate standard curve (RNA I mixed with 500 ng *in vitro* transcribed mRNA) (**Figure D25**). However, some of these samples produced very high chemiluminescent signal (especially aorta and brain) and were slightly above the maximum amount of RNA I used in our standard curve (2 pmol), so actual inosine concentrations may be slightly higher. Regardless, the EndoVLISA was consistently able to detect these tissue-level epitranscriptomic signatures, and

was highly comparable to large-scale RNA-seq analyses (**Figure D26**), further verifying our method as an effective and straightforward means for profiling global A-to-I editing signatures.

ADAR1 overexpression is rapidly emerging as a potential molecular mechanism for cancer progression, and the majority of cancer types display significantly upregulated A-to-I RNA editing levels.[20-21, 33, 79-80] Because of this tremendous potential as a clinical biomarker for disease characterization, we were interested in applying the EndoVLISA platform for identifying epitranscriptomic differences between healthy and diseased tissue. We first obtained total RNA from several normal human tissues as well as RNA samples from breast cancer, kidney renal cell carcinoma, lung adenocarcinoma, and liver hepatocellular carcinoma. We purified mRNA from each source through two rounds of polyT selection, and then performed the EndoVLISA workflow.

As shown in **Figure 5.5**, we detected significantly higher chemiluminescent signal in all 4 cancer types, reflecting large overall increases in global editing activity. While these tumor RNA samples were actually too high for accurate quantification using our typical standard curve, Net RLU values can be used to roughly estimate fold-increases in inosine content (**Figure 5.5**). In particular, breast, liver, and lung cancer samples exhibited the largest increases, displaying ~8, ~6, and ~5-fold upregulated activity, respectively. These observations were in close agreement with prior studies of RNA editing in all three cancer types,[32, 81-83] and additionally support large-scale bioinformatic profiling of editing changes in matched RNA-seq datasets from the Cancer Genome Atlas,[84] which identified significant upregulation of A-to-I activity in breast, liver, and lung cancer progression.[20, 33] Interestingly, different types of kidney carcinomas can exhibit hypo- or hyperediting signatures,[20, 31] and while we observed ~4-fold increase in EndoVLISA response in our tumor sample, RNA-seq analysis of a similar renal cancer found no significant difference in average *Alu* editing rates between normal and tumorigenic tissue.[33] However, this study also noted variable patterns between individuals and observed both over- and underediting phenotypes in renal carcinoma datasets (62 patients),[33] and so our results may simply reflect the composition of our own experimental sample. Additionally, profiling of Cancer Genome Atlas

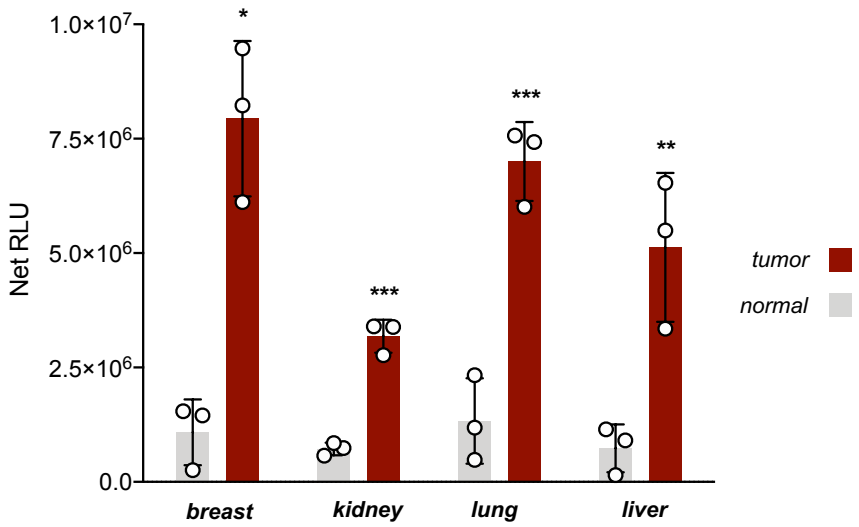


Figure 5.5 EndoVLISA detects upregulated A-to-I editing in several cancer types. mRNA from the indicated human tissues (normal = grey and tumor = red) was tested using EndoVLISA (0.5 µg mRNA per well, EndoV-MBP 1:1000, 1º antibody 1:1000, 2º HRP 1:40,000, West Atto substrate). Values represent net relative luminescent units (RLU, arbitrary units) cal-culated by subtracting appropriate blank wells (no RNA) from each set. Data points represent individual values from each well (n = 3 for each sample). Unpaired t-tests were computed in Prism between normal and tumor samples for all tissue types (* denotes $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$).

datasets interestingly revealed that while overall editing differentials were higher in kidney tumors compared to matched normal tissue controls, *Alu*-specific editing activity was significantly depleted in these samples.[20] These results suggest the presence of additional molecular mechanisms that regulate editing patterns in these cancer types and may point to a major limitation in relying solely on *Alu* editing signatures for indirectly inferring global cellular A-to-I editing activity. Expanding EndoVLISA profiling of tumor-specific editing signatures in high numbers of individuals will be a worthy pursuit for unequivocally resolving this heterogeneity and will likely enable identification of other important editing regulation mechanisms and drivers of cancer progression. Overall, these experiments demonstrated rapid, effective, and consistent EndoVLISA detection of upregulated A-to-I editing activity in several known cancer types, highlighting the powerful clinical and diagnostic potential of our method.

5.4 Conclusions:

A-to-I RNA editing is a key molecular event that significantly influences cellular function and disease progression. Despite this importance, our understanding of RNA editing regulation and its precise role in different biological pathways remains limited. Much of this uncertainty results from a lack of accessible tools to measure large-scale differences in A-to-I editing activity between different cell and tissue types. Gaining a deeper understanding of these and other RNA modification patterns would provide invaluable information on necessary cellular mechanisms and could potentially lead to the identification of novel druggable protein and RNA targets. Overall A-to-I editing activity is also a key biosignature for a number of disorders, and the ability to rapidly detect these changes could be powerfully leveraged toward new methods for disease diagnosis.

Here, we have addressed several of these needs by demonstrating proof-of-concept for a novel A-to-I editing bioassay based on the well-characterized ELISA format. Our method is the first assay of this type specifically designed to quantify A-to-I RNA editing, and enables direct immunodetection of global epitranscriptomic activity without sequencing. EndoVLISA is straightforward and exclusively uses commercially available components, positioning it as a highly accessible approach for measuring changes in A-to-I RNA editing. After carefully optimizing key parameters in our assay, we successfully validated its performance and demonstrated its use for detecting global cellular changes in RNA editing activity in both treated cell lines and primary tissue samples.

Looking to the future, we aim to deploy EndoVLISA for answering other long-standing questions in the field, and in particular envision that multiplexing our assay with existing ELISA kits specific for other RNA and DNA modifications, including N^6-methyladenosine (m^6A),[67-68] 5-methylcytosine (5mC),[69] 5-hydroxymethylcytosine (5hmC), and 5-carboxylcytosine (5caC),[70] will provide an unprecedented opportunity to elucidate functional relationships between different nucleic acid modifications. EndoVLISA is also well-poised for implementation in CRISPR-based knockout screens,[85] and will likely find utility in identifying protein regulators of ADAR activity that

influence global A-to-I signatures. Building off of our initial results, we anticipate that additional developmental efforts will yield increased benefit in both assay performance and versatility. While our present study confirms our previous observation that wild-type EndoV from *E. coli* provides sufficient affinity and selectivity,[54] we anticipate that performance will also be further enhanced as we explore EndoV variants from different species or perform engineering or evolution to further increase affinity and selectivity. Moreover, streamlining and automating our overall workflow for high-throughput drug screening could provide a significant technological advance for identifying ADAR-modulating pharmaceuticals, especially in non-engineered cell lines or primary tissues.

Finally, modifying the physical properties of our assay to incorporate colorimetric substrates or paper-based ELISA platforms could expand the utility of the EndoVLISA assay for creation of a point-of-care diagnostic. Together, this report details a simple yet powerful new tool to complement existing epitranscriptomic sequencing technologies. Our assay has broad versatility across many research disciplines, and we anticipate that EndoVLISA will dramatically improve our ability to rapidly measure global changes in A-to-I RNA editing across a diverse range of biological contexts.

5.5 Materials and Methods:

RNA Oligoribonucleotides

Oligonucleotide control strands used in this study were custom designed and purchased from Integrated DNA Technologies.

RNA A: 5’ GUGCCUUUAUGC**A**GCAAGGAUGCG 3’

RNA I: 5’ GUGCCUUUAUGC**I**GCAAGGAUGCG 3’

RNA Denaturation and 3’ Fluorescent Labeling

250 pmol (~2 µg) of RNA A or RNA I oligo was first combined with 62.5 µM $NaIO_4$ (Thermo Fisher Scientific) in a total volume of 40 µL of 10 mM sodium acetate buffer, pH 5.6 and incubated in the dark for 90 minutes at 25 ºC. RNA was then purified using the Monarch® RNA Cleanup Kit (New England Biolabs) and eluted in 35 µl nuclease-free water. This was then directly combined with 50 µl DMSO and 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and incubated for 2 hours at 50 ºC. Reactions were ethanol precipitated and reconstituted in 30 µL of nuclease-free water. Then was then directly combined with 0.2 mM Cy5 Hydrazide (Lumiprobe) in a total volume of 40 µL of 10 mM sodium acetate buffer, pH 5.6 and 7.5% DMSO. Reactions were incubated at 37 ºC for 2 hours and then ethanol precipitated, reconstituted in 30 µL of nuclease-free water, and then quantified using a NanoDrop spectrophotometer (Thermo Fisher). To validate covalent denaturation and labeling (**Figure D3d**), 2 pmol each of unmodified RNA, glyoxalated RNA, and dual glyoxalated and 3’ end labeled RNA were analyzed by 20% denaturing PAGE and stained with 1X SYBR gold (Thermo Fisher). The gel was then imaged with a GE Amersham Typhoon RGB scanner.

Microscale Thermophoresis (MST)

For our initial titration study (**Figure D4**), increasing total amounts of Cy5-labeled RNA I (0 - 90 ng) was added to 10 µl of 1X EndoV binding buffer (20 mM Tris, 137 mM NaCl, 3 mM KCl, 5 mM $CaCl_2$, 15 µM EDTA, 150 µM DTT, 0.025% Triton X-100, 30 µg/mL BSA, 7% glycerol, pH 7.4.). Samples were then loaded into NT.115 standard glass capillaries (Nanotemper) and scanned using a Nanotemper Monolith instrument at 5% LED intensity power. Peaks represent raw fluorescence values and capillary shape scans obtained directly from Monolith software. Scans were arranged in GraphPad Prism 9. To test binding affinity (**Figure D5a,b**) varying amounts of recombinant eEndoV-MBP fusion protein (New England Biolabs) were combined with 10 ng of Cy5-labeled and glyoxal denatured RNA A or RNA I targets in a final volume of 10 µL and allowed to incubate for 30 min at room temperature. Final buffer conditions in these samples were 20 mM Tris, 137 mM NaCl, 3 mM KCl, 5 mM $CaCl_2$, 15 µM EDTA, 150 µM DTT, 0.025% Triton X-100, 30 µg/mL BSA, 7% glycerol, pH 7.4. Samples were then loaded into NT.115 standard glass capillaries. MST experiments were performed using a Nanotemper Monolith NT.115 Pico instrument. All measurements were analyzed using the Pico-RED filter with 5% LED intensity and 40% laser power. Data were fitted using GraphPad Prism 9 analysis software to determine K_d values. Binding tests were performed in triplicate in separate trials. ΔF_{norm} was calculated by subtracting the blank (no EndoV) raw normalized fluorescence value (F_{norm}) from each test sample. Fold selectivity at each EndoV concentration was calculated as the average ΔF_{norm} of I/A.To test selectivity (**Figure D6**) varying ratios of RNA I was combined with RNA A (constant 10 ng) and mixed with 100 nM eEndoV-MBP in 10 µL 1X EndoV binding buffer. Samples (n = 3) were allowed to incubate for 30 minutes at room temperature and then analyzed by MST as described above.

RNA 3' Biotinylation and Streptavidin Plate Immobilization

To test immobilization and binding capacity of streptavidin-coated plates (**Figure D7**), 5 µg of RNA A was first fluorescently labeled with Cyanine 5 (Cy5) using the Label IT® Tracker™ Intracellular Nucleic Acid Localization Kit (Mirus Bio) according to the manufacturer's instructions. The oligonucleotide was then ethanol precipitated and analyzed by UV/Vis spectrophotometry to measure degree of labeling, confirming approximately 2-3 dye molecules per strand. Next, 250 pmol (~2 µg) of Cy5-labeled RNA A oligo was then combined with 62.5 µM $NaIO_4$ (Thermo Fisher Scientific) in a total volume of 40 µL of 10 mM sodium acetate buffer, pH 5.6 and incubated in the dark for 90 minutes at 25 ºC. RNA was then purified using the Monarch® RNA Cleanup Kit (New England Biolabs) and eluted in 35 µl nuclease-free water. This was then directly combined with 50 µl DMSO and 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and incubated for 2 hours at 50 ºC. Reactions were ethanol precipitated and reconstituted in 30 µL of nuclease-free water. Then was then directly combined with 25 mM biotin-dPEG®$_4$-hydrazide (Sigma Aldrich) in a total volume of 40 µL of 10 mM sodium acetate buffer, pH 5.6 and 10% DMSO. Negative control RNA reaction (- biotin) contained no biotin hydrazide. Reactions were incubated at 37 ºC for 2 hours, ethanol precipitated, reconstituted in 30 µL of nuclease-free water, and then quantified using a NanoDrop spectrophotometer (Thermo Fisher). Next, various dilutions of -/+ biotin RNA A were prepared in 1X binding/wash buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.01% Tween 20, pH 7.4). Each well of a Pierce™ Streptavidin Coated Plate, White, 96-Well (Thermo Fisher) was then washed two times with 200 µL of 1X binding/wash buffer. In triplicate wells, 100 µL of each RNA dilution was then added to each well and incubated for 1 hour with shaking at room temperature. Plate was then washed two times with 200 µL of 1X binding/wash buffer, and 100 µL of buffer was added to each well. Fluorescence was measured with a BioTek Cytation 5 plate reader using an excitation and emission wavelengths of 650 nm and 670 nm. Net relative fluorescence units (RFU, arbitrary units) were calculated by subtracting appropriate blank wells (buffer only, no RNA).

EndoVLISA Optimization

To optimize EndoVLISA performance (**Figures 5.2a-d, D8-13**), RNA A and RNA I stocks were first glyoxal denatured and 3' biotinylated as described above. For each EndoVLISA probing combination, 0, 0.1 and 2 pmol of RNA I and 2 pmol RNA A was added to duplicate wells of a white streptavidin-coated plate in 100 µL 1X binding/wash buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.01% Tween 20, pH 7.4) and incubated for 1 hour at room temperature with shaking. Plate was then washed two times with 200 µL of 1X binding/wash buffer, and each well and then received 100 µL of an EndoV-MBP (New England Biolabs) solution in 1X buffer (1:500, 1:1000, or 1:2000). Plates were incubated for 1 hour at room temperature with shaking, followed by 2x200 µL washes with 1X binding/wash buffer. Each well then received 100 µL of an anti-MBP primary mouse monoclonal antibody (New England Biolabs) solution in 1X buffer (1:1000, 1:2000, 1:4000, or 1:8000). After incubating for 1 hour at room temperature with shaking, plates were again washed two times with 200 µL 1X binding/wash buffer, and then 100 µL of a Goat anti-Mouse IgG (heavy + light chain) polyclonal HRP-conjugated secondary antibody (Thermo Fisher) solution in 1X buffer (1:10,000, 1:20,000 or 1:40,000) was added to appropriate wells. Plates were incubated at room temperature for 1 hour with shaking, followed by 2x200 µL washes with 1X binding/wash buffer. Each well was then emptied, and 100 µL of SuperSignal™ ELISA Pico Chemiluminescent Substrate working solution was added followed by shaking for 1 minute at room temperature. Light signal was then measured using a BioTek Cytation 5 platereader using a luminescence detection fiber optic filter (gain = 100). Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Pearson *r* *p*-value was computed in GraphPad Prism 9 using on-target RNA I signals from 0, 0.1 and 2 pmol RNA signals. Fold-selectivity was calculated as the net chemiluminescent signal of RNA I divided by RNA A (2 pmol/well). To acquire an image of chemiluminescent EndoVLISA detection (**Figure 5.2d**), different amounts of RNA A or I as indicated was first immobilized into a 96-well streptavidin coated plate, followed by immunodetection with different reagent combinations

(EndoV, 1º antibody, 2º HRP) as described above. Plate was then incubated with 100 µL Pico ELISA chemiluminescent substrate for 1 minute with shaking. Reactions were then transferred to a clear 96-well plate and chemiluminescent signal was captured using a GE Amersham Typhoon RGB scanner. Heatmap image was obtained using the Typhoon acquisition software and reflects relative luminescent intensities. Bar graph denotes net RLU values obtained from BioTek plate reader measurement.

EndoVLISA Linearity and Sensitivity

To assess EndoVLISA performance using optimal conditions identified from initial screen (**Figure D14**), decreasing amounts of biotinylated, glyoxal-denatured RNA I strand (0 – 2 pmol) were immobilized in white streptavidin-coated 96-well plates and then probed with the indicated detection reagent combinations (EndoV 1:1000, 1º antibody 1:1000, 2º HRP 1:20,000, ELISA Pico substrate) following the EndoVLISA procedure described earlier. Standard curve was then plotted in both linear and $\log_2$ scale using GraphPad Prism 9. Linear regression (R^2), pearson (r) correlation and p-value were computed in Prism. This same basic procedure was repeated when testing the SuperSignal™ West Atto chemiluminescent substrate (Thermo Fisher Scientific) and different amounts of 2º HRP (**Figures D15, D16**).

***In vitro* mRNA production**

A plasmid encoding GFP was purchased from AddGene (pET28 GFP, 60733). 1 µg of purified plasmid was then linearized by digestion with BamHI-HF (New England Biolabs) for 30 minutes at 37 ºC. Cut plasmid was then isolated using the Monarch® PCR Purification Kit (New England Biolabs). mRNA was then generated *in vitro* using the HiScribe™ T7 Quick High Yield RNA Synthesis Kit at 37 ºC overnight, and mRNA was purified using the Monarch® RNA Cleanup Kit (New England Biolabs) and quantified using a NanoDrop spectrophotometer.

EndoVLISA Selectivity with mRNA

To assess EndoVLISA performance in the presence of off-target mRNA (**Figure 5.2e, D17**), decreasing amounts of RNA I strand (0 – 1 pmol) were mixed in triplicate with 1 µg of *in vitro* transcribed mRNA. These samples were then subjected to periodate oxidation, glyoxal denaturation, and 3' biotinylation as described earlier. Each sample was then immobilized in white streptavidin-coated 96-well plates and then subjected to EndoVLISA detection (EndoV 1:1000, 1º antibody 1:1000, 2º HRP 1:40,000, SuperSignal™ West Atto substrate) following the general procedure described earlier. Standard curve was then plotted in GraphPad Prism 9, and linear regression (R^2), pearson (r) correlation and p-value were also computed in Prism.

Liquid Chromatography and Mass Spectrometry (LC/MS)

Ribonucleosides inosine, guanosine, adenosine, cytidine and uridine were purchased from Sigma Aldrich. Pseudouridine was purchased from Santa Cruz Biotechnology, and N^6-methyladenosine (m^6A) was purchased from MedChemExpress. Standard stock concentrations (1 mM) were prepared in pure nuclease-free water and diluted in water where necessary. Reversed-phase LC was performed on an attached Agilent 1260 Infinity II system using a 3 µM, 4.6 mm X 75 mm Atlantis T3, 100Å C18 column (Waters). All samples were analyzed using a linear mobile phase gradient from 0% to 20% acetonitrile in water + 0.1% formic acid over 10 minutes. All mass spectra were obtained using an Agilent 6320 time-of-flight (TOF) electrospray ionization (ESI) LC/MS instrument in positive ionization mode.

Cell Line Maintenance and Transfection

HEK293T cells (ATCC CRL-3216) were cultured in Dulbecco's Modified Eagle's Medium supplemented with 10% fetal bovine serum and maintained at 37 ºC, 5% CO_2. Cells were grown in T75 Nunc™ tissue culture-treated flasks. When ~50% confluent, cells were then transfected with increasing amounts of plasmid encoding a GFP-tagged ADAR1 enzyme (AddGene, pmGFP-

ADAR1-p110, #117928). Transfection was performed directly in flasks with indicated amount of plasmid (**Figure 5.3a**) using Lipofectamine 3000 (Thermo Fisher Scientific) according to the manufacturer's instructions. After ~48 h incubation at 37 ºC, 5% CO_2, cells were harvested by trypsinization and washed twice in cold 1X phosphate buffered saline.

Western Blotting

Whole cell lysates were first collected by adding 1mL of M-PER lysis reagent (Thermo Fisher Scientific) for each 100 mg (~100µL) of wet cell pellet. After pipetting to homogenize the pellet, mixture was incubated with end-over-end rotation for 10 minutes at room temperature. Cell debris was then removed by centrifugation at ~14,000 × *g* for 15 minutes, after which the supernatant was collected and protein content was estimated using Pierce™ 660nm Protein Assay Reagent (Thermo Fisher Scientific). To verify GFP-ADAR expression, 20 µl of this lysate was loaded into 384-well black plates in triplicate and measured using a BioTek Cytation 5 plate reader using excitation at 488 nm and emission at 509 nm (**Figure D23**). For western blotting analysis, 20 µg of whole 293T cell lysate from each transfection group was mixed with 1X reducing sample buffer (Thermo Fisher) and heated to 95 ºC for 5 minutes. Each sample was then resolved on a 4-20% Tris Glycine polyacrylamide gel (Thermo Fisher) along with a PAGERuler Plus Prestained Protein ladder (Thermo Fisher). After electrophoresis, the gel was removed from the cassette and briefly washed with diH_2O, after which it was transferred to a 0.2 µm nitrocellulose membrane (Invitrogen) for 2 hours at 25 V. The membrane was briefly washed with diH_2O, after which it was submerged for 1 hour in 1X blocking buffer comprised of a 5% non-fat dry milk solution (Kroger) in 1X TBST (Thermo Fisher, 25 mM Tris, 150 mM NaCl, 0.5% Tween 20, pH 7.4). The membrane was cut into two pieces and then probed with an anti-ADAR1 rabbit monoclonal antibody (Cell Signaling Technologies D7E2M) at a 1:1000 dilution and an anti-actin mouse monoclonal antibody (Thermo Fisher Scientific MA1-140) at a 1:5000 dilution in 1X blocking buffer for 1 hour with gentle agitation. After washing the blots 3x5 minutes in 1X TBST,

membranes were probed with a Goat anti-Rabbit IgG Alexa Fluor 647-conjugated secondary antibody (Thermo Fisher A21244) at a 1:1000 dilution and a Goat anti-Mouse IgG Alexa Fluor 488-conjugated secondary antibody (Thermo Fisher A11001) at a 1:1000 dilution in 1X blocking buffer for 1 hour protected from light. Membranes were washed 3x5 minutes in 1X TBST, followed by imaging of the blot using a GE Amersham Typhoon RGB scanner. Densitometric analysis was performed using ImageJ (Fiji).

293T mRNA Isolation, Processing and EndoVLISA detection

mRNA was collected from each treated flask by first isolating total RNA using the Monarch® Total RNA Miniprep Kit (New England Biolabs). Isolated material was then processed using two rounds of Oligo dT_{25} purification with the Magnetic mRNA Isolation Kit (New England Biolabs). After eluting in nuclease-free water, RNA concentration was estimated using a NanoDrop spectrophotometer. In separate reactions, 1 µg of collected mRNA was glyoxal denatured and biotinylated. Each sample was then immobilized in white streptavidin-coated 96-well plates and then subjected to EndoVLISA detection (EndoV 1:1000, 1º antibody 1:1000, 2º HRP 1:40,000, SuperSignal™ West Atto substrate) following the general procedure described earlier. A standard curve comprising 0-2 pmol of RNA I mixed with 1 µg *in vitro* transcribed mRNA was also processed and detected in parallel and used for estimating inosine content in treated mRNA samples (**Figure D24**). Standard curve was plotted in GraphPad Prism 9, and linear regression was used to correlate sample concentration.

RNA-seq and Alu Indexing

In triplicate, 50 ng of isolated mRNA from each 293T transfection group was used to prepare sequencing libraries with the SMARTer® Stranded Total RNA Sample Prep Kit - Low Input Mammalian 24 reactions, (Takara Bio); standard 8-bp i5 and i7 Illumina index barcodes and adapters were added to each library. All libraries were then sequenced using a NextSeq 500/550

High Output Kit v2.5 300 Cycles (Illumina) to produce paired end 150-bp reads (approximately 15M reads per sample). Raw FASTQ data were first trimmed to remove the first 3 bp using Trimmomatic[1] with parameter HEADCROP:3. QC was performed on data to check read quality (PHRED33) using FastQC [https://www.bioinformatics.babraham.ac.uk/projects/fastqc/] and MultiQC[2] before and after data trimming. Reads were then aligned to the human reference genome hg38 via STAR 2.5.2[3] with additional parameter --outFilterMatchNminOverLread 0.95 as this parameter is optimal for the detection of A-to-I editing[4]. Resulting .bam files were sorted, and had duplicates marked and removed using Samtools 1.3[5] and PicardTools 2.0.1 [http://broadinstitute.github.io/picard/] respectively. The RNA editing indexer package by Roth et al.[2,6] was used as written on the tool GitHub page to process an entire directory of samples using default settings to calculate the Alu Editing Index (AE) for each sample.

1) Bolger, A. M.; Lohse, M.; Usadel, B., Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014,** *30* (15), 2114-2120.
2) Ewels, P, Magnusson, M, and Kaller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics (2016). doi: 10.1093/bioinformatics/btw354.
3) Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635.
4) Roth, S.H., Levanon, E.Y. & Eisenberg, E. Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nat Methods* **16,** 1131–1138 (2019). https://doi.org/10.1038/s41592-019-0610-9.
5) Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, 1000 Genome Project Data Processing Subgroup,

The Sequence Alignment/Map format and SAMtools, *Bioinformatics*, Volume 25, Issue 16, 15 August 2009, Pages 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.

6) https://github.com/a2iEditing/RNAEditingIndexer

Human Tissue mRNA and Processing

Purified mRNA from human aorta, brain, breast (mammary gland), kidney, and skeletal muscle were obtained from Takara Bio (Clontech). In triplicate reactions, 500 ng of this material was glyoxal denatured and biotinylated. Each sample was then immobilized in white streptavidin-coated 96-well plates and then subjected to EndoVLISA detection (EndoV 1:1000, 1º antibody 1:1000, 2º antibody-HRP 1:40,000, SuperSignal™ West Atto substrate) following the general procedure described earlier. A standard curve comprising 0-2 pmol of RNA I mixed with 500 ng *in vitro* transcribed mRNA was also processed and detected in parallel and used for estimating inosine content in treated mRNA samples (**Figure D26**). Standard curve was plotted in GraphPad Prism 9, and linear regression was used to correlate sample concentration. For EndoVLISA detection in cancer, normal total RNA from breast (mammary gland), kidney, lung, and liver was purchased from Takara Bio (Clontech). Tumor total RNA from breast and kidney cancer biopsies were also purchased from Takara, while liver and lung total RNA was obtained from BioChain. 250 µg of each sample was subjected to two rounds of Oligo dT_{25} purification with the Magnetic mRNA Isolation Kit (New England Biolabs). After eluting in nuclease-free water, RNA concentration was estimated using a NanoDrop spectrophotometer. In separate reactions, 500 ng of collected mRNA was glyoxal denatured and biotinylated. Each sample was then immobilized in white streptavidin-coated 96-well plates and then subjected to EndoVLISA detection (EndoV 1:1000, 1º antibody 1:1000, 2º HRP 1:40,000, SuperSignal™ West Atto substrate) following the general procedure described earlier.

5.6 References

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

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

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

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

5. Porath, H. T.; Knisbacher, B. A.; Eisenberg, E.; Levanon, E. Y., Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. *Genome biology* **2017,** *18* (1), 1-12.

6. Ramaswami, G.; Li, J. B., RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic acids research* **2013,** *42* (D1), D109-D113.

7. Bazak, L.; Haviv, A.; Barak, M.; Jacob-Hirsch, J.; Deng, P.; Zhang, R.; Isaacs, F. J.; Rechavi, G.; Li, J. B.; Eisenberg, E., A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome research* **2014,** *24* (3), 365-376.

8. Kim, D. D.; Kim, T. T.; Walsh, T.; Kobayashi, Y.; Matise, T. C.; Buyske, S.; Gabriel, A., Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome research* **2004,** *14* (9), 1719-1725.

9. Eisenberg, E.; Levanon, E. Y., A-to-I RNA editing—immune protector and transcriptome diversifier. *Nature Reviews Genetics* **2018,** *19* (8), 473-490.

10. Liddicoat, B. J.; Piskol, R.; Chalk, A. M.; Ramaswami, G.; Higuchi, M.; Hartner, J. C.; Li, J. B.; Seeburg, P. H.; Walkley, C. R., RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **2015,** *349* (6252), 1115-1120.

11. Pestal, K.; Funk, C. C.; Snyder, J. M.; Price, N. D.; Treuting, P. M.; Stetson, D. B., Isoforms of RNA-editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. *Immunity* **2015,** *43* (5), 933-944.

12. Mannion, N. M.; Greenwood, S. M.; Young, R.; Cox, S.; Brindle, J.; Read, D.; Nellåker, C.; Vesely, C.; Ponting, C. P.; McLaughlin, P. J., The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell reports* **2014,** *9* (4), 1482-1494.

13. Wang, Q.; Miyakoda, M.; Yang, W.; Khillan, J.; Stachura, D. L.; Weiss, M. J.; Nishikura, K., Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene. *Journal of Biological Chemistry* **2004,** *279* (6), 4952-4961.

14. Hartner, J. C.; Schmittwolf, C.; Kispert, A.; Müller, A. M.; Higuchi, M.; Seeburg, P. H., Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. *Journal of Biological Chemistry* **2004,** *279* (6), 4894-4902.

15. Liddicoat, B. J.; Hartner, J. C.; Piskol, R.; Ramaswami, G.; Chalk, A. M.; Kingsley, P. D.; Sankaran, V. G.; Wall, M.; Purton, L. E.; Seeburg, P. H., Adenosine-to-inosine RNA editing by ADAR1 is essential for normal murine erythropoiesis. *Experimental hematology* **2016,** *44* (10), 947-963.

16. Shallev, L.; Kopel, E.; Feiglin, A.; Leichner, G. S.; Avni, D.; Sidi, Y.; Eisenberg, E.; Barzilai, A.; Levanon, E. Y.; Greenberger, S., Decreased A-to-I RNA editing as a source of keratinocytes' dsRNA in psoriasis. *RNA* **2018,** *24* (6), 828-840.

17. Roth, S. H.; Danan-Gotthold, M.; Ben-Izhak, M.; Rechavi, G.; Cohen, C. J.; Louzoun, Y.; Levanon, E. Y., Increased RNA editing may provide a source for autoantigens in systemic lupus erythematosus. *Cell reports* **2018,** *23* (1), 50-57.

18. Rice, G. I.; Kasher, P. R.; Forte, G. M.; Mannion, N. M.; Greenwood, S. M.; Szynkiewicz, M.; Dickerson, J. E.; Bhaskar, S. S.; Zampini, M.; Briggs, T. A., Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. *Nature genetics* **2012,** *44* (11), 1243-1248.

19. Maas, S.; Kawahara, Y.; Tamburro, K. M.; Nishikura, K., A-to-I RNA editing and human disease. *RNA biology* **2006,** *3* (1), 1-9.

20. Han, L.; Diao, L.; Yu, S.; Xu, X.; Li, J.; Zhang, R.; Yang, Y.; Werner, H. M. J.; Eterovic, A. K.; Yuan, Y.; Li, J.; Nair, N.; Minelli, R.; Tsang, Y. H.; Cheung, L. W. T.; Jeong, K. J.; Roszik, J.; Ju, Z.; Woodman, S. E.; Lu, Y.; Scott, K. L.; Li, J. B.; Mills, G. B.; Liang, H., The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer Cell* **2015,** *28* (4), 515-528.

21. Ishizuka, J. J.; Manguso, R. T.; Cheruiyot, C. K.; Bi, K.; Panda, A.; Iracheta-Vellve, A.; Miller, B. C.; Du, P. P.; Yates, K. B.; Dubrot, J., Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. *Nature* **2018,** *565* (7737), 43-48.

22. Nishikura, K., Functions and regulation of RNA editing by ADAR deaminases. *Annual review of biochemistry* **2010,** *79*, 321-349.

23. Mehler, M. F.; Mattick, J. S., Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiological reviews* **2007,** *87* (3), 799-823.

24. Wright, A. L.; Vissel, B., The essential role of AMPA receptor GluR2 subunit RNA editing in the normal and diseased brain. *Frontiers in molecular neuroscience* **2012,** *5*, 34.

25. Li, J. B.; Church, G. M., Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. *Nature neuroscience* **2013,** *16* (11), 1518.

26. Hwang, T.; Park, C.-K.; Leung, A. K.; Gao, Y.; Hyde, T. M.; Kleinman, J. E.; Rajpurohit, A.; Tao, R.; Shin, J. H.; Weinberger, D. R., Dynamic regulation of RNA editing in human brain development and disease. *Nature neuroscience* **2016,** *19* (8), 1093.

27. Tran, S. S.; Jun, H.-I.; Bahn, J. H.; Azghadi, A.; Ramaswami, G.; Van Nostrand, E. L.; Nguyen, T. B.; Hsiao, Y.-H. E.; Lee, C.; Pratt, G. A., Widespread RNA editing dysregulation in brains from autistic individuals. *Nature neuroscience* **2019,** *22* (1), 25.

28. Breen, M. S.; Dobbyn, A.; Li, Q.; Roussos, P.; Hoffman, G. E.; Stahl, E.; Chess, A.; Sklar, P.; Li, J. B.; Devlin, B., Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia. *Nature neuroscience* **2019,** *22* (9), 1402-1412.

29. Khermesh, K.; D'Erchia, A. M.; Barak, M.; Annese, A.; Wachtel, C.; Levanon, E. Y.; Picardi, E.; Eisenberg, E., Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *Rna* **2016,** *22* (2), 290-302.

30. Silvestris, D. A.; Picardi, E.; Cesarini, V.; Fosso, B.; Mangraviti, N.; Massimi, L.; Martini, M.; Pesole, G.; Locatelli, F.; Gallo, A., Dynamic inosinome profiles reveal novel patient stratification and gender-specific differences in glioblastoma. *Genome biology* **2019,** *20* (1), 33.

31. Paz, N.; Levanon, E. Y.; Amariglio, N.; Heimberger, A. B.; Ram, Z.; Constantini, S.; Barbash, Z. S.; Adamsky, K.; Safran, M.; Hirschberg, A.; Krupsky, M.; Ben-Dov, I.; Cazacu, S.; Mikkelsen, T.; Brodie, C.; Eisenberg, E.; Rechavi, G., Altered adenosine-to-inosine RNA editing in human cancer. *Genome Res* **2007,** *17* (11), 1586-95.

32. Chan, T. H.; Lin, C. H.; Qi, L.; Fei, J.; Li, Y.; Yong, K. J.; Liu, M.; Song, Y.; Chow, R. K.; Ng, V. H.; Yuan, Y. F.; Tenen, D. G.; Guan, X. Y.; Chen, L., A disrupted RNA editing balance mediated by ADARs (Adenosine DeAminases that act on RNA) in human hepatocellular carcinoma. *Gut* **2014,** *63* (5), 832-43.

33. Paz-Yaacov, N.; Bazak, L.; Buchumenski, I.; Porath, H. T.; Danan-Gotthold, M.; Knisbacher, B. A.; Eisenberg, E.; Levanon, E. Y., Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors. *Cell Rep* **2015,** *13* (2), 267-76.

34. Vlachogiannis, N. I.; Gatsiou, A.; Silvestris, D. A.; Stamatelopoulos, K.; Tektonidou, M. G.; Gallo, A.; Sfikakis, P. P.; Stellos, K., Increased adenosine-to-inosine RNA editing in rheumatoid arthritis. *Journal of autoimmunity* **2020,** *106*, 102329.

35. Hideyama, T.; Yamashita, T.; Aizawa, H.; Tsuji, S.; Kakita, A.; Takahashi, H.; Kwak, S., Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. *Neurobiology of disease* **2012,** *45* (3), 1121-1128.

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

37. Oakes, E.; Vadlamani, P.; Hundley, H. A., Methods for the Detection of Adenosine-to-Inosine Editing Events in Cellular RNA. In *mRNA Processing*, Springer: 2017; pp 103-127.

38. Roth, S. H.; Levanon, E. Y.; Eisenberg, E., Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nature methods* **2019**, 1-8.

39. Zhang, R.; Li, X.; Ramaswami, G.; Smith, K. S.; Turecki, G.; Montgomery, S. B.; Li, J. B., Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nature methods* **2014,** *11* (1), 51.

40. Giudice, C. L.; Silvestris, D. A.; Roth, S. H.; Eisenberg, E.; Pesole, G.; Gallo, A.; Picardi, E., Quantifying RNA Editing in Deep Transcriptome Datasets. *Frontiers in Genetics* **2020,** *11*.

41. Su, D.; Chan, C. T.; Gu, C.; Lim, K. S.; Chionh, Y. H.; McBee, M. E.; Russell, B. S.; Babu, I. R.; Begley, T. J.; Dedon, P. C., Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nature protocols* **2014,** *9* (4), 828.

42. Zhang, N.; Shi, S.; Jia, T. Z.; Ziegler, A.; Yoo, B.; Yuan, X.; Li, W.; Zhang, S., A general LC-MS-based RNA sequencing method for direct analysis of multiple-base modifications in RNA mixtures. *Nucleic acids research* **2019,** *47* (20), e125-e125.

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

44. Popova, A. M.; Williamson, J. R., Quantitative analysis of rRNA modifications using stable isotope labeling and mass spectrometry. *Journal of the American Chemical Society* **2014,** *136* (5), 2058-2069.

45. Laverdiere, I.; Caron, P.; Couture, F.; Guillemette, C.; Lévesque, E. r., Liquid chromatography–coupled tandem mass spectrometry based assay to evaluate inosine-5′-monophosphate dehydrogenase activity in peripheral blood mononuclear cells from stem cell transplant recipients. *Analytical chemistry* **2012,** *84* (1), 216-223.

46. Jimmerson, L. C.; Bushman, L. R.; Ray, M. L.; Anderson, P. L.; Kiser, J. J., A LC-MS/MS method for quantifying adenosine, guanosine and inosine nucleotides in human cells. *Pharmaceutical research* **2017,** *34* (1), 73-83.

47. Fritzell, K.; Xu, L.-D.; Otrocka, M.; Andréasson, C.; Öhman, M., Sensitive ADAR editing reporter in cancer cells enables high-throughput screening of small molecule libraries. *Nucleic acids research* **2019,** *47* (4), e22-e22.

48. Yao, M.; Hatahet, Z.; Melamede, R. J.; Kow, Y. W., Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3'endonuclease, from Escherichia coli. *Journal of Biological Chemistry* **1994,** *269* (23), 16260-16268.

49. Hitchcock, T. M.; Gao, H.; Cao, W., Cleavage of deoxyoxanosine-containing oligodeoxyribonucleotides by bacterial endonuclease V. *Nucleic acids research* **2004,** *32* (13), 4071-4080.

50. Dalhus, B.; Arvai, A. S.; Rosnes, I.; Olsen, Ø. E.; Backe, P. H.; Alseth, I.; Gao, H.; Cao, W.; Tainer, J. A.; Bjørås, M., Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nature structural & molecular biology* **2009,** *16* (2), 138.

51. Morita, Y.; Shibutani, T.; Nakanishi, N.; Nishikura, K.; Iwai, S.; Kuraoka, I., Human endonuclease V is a ribonuclease specific for inosine-containing RNA. *Nature communications* **2013,** *4*, 2273.

52. Vik, E. S.; Nawaz, M. S.; Andersen, P. S.; Fladeby, C.; Bjørås, M.; Dalhus, B.; Alseth, I., Endonuclease V cleaves at inosines in RNA. *Nature communications* **2013,** *4*, 2271.

53. Wu, J.; Samara, N. L.; Kuraoka, I.; Yang, W., Evolution of Inosine-Specific Endonuclease V from Bacterial DNase to Eukaryotic RNase. *Molecular cell* **2019,** *76* (1), 44-56. e3.

54. Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M., Selective enrichment of A-to-I edited transcripts from cellular RNA using Endonuclease V. *Journal of the American Chemical Society* **2020,** *142* (11), 5241-5251.

55. Knutson, S. D.; Heemstra, J. M., EndoVIPER-seq for Improved Detection of A-to-I Editing Sites in Cellular RNA. *Current Protocols in Chemical Biology* **2020,** *12* (2), e82.

56. Wienken, C. J.; Baaske, P.; Rothbauer, U.; Braun, D.; Duhr, S., Protein-binding assays in biological liquids using microscale thermophoresis. *Nature communications* **2010,** *1* (1), 1-7.

57. Jerabek-Willemsen, M.; André, T.; Wanner, R.; Roth, H. M.; Duhr, S.; Baaske, P.; Breitsprecher, D., MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure* **2014,** *1077*, 101-113.

58. Nishikura, K., A-to-I editing of coding and non-coding RNAs by ADARs. *Nature reviews Molecular cell biology* **2016,** *17* (2), 83.

59. Nakaya, K.; Takenaka, O.; Horinishi, H.; Shibata, K., Reactions of glyoxal with nucleic acids, nucleotides and their component bases. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis* **1968,** *161* (1), 23-31.

60. Qiu, C.; Liu, W.-Y.; Xu, Y.-Z., Fluorescence labeling of short RNA by oxidation at the 3′-end. In *RNA Nanotechnology and Therapeutics*, Springer: 2015; pp 113-120.

61. Clark, M. F.; Lister, R. M.; Bar-Joseph, M., ELISA techniques. *Methods in enzymology* **1986,** *118*, 742-766.

62. Engvall, E.; Perlmann, P., Enzyme-linked immunosorbent assay, ELISA: III. Quantitation of specific antibodies by enzyme-labeled anti-immunoglobulin in antigen-coated tubes. *The Journal of Immunology* **1972,** *109* (1), 129-135.

63. GeurtsvanKessel, C. H.; Okba, N. M.; Igloi, Z.; Bogers, S.; Embregts, C. W.; Laksono, B. M.; Leijten, L.; Rokx, C.; Rijnders, B.; Rahamat-Langendoen, J., An evaluation of COVID-19 serological assays informs future diagnostics and exposure assessment. *Nature communications* **2020,** *11* (1), 1-5.

64. Baran, P.; Hansen, S.; Waetzig, G. H.; Akbarzadeh, M.; Lamertz, L.; Huber, H. J.; Ahmadian, M. R.; Moll, J. M.; Scheller, J., The balance of interleukin (IL)-6, IL-6· soluble IL-6 receptor (sIL-6R), and IL-6· sIL-6R· sgp130 complexes allows simultaneous classic and trans-signaling. *Journal of Biological Chemistry* **2018,** *293* (18), 6762-6775.

65. De La Rica, R.; Stevens, M. M., Plasmonic ELISA for the ultrasensitive detection of disease biomarkers with the naked eye. *Nature nanotechnology* **2012,** *7* (12), 821-824.

66. Kobayashi, S.; Sakae, K.; Suzuki, Y.; Ishiko, H.; Kamata, K.; Suzuki, K.; Natori, K.; Miyamura, T.; Takeda, N., Expression of recombinant capsid proteins of chitta virus, a genogroup II Norwalk virus, and development of an ELISA to detect the viral antigen. *Microbiology and immunology* **2000,** *44* (8), 687-693.

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

68. Matsuzawa, S.; Wakata, Y.; Ebi, F.; Isobe, M.; Kurosawa, N., Development and validation of monoclonal antibodies against N6-methyladenosine for the detection of RNA modifications. *PloS one* **2019,** *14* (10), e0223197.

69. Tobias, I. C.; Kao, M.-M. C.; Parmentier, T.; Hunter, H.; LaMarre, J.; Betts, D. H., Targeted expression profiling reveals distinct stages of early canine fibroblast reprogramming are regulated by 2-oxoglutarate hydroxylases. *Stem cell research & therapy* **2020,** *11* (1), 1-23.

70. Chowdhury, B.; Cho, I.-H.; Hahn, N.; Irudayaraj, J., Quantification of 5-methylcytosine, 5-hydroxymethylcytosine and 5-carboxylcytosine from the blood of cancer patients by an enzyme-based immunoassay. *Analytica chimica acta* **2014,** *852*, 212-217.

71. Sommer, S. S.; Cohen, J. E., The size distributions of proteins, mRNA, and nuclear RNA. *Journal of molecular evolution* **1980,** *15* (1), 37-57.

72. Gerber, A. P.; Keller, W., An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **1999,** *286* (5442), 1146-1149.

73. Torres, A. G.; Piñeyro, D.; Filonava, L.; Stracker, T. H.; Batlle, E.; de Pouplana, L. R., A-to-I editing on tRNAs: biochemical, biological and evolutionary implications. *FEBS letters* **2014,** *588* (23), 4279-4286.

74. Torres, A. G.; Piñeyro, D.; Rodríguez-Escribà, M.; Camacho, N.; Reina, O.; Saint-Léger, A.; Filonava, L.; Batlle, E.; Ribas de Pouplana, L., Inosine modifications in human tRNAs are incorporated at the precursor tRNA level. *Nucleic acids research* **2015,** *43* (10), 5145-5157.

75. Bazak, L.; Levanon, E. Y.; Eisenberg, E., Genome-wide analysis of Alu editability. *Nucleic acids research* **2014,** *42* (11), 6876-6884.

76. Schaffer, A. A.; Kopel, E.; Hendel, A.; Picardi, E.; Levanon, E. Y.; Eisenberg, E., The cell line A-to-I RNA editing catalogue. *Nucleic acids research* **2020,** *48* (11), 5849-5858.

77. Galipon, J.; Ishii, R.; Suzuki, Y.; Tomita, M.; Ui-Tei, K., Differential binding of three major human ADAR isoforms to coding and long non-coding transcripts. *Genes* **2017,** *8* (2), 68.

78. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N., The genotype-tissue expression (GTEx) project. *Nature genetics* **2013,** *45* (6), 580-585.

79. Han, L.; Liang, H., RNA editing in cancer: mechanistic, prognostic, and therapeutic implications. *Molecular & cellular oncology* **2016,** *3* (2), e1117702.

80. Kung, C.-P.; Maggi Jr, L. B.; Weber, J. D., The role of RNA editing in cancer development and metabolic disorders. *Frontiers in endocrinology* **2018,** *9*, 762.

81. Fumagalli, D.; Gacquer, D.; Rothé, F.; Lefort, A.; Libert, F.; Brown, D.; Kheddoumi, N.; Shlien, A.; Konopka, T.; Salgado, R., Principles governing A-to-I RNA editing in the breast cancer transcriptome. *Cell reports* **2015,** *13* (2), 277-289.

82. Kang, L.; Liu, X.; Gong, Z.; Zheng, H.; Wang, J.; Li, Y.; Yang, H.; Hardwick, J.; Dai, H.; Poon, R. T.; Lee, N. P.; Mao, M.; Peng, Z.; Chen, R., Genome-wide identification of RNA editing in hepatocellular carcinoma. *Genomics* **2015,** *105* (2), 76-82.

83. Anadón, C.; Guil, S.; Simó-Riudalbas, L.; Moutinho, C.; Setien, F.; Martínez-Cardús, A.; Moran, S.; Villanueva, A.; Calaf, M.; Vidal, A., Gene amplification-associated overexpression of the RNA editing enzyme ADAR1 enhances human lung tumorigenesis. *Oncogene* **2016,** *35* (33), 4407-4413.

84. Weinstein, J. N.; Collisson, E. A.; Mills, G. B.; Shaw, K. R. M.; Ozenberger, B. A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J. M., The cancer genome atlas pan-cancer analysis project. *Nature genetics* **2013,** *45* (10), 1113-1120.

85. Shalem, O.; Sanjana, N. E.; Hartenian, E.; Shi, X.; Scott, D. A.; Mikkelsen, T. S.; Heckl, D.; Ebert, B. L.; Root, D. E.; Doench, J. G., Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **2014,** *343* (6166), 84-87.

Chapter 6

Thermoreversible Control of Nucleic Acid Structure and Function with Glyoxal Caging[1*]

6.1 Abstract

Controlling the structure and activity of nucleic acids expands their utility in therapeutics, nanotechnology and biocomputing. Several methods have been developed to impart responsiveness of DNA and RNA to small-molecule and light-based stimuli. However, heat-triggered control of nucleic acids remains largely unexplored, leaving a significant gap in responsive nucleic acid technology. Current technologies are also limited to natural nucleic acids and are often incompatible with polymerase-generated sequences. Here we show that glyoxal, a well-characterized compound that covalently attaches to the Watson-Crick-Franklin face of several nucleobases, addresses these limitations by thermoreversibly modulating the structure and activity of virtually any nucleic acid scaffold. Using a variety of constructs, we demonstrate that glyoxal is easily installed and potently disrupts nucleic acid structure and function. We also characterize the kinetics of decaging and show that activity can be restored via tunable thermal removal of glyoxal adducts. We further illustrate the versatility of this approach by reversibly caging a 2'-*O*-methylated RNA aptamer as well as synthetic threose nucleic acid (TNA) and peptide nucleic acid (PNA) scaffolds. We also show caging of a guide RNA for tunable and reversible control over CRISPR-Cas9 activity, and demonstrate glyoxal caging as an effective method for enhancing PCR specificity. We further cage an antisense oligonucleotide for time-release activation and titration of gene expression in living cells. Together, glyoxalation is a straightforward method for imparting reversible thermal responsiveness to any nucleic acid architecture, addressing a significant need in synthetic biology and offering a versatile new tool for constructing programmable nucleic acid components.

---

*Adapted from Ref. 1 with permission from Knutson, S. D.; Sanford, A. A.; Swenson, C. S.; Korn, M. K.; Manuel, B. A.; Heemstra, J. M. *Journal of the American Chemical Society*. **2020**, *142*(41), 17766–17781.

6.2 Introduction

Nucleic acids are highly versatile and dynamic biomolecules, exhibiting large data storage capacity and a high degree of structural complexity, in turn endowing them with diverse molecular recognition and catalytic activities. These attractive properties explain their ubiquitous presence in Nature and warrant investigation of nucleic acids as programmable synthetic components in biomedicine, nanotechnology and biocomputing.[2-5] Just as Nature engages temporal and spatial regulation to control gene expression, achieving tunable and predictable control over the structure and activity of nucleic acid constructs *in vitro* is vital toward their use in synthetic biology and biotechnology applications.

Several approaches have been explored to impart stimuli-responsive properties to nucleic acids and facilitate external control over their structure and function. Early methods to achieve chemically-triggered activation of nucleic acids incorporated trichloroethyl[6] or 4-nitrobenzyl[7] groups during solid-phase oligonucleotide synthesis. Resulting adducts prevent duplex formation and higher-order assembly, and can be subsequently removed by reducing agents, including zinc/acetic acid mixtures or sodium thiosulfate. Photoreversible blocking strategies have also been developed.[8-15] Similar to these previous methods, “photocaged” nucleobase derivatives are incorporated into oligonucleotide synthesis workflows, and can be uncaged with UV light to restore nucleic acid function. While effective, these protecting groups must be incorporated during oligonucleotide synthesis and are thus only applicable to relatively short strands. Moreover, the difficulty and cost of synthesizing these modified phosphoramidite monomers can be limiting, and thus it is essential to have information about the key residues to be caged in order to mask the activity of the sequence. More recent “cloaking” methods employ acylation of 2’ hydroxyls in RNA with either an azide-containing or *o*-nitroveratryl photoreactive reagent, which are then respectively removed by triphosphine-mediated Staudinger reduction or light-triggered decaging.[16-17] While these methods allow caging of both chemically and enzymatically synthesized transcripts, these acylation reagents can only be applied to RNA, and no comparable

methods are available for other nucleic acid scaffolds. Additionally, achieving full reactivation of cloaked RNA constructs can be challenging,[16-17] and UV removal of adducts in particular can both impart structural changes to certain nucleobases[18] and further require specialized and costly instrumentation.

In contrast to chemical or light-driven approaches, we envisioned that temperature could offer a straightforward external stimulus for controlling the caging of nucleic acids. Heat is relatively easy to introduce in a variety of settings, and a number of laboratory instruments that precisely control temperature are both commonplace and economical. Despite these advantages, heat-sensitive caging of nucleic acids remains largely unexplored. To address this need, we recognized that glyoxal could serve as an effective thermoresponsive caging group. Glyoxal reacts readily with nitrogen groups on the Watson-Crick-Franklin face of nucleobases to produce stable *bis*-hemiaminal adducts, and this reaction has been demonstrated in both RNA and DNA. Guanosine addition is typically the most stable and rapidly formed product, but longer reaction times (>30 min) and higher glyoxal concentrations (>0.1 M) can also lead to reaction of adenosine and cytidine (**Figure 6.1a**).[19-23] Adduct formation directly interferes with base pairing, and thus denatures overall secondary structure in nucleic acids. Attesting to the convenience and efficacy of this reaction, glyoxal has been utilized for several decades in electrophoresis analysis and structural probing of large RNAs,[24-25] and glyoxal can be used as a potent cell and tissue fixative prior to immunohistochemical staining.[26] Kethoxal, a chemical derivative of glyoxal, has also found broad utility for high-throughput mapping of RNA secondary structure.[27]

Key to use of glyoxal as a nucleic acid caging reagent is the ability to reverse adduct formation and restore base pairing. Fortuitously, glyoxal can be easily deprotected through a combination of heat and mild alkaline conditions (**Figure 6.1b**).[28] Taken together, the previous applications of glyoxal highlight several unique properties that we recognized could be leveraged for thermoresponsive caging: (1) glyoxalation is sufficient to disrupt strong nucleic acid secondary structure, (2) both installation and removal of glyoxal adducts proceed under mild reaction

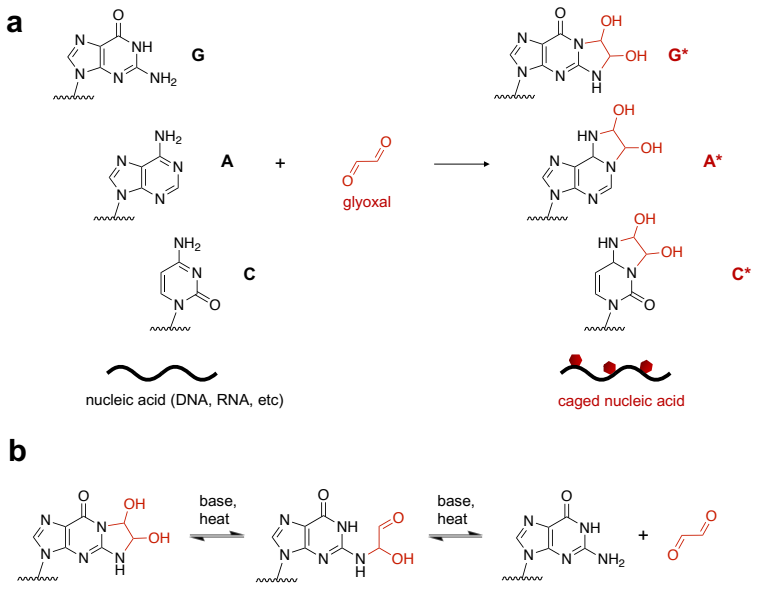


Figure 6.1. Glyoxal reactivity with nucleobases. a) Glyoxal forms *bis*-hemiaminal adducts (*) on guanine (G), adenosine (A), and cytidine (C) nucleobases. b) Glyoxal decaging occurs upon addition of heat under mild alkaline conditions.

conditions, (3) glyoxal reactivity towards nucleobases rather than backbone moieties suggests a general approach to cage a variety of different nucleic acid scaffolds including xenonucleic acids (XNAs), (4) glyoxal caging can be performed on both synthetic and naturally occurring nucleic acids, and (5) the reagent itself is both economical and commercially available.

Herein we explore the use of glyoxal for thermoresponsive control over nucleic acid structure and function. We show that glyoxal can effectively cage a wide range of natural and synthetic nucleic acid polymers, providing control over multiple functions including small-molecule binding, catalysis, antisense hybridization, interactions with proteins, and gene expression modulation in cells. Glyoxal caging is efficient and fully thermoresponsive, enabling reduction of function to undetectable levels and full reactivation. Moreover, restoration of function is tunable by temperature, pH, incubation time, and degree of caging. Together, this research addresses a significant gap in nucleic acid technology by providing a simple and effective way to impart thermoresponsive control over both chemically synthesized and enzymatically generated nucleic

acids. Additionally, glyoxal caging is applicable in virtually any nucleic acid scaffold, offering a versatile methodology for designing and implementing stimuli-responsive biomaterials with user-defined chemical architectures.

6.3 Results and Discussion

We first sought to estimate glycation kinetics in a simple DNA sequence (**Figure 6.2a**). Using previously established reaction conditions (1.3 M glyoxal in 50:50 DMSO:H_2O, 50 ºC)[29-30] we monitored molecular weight of our test strand using 20% denaturing PAGE and observed apparent full caging in as little as 30-40 minutes with an approximate $t_{1/2}$ of ~9 minutes (**Figure 6.2a**). We were next interested in defining conditions for removing these glyoxal groups, so we first exposed a fully caged strand to standard deprotection conditions (phosphate buffered saline pH 7.5, 70 °C)[30] and again monitored molecular weight shifts as a function of time. As shown in **Figure 6.2b**, we observed rapid removal of these groups with apparent full decaging in 30 minutes ($t_{1/2}$ ~11 minutes).

We next explored the effect of both pH and temperature parameters in a larger screen and estimated decaging half-life across these conditions (**Table 6.1, Figures E1-E4**). In these experiments, we observed that decaging rates were positively correlated to both pH and temperature. Excitingly, this provides a tunable deprotection rate in a variety of assay conditions and highlights the potential of glyoxal caging for constructing programmable nucleic acid-based clocks, thermometers, and pH responsive elements in nanodevices and biological circuits.[31-38] We also assessed room temperature stability of glyoxal adducts and observed minimal decaging over several days, indicating that glycation is stable in practical laboratory conditions until deprotection is triggered by elevated temperatures at the desired time (**Figure E5**). Although these PAGE experiments provide only a qualitative analysis of our model DNA strand, they also illustrate a general kinetic framework of glyoxal caging and decaging. Together, these results demonstrated

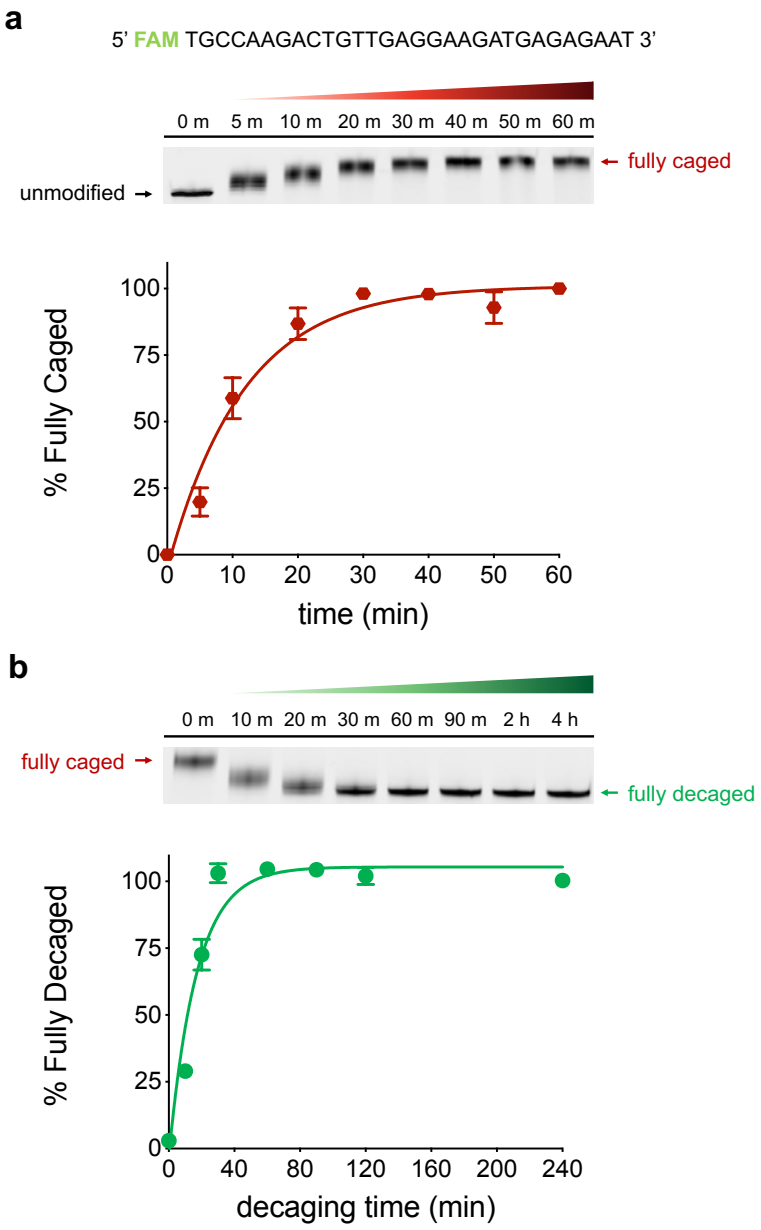


Figure 6.2. Glyoxal caging and decaging of a DNA substrate. a) Caging kinetics were monitored by treating 200 pmol of a FAM-labeled DNA substrate with 1.3 M glyoxal in 50:50 DMSO:H_2O at 50 ºC. At the indicated timepoints, reactions were analyzed using 20% denaturing PAGE. Mean values (n = 2) represent normalized percentages versus band intensity at t = 60 min with S.D. error bars. b) To estimate decaging, 20 pmol of a fully caged DNA substrate (1 h glyoxal treatment time) was incubated at 95 ºC, pH 7.5 for the indicated timepoints and analyzed using 20% denaturing PAGE. Values represent mean (n = 2) normalized percentages versus band intensity of an untreated DNA control. Error bars denote S.D.

	pH 6.5	pH 7.0	pH 7.5	pH 8.0
95 ºC	1.4 ± 0.34 min	1.1 ± 0.24 min	< 1 min	<< 1 min
70 ºC	52 ± 7.7 min	14 ± 2.7 min	11 ± 2.4 min	3.8 ± 1.6 min
50 ºC	14 ± 1.8 h	8.0 ± 1.7 h	4.6 ± 1.1 h	3.1 ± 0.70 h
37 ºC	14 ± 4.2 d	6.6 ± 1.4 d	3.0 ± 0.5 d	2.0 ± 0.33 d

Table 6.1. Half-life for glyoxal decaging with varying pH and temperature.

that glyoxal cage installation and removal were thermoreversible and tunable, and we turned to testing modulation of a wide range of functional nucleic acids.

As an initial target to functionally test our method, we selected the “broccoli” aptamer, a well-characterized fluorogenic RNA that binds the small molecule chromophore 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) to produce a fluorescent signal (**Figure 6.3a**).[39] We hypothesized that only partial caging would be sufficient to disrupt RNA folding and fluorogenic activity, so we first subjected the aptamer to increasing caging times and functionally tested fluorescence. We observed that only ~5 minutes of caging time reduced activity by > 98%, with complete loss of fluorescence at 10 minutes (**Figure E6**). While additional reaction times produced proportional increases in molecular weight corresponding to further glyoxalation (**Figure E6a**), these alterations provided no added functional benefit. Interestingly, we also noted that increased glycation resulted in proportional losses in SYBR Gold staining intensity, further highlighting potent disruption of nucleic acid and small molecule interactions. We utilized our minimal caging time (10 minutes) and subjected these constructs to rapid decaging conditions at 95 ºC, pH 7.5 with increasing reaction times (**Figure E7**). As expected, we observed a quick rise in signal from 0-2 minutes, and achieved complete restoration of fluorogenic activity after 2 minutes of decaging. We also observed a rapid drop in fluorescence enhancement with extended incubation times, which is likely due to base-mediated RNA hydrolysis at high temperatures,

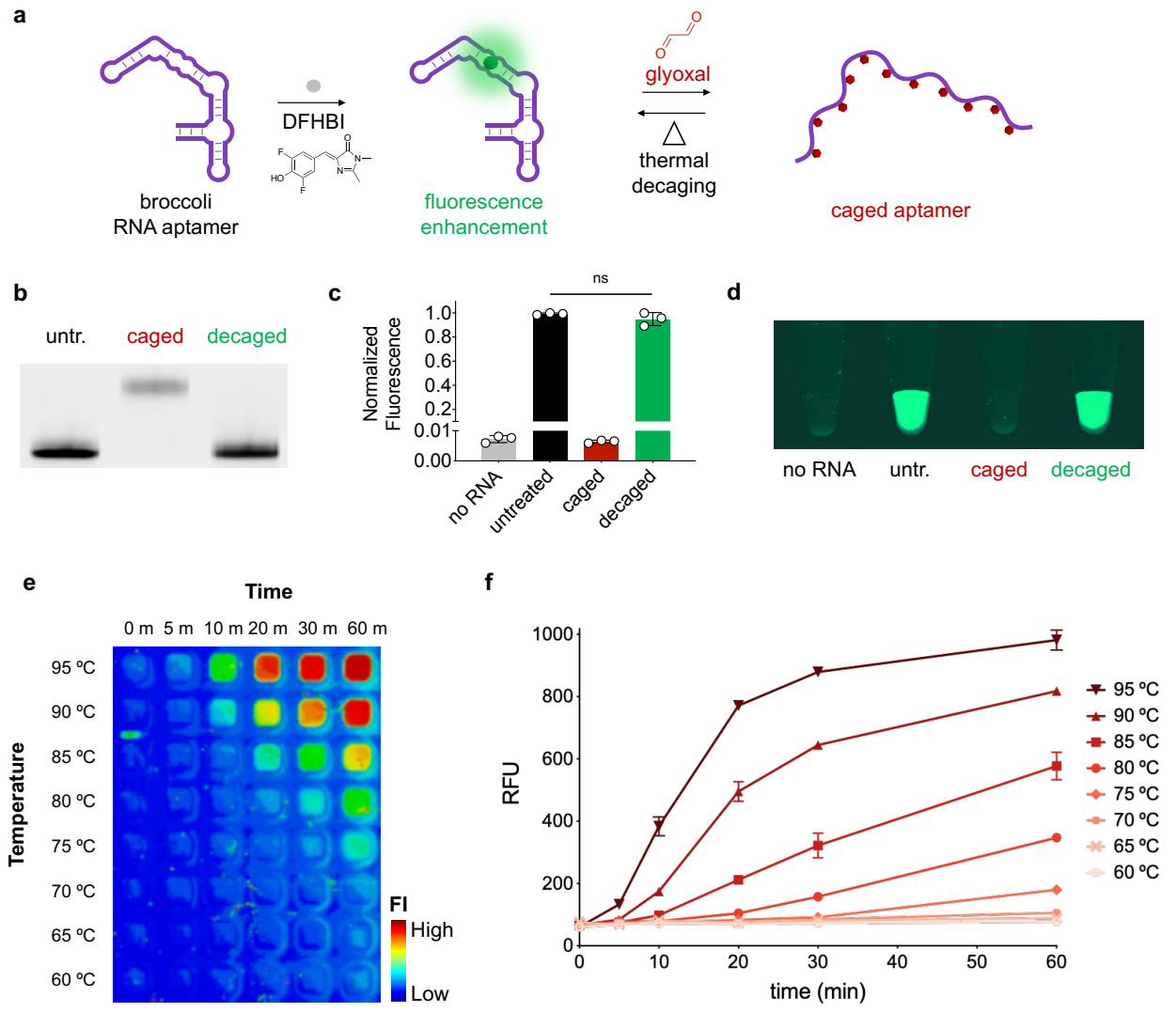


Figure 6.3. Glyoxal caging imparts thermoresponsive fluorogenic activity in the broccoli RNA aptamer. a) Schematic of the fluorogenic broccoli RNA aptamer. Glyoxal caging reversibly denatures and cages the aptamer, preventing fluorescent signal generation. b) 20% denaturing PAGE analysis of untreated (untr.), caged, and decaged RNA aptamers. 60 pmol (2 µg) of RNA aptamer was first treated with 1.3 M glyoxal in 50:50 DMSO:H_2O, 50 ºC for 10 minutes. To decage the aptamer, 20 pmol of caged RNA (10 minutes glyoxal treatment time) was incubated at 95 ºC, pH 7.5 for 2 minutes. c) Quantified fluorescence enhancement of untreated (untr.), caged, and decaged broccoli aptamers. 20 pmol of untreated, minimally caged (10 minute glyoxal treatment time), or decaged aptamer (2 minutes at 95 ºC, pH 7.5) was combined with 2 µM 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), 40 mM HEPES, 100 mM KCl, and 1 mM $MgCl_2$, pH 7.4 and incubated for 30 minutes at 37 °C prior to fluorescent measurement. Bars represent mean and S.D. from triplicate binding reactions. Unpaired t-test was performed between untreated and decaged samples. “ns” indicates no significant difference. d) Fluorescence enhancement reactions visualized with a typhoon gel imager. e) Minimally caged broccoli (10 minutes glyoxal treatment time) was combined with 2 µM DFHBI in 40 mM HEPES, 100 mM KCl, and 1 mM $MgCl_2$, pH 7.4 and incubated at various temperatures for increasing times. Reactions in 384-well plates were then visualized using a Typhoon gel imager. Heat map values represent fluorescent intensity generated using the acquisition software (FI = fluorescent intensity). f) Quantified broccoli aptamer fluorescence plotted against temperature and incubation time. Values represent mean and S.D. of 2 independent trials.

consistent with previously observed degradation of riboswitches in similar buffer conditions at extended heating times.[40] Using our optimized conditions (10 minutes glyoxal caging, 2 minutes decaging at 95 ºC, pH 7.5) we demonstrated both a full masking and complete restoration of broccoli fluorescence enhancement using glyoxal caging (**Figures 6.3b-d**). Together, these data demonstrate that glyoxalation is a rapid, potent, and fully reversible method to inhibit small molecule binding in aptamers, and is comparable to current caging methods which achieve similar modulation of activity.[16-17]

Given the simple fluorescent readout of the broccoli system along with our previous characterizations showing tunable thermal decaging kinetics, we next tested our ability to deprotect glyoxal adducts *in situ* for use as a one-pot molecular "thermometer" or "clock". In particular, accurate thermal sensing at the nanoscale is vital for developing new diagnostic and biocomputing platforms, and we envision that embedding caged fluorogenic aptamers into these materials presents a facile and economical strategy for monitoring thermal fluctuations in nanoconstructs. Additionally, compared to existing DNA-based molecular thermometers which largely rely on bulk-level analyses, the use of caged fluorogenic aptamers enables signal detection at single-molecule resolution, and incorporating caged fluorogenic aptamers into different materials will enable spatial sensing of temperature gradients across surfaces and tissues. Similarly, given the tunable deprotection kinetics we observed in a model DNA strand (**Table 6.1**), glyoxal-caged constructs can likely be implemented as "timers" and serve as accurate temporal recorders of nanoscale events when integrated into biomaterials or synthetic surfaces. Previous iterations of timers and thermometers have employed DNA hairpins,[33] nanoconstructs,[31, 34] and molecular beacons,[32] all of which introduce secondary structure changes to impart thermal sensitivity in nucleic acids and require extensive biophysical optimization. In comparison, our approach leverages glyoxal caging as a universal chemical means to modulate and convert a fluorogenic aptamer into a time-release thermosensitive biocomponent. We first combined minimally caged broccoli (10 minutes) with all necessary fluorogenic components (2 μM DFHBI,

40 mM HEPES, 100 mM KCl, 1 mM $MgCl_2$, pH 7.4), and then exposed these mixtures to a range of temperatures for increasing times. As shown in **Figures 6.3e-f**, we observed a highly predictable and proportional fluorogenic response to both of these variables, with optimal temporal linearity between ~80 ºC and ~85 ºC, and the most proportional thermal response following 60 minute incubation times. Overall, these results highlight the applicability of glyoxal-caged fluorogenic aptamers in designing molecular components in synthetic biology, including nucleic acid-based thermometers and timers. Further, these conditions are generalizable and can yield predictable changes in activity, allowing users to adjust the speed and intensity of responses by altering temperature, aptamer caging time, and/or buffer pH.

