The Role of ADARs in Innate Immunity and its Utility as a Therapeutic Tool

By

AGYA KARKI
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Peter A. Beal, Chair

Marie Heffern

Enoch P. Baldwin

Committee in Charge

2023

Dedicated to my immigrant parents who scarified everything. My father migrated to the United States to give his daughters a chance for better education, and I am happy to give this dissertation as a gift to him. This is also dedicated to my beautiful sisters and dear friend Rajiv who have selflessly supported me. I will always cherish my mentors and friends I have met along the way. To Erin, Peishan, and my lab-mates thank you for special memories, laughter, and support along the journey. Thank you to Prince and Victorio for always feeding me during hard times, I will cherish the friendship. Special thanks to my lab-mates and collaborators Herra, Kristen, and Sukanya, without them this thesis would not be possible.

The Role of ADARs in Innate Immunity and its Utility as a Therapeutic Tool

Abstract

The innate immune system relies on molecular sensors to detect specific patterns, like viral double-stranded RNA (dsRNA), triggering responses such as apoptosis and immune infiltration. Adenosine Deaminases Acting on RNA (ADARs) play a critical role by catalyzing site-selective hydrolytic deamination of adenosine (A) to inosine (I). This mechanism helps distinguish self from non-self RNA, preventing erroneous immune activation. However, ADAR1 loss-of-function mutations lead to Aicardi Goutières Syndrome (AGS), a severe autoimmune disorder in children. While most AGS-associated mutations in ADAR1 are within its catalytic domain, their precise effects on adenosine deamination remain unclear. This dissertation investigates four AGS-causing mutations (G1007R, R892H, K999N, and Y1112F) in ADAR1 p110 and truncated variants. Measurement of adenosine deamination rates using RNA substrates derived from human transcripts edited by ADAR1 (hGli1, 5-HT2cR) reveals that mutations affecting amino acids stabilizing the base-flipped conformation of the ADAR-RNA complex (G1007R and R892H) have the most significant impact on catalysis. The K999N mutation, near the RNA binding interface, contextually alters catalytic activity, while the Y1112F mutation exerts minimal effects. The characterization of ADAR1 disease-associated mutations necessitated several optimizations in purification conditions, construct design, and protein-RNA binding assay. These findings shed light on the context-dependent effects of AGS-causing mutations on adenosine deamination, advancing our structural understanding of ADAR1-mediated RNA editing and how mutations in ADAR1 catalytic domain may lead to disease phenotype.

From another perspective, ADAR-mediated RNA editing can rectify disease-causing mutations using complementary guide RNAs. However, ADARs exhibit a preference against RNA

sites containing 5'G or 5'C adjacent to the edited adenosine (5'GA and 5'CA sites). Several structural studies in our lab have suggested a steric clash between the exocyclic amine of guanosine and a critical glycine residue (G489) in ADAR's flipping loop. Recent work from our lab showcases the restoration of ADAR activity using guanosine or adenosine nucleosides that pair with 5'G in a syn conformation. This is supported by two high-resolution crystal structures demonstrating how a G_{syn}-G_{anti} and G_{syn}-AH^{+}_{anti} pairs alleviate clashes. Following on this work this thesis outlines efforts to discover guide RNAs enabling editing at 5'CA sites by evading G489 in a similar manner. We identified guanosine and adenosine analogs capable of robustly and selectively editing these sites through non-Watson-Crick pairing. In addition, a high-throughput method was implemented to identify unique sequence motifs that enable more precise editing at 5'GA sites. This dissertation also highlights the efforts to structurally characterize these motifs through X-ray crystallography. These combined efforts inform guide RNA design for disease targets featuring unfavorable 5' sequence contexts.