Encouraged by these results, we next sought to control the activity of a catalytic nucleic acid and chose a DNAzyme to demonstrate the versatility of our method across different nucleic acid backbones. Specifically, we employed the 10-23 DNAzyme which hybridizes to a ssDNA target and cleaves at an internal ribonucleotide (**Figure 6.4a**).[41] We first assessed caging kinetics of the 10-23 strand, observing expected increases in apparent size by denaturing PAGE (**Figure E8a**). We then tested the functional activity of increasingly caged DNAzyme samples and saw that target cleavage was significantly reduced after 20 minutes of caging, and completely ablated after 1 hour (**Figure E8b,c**). As expected, this structural shift was also reversible, and when we employed rapid decaging conditions (95 °C, pH 7.5) we observed restoration of the original size after 10 minutes (**Figure E9**). Using these parameters (20 minutes caging, 10 minutes decaging), we then demonstrated full reversibility of DNAzyme activity (**Figures 6.4b-d**). Untreated 10-23 DNAzyme cleaved ~45% of the target strand, whereas caged DNAzyme had no detectable activity. However, this activity was restored upon thermal decaging and displayed ~41% target cleavage with no significant difference compared to untreated DNAzyme. These results demonstrate that glyoxal caging is applicable to DNA substrates and can reversibly modulate catalytic cleavage activity. Moreover, we achieve comparable performance to alternative caging methods that require significantly more expensive reagents.[16-17, 42-43]

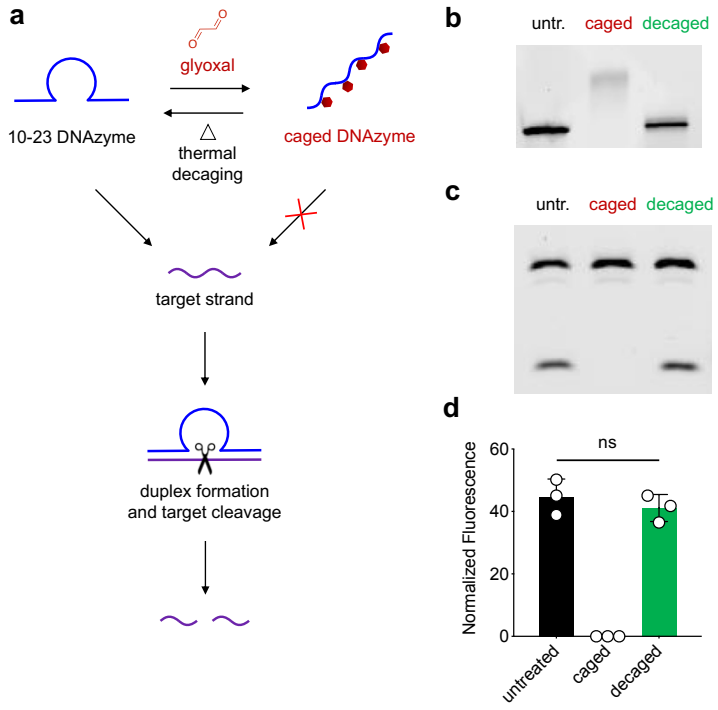


Figure 6.4. Reversible control of the RNA-cleaving 10-23 DNAzyme. a) Schematic of the 10-23 DNAzyme-mediated cleavage of the target strand following hybridization. Glyoxal reversibly inhibits hybridization and catalytic activity. b) 20% denaturing PAGE analysis of untreated (untr.), caged, and decaged 10-23 DNAzyme. 100 pmol of DNAzyme was first treated with 1.3 M glyoxal in 50:50 DMSO:H_2O, 50 ºC for 1 hour. To decage the DNAzyme, 10 pmol of caged DNA (1 hour glyoxal treatment time) was incubated at 95 ºC, pH 7.5 for 10 minutes. c,d) Functional activity of untreated (untr.), caged, and decaged 10-23 DNAzyme was tested by combining 15 pmol of untreated, caged, or decaged 10-23 DNAzyme with 1.5 pmol of a target DNA strand in 10 mM $MgCl_2$, 150 mM NaCl, 50 mM Tris-HCl, pH 7.5. Reactions was incubated for one hour at 37 °C, quenched with EDTA, and then analyzed with 12% denaturing PAGE. Percent target strand cleavage (n = 2) was quantified using band densitometry. Values represent mean and S.D. of 2 independent trials. Unpaired t-test was performed between untreated and decaged samples. "ns" indicates no significant difference.

After demonstrating reversible control of RNA and DNA constructs, we next sought to explore glyoxal caging of non-native XNA nucleic acid scaffolds. To test this, we first selected the ARC259 RNA aptamer targeting vascular endothelial growth factor (VEGF) (**Figure 6.5a**).[44] This particular aptamer was evolved using a fully 2′-*O*-methylated library and would enable us to not only assess disruption of aptamer-protein binding interactions, but also glyoxal caging on a heavily modified scaffold. We first verified that caging kinetics were similar with the ARC259 aptamer, and again saw a predictable rise in apparent molecular weight with increasing glyoxal caging times (**Figure E10a**). Next, we utilized fluorescence polarization (FP) to monitor ARC259-

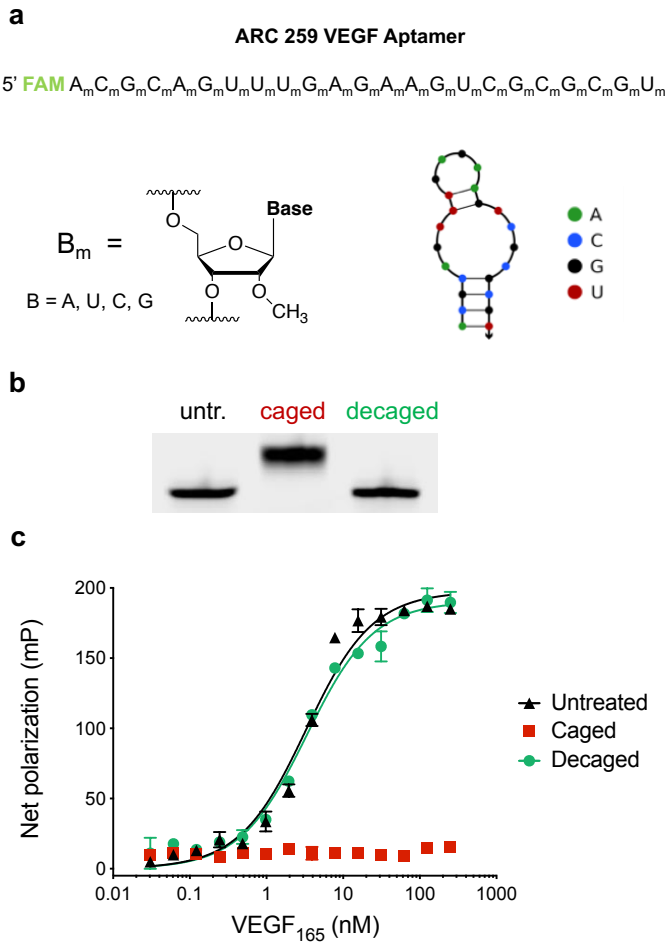


Figure 6.5. Glyoxal caging of a fully 2'-*O*-methylated RNA aptamer. a) Sequence and NUPACK[45] fold analysis of 2′-*O*-methylated ARC259 RNA aptamer. b) 20% PAGE analysis of untreated (untr.), caged and decaged ARC259 aptamer. 60 pmol of RNA aptamer was first treated with 1.3 M glyoxal in 50:50 DMSO:H_2O, 50 ºC for 40 minutes. To decage the aptamer, 20 pmol of caged ARC259 (40 minutes glyoxal treatment time) was incubated at 95 ºC, pH 7.5 for 5 minutes. c) Fluorescence polarization (FP) binding curves of untreated, caged, and decaged aptamer towards $VEGF_{165}$. A 10 nM solution of untreated, caged (40 minutes glyoxal treatment time), or decaged (5 minutes at 95 ºC, pH 7.5) ARC259 aptamer was combined with increasing amounts of recombinant human $VEGF_{165}$ and allowed to incubate at room temperature for 30 minutes. Binding reactions were then transferred to a 384-well black plate and fluorescence polarization was measured using Cytation 5 multi-mode plate reader. All values were normalized to a buffer blank and represent mean and S.D. of independent replicates (n = 3).

VEGF binding. Unsurprisingly, untreated aptamer displayed high affinity towards the target protein, and increasing caging times proportionally decreased this interaction, with ~40 minutes glyoxal treatment resulting in full loss of aptamer binding (**Figure E11**). We were next interested in restoring this activity, so we then subjected minimally caged ARC259 (40 minutes) to rapid decaging conditions (95 ºC, PBS pH 7.5) and observed a predicted drop in apparent molecular

weight in 20% denaturing PAGE with ~2-5 minutes decaging time (**Figure E10b**). Lastly, we compared the binding activity of untreated, caged, and decaged ARC259 in our FP assay (**Figure 6.5c**), and observed nearly identical binding affinities between untreated and decaged aptamers (K_D untreated = 3.27 ±0.59 nM, K_D decaged = 3.36 ±0.55 nM). Together, these results demonstrated the versatility of glyoxal caging towards chemically modified substrates, and further showed reversible control over protein-aptamer interactions. Additionally, to our knowledge, this is the first demonstration of post-synthetic caging of a 2' modified RNA aptamer, which has not been possible with previous cloaking methods that target 2'OH groups.[16-17]

Encouraged by the versatility of our approach, we next wanted to apply glyoxal caging toward non-canonical xenonucleic acid scaffolds. In particular, we targeted threose nucleic acid (TNA) and peptide nucleic acid (PNA). Compared to DNA or RNA, TNA is comprised of repeating threose sugars connected with alternating 2' to 3' phosphodiester bonds,[46] while PNA consists of amino ethyl glycine units to form a "peptide" rather than phosphodiester backbone (**Figure 6.6a**).[47] Due to their non-canonical chemical structures, both TNA and PNA are nuclease-resistant and are promising antisense oligo-nucleotide scaffolds. Additionally, these unique structural features are intriguing for demonstrating the flexibility of glyoxal caging with chemically modified substrates. As shown in **Figure E12**, we first assessed glyoxal caging kinetics with a model TNA oligonucleotide, and observed nearly identical kinetics compared to DNA. We next evaluated the effect of caging on hybridization between TNA and a DNA complement or scrambled oligonucleotide. We utilized microscale thermophoresis (MST) to measure duplex formation, as this method would preserve caging groups during analysis and has been previously demonstrated as a highly sensitive reporter of nucleic acid hybridization.[48-49] We observed steady disruption of duplex formation with increasing caging times on our TNA strand, with ~40 minutes of glyoxal treatment resulting in full ablation of hybridization capacity (**Figure E13**). Using these minimal caging conditions (40 minutes), we then used our rapid decaging protocol (5 min at 95 ºC in PBS,

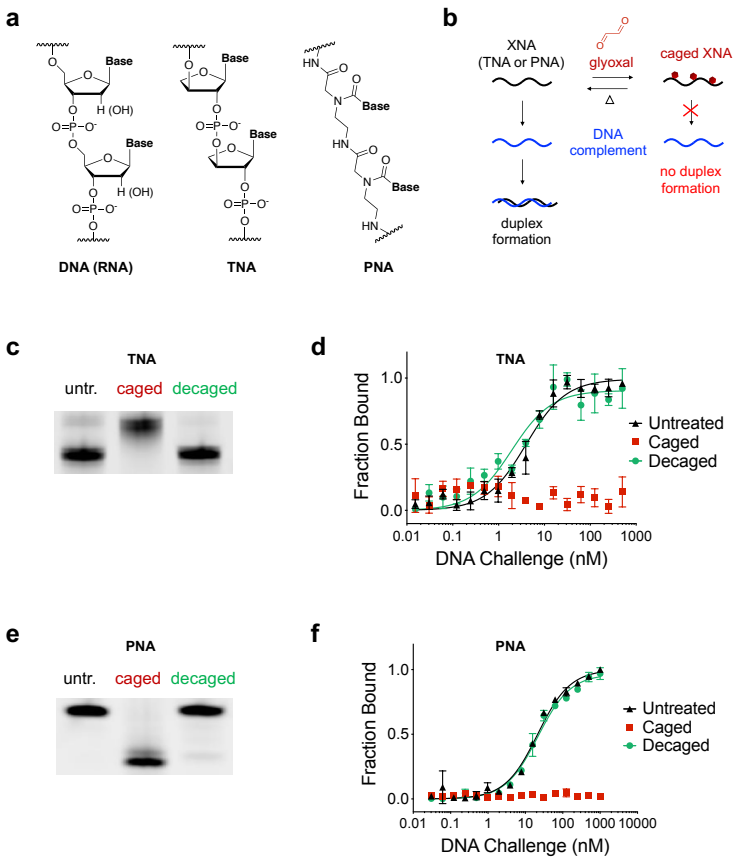


Figure 6.6. Glyoxal caging reversibly inactivates xenonucleic acid hybridization. a) Chemical structures of DNA/RNA alongside threose nucleic acid (TNA) and peptide nucleic acid (PNA) scaffolds. b) Heteroduplex formation between XNA strands and a DNA complement. Glyoxal caging reversibly inhibits duplex formation. c) 20% denaturing PAGE analysis of untreated (untr.), caged, and decaged TNA strands. 250 pmol of TNA strand was first treated with 1.3 M glyoxal in 50:50 DMSO:H_2O, 50 ºC for 40 minutes. To decage the TNA substrate, 20 pmol of caged strand (40 minutes glyoxal treatment time) was incubated at 95 ºC, pH 7.5 for 5 minutes. d) MST binding curves of untreated, caged (40 minutes glyoxal treatment time), and decaged (5 minutes at 95 ºC, pH 7.5) TNA strands when challenged with increasing amounts of DNA complement. Bars represent mean and S.E. from triplicate binding titration curves. e) 20% denaturing PAGE analysis of untreated (untr.), caged, and decaged PNA strands. 1 nmol of PNA was first treated with 1.3 M glyoxal in 50:50 DMSO:H_2O, 50 ºC for 20 minutes. To decage the PNA substrate, 20 pmol of caged strand (20 minutes glyoxal treatment time) was incubated at 95 ºC, pH 7.5 for 5 minutes. f) MST binding curves of untreated, caged, and decaged PNA strands when challenged with increasing amounts of DNA complement. Bars represent mean and S.E. from triplicate binding titration curves.

pH 7.5) to remove glyoxal adducts from TNA, observing full restoration in duplex formation (**Figure 6.6c,d**). To apply these experiments towards PNA, we then synthesized the "Nielsen decamer" sequence as a model strand (**Figure E14**).[50] Interestingly, when assessing glyoxal caging of PNA, we saw an unexpected drop in apparent molecular weight in 20% denaturing PAGE with increasing caging times (**Figure E16a**). While we were initially concerned about

possible degradation or hydrolysis, we did confirm an increase in mass corresponding to the addition of two glyoxal caging groups (**Figure E15**). Additionally, this effect was also reversible, and when we subjected a fully glyoxalated PNA strand to thermal decaging conditions at 95 ºC, pH 7.5, an opposite electrophoretic shift was observed, suggesting full removal of glyoxal groups (**Figure E16b**). Although uncertain, we hypothesize that glyoxal adducts may impart transient negative charges to the strand through cooperative hydrogen bonding between hydroxyl adducts to produce an ionizable proton (**Figure E17**). In the case of PNA where the only anion present is a terminal glutamate residue (**Figure E14a**), these alterations may contribute more significantly to electrophoretic shift than molecular weight changes. These analyses also yielded 3 visibly discrete bands, suggesting that guanosine addition products may be the predominant species in PNA glyoxal caging. When we tested hybridization capacity of increasingly caged PNA, we observed highly potent disruption following glyoxal treatment, observing a significant drop between 5 and 10 minutes caging time, and only 20 minutes required for full inhibition of duplex formation (**Figure E18**). Finally, we compared untreated, caged, and decaged PNA and again observed full restoration of duplex binding (**Figure 6.6e,f**). Together, our results show that glyoxal caging is a straightforward and robust method for reversibly modulating antisense interactions, and to our knowledge is the first demonstration of functional caging in non-canonical nucleic acid substrates.

Based on our results thus far showing reversible modulation of protein-nucleic acid interactions as well as the known molecular changes imparted by glyoxal addition on nucleobases, we were curious how caging would impact recognition by enzymes that interact with nucleic acid substrates (**Table 6.2**). RNase T1 is historically known to be affected by glyoxalation, which cleaves after both guanosine and inosine residues but is unable to digest caged guanosines, resulting in highly specific cleavage activity toward inosine.[29-30] We tested other enzymes starting with RNase H, which cleaves RNA substrates when hybridized to complementary DNA.[51] We first incubated a target RNA strand with or without glyoxal for 1 hour,

followed by hybridization to a complementary DNA strand and exposure to RNase H. Untreated heteroduplex was cleaved as expected, while caging of the ssRNA substrate prior to hybridization and digestion resulted in no detectable cleavage. Upon thermal decaging, full restoration of cleavage activity was achieved (**Figure E19**). Given our previous "one pot" thermoreversible readout using the fluorogenic broccoli RNA aptamer (**Figure 6.3e-f**), we were interested in recapitulating similar *in situ* decaging and cleavage using a thermostable RNase H.[52] To test this, we combined caged ssRNA, complementary DNA, and thermostable RNase H in 1X reaction buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, pH 8.3). Reactions were then separately heated to 95 °C for increasing time points, followed by one hour at 37 °C. As shown in Figure S20, we observed a steady and predictable increase in RNase H cleavage activity approaching 100% restoration over 10 minutes, indicative of proportional decaging of the target RNA strand with increased exposure to heat. These results were consistent with our previous "one pot" system measuring fluorogenic activity of the broccoli aptamer (**Figure 6.3e-f**), and we predict that both temperature and heat exposure time can be proportionally measured via RNase H cleavage and similarly leveraged towards construction of biological "timers" and "thermometers."

We next evaluated caging on the activity of RNase A, an endonuclease that cleaves single and double stranded RNA after cytidine and uridine nucleotides.[53] However, glyoxal addition was not able to inhibit activity, and distinct cleavage sites were present for all untreated, caged, and decaged RNAs (**Figure E21**). Similar results were observed with nuclease P1, a broadly active endonuclease that hydrolyzes phosphodiester bonds in 3′ → 5′ direction,[54] and glyoxal caging of ssDNA and ssRNA substrates still resulted in full digestion activity (**Figure E22**). We next tested DNase I, an endonuclease that preferentially cleaves double-stranded DNA substrates after pyrimidine nucleotides.[55] Similar to our previous experiments, we first treated a target DNA strand with or without glyoxal followed by hybridization to a complementary DNA and exposure to DNase I. As shown in Figure S23, glyoxal caging completely inhibited DNase I

Name	Type	Target	Inhibition?
RNase T1	Endonuclease	ssRNA after G residues	Partial*
RNase H	Endonuclease	RNA:DNA heteroduplexes	Yes
RNase A	Endonuclease	ssRNA, dsRNA	No
Nuclease P1	Endonuclease	ssRNA, ssDNA	No
DNase I	Endonuclease	ssDNA, dsDNA (preferred)	Yes
EcoRI	Endonuclease	dsDNA at 5' GAATTC 3'	Yes
RNase T	Exonuclease	3' exonuclease	No
Phosphodiesterase I	Exonuclease	5' exonuclease	No

Table 6.2. Glyoxalation reversibly modulates activity in several enzymes. *glyoxal reacts with G residues but not inosine sites, converting RNase T1 into an inosine-specific endonuclease.[29-30]

activity, and full cleavage of the target strand was observed upon thermal decaging. We observed similar results with the restriction endonuclease EcoRI, with glyoxal treatment reversibly modulating enzymatic cleavage of the target duplex (**Figure E24**).

Lastly, we were curious as to whether glyoxal interfered with exonuclease activity. We first tested RNase T, a 3'→ 5' exonuclease active on both ssRNA and ssDNA targets.[56-57] Unfortunately, caging exerted no effect on RNase T activity, and full hydrolysis of DNA and RNA substrates was observed regardless of glyoxal caging (**Figure E25**). Similar results were obtained when testing snake venom phosphodiesterase I which hydrolyzes ssDNA and ssRNA in the opposite 5' → 3' direction,[58] and glyoxal caging offered no protection toward enzymatic degradation of either DNA or RNA strand (**Figure E26**). While disappointing, these results are also somewhat unsurprising given that exonuclease sequence specificity is known to be broad and inherently promiscuous, and these enzymes interact primarily with the phosphate backbone irrespective of nucleobase or modification status.

As compiled in **Table 6.2**, it was apparent that glyoxal caging was most effective where nucleic acid secondary structure was required for substrate engagement and activity. In particular, RNase H, DNase I, and EcoRI all preferentially target duplex substrates, and glyoxal inhibition

can be mechanistically attributed toward preventing hybridization and subsequent presentation of preferred substrates. With the exception of RNase T1, all other enzymes tested appeared to have little sensitivity to glyoxal, as these likely prefer single stranded substrates and potentially act independent of nucleobase structure. Overall, our screen yielded distinct enzyme candidates that could be leveraged to produce detectable readouts, and further demonstrates the wider versatility of glyoxal caging towards thermoresponsive biodevice fabrication.

Based on our survey of reversible enzyme disruption using glyoxal caged substrates, we hypothesized that we could likely modulate CRISPR-Cas9 activity. In this system, a single guide RNA (sgRNA) acts as both an aptamer towards Cas9 to bind the nuclease as well as a programmable sequence-specific probe to bring the ribonucleoprotein (RNP) complex to a desired genetic locus for cleavage (**Figure 6.7a**).[59-60] These platforms have proved to be revolutionary towards perturbing biological circuits and have great promise in human gene therapy. As such, there is great interest in exerting tunable external control over CRISPR-based systems. Similar to previously described nucleic acid caging approaches, existing attempts have primarily focused chemical or light-based activation of sgRNAs.[61-63] However, heat activation of sgRNA again remains uninvestigated, and we predicted that glyoxal caging would provide a facile approach for imparting predictable thermal and temporal control. We envisioned that glyoxal cages would denature sgRNA secondary structure required for RNP complex formation while also strongly inhibiting basepairing at the DNA-gRNA interface, thus ablating Cas9 activity.

We first determined the minimal degree of caging required for shutdown of Cas9 cleavage of a 720 bp dsDNA target containing an enhanced green fluorescent protein (eGFP) coding region.[64] In contrast to the steady drop in activity with increasing caging times as seen in our previous experiments, we interestingly saw minimal inhibition of activity over 30 minutes of glyoxal caging followed by a drastic drop in cleavage after ~40 min–1 hour, and a full 2 hours of caging was necessary to completely shut down Cas9 activity (**Figure E27**). We hypothesize that this

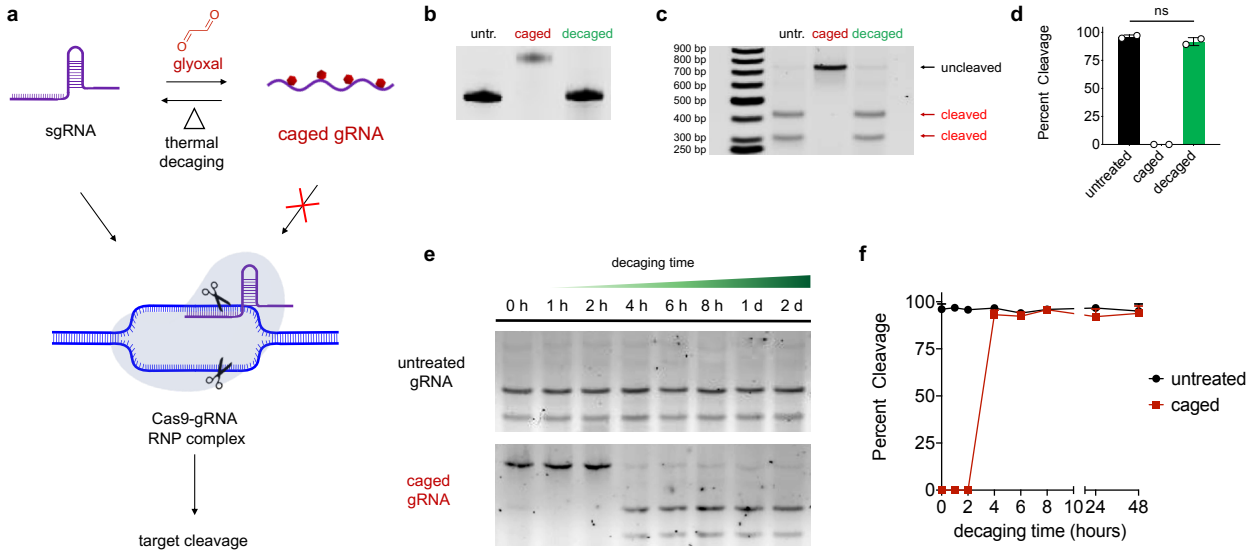


Figure 6.7. Thermal reactivation of CRISPR-Cas9 function. a) Schematic of reversible modulation of CRISPR-Cas9 cleavage. Glyoxal caging reversibly denatures and cages the sgRNA, preventing RNP complex formation and target cleavage. b) 20% denaturing PAGE analysis of untreated (untr.), caged, and decaged sgRNAs. 2 µg of sgRNA was first treated with 1.3 M glyoxal in 50:50 DMSO:H_2O, 50 ºC for 2 hours. To decage the substrate, 2 µg of caged sgRNA (2 hours glyoxal treatment time) was incubated at 95 ºC, pH 7.5 for 5 minutes. c) Functional Cas9 target cleavage with untreated (untr.), caged, and decaged sgRNAs visualized by 1% agarose gel and d) quantified using densitometry. Bars represent mean and S.D. from 2 independent trials. Unpaired t-test was performed between untreated and decaged samples. "ns" indicates no significant difference. e) Caged sgRNAs (2 hours glyoxal treatment time) were decaged at 37 ºC for increasing times and combined with Cas9 and dsDNA target followed by 1% agarose gel analysis and f) quantified using densitometry. Values represent mean and S.D. of 2 independent trials.

higher resistance to caging may be due in part to the uridine-rich nucleotide sequence of CRISPR sgRNAs, as these bases are not known to react with glyoxal.[19-23] Given our previous experiments with both broccoli and ARC259 RNA aptamers, it is also likely that partially caged sgRNAs are still functionally active with Cas9, which would explain maintenance of cleavage activity despite observed sgRNA caging via PAGE (**Figure E27a**). In our studies we utilized a commercially available Cas9 enzyme from *Streptococcus Pyogenes* which is known to exhibit single-turnover kinetics,[65] and so it is unlikely that individual RNP complexes would cleave multiple DNA targets. However, we did employ a moderate excess (~6-7-fold) of Cas9/sgRNA complex compared to DNA target, so these reaction conditions may have also contributed to our observed results. Glyoxal caging was nonetheless effective in completely inhibiting Cas9 cleavage, and so we next sought to thermally restore this activity. We first subjected caged sgRNAs to our rapid decaging

conditions at 95 ºC, pH 7.5, and when we combined these with Cas9 enzyme and dsDNA target, full activity was restored with 2 min of decaging time (**Figure E28**). These data also reflected our caging observations in that Cas9 activity did not steadily increase with decaging times but rather rapidly increased between 0.5-2 min. Additionally, our results corroborate our earlier notion that partially caged sgRNAs still promote Cas9 activity, as full cleavage was still observed from 2-10 min (**Figure E28b**) despite our PAGE shift analysis suggesting full decaging occurring at ~ 5 mins (**Figure E28a**). Regardless, when compared to untreated sgRNA there was also no significant difference in Cas9 cleavage activity when paired with a decaged sgRNA (**Figure 6.7b-d**), overall demonstrating that glyoxal caging is a fully reversible method of controlling CRISPR-Cas9 function.

Given the practical use of CRISPR-Cas9 systems for *in vivo* gene editing applications, we were curious to see if we could replicate thermal reactivation of Cas9 activity under "physiological" conditions. To explore this, we incubated caged sgRNA at 37 °C in PBS, pH 7.5 for increasing periods of time. We first assessed size shift by gel, which expectedly decreased over 2 days (**Figure E29**). We also tested each timepoint for functional Cas9 cleavage activity against an untreated sgRNA control, and we observed full restoration after 4 hours at 37 °C (**Figure 6.7e,f**). Together, these results demonstrate the feasibility of our approach for caging and thermally reactivating CRISPR-Cas9 systems for potential use in biological circuits, nanodevices, and *in vivo* applications. In particular, employing a sgRNA with predictable reactivation times is likely to be useful for both *in cellulo* and eventual *in vivo* contexts, where steady release of active sgRNA provides temporal control over effector function. We anticipate that this would allow tunable gene knockdown, activation, and inhibition across multiple Cas9 platforms. Additionally, eventual *in vivo* delivery of an inactive sgRNA construct would likely reduce premature activity at undesired off-target locations proximal to the delivery site and provide more time for on-target cellular localization and uptake. Additionally, in our proof-of-concept demonstration we used a standard DNA cleavage readout. However, this is only one application of this platform, and glyoxal caging

can likely also be used in applications such as gene activation,[66] gene interference,[67] base editing,[68] or prime editing,[69] which rely on dead Cas9 fusions. In addition, designer systems have been developed that combine guide RNA targeting with fully human effector protein components,[70] and we envision the use of glyoxal caging as a useful and straightforward method for tuning activity in these systems.

Polymerase chain reaction (PCR) is a widely adopted molecular biology technique and a powerful diagnostic method for quantifying gene expression[71] and detecting genetic material.[72] While highly sensitive and efficient, PCR is also prone to non-specific amplification of off-target DNA products. Many of these issues arise from undesired molecular interactions, including primer-dimer formation and non-specific annealing within DNA templates, which primarily occur during lower-temperature steps in early PCR rounds.[73] To mitigate these effects, “hot start” PCR assays are often employed wherein DNA polymerase is inactivated with aptamers and/or antibodies to prevent primer extension until high temperatures are reached.[74-75] While this technique can reduce some PCR artifacts, hot start polymerases are also variably effective for primer-dimers and off-target products, and the added cost of these neutralizing biocomponents can be prohibitive for some applications. Alternatively, primers can be chemically modified with photoactivatable[76] or thermolabile[77] adducts to provide similar improvements in PCR performance. Although these modifications are more effective for directly addressing issues with mis-priming, these methods also suffer from similar practical limitations described above, as they necessitate costly incorporation of chemical groups during solid-phase oligonucleotide synthesis and typically require extensive optimization. Conversely, glyoxalation is cost-effective, tunable, and can be applied toward a variety of modified nucleic acid scaffolds, and hence may offer greater applicability in improving PCR performance. Because glyoxal adducts potently disrupt nucleic acid hybridization and are fully reversible using heat, we were interested to see if this method would prevent off-target primer interactions and improve overall amplification specificity (**Figure 6.8a**).

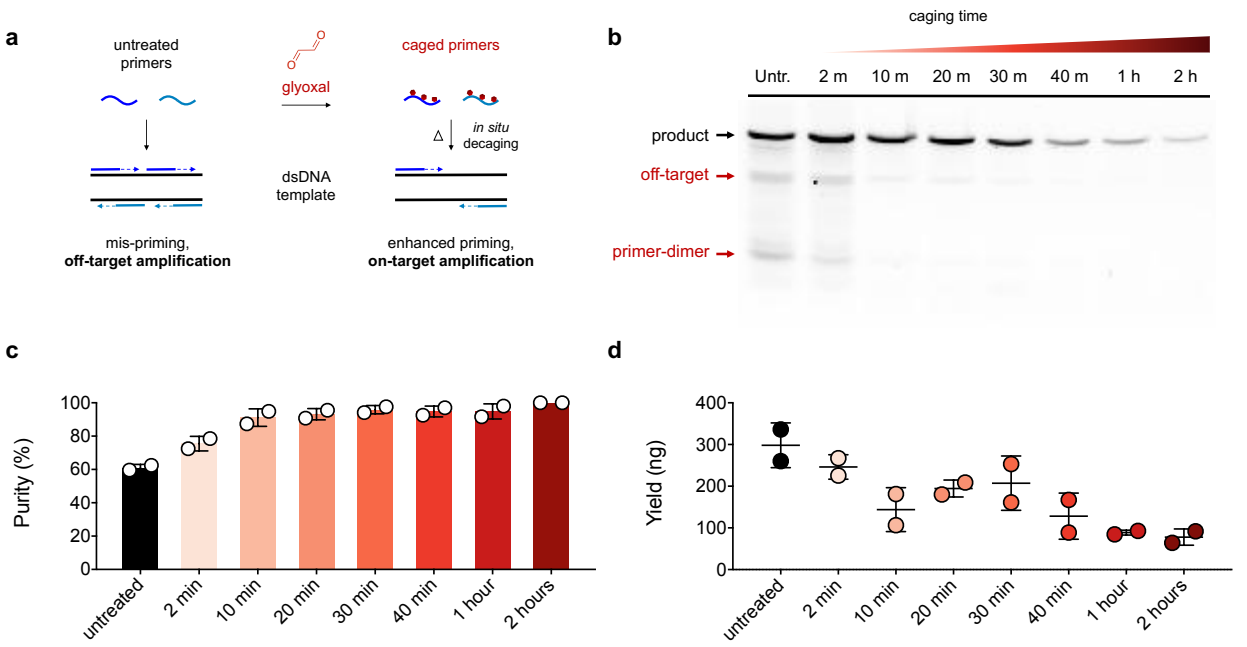


Figure 6.8. Glyoxal treated primers enhance PCR specificity. a) Schematic of PCR amplification using a dsDNA template with untreated and caged primers, which reduce mis-priming and primer-dimers. b) 1% agarose gel analysis of PCR reactions to generate a 653 bp *ACTB* product from human genomic DNA with untreated (untr.) and increasingly caged primers. c) Densitometric quantification of on-target PCR amplification purity with increasingly caged primers. Bars represent mean and S.D. from 2 independent trials. d) Overall PCR yield following amplification and purification. Bars represent mean and S.D. from 2 independent trials.

As a model system, we amplified a ~653 bp segment of the β-actin gene (*ACTB*) using human total DNA and a standard *Thermus aquaticus* (*Taq*) polymerase, a combination which was found to be prone to both mis-priming and primer-dimers.[77] Additionally, a previous study attempted chemical modification of DNA primers and found this mildly improved PCR performance, and we were interested to see if glyoxalated primers would be better suited for addressing mis-priming issues in PCR (**Figure 6.8a**). We first attempted amplification using standard *Taq* PCR reagents, and while we observed the expected target band, we also experienced general "smearing" in reactions as well as the presence of an off-target amplification product at ~400 bp. Additionally, in the absence of any DNA template, products arising from primer-dimer formation were identified at ~150 bp (**Figure E30a**). Given that antibody-neutralized hot start *Taq* is commonly employed to mitigate these issues, we repeated this experiment using

these materials, and while we saw a slight reduction in the formation of off-target products, this ultimately provided little benefit in overall amplification purity (**Figure E30a,b**). We next functionally assessed glyoxal treatment by separately caging either the forward or reverse DNA primer. Interestingly, treating the forward primer produced almost no effect (**Figure E31**), and although reverse primer caging did somewhat enhance amplification specificity (~60% to ~85%), we were disappointed that only moderate improvements were observed for all caging timepoints tested (**Figure E32**). Surprisingly, when we caged both primers prior to PCR, we saw a significant improvement in amplification specificity (**Figure 6.8b,c**). This effect was particularly evident between 0-10 minutes glyoxal treatment time, where observed purity values increased from ~60% to >90%. Moreover, extended primer caging times beyond 30-60 minutes resulted in PCR yields approaching >95% purity. Importantly, glyoxal caging almost completely eliminated both off-target amplification and primer-dimer issues (**Figure 6.8b**), and outperformed previous attempts using thermolabile chemical adducts, which did not suppress these artifacts in the same system.[77]

We hypothesized that this enhancement was due to a slow, iterative release of active primer throughout PCR cycles. To verify this, we glyoxalated both primers for 10 minutes and then monitored primer decaging in our reaction buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, pH 8.3) during the initial denaturation step and subsequent PCR cycles. Interestingly, full decaging of primers was not observed until cycle 20 as monitored by denaturing PAGE shift (**Figure E33**), supporting the idea that stoichiometric limiting of available active primer may contribute to enhanced specificity. We highlight that our glyoxal system is the first to be able to achieve this type of time-released activation of primers across multiple PCR cycles. We also quantified total dsDNA yields after purifying each PCR reaction, and although this led to an expected proportional drop in recovered product, this comes with the large benefit of significantly increased purity (**Figure 6.8c,d**). Primer concentration is often the limiting reagent for total amplification yield, and given our analysis of primer decaging, it is likely that very long caging times result in a proportion of primers that remain caged throughout the entire thermal cycling

program. A larger number of cycles would likely activate these remaining caged primers and may improve yields. Overall, our results demonstrated that glyoxal is a straightforward and tunable modification for enhancing PCR amplification specificity, and will likely benefit both standard molecular biology assays as well as high-accuracy PCR-based diagnostics.

After exploring a wide range of *in vitro* applications, we were interested in demonstrating the ability of glyoxal-caged nucleic acids to modulate function in a cellular environment. Because glyoxal disrupts Watson-Crick-Franklin base pairing, we hypothesized that an antisense oligonucleotide (ASO) would be strongly affected by caging and that we could reversibly modulate its overall function. Titratable gene interference is also emerging as a powerful strategy for profiling cellular proteomes and engineering metabolic pathways,[78-79] and we were curious if a glyoxal-caged ASO would allow us to exert tunable and predictable control over gene expression in living cells. To test this idea, we identified a chemically modified "gapmer" ASO optimized to silence eGFP expression by targeting and degrading its cognate mRNA via the canonical RNase H-mediated decay pathway (**Figure 6.9a**).[80-81] This ASO class is comprised of locked-nucleic acid (LNA) nucleotides at each terminus and contains phosphorothioate linkages throughout the strand, endowing these molecules with high endonuclease resistance and biostability while retaining functional compatibility with cellular gene silencing machinery.[82-84] Using this system, we reasoned that glyoxal caging would temporarily inhibit the ability of the ASO to hybridize to its target mRNA, thus blocking gene suppression activity (**Figure 6.9b**). Based on our previous kinetic analyses for decaging RNA constructs at 37 ºC (**Table 6.1 and Figure 6.7e**), we also hypothesized that this activity could be restored to shut down GFP synthesis in a predictable and tunable manner. Although a widely-adopted and useful fluorescent reporter, GFP also displays remarkable stability ($t_{1/2}$ ~26 h) and accumulates at high levels in the cytoplasm,[85] which can interfere with detecting transient changes in mRNA levels. To circumvent this, GFP is often destabilized by appending a proteolytic "degron" sequence to promote rapid turnover and facilitate temporal detection of transcriptional activity.[86-87] We were concerned that high GFP stability would

mask our ability detect ASO function and reactivation, and so to avoid this outcome and enable high-resolution tracking of mRNA degradation in real-time, we appended a similar C-terminal degron tag (supplemental methods).

After inserting this construct into a cytomegalovirus (CMV) mammalian expression plasmid, we first confirmed reasonable transfection efficiency and GFP expression in HEK293T cells by fluorescence microscopy (**Figure E34**). We next wanted to identify an optimal ASO concentration sufficient for silencing GFP expression, as well as validate the specificity of this previously tested sequence.[80] As shown in **Figure E35**, cells treated with ~250-500 nM of a GFP-targeting ASO exhibited virtually no detectable GFP signal, while a scrambled sequence produced no GFP inhibition across all concentrations tested, indicating potent ASO-mediated silencing of GFP synthesis in a sequence-specific manner. Using these optimal conditions (250 nM ASO) we next created a library of increasingly caged ASO samples, and assessed the impact of glyoxal caging on GFP silencing by co-transfecting HEK293T cells with both the pCMV-GFP vector and respective ASOs. We then imaged all cells after a 12 h recovery period and interestingly saw that ~6-8 h glyoxal treatment time was necessary to completely ablate ASO function and produce GFP signal commensurate with cells receiving only plasmid (**Figure E37**). This treatment time was unexpectedly long compared to some of our previous applications, and while it is possible that some decaging occurs within the 12 h window between transfection and imaging, we also hypothesize that RNase H may tolerate partially caged ASOs, consistent with previous observations that DNA:RNA heteroduplex regions as short as ~4 bp are sufficient for RNA cleavage.[88-90] Ultimately, we were still able to completely inhibit ASO function with glyoxal caging.

We were also initially concerned that glyoxalation may simply prevent ASO strands from entering cells, and so we separately exposed HEK293T cells to Cy5-labeled ASOs either unmodified or treated with glyoxal (8 h). As shown in **Figure E38**, we observed similar internalized fluorescent signal in both untreated and fully glyoxalated samples, and our overall results were consistent with previous studies using these ASO constructs.[80] Combined with our functional

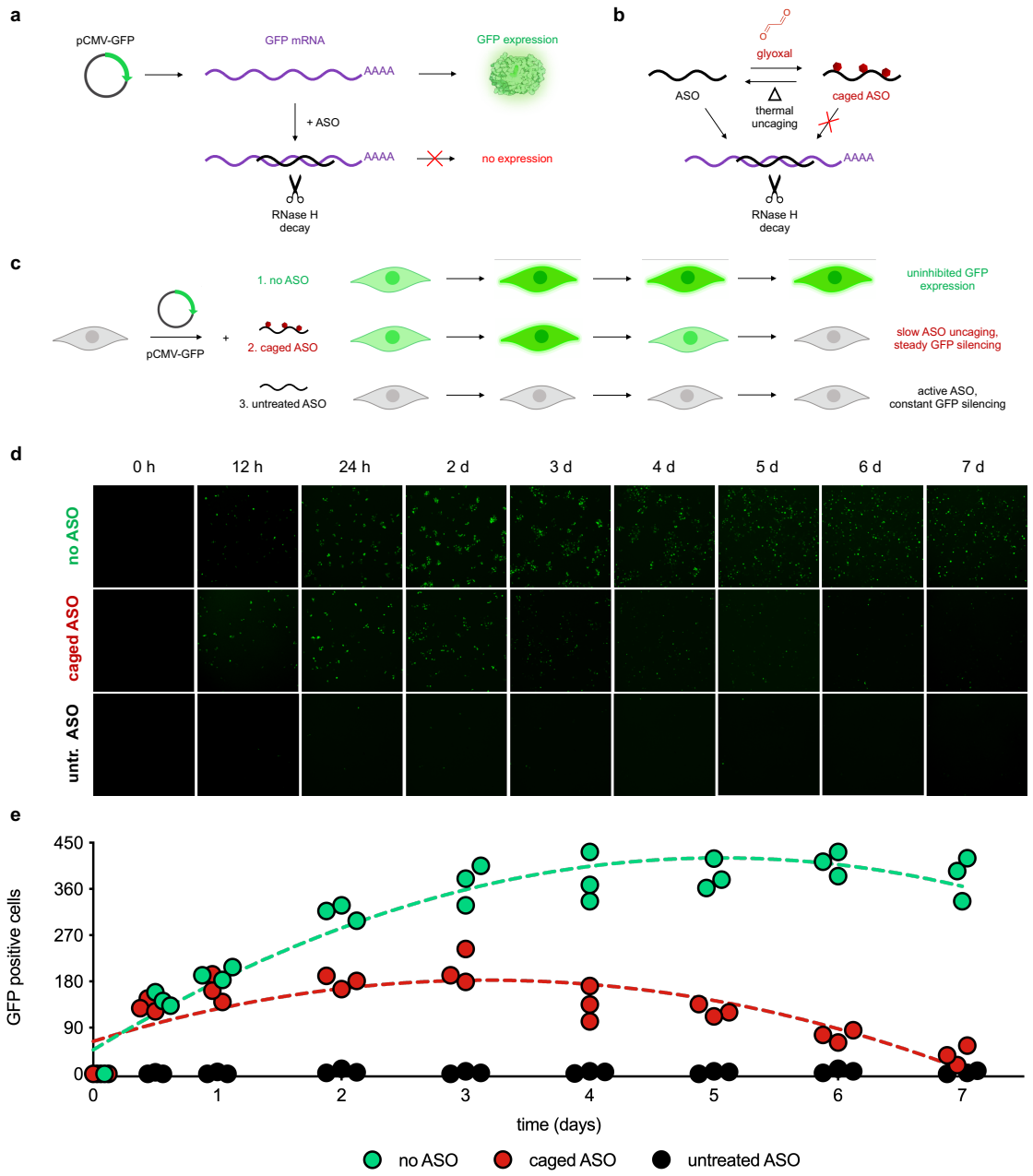


Figure 6.9. Thermal reactivation of antisense oligonucleotide (ASO) function *in cellulo*. a) Schematic of ASO gene silencing mechanism through antisense mRNA hybridization and RNase H-mediated decay. b) Glyoxal caging reversibly blocks ASO hybridization, preventing silencing complex formation and target cleavage. c) Experimental workflow for *in cellulo* time-release ASO decaging. At $t = 0$, HEK293T cells were transfected with a pCMV-GFP plasmid as well as 250 nM glyoxal caged ASO (8 hours glyoxal treatment time, red), untreated (untr.) ASO (black) or no ASO (green). Cells receiving no ASO exhibit uninhibited GFP expression throughout the experiment, while cells given untreated (untr.) ASO produce no detectable GFP due to constant gene silencing. Cells treated with caged ASO (8 hours glyoxal treatment time) exhibit initial increase in fluorescence due to GFP expression followed by a delayed suppression of GFP signal due to time-release reactivation of glyoxalated ASO. d) Representative live-cell fluorescence microscopy images (4X magnification) taken throughout the course of the experiment. e) Quantification of GFP-positive cells in each field across treatment groups during the experimental time course. Circles represent individual wells (n = 3) from a 96-well plate. Curves are overlayed with a second-order polynomial fit (dashed lines).

results demonstrating GFP silencing in transfected cells, it is likely that caging exerts minimal effect on cellular uptake.

We were next interested in restoring ASO activity *in cellulo*, and so we performed a time-course experiment to monitor GFP expression in HEK293T cells. On day 0, cells received either 1) GFP plasmid alone, 2) plasmid and a fully caged ASO (8 h glyoxal treatment time), or 3) plasmid and an untreated active ASO (**Figure 6.9c**). We periodically imaged these samples over the course of 7 days and then used ImageJ[91] to quantify and plot GFP-positive cells in images across groups. (**Figure 6.9d,e**). Cells receiving a fully caged ASO exhibited similar increases in GFP expression throughout the first 24 h of the experiment compared to cells receiving only plasmid, suggesting full glyoxal inhibition of ASO function during this period. However, in contrast to control cells which exhibited an increased GFP-positive population and sustained expression throughout days 2-7, cells treated with a caged ASO displayed a steady decrease in GFP-positive cells for the remainder of the time-course experiment, indicating ASO thermal uncaging and reactivation of gene silencing function, with approximately ~90% restoration in function by day 7 (**Figure 6.9d,e**). Conversely, cells receiving an untreated ASO displayed virtually no GFP expression for the entirety of the experiment, confirming that this ASO concentration (250 nM) was sufficient for gene silencing in proliferating cells across all timepoints.

To provide a comparison of decaging in these conditions, we also prepared several caged (8 h treatment time) ASO samples in complete DMEM and monitored decaging at 37 ºC using denaturing PAGE (**Figure E39**). Interestingly, these data suggest that full ASO decaging may be complete in ~1-2 days, which was somewhat faster than both our previous kinetic analysis on a model DNA strand (**Table 6.1**) as well as our functional gene silencing results (**Figure 6.9d,e**). To recapitulate *in cellulo* caging conditions, the amount of nucleic acid present in each decaging reaction (250 nM, 1.3 ng/μL) was lower than these previous tests (400 nM, ~4 ng/μL), suggesting that the concentration of caged construct may be an important consideration for thermoreversibility. Additionally, there may also be a “buffer effect” (complete DMEM vs PBS)

that could alter the overall kinetics of glyoxal release. While still somewhat faster than our functional *in cellulo* results, these reactions are also a simple representation of a cellular context and we hypothesize that ASO decaging is likely influenced by intracellular pH gradients, which can range from pH 4.5-8 across organelle compartments.[92-93] Moreover, the process of ASO uptake, cytosolic partitioning, and RNA silencing complex formation is a multiphase process with its own kinetics, and additionally competes with multiple cell divisions that can dilute these materials. Ultimately, these data provide a rough kinetic estimate for decaging and importantly illustrate analytical confirmation that decaging occurs in these conditions (**Figure E39**).

In nearly all of our previously explored *in vitro* applications, glyoxal treatment time was an adjustable variable to produce proportional decreases in nucleic acid activity, and we were interested if we could similarly tune GFP expression by varying the degree of ASO caging. In parallel with our kinetic analysis using a fully caged ASO *in cellulo,* we also monitored functional gene silencing activity of HEK293T cells co-transfected with both GFP plasmid and increasingly caged ASOs (5 min–8 h glyoxal treatment time). Interestingly, when combined with our previous data, the degree of glyoxal caging in different ASO samples produced a clear correlation in the number of GFP-positive cells in each well, demonstrating that glyoxalation is a feasible method for tuning protein expression levels within cell populations (**Figures E40, E41**). ASOs treated with glyoxal for 0–30 minutes produced minimal inhibition, while 1-8 hours resulted in relative increases in GFP expression. Additionally, these samples all displayed proportional reactivation of gene silencing activity across 7 days, demonstrating the use of this method to both delay gene suppression as well as modulate overall protein synthesis levels. Although we did not observe full restoration of ASO activity across all caging treatment times (**Figures 6.9e, E40**), this would likely be seen in longer decaging experiments past 7 days. We also performed these experiments using highly proliferative immortalized cells (HEK293T), and the predictable reactivation we observed over several days invites interesting use of these constructs to modulate gene expression levels in both terminally differentiated and/or primary cells.

Although a natural metabolite produced from both sugar and lipid oxidation,[94-96] glyoxal can be cytotoxic at higher concentrations,[97-98] and we wanted to verify that slow intracellular release of small amounts glyoxal was tolerable *in cellulo*. Qualitatively, we observed no differences in morphology or growth rate between any cell groups receiving differentially caged ASOs, and all treated cells displayed roughly equivalent doubling times and were confluent by end of the experiment (**Figures E40**). At the conclusion of day 7, we also tested viability and mitochondrial function of treated cells using a water-soluble tetrazolium assay,[99] and we observed that all ASOs, regardless of degree of caging, produced no significant decreases in viability (**Figure E41**). In parallel, we also evaluated glyoxal alone in a dose-response viability assay, and did confirm that the reagent itself is cytotoxic, albeit at much higher concentrations (IC_{50} = 310±83 μM, Figure S42). In our example ASO, even if all nucleobase positions contained a glyoxal adduct (16 nt), glyoxal levels would be ~4 μM, well below the concentration where adverse effects are observed. Additionally, the slow release of glyoxal may enable it to be cleared by existing cellular metabolism pathways. Careful monitoring of potential toxicity would obviously be wise when applying glyoxal-modified nucleic acids in pharmaceutical formulations, but in our experiments we saw no observable cytotoxicity from our caged ASOs, demonstrating their compatibility and utility to perturb gene expression in cell populations. Additionally, glyoxal-caged nucleic acid constructs are generalizable, and thus can be used broadly for kinetically tuning gene expression to better understand cellular signaling mechanisms as well as engineer metabolic pathways.

6.4 Conclusions

Nucleic acids are versatile and attractive materials for constructing programmable and responsive elements in biomedicine, data storage, biocomputing, and nanotechnology. The ability to control nucleic acid structure and activity is essential in the design and implementation of such systems, and significant effort has been dedicated toward imparting stimuli-responsiveness in nucleic acid constructs. However, the majority of these efforts have focused on chemical or light-based reactivation of caged nucleic acids, and limitations remain in that these caging systems are not compatible with enzymatically-derived DNA substrates and have yet to be demonstrated with XNAs. As a third major stimuli source, heat remains surprisingly unexplored as a nucleic acid uncaging element, despite the high degree of control that is possible and its widespread use in laboratory settings.

In this work we show that glyoxal, a chemical denaturant used for decades in molecular biology assays, can impart thermoreversible inhibition of both the structure and activity of a variety of nucleic acid scaffolds. We first demonstrate facile glyoxal attachment on model DNA strands and show tunable addition of caging groups. We then extensively parameterize the removal of these cages through adjustable combinations of pH, temperature, and incubation times. Glyoxal cages are potent disruptors of nucleic acid secondary structure, and we show full inhibition of small molecule aptamer interactions as well as DNAzyme-based catalysis. In addition, because glyoxal reacts with nucleobase moieties rather than backbone functional groups, we show that thermoreversible caging can be easily applied toward natural DNA and RNA scaffolds as well as heavily modified backbones and non-canonical XNA polymers. We also explore glyoxal-based interference with enzyme activity, and identify several enzyme candidates that display reversible function toward caged nucleic acid substrates, including several endonucleases as well as the CRISPR-Cas9 gene editing platform. Because glyoxalation is heat-reversible, we also demonstrate that caged primers provide significantly enhanced specificity in PCR amplification reactions. Lastly, we treat an antisense oligonucleotide with glyoxal and show that these

constructs enable potent disruption of function as well as tunable activation and titration of gene expression levels in living mammalian cells. Together, we demonstrate that thermoreversible glyoxal caging can be easily applied for tunable inhibition and full reactivation of nucleic acid function in a suite of contexts, establishing a straightforward and effective framework for use in a variety of potential synthetic biology and biotechnology applications.

6.5 Materials and Methods

Glyoxal Caging Kinetics

For visualization and quantification of caging kinetics, a custom designed DNA oligonucleotide was purchased from Integrated DNA Technologies (IDT), containing a 6-FAM (Fluorescein) label at the 5' terminus as shown below.

5' FAM TGCCAAGACTGTTGAGGAAGATGAGAGAAT 3'

In triplicate, 0.2 nmol of the test DNA strand was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO. Reactions were brought to a final volume of 100 µL with nuclease-free water, and incubated at 50 ºC for 0, 5, 10, 20, 30, 40, 50, and 60 minutes. At each time point, reactions were ethanol precipitated and reconstituted in 60 µL of nuclease-free water. 10 pmol of each purified reaction was then mixed with 2X RNA loading dye (New England Biolabs) and then separated with 20% PAGE and imaged with a GE Amersham Typhoon RGB scanner using a 488 nm excitation laser and the Cy2 525BP20 emission filter. Fiji (ImageJ) was used to calculate densitometric intensity of each reaction and normalizing to the fully caged DNA band at 60 min to estimate percent conversion. Caging half-times was estimated based on pseudo-first order rate kinetics using a nonlinear curve fit in GraphPad Prism.

Decaging Kinetic Assays

In duplicate, 20 pmol of a fully caged (1 hour treatment) test DNA strand was incubated in a final volume of 50 µL of phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4) adjusted to pH 6.5, 7.0, 7.5, or 8.0 where appropriate. Samples were incubated at various temperatures (25 ºC, 37 ºC, 50 ºC, 70 ºC, 95 ºC) in a thermal cycler, and 3 µL samples were taken at the indicated time points and mixed with 2X RNA loading dye (New England Biolabs) and then separated with 20% PAGE and imaged with a GE Amersham Typhoon RGB scanner using a 488 nm excitation laser and the Cy2 525BP20 emission filter. Fiji (ImageJ) was used to calculate densitometric intensity of each reaction and normalizing to an untreated DNA band to estimate percent fully decaged. For stability at 25 ºC, percent remaining was calculated using the band intensity of each reaction normalized to the fully caged DNA strand at T = 0 hours. Decaging half-times were estimated based on pseudo-first order rate kinetics using a nonlinear curve fit in GraphPad Prism.

Broccoli Aptamer Preparation

A dsDNA template gBlock containing a T7 RNA polymerase promoter (underlined) was purchased from Integrated DNA Technologies (IDT) as shown below.

5’ GCTAG*TAATACGACTCACTATA*GTTGCCATGTGTATGTGGGAGACGGTCGGGT
CCAGATATTCGTATCTGTCGAGTAGAGTGTGGGCTCCCACATACTCTGATGATCCTTCGGGATC
ATTCATGGC 3’

Template was then amplified with a forward primer (5’GCTAGTAATACGACTCACTATA GGGTTGCC 3’) and reverse primer 5’ GCCATGAATGATCCCGAAGGATCATCA 3’) using HotStart Taq DNA polymerase (New England Biolabs) according to the manufacturer’s instructions, using the following PCR program: 94 °C for 3 min, followed by 30 cycles of (94 °C for 1 m, 57 °C for 45 s, 68 °C for 1 m), 68 °C for 5 min. PCR reactions were then purified using

the Monarch PCR & DNA Cleanup Kit (New England Biolabs). RNA was then synthesized and purified using the HiScribe T7 High Yield RNA Synthesis Kit and Monarch RNA Cleanup Kit (New England Biolabs).

Broccoli Aptamer Caging and Functional Fluorogenic Analysis

60pmol (~2 µg) of broccoli RNA aptamer was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO. Reactions were brought to a final volume of 100 µL with nuclease-free water and reacted at 50 ºC for 0 min, 30 sec, 5 min, 10 min, 15 min, 20 min, 25 min, 30 min, and 1 hour. At each time point, reactions were ethanol precipitated and reconstituted in 25 µL of nuclease-free water. 10 pmol of each purified reaction was then mixed with 2X RNA loading dye (New England Biolabs) and then separated with 10% denaturing PAGE. Gels were then stained with 1X SYBR Gold solution (Thermo Fisher Scientific) for 20 minutes, and visualized with the Typhoon RGB scanner. Functional fluoregenic activity of the broccoli RNA aptamer was performed by mixing 20 pmol of untreated, caged, or decaged aptamer with 2 µM 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI, Sigma Aldrich), 40 mM HEPES, 100 mM KCl, 1 mM $MgCl_2$, pH 7.4 in a final volume of 20 µL. Reactions were incubated for 30 minutes at 37 °C. Reaction tubes were imaged with the GE Amersham Typhoon RGB scanner using a 488 nm excitation laser and the Cy2 525BP20 emission filter. Fluorescence was also quantified by transferring the reactions to 384-well black plates (Greiner) and measuring intensity on a Cytation 5 multi-mode plate reader (BioTek) using excitation at 447 nm and emission at 501 nm.

Broccoli Aptamer Decaging

To remove glyoxal adducts, 20 pmol of minimally caged (10 minute glyoxal treatment) broccoli RNA aptamer was added to a final volume of 50 µL of phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.5). Samples were incubated in a thermal cycler at 95 ºC for indicated times and purified by ethanol precipitation. 10 pmol of the collected RNAs were then combined with DFHBI and analyzed as described above.

***In Situ* Broccoli Aptamer Decaging**

20 pmol of minimally caged broccoli RNA aptamer was combined with 2 µM DFHBI, 40 mM HEPES, 100 mM KCl, 1 mM $MgCl_2$, pH 7.4 in a final volume of 20 µL. Reactions were incubated in a thermal cycler set at 60 ºC, 65 ºC, 70 ºC, 75 ºC, 80 ºC, 85 ºC, 90 ºC, and 95 ºC and allowed to react for 0, 5, 10, 20, 30 and 60 minutes. Reactions were quenched by freezing at -80 ºC. Reactions were then thawed at room temperature and transferred to a 384-well black plate with a clear bottom (Greiner) and imaged and analyzed as described earlier. Heatmap image was obtained using the Typhoon acquisition software and reflects relative fluorescent intensities.

10-23 DNAzyme Caging and Functional Assay

The 10-23 DNAzyme oligonucleotide was purchased from IDT as shown below:

5’ - AGGACGGGAGGCTAGCTACAACGAGTGGTTGCC- 3’

100 pmol of DNAzyme was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO. Reactions were brought to a final volume of 100 µL with nuclease-free water and reacted at 50 ºC for 0 min, 1 min, 5 min, 10 min, 20 min, 30 min, 40 min, and 1 hour. At each time point, reactions were ethanol precipitated and reconstituted nuclease-free water at a final concentration of 1 µM. 10 pmol of each purified reaction was then mixed with 2X RNA loading dye (New England Biolabs) and then separated with 20% denaturing PAGE. Gels were then

stained with 1X SYBR Gold solution (Thermo Fisher Scientific) for 20 minutes, and visualized with the Typhoon RGB scanner. To test functional cleavage activity, 15 pmol of untreated or caged 10-23 DNAzyme were added to 1.5 pmol of target strand: (5' FAM GGCAACCACrGTCCCGTCCT BHQ1 3', rG indicates ribonucleotide guanosine) in DNAzyme buffer (10 mM $MgCl_2$, 150 mM NaCl, 50 mM Tris-HCl, pH 7.5) at a total volume of 50 µL. The reaction was incubated in a thermocycler for one hour at 37 °C and quenched with addition of 1 µL of 125 mM EDTA. All samples were then analyzed with 12% denaturing PAGE, stained with 1X SYBR Gold solution (Thermo Fisher Scientific) for 20 minutes and visualized with the Typhoon RGB scanner. Percent cleavage was calculated by densitometry analysis of the cleaved band/sum using ImageJ software.

10-23 DNAzyme Decaging

To decage the 10-23 DNAzyme, we first determined the minimum amount of thermal decaging time required for the gel shift to return to the original size. 10 pmol DNAzyme was suspended in 1X PBS and incubated at 95 °C for 30 seconds, 1 min, 2 min, 3 min, 5 min, 8 min, and 10 mins. The samples were ethanol precipitated and resuspended at a concentrated of 1 µM. The size shift from glyoxal removal was monitored by 20% denaturing PAGE. Optimal timepoints for caging (1 h) and decaging (10 minute) were then analyzed for size shift with untreated 10-23 DNAzyme on 20% denaturing PAGE. For full tuning of activity, we performed the cleavage assay with untreated, caged, and decaged 10-23 DNAzyme. 15 pmol of untreated, caged, or decaged 10-23 DNAzyme were added to 1.5 pmol of target strand in DNAzyme buffer in a total volume of 50 µL. The reaction was incubated for one hour at 37 °C and quenched with the addition of 1 µL of EDTA at 125 mM. All samples were then analyzed on 12% denaturing PAGE. Densitometry analysis was performed using ImageJ software. Percent cleavage was calculated by dividing the band intensity of the cleaved product by the sum of all band intensities in that lane.

ARC259 Aptamer Caging

ARC259 2′-*O*-Methylated RNA aptamer was purchased from Integrated DNA Technologies, containing a 6-FAM (Fluorescein) label at the 5’ terminus as shown below (m indicates 2′-*O*-Methylated modification).

5’ FAM mAmCmGmCmAmGmUmUmUmGmAmGmAmAmGmUmCmGmCmGmCmGmU 3′

60pmol (~2 µg) of ARC259 aptamer was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO. Reactions were brought to a final volume of 100 µL with nuclease-free water and reacted at 50 ºC for 0 min, 30 sec, 5 min, 10 min, 15 min, 20 min, 25 min, 30 min, and 1 hour. At each time point, reactions were ethanol precipitated and reconstituted in 25 µL of nuclease-free water. 5 pmol of each reaction was analyzed by 20% denaturing PAGE.

ARC259 Decaging

20 pmol of minimally caged (40 minute glyoxal treatment) ARC259 aptamer was added to a final volume of 50 µL of phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.5). Samples were incubated in a thermal cycler at 95 ºC for indicated times and purified by ethanol precipitation. 10 pmol of the collected RNAs were then analyzed by 20% denaturing PAGE as described above.

VEGF Fluorescence Polarization Binding Assay

A 20 nM solution of untreated, caged, or decaged ARC259 aptamer was prepared in 1X binding buffer (phosphate buffered saline 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO4, pH 7.5 with 0.05% Tween 20). A 30 µM solution of recombinant human VEGF165 (Peprotech) or bovine serum albumin (BSA, Sigma Aldrich) was prepared in 1X binding buffer and serially diluted 1:1 several times. Dilutions were combined with an equal volume of 20 nM ARC259 to yield final conditions in all samples of 10 nM ARC259, 1X binding buffer, and protein

(VEGF or BSA) ranging from ~30 pM to 250 nM in a final volume of 40 µL. Binding reactions were allowed to incubate at room temperature for at least 30 minutes, and then transferred to a 384-well black plate. Fluorescence polarization was measured using Cytation 5 multi-mode plate reader (BioTek) equipped with a Blue/UV FP filter cube. All measurements were performed using a 360/40 excitation and 460/40 emission filter set in combination with a 400 nm cut off dichroic mirror. Net values were computed by subtracting FP values calculated from blank (buffer) wells.

TNA Oligonucleotide Synthesis and Labeling

TNA phosphoramidites were synthesized according to previously published methods for solid phase synthesis of TNA oligonucleotides.[1] A custom TNA oligonucleotide was synthesized and purchased from the University of Utah DNA/Peptide Synthesis Core Facility as shown below:

3’ CATGACATGAGCTAACCAGACAG 2’

TNA was fluorescently labeled with Cyanine 5 (Cy5) using the Label IT® Tracker™ Intracellular Nucleic Acid Localization Kit (Mirus Bio) according to the manufacturer’s instructions. The oligonucleotide was ethanol precipitated and analyzed by UV/Vis spectrophotometry to measure degree of labeling, confirming approximately 2 dye molecules per strand.

1. Zhang, S.; Chaput, J. C., Synthesis of Threose Nucleic Acid (TNA) phosphoramidite monomers and oligonucleotide polymers. Current protocols in nucleic acid chemistry 2012, 50 (1), 4.51. 1-4.51. 26.

TNA Caging

250 pmol of TNA oligonucleotide was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO. Reactions were brought to a final volume of 100 µL with nuclease-free water and reacted at 50 ºC for 0 minutes, 5 minutes, 10 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, and 1 hour. At each time point, reactions were ethanol precipitated and reconstituted in 25 µL of nuclease-free water. 5 pmol of each reaction was analyzed by 20% denaturing PAGE and imaged with a GE Amersham Typhoon RGB scanner using a 635 nm excitation laser and the Cy5 670BP30 emission filter.

TNA Decaging

20 pmol of minimally caged (40 minute glyoxal treatment) TNA oligonucleotide was added to a final volume of 50 µL of phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.5). Samples were incubated in a thermal cycler at 95 ºC for 5 minutes and purified by ethanol precipitation. 5 pmol of collected TNA was then analyzed by 20% denaturing PAGE as described above.

TNA Hybridization Assays

To test hybridization, a full complement and scrambled DNA oligonucleotide was purchased from Integrated DNA Technologies as shown below.

TNA complement: 5’ CTGTCTGGTTAGCTCATGTCATG 3’

TNA scrambled: 5’ ACTCTGTTCGGTACTGGTCTTG 3’

For each hybridization test, a 10 nM solution of untreated, caged, or decaged Cy5-labeled TNA strand was prepared in 1X binding buffer (40 mM HEPES, 100 mM KCl, 1 mM $MgCl_2$, 0.05% Tween 20, pH 7.4). A 1 µM solution of complement and scrambled DNA was prepared in 1X

binding buffer and serially diluted 1:1 several times. Dilutions were combined with an equal volume of 10 nM TNA solution to yield final conditions in all samples of 5 nM TNA strand, 1X binding buffer, and DNA challenge (complement or scramble) ranging from ~152 pM to 500 nM in a final volume of 100 µL. Samples were incubated at room temperature for 30 minutes and then loaded into NT.115 standard glass capillaries. MST experiments were performed using a Nanotemper Monolith NT.115 Pico instrument. All measurements were analyzed using the Pico-RED filter with 20% LED intensity and low laser power.

PNA Oligomer Synthesis

A fluorescently labeled PNA strand was synthesized using a standard solid-phase synthesis protocol on a Biotage SP wave semi-automatic synthesizer. Sequence is shown below. FAM denotes fluorescein, and E denotes a single glutamate residue (structure shown in Figure S14a).

N - FAM GTAGATCACT E - C

Synthesis began by loading 69.1 mg of a rink amide MBHA resin (0.52 mmol/g) with 5 µmols of Fmoc-L-glutamic acid γ-*tert*-butyl ester using 1.5 eq. HATU, 1.5 eq. DIPEA, and 1.5 eq. 2,6-lutidine in 200 µL dry NMP for 1 hour followed by a 1 hour capping step using a solution of 9% acetic anhydride and 13% 2,6-lutidine in DMF. The resin was then deprotected with a solution of 25% piperidine in DMF. For monomer couplings, 5 eq. of monomer was pre-activated for 10 minutes with 5 eq. HATU, 5 eq. DIPEA, and 5 eq. 2,6-lutidine in 400 µL NMP before addition to the resin. Coupling proceeded with microwave-assistance at 75°C for 6 min. The resin was then washed (5x1 mL DMF), capped using the capping solution (2x5 min with 1 mL each), washed (5x1 mL DMF, 3x1 mL DCM, 3x1 mL DMF), deprotected with deprotection solution (3x2 min with 1mL each), and washed (5x1 mL DMF, 3x1 mL DCM, 3x1 mL DMF) to complete a coupling cycle.

Upon completion of synthesis, the resin was washed with DCM and dried before cleavage using a solution of 2.5% H_2O and 2.5% TIS in TFA. The crude oligomer was ether precipitated, washed with ether, and dried for purification. Purification was performed by reverse-phase HPLC using an Agilent Eclipse XDB-C18 5 μm, 9.4x250 mm column at 60°C with a flow rate of 2 mL/min, monitored at 260 nm using a linear gradient (10%-40% in 15 min) of 0.1% TFA/acetonitrile in 0.1% TFA/H_2O. Identity of pure oligomer was confirmed using an Agilent 6230 electrospray ionization time-of-flight (ESI-TOF) mass spectrometer.

Abbreviations:

Fmoc, fluorenylmethyloxycarbonyl; HATU, 1- [Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate; DIPEA, diisopropylethylamine; NMP, N-methyl-2-pyrrolidone; DCM, dichloromethane; TIS, triisopropylsilane; TFA, trifluoroacetic acid.