TABLE OF CONTENTS

List of figures

List of Tables

List of Abbreviation

2-AP ..... 2-amino purine

5-HT_{2C} ..... 5-Hydroxytryptamine receptor 2C

8-azaN 8-azanebularine

AATD ..... Alpha-1 antitrypsin deficiency

ADAR ..... Adenosine Deaminase Acting on RNA

ADAR1d ..... Adenosine Deaminase of hADAR1

ADAR1 R3D .....
. N-terminal deletion mutant containing only one RNA binding domain of human ADAR1

ADAR1 R2D .....
. N-terminal deletion mutant containing only one RNA binding domain of human ADAR2

AGS ..... Aicardi Goutières Syndrome

ALS ..... Amyotrophic lateral sclerosis

AZIN1 ..... antizyme inhibitor 1

BME ..... β-mercaptoethanol

DSH ..... Dyschromatosis Symmetrica Hereditarian

dsRBD ..... Double-stranded RNA binding domains

dsRNA ..... Double stranded RNA

EMERGe ..... En Masse Evaluation of Guide RNA (EMERGe)

EMSA ..... Electro mobility shift assay

GluA2 ..... Glutamate inotropic receptor AMPA type subunit 2

hADAR2d ..... human ADAR2 deaminase domain

hGli1 ..... human glioma-associated oncogene1

IFN ..... Interferon

IP6 ..... Inositol hexakiphosphate

LINEs ..... Long interspersed elements

MBP ..... Maltose binding protein

MDA5 ..... Melanoma differentiation antigen 5

MECP2 ..... Methyl CPG binding protein 2

miRNA ..... Micro-RNA

PKR ..... Protein Kinase R

RNA ..... Ribonucleic acid

rRNA ..... Ribosomal RNA

SEC ..... Size exclusion column

S_NAR ..... Substitution-Nucleophilic Aromatic Reaction

snoRNA ..... Small nucleolar RNA

snRNA ..... Small nuclear RNA

T_M ..... Thermal melting temperature

tRNA ..... Transfer RNA

Z-DNA/RNA ..... Z binding domain

Chapter 1

RNA Editing by Adenosine Deaminases Acting on RNA and Immune Regulation

1.1 The Epitranscriptome.

Advances in detection methods have unveiled more than 150 modifications in RNA, many playing crucial roles in determining cellular functions and fate[1]. These modifications encompass chemical alterations and changes to different nucleobases, collectively referred to as the epitranscriptomics[2]. These dynamic changes, some reversible and others not, further enrich the vast sequence landscape of RNA[2]. While these alterations are primarily found in molecules like transfer RNA (tRNA), ribosomal RNA (rRNA), small nucleolar RNA (snoRNA), micro RNA (miRNA), and small nuclear RNA (snRNA), modern high-throughput sequencing techniques have also enabled their mapping in mRNA[1,2]. Understanding these modifications' precise functions and their links to diseases remains a dynamic and ongoing field of research[3].

The introduction and removal of RNA modifications involves a sequence of coordinated actions: writer enzymes introduce the modifications, reader enzymes interpret their effects, and eraser proteins often reverse these changes[2,3]. This complex interplay of modification processes intricately shapes RNA's regulatory mechanisms and its overall functionality. Among the most extensively studied RNA modifications are N6-methyladenosine (m6A), 5-methylcytosine (m5C), 2'-O-methylation (Nm), pseudouridine, and the conversion of adenosine (A) to inosine (I)[1,2,3]. Dysregulation of these modifications has been closely linked to various diseases. For example, disruptions in m6A pathways are associated with neurological disorders, cellular differentiation, and embryo development. Furthermore, the presence of pseudouridine has been shown to enhance

the evasion of synthetic mRNAs from the innate immune system while improving their stability. This strategic application was notably employed in COVID-19 vaccines to enhance their efficacy, as demonstrated by Kariko et al[4].