PNA Caging

1 nmol of PNA oligonucleotide was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and brought to a final volume of 100 µL with nuclease-free water. Samples were reacted at 50 ºC for 0 minutes, 5 minutes, 10 minutes, 20 minutes, 30 minutes, 1 hour, 2 hours, and 4 hours. At each time point, reactions were purified with reverse-phase HPLC using an Agilent Eclipse Plus C18 3.5 μm, 4.6x150 mm column at 60°C with a flow rate of 1 mL/min, monitored at 260 nm using a linear gradient (10% - 50% in 20 min) of 0.1% TFA/acetonitrile in 0.1% TFA/H_2O. Collected fractions were dried under vacuum and then resuspended in 50 µl phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.5). 10 pmol of each reaction was analyzed by 20% denaturing PAGE as described earlier.

PNA Decaging

250 pmol of caged (2 h glyoxal treatment) PNA strand was added to a final volume of 50 µL of phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.5). Samples were incubated in a thermal cycler at 95 ºC for 0 minutes, 1 minutes, 2 minutes, 5 minutes, 10 minutes and 20 minutes. 5 pmol of each reaction was then analyzed by 20% denaturing PAGE as described above.

PNA Hybridization Assays

To test hybridization, a full complement and scrambled DNA oligonucleotide was purchased from Integrated DNA Technologies as shown below.

PNA complement: 5’ AGTGATCTAC 3’

PNA scrambled: 5’ CTATGGTACA 3’

For each hybridization test, a 20 nM solution of untreated, caged, or decaged Cy5-labeled PNA strand was prepared in 1X binding buffer (40 mM HEPES, 100 mM KCl, 1 mM $MgCl_2$, 0.05% Tween 20, pH 7.4). A 1 µM solution of complement and scrambled DNA was prepared in 1X binding buffer and serially diluted 1:1 several times. Dilutions were combined with an equal volume of 10 nM PNA solution to yield final conditions in all samples of 5 nM PNA strand, 1X binding buffer, and DNA challenge (complement or scramble) ranging from ~152 pM to 500 nM in a final volume of 100 µL. Samples were incubated at room temperature for 30 minutes and then loaded into NT.115 standard glass capillaries. MST experiments were performed using a Nanotemper Monolith NT.115 Pico instrument. All measurements were analyzed using the Pico-RED filter with 20% LED intensity and low laser power.

RNase H assays

An RNA oligonucleotide containing a 6-FAM modifier was purchased from IDT as shown below:

5’ FAM rArArGrCrArGrCrArGrGrCrUrArUrGrUrUrArGrArArCrArArU 3’

To demonstrate duplex requirement for activity, 5 pmols of this RNA was then hybridized to 5 pmols of complementary DNA (5’- ATTGTTCTAACATAGCCTGCTGCTT -3’) in 1X RNase H buffer (New England Biolabs) in a total volume of 50 µL for 30 minutes at 37 °C. After hybridization, 5 units of RNase H (New England Biolabs) were added to 10 µL of each sample. Reactions were incubated for one hour at 37 °C and halted with addition of 1 µL of 125 mM EDTA. Samples were then analyzed with 12 % non-denaturing native PAGE.

To inhibit RNase H activity through caging, 100 pmol of RNA was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO and brought to a final volume of 100 µL with nuclease-free water. RNA was caged at 50 °C for 2 h. RNA was then ethanol precipitated and resuspended in nuclease free water at a final concentration of 1 µM. 5 pmols of RNA was then hybridized to 5 pmol of complementary DNA (5’- TTCGTCGTCCGATACAATCTTGTTA -3’) in 1X RNase H buffer (New England Biolabs) in a total volume of 50 µL for 30 minutes at 37 °C. After hybridization, 5 units of RNase H (New England Biolabs) were added to 10 µL of each sample. Reactions were incubated for one hour at 37 °C and halted with addition of 1 µL of 125 mM EDTA. Densitometry analysis was performed using ImageJ software. Percent cleavage was calculated by dividing the band intensity of the cleaved product by the sum of all band intensities in that lane. To decage the strand, 10 pmol of caged RNA was incubated in 1X PBS, pH 7.5 at 95 °C for 5 minutes. RNA was then ethanol precipitated and resuspended at a concentrated of 1 µM. RNA size shift was analyzed via 20% denaturing PAGE. For full tuning of RNase H activity, we performed the cleavage assay with untreated, caged, and fully decaged RNA. To separate tubes,

5 pmol of the untreated, caged, or decaged RNA were hybridized to 5 pmol of complementary DNA in 1X RNase H1 buffer in a total volume of 50 µL at 37 °C for 30 minutes. RNase H1 (5 units) was added to 10 µL of each reaction and incubated for one hour at 37 °C. Cleavage was halted with the addition of 1 µL of 125 mM EDTA.

Thermostable RNase

For one pot decaging of Thermostable RNase H activity, 10 pmol of untreated, caged, or decaged RNA were hybridized with 10 pmol of complementary DNA in 1X RNase H buffer in a total volume of 100 µL at 37 °C for 30 minutes. 5 units of Thermostable RNase H (New England Biolabs) was added to 10 µL of each duplex (untreated, caged, decaged) and incubated at 90 °C for 0, 0.5, 1, 2, 3, 5, 8, 10 minutes followed by incubation at 37 °C for one hour. Cleavage was halted with the addition of 1 µL of 125 mM EDTA. Activity was then measured using 12% denaturing PAGE as described earlier.

RNase A Assay

To test RNase A, we used the same RNA sequence as shown above in RNase H assays. Caging and decaging conditions were also identical as described earlier. RNase A (Sigma Aldrich, ~0.02 units at 0.5 µg/mL) was added to 1 pmol of untreated, caged, or decaged RNA diluted to 10 µl in 1X PBS, pH 7.5. Reactions were incubated for one hour at 37 °C and halted with the addition of 1 µL of 125 mM EDTA. Activity was then measured using 12% denaturing PAGE as described earlier.

Nuclease P1 Assay

To test Nuclease P1, we used the same fluorescein labeled RNA as shown earlier in RNase H assays (5' FAM rArArGrCrArGrCrArGrGrCrUrArUrGrUrUrArGrArArCrArArU 3') and the same fluorescein labeled DNA as used in glyoxal caging kinetics (5' FAM TGCCAAGACTGTTGAGGAAGATGAGAGAAT 3'). Caging and decaging conditions were also identical as described earlier. Nuclease P1 (New England Biolabs, 1 unit) was added to 1 pmol of untreated, caged, or decaged RNA in a total volume of 10 µL in 1X Nuclease P1 buffer (New England Biolabs). Reactions were incubated for 15 minutes at 37 °C and halted by addition of 1 µL of 125 mM EDTA. All samples were then analyzed on 12% denaturing PAGE.

DNase I Assay

To test DNase I, we used the fluorescein labeled DNA sequence as described in the glyoxal caging kinetics section (5' FAM TGCCAAGACTGTTGAGGAAGATGAGAGAAT 3'). To inhibit DNase activity through caging, 100 pmol of target DNA was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO and brought to a final volume of 100 µL with nuclease-free water. DNA was caged at 50 °C for 1 h. DNA was then ethanol precipitated and resuspended in nuclease free water at a final concentration of 1 µM. To decage the strand, 10 pmol of caged DNA was incubated in 1X PBS, pH 7.5 at 95 °C for 5 minutes. DNA was then ethanol precipitated and resuspended at a concentrated of 1 µM. In separate tubes, 10 pmol of untreated, caged, and decaged labeled strand was combined with 10 pmol of a complementary DNA strand (5' ATTCTCTCATCTTCCTCAACAGTCTTGGCA 3') in 100 µL 1X DNase buffer (Thermo Fisher) at 37 °C for 30 minutes. After hybridization, 0.2 units of DNase I (Thermofisher) were added to 10 µL of untreated, caged, or decaged duplexes. Reactions were incubated for one hour at 37 °C and halted with addition of 1 µL of EDTA at 125 mM. All samples were then analyzed with 12% denaturing PAGE.

EcoRI Assay

An DNA oligonucleotide containing a 6-FAM modifier was purchased from IDT as shown below (cut site underlined):

5’ FAM TGCCGTACCAGAATTCGCTTAGATGT 3’

To inhibit restriction endonuclease activity through caging, 100 pmol of DNA was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO and brought to a final volume of 100 µL with nuclease-free water. DNA was caged at 50 °C for 2 h. DNA was then ethanol precipitated and resuspended in nuclease free water at a final concentration of 1 µM. To decage the strand, 10 pmol of caged DNA was incubated in 1X PBS, pH 7.5 at 95 °C for 5 minutes. DNA was then ethanol precipitated and resuspended at a concentrated of 1 µM. In separate tubes, 10 pmol of the untreated, caged, or decaged DNA was hybridized to 10 pmol of complementary DNA (5’ ACATCTAAGCGAATTCTGGTACGGCA 3’) in 1X CutSmart buffer (New England Biolabs) in a total volume of 100 µL for 30 minutes at 37 °C. After hybridization, 20 units of EcoRI-HF (New England Biolabs) was added to 10 µL of each sample. Reactions were incubated for one hour at 37 °C and halted with the addition of 1 µL of 125 mM EDTA. All samples were then analyzed on 12% denaturing PAGE.

RNase T Assay

To test RNase T, we used the same fluorescein labeled RNA as shown earlier in RNase H assays (5’ FAM rArArGrCrArGrCrArGrGrCrUrArUrGrUrUrArGrArArCrArArU 3’) and the same fluorescein labeled DNA as used in glyoxal caging kinetics (5’ FAM TGCCAAGACTGTTGAGGAAGATGAGAGAAT 3’). Caging and decaging conditions were also identical as described earlier. RNase T (New England Biolabs, 5 units) was added to 1 pmol of untreated, caged, or decaged RNA and DNA where appropriate in 10 µL 1X NEBuffer™ 4 (New

England Biolabs). Reactions were incubated for one hour at 25 °C and halted with the addition of 1 µL of 125 mM EDTA. All samples were then analyzed on 12% denaturing PAGE.

Phosphodiesterase I Assay

To test phosphodiesterase I, we used the same fluorescein labeled RNA as shown earlier in RNase H assays (5' FAM rArArGrCrArGrCrArGrGrCrUrArUrGrUrUrArGrArArCrArArU 3') and the same fluorescein labeled DNA as used in glyoxal caging kinetics (5' FAM TGCCAAGACTGTTGAGGAAGATGAGAGAAT 3'). Caging and decaging conditions were also identical as described earlier. Snake venom phosphodiesterase I isolated from *Crotalus adamanteus* (Sigma Aldrich, 0.005 units) was added to 1 pmol of untreated, caged, or decaged RNA and DNA where appropriate in 10 µL 1X CutSmart buffer (New England Biolabs). Reactions were incubated for one hour at 37 °C and halted with the addition of 1 µL of 125 mM EDTA. All samples were then analyzed on 12% denaturing PAGE.

Cas9 sgRNA Preparation

The following dsDNA gBlock containing a T7 RNA polymerase promoter (underlined) was purchased from IDT as shown below.

5'CCCGGGTTC*TAATACGACTCACTATA*GGAGCGCACCATCTTCTTCAGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG CTTTT 3'

Template was then amplified with a forward primer (5' CCCGGGTTCTAATACGACTCACTATAG 3') and reverse primer 5' AAAAGCACCGACTCGGTGC 3') using HotStart Taq DNA polymerase (New England Biolabs) according to the manufacturer's instructions, using the following PCR program: 94 °C for 3 min, followed by 30 cycles of (94 °C for 1 m, 56 °C for 45 s, 68 °C for 1 m), 68 °C for 5 m. PCR reactions were then purified using the Monarch PCR & DNA Cleanup Kit (New

England Biolabs). DNA was then loaded onto a 1% agarose gel and the desired 126 bp band was excised from the gel and purified using the Monarch DNA Gel Extraction Kit. RNA was then synthesized and purified using the HiScribe T7 High Yield RNA Synthesis Kit and Monarch RNA Cleanup Kit (New England Biolabs) to yield the full sgRNA sequence shown below:

5’GGAGCGCACCAUCUUCUUCAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG
UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU 3’

dsDNA Target Preparation

A 720 bp region of dsDNA was amplified from a plasmid encoding eGFP (Addgene #60733) sgRNA docking region in bold, cleavage site underlined:

5’ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGG
ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC
TACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
ACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATG
AAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCA**GGAGCGCACCATC**
TTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACAC
CCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGG
GGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGC
TCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGAC
AACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCAC
ATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTAC
AAGTAA 3’

Plasmid template (20 ng) was amplified with Hot Start Taq DNA Polymerase according to the instructions with a forward primer (5' ATGGTGAGCAAGGGCGAGGA 3') and a reverse primer (5' TTACTTGTACAGCTCGTCCATGCCGAGAG 3'). The following PCR steps were used: 94°C for 3 m, followed by 30 cycles of (94°C for 1 m, 60°C for 45 s, 68°C for 1 m), 68°C for 5 m. The PCR product was then purified using the Monarch PCR & DNA cleanup Kit.

Cas9 Cleavage Assays

For all CRISPR Cas9 cleavage reactions, 400 ng (12.44 pmol) of sgRNA was incubated with 500 ng (~3 pmol) of purified recombinant *Streptococcus pyogenes* Cas9 (IDT) in 10 µL 1X PBS, pH 7.5. Samples were incubated for 10 minutes at room temperature to allow for formation of RNP complex. Each RNP complex (10 µL) was then added to 200 ng of dsDNA target in a total volume of 20 µL 1X CRISPR Cas9 buffer (200 mM HEPES, 1 M NaCl, 50 mM $MgCl_2$, 1 mM EDTA, pH 6.5). The reactions were incubated at 37 °C for 2 hours and halted with the addition of 1 µL proteinase K (20 mg/mL, Thermo Fisher) and a 10-minute incubation at 56 °C. All samples were diluted with 20 µL nuclease free water and analyzed on 1% agarose gel with 1X SYBR Safe. Densitometry analysis was performed using ImageJ software. Percent cleavage was calculated by dividing the band intensity of the two cleaved product bands by the sum of all band intensities in that lane.

To inhibit Cas9 activity through caging, 2 µg sgRNA was added to separate PCR tubes and mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO and brought to a final volume of 100 µL with nuclease-free water. Reactions were incubated at 50 °C for 0 min, 1 min, 5 min, 10 min, 20 min, 30 min, 40 min, 1 h, 2 h, 4 h, 6 h, and 8 h. Immediately after each timepoint, each sample was ethanol precipitated and resuspended in nuclease free water to 100 ng/µL. The size shift from glyoxal addition to sgRNAs was then monitored by 10% denaturing PAGE stained with 1X SYBR Gold solution for 20 minutes.

To decage the sgRNA, 2 µg sgRNA was added to 100 µL 1X PBS, pH 7.5 and incubated at 95 °C for 0, 0.5, 1, 2, 3, 5, 8, and 10 minutes. Samples were ethanol precipitated and resuspended in nuclease free water to 100 ng/µL. Size shift from glyoxal removal was then monitored by 10% denaturing PAGE stained with 1X SYBR Gold solution for 20 minutes. Optimal timepoints for caging (2 h) and decaging (5 min) were then analyzed for size shift with untreated sgRNA on 10% denaturing PAGE with SYBR Gold staining as previously described.

For sgRNA decaging at 37 °C, 1 µg untreated or caged sgRNA was combined in a total of 100 µL 1X PBS and incubated at 37 °C for 0 h, 1 h, 2 h, 4 h, 6 h, 8 h, 1 d, and 2 d. The samples were ethanol precipitated and resuspended in nuclease free water to 100 ng/µL. The size shift from glyoxal removal was then monitored by running the samples on 10% denaturing PAGE stained with SYBR Gold. Each of the timepoints for both untreated and caged sgRNA were tested as described above for Cas9 cleavage activity and analyzed by 1% agarose gel.

β-actin PCR

The following DNA oligonucleotide primers were purchase from IDT as shown below:

Forward primer: 5’ AGAGATGGCCACGGCTGCTT 3’

Reverse primer: 5’ ATTTGCGGTGGACGATGGAG 3’

For the initial test, a 653 bp segment of the *ACTB* gene was amplified from 0 ng, 0.1 ng, 1 ng, 10 ng, or 100 ng human genomic DNA (Promega) with standard *Taq* DNA polymerase (New England Biolabs) according to the manufacturer’s instructions using a final primer concentration of 0.2 µM in each reaction. Thermal cycler conditions were: 94 °C for 2 min, followed by 35 cycles of (94 °C for 30 s, 60 °C for 30 s, 72 °C for 45 s), followed by a final extension at 72 °C for 7 min. 20 µL of each crude PCR reaction were analyzed on 1% agarose gel with 1X SYBR Safe and visualized with a Typhoon RGB scanner. Percent purity was calculated by densitometry analysis using ImageJ software by dividing the target band intensity by the sum of all lane bands. For PCR

reactions using Hot Start *Taq* (New England Biolabs), this experimental set up was repeatedly identically.

Primer Caging

For single primer caging, 152 pmol (1 µg) or either forward or reverse primer was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO. All reactions were brought to a final volume of 100 µL with nuclease-free water and reacted at 50 ºC for 0 min, 2 min, 10 min, 20 min, 30 min, 40 min, 1 hour, and 2 hour. At each time point, reactions were ethanol precipitated and reconstituted with equimolar amounts of the opposite primer in nuclease-free water at a final stock concentration of 1 µM. For caging of both primers simultaneously, 1 µg of both forward and reverse primer (2 µg total) was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO. All reactions were brought to a final volume of 100 µL with nuclease-free water and reacted at 50 ºC for 0 min, 2 min, 10 min, 20 min, 30 min, 40 min, 1 hour, and 2 hour. At each time point, reactions were ethanol precipitated and reconstituted with nuclease-free water at a final concentration of 1 µM. An untreated primer mix was also prepared using 1 µg each of forward and reverse primer mixed with 50 µL DMSO, followed by immediate ethanol precipitation. To assess primer caging by gel, 5 pmol of each purified reaction from the caged primer mix was then mixed with 2X RNA loading dye (New England Biolabs) and then separated with 20% denaturing PAGE. Gels were then stained with 1X SYBR Gold solution (ThermoFisher Scientific) for 20 minutes and visualized with the Typhoon RGB scanner.

For functional PCR testing of the single or double caged primer mix, *ACTB* was amplified using standard Taq polymerase as described above with 100 ng human genomic DNA and a final primer concentration of 0.2 µM. To assess overall yield, PCR reactions were first purified using the Monarch PCR & DNA Cleanup Kit (New England Biolabs) and then dsDNA was quantified using a BioTek Take3 spectrophotometer. Target yield was calculated by multiplying quantified mass (ng) by the purity percentage.

Primer Decaging

10 pmol of caged primer mix (10 minute glyoxal treatment) was added to 25 µL of 1X Standard Taq Reaction buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, pH 8.3). Samples were incubated for the specified thermocycler steps and then 20 ul of each sample was analyzed by 20% denaturing PAGE as described above.

ASOs and Plasmids

The following LNA gapmer oligonucleotides were purchased from IDT as shown below. "+" indicates LNA nucleobases and "*" denotes phosphorothioate modification.

Anti-eGFP: 5' +G*+A*+A*C*T*T*C*A*G*G*G*T*C*+A*+G*+C 3'

Scramble: 5' +A*+G*+G*A*C*G*A*C*T*C*T*A*G*+G*+C*+T 3'

To construct a destabilized eGFP vector, a pCMV plasmid harboring the eGFP coding sequence (https://www.addgene.org/11153/) was digested with restriction enzymes bsrGI and NotI (New England Biolabs) and a dsDNA sequence encoding the CL1 degron tag (ACKNWFSSLSHFVIHL) was cloned into the vector at the eGFP C-terminus. Correct insertion was verified by Sanger sequencing (full plasmid map is included in additional supplemental materials as a SnapGene file).

Transfection and Live-Cell Imaging

HEK293T cells (ATCC CRL-3216) were cultured in Dulbecco's Modified Eagle's Medium supplemented with 10% fetal bovine serum and maintained at 37 ºC, 5% CO_2. Cells were seeded at 10,000 cells/well in tissue culture-treated clear polysterene 96-Well Plates (Costar) followed by overnight recovery. For initial tests of GFP expression, 200 ng of constructed eGFP plasmid was introduced into cells with Lipofectamine 3000 (Thermo Fisher Scientific) according to the manufacturer's instructions. After ~12 h incubation at 37 ºC, 5% CO_2, cells were then imaged at

4X magnification using a Biotek FX Lionheart automated live-cell microscope, and images were processed using ImageJ.

To optimize ASO delivery, cells were seeded into 96-well plates and transfected with 200 ng of plasmid using Lipofectamine 3000 as described above. Cells were then exposed to a range of concentrations (0 – 500 nM) of both anti-eGFP and scrambled ASO in 0.2 mL DMEM per well. After ~12 h incubation at 37 ºC, 5% CO_2, cells were then imaged using a Biotek FX Lionheart as described.

Increasingly caged ASO samples were generated by combining 2 µg of anti-eGFP ASO with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO. All reactions were brought to a final volume of 100 µL with nuclease-free water and reacted at 50 ºC for 0 min, 5 min, 10 min, 20 min, 30 min, 1 hour, 2 hours, 4 hours, 6 hours, and 8 hours. At each time point, reactions were ethanol precipitated and reconstituted in nuclease free water. To assess caging by gel, 5 pmol of each purified reaction was then mixed with 2X RNA loading dye (New England Biolabs) and then separated with 20% denaturing PAGE. Gels were then stained with 1X SYBR Gold solution (ThermoFisher Scientific) for 20 minutes and visualized with the Typhoon RGB scanner.

To assess cellular uptake, anti-eGFP ASO was first fluorescently labeled with Cyanine 5 (Cy5) using the Label IT® Tracker™ Intracellular Nucleic Acid Localization Kit (Mirus Bio) according to the manufacturer's instructions and purified by ethanol precipitation. A portion of this material was set aside (untreated ASO), and in separate PCR tubes, 2 µg ASO was mixed with 14.5 µL of a 40% glyoxal solution (Sigma Aldrich) and 50 µL DMSO and brought to a final volume of 100 µL with nuclease-free water. Reactions were incubated at 50 °C 8 h, after which it was ethanol precipitated and resuspended in nuclease free water. HEK293T cells were seeded into 96-well plates as described previously, and cells were exposed to untreated of glyoxal caged ASO at a final concentration of 250 nM in 0.2 mL DMEM per well. Cells were incubated overnight at 37

ºC, 5% CO_2, and each well was carefully washed three times with 0.2 mL of prewarmed fresh DMEM, followed by fluorescent microscopy using the Biotek FX lionheart.

To assess GFP expression across different time points, HEK293T cells were seeded into 96-well plates as described previously, and cells were transfected with 200 ng of plasmid using Lipofectamine 3000. Cells were then exposed to 250 nM of increasingly caged anti-eGFP in 0.2 mL DMEM per well. At the indicated time points cells were then imaged using a Biotek FX Lionheart as described. On days 4 and 7 of the experiment, 100 µL of media was carefully removed from the top of each well and replaced with prewarmed fresh complete DMEM. Images were acquired at 4x magnification in the center of each well, and exposure parameters were identical across cell samples, with GFP exposure times always set at each time point according to “no ASO” sample wells. Raw images were then processed and colorized in ImageJ. To calculate GFP-positive cells per well, fluorescent image thresholds were first normalized using the Huang algorithm, followed by “analyze” particles” ImageJ plugin for cell quantification using a lower area range set at 10 pixels/sq-inch.

Cell Viability Assays

To test viability, plates from the full time-course decaging experiment were tested using WST-1 (Sigma Aldrich) according to the manufacturer’s protocol. In brief, on day 7 after acquiring final images, media was carefully replaced with fresh 100 µL DMEM, and 10 µL of the WST-1 reagent was added to each well. Plates were incubated at 37 ºC, 5% CO_2 for four hours, and absorbance of each well at 440 nm was measured. Separately, glyoxal alone was also tested after seeding HEK293T cells at 10,000 cells per well as described previously. Cells were then exposed to a range of glyoxal concentrations (0 – 1 mM) in 0.2 mL DMEM. On days 4 and 7 of the experiment, 100 µL of media was carefully removed from the top of each well and replaced with prewarmed fresh complete DMEM supplemented with the appropriate amount of glyoxal. After 7 days, viability was assayed using WST-1. Percent viability was calculated as the net

absorbance value compared to untreated HEK293T cells. IC_{50} value (mean with 95% confidence interval) was calculated using a dose-response curve fit in Prism.

6.6 References

1. 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. *Journal of the American Chemical Society* **2020,** *142* (41), 17766-17781.

2. Seeman, N. C.; Sleiman, H. F., DNA nanotechnology. *Nature Reviews Materials* **2018,** *3* (1), 17068.

3. Guo, P., The emerging field of RNA nanotechnology. *Nature nanotechnology* **2010,** *5* (12), 833.

4. Wu, C.; Wan, S.; Hou, W.; Zhang, L.; Xu, J.; Cui, C.; Wang, Y.; Hu, J.; Tan, W., A survey of advancements in nucleic acid-based logic gates and computing for applications in biotechnology and biomedicine. *Chemical Communications* **2015,** *51* (18), 3723-3734.

5. Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M., The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology* **2019**, 1.

6. Höbartner, C.; Mittendorfer, H.; Breuker, K.; Micura, R., Triggering of RNA secondary structures by a functionalized nucleobase. *Angewandte Chemie International Edition* **2004,** *43* (30), 3922-3925.

7. Ikeda, M.; Kamimura, M.; Hayakawa, Y.; Shibata, A.; Kitade, Y., Reduction-Responsive Guanine Incorporated into G-Quadruplex-Forming DNA. *ChemBioChem* **2016,** *17* (14), 1304-1307.

8. Govan, J. M.; Young, D. D.; Lusic, H.; Liu, Q.; Lively, M. O.; Deiters, A., Optochemical control of RNA interference in mammalian cells. *Nucleic acids research* **2013,** *41* (22), 10518-10528.

9. Mikat, V.; Heckel, A., Light-dependent RNA interference with nucleobase-caged siRNAs. *Rna* **2007,** *13* (12), 2341-2347.

10. Lucas, T.; Schäfer, F.; Müller, P.; Eming, S. A.; Heckel, A.; Dimmeler, S., Light-inducible antimiR-92a as a therapeutic strategy to promote skin repair in healing-impaired diabetic mice. *Nature communications* **2017,** *8*, 15162.

11. Chaulk, S. G.; MacMillan, A. M., Synthesis of oligo-RNAs with photocaged adenosine 2′-hydroxyls. *Nature protocols* **2007,** *2* (5), 1052.

12. Matsushita-Ishiodori, Y.; Ohtsuki, T., Photoinduced RNA interference. *Accounts of chemical research* **2012,** *45* (7), 1039-1047.

13. Meyer, A.; Mokhir, A., RNA interference controlled by light of variable wavelength. *Angewandte Chemie* **2014,** *126* (47), 13054-13057.

14. Lu, J.; Koo, S. C.; Li, N.-S.; Piccirilli, J. A., Synthesis of 2′-O-Photocaged Ribonucleoside Phosphoramidites. *Nucleosides, Nucleotides and Nucleic Acids* **2015,** *34* (2), 114-129.

15. Tan, Z.; Feagin, T. A.; Heemstra, J. M., Temporal control of aptamer biosensors using covalent self-caging to shift equilibrium. *Journal of the American Chemical Society* **2016,** *138* (20), 6328-6331.

16. Kadina, A.; Kietrys, A. M.; Kool, E. T., RNA Cloaking by Reversible Acylation. *Angewandte Chemie International Edition* **2018,** *57* (12), 3059-3063.

17. Velema, W. A.; Kietrys, A. M.; Kool, E. T., RNA control by photoreversible acylation. *Journal of the American Chemical Society* **2018,** *140* (10), 3491-3495.

18. Durbeej, B.; Eriksson, L. A., Reaction mechanism of thymine dimer formation in DNA induced by UV light. *Journal of Photochemistry and Photobiology A: Chemistry* **2002,** *152* (1-3), 95-101.

19. Staehelin, M., Inactivation of virus nucleic acid with glyoxal derivatives. *Biochimica et biophysica acta* **1959,** *31* (2), 448-454.

20. Nakaya, K.; Takenaka, O.; Horinishi, H.; Shibata, K., Reactions of glyoxal with nucleic acids, nucleotides and their component bases. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis* **1968,** *161* (1), 23-31.

21. Shapiro, R.; Cohen, B. I.; Shiuey, S.-J.; Maurer, H., Reaction of guanine with glyoxal, pyruvaldehyde, and kethoxal, and the structure of the acylguanines. Synthesis of N2-alkylguanines. *Biochemistry* **1969,** *8* (1), 238-245.

22. Shapiro, R.; Cohen, B. I.; Clagett, D. C., Specific acylation of the guanine residues of ribonucleic acid. *Journal of Biological Chemistry* **1970,** *245* (10), 2633-2639.

23. Aubert, M.; Bellemare, G.; Monier, R., Selective reaction of glyoxal with guanine residues in native and denatured Escherichia coli 5S RNA. *Biochimie* **1973,** *55* (2), 135-142.

24. McMaster, G. K.; Carmichael, G. G., Analysis of single-and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proceedings of the National Academy of Sciences* **1977,** *74* (11), 4835-4838.

25. Murant, A.; Taylor, M.; Duncan, G.; Raschke, J., Improved estimates of molecular weight of plant virus RNA by agarose gel electrophoresis and electron microscopy after denaturation with glyoxal. *Journal of General Virology* **1981,** *53* (2), 321-332.

26. Bussolati, G.; Annaratone, L.; Berrino, E.; Miglio, U.; Panero, M.; Cupo, M.; Gugliotta, P.; Venesio, T.; Sapino, A.; Marchiò, C., Acid-free glyoxal as a substitute of formalin for structural and molecular preservation in tissue samples. *PloS one* **2017,** *12* (8), e0182965.

27. Weng, X.; Gong, J.; Chen, Y.; Wu, T.; Wang, F.; Yang, S.; Yuan, Y.; Luo, G.; Chen, K.; Hu, L., Keth-seq for transcriptome-wide RNA structure mapping. *Nature Chemical Biology* **2020,** *16* (5), 489-492.

28. Burnett, W. V., Northern blotting of RNA denatured in glyoxal without buffer recirculation. *BioTechniques* **1997,** *22* (4), 668-671.

29. Morse, D. P.; Bass, B. L., Detection of inosine in messenger RNA by inosine-specific cleavage. *Biochemistry* **1997,** *36* (28), 8429-8434.

30. Cattenoz, P. B.; Taft, R. J.; Westhof, E.; Mattick, J. S., Transcriptome-wide identification of A> I RNA editing sites by inosine specific cleavage. *Rna* **2013,** *19* (2), 257-270.

31. Gareau, D.; Desrosiers, A.; Vallée-Bélisle, A., Programmable quantitative DNA nanothermometers. *Nano letters* **2016,** *16* (7), 3976-3981.

32. Ke, G.; Wang, C.; Ge, Y.; Zheng, N.; Zhu, Z.; Yang, C. J., L-DNA molecular beacon: a safe, stable, and accurate intracellular nano-thermometer for temperature sensing in living cells. *Journal of the American Chemical Society* **2012,** *134* (46), 18908-18911.

33. Jonstrup, A. T.; Fredsøe, J.; Andersen, A. H., DNA hairpins as temperature switches, thermometers and ionic detectors. *Sensors* **2013,** *13* (5), 5937-5944.

34. Ebrahimi, S.; Akhlaghi, Y.; Kompany-Zareh, M.; Rinnan, Å., Nucleic acid based fluorescent nanothermometers. *ACS nano* **2014,** *8* (10), 10372-10382.

35. Modi, S.; Swetha, M.; Goswami, D.; Gupta, G. D.; Mayor, S.; Krishnan, Y., A DNA nanomachine that maps spatial and temporal pH changes inside living cells. *Nature nanotechnology* **2009,** *4* (5), 325-330.

36. Chen, X.; Chen, T.; Ren, L.; Chen, G.; Gao, X.; Li, G.; Zhu, X., Triplex DNA nanoswitch for pH-sensitive release of multiple cancer drugs. *ACS nano* **2019,** *13* (6), 7333-7344.

37. Deng, Y.; Ma, L.; Han, Q.; Yu, C.; Johnson-Buck, A.; Su, X., DNA-Templated Timer Probes for Multiplexed Sensing. *Nano Letters* **2020,** *20* (4), 2688-2694.

38. Schaffter, S. W.; Schulman, R., Building in vitro transcriptional regulatory networks by successively integrating multiple functional circuit modules. *Nature chemistry* **2019,** *11* (9), 829-838.

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

40. Wilson, S. C.; Cohen, D. T.; Wang, X. C.; Hammond, M. C., A neutral pH thermal hydrolysis method for quantification of structured RNAs. *Rna* **2014,** *20* (7), 1153-1160.

41. Santoro, S. W.; Joyce, G. F., A general purpose RNA-cleaving DNA enzyme. *Proceedings of the national academy of sciences* **1997,** *94* (9), 4262-4266.

42. Young, D. D.; Lively, M. O.; Deiters, A., Activation and deactivation of DNAzyme and antisense function with light for the photochemical regulation of gene expression in mammalian cells. *Journal of the American Chemical Society* **2010,** *132* (17), 6183-6193.

43. Hwang, K.; Wu, P.; Kim, T.; Lei, L.; Tian, S.; Wang, Y.; Lu, Y., Photocaged DNAzymes as a general method for sensing metal ions in living cells. *Angewandte Chemie International Edition* **2014,** *53* (50), 13798-13802.

44. Burmeister, P. E.; Lewis, S. D.; Silva, R. F.; Preiss, J. R.; Horwitz, L. R.; Pendergrast, P. S.; McCauley, T. G.; Kurz, J. C.; Epstein, D. M.; Wilson, C., Direct in vitro selection of a 2′-O-methyl aptamer to VEGF. *Chemistry & biology* **2005,** *12* (1), 25-33.

45. Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A., NUPACK: analysis and design of nucleic acid systems. *Journal of computational chemistry* **2011,** *32* (1), 170-173.

46. Schöning, K.-U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A., Chemical etiology of nucleic acid structure: the α-threofuranosyl-(3'→ 2') oligonucleotide system. *Science* **2000,** *290* (5495), 1347-1351.

47. Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H., Peptide nucleic acids (PNA). Oligonucleotide analogs with an achiral peptide backbone. *Journal of the American Chemical Society* **1992,** *114* (5), 1895-1897.

48. Jerabek-Willemsen, M.; André, T.; Wanner, R.; Roth, H. M.; Duhr, S.; Baaske, P.; Breitsprecher, D., MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure* **2014,** *1077*, 101-113.

49. Jacob, D.; Thüring, K.; Galliot, A.; Marchand, V.; Galvanin, A.; Ciftci, A.; Scharmann, K.; Stock, M.; Roignant, J. Y.; Leidel, S. A., Absolute quantification of noncoding RNA by microscale thermophoresis. *Angewandte Chemie International Edition* **2019**.

50. Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M.; Norde, B., DNA-like double helix formed by peptide nucleic acid. *Nature* **1994,** *368* (6471), 561.

51. Miller, H. I.; Riggs, A. D.; Gill, G. N., Ribonuclease H (Hybrid) in Escherichia coli IDENTIFICATION AND CHARACTERIZATION. *Journal of Biological Chemistry* **1973,** *248* (7), 2621-2624.

52. Itaya, M.; Kondo, K., Molecular cloning of a ribonuclease H (RNase HI) gene from an extreme thermophile Thermus thermophilus HB8: a thermostable RNase H can functionally replace the Escherichia coli enzyme in vivo. *Nucleic acids research* **1991,** *19* (16), 4443-4449.

53. Findly, D.; Herries, D.; Mathias, A.; Rabin, B.; Ross, C., The active site and mechanism of action of bovine pancreatic ribonuclease. *Nature* **1961,** *190* (4778), 781-784.

54. Fujimoto, M.; Kuninaka, A.; Yoshino, H., Substrate specificity of nuclease P1. *Agricultural and Biological Chemistry* **1974,** *38* (9), 1555-1561.

55. Vanecko, S.; Laskowski Sr, M., Studies of the specificity of deoxyribonuclease I. II. Hydrolysis of oligonucleotides carrying a monoesterified phosphate on carbon 3'. *Journal of Biological Chemistry (US)* **1961,** *236*.

56. Deutscher, M. P.; Marlor, C. W.; Zaniewski, R., Ribonuclease T: new exoribonuclease possibly involved in end-turnover of tRNA. *Proceedings of the National Academy of Sciences* **1984,** *81* (14), 4290-4293.

57. Viswanathan, M.; Lanjuin, A.; Lovett, S. T., Identification of RNase T as a high-copy suppressor of the UV sensitivity associated with single-strand DNA exonuclease deficiency in Escherichia coli. *Genetics* **1999,** *151* (3), 929-934.

58. Williams, E. J.; Sung, S.-C.; Laskowski, M., Action of venom phosphodiesterase on deoxyribonucleic acid. *Journal of Biological Chemistry* **1961,** *236* (4), 1130-1134.

59. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E., A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science* **2012,** *337* (6096), 816-821.

60. Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A., Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013,** *339* (6121), 819-823.

61. Zhou, W.; Brown, W.; Bardhan, A.; Delaney, M.; Ilk, A. S.; Rauen, R. R.; Kahn, S. I.; Tsang, M.; Deiters, A., Spatiotemporal Control of CRISPR/Cas9 Function in Cells and Zebrafish using Light-Activated Guide RNA. *bioRxiv* **2019**, 831974.

62. Wang, S.-R.; Wu, L.-Y.; Huang, H.-Y.; Xiong, W.; Liu, J.; Wei, L.; Yin, P.; Tian, T.; Zhou, X., Conditional control of RNA-guided nucleic acid cleavage and gene editing. *Nature Communications* **2020,** *11* (1), 1-10.

63. Habibian, M.; McKinlay, C.; Blake, T. R.; Kietrys, A. M.; Waymouth, R. M.; Wender, P. A.; Kool, E. T., Reversible RNA acylation for control of CRISPR–Cas9 gene editing. *Chemical Science* **2020**.

64. Lin, B.; An, Y.; Meng, L.; Zhang, H.; Song, J.; Zhu, Z.; Liu, W.; Song, Y.; Yang, C., Control of CRISPR-Cas9 with small molecule-activated allosteric aptamer regulating sgRNAs. *Chemical Communications* **2019,** *55* (81), 12223-12226.

65. Raper, A. T.; Stephenson, A. A.; Suo, Z., Functional insights revealed by the kinetic mechanism of CRISPR/Cas9. *Journal of the American Chemical Society* **2018,** *140* (8), 2971-2984.

66. Konermann, S.; Brigham, M. D.; Trevino, A. E.; Joung, J.; Abudayyeh, O. O.; Barcena, C.; Hsu, P. D.; Habib, N.; Gootenberg, J. S.; Nishimasu, H., Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **2015,** *517* (7536), 583.

67. Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **2013,** *152* (5), 1173-1183.

68. Komor, A. C.; Kim, Y. B.; Packer, M. S.; Zuris, J. A.; Liu, D. R., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **2016,** *533* (7603), 420-424.

69. Anzalone, A. V.; Randolph, P. B.; Davis, J. R.; Sousa, A. A.; Koblan, L. W.; Levy, J. M.; Chen, P. J.; Wilson, C.; Newby, G. A.; Raguram, A., Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **2019,** *576* (7785), 149-157.

70. Rauch, S.; He, E.; Srienc, M.; Zhou, H.; Zhang, Z.; Dickinson, B. C., Programmable RNA-guided RNA effector proteins built from human parts. *Cell* **2019,** *178* (1), 122-134. e12.

71. Wang, A. M.; Doyle, M. V.; Mark, D. F., Quantitation of mRNA by the polymerase chain reaction. *Proceedings of the National Academy of Sciences* **1989,** *86* (24), 9717-9721.

72. Chan, J. F.-W.; Yip, C. C.-Y.; To, K. K.-W.; Tang, T. H.-C.; Wong, S. C.-Y.; Leung, K.-H.; Fung, A. Y.-F.; Ng, A. C.-K.; Zou, Z.; Tsoi, H.-W., Improved molecular diagnosis of COVID-19 by the novel, highly sensitive and specific COVID-19-RdRp/Hel real-time reverse transcription-PCR assay validated in vitro and with clinical specimens. *Journal of Clinical Microbiology* **2020,** *58* (5).

73. Rychlik, W., Selection of primers for polymerase chain reaction. *Molecular biotechnology* **1995,** *3* (2), 129-134.

74. Kellogg, D.; Rybalkin, I.; Chen, S.; Mukhamedova, N.; Vlasik, T.; Siebert, P.; Chenchik, A., TaqStart Antibody:" hot start" PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. *Biotechniques* **1994,** *16* (6), 1134-1137.

75. Lin, Y.; Jayasena, S. D., Inhibition of multiple thermostable DNA polymerases by a heterodimeric aptamer. *Journal of molecular biology* **1997,** *271* (1), 100-111.

76. Young, D. D.; Edwards, W. F.; Lusic, H.; Lively, M. O.; Deiters, A., Light-triggered polymerase chain reaction. *Chemical communications* **2008,** (4), 462-464.

77. Lebedev, A. V.; Paul, N.; Yee, J.; Timoshchuk, V. A.; Shum, J.; Miyagi, K.; Kellum, J.; Hogrefe, R. I.; Zon, G., Hot start PCR with heat-activatable primers: a novel approach for improved PCR performance. *Nucleic acids research* **2008,** *36* (20), e131-e131.

78. Jost, M.; Santos, D. A.; Saunders, R. A.; Horlbeck, M. A.; Hawkins, J. S.; Scaria, S. M.; Norman, T. M.; Hussmann, J. A.; Liem, C. R.; Gross, C. A., Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nature Biotechnology* **2020,** *38* (3), 355-364.

79. Michaels, Y. S.; Barnkob, M. B.; Barbosa, H.; Baeumler, T. A.; Thompson, M. K.; Andre, V.; Colin-York, H.; Fritzsche, M.; Gileadi, U.; Sheppard, H. M., Precise tuning of gene expression levels in mammalian cells. *Nature communications* **2019,** *10* (1), 1-12.

80. Castanotto, D.; Lin, M.; Kowolik, C.; Wang, L.; Ren, X.-Q.; Soifer, H. S.; Koch, T.; Hansen, B. R.; Oerum, H.; Armstrong, B., A cytoplasmic pathway for gapmer antisense oligonucleotide-mediated gene silencing in mammalian cells. *Nucleic acids research* **2015,** *43* (19), 9350-9361.

81. Liang, X.-H.; Sun, H.; Nichols, J. G.; Crooke, S. T., RNase H1-dependent antisense oligonucleotides are robustly active in directing RNA cleavage in both the cytoplasm and the nucleus. *Molecular Therapy* **2017,** *25* (9), 2075-2092.

82. Stein, C.; Hansen, J. B.; Lai, J.; Wu, S.; Voskresenskiy, A.; H∅ g, A.; Worm, J.; Hedtjärn, M.; Souleimanian, N.; Miller, P., Efficient gene silencing by delivery of locked nucleic acid antisense oligonucleotides, unassisted by transfection reagents. *Nucleic acids research* **2010,** *38* (1), e3-e3.

83. Petersen, M.; Wengel, J., LNA: a versatile tool for therapeutics and genomics. *Trends in biotechnology* **2003,** *21* (2), 74-81.

84. Lim, K. R. Q.; Maruyama, R.; Echigoya, Y.; Nguyen, Q.; Zhang, A.; Khawaja, H.; Chandra, S. S.; Jones, T.; Jones, P.; Chen, Y.-W., Inhibition of DUX4 expression with antisense LNA gapmers as a therapy for facioscapulohumeral muscular dystrophy. *Proceedings of the National Academy of Sciences* **2020,** *117* (28), 16509-16515.

85. Zimmer, M., Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chemical reviews* **2002,** *102* (3), 759-782.

86. He, L.; Binari, R.; Huang, J.; Falo-Sanjuan, J.; Perrimon, N., In vivo study of gene expression with an enhanced dual-color fluorescent transcriptional timer. *Elife* **2019,** *8*, e46181.

87. Corish, P.; Tyler-Smith, C., Attenuation of green fluorescent protein half-life in mammalian cells. *Protein engineering* **1999,** *12* (12), 1035-1040.

88. Kiełpiński, Ł. J.; Hagedorn, P. H.; Lindow, M.; Vinther, J., RNase H sequence preferences influence antisense oligonucleotide efficiency. *Nucleic acids research* **2017,** *45* (22), 12932-12944.

89. Magner, D.; Biala, E.; Lisowiec-Wachnicka, J.; Kierzek, R., Influence of mismatched and bulged nucleotides on SNP-preferential RNase H cleavage of RNA-antisense gapmer heteroduplexes. *Scientific reports* **2017,** *7* (1), 1-16.

90. Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W., Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **2005,** *121* (7), 1005-1016.

91. Grishagin, I. V., Automatic cell counting with ImageJ. *Analytical biochemistry* **2015,** *473*, 63-65.

92. Asokan, A.; Cho, M. J., Exploitation of intracellular pH gradients in the cellular delivery of macromolecules. *Journal of pharmaceutical sciences* **2002,** *91* (4), 903-913.

93. Näreoja, T.; Deguchi, T.; Christ, S.; Peltomaa, R.; Prabhakar, N.; Fazeli, E.; Perälä, N.; Rosenholm, J. M.; Arppe, R.; Soukka, T., Ratiometric sensing and imaging of intracellular pH using polyethylenimine-coated photon upconversion nanoprobes. *Analytical chemistry* **2017,** *89* (3), 1501-1508.

94. Kulkarni, C.; Nadtochiy, S. M.; Kennedy, L.; Zhang, J.; Chhim, S.; Alwaseem, H.; Murphy, E.; Fu, D.; Brookes, P. S., ALKBH7 mediates necrosis via rewiring of glyoxal metabolism. *bioRxiv* **2020**.

95. Lee, C.; Kim, I.; Lee, J.; Lee, K.-L.; Min, B.; Park, C., Transcriptional activation of the aldehyde reductase YqhD by YqhC and its implication in glyoxal metabolism of Escherichia coli K-12. *Journal of bacteriology* **2010,** *192* (16), 4205-4214.

96. Mlakar, A.; Batna, A.; Dudda, A.; Spiteller, G., Iron (II) ions induced oxidation of ascorbic acid and glucose. *Free radical research* **1996,** *25* (6), 525-539.

97. Sliman, S. M.; Eubank, T. D.; Kotha, S. R.; Kuppusamy, M. L.; Sherwani, S. I.; Butler, E. S. O. C.; Kuppusamy, P.; Roy, S.; Marsh, C. B.; Stern, D. M., Hyperglycemic oxoaldehyde, glyoxal, causes barrier dysfunction, cytoskeletal alterations, and inhibition of angiogenesis in vascular endothelial cells: aminoguanidine protection. *Molecular and cellular biochemistry* **2010,** *333* (1-2), 9-26.

98. Shangari, N.; O'Brien, P. J., The cytotoxic mechanism of glyoxal involves oxidative stress. *Biochemical pharmacology* **2004,** *68* (7), 1433-1442.

99. Tominaga, H.; Ishiyama, M.; Ohseto, F.; Sasamoto, K.; Hamamoto, T.; Suzuki, K.; Watanabe, M., A water-soluble tetrazolium salt useful for colorimetric cell viability assay. *Analytical Communications* **1999,** *36* (2), 47-50.

Chapter 7

Conclusions and Future Perspectives

RNA is a critical biomolecule in all life on Earth, serving as both an information-storage polymer as well as a direct catalytic participant in cellular reactions. Although initially biosynthesized from the four canonical nucleosides (A, U, C, and G), RNA is biochemically modified by a number of enzymes to significantly expand the nucleobase content of individual transcripts. We are just beginning to understand the biological importance of these modifications, and they have profound clinical implications for disease diagnosis and management. Adenosine-to-Inosine (A-to-I) RNA editing is one of the most widespread, conserved, and impactful of these RNA modifications, and serves a number of critical functions in the cell.[1-2] We now understand that A-to-I editing in mRNA actively recodes proteins in key pathways, facilitating proper cell and tissue function throughout the body. [2-3] It is also increasingly clear that A-to-I editing plays a vital role in regulating the innate immune system and is often dysregulated in autoimmune disease and several cancer types.[4-8]

Despite this broad functional importance, numerous major technical limitations render it difficult to study or harness A-to-I editing, and several key questions remain. First, because it is difficult to comprehensively map editing sites in total RNA, the true prevalence and landscape of A-to-I editing in the human transcriptome is still unknown. Although current databases have identified ~5-10 million editing sites in human RNA, this number is estimated to be ~10-fold higher.[9] While the vast majority of these sites occur in repetitive regions, a small number of protein-recoding events produce significant biological consequences.[2-3] Because technical limitations in existing methods continue to make the detection these rare sites problematic, it it likely that many of these events remain undiscovered. Although not as impactful on an individual basis, repetitive Alu editing sites across the transcriptome also collectively contribute to overall tissue development and immune system regulation.[4-8] While potential functional roles have been

hypothesized for the distinct global editing patterns observed between different cell and tissue types, the biological significance of these differences remain somewhat mysterious.

Large-scale analyses are beginning to shed light on ADAR-targeting mechanisms, but it is also unclear why certain sites are edited over others and what precise function each individual editing serves. While some sites are likely to be critical, it may also be possible that a large number of sites under "stochastic" editing based on RNA availability and structure equilibriums. ADAR expression is also poorly correlated with overall inosine content in cellular RNA,[2] and there is great interest in identifying mechanisms that regulate both ADAR activity and overall editing levels. Sampling a large number of biological contexts will likely reveal additional editing patterns and hint at context-specific functions and regulatory mechanisms, but the present requirement for lengthy and high-cost RNA-seq limits our ability to efficiently explore these settings.

Lastly, while it is known that dysregulated editing is linked with numerous diseases,[10-18] the exact molecular and cellular mechanisms responsible for this relationship are also poorly understood. We also have indirect evidence that different cell types and tissues display vastly different editing patterns, but there are no approaches to measure these differences. As a result, we have little understanding of how this heterogeneity affects overall tissue and organ function which might contribute to disease. The ability to pharmacologically control global editing levels is also of great interest to the field, and would enhance our ability to treat disease and elucidate these tissue-level consequences.

In this dissertation, I have described the development of a suite of technologies for probing A-to-I RNA editing and addressing many of these methodological limitations. In Chapters 2 and 3, we designed and tested acrylamide derivatives as an improved chemical approach for inosine detection in RNA. We next investigated Endonuclease V as a protein-based detection platform, illustrating its ability to enrich edited transcripts prior to RNA-seq (Chapter 4) as well as profile large-scale changes in transcriptome-wide A-to-I signatures (Chapter 5). Finally, we were inspired by our use of glyoxal in these earlier pursuits and leveraged this chemical labeling step as a way

to impart expanded functionality to a variety of substrates, in turn serving as a “synthetic” nucleic acid modification (Chapter 6). Here, we discuss how these different methodologies can address knowledge gaps in key areas surrounding nucleic acid biology and RNA editing, and we also outline how these platforms can be improved for more robust implementation in a variety of contexts.

7.1 Chemical Profiling of ADAR Mechanisms and Substrate Preferences

Inosine can be chemically labeled with acrylonitrile or acrylamide derivatives to form N^1 addition products (Chapters 2 and 3). While our lab as well as other groups have demonstrated feasibility for modifying simple RNA substrates, these reagents also display off-target labeling with other nucleobases and are generally intractable for use with complex RNA samples.[19-22] Despite this limitation, we found great utility with this reagent for profiling *in vitro* deamination kinetics with a chemically synthesized RNA substrate, and detected significant catalytic differences between wild-type and mutant ADAR1 isoforms (Chapter 3). ADAR targeting mechanisms and RNA substrate recognition preferences are still not well understood, and we envision that our acrylamidofluorescein or EPhAA reagents would serve as a valuable assay for rapidly detecting these properties (**Figure 7.1**). In particular, it would be economical to screen small libraries RNA substrates to determine which motifs and structures are robustly edited, and what minimal sequences are required for recognition. Similar to our earlier comparison of wild-type and E1008Q ADAR1 (Chapter 3), this platform can also be easily applied to biochemically characterizing a panel of enzyme mutants for elucidating how different ADAR isoforms interact with RNA substrates and which amino acids are key for catalysis. Additionally, while this method is not immediately tractable for high-throughput drug screening, we do envision this assay as an facile downstream step for validating potential small-molecule hits that influence ADAR activity and rapidly characterizing various pharmacological properties (K_d, K_i, IC_{50}, etc).

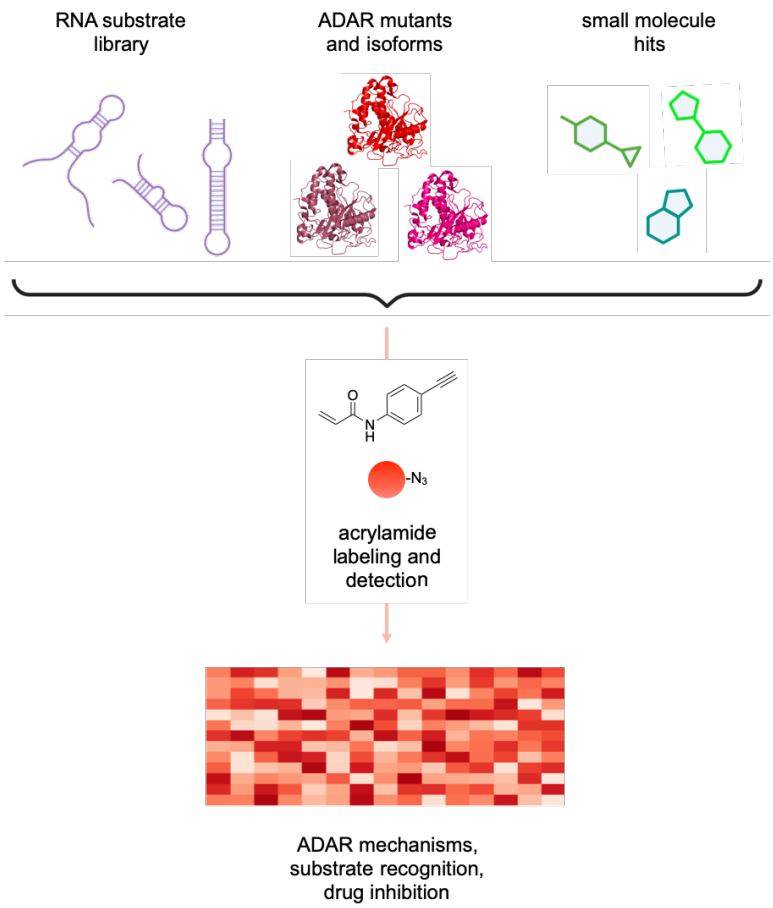


Figure 7.1. Acrylamide labeling enables efficient profiling of RNA substrate recognition, ADAR mutant characteristics, and small-molecule hit validation for modulating editing activity.

7.2 Engineering EndoV for Enhanced Inosine Recognition

As described in Chapters 4 and 5, the unique molecular recognition capabilities of EndoV can be leveraged for selective enrichment and global detection of A-to-I editing, improving high-throughput mapping of editing sites and monitoring transcriptome-wide modification signatures.[23] While both technologies demonstrated feasibility and effectiveness in their respective applications compared to current state-of-the-art methods, we employed a commercially available, recombinant EndoV from *E.* coli (New England Biolabs) in these studies,[23] and we recognize that there remains significant potential for performance improvements. The most critical component in the EndoVIPER and EndoVLISA protocol is the enzyme itself, and so future experiments will focus

on the EndoV-inosine molecular interaction and deploy protein engineering and evolution strategies to improve binding affinity and selectivity for inosine-containing RNAs.

EndoV is highly conserved in Nature and present across all domains of life, and we hypothesize that other homologs may have increased affinity and selectivity for inosine in RNA.[24] Work is currently underway to sample broadly across this diversity of naturally occurring enzymes, and initial screening experiments will directly compare EndoV homologs from each of the three domains as well as different fungi, plants, and animals: 2 archaea organisms (*Pyrobaculum islandicum* and *Archaeoglobus fulgidus*), 4 prokaryotic species (*Thermotoga maritima, Thermus thermophilus, Chloroflexus aurantiacus, and Escherichia coli* as a baseline control), as well as 7 model eukaryotes (*Schizosaccharomyces pombe*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Xenopus tropicalis*, *Gallus gallus*, *Mus musculus,* and *Homo sapiens*). Assessing the binding affinity of each EndoV isoform using microscale thermophoresis (MST) and performing functional *in vitro* pulldown experiments using both ssRNA A and ssRNA I will enable us to quantify and demonstrate any performance differences between enzyme variants.

Using the well characterized EndoV crystal structures from *Thermotoga maritima*[25] and *Mus musculus*,[24] we also identified 6 key residues that collectively form an "inosine-binding pocket" (Tyr 80, Leu 85, Gln 112, His 116, Ile 122, and Leu 142) (**Figure 7.2a**). Importantly, these residues

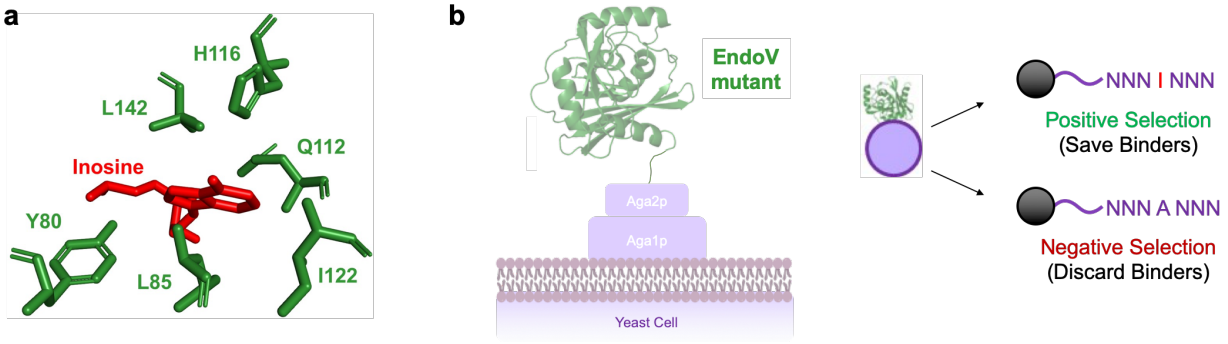


Figure 7.2. EndoV engineering and evolution. a) Crystal structure of *T. maritima* EndoV highlighting amino acids (green) that interact with inosine (red). b) EndoV selection using yeast display.

are not directly involved in binding metal cofactors, and although these sites are conserved among many species, they also can vary considerably across domains,[24] and thus we hypothesize that mutation at these positions could significantly alter binding affinity and selectivity. Current efforts in our lab are devoted to generating a library of EndoV mutants from these residues, and screening their affinity for inosine-containing RNA using yeast display (**Figure 7.2b**).[26] We are also pursuing a directed evolution approach wherein each variant will be expressed on the surface of a yeast cell and undergo both positive and negative selection with immobilized oligonucleotides. Together, these efforts will ideally yield a highly selective and potent EndoV scaffold for future use in inosine-detection platforms.

7.3 Deep RNA sequencing using an Optimized EndoVIPER Workflow

Despite tens of millions of potential editing sites, these events are quite rare in the context of total cellular RNA (~0.01–<0.0001% of all nucleotides),[2-3] and it has been technically challenging to comprehensively map these events using RNA-seq. Utilizing techniques that improve the efficiency and sensitivity of A-to-I site identification, including EndoVIPER[23] or microfluidics-based multiplex PCR sequencing,[27] will likely improve our ability to identify new candidate sites. Using optimized EndoV variants, we plan to perform a series of large scale, high-depth EndoVIPER-seq (100M+ reads/sample) experiments to explore and map of RNA editing sites in human mRNA. Extensive efforts will then be devoted to identifying patterns between different biological samples, as well as determining which sites are physiologically relevant or implicated in disease. Surprisingly, there have also been very few studies focusing specifically on A-to-I editing in small non-coding RNAs, including microRNAs and small interfering RNAs, which have been demonstrated to both undergo editing as well as influence overall cellular behavior.[6, 28] We similarly predict that our high-resolution protein-based detection platform will also identify crucial editing sites in these transcript classes.

7.4 Elucidating the Natural Function of EndoV in Humans

Although we successfully utilized EndoV for developing inosine-detection platforms (Chapters 4 and 5), the biological functions of this enzyme in humans remain mysterious. EndoV is present across all types of organisms, and while protein sequences are highly conserved across different species, evolution has gradually shifted the functional roles of these enzymes in their respective niches, and their biochemical preferences for different substrates have inexorably changed.[24, 29-31] In bacteria, EndoV cleaves downstream of inosine lesions resulting from oxidative damage in DNA to promote base excision repair.[32] Despite this preference for DNA, previous characterizations also showed that *Escherichia coli* EndoV (eEndoV) is both specific and active toward inosine in single-stranded RNA (ssRNA).[30-31] In humans and other higher eukaryotes, this substrate preference has interestingly shifted toward A-to-I edited RNA. Several groups independently demonstrated that human EndoV is an inosine-specific ribonuclease and cleaves A-to-I edited RNAs *in vitro*,[30-31] but recapitulating this activity *in vivo* has yielded confounding results in that enzyme levels do not correlate with overall inosine abundance in RNA.[33] It is likely that EndoV activity is regulated or suppressed in some way, and RNA structure may play a key role in determining which transcripts are degraded by EndoV. Our explorations of EndoV also relied heavily on controlling ionic conditions, with Ca^{2+} and Mg^{2+} dictating enzyme binding or cleavage, respectively. Although it is easy to modulate this activity *in vitro* by changing cation concentrations in different buffers, it is also interesting to speculate whether ion switching is occurring in the cytosol to control EndoV activity. To our knowledge, it is unknown what proportion of these different cations are complexed with EndoV *in vivo*, and characterizing these effects may yield insight into how enzyme activity is regulated. Overall, EndoV has emerged as a powerful molecular recognition platform for characterizing A-to-I editing, and determining its natural biological functions in humans is an area of high interest in the RNA editing field. Elucidating these roles are likely to yield critical information for understanding epitranscriptomics and its relationship to RNA metabolism.

7.5 Implementation of the EndoVLISA Bioassay

Profiling transcriptome-level A-to-I signatures is key to understanding how global editing activity is regulated in different cell types. Similarly, these measurements are also emerging as a key biomarker for several diseases,[13, 18, 34-35] but there are few methods available that can cheaply and rapidly detect these changes. Protein-based platforms that do not rely on sequencing are well poised for addressing both of these major pursuits, enabling facile characterization of global editing changes as well as rapid diagnostic detection of disease-relevant epitranscriptomic signatures.

Recent large-scale screens have identified a number of RNA-binding proteins that may act as *cis*- and *trans*- regulators to influence A-to-I editing levels, and our EndoVLISA platform would enable further validation of these potential hits. For example, skeletal muscle exhibits especially low editing activity in humans, and this is explained in part by high expression of the protein AIMP2, appears to facilitate rapid degradation of ADAR enzymes and negatively regulate overall A-to-I editing levels.[2] Additionally, TDP-43, DROSHA, NF45/90 and Ro60 have also been implicated in editing regulation, and interestingly influence these levels through a combination of modulating ADAR1 expression, directly interacting with the enzyme, or binding Alu elements to out-compete ADAR recognition.[36] The EndoVLISA platform is well-suited to answer key questions surrounding of this regulation of ADAR activity, and is highly compatible with overexpression and knockout of these candidate proteins in model cell lines. CRISPR knockout screens are also of interest with this assay and would facilitate discovery of new proteins that perturb global A-to-I signatures in different cell and tissue types.

ADAR1 overexpression and upregulation of editing activity is also a hallmark of cancer[34-35] and autoimmune disease,[37-39] and we found that EndoVLISA could reliably detect these changes in several cancer types from different human tissues (Chapter 5). Human brain development and overall nervous system function are also reliant on A-to-I editing, and malfunctions in this process have been implicated in epilepsy, amyotrophic lateral sclerosis,

glioblastoma, schizophrenia, autism, and Alzheimer's disease. [10-17] Deploying EndoVLISA to these and other disease classes would be an interesting pursuit in applying our method for clinical applications. Similarly, long-term plans include potential commercialization of this assay as well as modifying its physical properties toward colorimetric or paper-based systems which would enable instrument-free, point-of-care analysis.

7.6 Visualizing Global A-to-I Editing Patterns with EndoV Immunostaining

Immunostaining has provided significant insight into the mechanisms and functions of N^6-methyladenosine (m^6A),[40] but no comparable method exists for A-to-I editing. Building off of our results in developing an immunoprecipitation (EndoVIPER) and immunosorbency assay (EndoVLISA), we are very interested in leveraging EndoV binding in an immunostaining workflow to enable spatial visualization of A-to-I RNA editing in human cells. Specifically, EndoV fused to an affinity tag (e.g. maltose-binding protein, MBP) will be used to bind inosine-containing RNAs, followed by staining with primary and secondary antibodies (**Figure 7.3**). Work is currently underway to validate a full immunofluorescence workflow in fixed HEK293T cells having ADAR1 up- and downregulated, ideally demonstrating that signal generation is proportional to cellular editing activity. Using secondary antibodies conjugated to different dyes (e.g. Alexa Fluor 647), we will also explore the utility of our method with super-resolution imaging platforms, which will

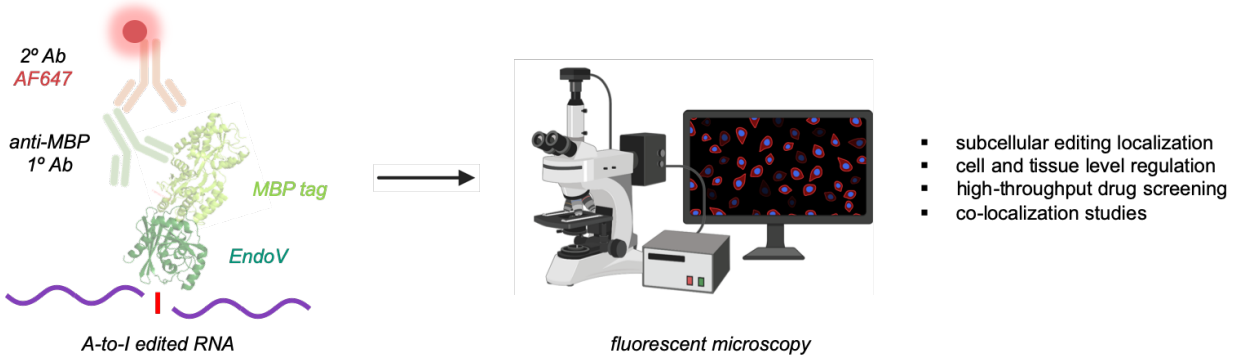


Figure 7.3. EndoV immunostaining for visualizing A-to-I RNA editing. In a similar manner to the EndoVLISA platform, edited transcripts can be immunstained with EndoV, anti-MBP primary antibody, and a fluorophore-conjugated secondary antibody to enable high-resolution fluorescence microscopy.

excitingly enable us to directly visualize subcellular editing dynamics with high precision. A succcessful immunostaining workflow will also facilitate histological staining of tissue sections using our method along with cell-type specific protein markers, providing a detailed portrait of differential editing activity between different cells within the same tissue. Additionally, while EndoVLISA is tractable for high-throughput drug screening, this approach may be more feasible for testing very large compound libraries using automated image acquisition and analysis pipelines.

7.7 Parallel Study of A-to-I Editing and Subcellular RNA Localization

RNA sequence, function, and translational efficiency are all actively modulated in the cell though the localization and editing of transcripts.[1, 41] These processes are essential for cellular function and are dysregulated in many diseases, including neurodegenerative disorders and cancers. For example, subcellular mRNA transport facilitates spatial control of protein synthesis as well as regulation of cellular stress responses.[42-45] RNA editing primarily influences cellular immune activation,[7, 46] but also has an important role in recoding protein sequences and modulating overall gene expression levels.[1-2, 8, 28] On their own, each of these processes are critical for a number of cellular pathways, and recent studies have offered groundbreaking evidence that RNA trafficking and RNA editing serve as synergistic mechanisms for mutually regulating and influencing cellular function.[8, 47-50] Despite their importance, significant gaps remain in our understanding of the timing, mechanisms, regulation, and consequences of these events. Methods have been reported that enable researchers to label and image specific RNAs in living cells[51] or sequence edited transcripts from cell lysates.[52] However, significant technological limitations still exist, and addressing these gaps holds promise for the development of new therapeutics and diagnostics.

The Heemstra lab has developed a powerful method in which the aptamer sequence for the fluorogenic dye malachite green (MG)[53] is genetically fused to an RNA of interest, and addition

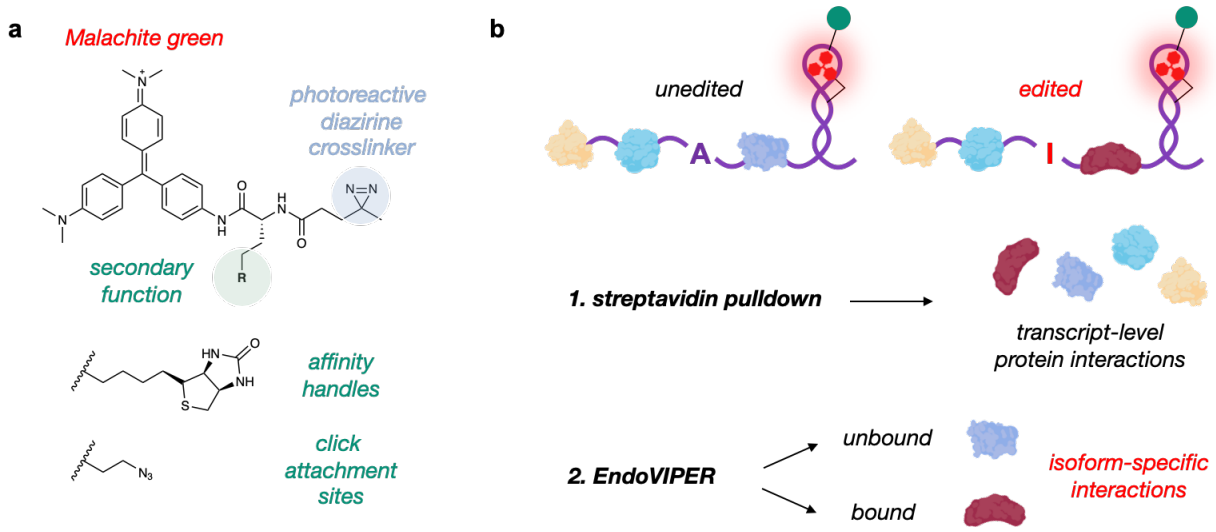


Figure 7.4. Combining aptamer-based RNA tagging with EndoV pulldown enables identification of isoform specific proteins. Malachite green (MG) diazirine functionalized with affinity handles (including biotin or click-compatible groups) binds and covalently lables aptamer-tagged mRNA. Sequential pulldown with streptavidin and EndoVIPER enables identification of isoform-specific RNA-protein interactions.

of MG-diazirine (MGD) followed by UV irradiation covalently attaches the dye to the RNA.[54] Based on this technology as well as our EndoV platforms, our lab is uniquely poised to address key biological questions surrounding the relationship between RNA editing and RNA trafficking dynamics. In particular, combination of our potential EndoV immunostaining workflow and our aptamer-based imaging platform will enable us to explore the relationship between RNA editing status and subcellular location.

Protein interactions facilitate biomolecule trafficking, and identifying transcript-specific RNA-binding proteins is critical for understanding RNA localization mechanisms and their relationship with epitranscriptomic modifications. A key advantage of our covalent labeling approach is the ability to design probes containing affinity handles for isolating labeled transcripts with high efficiency. We hypothesize that when combined with our existing EndoVIPER method, this will enable identification of protein-RNA interactions that are induced or repressed upon A-to-I editing. (**Figure 7.4**). Importantly, this sequential labeling and pulldown approach will enable identification of isoform-specific RNA binding proteins as well as common interactors regardless

of editing status. Initial efforts will validate the power of this technology, providing a foundation for experiments aimed at discovering new protein binding partners responsible for editing-dependent localization of cellular transcripts.

7.8 Directed Evolution of Deaminase Ribozymes for Site-Directed RNA Editing

Given the broad potential for recoding disease-causing protein mutations, site-directed A-to-I RNA editing is an attractive therapeutic platform.[55] This has been previously achieved through fusion of ADAR deaminase domains to various RNA targeting systems, including BoxB-λN hairpins,[56] MS2 phage proteins,[57] human effector proteins,[58] and Cas9.[59] Alternatively, endogenous ADAR enzymes can be recruited to targets by introducing chemically modified guide RNAs into cells that mimic natural ADAR recognition motifs.[60-63] While significant progress has been made in designing and implementing these constructs, both of these approaches are reliant on incomplete knowledge concerning the basic mechanisms of how ADARs interact with and edit different RNA substrates. As a result, these platforms suffer from both low editing efficiency at desired locations as well as promiscuous and dangerous off-target editing activity. Additionally, these systems have significant limitations in delivering multiple large protein components or plasmids to cells.

Based on our EndoV platform, we aim to develop a completely novel editing platform, and propose an innovative approach to *in vitro* selection in which EndoV binding is used to enrich ribozyme sequences capable of catalyzing adenosine deamination. The central premise for the *in vitro* selection of functional nucleic acids is that given a large (generally ~10^{14}) pool of diverse sequences, the likelihood is high that some sequences show the desired activity, and these can be enriched through iterative rounds of selection and amplification. To generate adenosine deaminase ribozymes using this process, we have designed the selection scheme outlined in (**Figure 7.5**). We will begin with a DNA library having an N_{40} random region flanked by primer

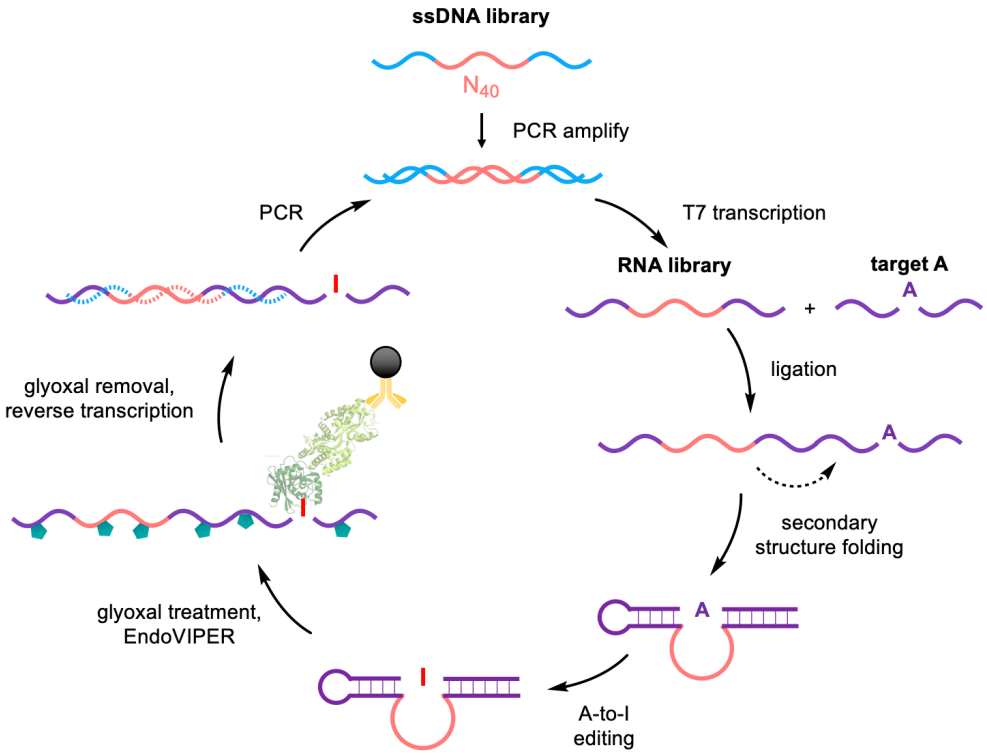


Figure 7.5. Selection of deaminase ribozymes using EndoVIPER.

binding sites, and this library will be amplified to generate double-stranded DNA (dsDNA) and install a T7 promoter sequence. The dsDNA library will then be transcribed using T7 RNA polymerase to generate a corresponding RNA library, and the "target A" substrate strand will be attached using T4 RNA ligase. The ligated library will be incubated at 37 °C in buffer mimicking intracellular salt concentrations to allow for the desired deamination reaction to take place. To isolate the functional sequences, we intend to employ our EndoVIPER workflow as described in the Chapter 4. Finally, we will perform RT-PCR using primers that are complementary to the regions flanking the N_{40} library, as this will amplify the library portion and enable the attachment of a new adenosine-containing target substrate strand.

The resulting RNA catalysts will not rely on synthetic or host ADAR machinery, and this is anticipated to provide a safer and more specific method for site-directed RNA editing. Additionally,

ribozymes combine both sequence targeting and catalytic activity into a single biomolecule, which may lead to greater A-to-I editing efficiency than existing platforms that require RNA-protein complexes. Finally, ribozymes (20-100 nt) are a fraction of the size of typical effector protein fusion constructs (>4500 nt), enabling versatile multiplexed targeting and ease of delivery in viral vectors.[64-65] The proposed research represents the first fully orthogonal version of A-to-I editing machinery, and this is anticipated to offer a number of technological advantages over existing methods.

7.9 Single-cell Profiling of A-to-I Editing

While the vast majority of sites reside in untranslated RNA regions and serve to regulate cellular immune responses,[5, 46, 66] A-to-I editing also occurs in coding regions and directly alters amino acid sequences to produce different protein isoforms.[6] These protein recoding events, while rare, generate complexity and diversity in the cellular proteome. In particular, several neurotransmitter receptors[67-70] and ion channels[71] require mRNA recoding for proper neuronal activity, and dysregulation of this process is directly responsible for brain development disorders,[13-14] glioblastoma,[72] and neurodegeneration.[16, 73] Overall editing activity also varies between different tissue and cell types, underscoring the vital role of A-to-I editing in creating intercellular diversity and specialized tissue function.[2, 74] Interestingly, these recoding patterns also change significantly during tissue development and are often required for stem cell differentiation.[2, 75-79] Despite this importance, the majority of studies characterizing A-to-I patterns rely on bulk measurement methods, including Sanger or next-generation RNA sequencing, which can only measure average RNA editing frequencies across large groups of pooled cells (Fig. 1b). To achieve better resolution, different cell types can be sorted prior to RNA-seq, and these efforts have confirmed that A-to-I recoding varies substantially between populations.[80-82] However, these approaches still rely on pooled populations and are unable to detect changes between individual

cells. Another recent report analyzed a single-cell RNA-seq (scRNA-seq) dataset generated using the Fluidigm[83] platform and found that A-to-I activity can differ substantially between individual cells of the same subtype,[84] further highlighting its importance for generating functional diversity within tissues. However, scRNA-seq platforms that rely on well-based cell sorting, including Fluidigm,[83] ICELL8,[85] Seq-Well,[86] and sci-RNA-seq[87], require high cost dispensing of individual cells into plates which limits practical throughput to hundreds of cells per run. Presently, there are no facile high-throughput approaches for profiling A-to-I RNA editing in individual cells. As a result, our understanding of how A-to-I recoding actively contributes to tissue development and function is limited, and the mechanisms by which these events regulate and influence the expression of other genes is not well known. The ability to probe RNA editing frequencies at large scales and single-cell resolution would dramatically improve our understanding of these key processes and enable advances in diagnostic and therapeutic interventions.

Microfluidic technologies, including Drop-seq,[88] Chromium (10X Genomics),[89] and inDrop (1CellBio),[90] can routinely interrogate >10^4 cells, positioning these as the most powerful current systems for scRNA-seq.[91-92] In these assays, a single cell and a barcoded bead are co-encapsulated in a ~1 nL droplet using a microfluidic chip. The cell is then lysed, enabling RNA transcripts to bind to a bead decorated with oligonucleotide strands, each containing a unique barcode sequence and a polyT primer that hybridizes to the 3' polyA tails of mRNA transcripts. cDNA from single cells is then generated from each RNA transcript, PCR amplified, and subjected to high-throughput sequencing. However, these methods require enzymatic tagmentation[93] to prepare libraries for Illumina sequencing, and hence only retain a small "window" (~100 nt) of the 5' or 3' regions of the RNA molecule. While this is a suitable approach for identifying transcripts and quantifying gene expression in single cells, this window omits the majority of the mRNA sequence, including all putative A-to-I recoding sites (**Figure 7.6a**). An effective solution to this problem is to position RNA editing events within the sequencing window using gene-specific

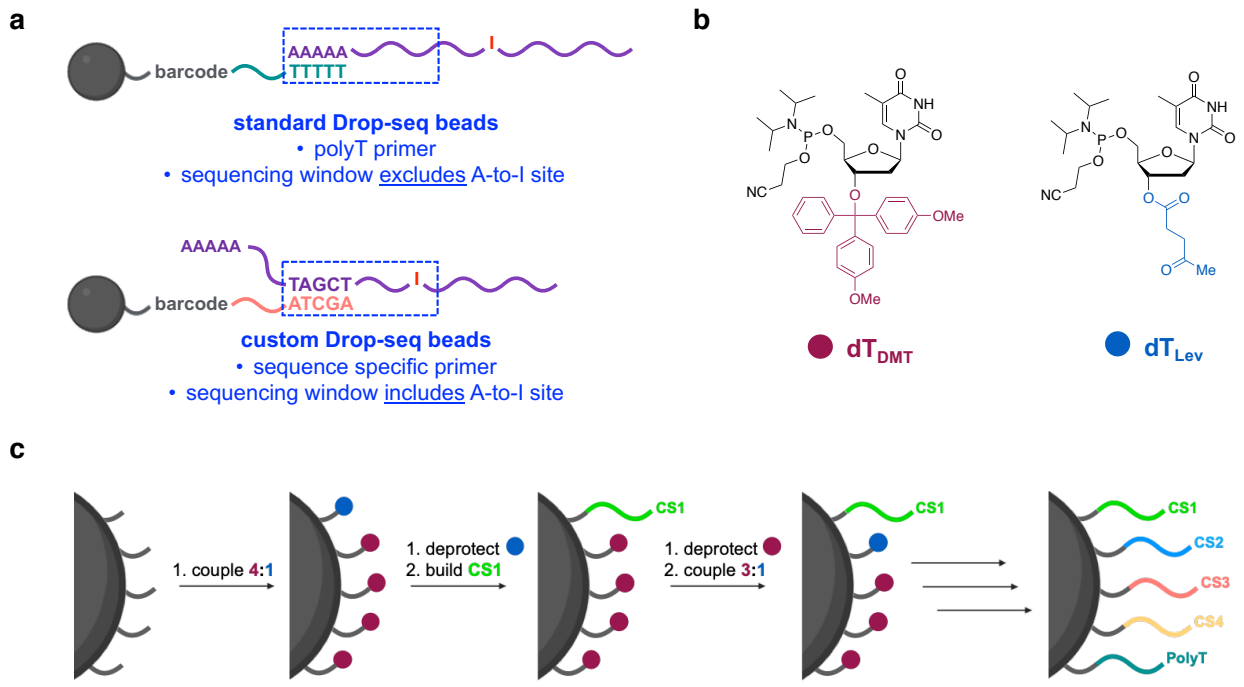


Figure 7.6. Engineering Drop-seq beads to measure A-to-I recoding events in single cells. a) Standard Drop-seq beads can only capture short sequencing windows from RNA termini, while custom beads with sequence-specific probes detect A-to-I editing status within short read windows. b) Orthogonal DMT and Lev 3' protecting groups on reverse phosphoramidite monomers (deoxythymine, dT) enable c) stepwise solid-phase synthesis of custom multiplexed beads with gene-specific capture strands (CS1, 2, etc).

capture strand (CS) primers (**Figure 7.6a**). Towards profiling individual A-to-I recoding sites in single cells, we envisioned the synthesis of Drop-seq beads containing multiple discrete capture strands for site-specific quantification of RNA editing rates. Additionally, polyT strands on the bead would simultaneously enable overall transcriptome characterization and identify potential correlations between editing frequencies at specific sites and overall gene expression patterns. Although seemingly simple, split-and-pool synthesis is required for creating bead barcode sequences and unique molecular identifiers (UMIs) for each strand, and thus only one sequence (usually either polyT or a single custom sequence) can be made per bead.[88] We recently developed a straightforward method for synthesizing multiplexed beads and demonstrated their use for specific capture and sequencing of T-cell receptor mRNA isoforms.[94] In this approach,

standard split-pool synthesis is first used to generate barcoded beads, which are then reacted with defined mixtures of 5' phosphoramidite monomers containing orthogonal 3' protecting groups. In particular, the first monomer contains the standard dimethoxytrityl (DMT) cap, which is removed under mildly acidic conditions. The second monomer contains a 2'-levulinyl (Lev) group, which is stable under DMT acidic deprotection and can be orthogonally removed using aqueous hydrazine (**Figure 7.6b**).[94-95] Using defined mixtures of these monomers allows selective deprotection of a portion of the barcoded strands and enables synthesis of a specific capture sequence. Remaining DMT groups are removed, and the beads are again blocked with a predefined mixture of DMT/Lev monomers, followed by selective deprotection and synthesis of the next capture strand. This entire process is then repeated iteratively for each desired sequence (**Figure 7.6c**).

As an biological target for single-cell profiling of RNA editing, myeloid leukemia is an interesting context, which is a highly aggressive blood cancer that displays marked RNA editing malfunctions.[35, 96] Cancer is also an inherently heterogeneous disease, and probing the epitranscriptomic activity of individual cells will shed light how these changes might contribute to overall disease characteristics. This custom Drop-seq approach can be used to track individual cell trajectories throughout myeloid differentiation and identify potential relationships between A-to-I recoding events and changes in gene expression patterns. This paradigm presents a potential new diagnostic tool for understanding cellular disease, and could provide the first high-throughput profiling of RNA editing at the single-cell level to generate a high-resolution transcriptomic portrait of myeloid cell differentiation in a leukemia model. This method is also a generalizable framework for constructing multiplexed, user-defined Drop-seq beads for monitoring mRNA isoforms and single-nucleotide changes in individual cells. This tool will likely find us for interrogating a multitude of RNA editing events in nearly any biological context, offering a powerful new method for linking epitranscriptomic changes with intercellular diversity.

7.10 Improved Control over Glyoxal Caging and Decaging

In Chapter 6, we described the use of the covalent denaturant glyoxal as a means of imparting thermoresponsive activity in variety of nucleic acid constructs. While this was effective in reversibly controlling the activity of RNA, DNA, 2' modified scaffolds, and XNAs, we were unable to control the precise location of where glyoxal adducts were added. This will be crucial for effectively controlling the structure and function of certain nucleobase constructs, and will require creative solutions to implement this idea. As an example, we did explore glyoxal treatment of mRNA and intended to apply this caging principle to reversibly inhibit expression of a GFP reporter gene (**Figure 7.7a**) We interestingly found that caging was highly potent in inhibiting expression, with < 30 seconds treatment time needed to completely ablate GFP production (**Figure 7.7b**).

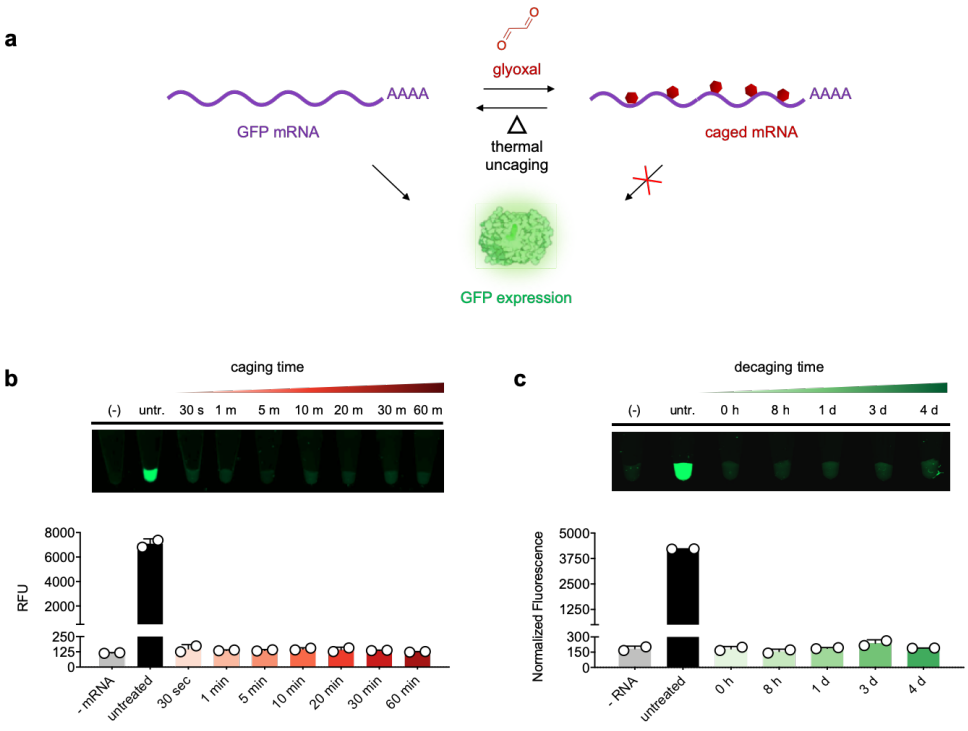


Figure 7.7. Glyoxal caging of mRNA for controlling gene expression. a) mRNA encoding GFP can be reversibly caged with glyoxal treatment. b) Increasing glyoxal treatment times inhibit GFP synthesis in a cell free protein expression system. c) Minimally caged mRNA (30 seconds treatment time) was incubated in PBS at pH 7.5 for the indicated time periods and then GFP expression was attempted *in vitro*.

While inhibiting activity is important, the key step in our overall approach is restoring activity upon thermal decaging. However, we found that we could not effectively regain gene expression function, even after incubating minimally caged mRNA at pH 7.5 and 37 ºC for several days. We hypothesize that mRNA may be especially difficult for this prospect because 1) even a single adduct in the coding region would likely inhibit translation of a full protein, and 2) inherent mRNA instability in decaging conditions may lead to significant drop in overall transcript levels.

To address this, a potential solution is to both reduce the overall amount of glyoxal adducts introduced as well exert more control over where they are added. Similar to other lithography techniques for nanotechnology applications, we envision that a majority of the mRNA molecule can be "masked" with complementary DNA, resulting in glyoxal caging of a small portion of the transcript (**Figure 7.8**).

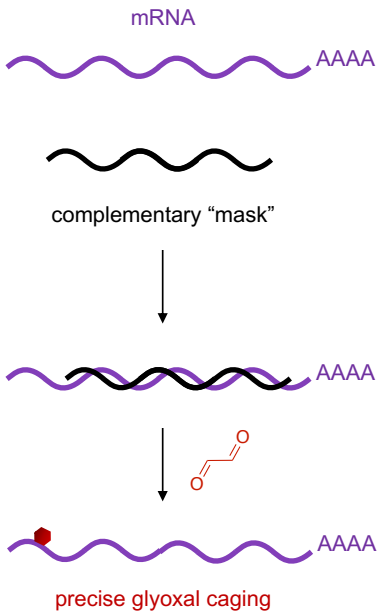


Figure 7.8. Regio-specific glyoxal caging using chemical lithography. Target mRNA can first be "masked" using a complementary sequence, exposing only a small portion of the sequence for glyoxal treatment and enabling faster decaging response.

When considering the mechanism of glyoxal introduction, it also theoretically possible for an enzyme to catalytically install glyoxal. Similarly, a ribozyme or DNAzyme could be evolved to facilitate adduct formation, and when combined with site- or sequence-specific targeting modalities, this could enable precise caging of individual nucleobases in a genetically encodable manner. Lastly, a major limitation of our method is the reliance on slow decaging from alkaline buffer conditions (Chapter 6). While this predictable and tunable decaging is an advantage in many contexts, the ability to rapidly remove glyoxal adducts and restore nucleic acid activity immediately would be a transformative advance for this technology. Work is currently underway to derivatize glyoxal to enable this, and we are currently exploring both small-molecule and light-triggered release of caging groups.

7.11 Glyoxal Caging of Anti-Viral Drugs

In our earlier exploration of glyoxal (Chapter 6), we found that the reagent was able to robustly react with adenosine (A), cytidine (C) and guanosine (G) nucleobases.[97] Interestingly, many anti-viral and chemotherapeutic drugs are based on these nucleobase scaffolds (**Figure 7.9**),[98] and we hypothesize that glyoxal addition onto these small-molecules will enable time-release activation and improved therapeutic properties. Additionally, glyoxal can be chemically derivatized to include click-compatible handles or attachment sites for secondary payloads, which could expand the functionality of these small molecules. In particular, we have shown that azidophenylglyoxal reacts similarly to glyoxal and could enable cell-specific targeting of these drugs to alleviate cytotoxic off-target effects. Kethoxal-based scaffolds can be modified in a similar manner,[99] and current work in our lab aims to explore these reagents to modulate nucleoside analogue activity.

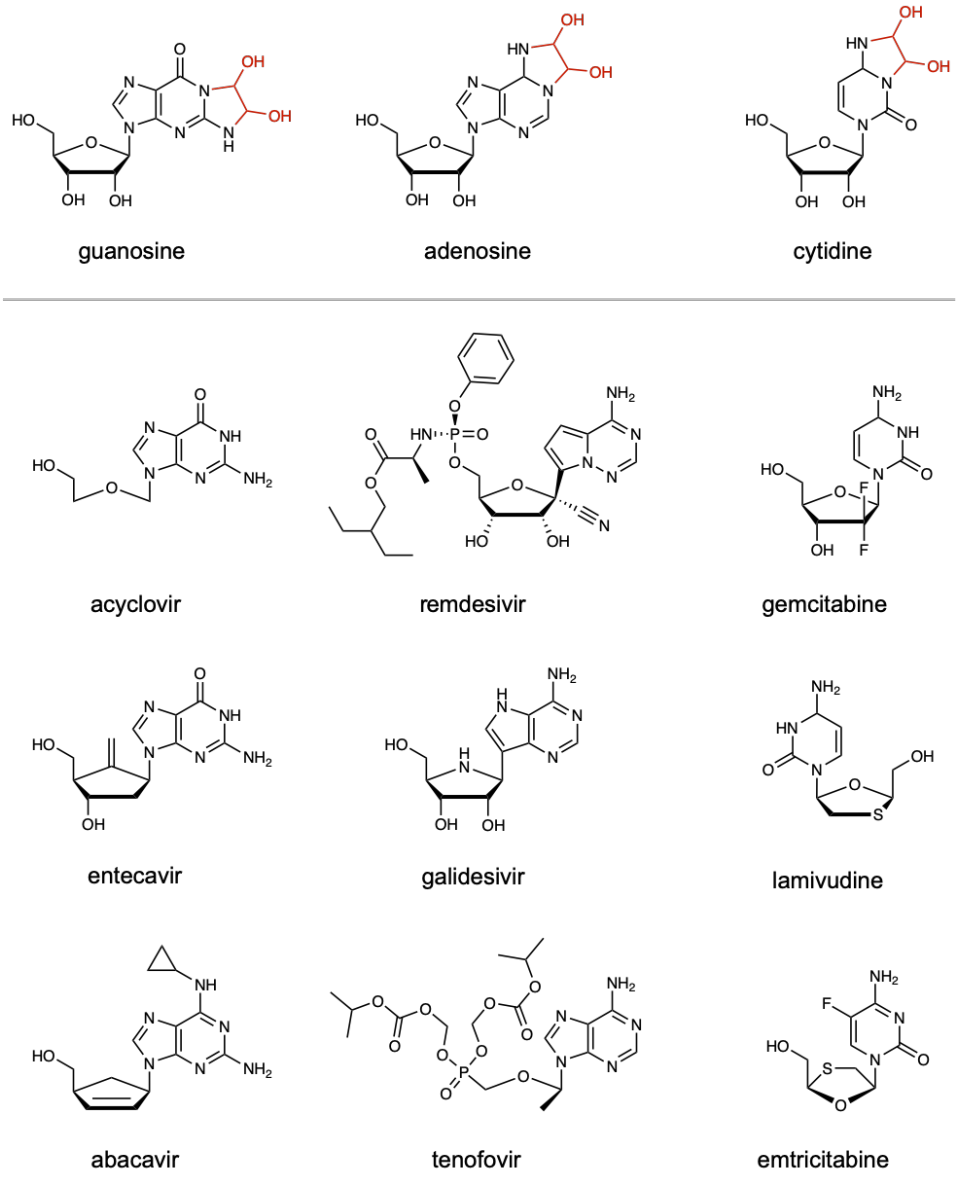


Figure 7.9. Glyoxal caging of nucleoside analogue drugs. Canonical nucleosides guanosine, adenosine, and cytidine shown with glyoxal addition products (red). Various candidate nucleoside analogue drugs are shown in the bottom panel.

7.12 References

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

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

3. Mansi, L.; Tangaro, M. A.; Lo Giudice, C.; Flati, T.; Kopel, E.; Schaffer, A. A.; Castrignanò, T.; Chillemi, G.; Pesole, G.; Picardi, E., REDIportal: millions of novel A-to-I RNA editing events from thousands of RNAseq experiments. *Nucleic Acids Research* **2020,** *in press*.

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

5. Mannion, N. M.; Greenwood, S. M.; Young, R.; Cox, S.; Brindle, J.; Read, D.; Nellåker, C.; Vesely, C.; Ponting, C. P.; McLaughlin, P. J., The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell reports* **2014,** *9* (4), 1482-1494.

6. Nishikura, K., A-to-I editing of coding and non-coding RNAs by ADARs. *Nature reviews Molecular cell biology* **2016,** *17* (2), 83.

7. Ishizuka, J. J.; Manguso, R. T.; Cheruiyot, C. K.; Bi, K.; Panda, A.; Iracheta-Vellve, A.; Miller, B. C.; Du, P. P.; Yates, K. B.; Dubrot, J., Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. *Nature* **2018,** *565* (7737), 43-48.

8. Chung, H.; Calis, J. J.; Wu, X.; Sun, T.; Yu, Y.; Sarbanes, S. L.; Thi, V. L. D.; Shilvock, A. R.; Hoffmann, H.-H.; Rosenberg, B. R., Human ADAR1 prevents endogenous RNA from triggering translational shutdown. *Cell* **2018,** *172* (4), 811-824. e14.

9. Picardi, E.; D'Erchia, A. M.; Lo Giudice, C.; Pesole, G., REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic acids research* **2016,** *45* (D1), D750-D757.

10. Mehler, M. F.; Mattick, J. S., Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiological reviews* **2007,** *87* (3), 799-823.

11. Wright, A. L.; Vissel, B., The essential role of AMPA receptor GluR2 subunit RNA editing in the normal and diseased brain. *Frontiers in molecular neuroscience* **2012,** *5*, 34.

12. Li, J. B.; Church, G. M., Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. *Nature neuroscience* **2013,** *16* (11), 1518.

13. Hwang, T.; Park, C.-K.; Leung, A. K.; Gao, Y.; Hyde, T. M.; Kleinman, J. E.; Rajpurohit, A.; Tao, R.; Shin, J. H.; Weinberger, D. R., Dynamic regulation of RNA editing in human brain development and disease. *Nature neuroscience* **2016,** *19* (8), 1093.

14. Tran, S. S.; Jun, H.-I.; Bahn, J. H.; Azghadi, A.; Ramaswami, G.; Van Nostrand, E. L.; Nguyen, T. B.; Hsiao, Y.-H. E.; Lee, C.; Pratt, G. A., Widespread RNA editing dysregulation in brains from autistic individuals. *Nature neuroscience* **2019,** *22* (1), 25.

15. Breen, M. S.; Dobbyn, A.; Li, Q.; Roussos, P.; Hoffman, G. E.; Stahl, E.; Chess, A.; Sklar, P.; Li, J. B.; Devlin, B., Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia. *Nature neuroscience* **2019,** *22* (9), 1402-1412.

16. Khermesh, K.; D'Erchia, A. M.; Barak, M.; Annese, A.; Wachtel, C.; Levanon, E. Y.; Picardi, E.; Eisenberg, E., Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *Rna* **2016,** *22* (2), 290-302.

17. Silvestris, D. A.; Picardi, E.; Cesarini, V.; Fosso, B.; Mangraviti, N.; Massimi, L.; Martini, M.; Pesole, G.; Locatelli, F.; Gallo, A., Dynamic inosinome profiles reveal novel patient stratification and gender-specific differences in glioblastoma. *Genome biology* **2019,** *20* (1), 33.

18. Slotkin, W.; Nishikura, K., Adenosine-to-inosine RNA editing and human disease. *Genome medicine* **2013,** *5* (11), 105.

19. Knutson, S. D.; Ayele, T. M.; Heemstra, J. M., Chemical Labeling and Affinity Capture of Inosine-Containing RNAs Using Acrylamidofluorescein. *Bioconjugate chemistry* **2018,** *29* (9), 2899-2903.

20. Knutson, S. D.; Korn, M. M.; Johnson, R. P.; Monteleone, L. R.; Dailey, D. M.; Swenson, C. S.; Beal, P. A.; Heemstra, J. M., Chemical Profiling of A-to-I RNA Editing Using a Click-Compatible Phenylacrylamide. *Chemistry–A European Journal* **2020,** *26* (44), 9874-9878.

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

22. Li, Y.; Göhl, M.; Ke, K.; Vanderwal, C. D.; Spitale, R. C., Identification of Adenosine-to-Inosine RNA Editing with Acrylonitrile Reagents. *Organic letters* **2019,** *21* (19), 7948-7951.

23. Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M., Selective enrichment of A-to-I edited transcripts from cellular RNA using Endonuclease V. *Journal of the American Chemical Society* **2020,** *142* (11), 5241-5251.

24. Wu, J.; Samara, N. L.; Kuraoka, I.; Yang, W., Evolution of Inosine-Specific Endonuclease V from Bacterial DNase to Eukaryotic RNase. *Molecular cell* **2019,** *76* (1), 44-56. e3.

25. Dalhus, B.; Arvai, A. S.; Rosnes, I.; Olsen, Ø. E.; Backe, P. H.; Alseth, I.; Gao, H.; Cao, W.; Tainer, J. A.; Bjørås, M., Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nature structural & molecular biology* **2009,** *16* (2), 138.

26. Cherf, G. M.; Cochran, J. R., Applications of yeast surface display for protein engineering. In *yeast surface display*, Springer: 2015; pp 155-175.

27. Zhang, R.; Li, X.; Ramaswami, G.; Smith, K. S.; Turecki, G.; Montgomery, S. B.; Li, J. B., Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nature methods* **2014,** *11* (1), 51.

28. Kawahara, Y.; Zinshteyn, B.; Sethupathy, P.; Iizasa, H.; Hatzigeorgiou, A. G.; Nishikura, K., Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* **2007,** *315* (5815), 1137-1140.

29. Fladeby, C.; Vik, E. S.; Laerdahl, J. K.; Neurauter, C. G.; Heggelund, J. E.; Thorgaard, E.; Strøm-Andersen, P.; Bjørås, M.; Dalhus, B.; Alseth, I., The human homolog of Escherichia coli

endonuclease V is a nucleolar protein with affinity for branched DNA structures. *PLoS One* **2012,** *7* (11), e47466.

30. Morita, Y.; Shibutani, T.; Nakanishi, N.; Nishikura, K.; Iwai, S.; Kuraoka, I., Human endonuclease V is a ribonuclease specific for inosine-containing RNA. *Nature communications* **2013,** *4*, 2273.

31. Vik, E. S.; Nawaz, M. S.; Andersen, P. S.; Fladeby, C.; Bjørås, M.; Dalhus, B.; Alseth, I., Endonuclease V cleaves at inosines in RNA. *Nature communications* **2013,** *4*, 2271.

32. Yao, M.; Hatahet, Z.; Melamede, R. J.; Kow, Y. W., Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3'endonuclease, from Escherichia coli. *Journal of Biological Chemistry* **1994,** *269* (23), 16260-16268.

33. Kong, X. Y.; Vik, E. S.; Nawaz, M. S.; Berges, N.; Dahl, T. B.; Vågbø, C.; Suganthan, R.; Segers, F.; Holm, S.; Quiles-Jiménez, A., Deletion of Endonuclease V suppresses chemically induced hepatocellular carcinoma. *Nucleic acids research* **2020,** *48* (8), 4463-4479.

34. Maas, S.; Kawahara, Y.; Tamburro, K. M.; Nishikura, K., A-to-I RNA editing and human disease. *RNA biology* **2006,** *3* (1), 1-9.

35. Han, L.; Diao, L.; Yu, S.; Xu, X.; Li, J.; Zhang, R.; Yang, Y.; Werner, H. M. J.; Eterovic, A. K.; Yuan, Y.; Li, J.; Nair, N.; Minelli, R.; Tsang, Y. H.; Cheung, L. W. T.; Jeong, K. J.; Roszik, J.; Ju, Z.; Woodman, S. E.; Lu, Y.; Scott, K. L.; Li, J. B.; Mills, G. B.; Liang, H., The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer Cell* **2015,** *28* (4), 515-528.

36. Quinones-Valdez, G.; Tran, S. S.; Jun, H.-I.; Bahn, J. H.; Yang, E.-W.; Zhan, L.; Brümmer, A.; Wei, X.; Van Nostrand, E. L.; Pratt, G. A., Regulation of RNA editing by RNA-binding proteins in human cells. *Communications biology* **2019,** *2* (1), 1-14.

37. Shallev, L.; Kopel, E.; Feiglin, A.; Leichner, G. S.; Avni, D.; Sidi, Y.; Eisenberg, E.; Barzilai, A.; Levanon, E. Y.; Greenberger, S., Decreased A-to-I RNA editing as a source of keratinocytes' dsRNA in psoriasis. *RNA* **2018,** *24* (6), 828-840.

38. Roth, S. H.; Danan-Gotthold, M.; Ben-Izhak, M.; Rechavi, G.; Cohen, C. J.; Louzoun, Y.; Levanon, E. Y., Increased RNA editing may provide a source for autoantigens in systemic lupus erythematosus. *Cell reports* **2018,** *23* (1), 50-57.

39. Rice, G. I.; Kasher, P. R.; Forte, G. M.; Mannion, N. M.; Greenwood, S. M.; Szynkiewicz, M.; Dickerson, J. E.; Bhaskar, S. S.; Zampini, M.; Briggs, T. A., Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. *Nature genetics* **2012,** *44* (11), 1243-1248.

40. Fu, Y.; Zhuang, X., m 6 A-binding YTHDF proteins promote stress granule formation. *Nature Chemical Biology* **2020,** *16*, 955-963.

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

42. Katz, Z. B.; Wells, A. L.; Park, H. Y.; Wu, B.; Shenoy, S. M.; Singer, R. H., β-Actin mRNA compartmentalization enhances focal adhesion stability and directs cell migration. *Genes & development* **2012,** *26* (17), 1885-1890.

43. Broadus, J.; Fuerstenberg, S.; Doe, C. Q., Staufen-dependent localization of prospero mRNA contributes to neuroblast daughter-cell fate. *Nature* **1998,** *391* (6669), 792-795.

44. Jansova, D.; Tetkova, A.; Koncicka, M.; Kubelka, M.; Susor, A., Localization of RNA and translation in the mammalian oocyte and embryo. *PLoS One* **2018,** *13* (3), e0192544.

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

46. Liddicoat, B. J.; Piskol, R.; Chalk, A. M.; Ramaswami, G.; Higuchi, M.; Hartner, J. C.; Li, J. B.; Seeburg, P. H.; Walkley, C. R., RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **2015,** *349* (6252), 1115-1120.

47. Vallecillo-Viejo, I. C.; Liscovitch-Brauer, N.; Diaz Quiroz, J. F.; Montiel-Gonzalez, M. F.; Nemes, S. E.; Rangan, K. J.; Levinson, S. R.; Eisenberg, E.; Rosenthal, J. J., Spatially regulated editing of genetic information within a neuron. *Nucleic Acids Research* **2020,** *48* (8), 3999-4012.

48. Mellis, I. A.; Gupte, R.; Raj, A.; Rouhanifard, S. H., Visualizing adenosine-to-inosine RNA editing in single mammalian cells. *Nature methods* **2017,** *14* (8), 801.

49. Ng, S. K.; Weissbach, R.; Ronson, G. E.; Scadden, A., Proteins that contain a functional Z-DNA-binding domain localize to cytoplasmic stress granules. *Nucleic acids research* **2013,** *41* (21), 9786-9799.

50. Weissbach, R.; Scadden, A., Tudor-SN and ADAR1 are components of cytoplasmic stress granules. *Rna* **2012,** *18* (3), 462-471.

51. Xia, Y.; Zhang, R.; Wang, Z.; Tian, J.; Chen, X., Recent advances in high-performance fluorescent and bioluminescent RNA imaging probes. *Chemical Society Reviews* **2017,** *46* (10), 2824-2843.

52. Oakes, E.; Vadlamani, P.; Hundley, H. A., Methods for the Detection of Adenosine-to-Inosine Editing Events in Cellular RNA. In *mRNA Processing*, Springer: 2017; pp 103-127.

53. Babendure, J. R.; Adams, S. R.; Tsien, R. Y., Aptamers switch on fluorescence of triphenylmethane dyes. *J. Am. Chem. Soc.* **2003,** *125* (Copyright (C) 2011 American Chemical Society (ACS). All Rights Reserved.), 14716-14717.

54. 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. *bioRxiv* **2020**.

55. Montiel-Gonzalez, M. F.; Quiroz, J. F. D.; Rosenthal, J. J., Current strategies for site-directed RNA editing using ADARs. *Methods* **2019,** *156*, 16-24.

56. Montiel-Gonzalez, M. F.; Vallecillo-Viejo, I.; Yudowski, G. A.; Rosenthal, J. J., Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proceedings of the National Academy of Sciences* **2013,** *110* (45), 18285-18290.

57. Katrekar, D.; Chen, G.; Meluzzi, D.; Ganesh, A.; Worlikar, A.; Shih, Y.-R.; Varghese, S.; Mali, P., In vivo RNA editing of point mutations via RNA-guided adenosine deaminases. *Nature methods* **2019,** *16* (3), 239-242.

58. Rauch, S.; He, E.; Srienc, M.; Zhou, H.; Zhang, Z.; Dickinson, B. C., Programmable RNA-guided RNA effector proteins built from human parts. *Cell* **2019,** *178* (1), 122-134. e12.

59. Cox, D. B.; Gootenberg, J. S.; Abudayyeh, O. O.; Franklin, B.; Kellner, M. J.; Joung, J.; Zhang, F., RNA editing with CRISPR-Cas13. *Science* **2017**, eaaq0180.

60. Vogel, P.; Schneider, M. F.; Wettengel, J.; Stafforst, T., Improving Site-Directed RNA Editing In Vitro and in Cell Culture by Chemical Modification of the GuideRNA. *Angewandte Chemie International Edition* **2014,** *53* (24), 6267-6271.

61. Merkle, T.; Merz, S.; Reautschnig, P.; Blaha, A.; Li, Q.; Vogel, P.; Wettengel, J.; Li, J. B.; Stafforst, T., Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. *Nature biotechnology* **2019,** *37* (2), 133-138.

62. Vogel, P.; Moschref, M.; Li, Q.; Merkle, T.; Selvasaravanan, K. D.; Li, J. B.; Stafforst, T., Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nature methods* **2018,** *15* (7), 535-538.

63. Qu, L.; Yi, Z.; Zhu, S.; Wang, C.; Cao, Z.; Zhou, Z.; Yuan, P.; Yu, Y.; Tian, F.; Liu, Z., Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. *Nature biotechnology* **2019,** *37* (9), 1059-1069.

64. Asokan, A.; Schaffer, D. V.; Samulski, R. J., The AAV vector toolkit: poised at the clinical crossroads. *Molecular Therapy* **2012,** *20* (4), 699-708.

65. Kotterman, M. A.; Schaffer, D. V., Engineering adeno-associated viruses for clinical gene therapy. *Nature Reviews Genetics* **2014,** *15* (7), 445-451.

66. Chalk, A. M.; Taylor, S.; Heraud-Farlow, J. E.; Walkley, C. R., The majority of A-to-I RNA editing is not required for mammalian homeostasis. *Genome biology* **2019,** *20* (1), 1-14.

67. Sommer, B.; Köhler, M.; Sprengel, R.; Seeburg, P. H., RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **1991,** *67* (1), 11-19.

68. Lomeli, H.; Mosbacher, J.; Melcher, T.; Hoger, T.; Kuner, T.; Monyer, H.; Higuchi, M.; Bach, A.; Seeburg, P. H., Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* **1994,** *266* (5191), 1709-1713.

69. Burns, C. M.; Chu, H.; Rueter, S. M.; Hutchinson, L. K.; Canton, H.; Sanders-Bush, E.; Emeson, R. B., Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **1997,** *387* (6630), 303-308.

70. Daniel, C.; Wahlstedt, H.; Ohlson, J.; Björk, P.; Öhman, M., Adenosine-to-inosine RNA editing affects trafficking of the γ-aminobutyric acid type A (GABAA) receptor. *Journal of Biological Chemistry* **2011,** *286* (3), 2031-2040.

71. Bhalla, T.; Rosenthal, J. J.; Holmgren, M.; Reenan, R., Control of human potassium channel inactivation by editing of a small mRNA hairpin. *Nature structural & molecular biology* **2004,** *11* (10), 950-956.

72. Oakes, E.; Anderson, A.; Cohen-Gadol, A.; Hundley, H. A., Adenosine Deaminase That Acts on RNA 3 (ADAR3) Binding to Glutamate Receptor Subunit B Pre-mRNA Inhibits RNA Editing in Glioblastoma. *Journal of Biological Chemistry* **2017,** *292* (10), 4326-4335.

73. Kawahara, Y.; Ito, K.; Sun, H.; Aizawa, H.; Kanazawa, I.; Kwak, S., Glutamate receptors: RNA editing and death of motor neurons. *Nature* **2004,** *427* (6977), 801-802.

74. Roth, S. H.; Levanon, E. Y.; Eisenberg, E., Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nature methods* **2019**, 1-8.

75. Wahlstedt, H.; Daniel, C.; Ensterö, M.; Öhman, M., Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. *Genome research* **2009,** *19* (6), 978-986.

76. Shtrichman, R.; Germanguz, I.; Mandel, R.; Ziskind, A.; Nahor, I.; Safran, M.; Osenberg, S.; Sherf, O.; Rechavi, G.; Itskovitz-Eldor, J., Altered A-to-I RNA editing in human embryogenesis. *PLoS One* **2012,** *7* (7).

77. Goldstein, B.; Agranat-Tamir, L.; Light, D.; Zgayer, O. B.-N.; Fishman, A.; Lamm, A. T., A-to-I RNA editing promotes developmental stage–specific gene and lncRNA expression. *Genome research* **2017,** *27* (3), 462-470.

78. Wang, Q.; Khillan, J.; Gadue, P.; Nishikura, K., Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. *Science* **2000,** *290* (5497), 1765-1768.

79. Liddicoat, B. J.; Hartner, J. C.; Piskol, R.; Ramaswami, G.; Chalk, A. M.; Kingsley, P. D.; Sankaran, V. G.; Wall, M.; Purton, L. E.; Seeburg, P. H., Adenosine-to-inosine RNA editing by ADAR1 is essential for normal murine erythropoiesis. *Experimental hematology* **2016,** *44* (10), 947-963.

80. Gal-Mark, N.; Shallev, L.; Sweetat, S.; Barak, M.; Li, J. B.; Levanon, E. Y.; Eisenberg, E.; Behar, O., Abnormalities in A-to-I RNA editing patterns in CNS injuries correlate with dynamic changes in cell type composition. *Scientific reports* **2017,** *7* (1), 1-12.

81. Harjanto, D.; Papamarkou, T.; Oates, C. J.; Rayon-Estrada, V.; Papavasiliou, F. N.; Papavasiliou, A., RNA editing generates cellular subsets with diverse sequence within populations. *Nature communications* **2016,** *7*, 12145.

82. Sapiro, A. L.; Shmueli, A.; Henry, G. L.; Li, Q.; Shalit, T.; Yaron, O.; Paas, Y.; Li, J. B.; Shohat-Ophir, G., Illuminating spatial A-to-I RNA editing signatures within the Drosophila brain. *Proceedings of the National Academy of Sciences* **2019,** *116* (6), 2318-2327.

83. Islam, S.; Zeisel, A.; Joost, S.; La Manno, G.; Zajac, P.; Kasper, M.; Lönnerberg, P.; Linnarsson, S., Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature methods* **2014,** *11* (2), 163.

84. Picardi, E.; Horner, D. S.; Pesole, G., Single-cell transcriptomics reveals specific RNA editing signatures in the human brain. *Rna* **2017,** *23* (6), 860-865.

85. Goldstein, L. D.; Chen, Y.-J. J.; Dunne, J.; Mir, A.; Hubschle, H.; Guillory, J.; Yuan, W.; Zhang, J.; Stinson, J.; Jaiswal, B., Massively parallel nanowell-based single-cell gene expression profiling. *BMC genomics* **2017,** *18* (1), 519.

86. Gierahn, T. M.; Wadsworth II, M. H.; Hughes, T. K.; Bryson, B. D.; Butler, A.; Satija, R.; Fortune, S.; Love, J. C.; Shalek, A. K., Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nature methods* **2017,** *14* (4), 395-398.

87. Cao, J.; Packer, J. S.; Ramani, V.; Cusanovich, D. A.; Huynh, C.; Daza, R.; Qiu, X.; Lee, C.; Furlan, S. N.; Steemers, F. J., Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **2017,** *357* (6352), 661-667.

88. Macosko, E. Z.; Basu, A.; Satija, R.; Nemesh, J.; Shekhar, K.; Goldman, M.; Tirosh, I.; Bialas, A. R.; Kamitaki, N.; Martersteck, E. M., Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **2015,** *161* (5), 1202-1214.

89. Zheng, G. X.; Terry, J. M.; Belgrader, P.; Ryvkin, P.; Bent, Z. W.; Wilson, R.; Ziraldo, S. B.; Wheeler, T. D.; McDermott, G. P.; Zhu, J., Massively parallel digital transcriptional profiling of single cells. *Nature communications* **2017,** *8* (1), 1-12.

90. Klein, A. M.; Mazutis, L.; Akartuna, I.; Tallapragada, N.; Veres, A.; Li, V.; Peshkin, L.; Weitz, D. A.; Kirschner, M. W., Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **2015,** *161* (5), 1187-1201.

91. Ding, J.; Adiconis, X.; Simmons, S. K.; Kowalczyk, M. S.; Hession, C. C.; Marjanovic, N. D.; Hughes, T. K.; Wadsworth, M. H.; Burks, T.; Nguyen, L. T., Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nature biotechnology* **2020**, 1-10.

92. Ziegenhain, C.; Vieth, B.; Parekh, S.; Reinius, B.; Guillaumet-Adkins, A.; Smets, M.; Leonhardt, H.; Heyn, H.; Hellmann, I.; Enard, W., Comparative analysis of single-cell RNA sequencing methods. *Molecular cell* **2017,** *65* (4), 631-643. e4.

93. Adey, A.; Morrison, H. G.; Xun, X.; Kitzman, J. O.; Turner, E. H.; Stackhouse, B.; MacKenzie, A. P.; Caruccio, N. C.; Zhang, X.; Shendure, J., Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome biology* **2010,** *11* (12), R119.

94. Hanson, W. M.; Chen, Z.; Jackson, L. K.; Attaf, M.; Sewell, A. K.; Heemstra, J. M.; Phillips, J. D., Reversible oligonucleotide chain blocking enables bead capture and amplification of T-cell receptor α and β chain mRNAs. *Journal of the American Chemical Society* **2016,** *138* (35), 11073-11076.

95. Katolik, A.; Johnsson, R.; Montemayor, E.; Lackey, J. G.; Hart, P. J.; Damha, M. J., Regiospecific solid-phase synthesis of branched oligoribonucleotides that mimic intronic lariat RNA intermediates. *The Journal of organic chemistry* **2014,** *79* (3), 963-975.

96. Jiang, Q.; Crews, L. A.; Barrett, C. L.; Chun, H. J.; Court, A. C.; Isquith, J. M.; Zipeto, M. A.; Goff, D. J.; Minden, M.; Sadarangani, A.; Rusert, J. M.; Dao, K. H.; Morris, S. R.; Goldstein, L. S.; Marra, M. A.; Frazer, K. A.; Jamieson, C. H., ADAR1 promotes malignant progenitor reprogramming in chronic myeloid leukemia. *Proc Natl Acad Sci U S A* **2013,** *110* (3), 1041-6.

97. 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. *Journal of the American Chemical Society* **2020,** *142* (41), 17766-17781.

98. 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 research* **2018,** *154*, 66-86.

99. Weng, X.; Gong, J.; Chen, Y.; Wu, T.; Wang, F.; Yang, S.; Yuan, Y.; Luo, G.; Chen, K.; Hu, L., Keth-seq for transcriptome-wide RNA structure mapping. *Nature Chemical Biology* **2020,** *16* (5), 489-492.

Appendix A:

Omitted Data from Chapter 2

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

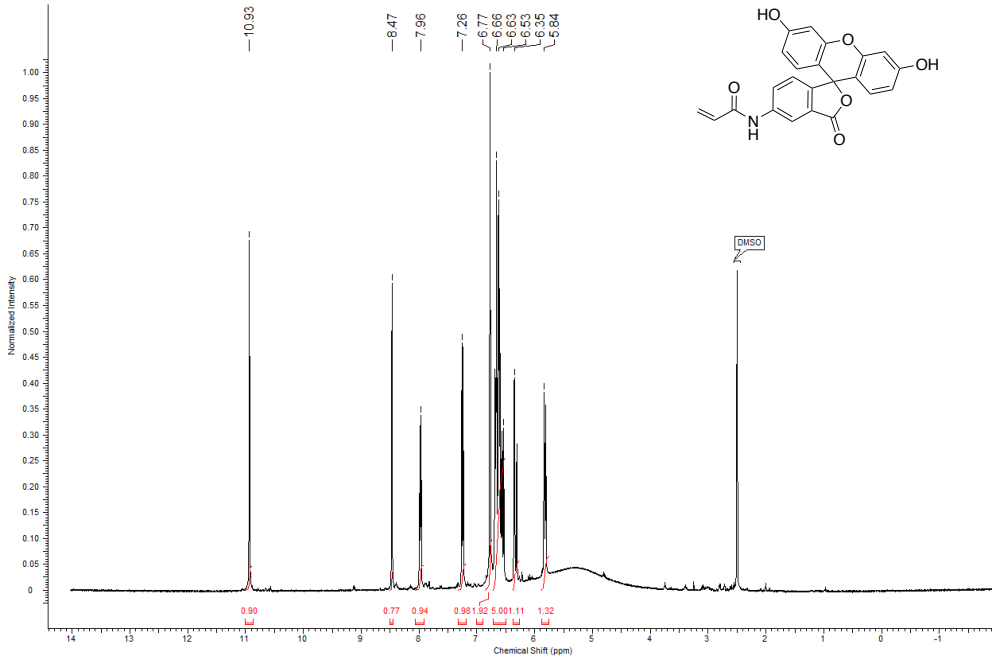


Figure A1. ^{1}H NMR spectrum of acrylamidofluorescein (400 MHz, DMSO-d6).

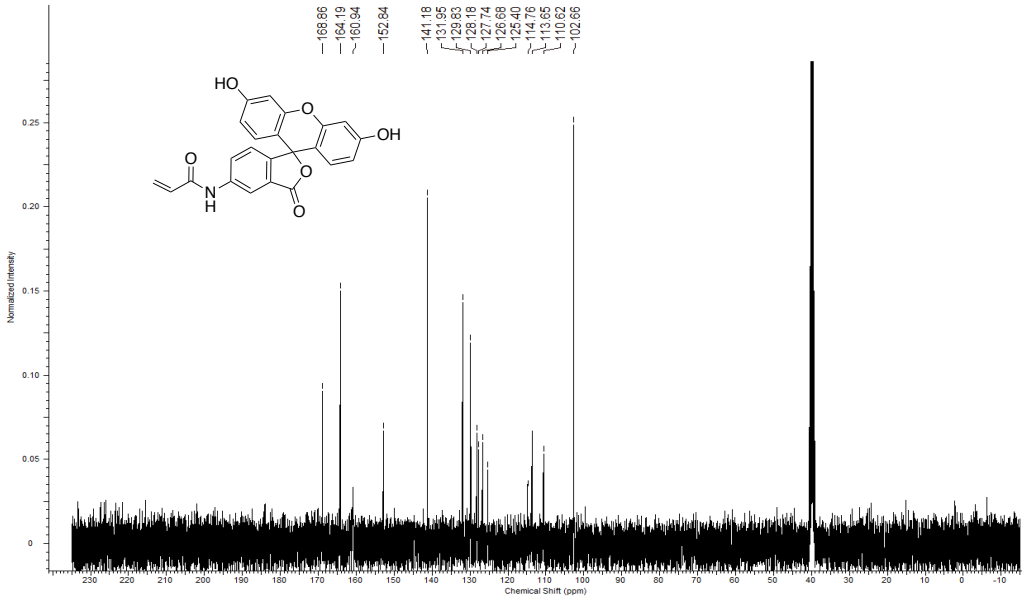


Figure A2. ^{13}C NMR (spectrum of acrylamidofluorescein 400 MHz, DMSO-d6).

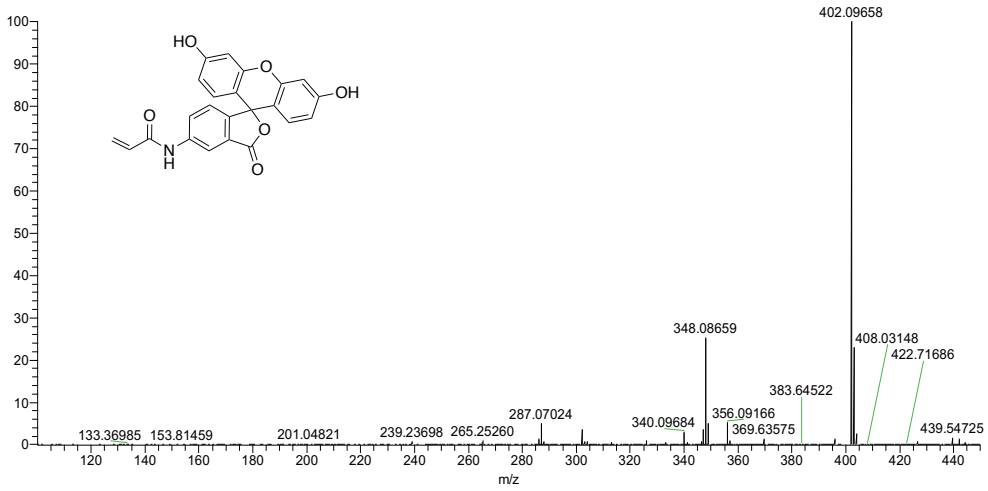
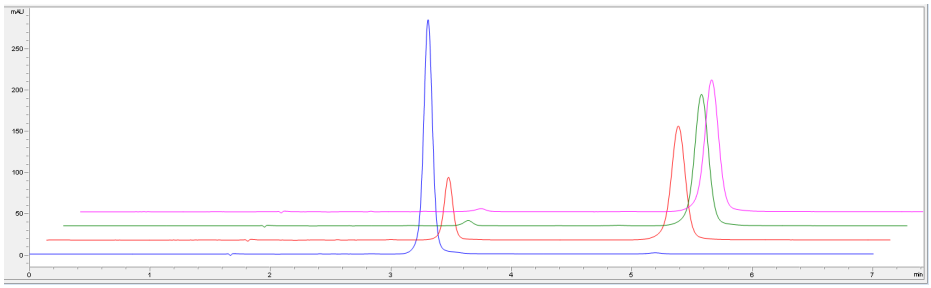


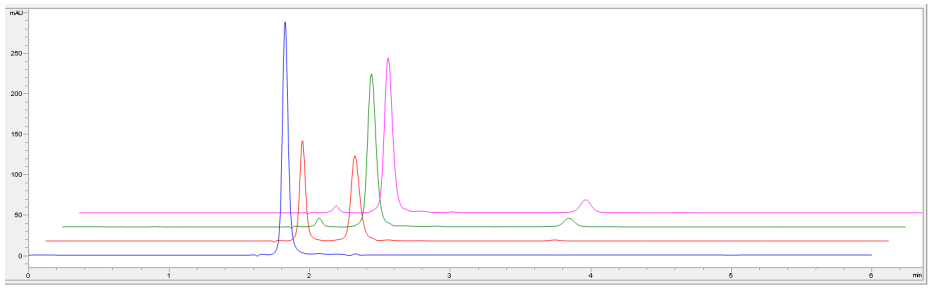
Figure A3. ESI-MS spectrum of acrylamidofluorescein. Calculated $(M+H)^+$ for $C_{23}H_{15}NO_6$ 402.09776

Figure A4. Representative HPLC traces of ribonucleoside reactivity with acrylonitrile. All chromatograms were monitored at 254nm. Blue = 0 hours, red = 1 hour, green = 8 hours, pink = 24 hours.

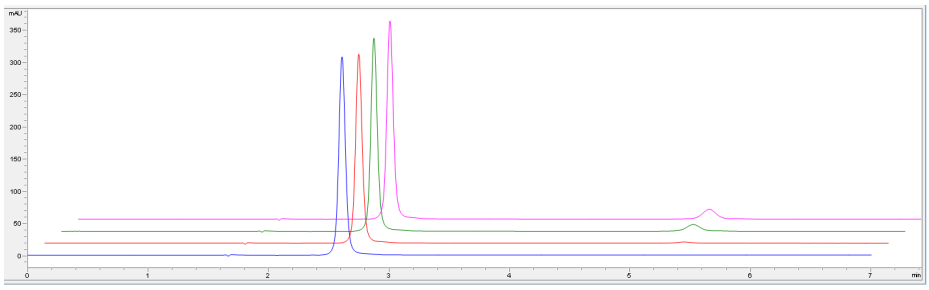
a) Inosine



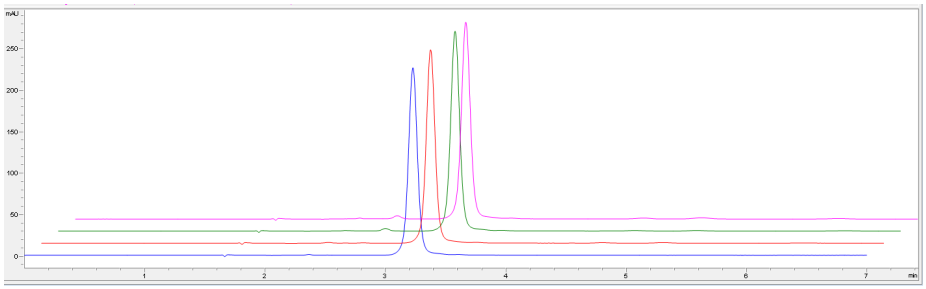
b) Pseudouridine



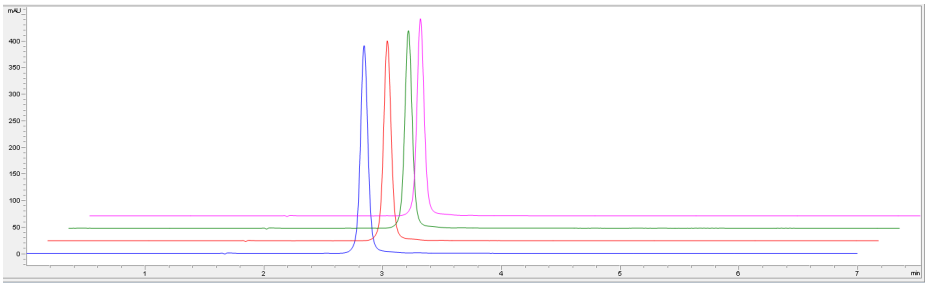
c) Uridine



d) Guanosine



e) Adenosine



f) Cytidine

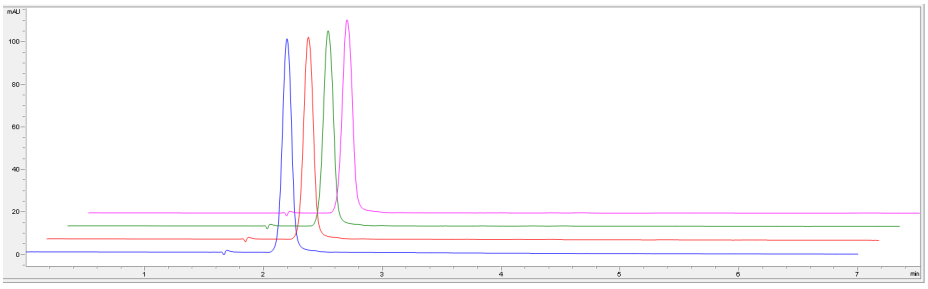
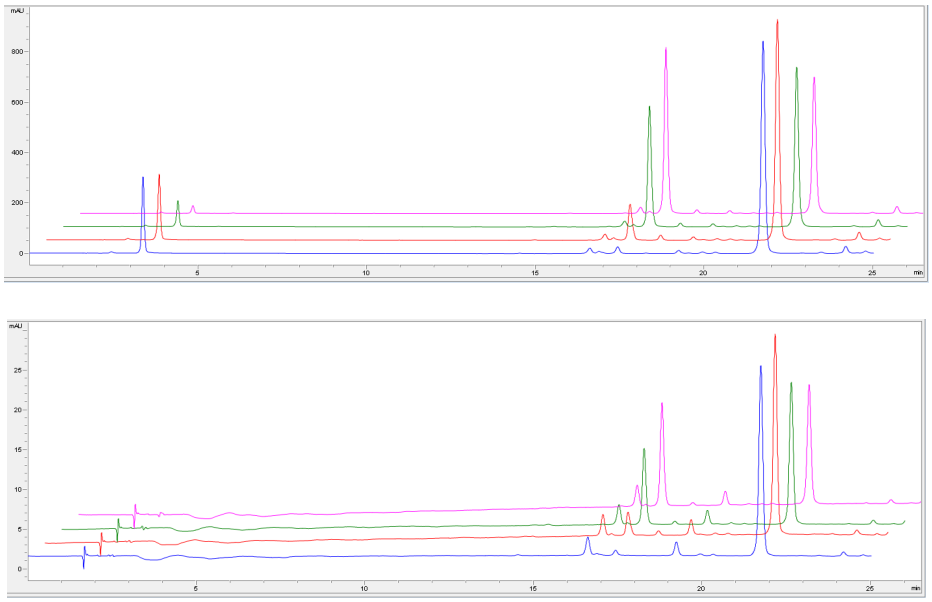


Figure A5. Representative HPLC traces of ribonucleoside reactivity with acrylamidofluorescein. All chromatograms were monitored at 254nm. Blue = 0 hours, red = 1 hour, green = 8 hours, pink = 24 hours.

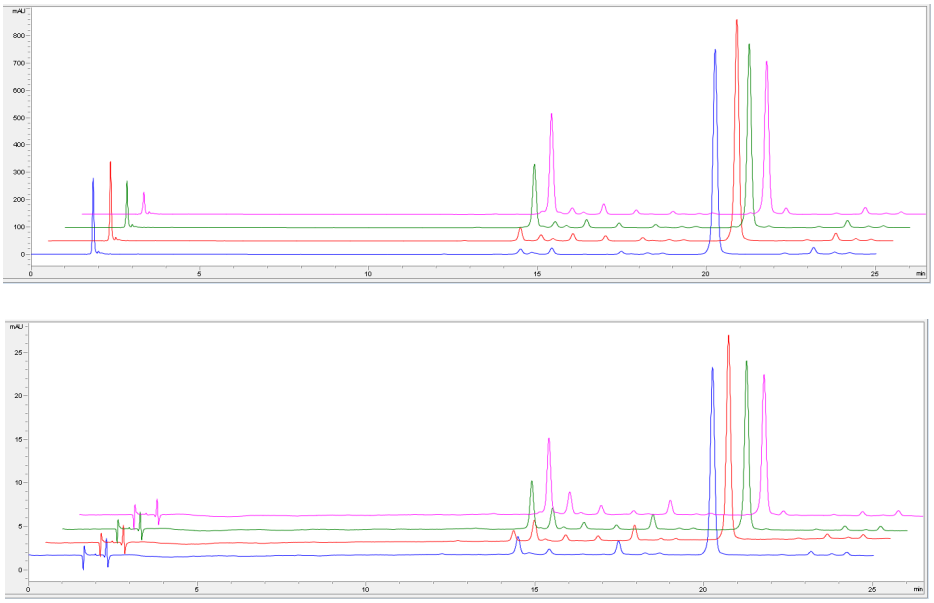
a) Inosine



254nm

494nm

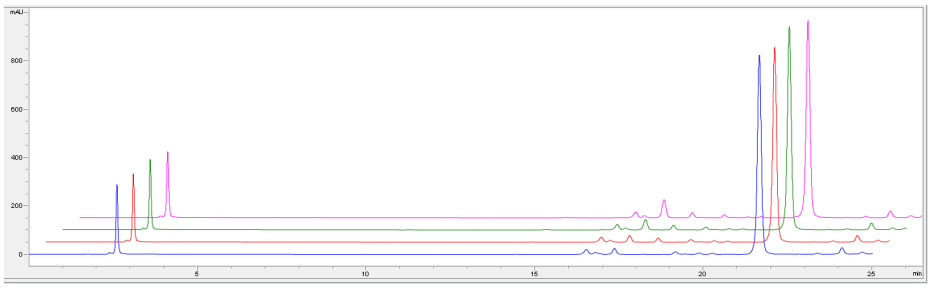
b) Pseudouridine



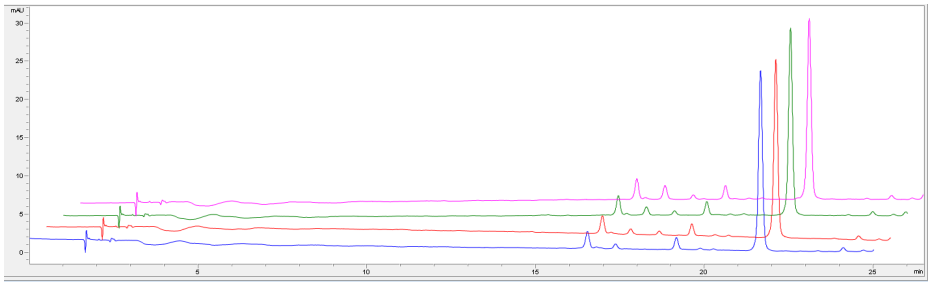
254nm

494nm

c) Uridine

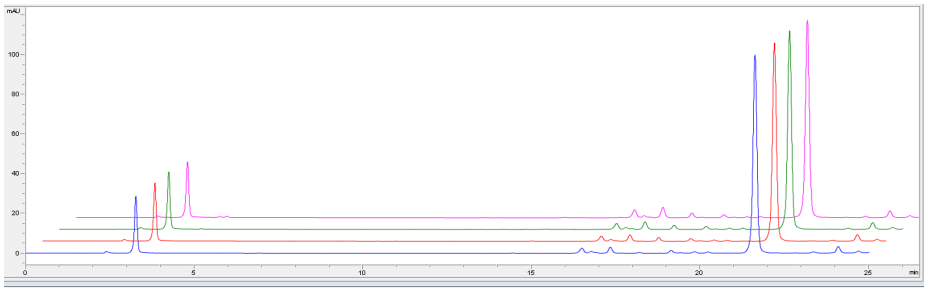


254nm

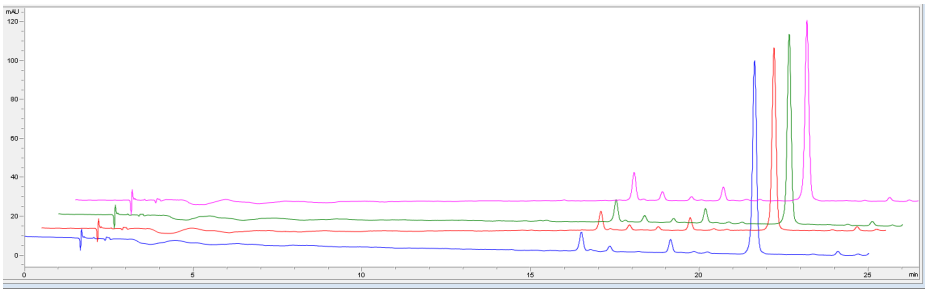


494nm

d) Guanosine

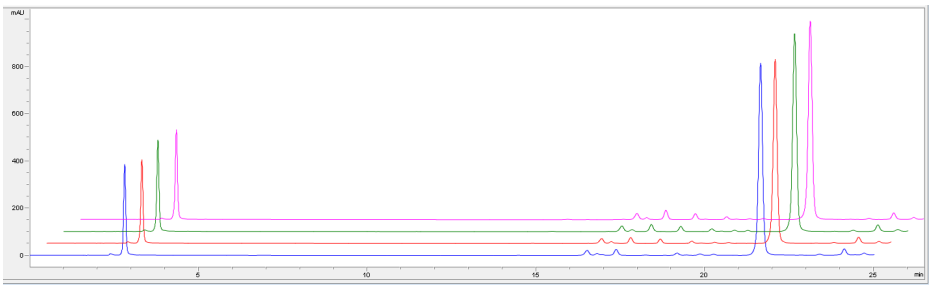


254nm

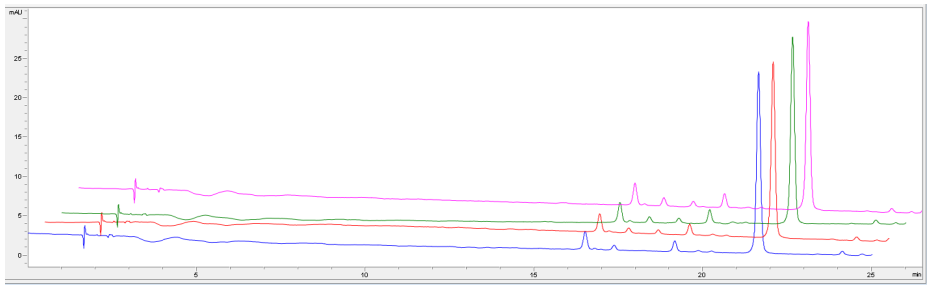


494nm

e) Adenosine

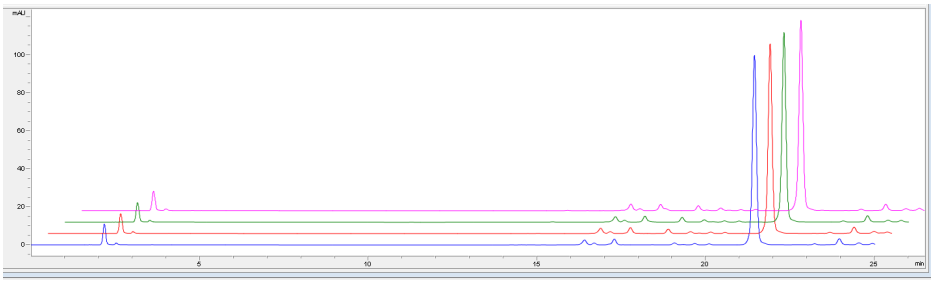


254nm

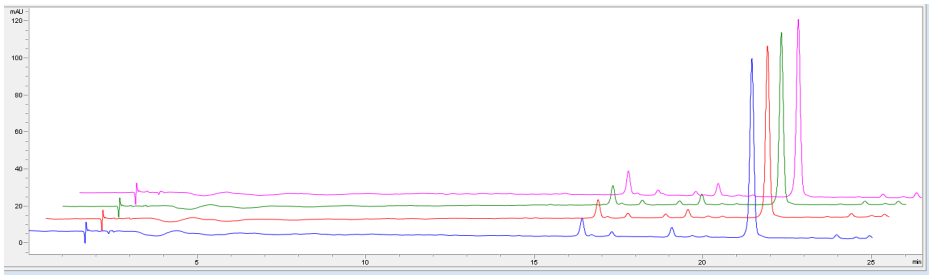


494nm

f) Cytidine

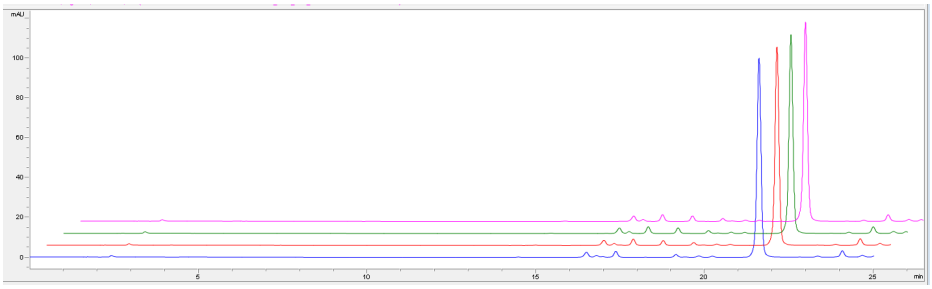


254nm

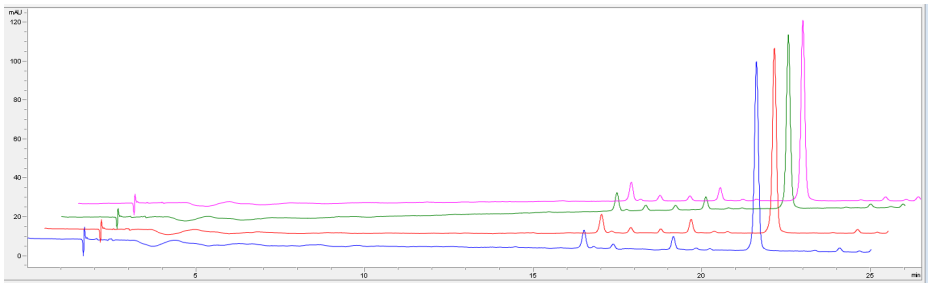


494nm

g) Acrylamidofluorescein alone (no ribonucleoside)



254nm



494nm

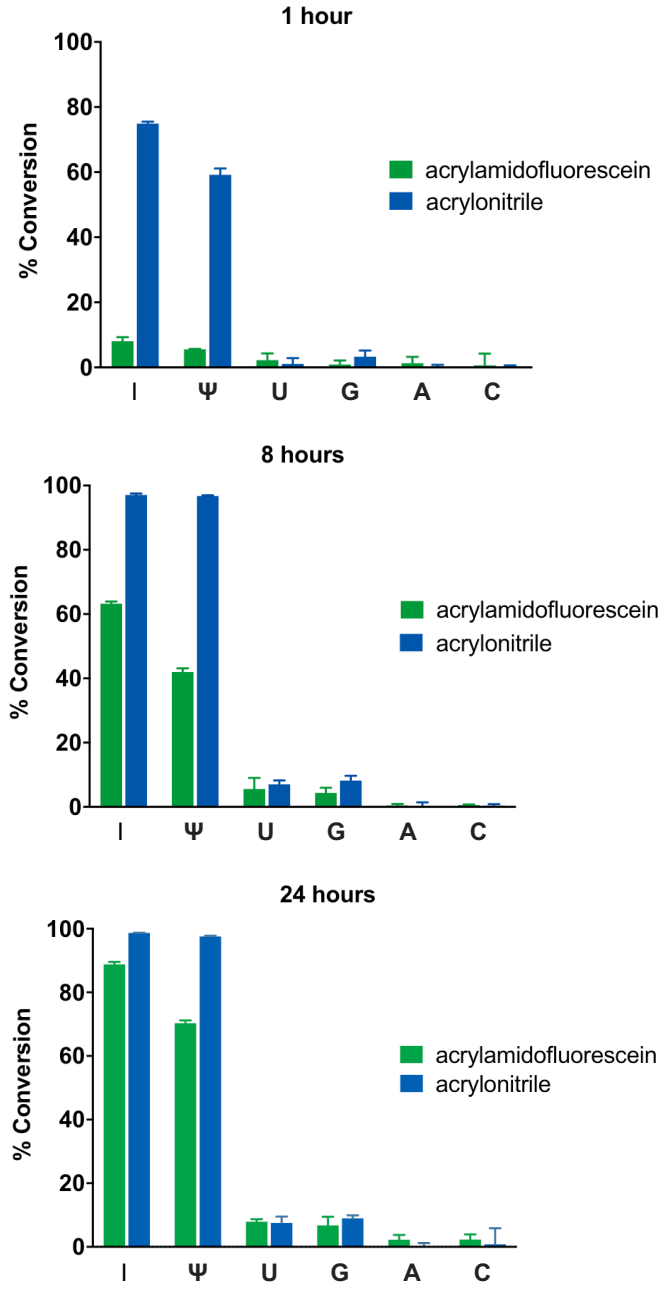


Figure A6. Percent conversion of ribonucleosides when reacted with acrylamidofluorescein (green bars) or acrylonitrile (blue bars) after 1, 8 and 24 hours at 70 °C, pH 8.6.

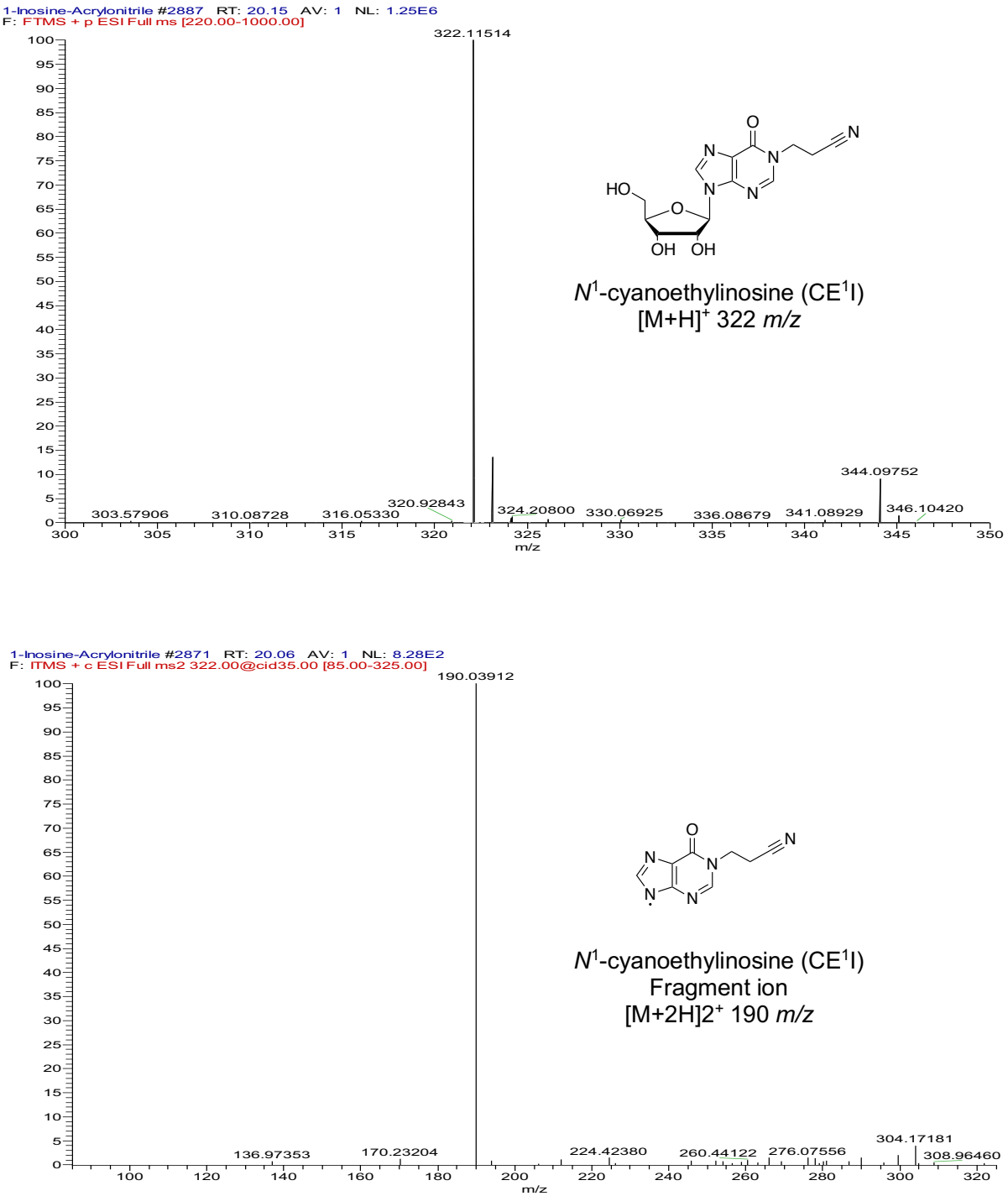
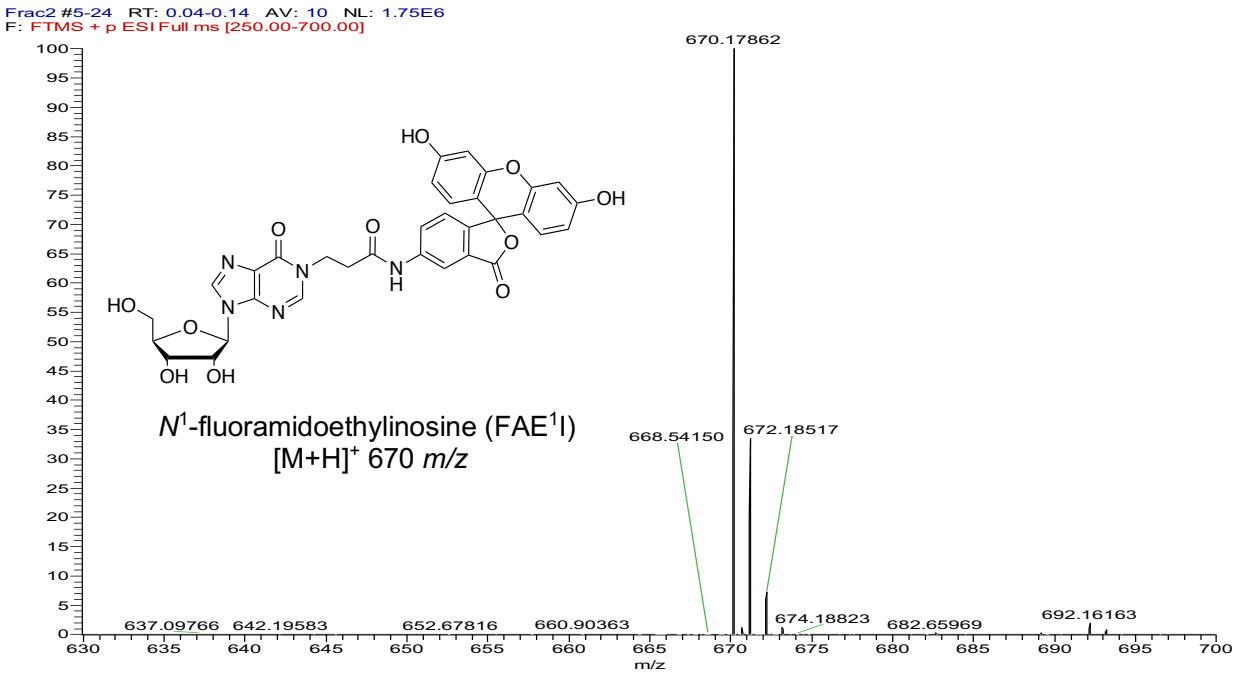


Figure A7. ESI-MS and MS/MS spectra of isolated product fraction for the reaction of inosine and acrylonitrile.



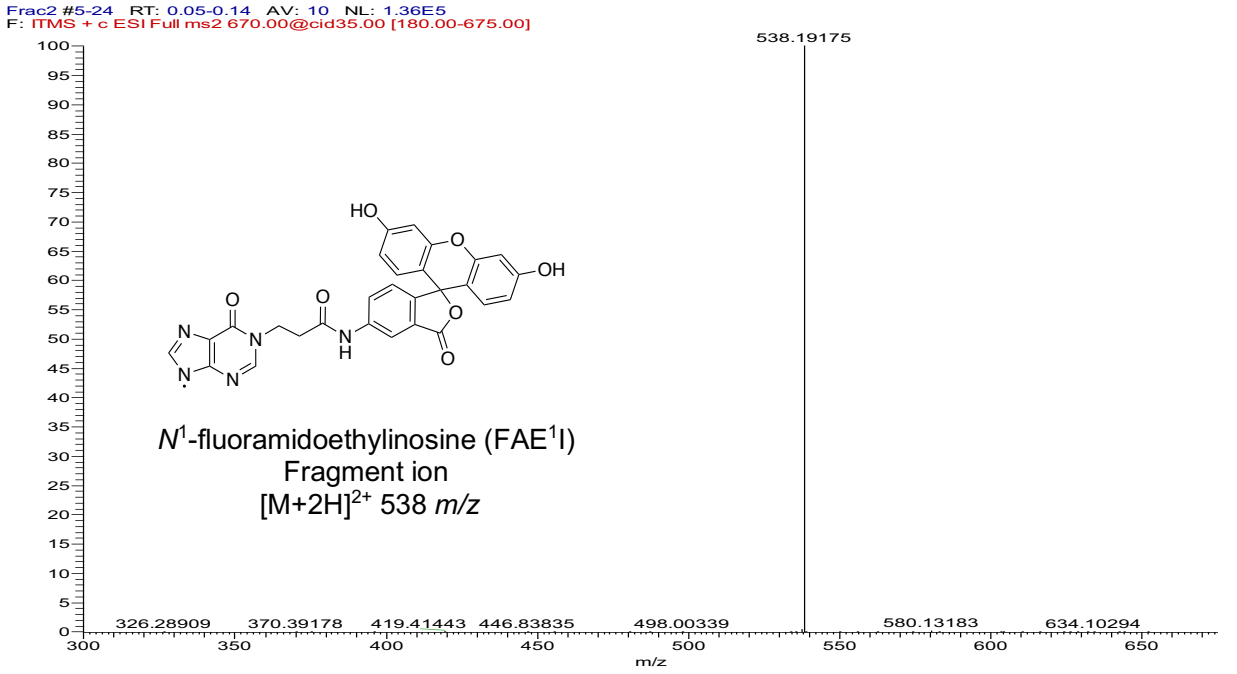
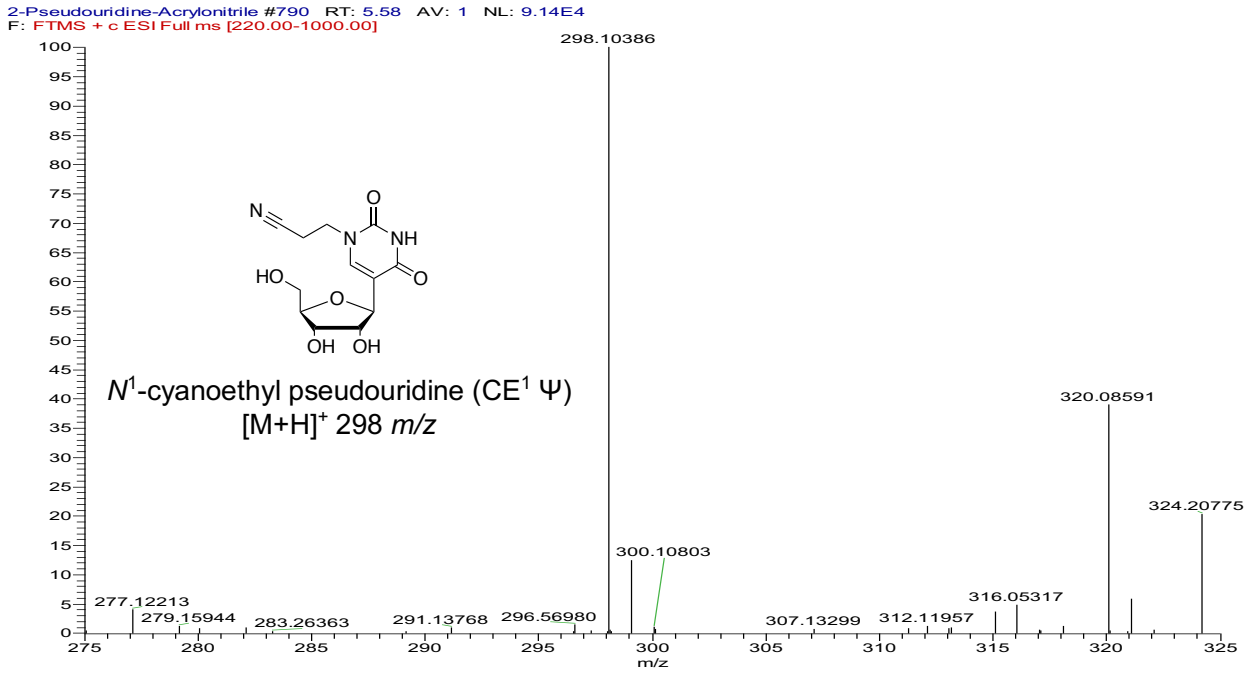


Figure A8. ESI-MS and MS/MS spectra of isolated product fraction for the reaction of inosine and acrylamidofluorescein.



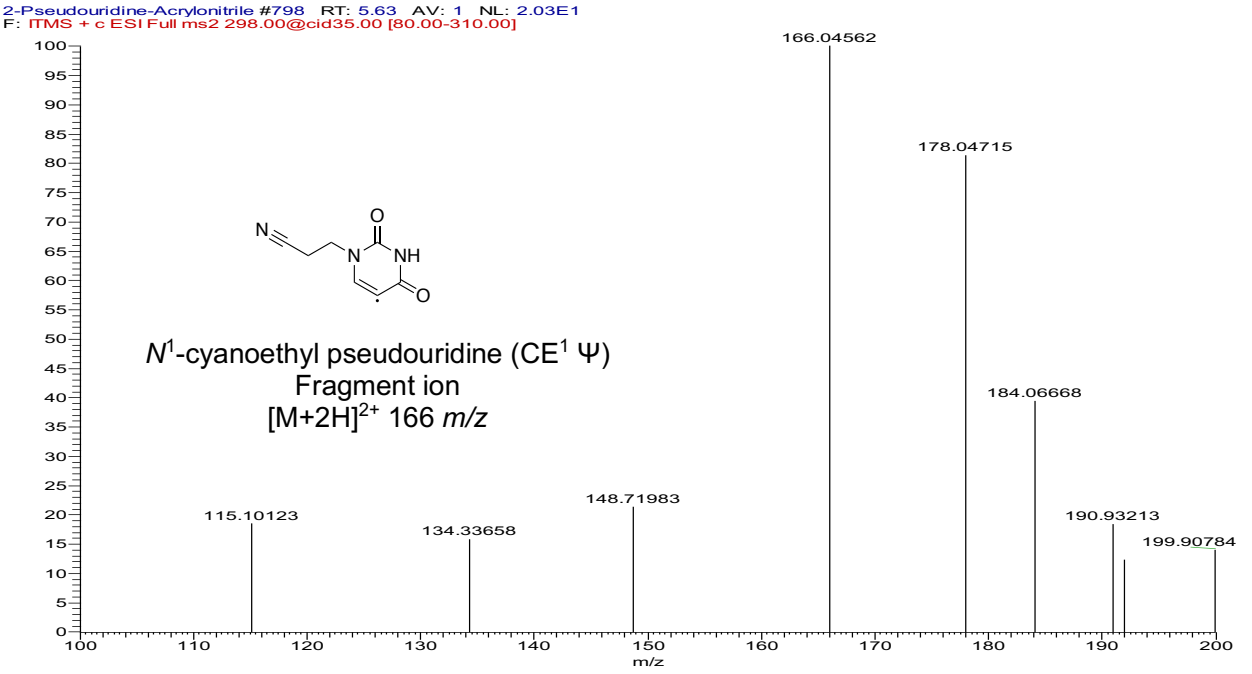


Figure A9. ESI-MS and MS/MS spectra of isolated product fraction for the reaction of pseudouridine and acrylonitrile.

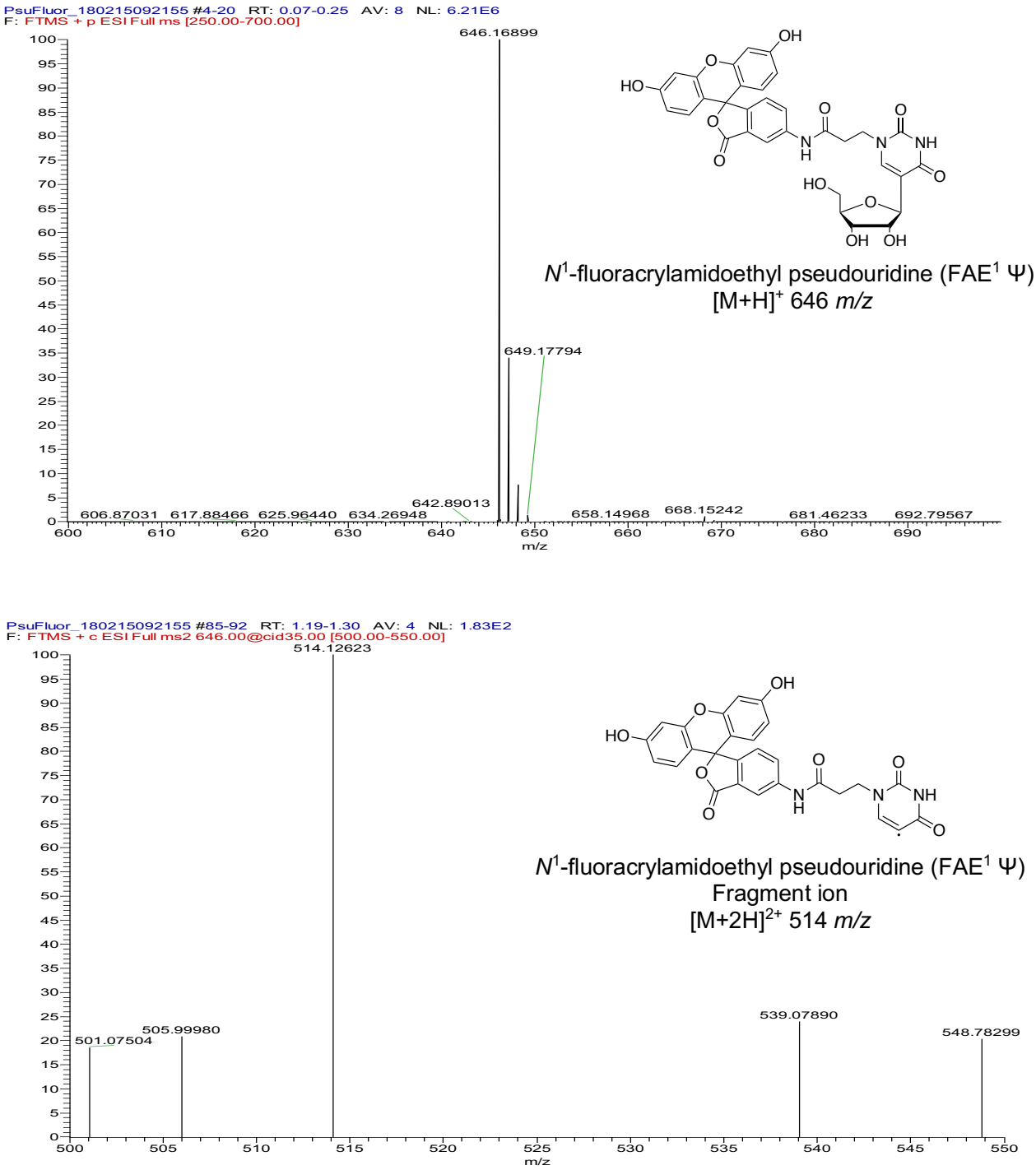


Figure A10. ESI-MS and MS/MS spectra of isolated product fraction for the reaction of pseudouridine and acrylamidofluorescein.

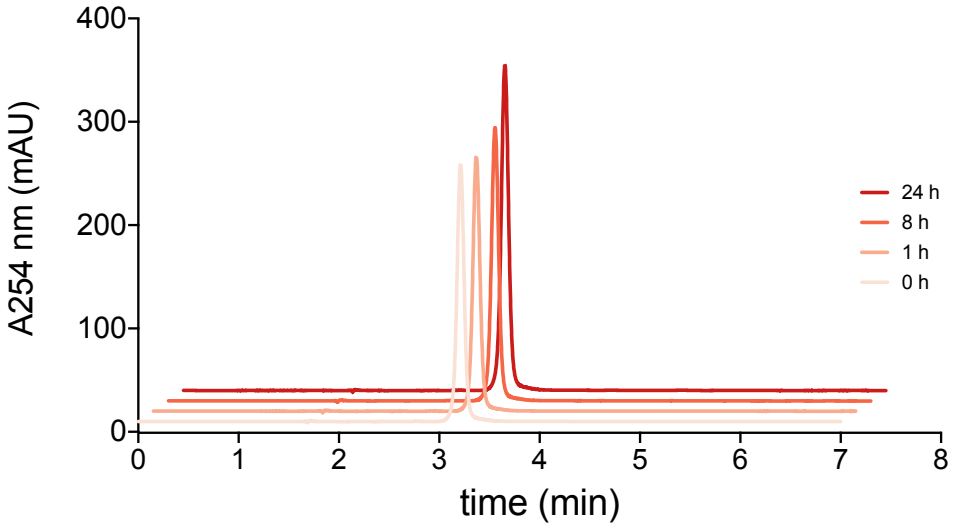
Appendix B:

Omitted Data from Chapter 3

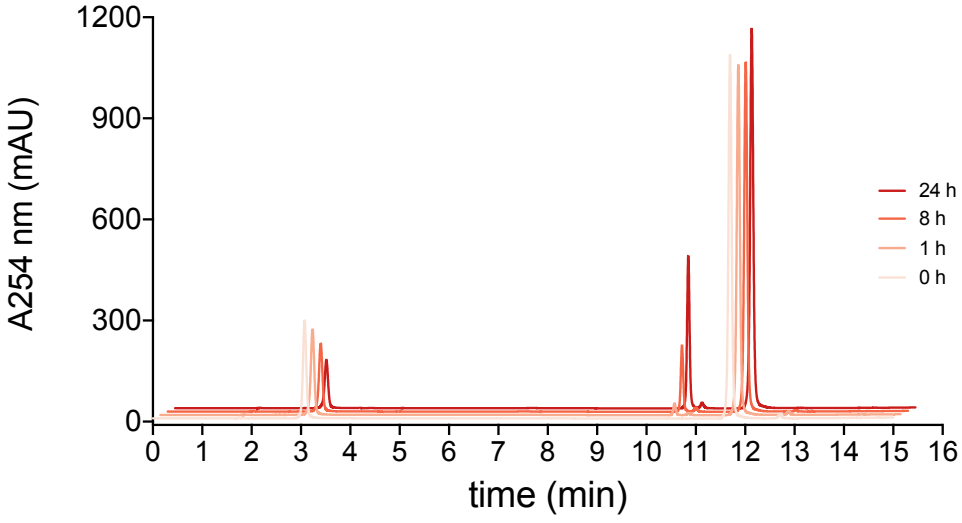
Chemical Profiling of A-to-I RNA Editing Using a Click-Compatible Phenylacrylamide

Figure B1. Representative HPLC traces of inosine reactivity with different acrylamide derivatives. Reaction times (0-24 h) indicated by included legends. (Below).

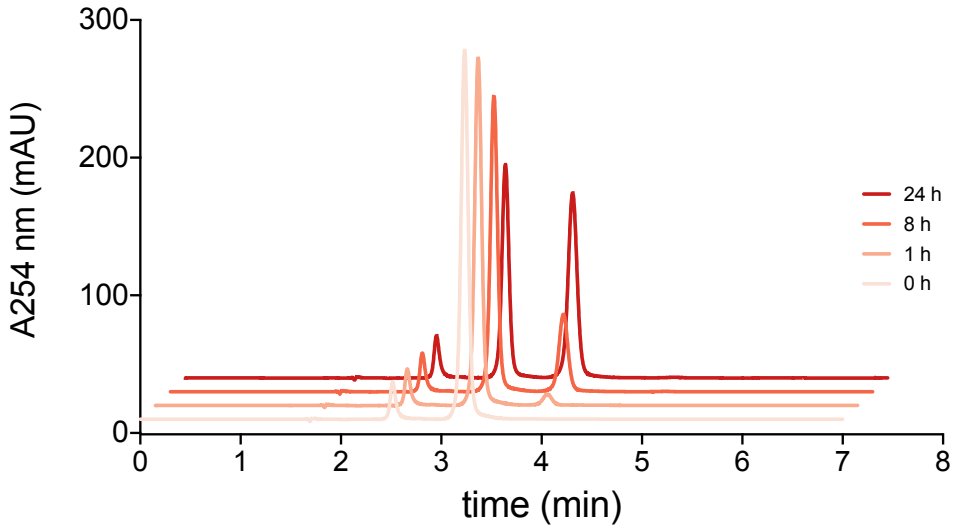
a) Inosine alone



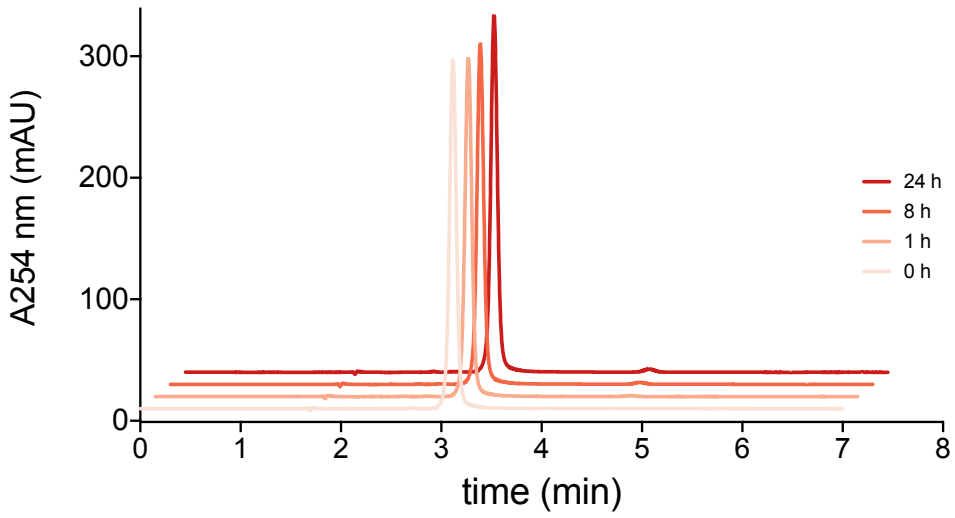
b) Inosine + *N*-phenylacrylamide



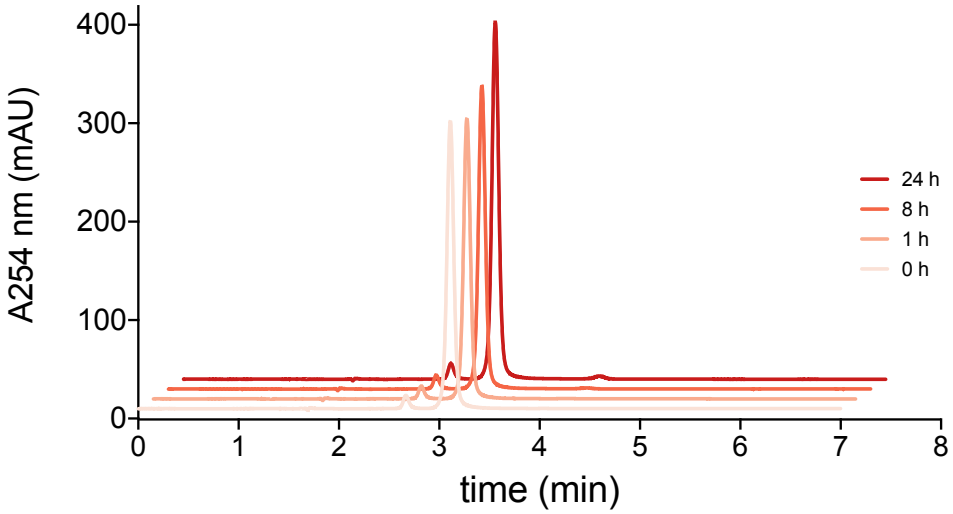
c) Inosine + acrylamide



d) Inosine + mPEG-acrylamide



e) Inosine + *N*-hydroxyethylacrylamide



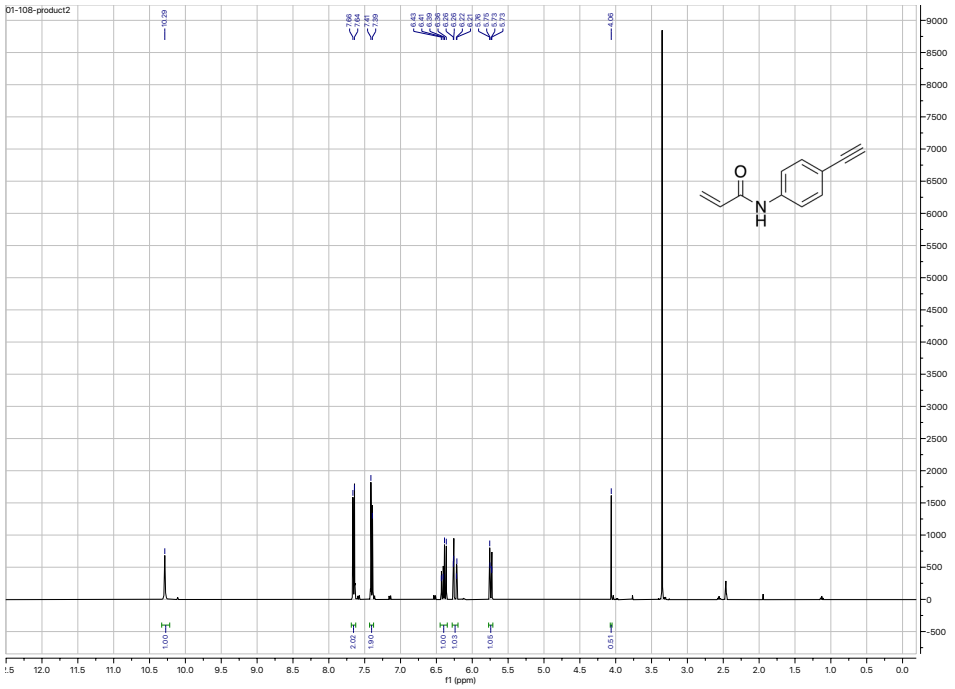


Figure B2. ^{1}H NMR spectrum of EPhAA (400 MHz, DMSO-d6).

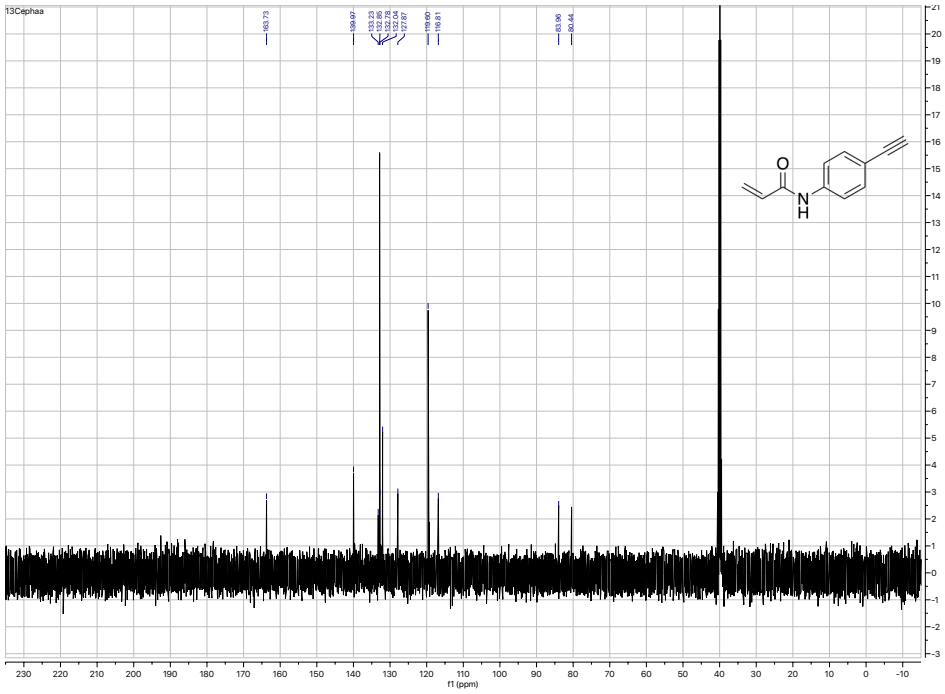


Figure B3. ^{13}C NMR spectrum of EPhAA (400 MHz, DMSO-d6).

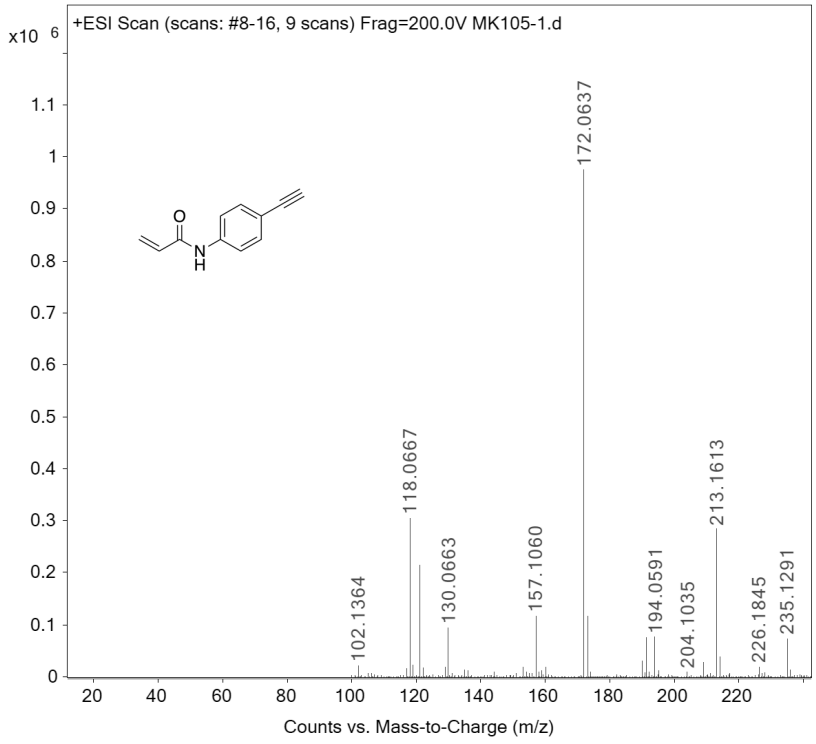
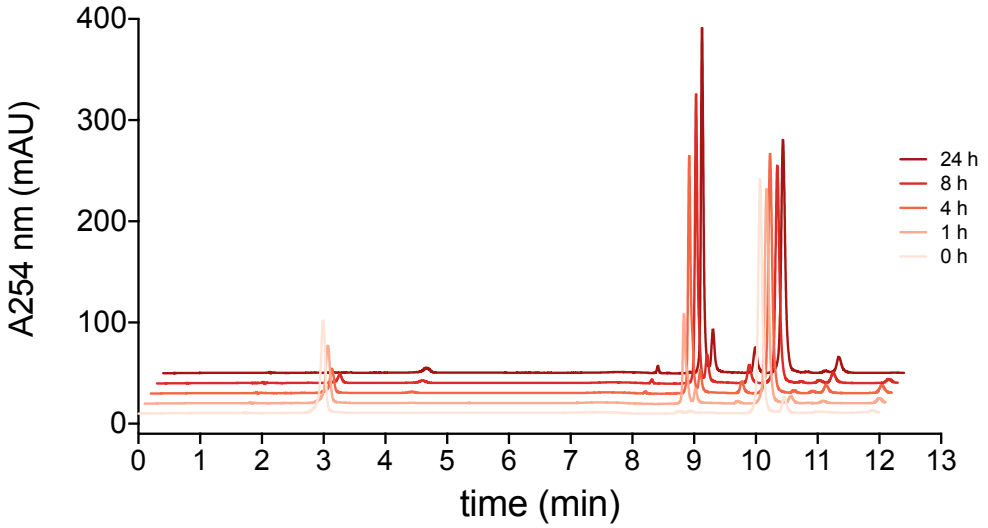


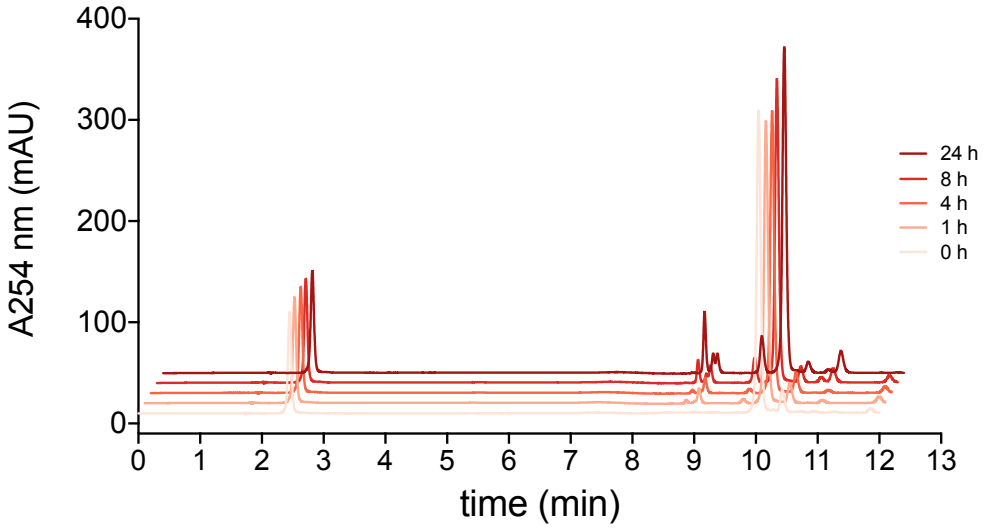
Figure B4. ESI-MS spectrum of EPhAA. Calculated $(M+H)^+$ for $C_{11}H_{10}NO$: 172.0762.

Figure B5. Representative HPLC traces of ribonucleoside reactivity with ethynylphenylacrylamide. Reaction times (0-24 h) indicated by included legends. (Below).

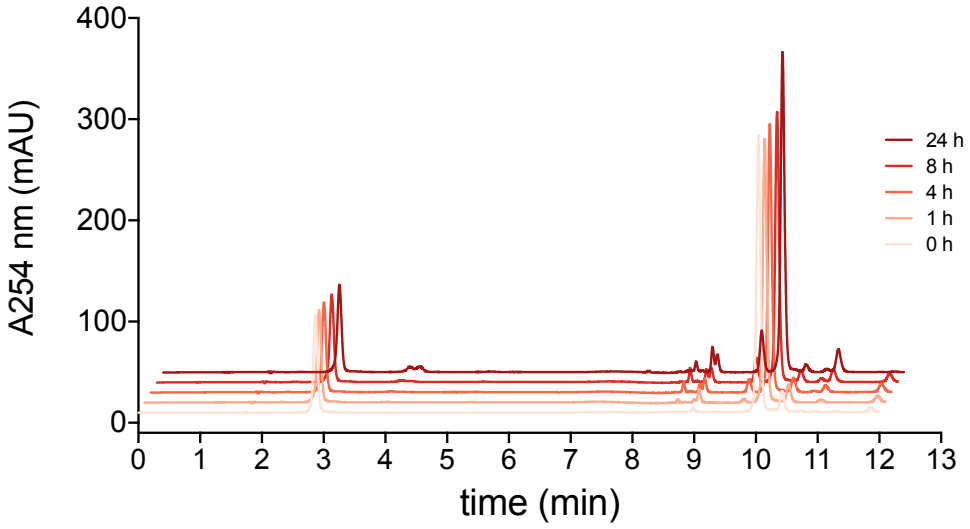
a) inosine



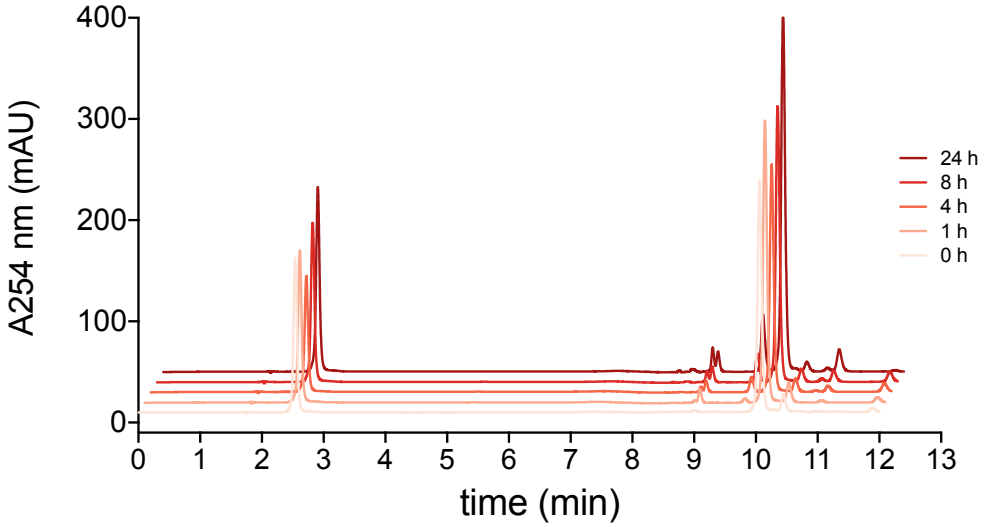
b) uridine



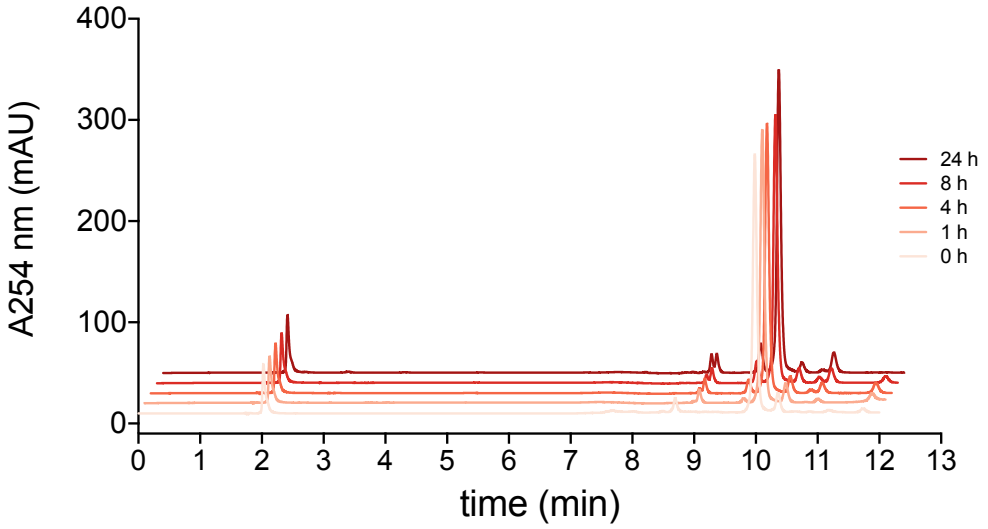
c) guanosine



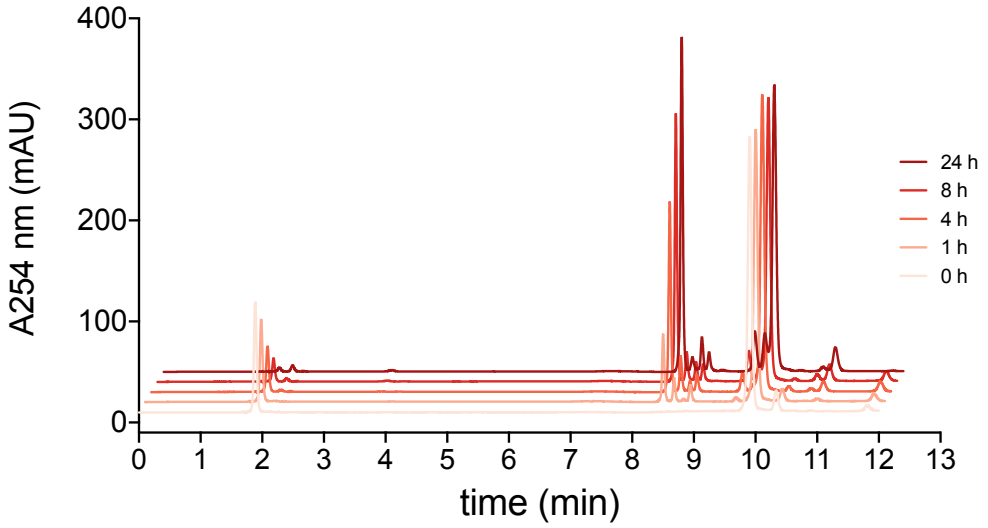
d) adenosine



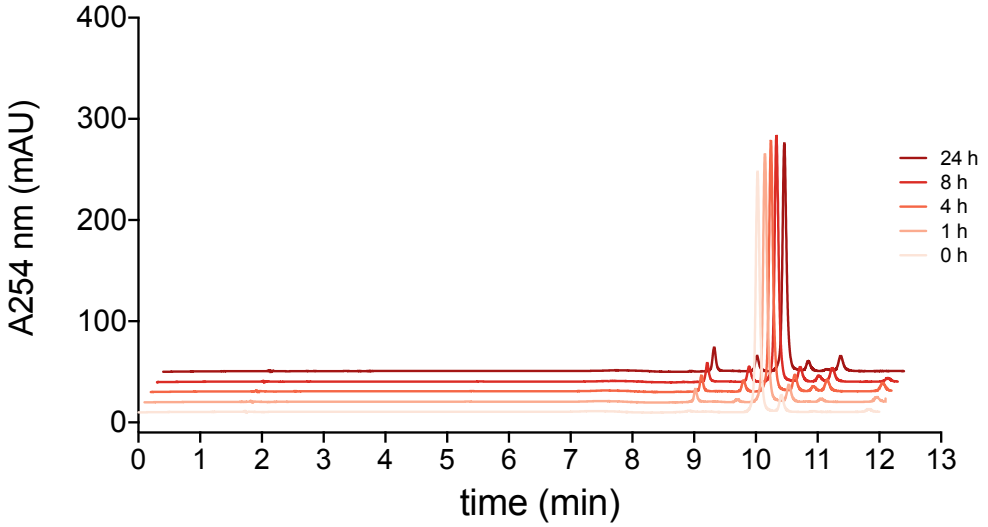
e) cytidine



f) pseudouridine



g) ethynylphenylacrylamide alone



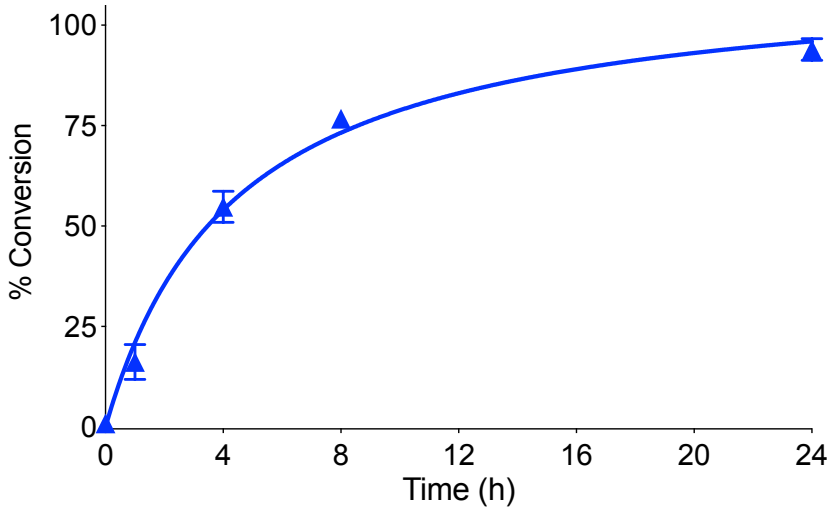
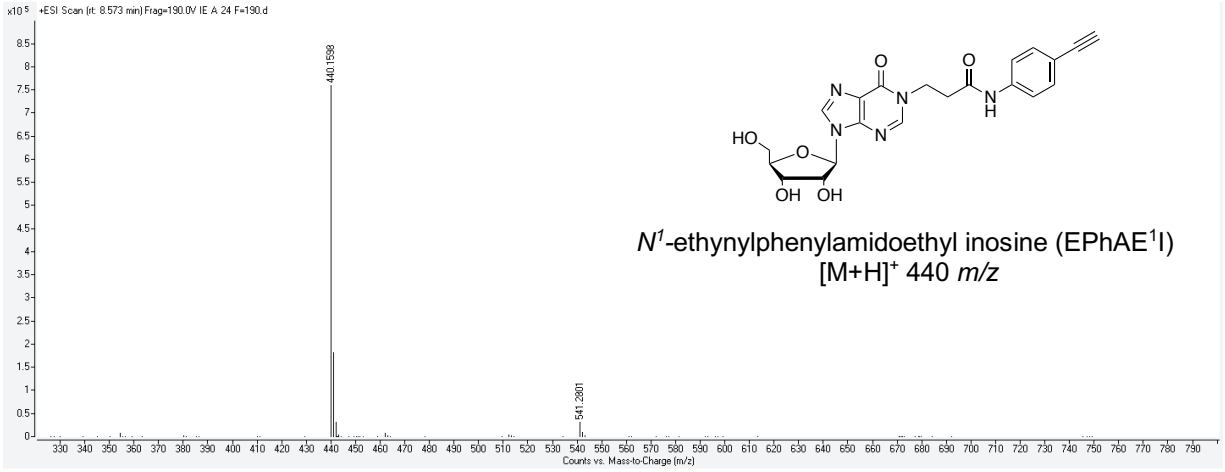


Figure B6. EPhAA reactivity with pseudouridine over 24 hours as monitored by HPLC. Values represent mean with S.D. error bars (n = 3).



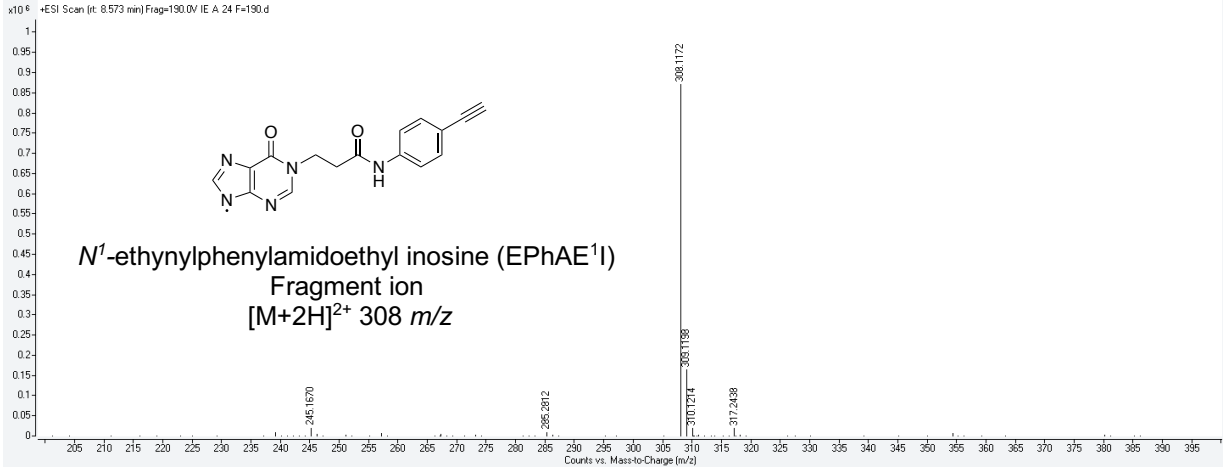
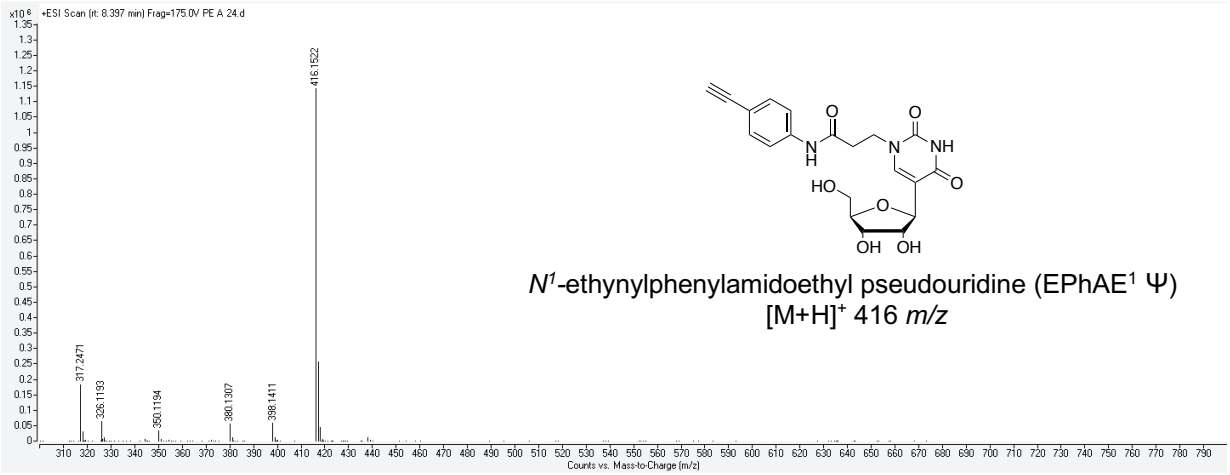


Figure B7. ESI-MS and MS/MS spectra of isolated product fraction for the reaction of inosine and ethynylphenylacrylamide.



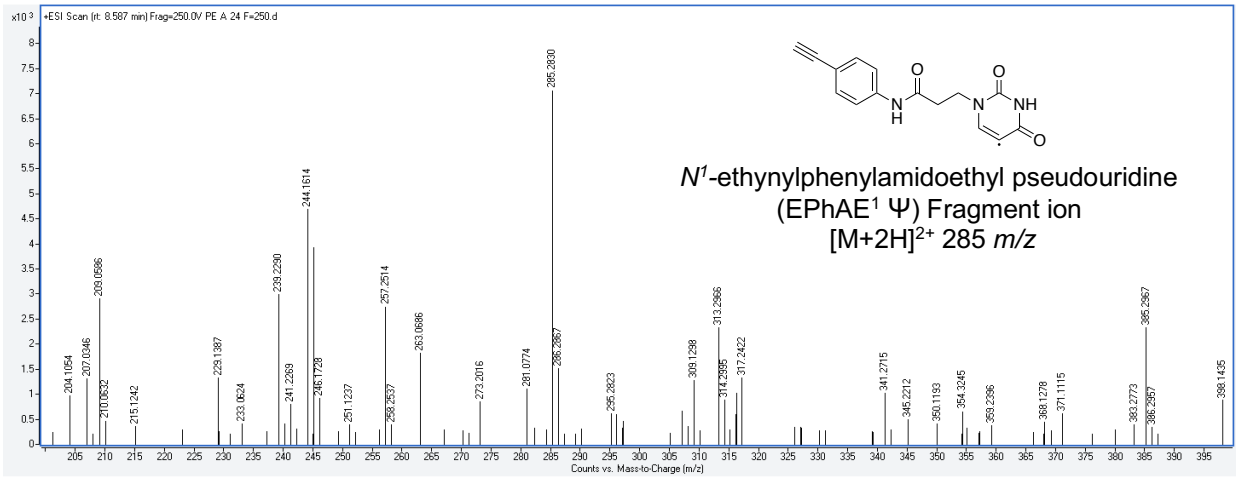


Figure B8. ESI-MS and MS/MS spectra of isolated product fraction for the reaction of pseudouridine and ethynylphenylacrylamide.

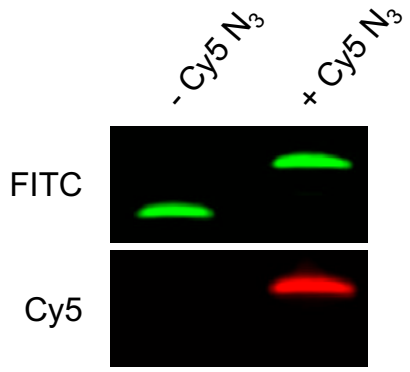


Figure B9. Denaturing PAGE analysis of CuAAC reactions using an alkyne-functionalized DNA with and without Cy5-N_3.

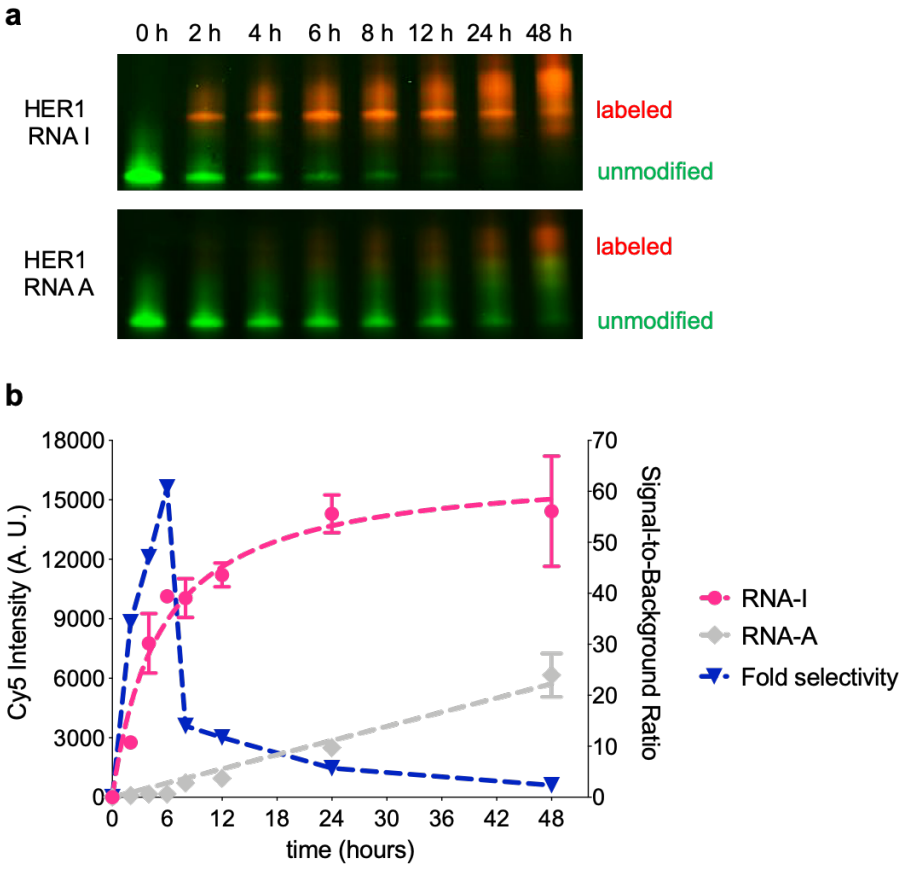


Figure B10. Selectivity for inosine as a function of EPhAA labeling time. a) denaturing PAGE analysis of RNA I and RNA A labeling reactions, showing both Cy5 fluorescence (red) and SYBR gold (green) staining. b) Densitometric Cy5 intensity quantification of RNA I (pink) and RNA A (gray) and fold selectivity (A/I ratio, blue). Values represent mean intensities (arbitrary units, A.U.) with S.D. error bars (n = 2).

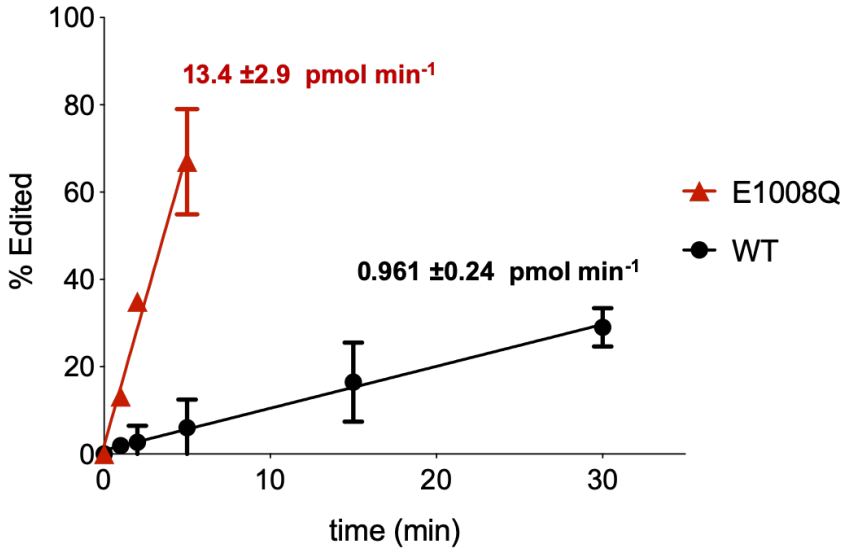


Figure B11. Initial deamination velocities for WT and E1008Q hADAR1 enzymes. Regression analysis for linear portions of each editing curve. Values represent mean with S.D. error bars. Velocities represent mean with 95% confidence intervals (n = 2).

Appendix C:

Omitted Data from Chapter 4

Selective Enrichment of A-to-I Edited Transcripts from Cellular RNA Using Endonuclease V

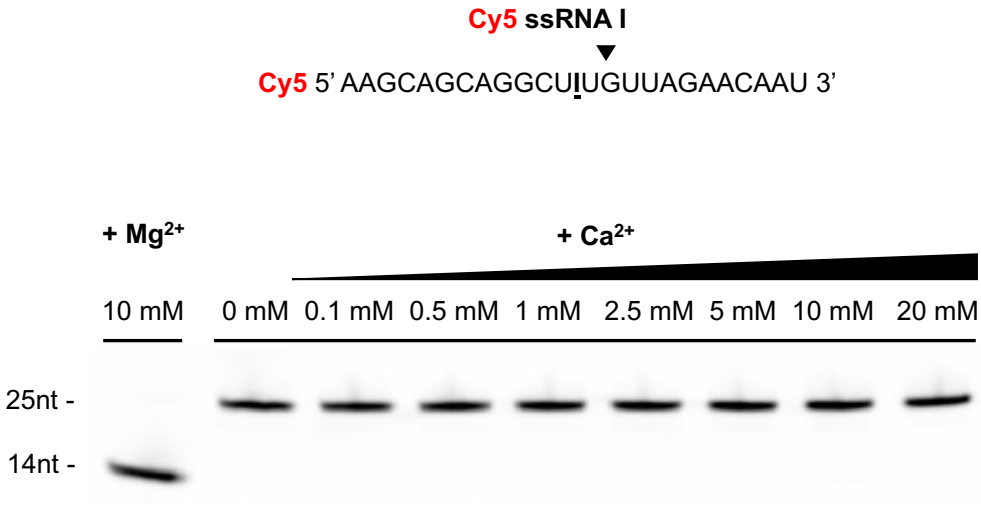


Figure C1. EndoV does not cleave RNA in the presence of Ca^{2+}. Cy5 labeled ssRNA I (sequence shown at top with cleavage site indicated by black arrow) was incubated with EndoV and either 10 mM Mg_2Cl or increasing amounts of Ca_2Cl as indicated. Cleavage is observed only in the presence of Mg^{2+}. Reactions were incubated at room temperature for 3 hours and resolved by 10 % denaturing PAGE.

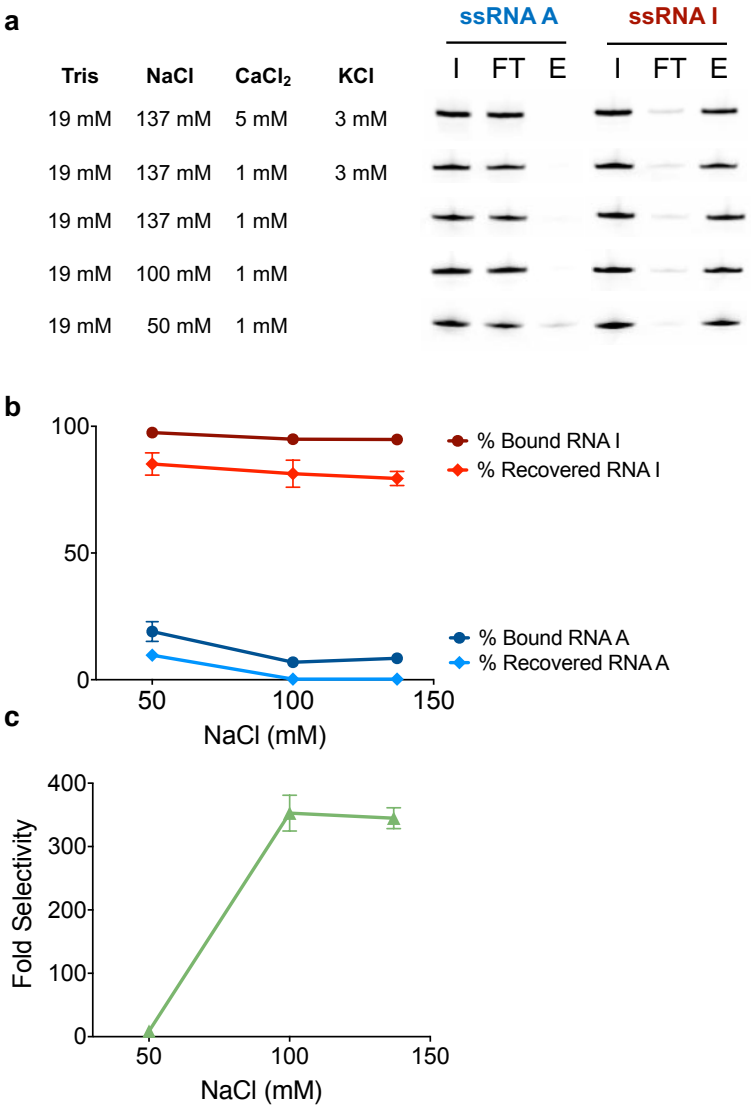


Figure C2. Effect of binding buffer ionic strength on EndoVIPER performance. a) Representative PAGE analysis of initial (I), flowthrough (FT) and eluate (E) EndoVIPER fractions when tested with varying concentrations of $CaCl_2$, NaCl, and KCl in binding buffer. b-c) Densitometric analysis of EndoVIPER efficiency and selectivity as a function of NaCl concentration. Values represent mean with standard deviation. (n = 2)

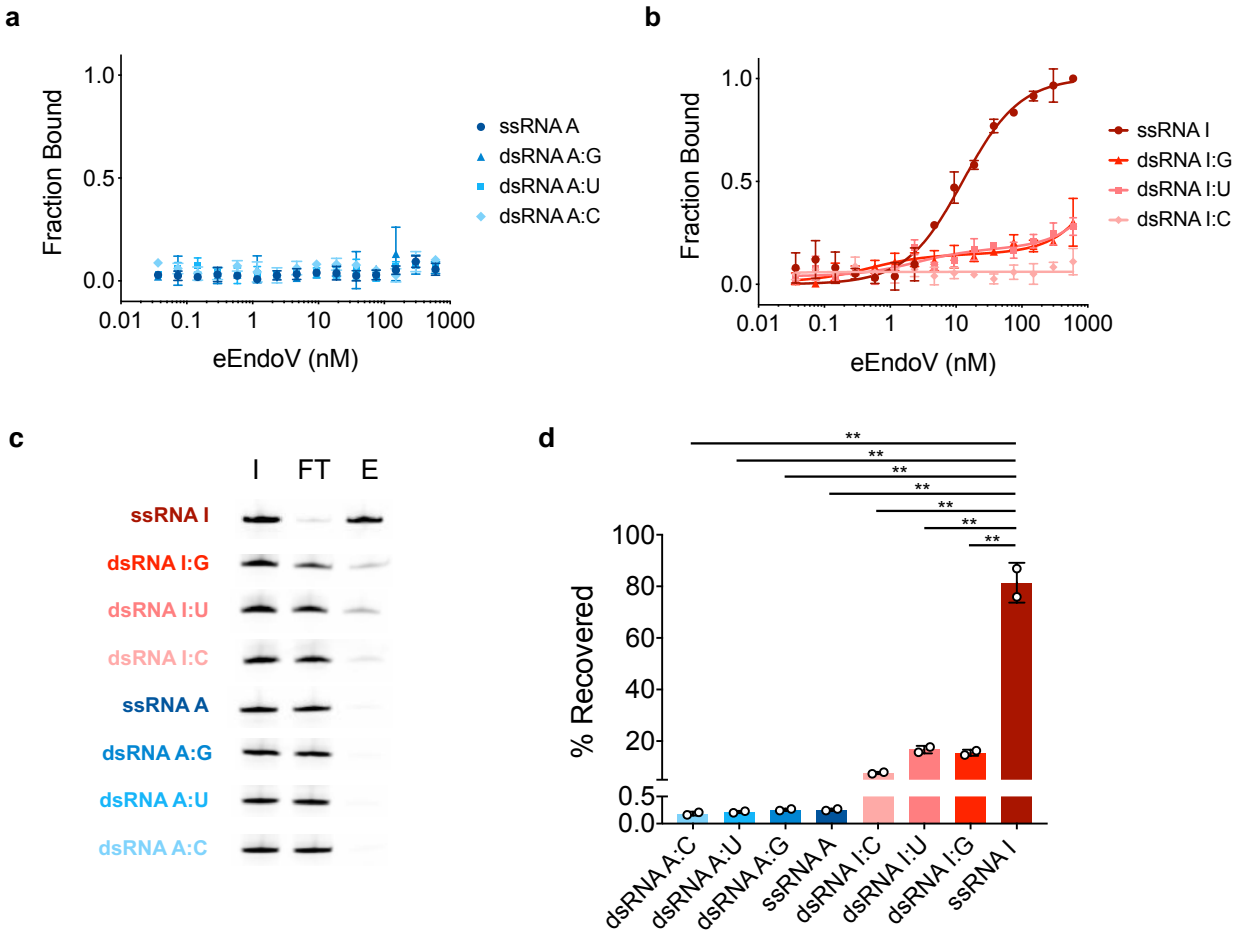


Figure C3. Reduced ionic strength buffer (19 mM Tris, 100 mM NaCl, 1mM $CaCl_2$, pH 7.4 does not alleviate eEndoV structural binding preferences toward ssRNA. Using buffer formulations from Fig S1, we performed MST analysis of eEndoV binding affinity towards a) dsRNA A and b) dsRNA I targets. Values represent mean with standard deviation. (n = 3) c) Representative PAGE analysis of initial (I), flowthrough (FT) and eluate (E) EndoVIPER fractions when tested with various dsRNA targets. d) Densitometric analysis of EndoVIPER efficiency for dsRNA targets. Values represent mean with standard deviation (n = 2). Unpaired t-test was performed for all samples against ssRNA I pulldowns showing significantly different yields (** denotes $p = 0.0045$ for all comparisons).

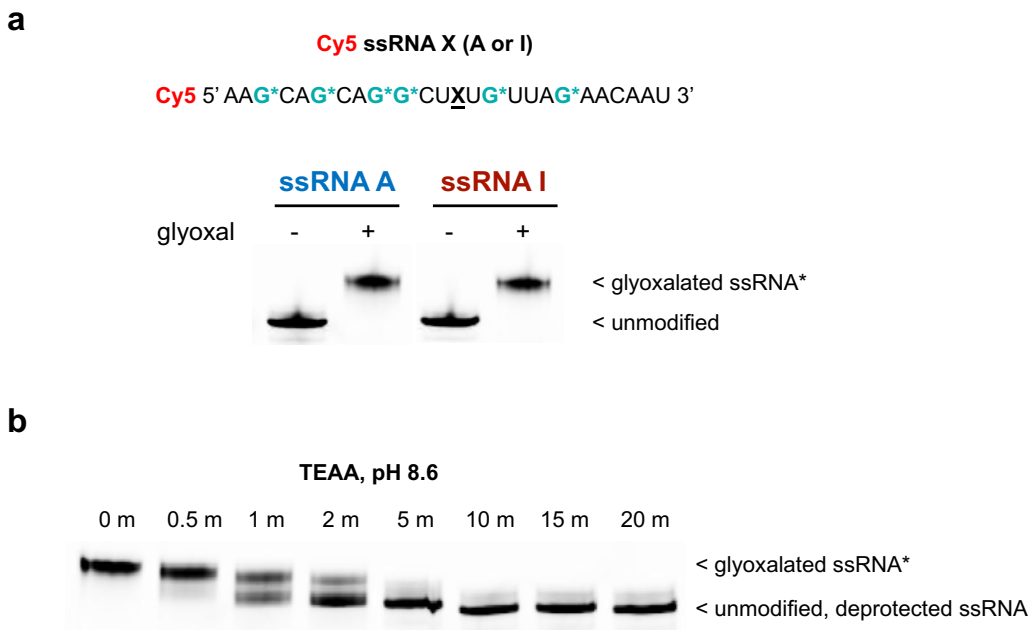


Figure C4. Glyoxal reversibly reacts with guanine residues on RNA and disrupts secondary structure. a) Oligoribonucleotide test sequences with glyoxalated guanosine residues (G*) and 20% PAGE analysis of reactions ssRNAs with glyoxal illustrating upward molecular weight gel shift. b) Kinetic analysis of glyoxal deprotection conditions. 10pmol of a glyoxalated ssRNA was incubated at 95 ºC for the indicated time periods and analyzed with 20% PAGE.

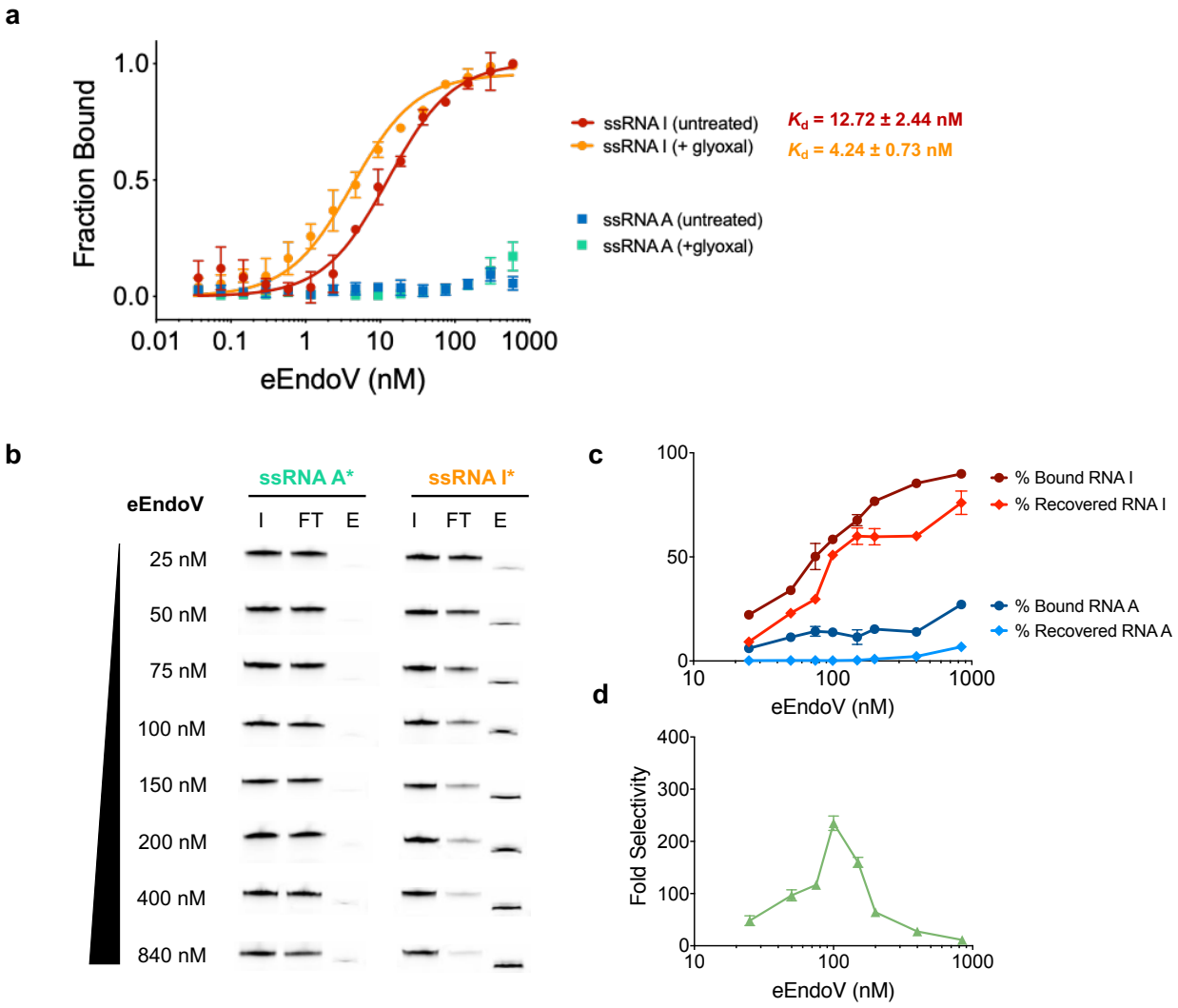


Figure C5. Glyoxal treatment is compatible with EndoVIPER. a) MST analysis of eEndoV binding affinity towards glyoxal treated and unmodified ssRNA A and ssRNA I. Values represent mean with standard deviation. K_d denotes mean with 95% confidence interval (n = 3). b) Representative PAGE analysis of initial (I), flowthrough (FT) and eluate (E) EndoVIPER fractions when tested with increasing concentrations of eEndoV against glyoxal-treated ssRNA targets. c-d) Densitometric analysis of EndoVIPER efficiency and selectivity for glyoxal-treated ssRNA targets. Values represent mean with standard deviation. (n = 2)

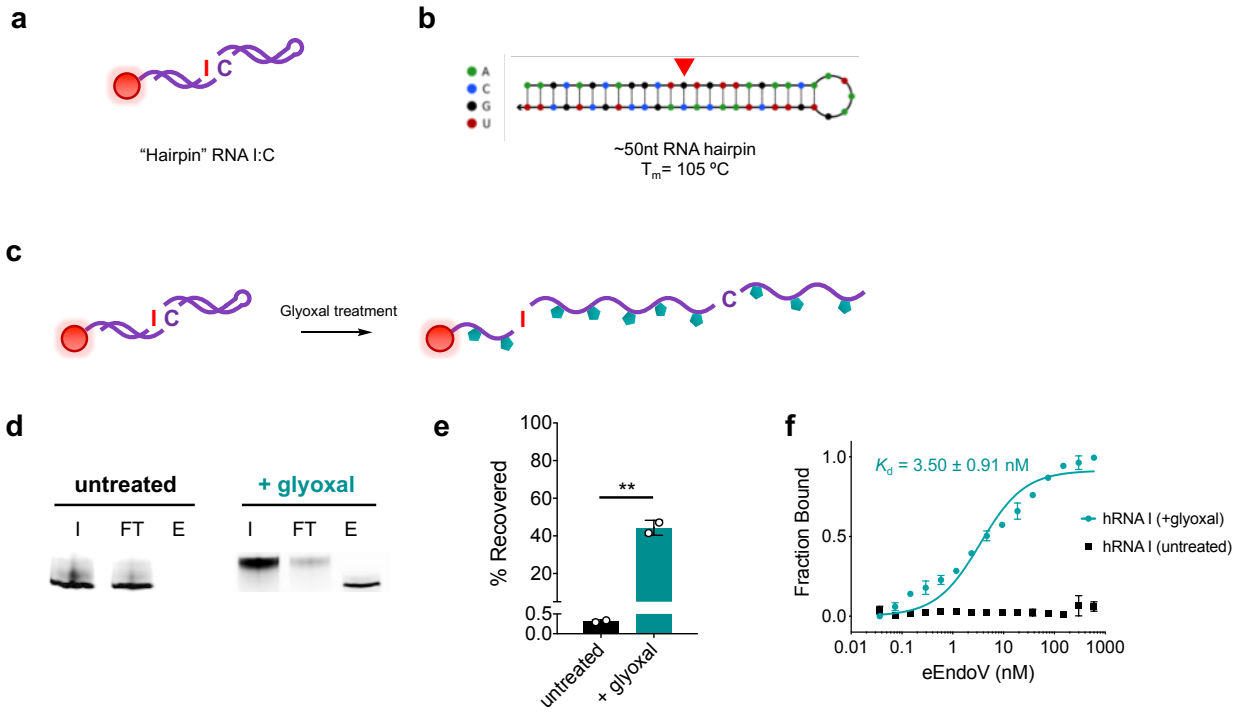


Figure C6. Glyoxal treatment with EndoVIPER enables robust binding and pulldown efficiency in RNAs with high degrees of secondary structure. a) Schematic of test hairpin RNA I substrate (hRNA I) labeled with Cy5, and b) NUPACK analysis illustrating calculated secondary structure and melting temperature. Inosine site is indicated by the red arrow. c) Glyoxal treatment disrupts secondary structure formation, unfolding the hairpin and providing access for inosine binding and recognition. d) Representative PAGE analysis of initial (I), flowthrough (FT) and eluate (E) EndoVIPER fractions when tested with glyoxal-treated or unmodified hRNA I targets. e) Densitometric analysis of EndoVIPER efficiency for glyoxal-treated or unmodified hRNA I targets. Values represent mean (n = 2) with standard deviation. Unpaired t-test was performed between untreated and glyoxal pulldowns (** denotes p = 0.004). f) MST analysis of eEndoV binding affinity towards glyoxal treated or unmodified hRNA I. Values represent mean with standard deviation. K_d denotes mean with 95% confidence interval (n = 3).

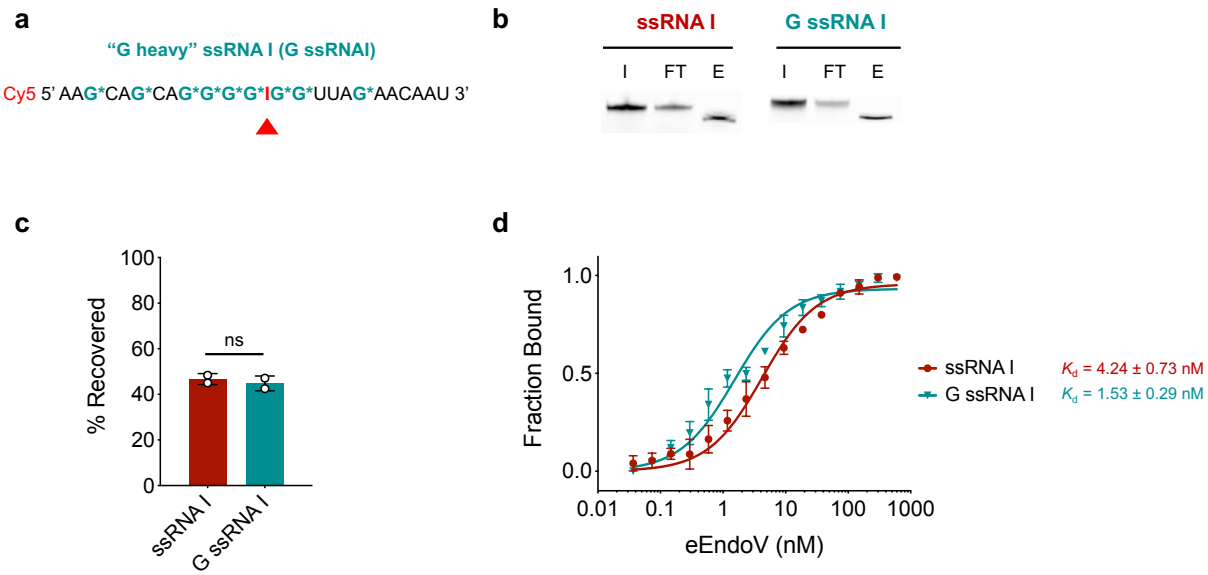


Figure C7. Minimal binding bias when using EndoVIPER on a G heavy RNA substrate. a) "G heavy" oligoribonucleotide test sequence (G ssRNA I) with glyoxalated G nucleotides (G*) highlighted surrounding an inosine site (red arrow). b) Representative PAGE analysis of initial (I), flowthrough (FT) and eluate (E) EndoVIPER fractions when tested with glyoxal-treated ssRNA I or G ssRNA I targets. c) Densitometric analysis of EndoVIPER efficiency for glyoxal-treated targets. Values represent mean (n = 2) with standard deviation. Unpaired t-test was performed between ssRNA I and G ssRNA I pulldowns ("ns" denotes no significant difference). d) MST analysis of eEndoV binding affinity towards glyoxal treated ssRNA I or G ssRNA I targets. Values represent mean with standard deviation. K_d denotes mean with 95% confidence interval (n = 3).

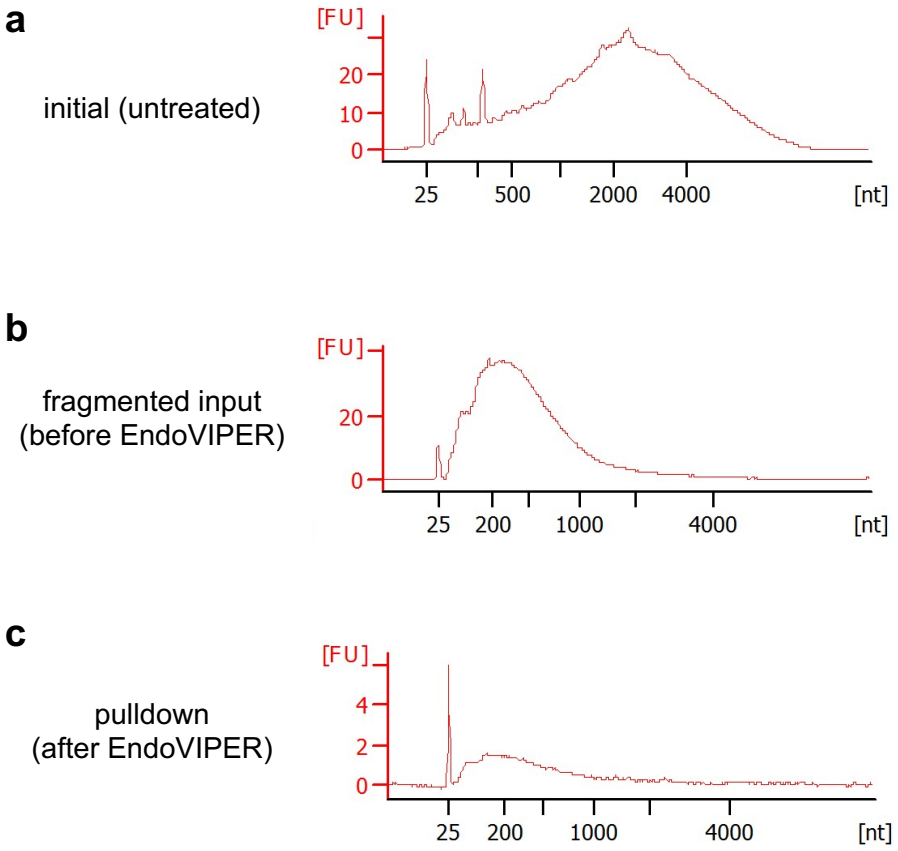


Figure C8. Characterizing RNA size distribution in both input and EndoVIPER pulldowns. Size distribution traces of human brain mRNA used in EndoVIPER and RNA-seq. Initial starting material (untreated) was assessed to confirm material was intact and not degraded. mRNA was then fragmented for ~1 minute at 94 ºC with NEBNext® Magnesium RNA Fragmentation module (New England Biolabs) and assessed for size. Material collected as output from EndoVIPER pulldown was also assessed for size. Traces were obtained on an Agilent 2100 bioanalyzer using RNA 6000 Pico Kit (Agilent).

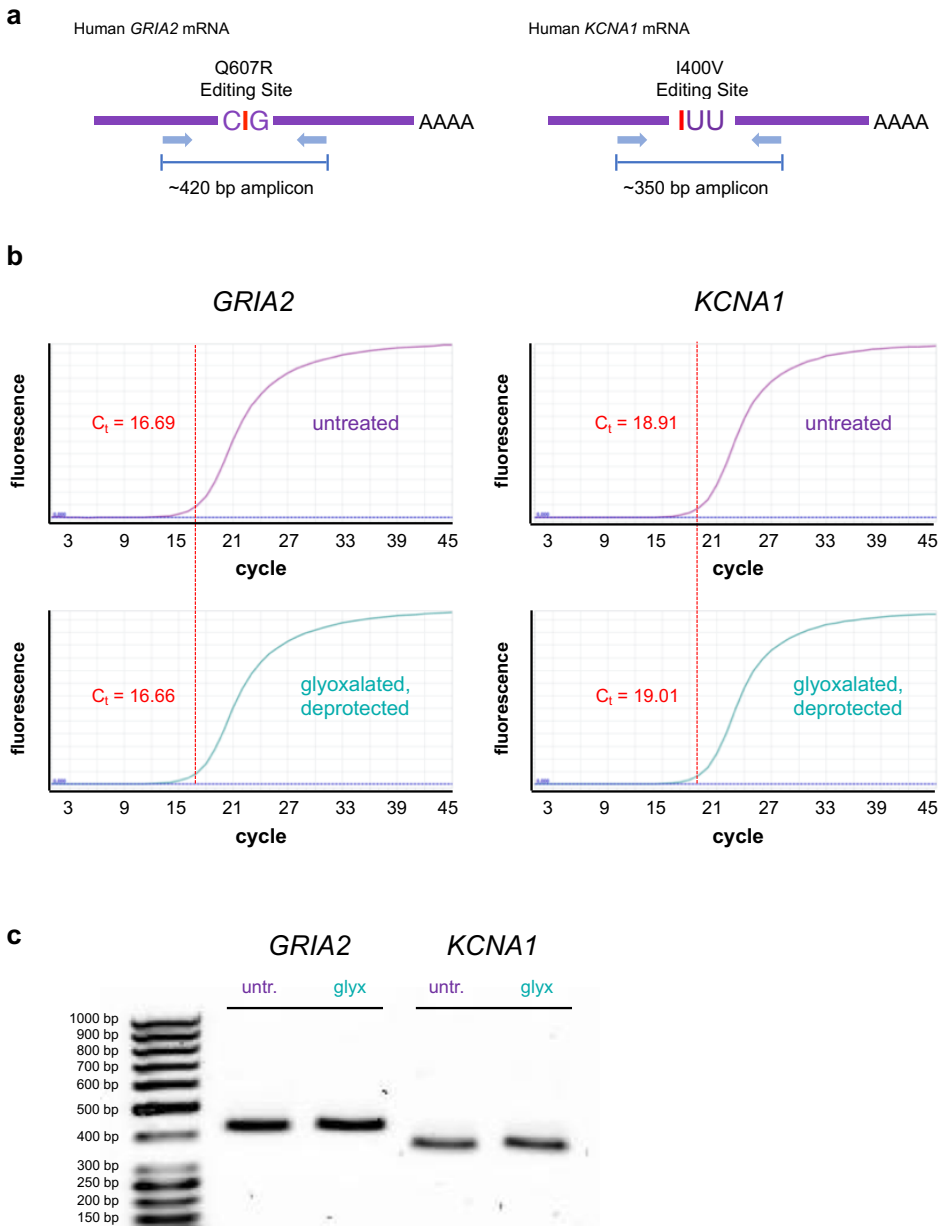
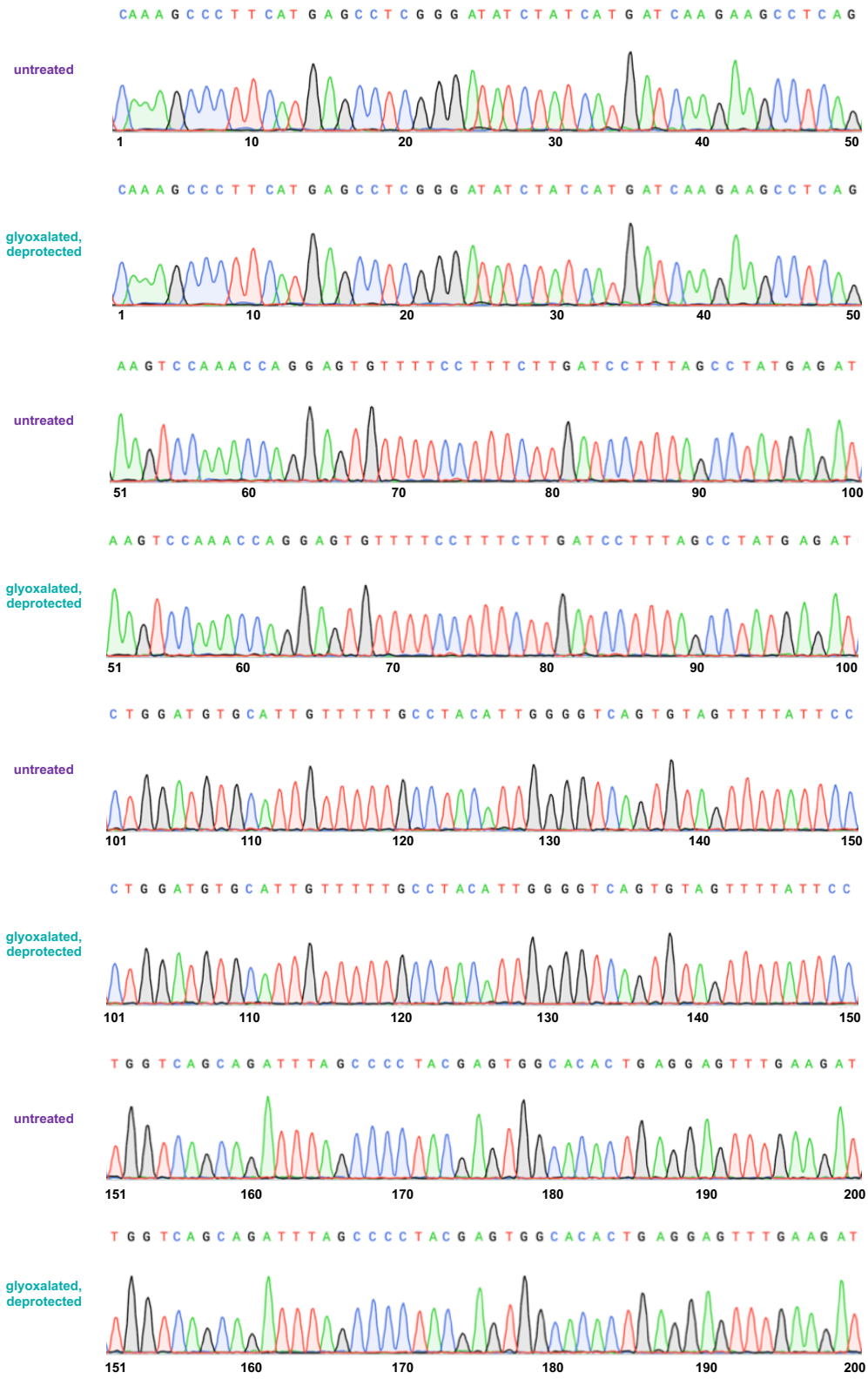
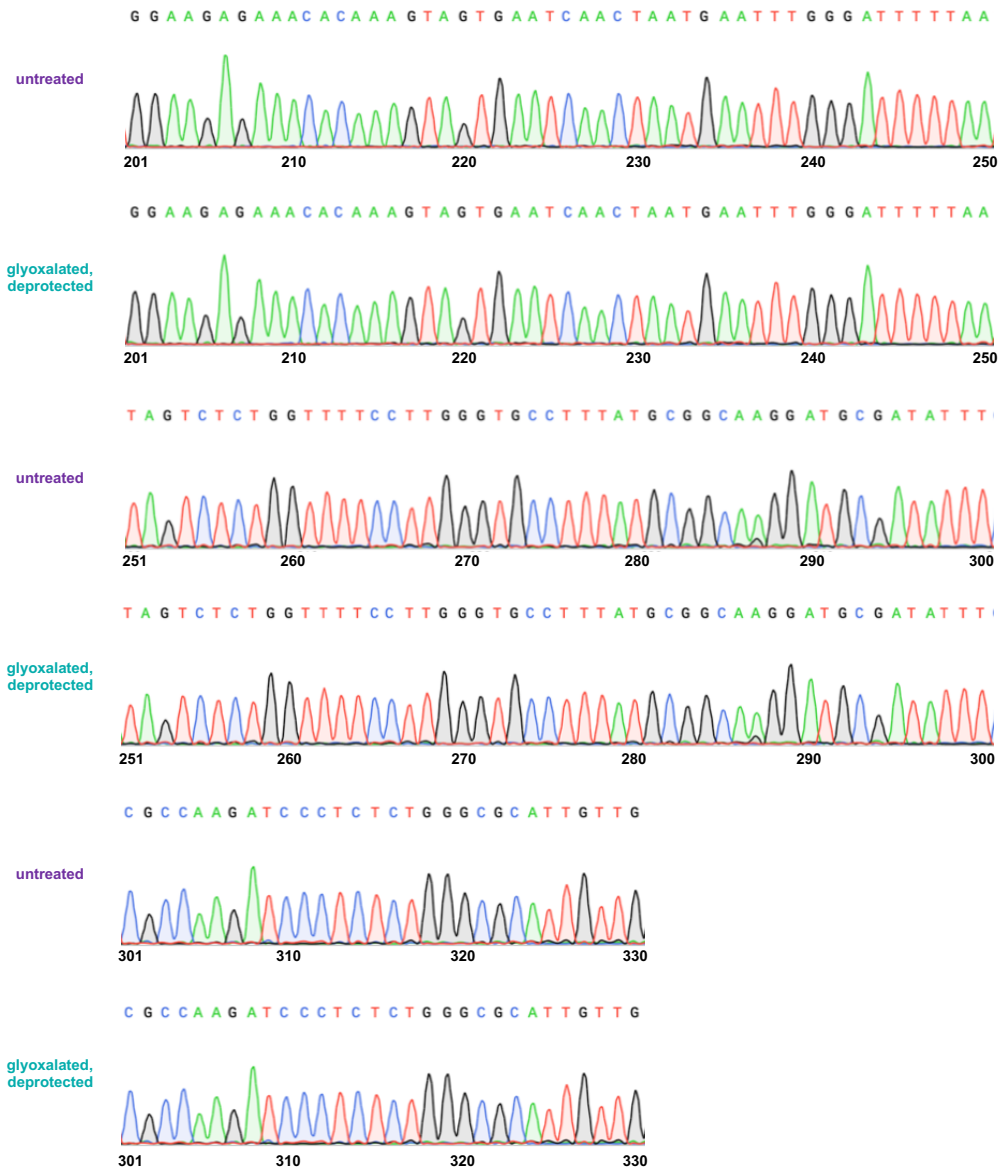


Figure C9. Glyoxal treatment of mRNA is fully reversible and does not effect RT or PCR performance. a) PCR amplicons were generated to target known editing sites in both *GRIA2* and *KCNA1* mRNA transcripts. b) Untreated as well as glyoxal denatured and fully deprotected human brain mRNA was reverse transcribed and PCR amplified using gene specific primers for both *GRIA2* and *KCNA1*. Amplicons were monitored using real-time PCR to determine cycle threshold (C_t) and overall performance. Traces were generated using a LightCycler® 96 instrument (Roche). c) Amplified DNA was purified and analyzed by 1% agarose gel to assess amplicon purity.

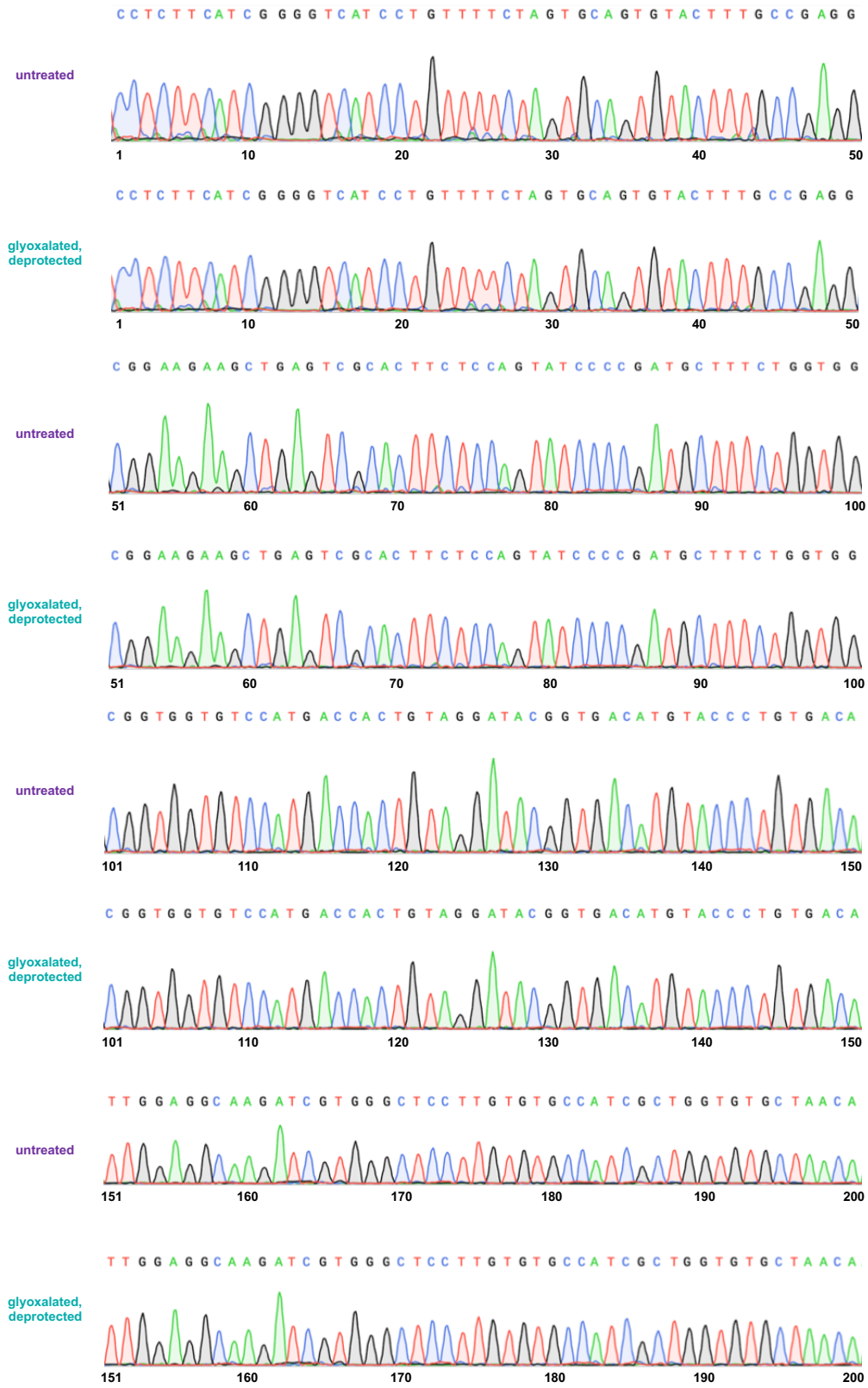
GRIA2


CAAAGCCCTTTCATGAGCCTCGGGGATATCTATCATGATCAAGAAGCCTCAG
untreated
1 10 20 30 40 50
CAAAGCCCTTTCATGAGCCTCGGGGATATCTATCATGATCAAGAAGCCTCAG
glyoxalated, deprotected
1 10 20 30 40 50
AAGTCCAAACCAGGAGTGTTTTCCTTTTCTTGATCCTTTAGCCTATGAGAT
untreated
51 60 70 80 90 100
AAGTCCAAACCAGGAGTGTTTTCCTTTTCTTGATCCTTTAGCCTATGAGAT
glyoxalated, deprotected
51 60 70 80 90 100
CTGGATGTGCATTGTTTTTGCCTACATTGGGGTCAGTGTAGTTTTATTCC
untreated
101 110 120 130 140 150
CTGGATGTGCATTGTTTTTGCCTACATTGGGGTCAGTGTAGTTTTATTCC
glyoxalated, deprotected
101 110 120 130 140 150
TGGTCAGCAGATTTAGCCCCTACGAGTGGCACACTGAGGAGTTTGAAGAT
untreated
151 160 170 180 190 200
TGGTCAGCAGATTTAGCCCCTACGAGTGGCACACTGAGGAGTTTGAAGAT
glyoxalated, deprotected
151 160 170 180 190 200

GRIA2 (continued)



KCNA1


C C T C T T C A T C G G G G T C A T C C T G T T T T C T A G T G C A G T G T A C T T T G C C G A G G
untreated
1 10 20 30 40 50
C C T C T T C A T C G G G G T C A T C C T G T T T T C T A G T G C A G T G T A C T T T G C C G A G G
glyoxalated, deprotected
1 10 20 30 40 50
C G G A A G A A G C T G A G T C G C A C T T C T C C A G T A T C C C C G A T G C T T T C T G G T G G
untreated
51 60 70 80 90 100
C G G A A G A A G C T G A G T C G C A C T T C T C C A G T A T C C C C G A T G C T T T C T G G T G G
glyoxalated, deprotected
51 60 70 80 90 100
C G G T G G T G T C C A T G A C C A C T G T A G G A T A C G G T G A C A T G T A C C C T G T G A C A
untreated
101 110 120 130 140 150
C G G T G G T G T C C A T G A C C A C T G T A G G A T A C G G T G A C A T G T A C C C T G T G A C A
glyoxalated, deprotected
101 110 120 130 140 150
T T G G A G G C A A G A T C G T G G G C T C C T T G T G T G C C A T C G C T G G T G T G C T A A C A
untreated
151 160 170 180 190 200
T T G G A G G C A A G A T C G T G G G C T C C T T G T G T G C C A T C G C T G G T G T G C T A A C A
glyoxalated, deprotected
151 160 170 180 190 200

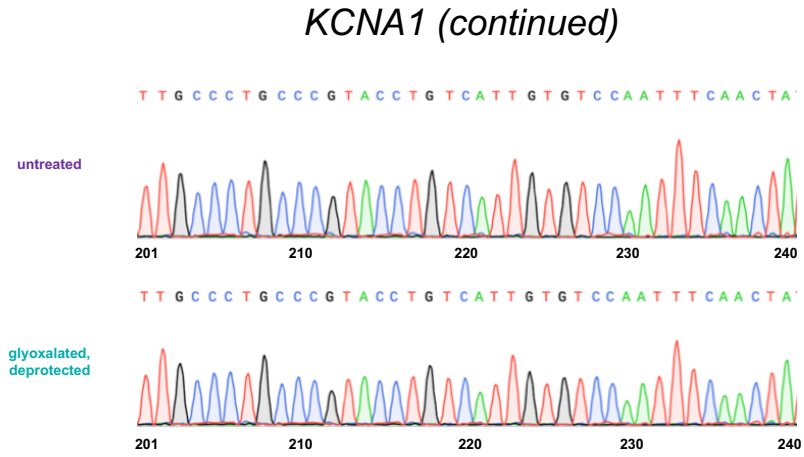


Figure C10. Glyoxal treatment of mRNA is reversible and does not affect sequencing performance. Purified RT-PCR amplicons for both *GRIA2* and *KCNA1* were subjected to Sanger sequencing. Representative traces were visualized and compared using SnapGene Viewer.

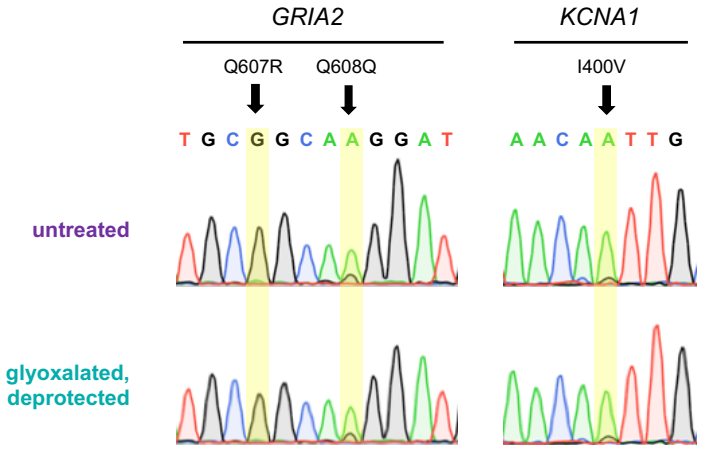


Figure C11. Glyoxal treatment of mRNA is reversible and does not affect detection of A-to-I editing. Known protein recoding editing sites (yellow) in both *GRIA2* and *KCNA1* mRNA transcripts are visible in Sanger traces regardless of glyoxal treatment and removal.

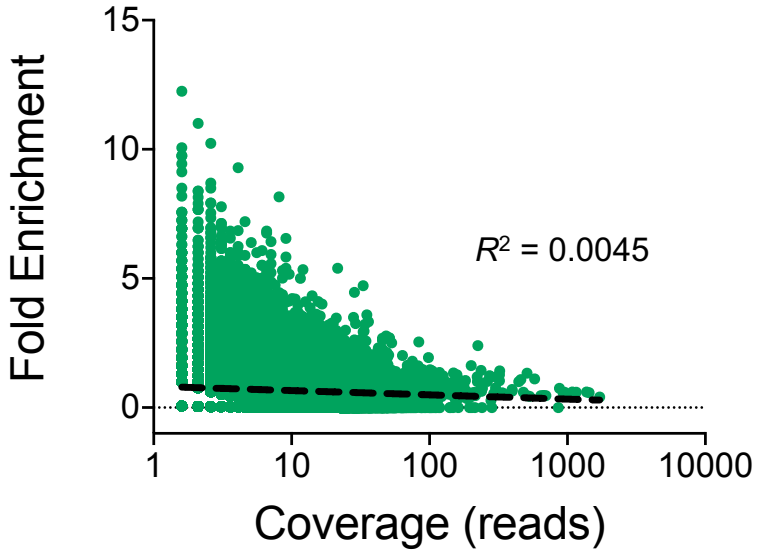


Figure C12. Semi-log scatter plot of transcript abundance vs fold enrichment. All enriched sites found in RNA-seq replicates (n = 31,309 sites) were plotted against fold enrichment scores. Semi-log regression (black dashed line) was calculated using GraphPad Prism 8.

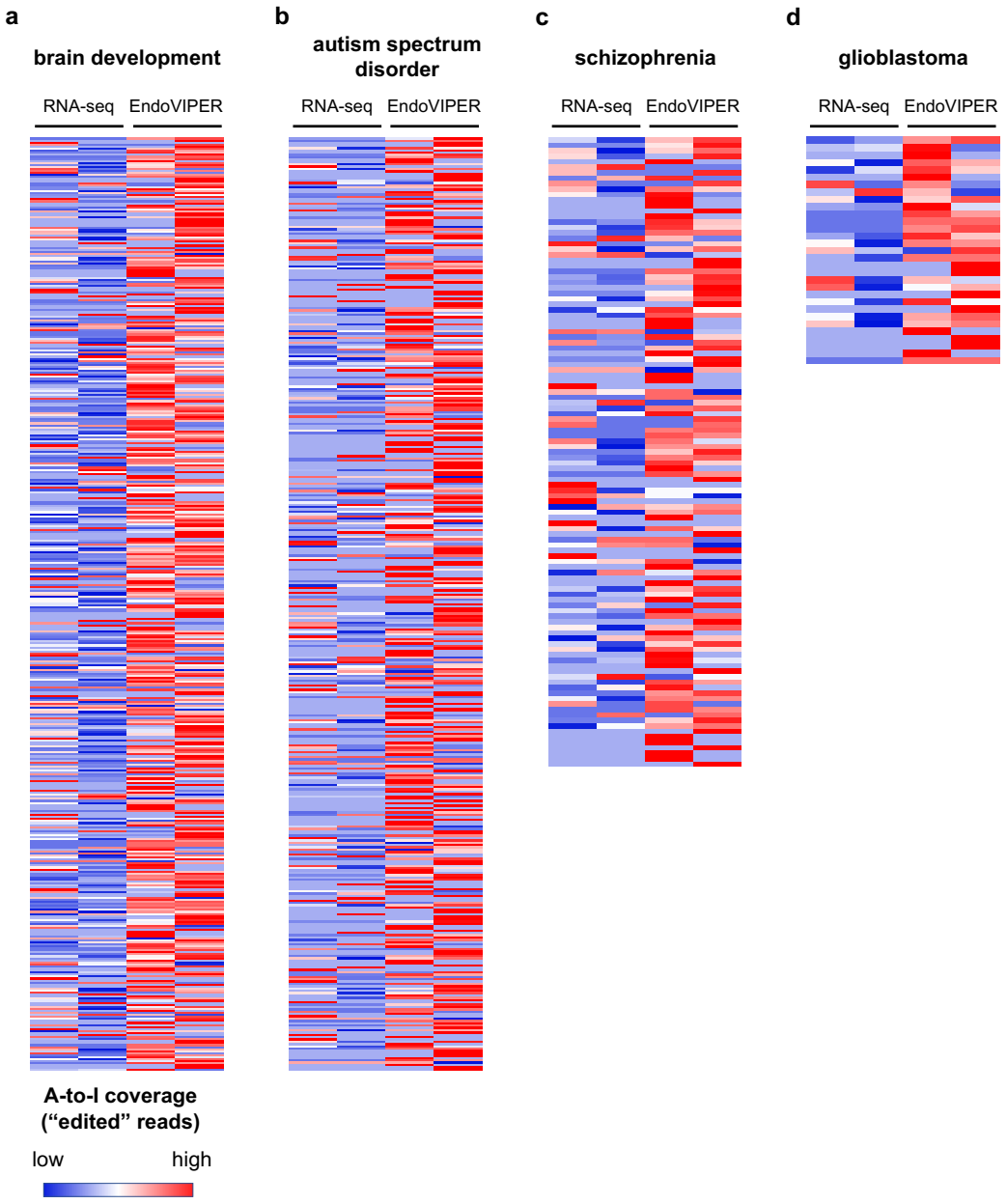


Figure C13. EndoVIPER-seq enhances detection of edited transcript isoforms. Detection of "edited" reads in upregulated RNA editing sites of interest in a) brain development (462 sites), b) autism spectrum disorder (403 sites), c) schizophrenia (115 sites) and d) protein recoding events in glioblastoma (31 sites). Edited reads were calculated from total read coverage and calculated editing rate at each site. Heatmap columns display both replicate datasets for RNA-seq and EndoVIPER samples, and each row denotes an individual site and scaled to illustrates low (blue) and high (red) read coverage between groups.

Appendix D:

Omitted Data from Chapter 5

Direct Immunodetection of Global A-to-I RNA Editing Activity with a Chemiluminescent Bioassay

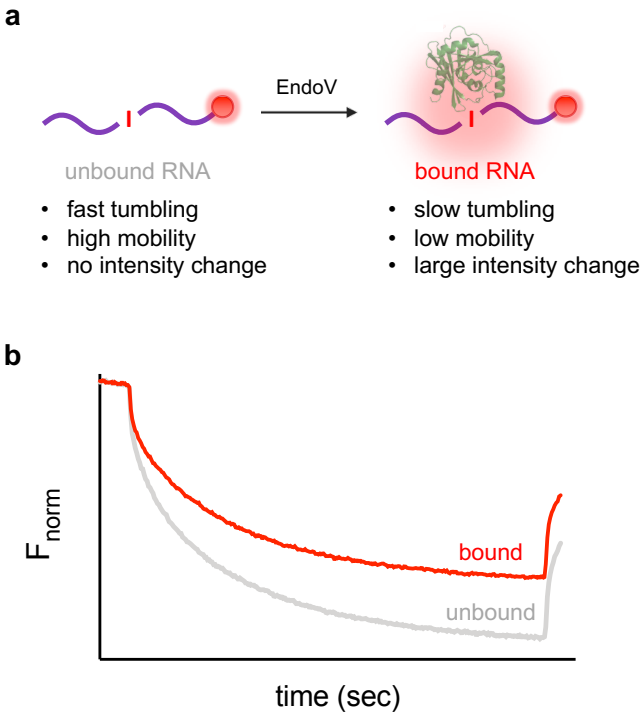


Figure D1. Microscale thermophoresis (MST) can detect binding of EndoV to inosine-containing RNA. a) Overall principle of MST detection, wherein a fluorophore-tagged ligand (RNA containing inosine) undergoes changes in mobility, solvation, and fluorescent intensity upon binding to a larger protein (EndoV). b) Representative MST data in the unbound (gray) and bound state (red). Binding events can be monitored and quantified as changes in normalized fluorescence (F_{norm}) when a temperature gradient is applied to the sample over time.

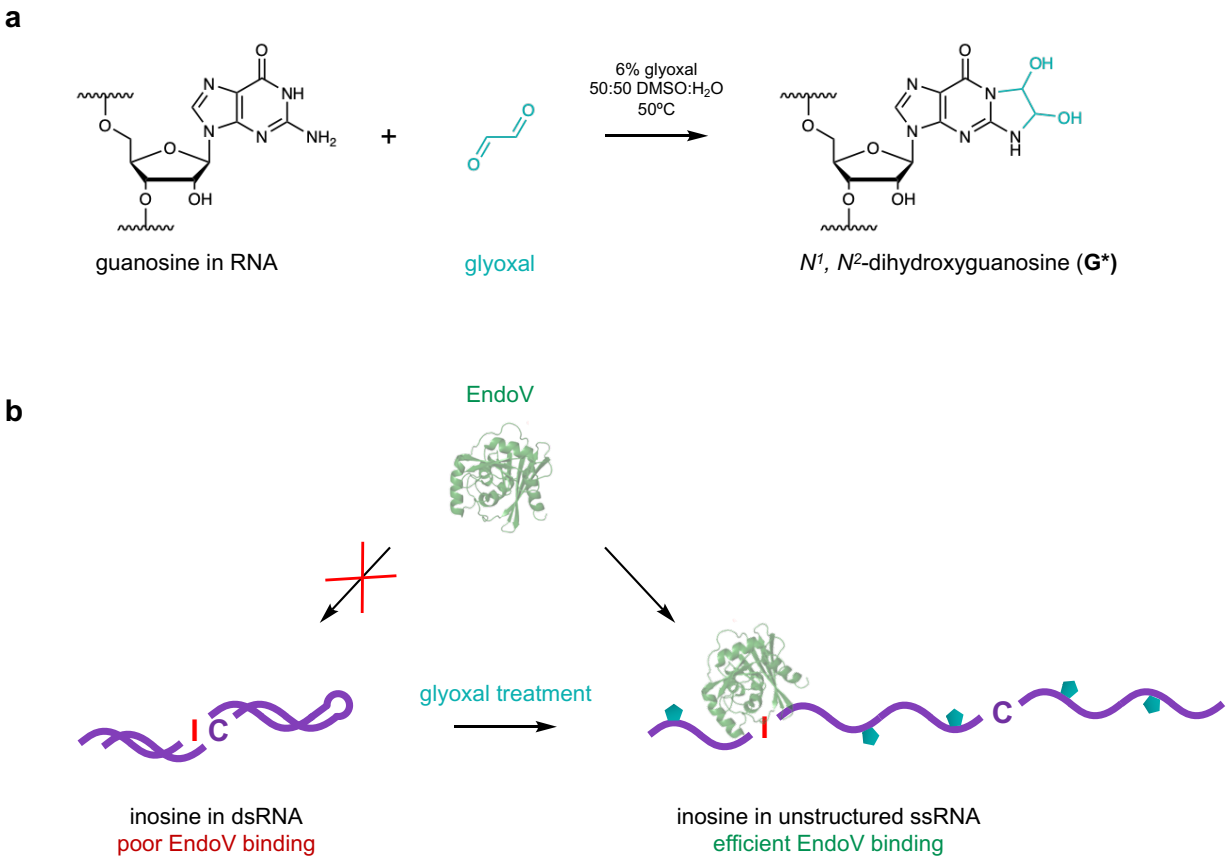


Figure D2. Glyoxal denaturation eliminates secondary structure in RNA and enhances EndoV binding. a) Schematic and reaction conditions for covalent glyoxal addition onto the Watson–Crick–Franklin face of guanosine residues, forming a N^1,N^2-dihydroxyguanosine adduct. b) EndoV binds poorly to inosine residing in highly structured dsRNA. Glyoxal potently disrupts RNA secondary structure and does not react with inosine, enabling efficient EndoV binding.

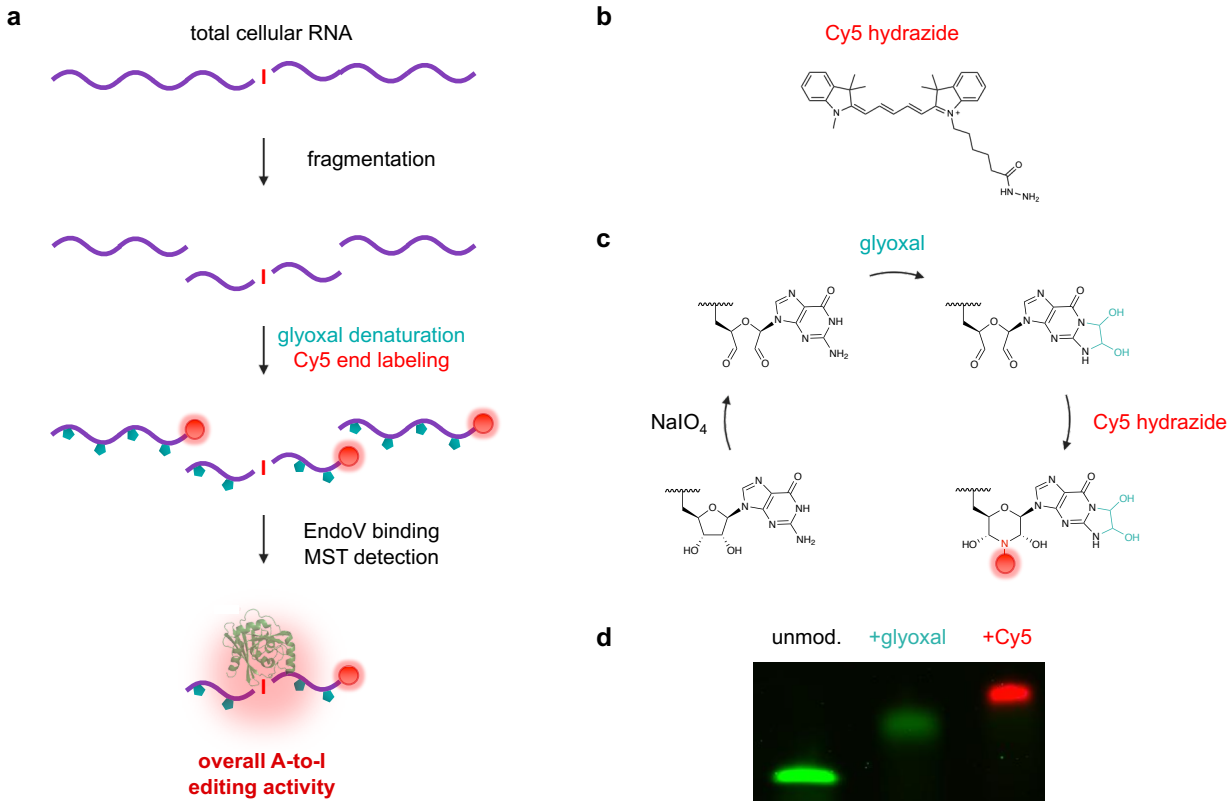


Figure D3. Proposed initial workflow for detecting global A-to-I editing activity in cellular RNA using MST. a) To measure inosine content, cellular RNA is first fragmented into smaller strands, followed by sequential denaturation and fluorescent labeling. To label each strand, b) Cy5 hydrazide is used in combination with c) oxidation of 3' OH groups using sodium metaperiodate ($NaIO_4$), glyoxal denaturation, and coupling of Cy5 hydrazide onto the 3' dialdehyde. d) 20% denaturing PAGE analysis illustrating an unmodified (unmodified) test RNA strand sequentially treated with $NaIO_4$, glyoxal and Cy5-hydrazide.

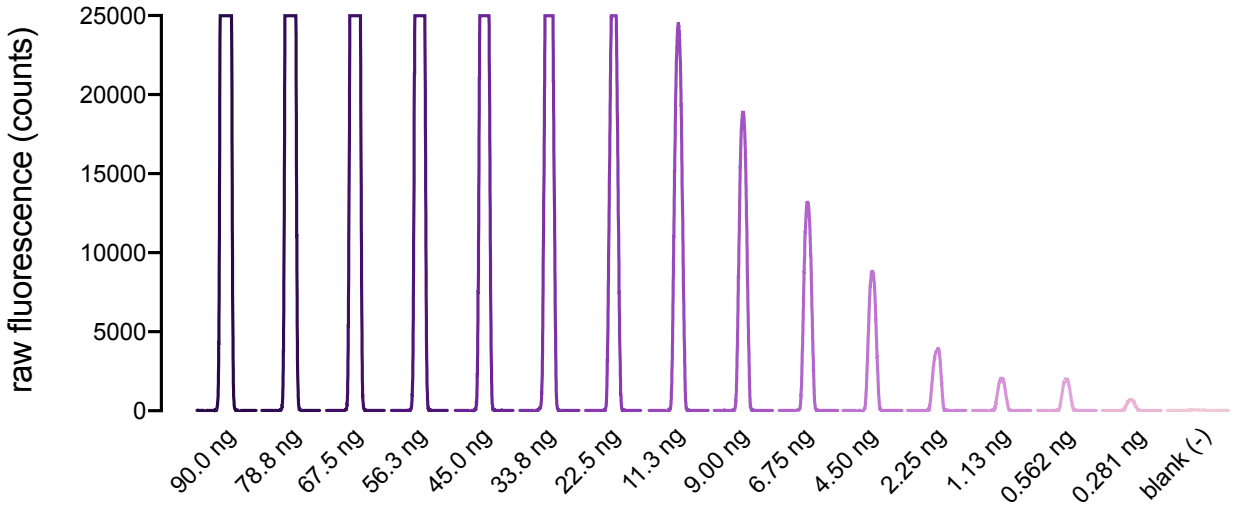


Figure D4. Titrating Cy5-labeled RNA for MST detection. Increasing total amounts of Cy5-labeled RNA l (ng, x-axis) was added to 10 µl of 1X buffer (19 mM Tris, 137 mM NaCl, 3 mM KCl, 5 mM CaCl2, 15 µM EDTA, 150 µM DTT, 0.025% Triton X-100, 30 µg/mL BSA, 7% glycerol, pH 7.4.). Samples were then loaded into standard MST capillaries (Nanotemper) and scanned using a Nanotemper Monolith instrument at 5% laser intensity power. Peaks represent raw fluorescence values and capillary shape scans obtained directly from Monolith software. Blank sample contains only buffer.

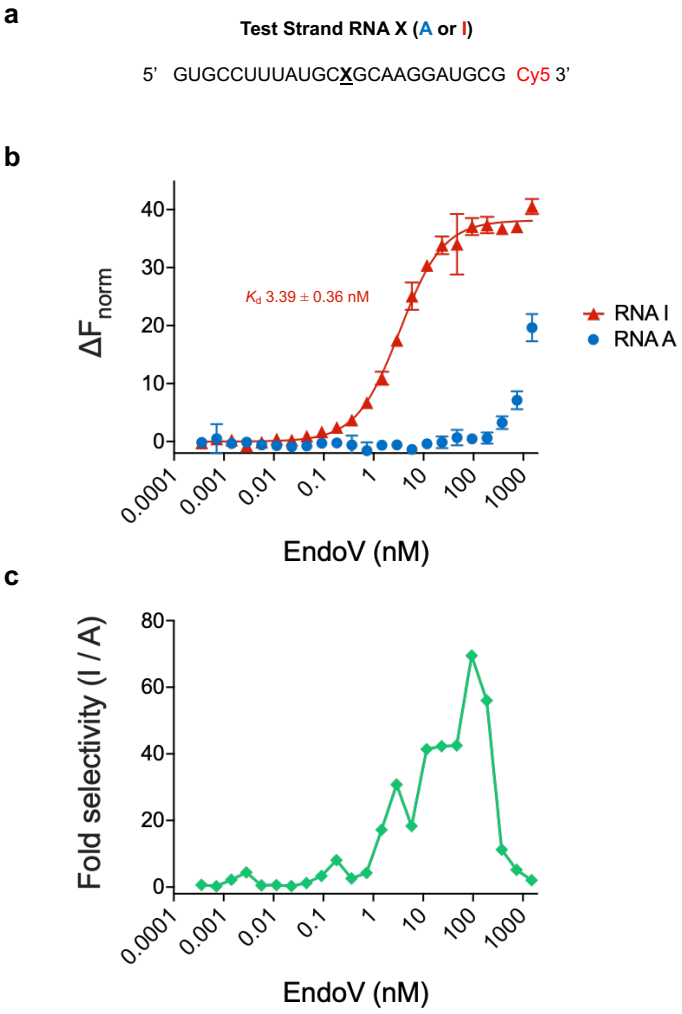
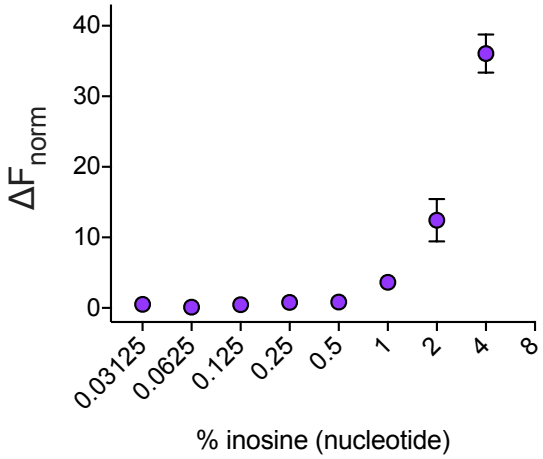


Figure D5. Measuring selectivity of EndoV for inosine-containing RNAs. a) Sequence of test RNA strands with 3' Cy5 label. X indicates position of A or I. b) MST analysis of RNA I and RNA A binding to increasing amounts of EndoV. ΔF_{norm} is calculated by subtracting the blank (no EndoV) raw normalized fluorescence value (F_{norm}) from the from each test sample. Values represent mean (n = 3) and error bars denote S.D. c) Fold selectivity is calculated as the average ΔF_{norm} of I/A for each EndoV concentration.



RNA I : RNA A (ratio)	percent inosine
1:0	4.17%
1:1	2.08%
1:2	1.04%
1:4	0.52%
1:8	0.26%
1:16	0.13%
1:32	0.06%
1:64	0.03%

Figure D6. Measuring sensitivity of EndoV for detecting decreasing amounts of inosine by MST. RNA A and RNA I (10 ng total) were mixed in different ratios according to the table, combined with 100 nM EndoV, and analyzed with MST. ΔF_{norm} is calculated by subtracting the blank (no EndoV) raw normalized fluorescence value (F_{norm}) from the from each test sample. Values represent mean (n = 3) and error bars denote S.D.

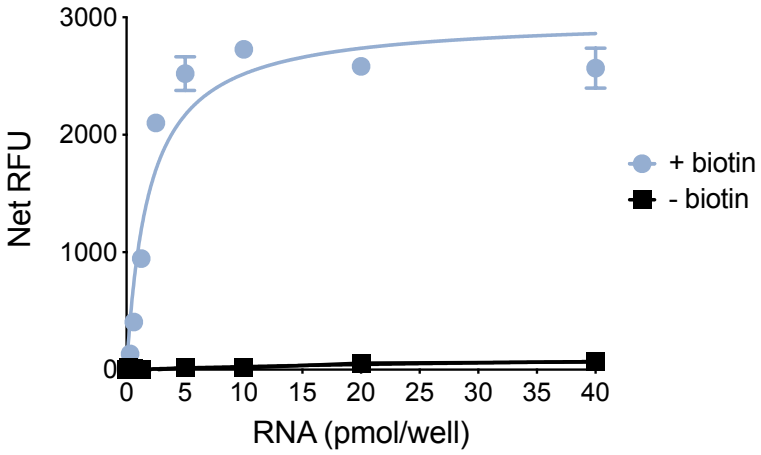


Figure D7. Estimated RNA immobilization capacity onto streptavidin-coated plates. A 24 nt RNA I strand was Cy5 labeled, glyoxal denatured, and treated with and without biotin hydrazide labeling. Increasing amounts (0 – 40 pmol) of strand were loaded to each well of a 96-well streptavidin coated plate and incubated at room temperature for 1 hour with gentle shaking. Each well was then washed three times and fluorescence was measured using a BioTek Cytation 5 plate reader. Values represent net relative fluorescence units (RFU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA). Values represent mean with standard deviation (n = 3).

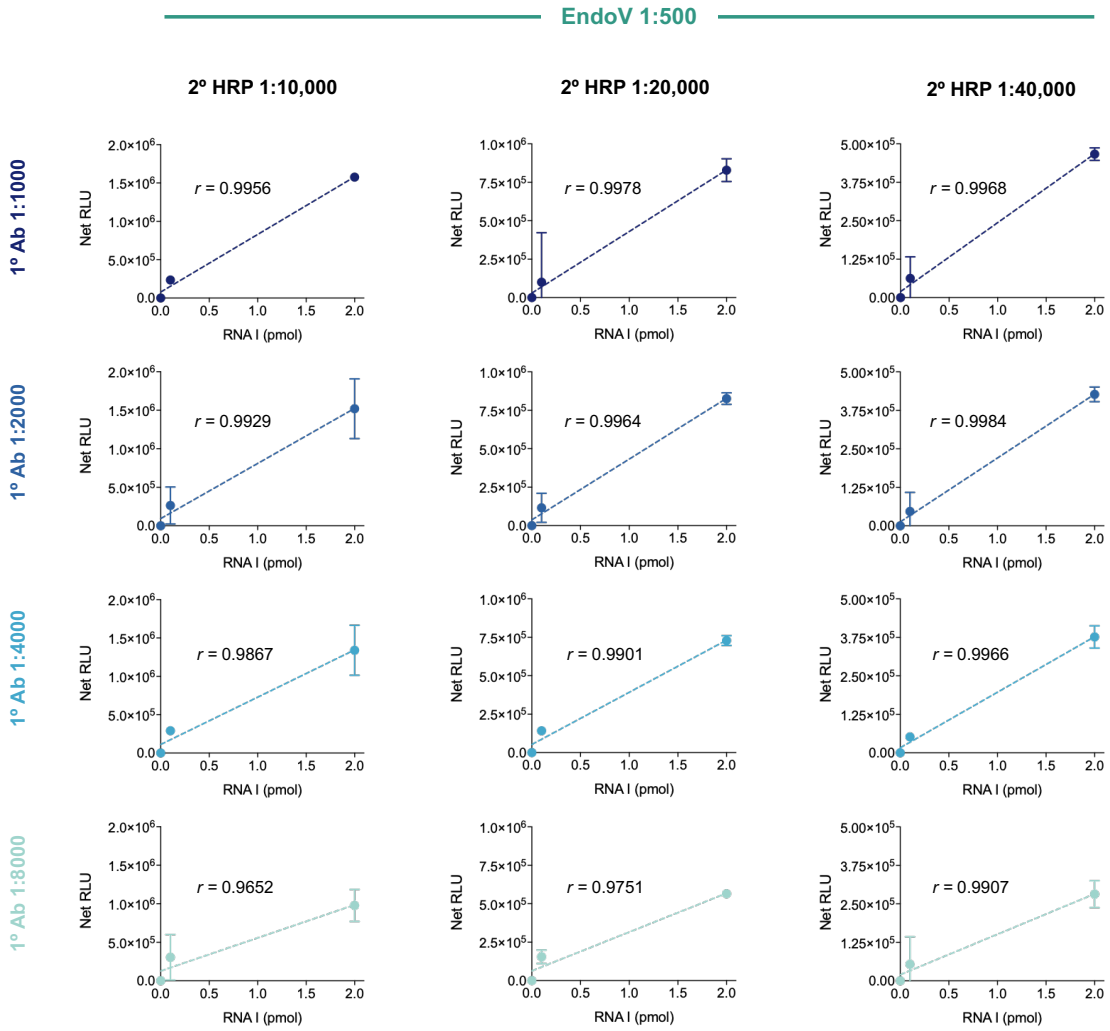


Figure D8. Conditional screen to optimize EndoVLISA linearity using EndoV-MBP at 1:500. In duplicate, different amounts of biotinylated, glyoxal-denatured RNA I strand (0, 0.1 and 2 pmol) were immobilized in white streptavidin-coated 96-well plates and then incubated with a solution of EndoV-MBP fusion protein at 1:500 in 1X buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.05% Tw 20, pH 7.5). Different combinations of MBP-targeting 1º antibody (1:1000 – 1:8000) and 2º HRP-conjugated antibody (1:10,000 – 1:40,000) were then used to probe the plate. SuperSignal™ ELISA Pico chemiluminescent substrate (Thermo Fisher Scientific) was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Data points represent mean (n = 2) and error bars denote 95% confidence intervals. Dashed line indicates simple linear regression estimation, and pearson correlation coefficients (*r*) were calculated in GraphPad Prism.

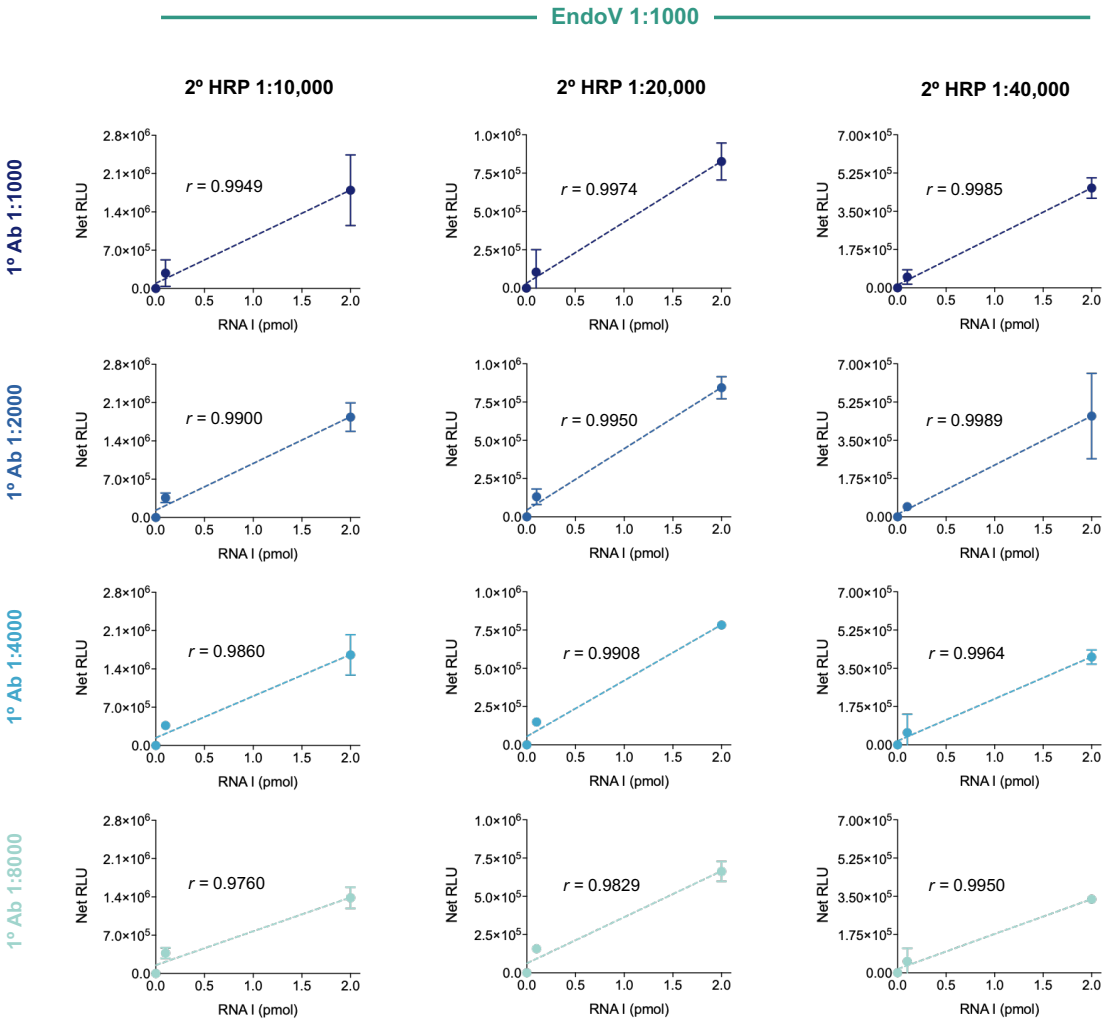


Figure D9. Conditional screen to optimize EndoVLISA linearity using EndoV-MBP at 1:1000. In duplicate, different amounts of biotinylated, glyoxal-denatured RNA I strand (0, 0.1 and 2 pmol) were immobilized in white streptavidin-coated 96-well plates and then incubated with a solution of EndoV-MBP fusion protein at 1:1000 in 1X buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.05% Tw 20, pH 7.5). Different combinations of MBP-targeting 1º antibody (1:1000 – 1:8000) and 2º HRP-conjugated antibody (1:10,000 – 1:40,000) were then used to probe the plate. SuperSignal™ ELISA Pico chemiluminescent substrate (Thermo Fisher Scientific) was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Data points represent mean (n = 2) and error bars denote 95% confidence intervals. Dashed line indicates simple linear regression estimation, and pearson correlation coefficients (*r*) were calculated in GraphPad Prism.

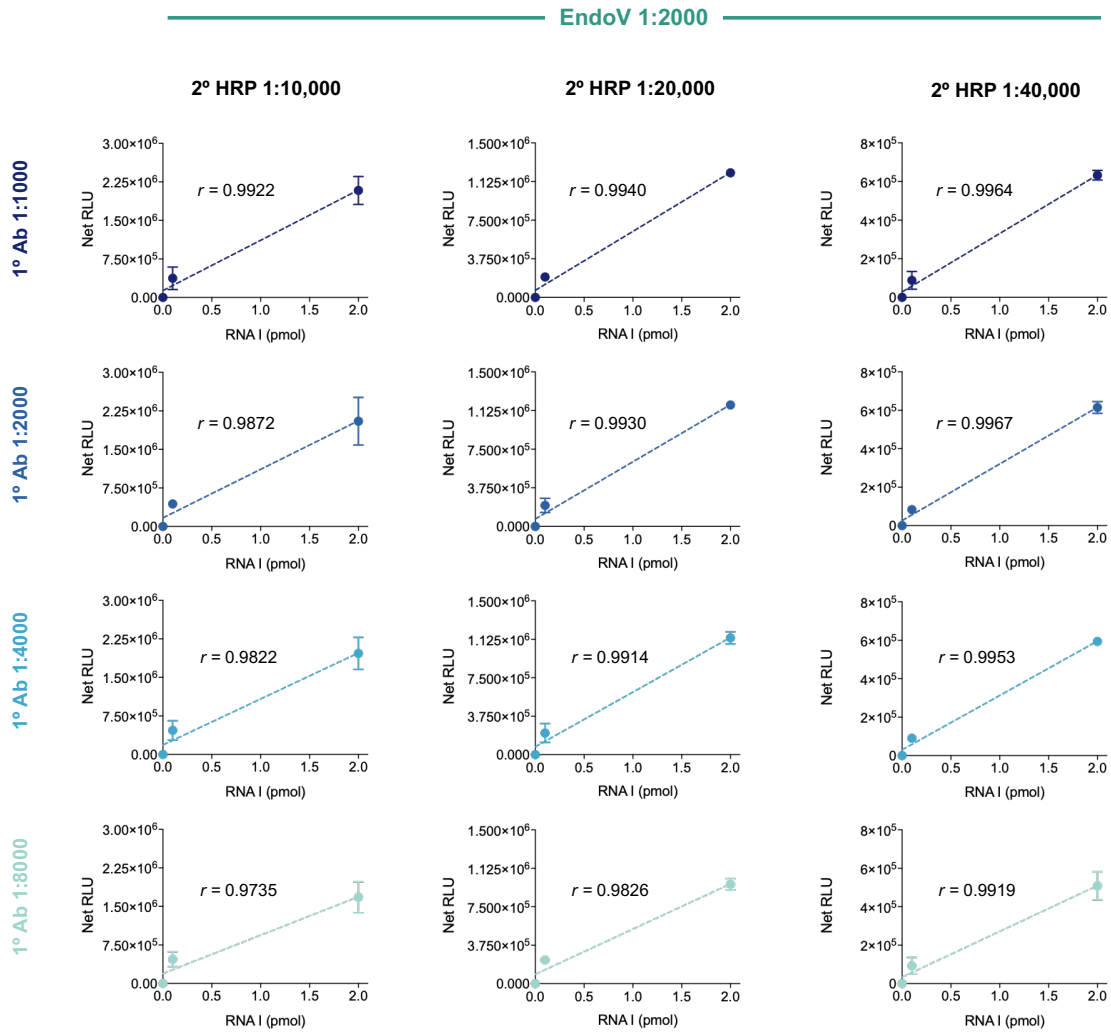


Figure D10. Conditional screen to optimize EndoVLISA linearity using EndoV-MBP at 1:2000. In duplicate, different amounts of biotinylated, glyoxal-denatured RNA I strand (0, 0.1 and 2 pmol) were immobilized in white streptavidin-coated 96-well plates and then incubated with a solution of EndoV-MBP fusion protein at 1:2000 in 1X buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.05% Tw 20, pH 7.5). Different combinations of MBP-targeting 1º antibody (1:1000 – 1:8000) and 2º HRP-conjugated antibody (1:10,000 – 1:40,000) were then used to probe the plate. SuperSignal™ ELISA Pico chemiluminescent substrate (Thermo Fisher Scientific) was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Data points represent mean (n = 2) and error bars denote 95% confidence intervals. Dashed line indicates simple linear regression estimation, and pearson correlation coefficients (*r*) were calculated in GraphPad Prism.

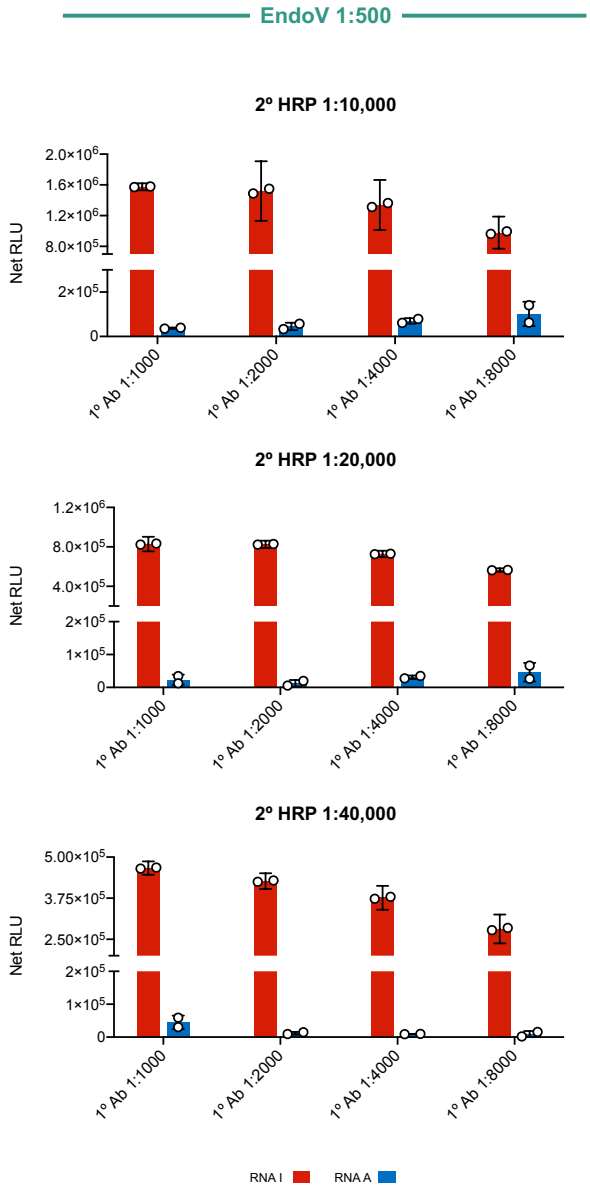


Figure D11. Conditional screen to optimize EndoVLISA selectivity using EndoV-MBP at 1:500. In duplicate, 2 pmol of biotinylated, glyoxal-denatured RNA A (blue) or RNA I (red) strand were immobilized in white streptavidin-coated 96-well plates and then incubated with a solution of EndoV-MBP fusion protein at 1:500 in 1X buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.05% Tw 20, pH 7.5). Different combinations of MBP-targeting 1º antibody (1:1000 – 1:8000) and 2º HRP-conjugated antibody (1:10,000 – 1:40,000) were then used to probe the plate. SuperSignal™ ELISA Pico chemiluminescent substrate (Thermo Fisher Scientific) was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Data points represent mean (n = 2) and error bars denote 95% confidence intervals.

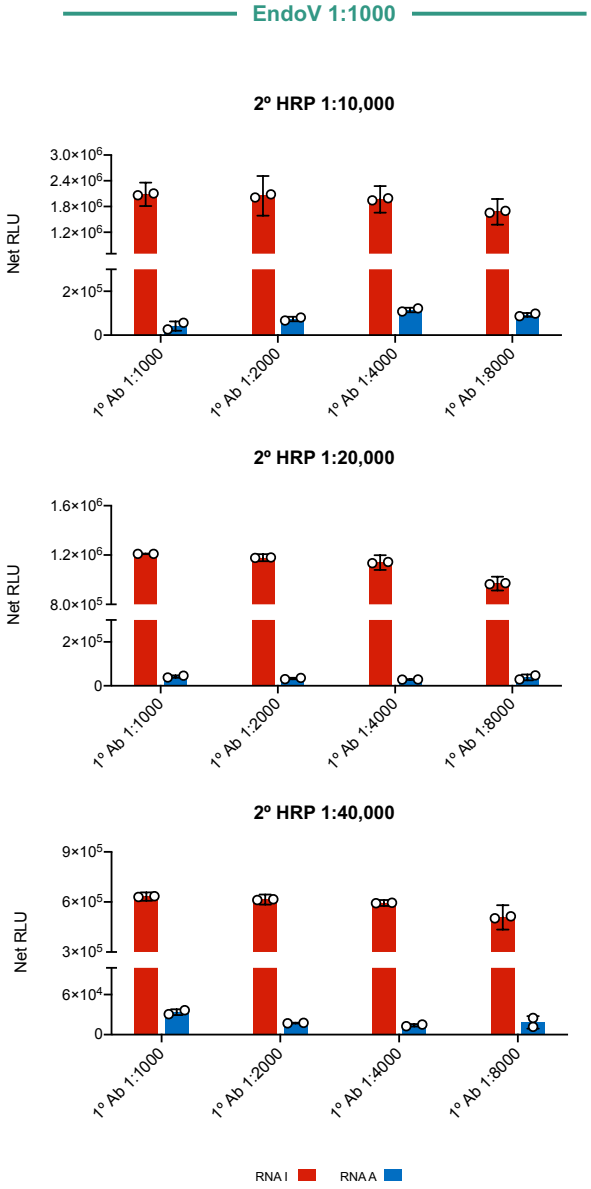


Figure D12. Conditional screen to optimize EndoVLISA selectivity using EndoV-MBP at 1:1000. In duplicate, 2 pmol of biotinylated, glyoxal-denatured RNA A (blue) or RNA I (red) strand were immobilized in white streptavidin-coated 96-well plates and then incubated with a solution of EndoV-MBP fusion protein at 1:1000 in 1X buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.05% Tw 20, pH 7.5). Different combinations of MBP-targeting 1º antibody (1:1000 – 1:8000) and 2º HRP-conjugated antibody (1:10,000 – 1:40,000) were then used to probe the plate. SuperSignal™ ELISA Pico chemiluminescent substrate (Thermo Fisher Scientific) was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Datapoints represent mean (n = 2) and error bars denote 95% confidence intervals.

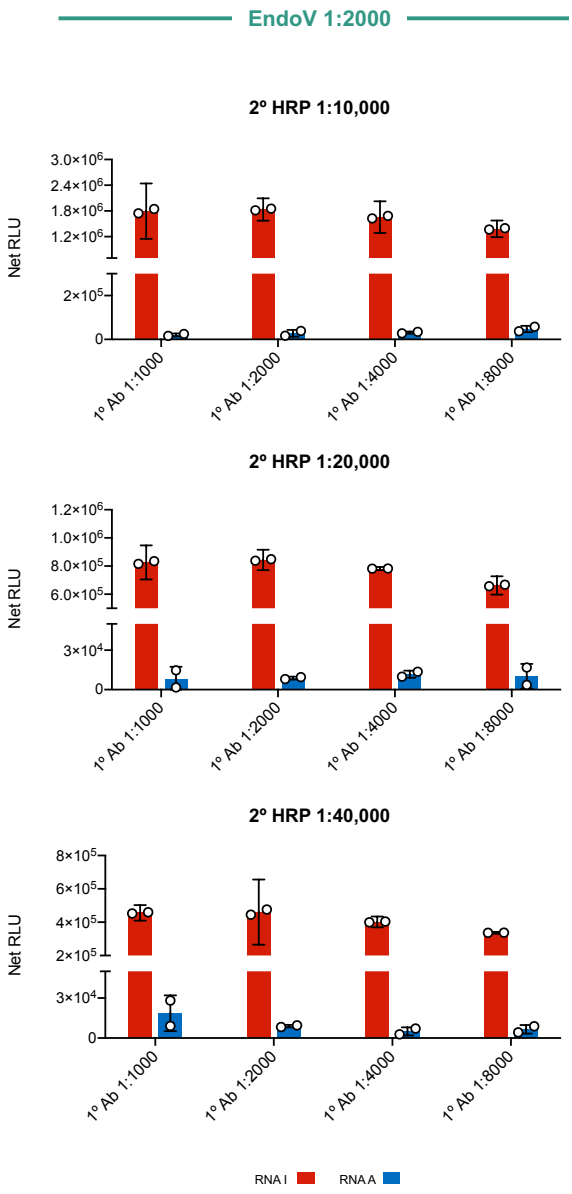


Figure D13. Conditional screen to optimize EndoVLISA selectivity using EndoV-MBP at 1:2000. In duplicate, 2 pmol of biotinylated, glyoxal-denatured RNA A (blue) or RNA I (red) strand were immobilized in white streptavidin-coated 96-well plates and then incubated with a solution of EndoV-MBP fusion protein at 1:2000 in 1X buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.05% Tw 20, pH 7.5). Different combinations of MBP-targeting 1º antibody (1:1000 – 1:8000) and 2º HRP-conjugated antibody (1:10,000 – 1:40,000) were then used to probe the plate. SuperSignal™ ELISA Pico chemiluminescent substrate (Thermo Fisher Scientific) was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Data points represent mean (n = 2) and error bars denote 95% confidence intervals.

EndoV	1º Ab	2º HRP	Linearity (*p* value)	Fold-selectivity
1:1000	1:1000	1:20K	0.046	101.467
1:1000	1:4000	1:40K	0.0538	80.568
1:500	1:1000	1:20K	0.0423	75.596
1:1000	1:2000	1:40K	0.0303	51.554
1:500	1:2000	1:40K	0.0356	35.459
1:1000	1:1000	1:40K	0.0347	24.508

Table D1. Summary of top-performing EndoVLISA component combinations.

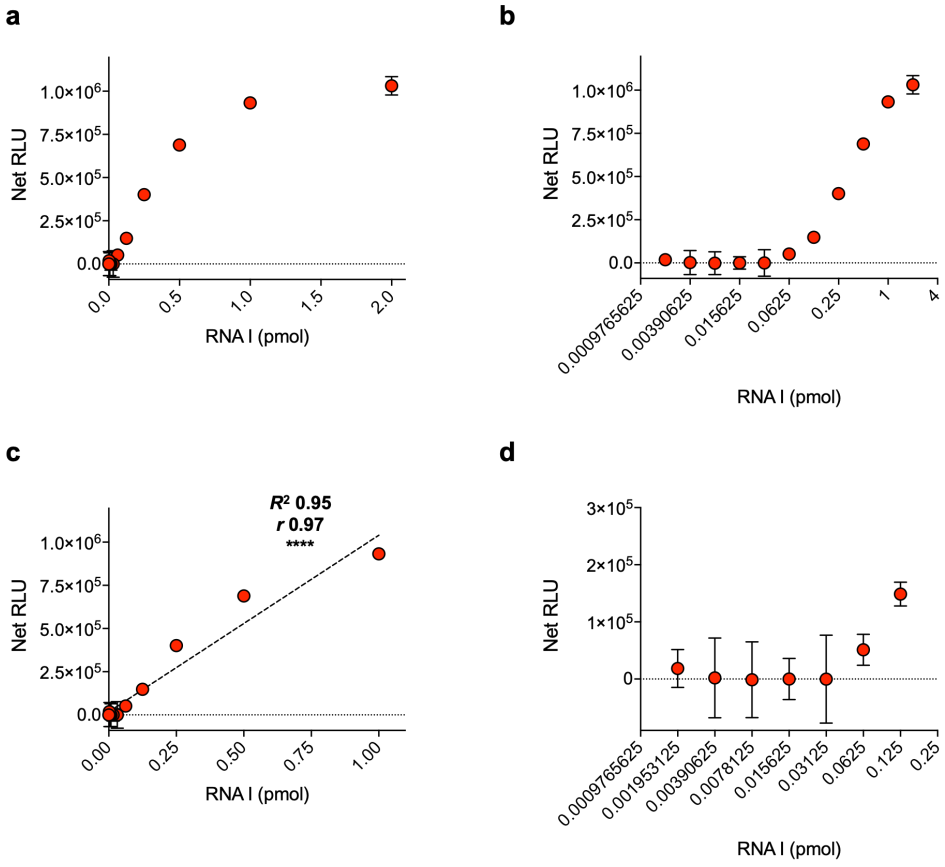


Figure D14. Linearity and sensitivity in detecting inosine. Decreasing amounts of biotinylated, glyoxal-denatured RNA I strand were immobilized in white streptavidin-coated 96-well plates and then incubated with a solution of EndoV-MBP fusion protein at 1:1000 in 1X buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.05% Tw 20, pH 7.5). Wells were then probed with MBP-targeting 1º antibody (1:1000) and 2º HRP-conjugated antibody (1:20,000), and then SuperSignal™ ELISA Pico chemiluminescent substrate (Thermo Fisher Scientific) was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Full RNA I standard curve in a) linear and b) log_2 scale. c) Partial standard curve (RNA I <1 pmol) in linear scale. Linear regression (black dashed line), R2 and pearson (*r*) correlation (**** denotes $p < 0.0001$) were computed in Prism. d) Partial standard curve (RNA I < 0.125 pmol) in log_2 scale to estimate limit of detection above blank. All values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting blank wells (no RNA). Data points represent mean (n = 3) and error bars denote 95% confidence intervals.

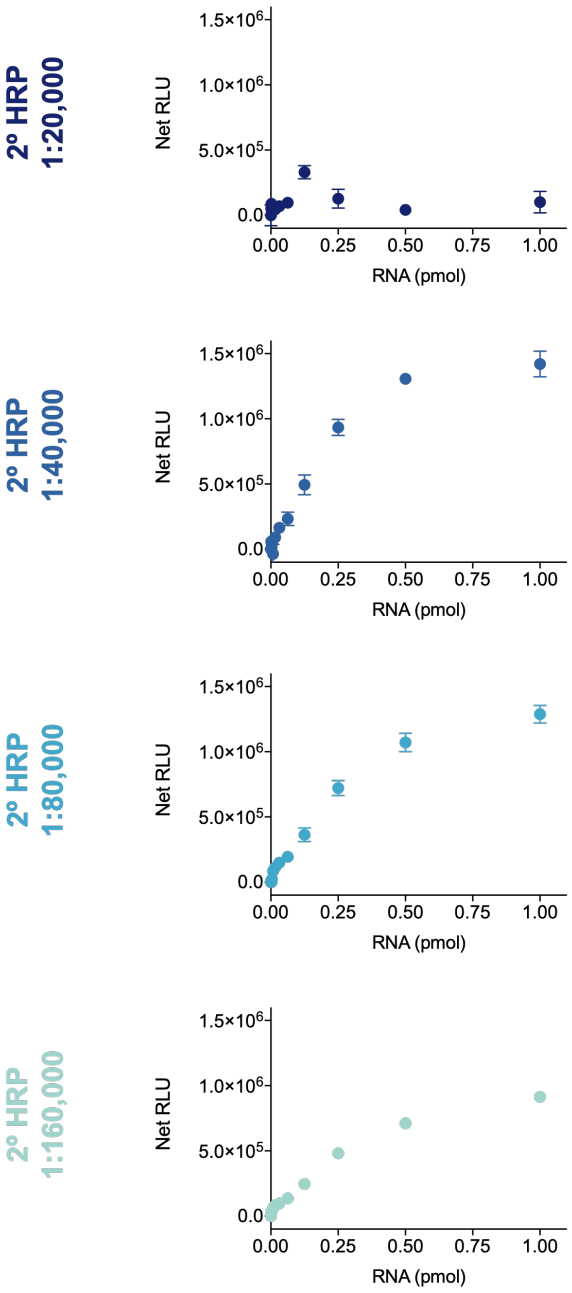


Figure D15. Testing 2º antibody-HRP concentrations with a high-sensitivity SuperSignal™ West Atto chemiluminescent substrate. Decreasing amounts of biotinylated, glyoxal-denatured RNA I strand were immobilized in white streptavidin-coated 96-well plates and then incubated with a solution of EndoV-MBP fusion protein at 1:1000 in 1X buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.05% Tw 20, pH 7.5). Wells were then probed with MBP-targeting 1º antibody (1:1000) and different amounts of 2º HRP-conjugated antibody (1:20,000-160,000), and then SuperSignal™ West Atto chemiluminescent substrate (Thermo Fisher Scientific) was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Datapoints represent mean (n = 3) and error bars denote 95% confidence intervals.

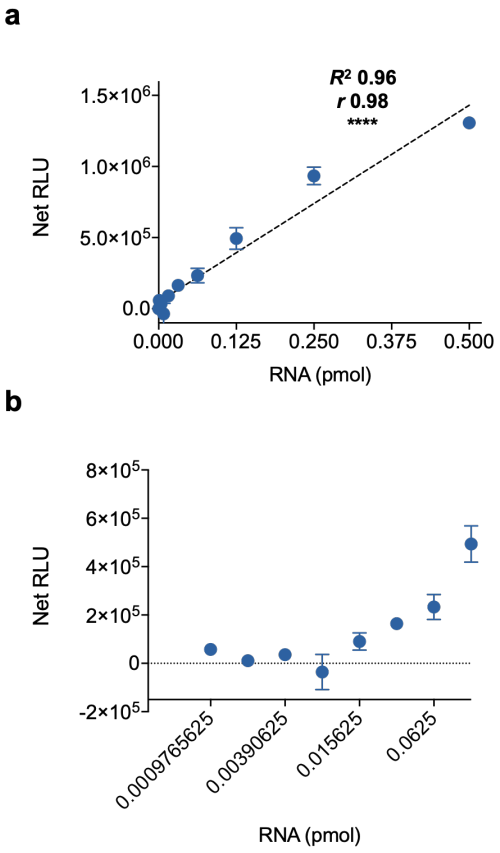


Figure D16. Linearity and sensitivity of EndoVLISA with 1:40,000 2º antibody-HRP and high-sensitivity SuperSignal™ West Atto chemiluminescent substrate. a) Full RNA I standard curve in linear scale. Linear regression (black dashed line), R^2 and pearson (r) correlation (**** denotes $p < 0.0001$) were computed in Prism. b) Partial standard curve (RNA I < 0.125 pmol) in $\log_2$ scale to estimate limit of detection above blank. All values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting blank wells (no RNA). Data points represent mean (n = 3) and error bars denote 95% confidence intervals.

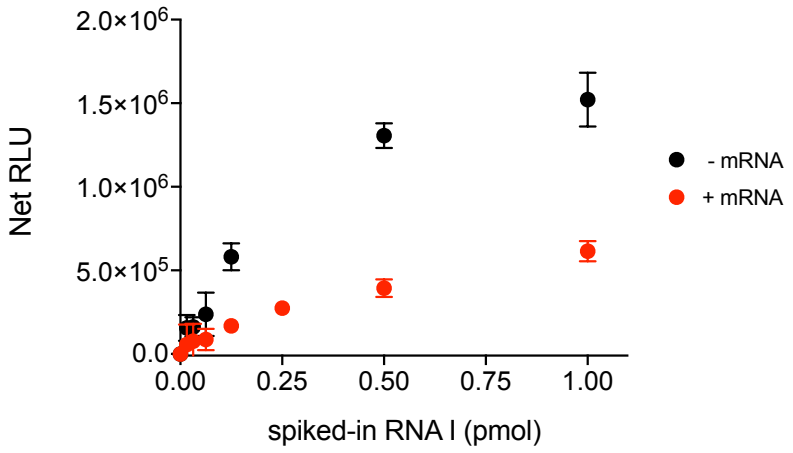


Figure D17. EndoVLISA performance in the presence of mRNA. Decreasing amounts RNA I strand was mixed with 1 µg *in vitro* transcribed mRNA and then glyoxal denatured and biotinylated. Samples were then immobilized in white streptavidin-coated 96-well plates and incubated with a solution of EndoV-MBP fusion protein at 1:1000 in 1X buffer (20 mM Tris, 100 mM NaCl, 1 mM $CaCl_2$, 0.05% Tw 20, pH 7.5). Wells were then probed with MBP-targeting 1º antibody (1:1000) and 2º HRP-conjugated antibody (1:40,000), and then SuperSignal™ West Atto chemiluminescent substrate (Thermo Fisher Scientific) was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Data points represent mean (n = 3) and error bars denote 95% confidence intervals.

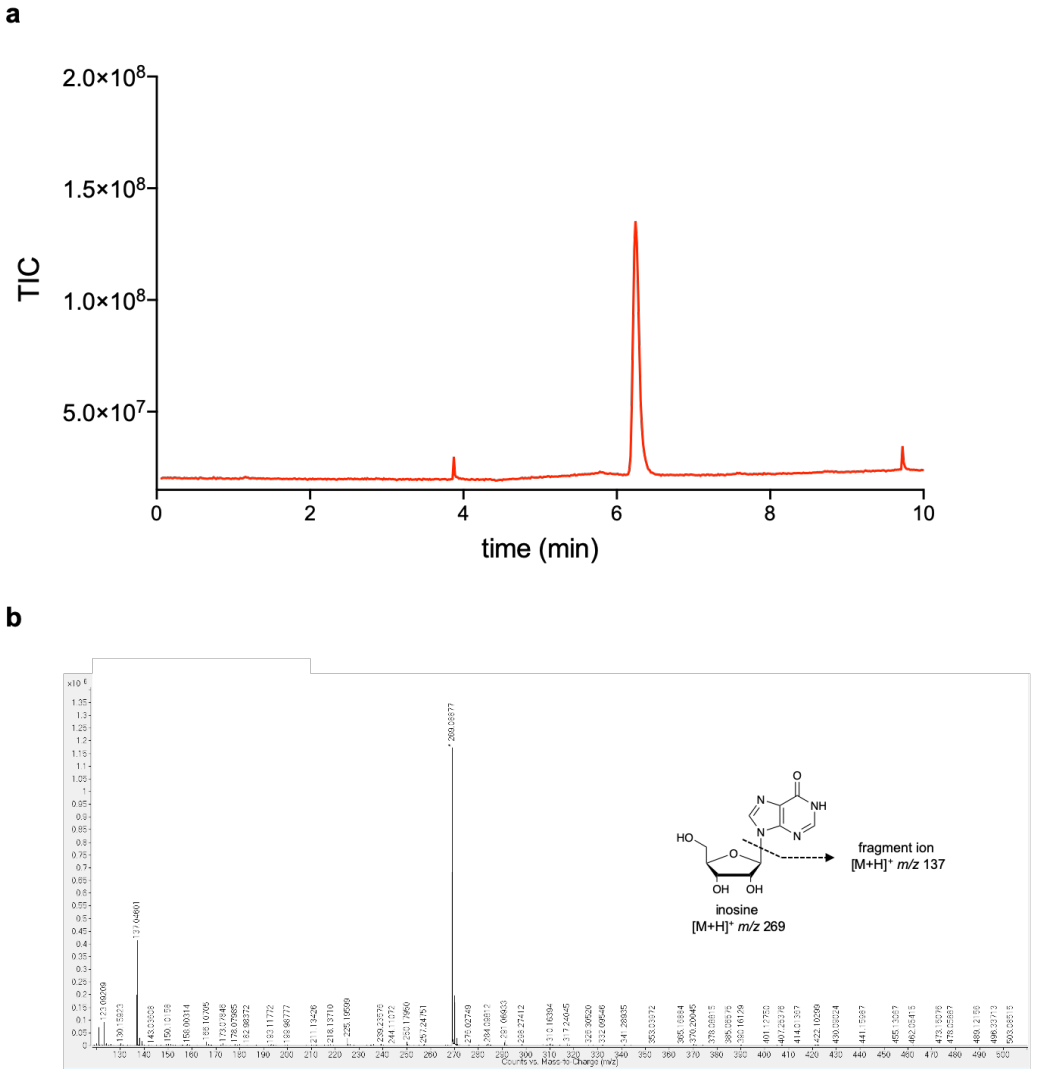


Figure D18. Inosine detection using LC/MS. a) Chromatography trace of inosine. 1 nmol inosine in water was separated using an Atlantis T3 column and total ion count (TIC) monitored over time. b) Mass spectrum of separated inosine (6.2 – 6.3 min retention time), with both the parent and fragment ion detected. Y axis denotes TIC values.

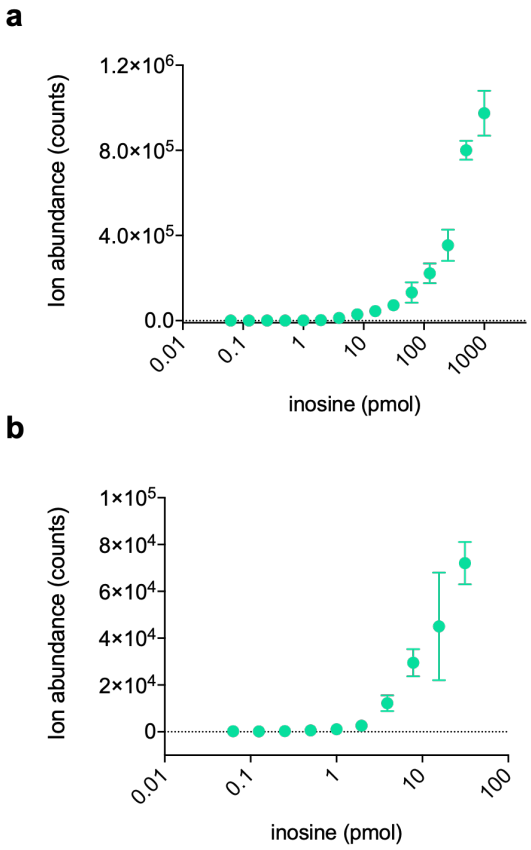


Figure D19. Inosine detection sensitivity using LC/MS. Decreasing amounts of inosine ribonucleoside was injected and analyzed by LC/MS. a) Full and b) partial (inosine < 100 pmol) standard curve in $\log_{10}$ scale. All values represent ion abundance counts from raw spectra. Data points represent mean (n = 3) and error bars denote 95% confidence intervals. Mass spectrum of inosine (6.2 – 6.3 min retention time) was used to quantify abundance of the parent ion (*m/z* 269).

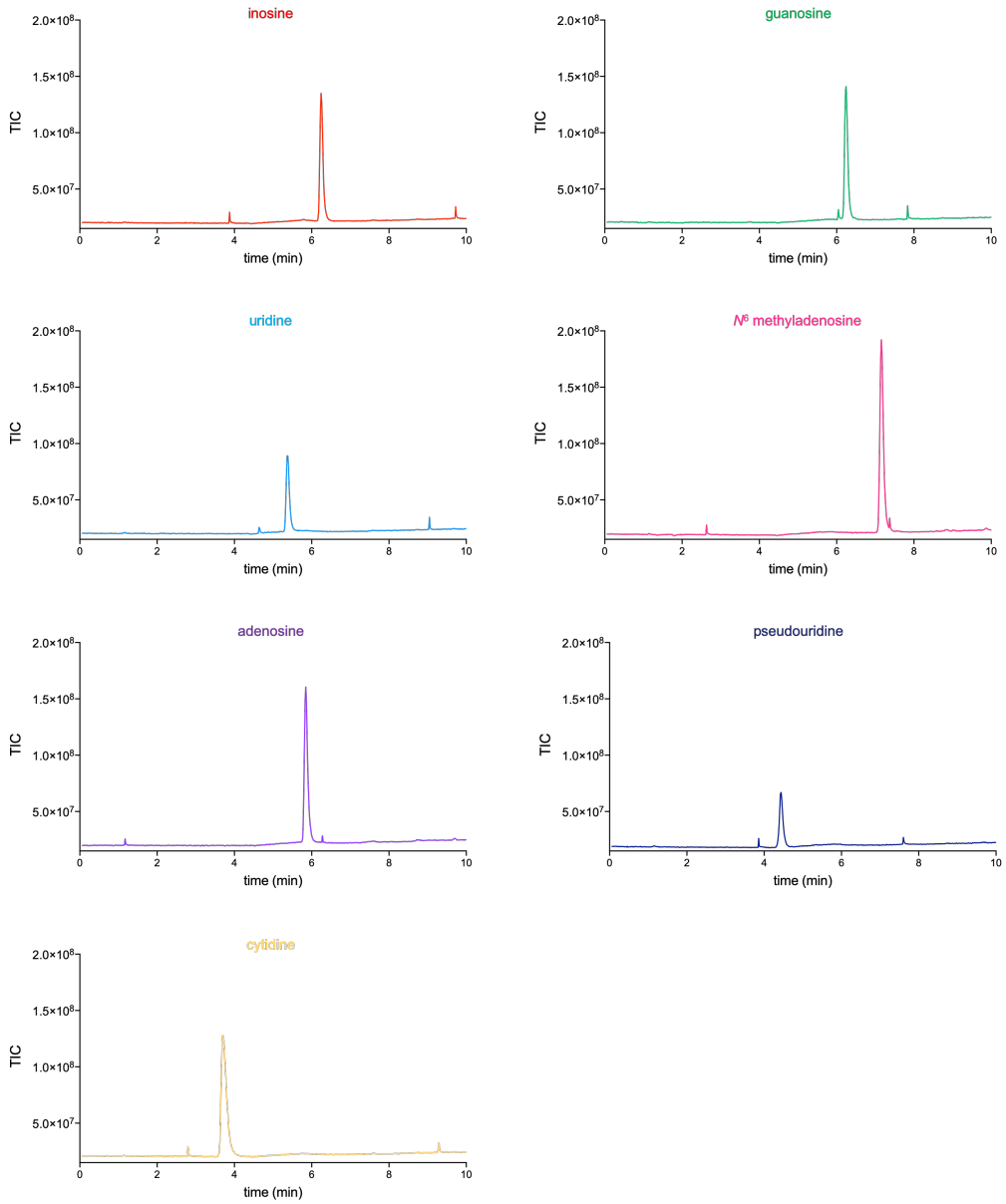


Figure D20. Chromatographic separation of ribonucleosides using LC/MS. 1 nmol of each ribonucleoside in water was separated using an Atlantis T3 column and total ion count (TIC) monitored over time.

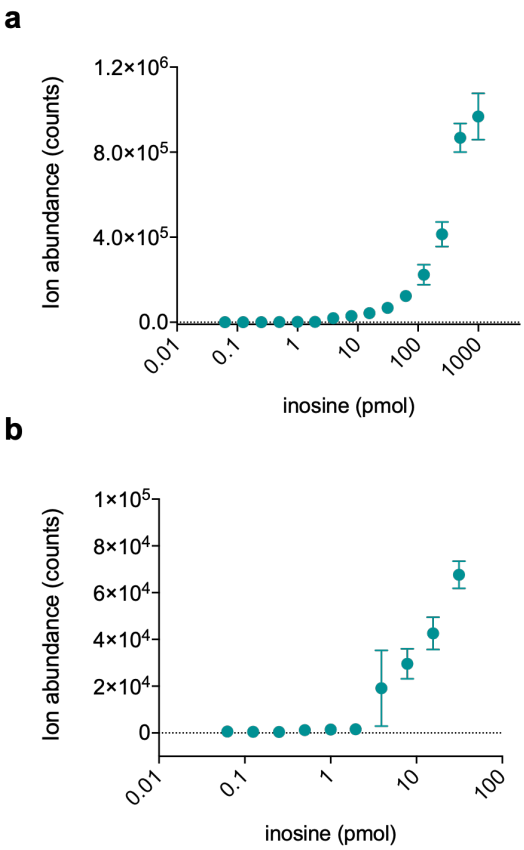


Figure D21. Inosine detection sensitivity in complex mixtures using LC/MS. Decreasing amounts of inosine ribonucleoside was spiked into 1 µg of an equimolar mixture of A, U, C, G, m^6A and PsU and then injected and analyzed by LC/MS. a) Full and b) partial (inosine < 100 pmol) standard curve in $\log_{10}$ scale. All values represent ion abundance counts from raw spectra. Data points represent mean (n = 3) and error bars denote 95% confidence intervals. Mass spectrum of inosine (6.2 – 6.3 min retention time) was used to quantify abundance of the parent ion (*m/z* 269).

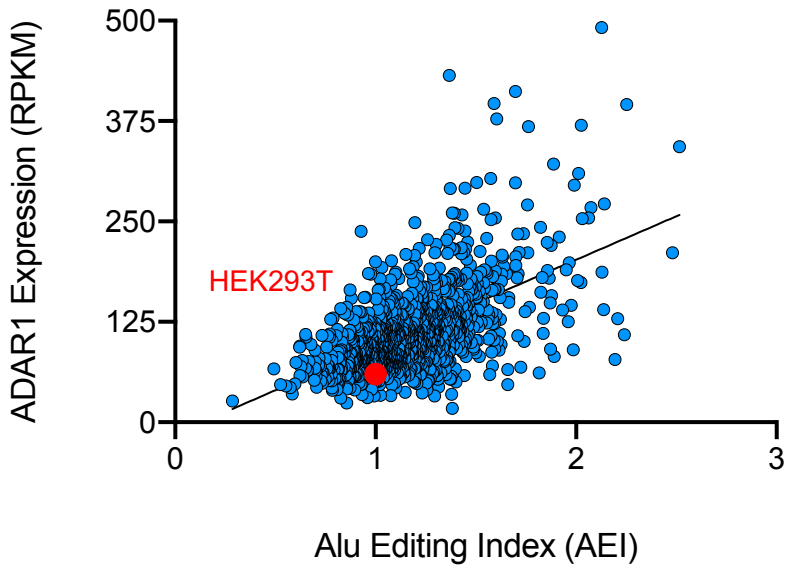


Figure D22. Overall editing activity in cell lines. Global A-to-I rates in 1147 different cell lines were estimated from RNA-seq datasets using the Alu Editing Index (AEI) and correlated with ADAR1 expression levels (reads per kilobase of transcript, per million mapped reads, RPKM). Reproduced in GraphPad Prism using data in supplementary table 4 from ref 75. Red dot indicates position of HEK293T cell line.

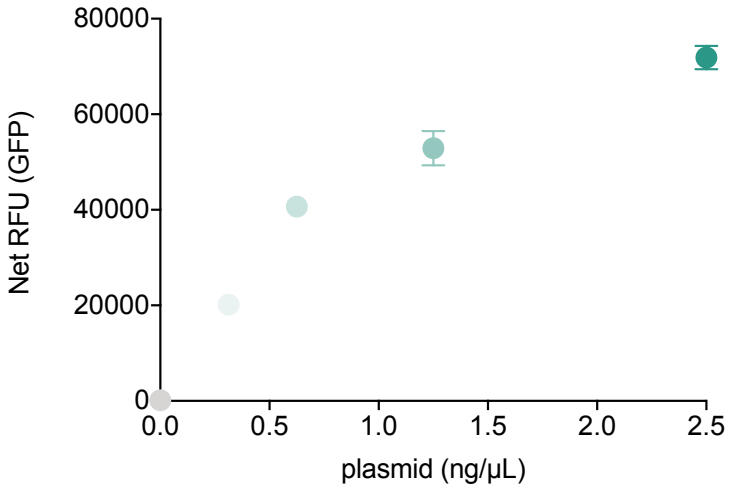


Figure D23. Increased ADAR1 expression in transfected HEK293T cells. Whole cell lysate was collected from HEK293T cells treated with increasing amounts of a GFP-ADAR1p110 plasmid. In triplicate, 20 μl of lysate was loaded into 384-well black plates and measured using a BioTek Cytation 5 plate reader using excitation at 488 nm and emission at 509 nm. Values represent mean (n = 3 wells) and error bars denote S.D. Net relative fluorescence units (RFU, arbitrary units) is calculated by subtracting blank values (lysis buffer).

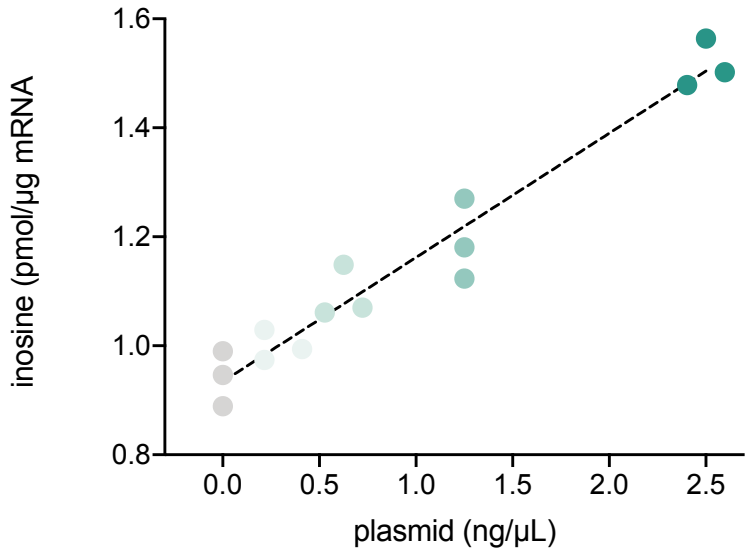


Figure D24. Quantifying inosine levels in HEK293Tcells overexpressing ADAR1. mRNA was isolated from HEK293T cells treated with increasing amounts of a GFP-ADAR1p110 plasmid. mRNA was then biotinylated, glyoxal-denatured, and then 1 µg was immobilized in each well of a white streptavidin-coated 96-well plate. Wells were then probed with a 1:1000 EndoV-MBP solution followed by MBP-targeting 1º antibody (1:1000) and 2º HRP-conjugated antibody (1:40,000). SuperSignal™ West Atto chemiluminescent substrate was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Inosine (pmol/µg mRNA) was calculated based off a standard curve containing known concentrations of RNA I control strand (0-2 pmol) mixed with 1 µg *in vitro* transcribed mRNA. Values represent net relative luminescent units (RLU, arbitrary units) calculated by subtracting appropriate blank wells (no RNA) from each set. Data points represent individual values from each well (n = 3 for each treatment group).

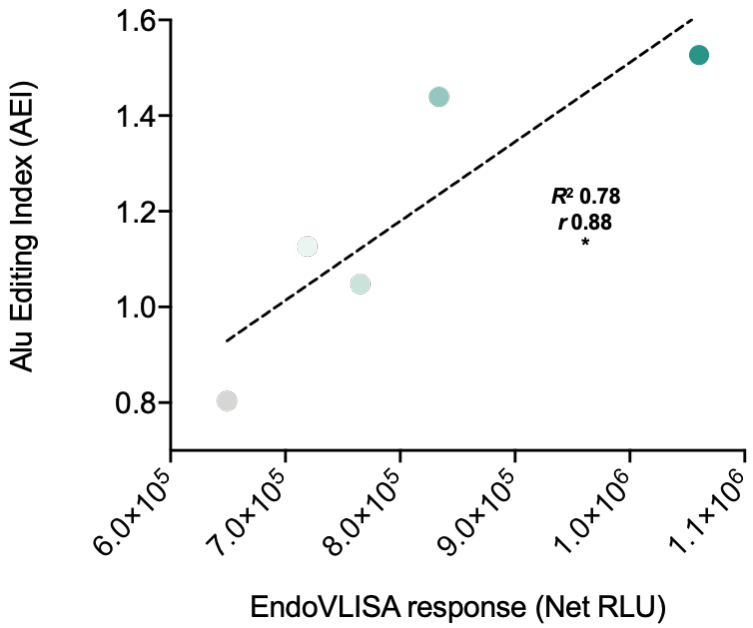


Figure D25. Comparing RNA-seq (Alu editing index, AEI) and EndoVLISA in detecting increased RNA editing signatures. Average AEI and EndoVLISA signal (net relative luminescent units, Net RLU) were plotted for each transfection amount (0, 0.31, 0.62, 1.25, 2.5 ng/μL plasmid). Values represent means from triplicate experiments. Linear regression (black dashed line), R^2 and pearson (r) correlation (* denotes $p < 0.05$) were computed in Prism.

Component	Product (vendor and SKU)	Unit Price ($)	Price per sample ($)
library preparation kit	SMARTer® Stranded Total RNA Sample Prep Kit - Low Input Mammalian 24 reactions, (Takara Bio 634861)	$2895.00	$120.63
Sequencing flow cell	NextSeq 500/550 High Output Kit v2.5 300 Cycles (Illumina 20024908)	$4867.00	$304.19*
			Total $419.15

Table D2. Components needed for RNA-seq analysis of A-to-I editing. *Price per sample is variable depending on multiplexing in RNA-seq on the same flow cell (16 samples were used in this study).

Component	Product (vendor and SKU)	Unit Price ($)	Price per sample* ($)
Biotin label	Biotin-dPEG$_4$-hydrazide, 50 mg (Sigma Aldrich QBD10219)	$257.00	$0.05
Plate	Streptavidin 96-well white (5 pack, Thermo Fisher 15218)	$212.00	$0.44
EndoV	Recombinant EndoV-MBP (New England Biolabs, M0305S)	$76.00	$0.30
1º antibody	Anti-MBP Monoclonal Antibody (New England Biolabs, E8032S)	$181.00	$0.72
2º HRP	Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP	$204.00	$0.08
Substrate	SuperSignal™ West Atto Ultimate Sensitivity Substrate (100 mL	$400.00	$0.40
			Total $2.00

Table D3. Components needed for EndoVLISA detection of A-to-I editing. *Calculated based on necessary amounts for processing 1 µg of mRNA and performing EndoVLISA detection in one-well of a 96-well plate.

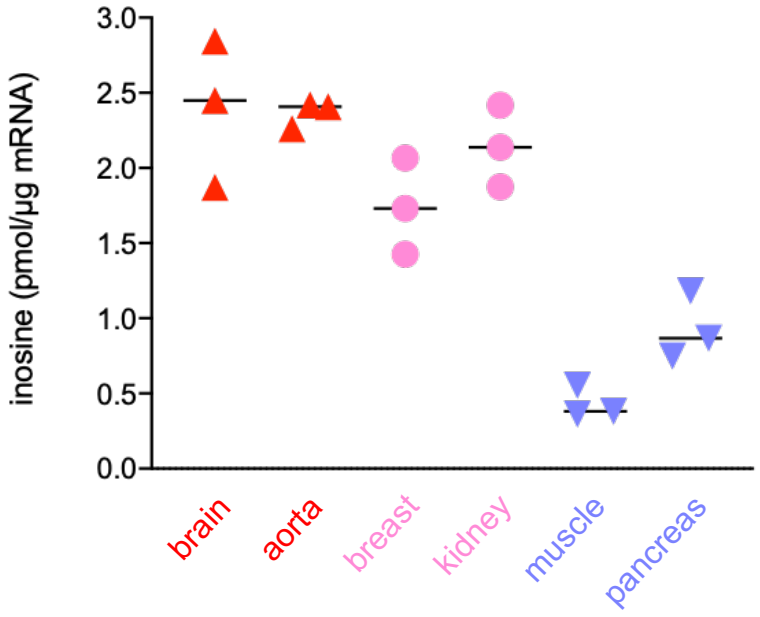


Figure D26. Quantifying inosine levels in mRNA isolated from human tissue. Purified mRNA from the indicated tissues was purchased from Takara Bio and then biotinylated and glyoxal-denatured. 500 ng was immobilized in each well of a white streptavidin-coated 96-well plate, and wells were then probed with a 1:1000 EndoV-MBP solution followed by MBP-targeting 1º antibody (1:1000) and 2º HRP-conjugated antibody (1:40,000). SuperSignal™ West Atto chemiluminescent substrate was then added to each well and allowed to incubate with shaking for 1 minute, followed by luminescent measurement using a BioTek Cytation 5 plate reader. Inosine (pmol/µg mRNA) was calculated based off a standard curve containing known concentrations of RNA I control strand (0-2 pmol) mixed with 500 ng *in vitro* transcribed mRNA. Data points represent individual values from each well (n = 3 for each sample).

Appendix E:

Omitted Data from Chapter 6

Thermoreversible Control of Nucleic Acid Structure and Function with Glyoxal Caging

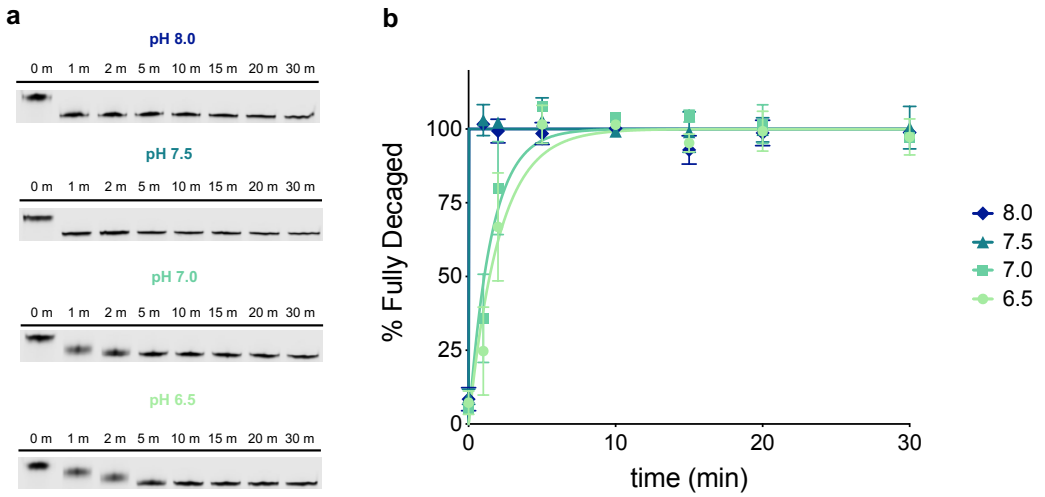


Figure E1. Decaging kinetics of a fully glyoxalated DNA strand at 95 °C with increasing pH. a) Representative images of decaging with 20% PAGE analysis. 20 pmol of a fully caged DNA strand was incubated at 95 ºC for the indicated times and immediately loaded onto a 20% polyacrylamide gel. b) Densitometric quantification of caging as a function of time. Values represent mean (n = 2) normalized percentages versus band intensity of an untreated DNA control. Error bars denote standard deviation.

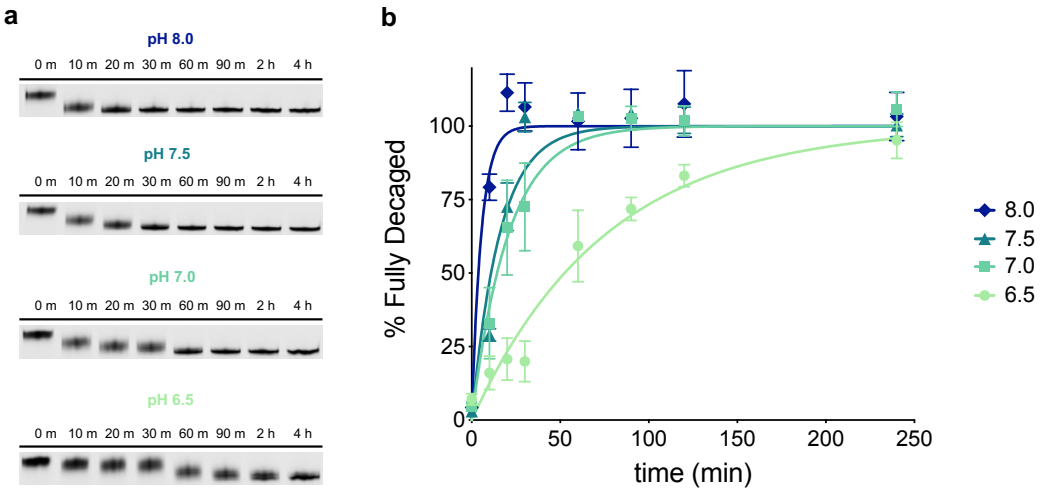


Figure E2. Decaging kinetics of a fully glyoxalated DNA strand at 70 °C with increasing pH. a) Representative images of decaging with 20% PAGE analysis. 20 pmol of a fully caged DNA strand was incubated at 70 ºC for the indicated times and immediately loaded onto a 20% polyacrylamide gel. b) Densitometric quantification of caging as a function of time. Values represent mean (n = 2) normalized percentages versus band intensity of an untreated DNA control. Error bars denote standard deviation.

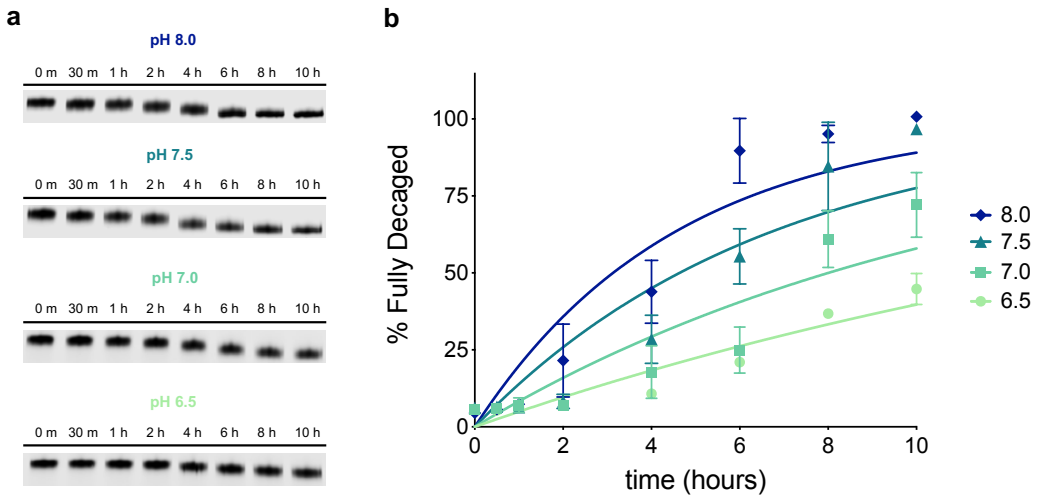


Figure E3. Decaging kinetics of a fully glyoxalated DNA strand at 50 °C with increasing pH. a) Representative images of decaging with 20% PAGE analysis. 20 pmol of a fully caged DNA strand was incubated at 50 ºC for the indicated times and immediately loaded onto a 20% polyacrylamide gel. b) Densitometric quantification of caging as a function of time. Values represent mean (n = 2) normalized percentages versus band intensity of an untreated DNA control. Error bars denote standard deviation.

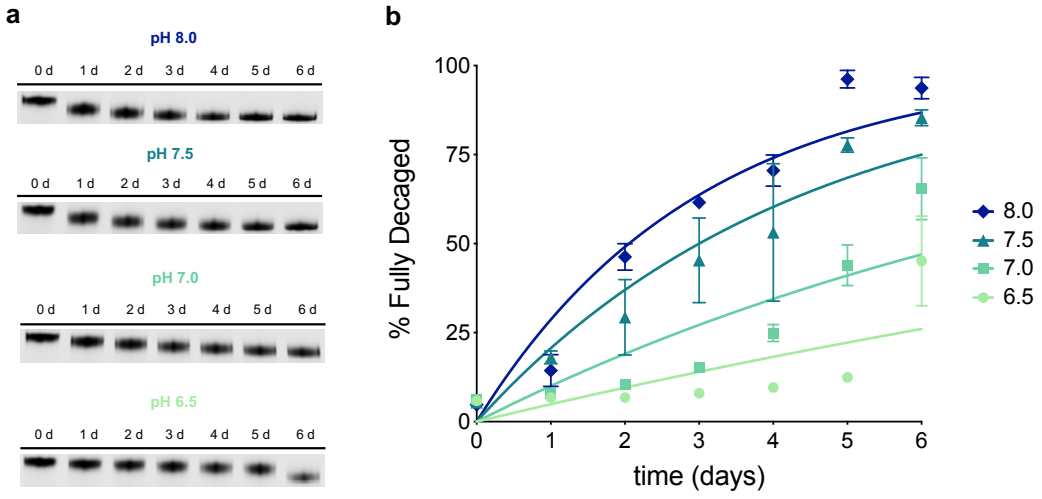


Figure E4. Decaging kinetics of a fully glyoxalated DNA strand at 37 °C with increasing pH. a) Representative images of decaging with 20% PAGE analysis. 20 pmol of a fully caged DNA strand was incubated at 37 ºC for the indicated times and immediately loaded onto a 20% polyacrylamide gel. b) Densitometric quantification of caging as a function of time. Values represent mean (n = 2) normalized percentages versus band intensity of an untreated DNA control. Error bars denote standard deviation.

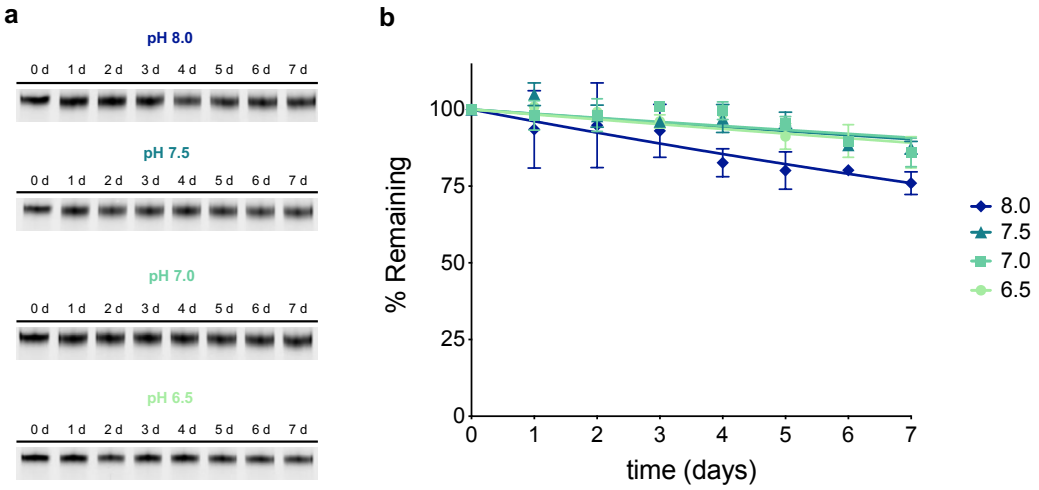


Figure E5. Room temperature stability of a glyoxalated DNA strand with increasing pH. a) Representative images of decaging with 20% PAGE analysis. 20 pmol of a fully caged DNA strand was incubated at room temperature for the indicated times and immediately loaded onto a 20% polyacrylamide gel. b) Densitometric quantification of caging as a function of time. Values represent mean (n = 2) normalized percentages versus band intensity of the fully caged DNA strand at T = 0 hours. Error bars denote standard deviation.

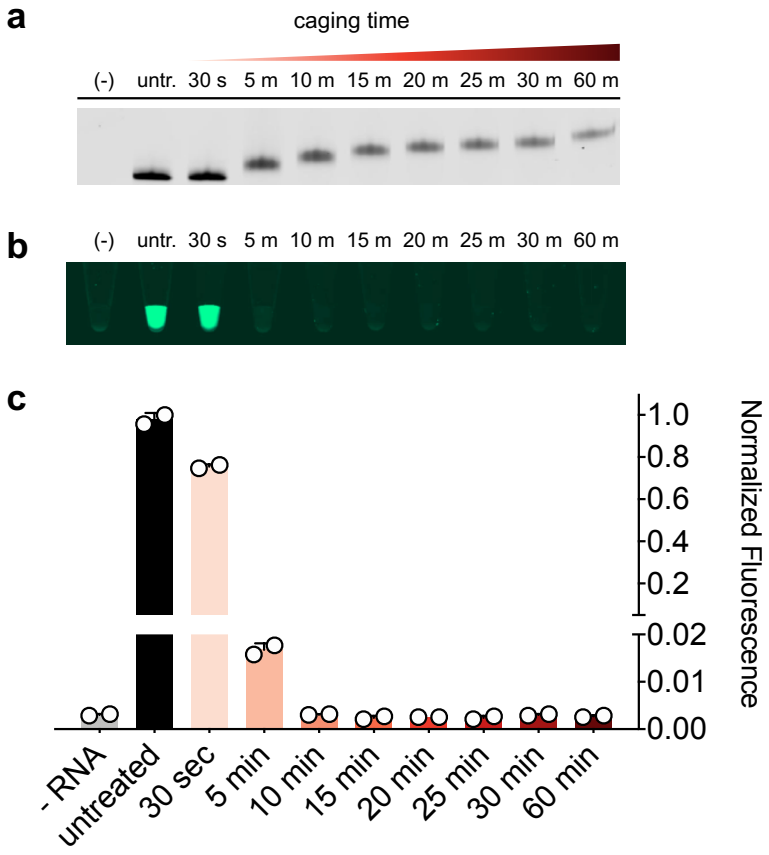


Figure E6. Fluorogenic performance of the broccoli RNA aptamer with increasing caging times. a) 20% PAGE analysis of broccoli RNA aptamer (104nt) after various amounts of glyoxal caging, illustrating an increase in apparent molecular weight. b,c) Functional performance of broccoli aptamers with increasing degrees of caging, demonstrating that only partial glyoxalation (5–10 minutes) is necessary to ablate fluorogenic activity.

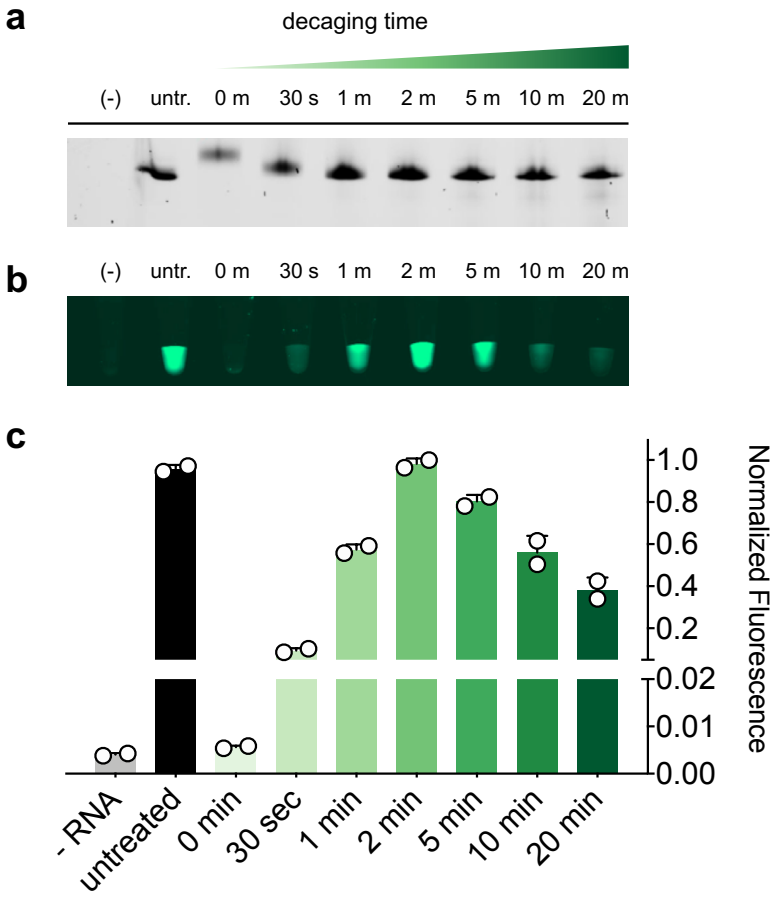


Figure E7. Restoration of fluorogenic activity in a caged broccoli RNA aptamer as a function of increasing decaging times. a) 20% PAGE analysis of caged broccoli RNA aptamer (104 nt) after increasing decaging times at 95 ºC pH 7.5 illustrating a decrease in apparent molecular weight. b,c) Fluorescent activity of caged broccoli aptamers as a function of decaging times, demonstrating that ~2 minutes is optimal for full restoration of fluorogenic activity.

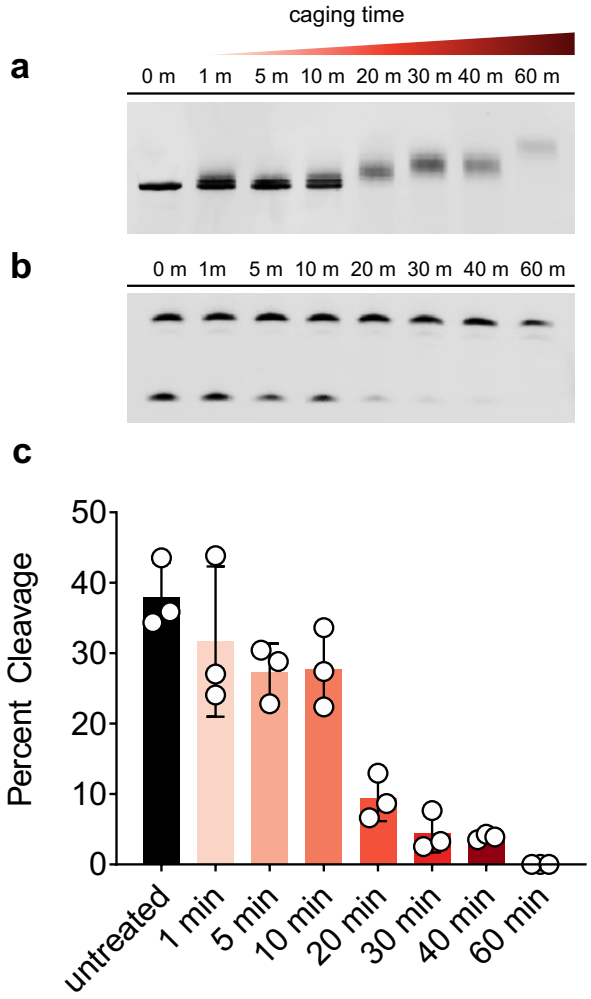


Figure E8. Glyoxalation disrupts 10-23 DNAzyme structure and target cleavage. a) 20% PAGE analysis of the 10-23 DNAzyme with increasing amounts of glyoxal caging, illustrating an increase in apparent molecular weight. b) 12% PAGE analysis monitoring target cleavage by increasingly caged 10-23 DNAzyme. c) Densitometric quantification of caged DNAzyme activity. Band intensity was used to quantify activity as percent target cleavage (n = 3). Error bars denote standard deviation.

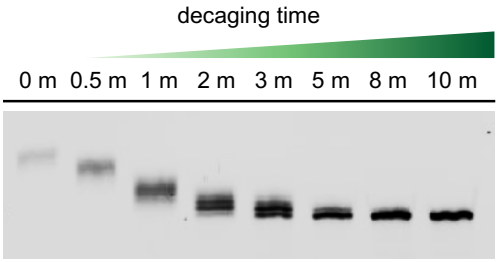


Figure E9. DNAzyme decaging kinetics. 20% PAGE analysis of caged 10-23 DNAzyme after increasing decaging times at 95 ºC pH 7.5 illustrating a decrease in apparent molecular weight.

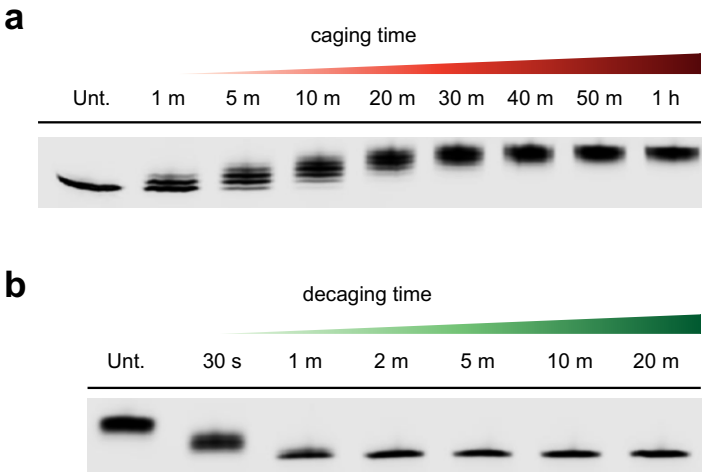


Figure E10. Glyoxal caging and decaging of a fully 2'-*O*-methylated RNA aptamer. a) 20% PAGE analysis of caged ARC259 (23nt) after increasing caging times, illustrating an increase in apparent molecular weight. b) 20% PAGE analysis of caged ARC259 after increasing decaging times at 95 ºC, pH 7.5 illustrating a decrease in apparent molecular weight.

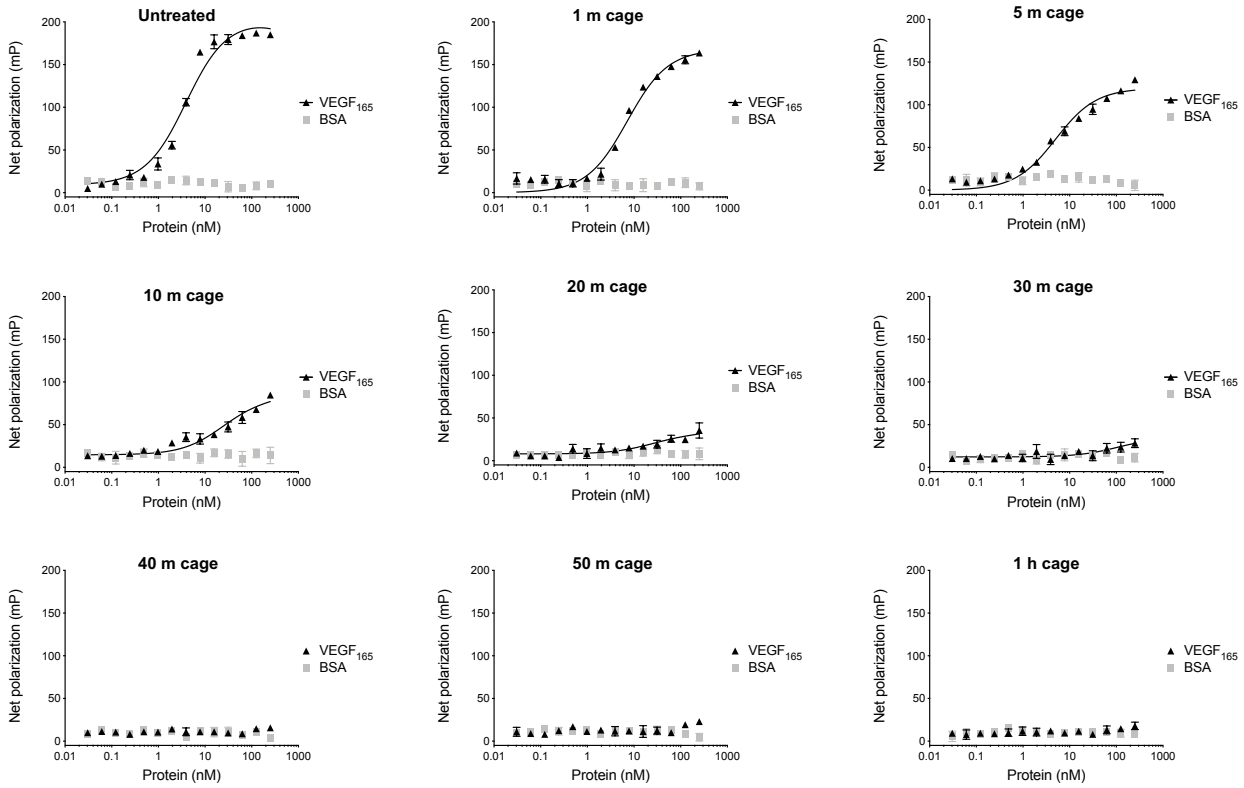


Figure E11. Functional binding activity of increasingly caged ARC259. Fluorescence polarization (FP) curves of aptamer binding towards $VEGF_{165}$ or bovine serum albumin (BSA). All values were normalized to a buffer blank and represent mean (n = 3) of independent replicates.

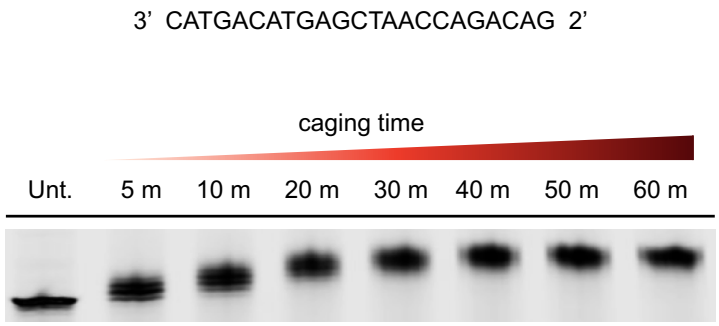


Figure E12. Glyoxal caging of a TNA oligonucleotide. Sequence of model TNA strand (23 nt) and 20% PAGE analysis after increasing caging times, illustrating an increase in apparent molecular weight.

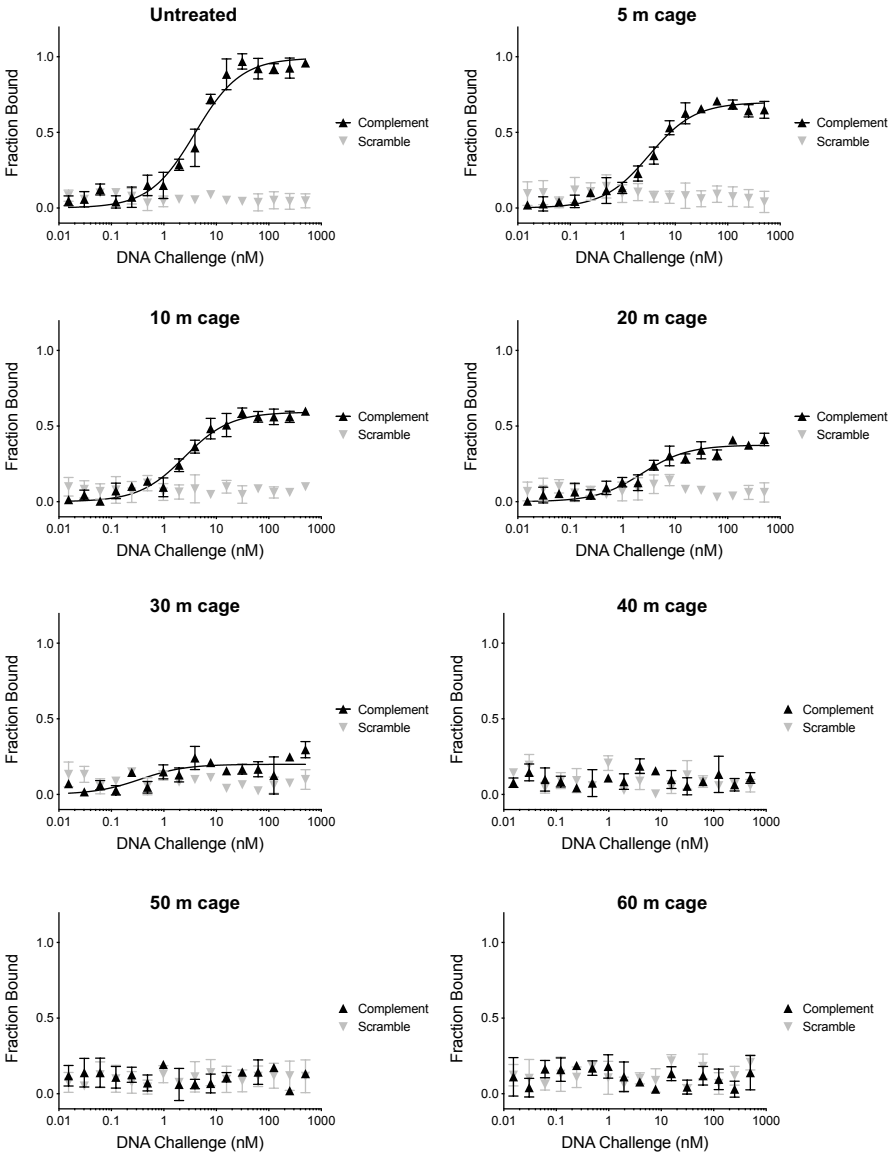


Figure E13. Glyoxal caging of TNA disrupts hybridization to DNA. Microscale thermophoresis (MST) of TNA binding towards a complementary (black) or scrambled DNA sequence (grey) with increasing caging times. Values represent mean (n = 3) of independent replicates.

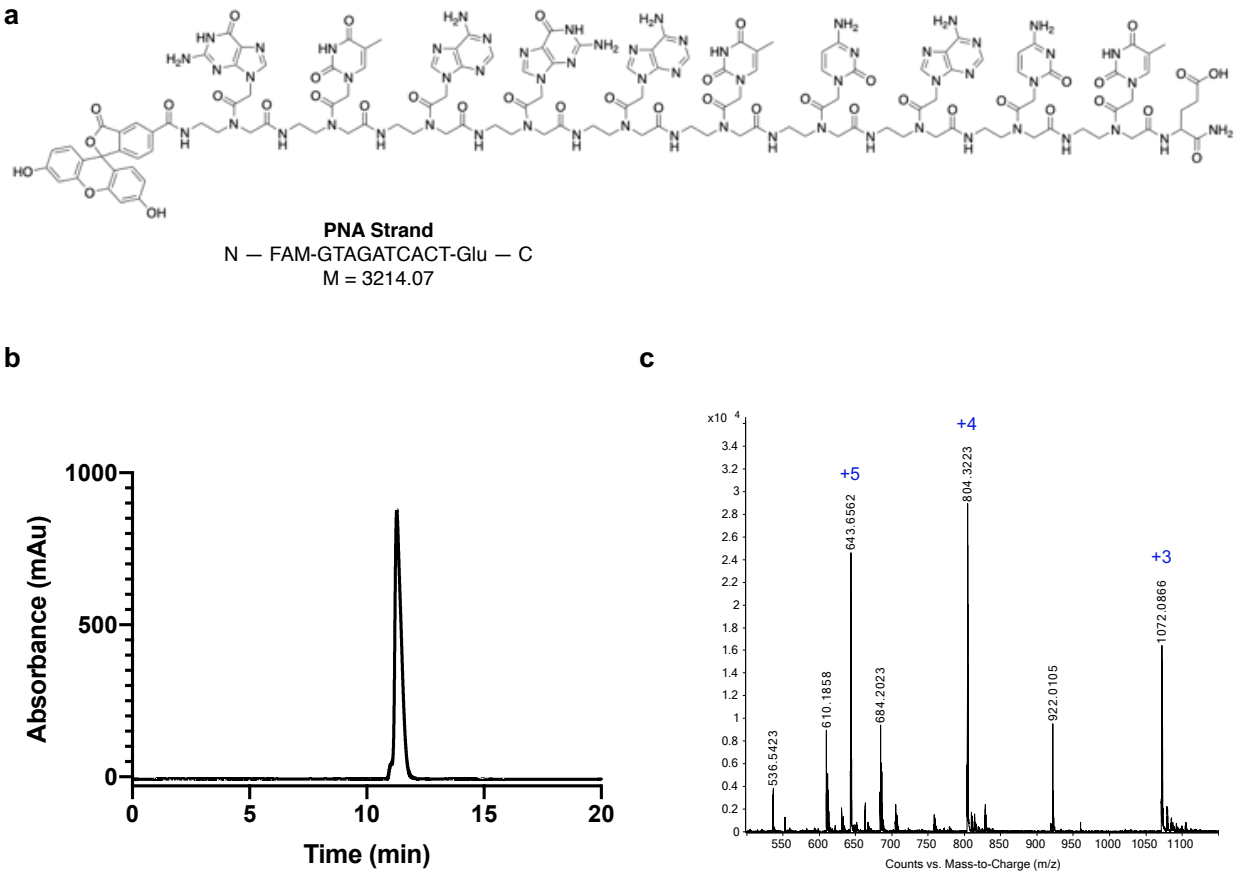


Figure E14. Analytical characterization of synthesized model PNA strand. a) Chemical structure and sequence of model PNA strand used in the study. Glu = glutamic acid. b) HPLC (A260 nm) and c) TOF ESI-MS analysis of purified PNA.

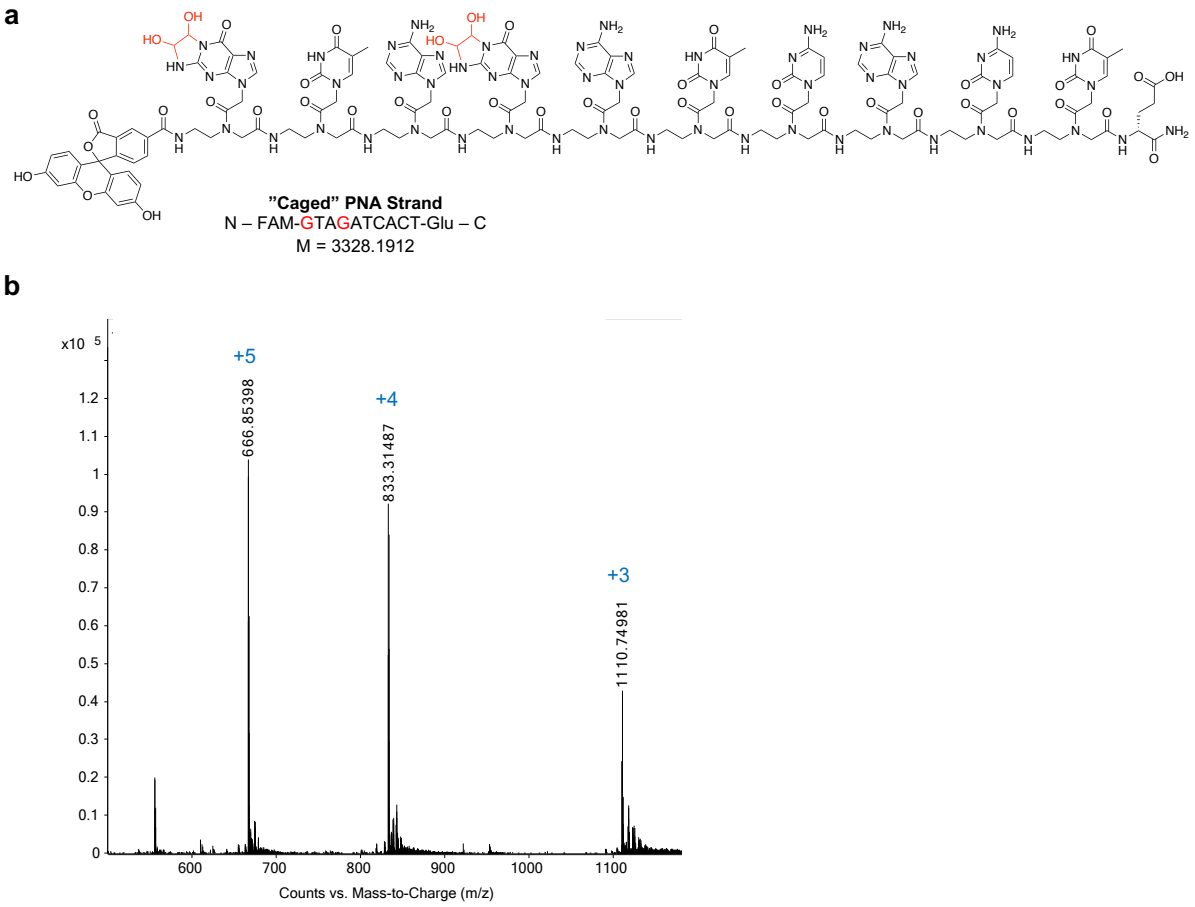


Figure E15. Analytical characterization of glyoxal-treated PNA. a) Chemical structure and sequence of PNA strand with putative glyoxal adducts (red). Glu = glutamic acid. b) TOF ESI-MS analysis of purified PNA treated with glyoxal.

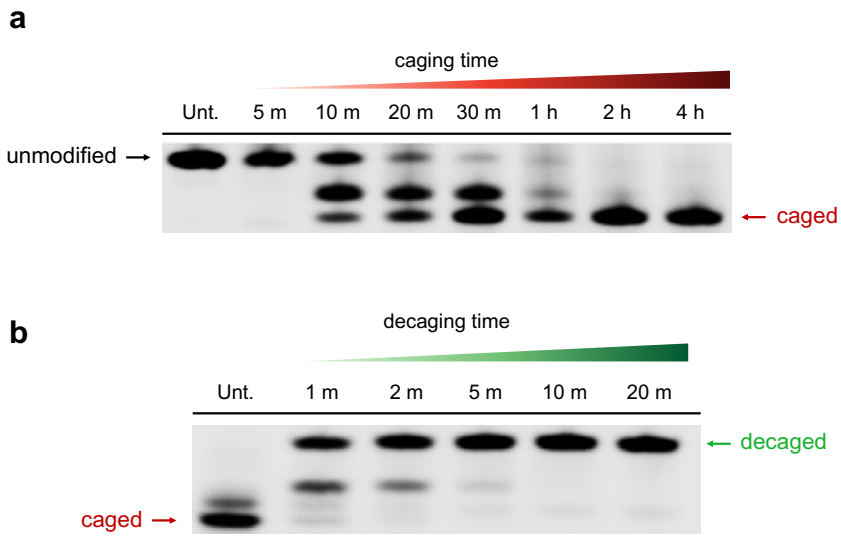


Figure E16. PAGE analysis of PNA caging and decaging. a) 20% PAGE analysis after increasing caging times, illustrating observed downward shift in electrophoretic mobility. b) 20% PAGE analysis of caged PNA after increasing decaging times at 95 ºC pH 7.5, illustrating restoration in electrophoretic mobility.

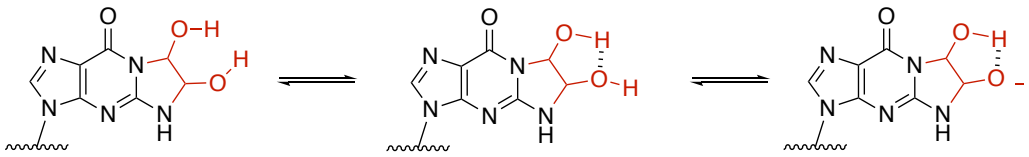


Figure E17. Proposed molecular mechanism of PNA mobility shifts. Bis-hemiaminal adducts introduced by glyoxal may result in an ionizable proton and overall gain in negative charge through hydrogen bond sharing between hydroxyl moieties.

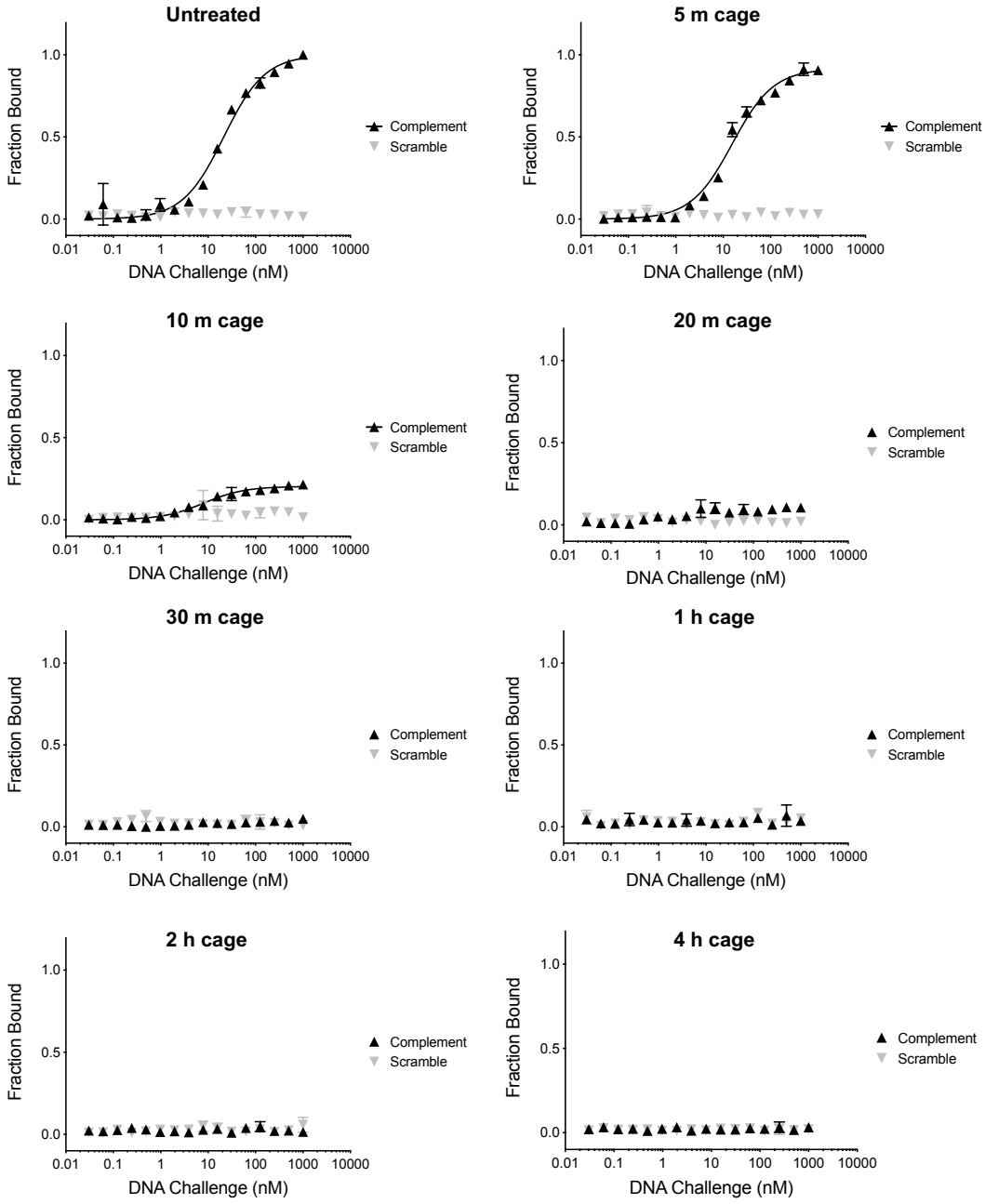


Figure E18. Glyoxal caging of PNA caging inhibits hybridization to DNA. a) Microscale thermophoresis (MST) of PNA binding towards a complementary (black) or scrambled DNA sequence (grey) with increasing caging times. Values represent mean (n = 3) of independent replicates.

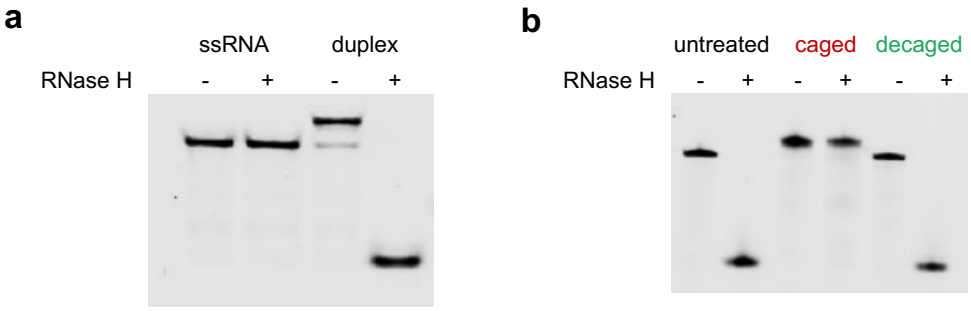


Figure E19. Reversible control of RNase H with glyoxal caging. a) 12% native PAGE gel showing requirement of RNA:DNA duplexes for RNase H cleavage activity. b) 12% denaturing PAGE gel of RNase H mediated target cleavage of untreated, caged, and decaged RNA.

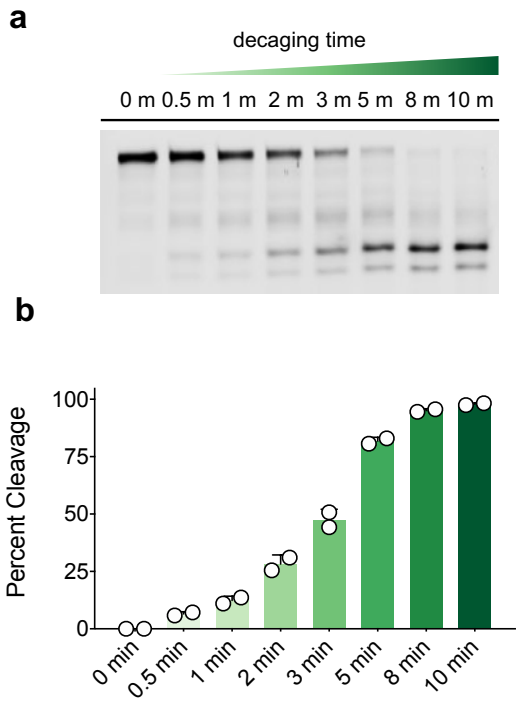


Figure E20. One-pot activation of thermostable RNase H. a) 12% PAGE analysis and b) quantified percent target cleavage (n = 2) by thermostable RNase H with increased decaging time at 95 ºC.

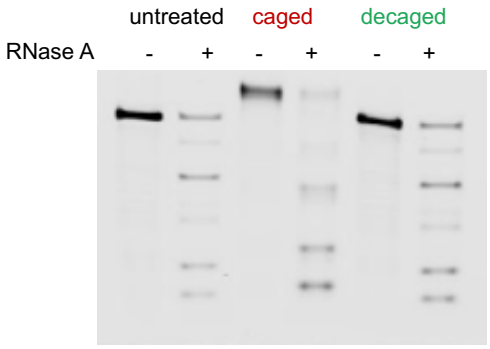


Figure E21. Glyoxal does not inhibit RNase A. 12% PAGE analysis of RNase A activity towards untreated, caged, and decaged RNA, illustrating cleavage of all substrates.

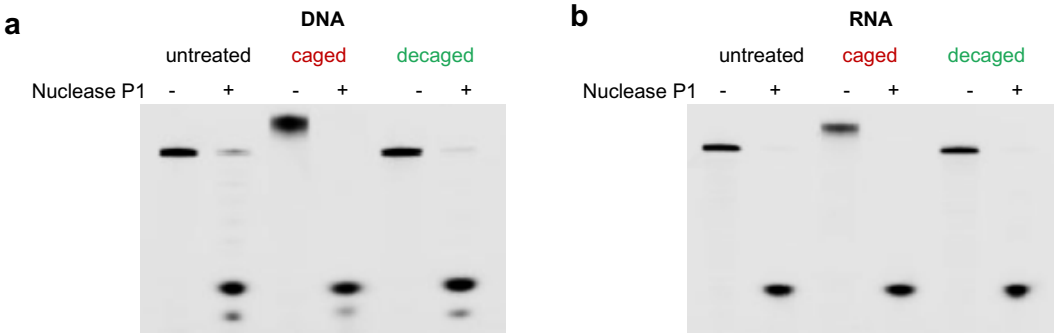


Figure E22. Glyoxal does not inhibit Nuclease P1. 12% PAGE gel of nuclease P1 cleavage of untreated, caged, and decaged a) DNA and b) RNA.

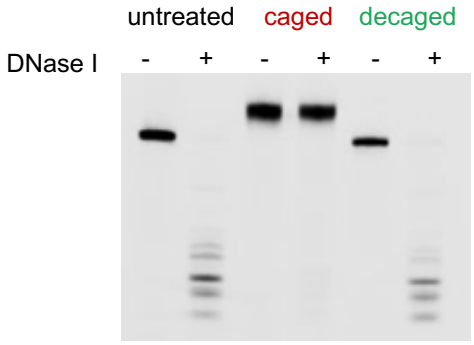


Figure E23. Reversible control of DNase I recognition and cleavage. 12% PAGE gel of DNase I cleavage of untreated, caged, and decaged DNA.

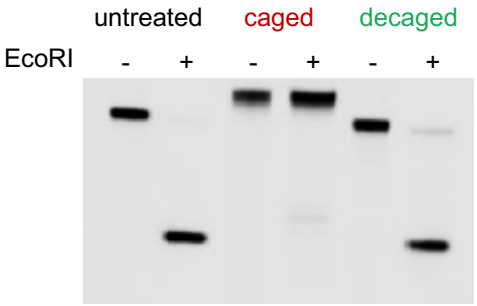


Figure E24. Reversible control of EcoRI. 12% PAGE gel of EcoRI mediated target cleavage by untreated, caged, and decaged DNA.

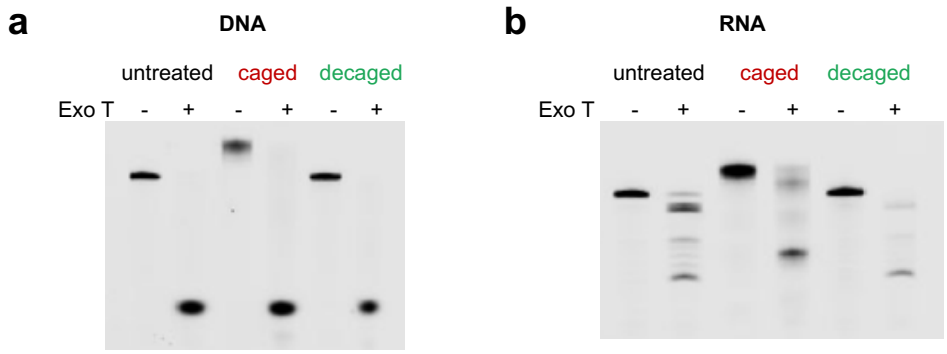


Figure E25. Glyoxal does not inhibit RNase T. 12% PAGE gel of RNase T cleavage of untreated, caged, and decaged a) DNA and b) RNA.

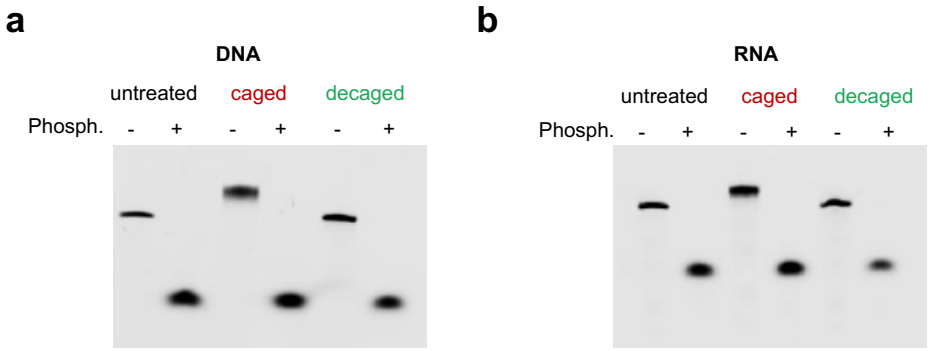


Figure E26. Glyoxal does not inhibit snake venom phosphodiesterase I. 12% PAGE gel of phosphodiesterase cleavage of untreated, caged, and decaged a) DNA and b) RNA.

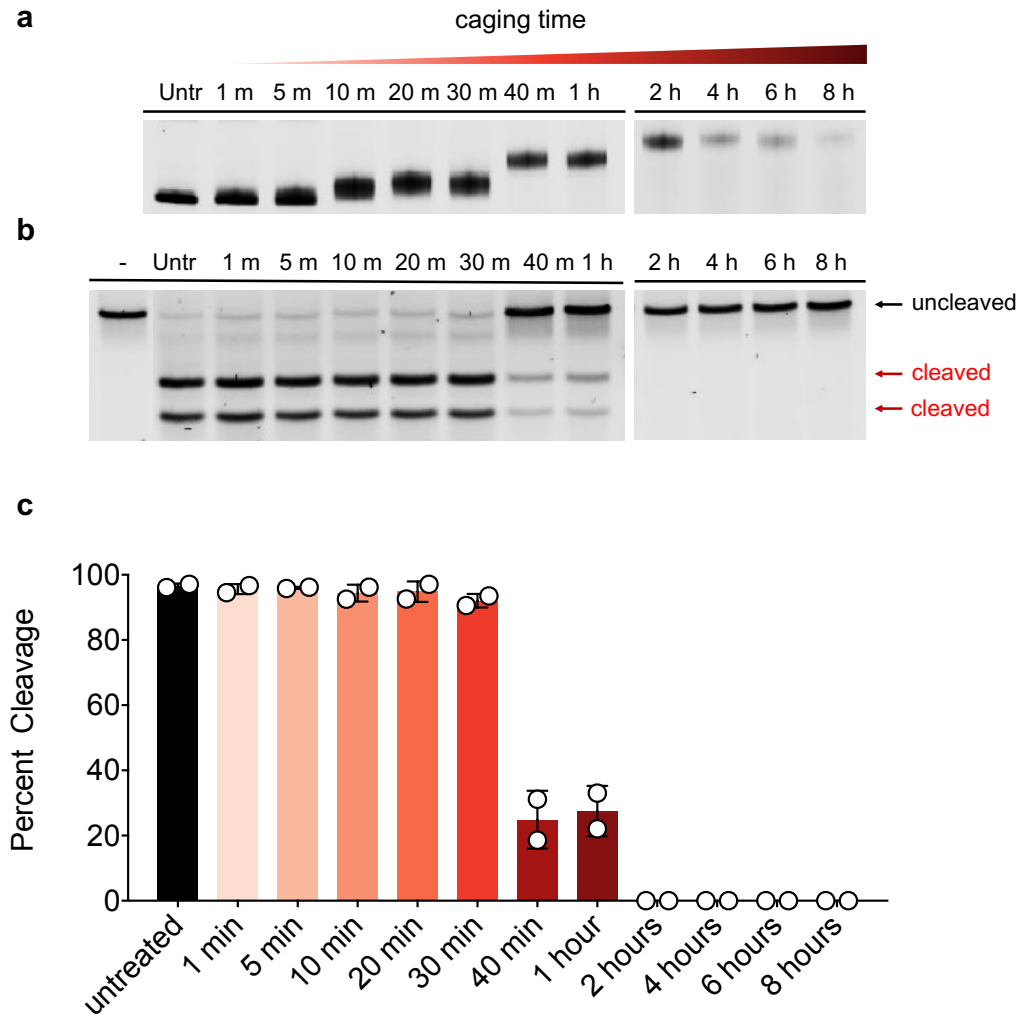


Figure E27. Glyoxalation disrupts sgRNA and Cas9 mediated target cleavage. a) Caging kinetics of sgRNA were monitored by 10% PAGE gel illustrating an increase in apparent molecular weight. b) 1% agarose gel analysis monitoring dsDNA target cleavage by increasingly caged sgRNA. "-" indicates no sgRNA or RNP included in reaction. c) Densitometric quantification of Cas9 activity with increasingly caged sgRNA. Band intensity was used to quantify activity as percent target cleavage (n = 2). Error bars denote standard deviation.

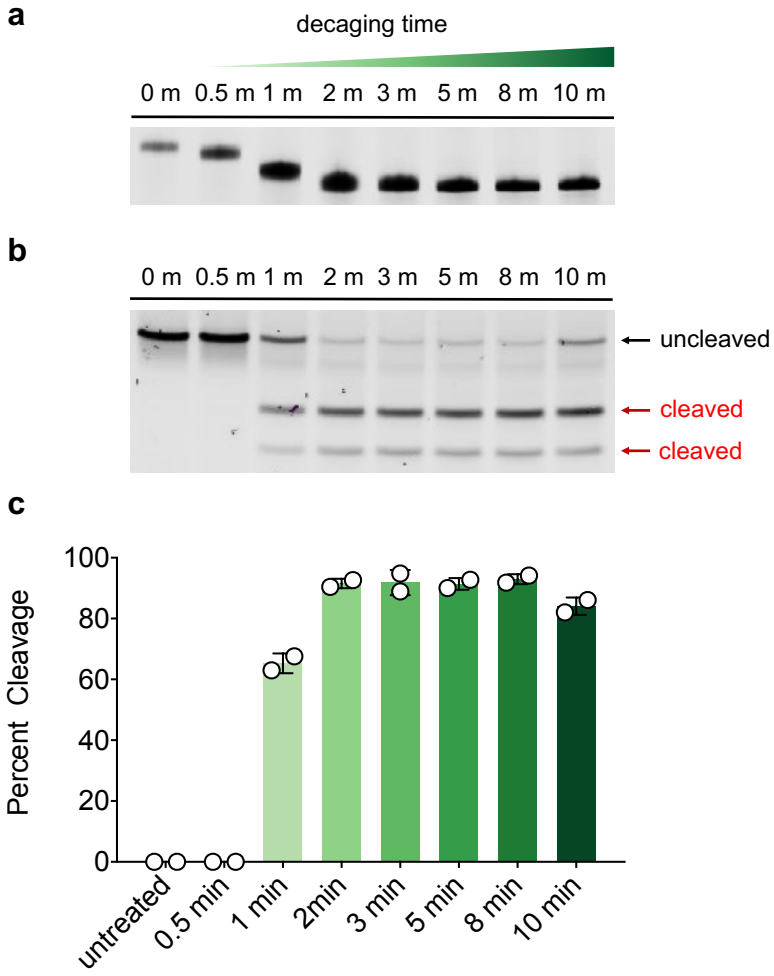


Figure E28. Glyoxalation of sgRNA is reversible with rapid decaging. a) Decaging kinetics of sgRNA at 95 °C were monitored by 10% PAGE gel illustrating decrease in apparent molecular weight. b) 1% agarose gel analysis monitoring target cleavage by increasingly decaged sgRNA. c) Densitometric quantification of caged sgRNA- Cas9 activity. Band intensity was used to quantify activity as percent target cleavage (n = 2). Error bars denote standard deviation.

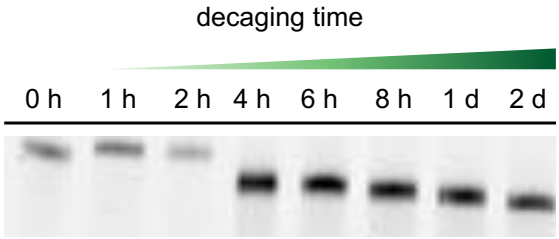


Figure E29. Glyoxalation of sgRNA is reversible with slow decaging. Decaging kinetics of sgRNA at 37 °C were monitored by 10% PAGE gel illustrating decrease in apparent molecular weight.

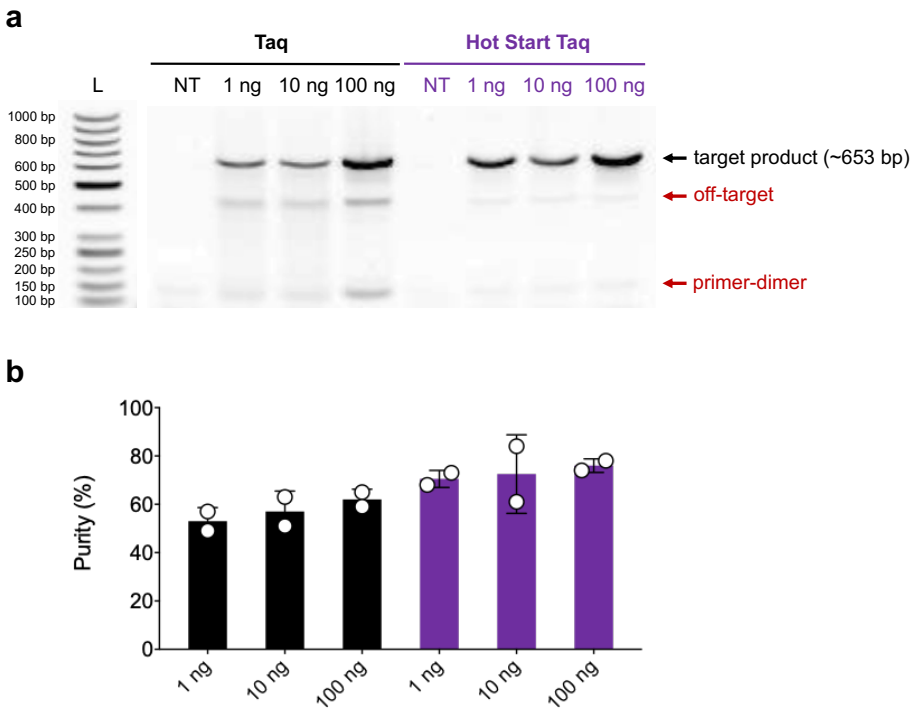


Figure E30. Amplification of β-actin from human genomic DNA using *Taq* and Hot Start *Taq* polymerases. a) PCR amplification with increasing genomic DNA template visualized by 1% agarose gel stained with SYBR Safe (NT = no template). Expected target band (~653 bp) is indicated by black arrows, while off-target and putative primer-dimer bands are labeled with red arrows. b) Densitometric quantification of amplificon purity. Bars represent mean and S.D. from 2 independent trials.

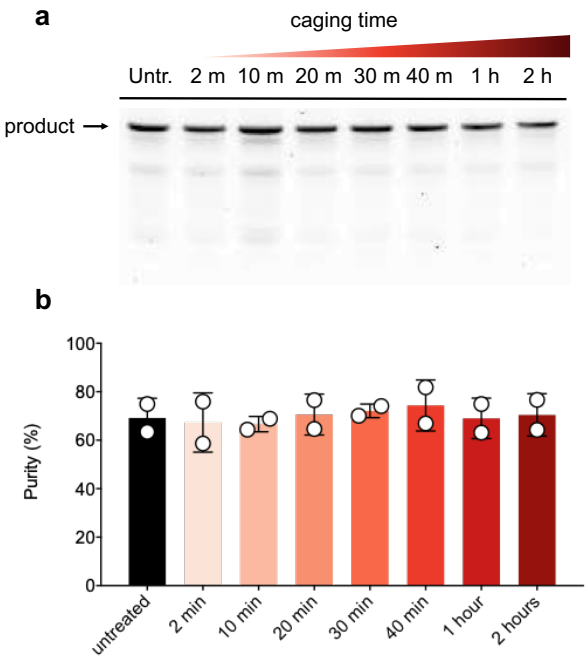


Figure E31. Forward primer glyoxal treatment and PCR specificity. a) PCR reactions using increasingly caged forward primer as visualized by 1% agarose gel and b) quantified using densitometry. Bars represent mean and S.D. from 2 independent trials.

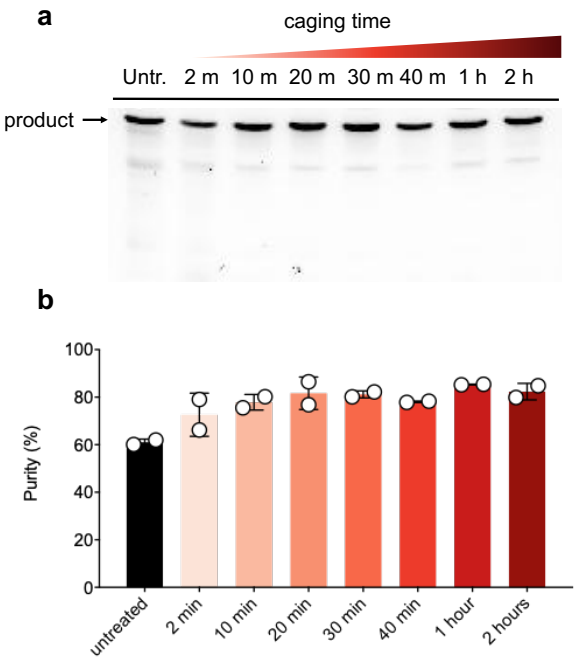


Figure E32. Reverse primer glyoxal treatment and PCR specificity a) PCR reactions with increasingly caged reverse primer as visualized by 1% agarose gel and b) quantified using densitometry. Bars represent mean and S.D. from 2 independent trials.

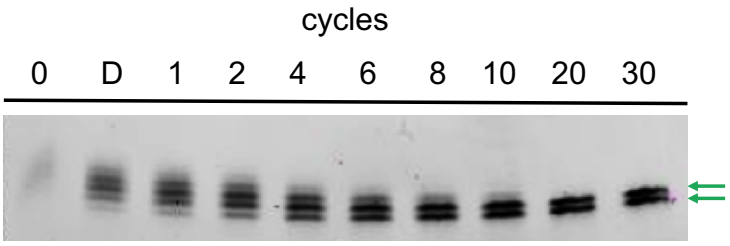


Figure E33. Primer decaging during PCR. 20% PAGE analysis of caged forward/primer mix after increasing PCR cycles in 1X PCR buffer, illustrating a decrease in apparent molecular weight. D = initial denaturation step (94 ºC for 2 min). Reactions contain a mix of two primers, resulting in two distinct bands in each lane. Green arrows denote fully decaged forward and primer.

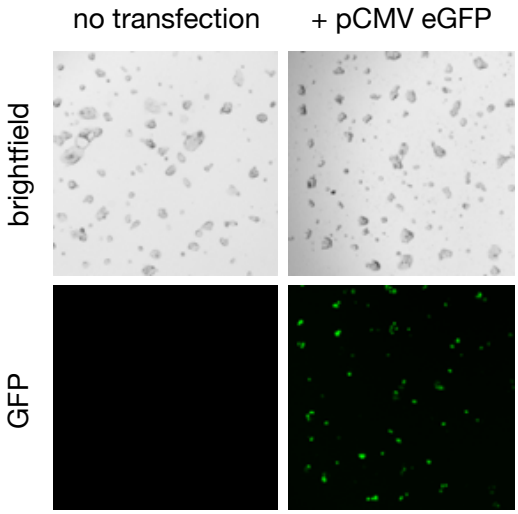


Figure E34. Confirming transfection efficiency and specificity. HEK293T cells were transfected with 200 ng of a pCMV vector encoding eGFP. Control cells received PBS. After 12 hours, GFP expression was detected by live-cell imaging using a BioTek Lionheart FX automated microscope.

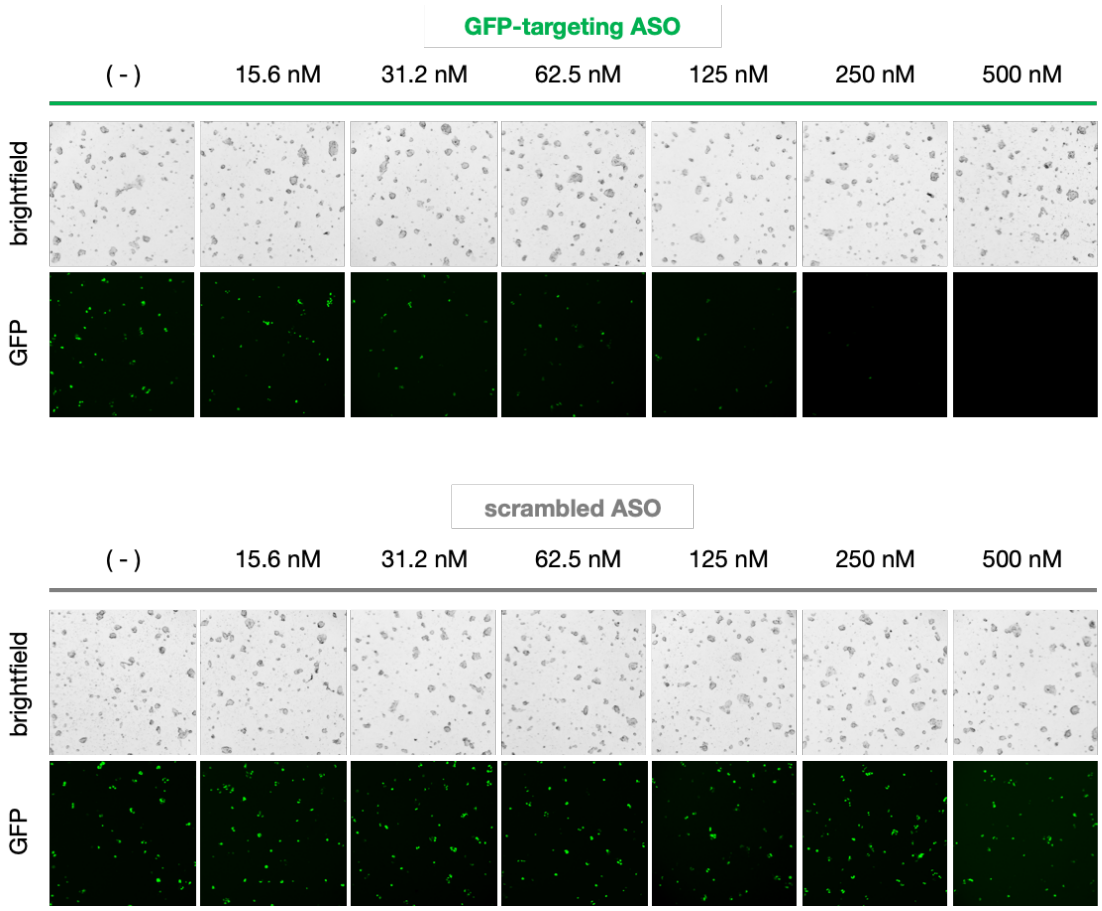


Figure E35. Optimizing ASO concentration and confirming sequence specificity. HEK293T cells were co-transfected with both a pCMV vector encoding eGFP (200 ng) as well as increasing amounts of either an eGFP-targeting or scrambled ASO ("-" denotes no ASO). After 24 hours, GFP expression was detected by live-cell imaging using a BioTek Lionheart FX automated microscope (4X magnification).

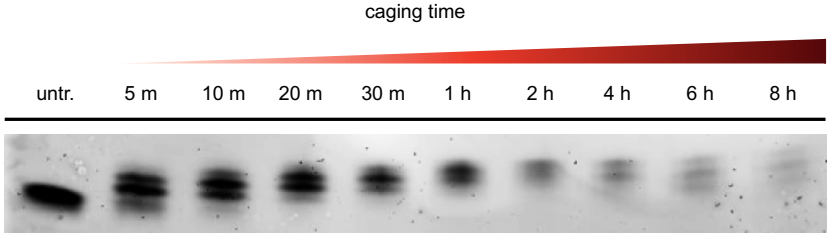


Figure E36. Caging kinetics of an eGFP-targeting ASO as monitored by 20% PAGE gel illustrating an increase in apparent molecular weight.

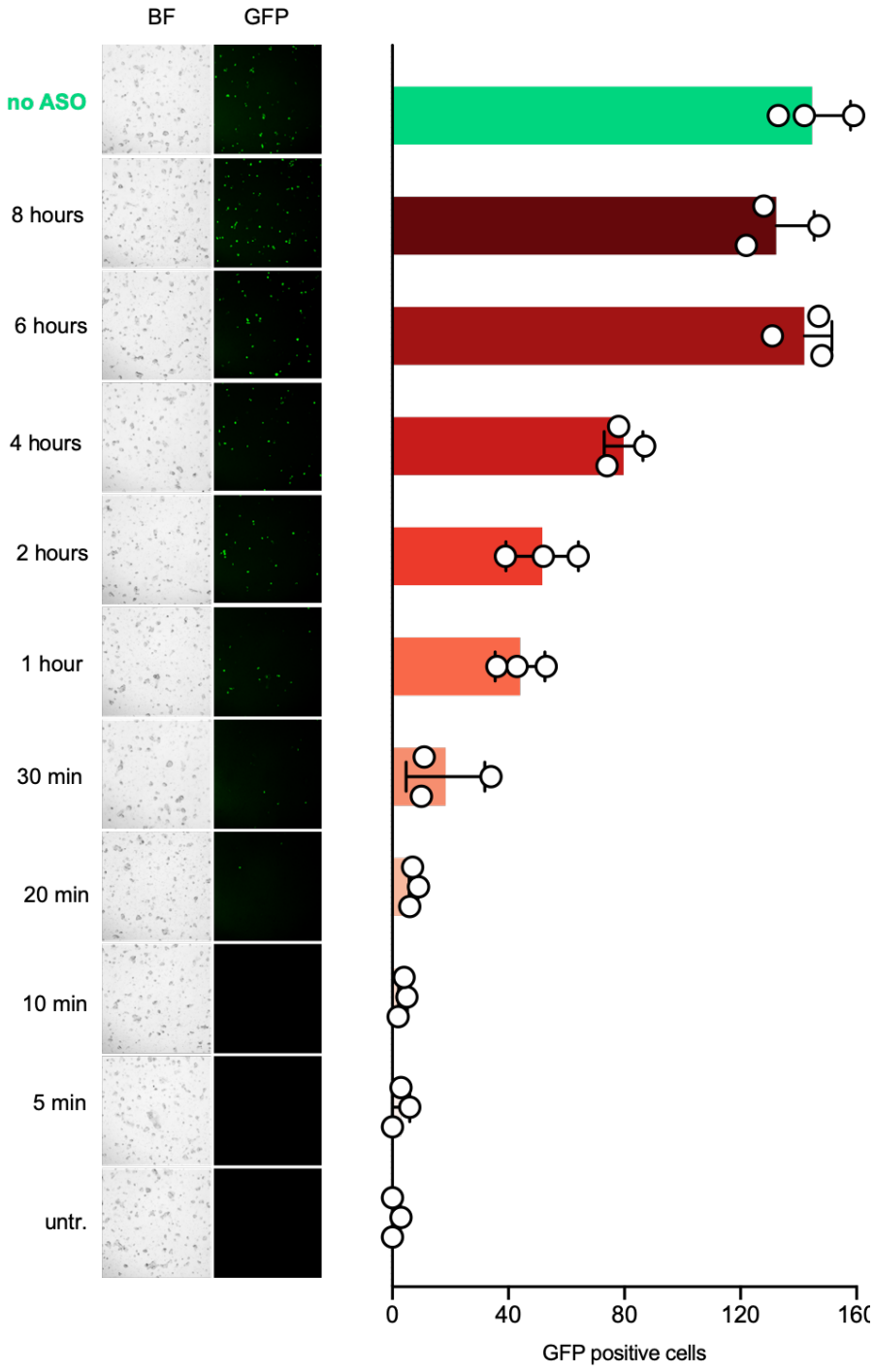


Figure E37. Optimizing ASO glyoxal treatment time for inhibiting gene silencing. HEK293T cells were co-transfected with both a pCMV-eGFP vector (200 ng) as well as 250 nM increasingly caged eGFP-targeting ASO. After 12 hours, expression was detected by live-cell imaging using a BioTek Lionheart FX automated microscope (4X magnification). GFP positive cells per field were counted using ImageJ, and values represent mean with S.D (n = 3 wells). BF = brightfield.

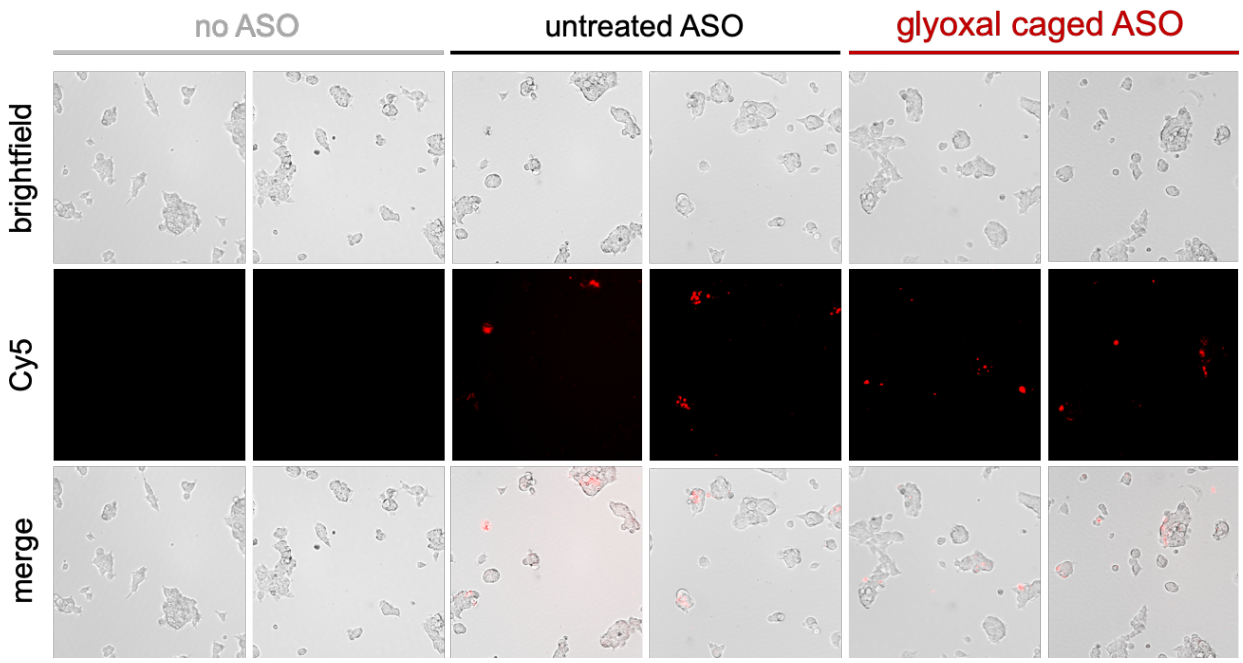


Figure E38. Untreated and glyoxal caged ASOs are cell permeable. LNA ASOs were first Cy5-labeled, followed by either no treatment or glyoxal caging for 8 h. HEK293T cells were incubated with 250 nM respective ASO for 18 hour. Media was then removed, and cells were washed 3x with fresh media. Cells were then imaged using a BioTek Lionheart FX automated microscope (10X magnification).

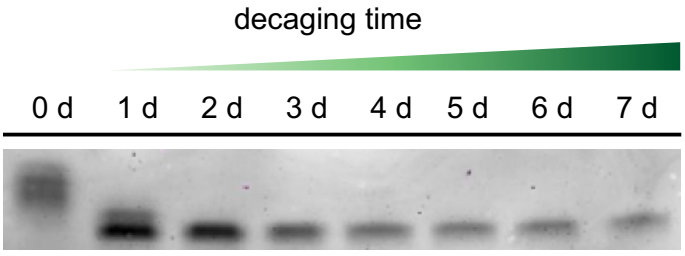


Figure E39. ASO glyoxalation is reversible with slow decaging. Decaging kinetics of fully caged (8 hour treatment time) ASO at 37 °C in complete DMEM were monitored by 20% PAGE gel illustrating decrease in apparent molecular weight.

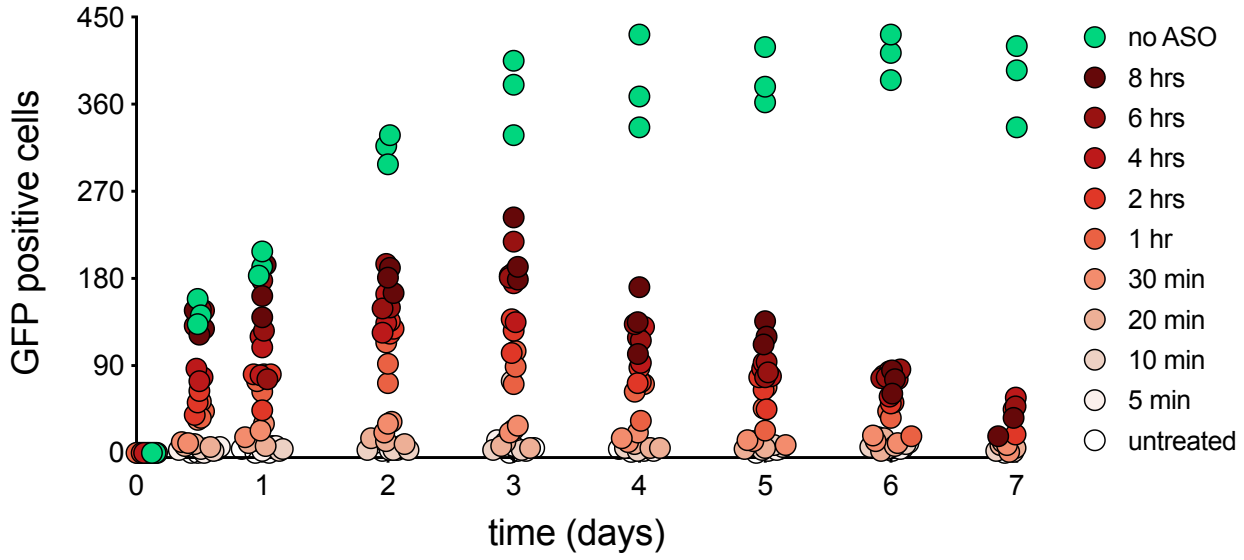
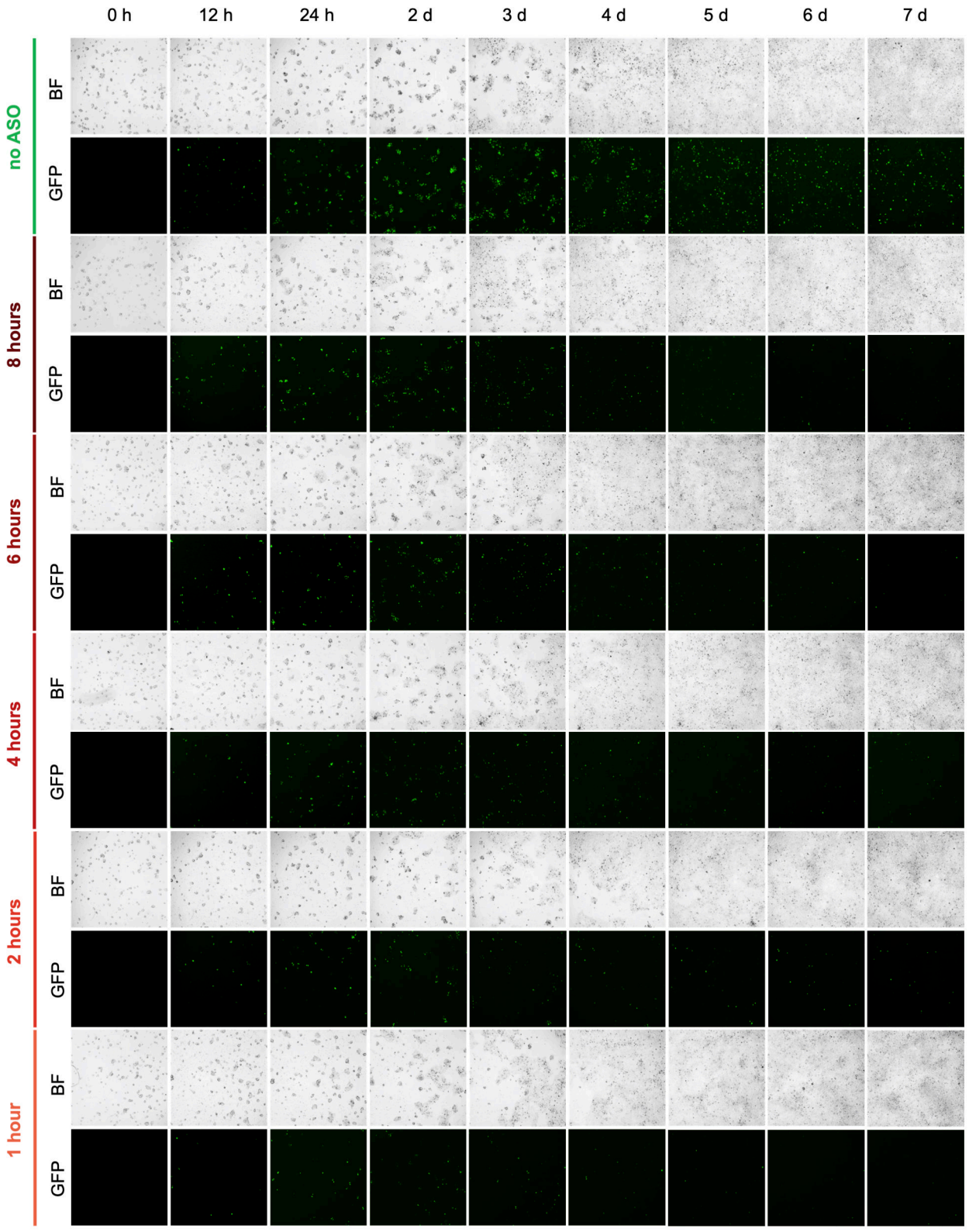


Figure E40. Increasingly caged ASO proportionally tunes the amount of GFP positive cells. At *t* = 0, HEK293T cells were transfected with a pCMV-GFP plasmid as well as 250 nM of untreated or increasingly glyoxal caged ASO. Graph shows quantification of GFP-positive cells in each field across treatment groups during the experimental time course. Circles represent individual wells (n = 3) from a 96-well plate.


0 h
12 h
24 h
2 d
3 d
4 d
5 d
6 d
7 d
no ASO
8 hours
6 hours
4 hours
2 hours
1 hour
BF
GFP

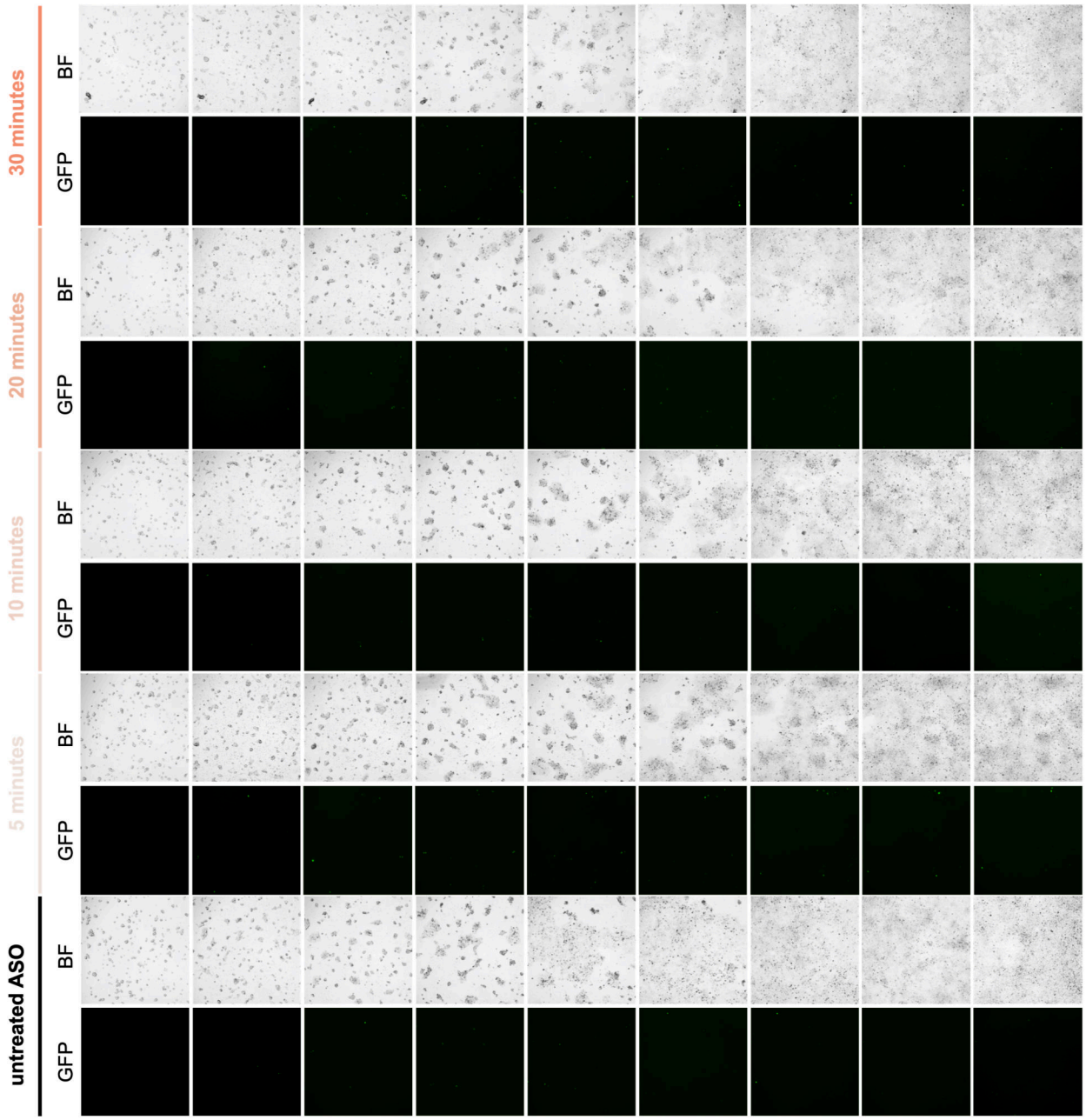


Figure E41. Representative live-cell fluorescence microscopy images during *in cellulo* ASO decaging. HEK293T cells were co-transfected with both a pCMV-eGFP vector (200 ng) as well as 250 nM increasingly caged eGFP-targeting ASO. At the indicated time points, expression was detected by live-cell imaging using a BioTek Lionheart FX automated microscope (4X magnification). BF = brightfield. Media was replaced on days 4 and 7.

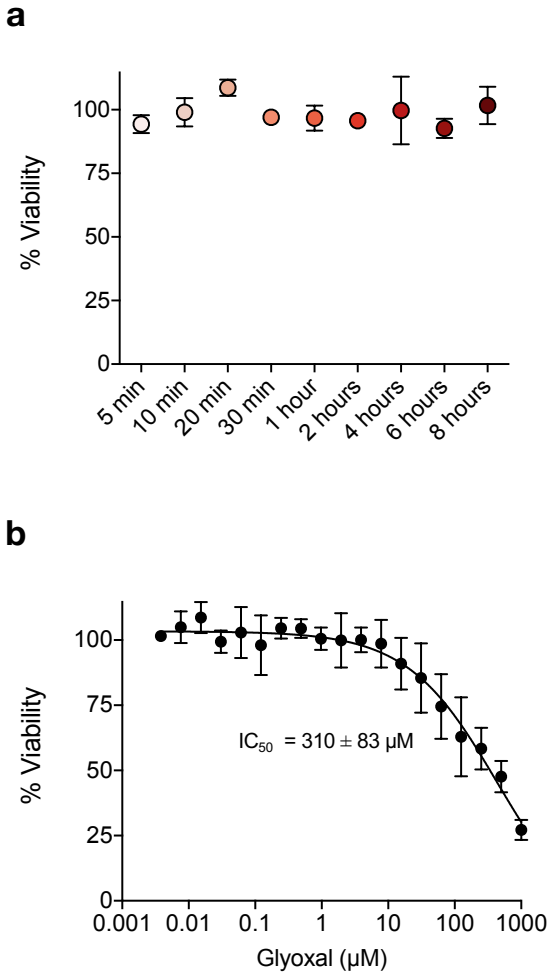


Figure E42. Caged ASOs do not affect cell viability. a) On day 7, plates from the time course decaging experiment were tested using a WST-1 assay to measure cell viability. Values represent mean and S.D. of 3 wells calculated as a percentage of untreated control cells. b) In parallel, HEK293T cells were seeded in a 96-well plate at 10,000 cells/well and incubated with a range of glyoxal concentrations for 7 days. Media was replaced on days 4 and 7, and viability was then measured using a WST-1 assay. Values represent mean and S.D. of 3 wells calculated as a percentage of untreated control cells. IC_{50} value (mean with 95% confidence interval) was calculate using a dose-response curve fit in Prism.