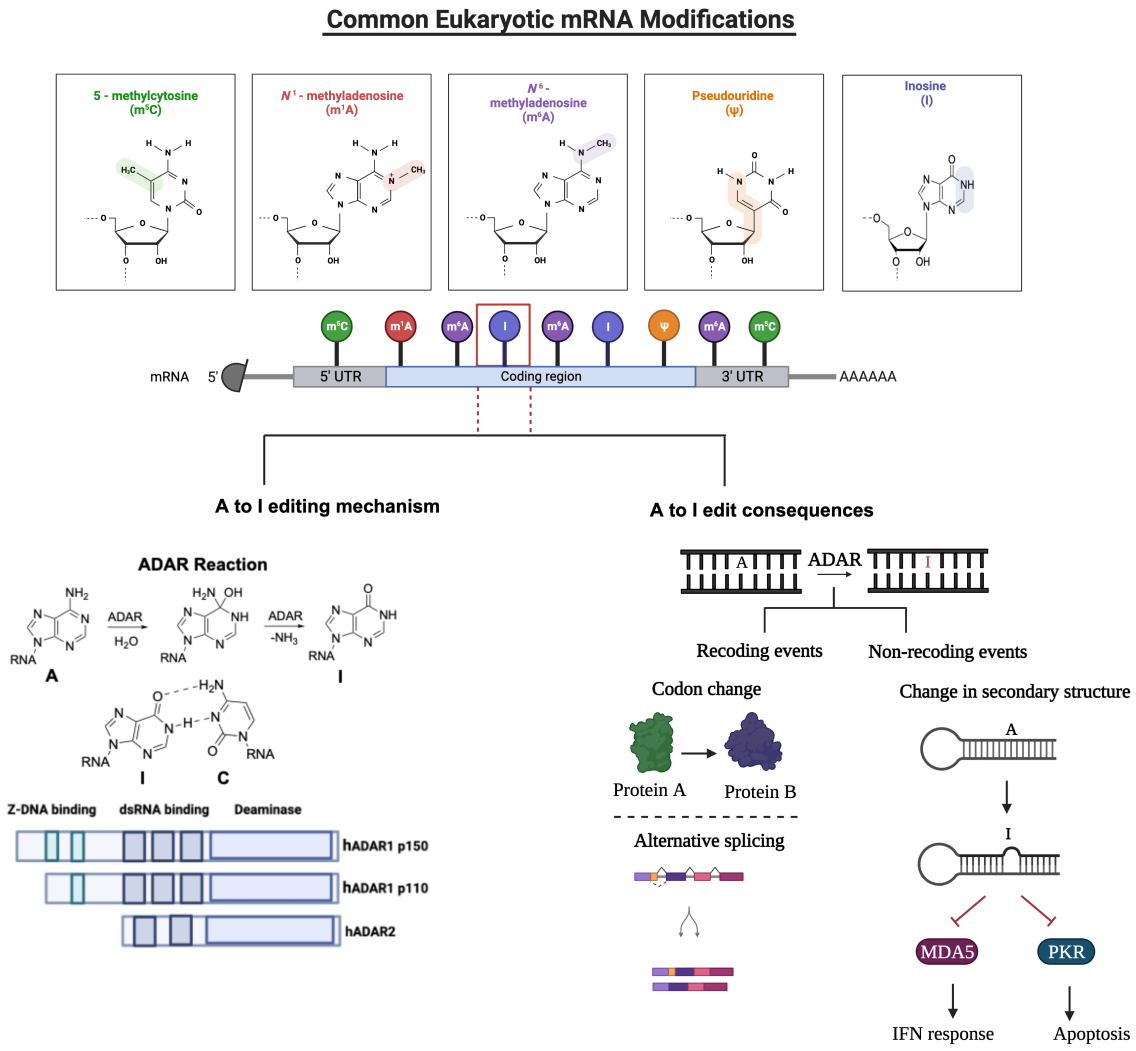


Figure 1.1. Highlighted are common modifications found within an mRNA that contribute to its structural diversity[1,2,3]. An interesting modification is the enzymatic conversion of adenosine to inosine catalyzed ADAR enzymes. Two active isoforms ADAR1 and ADAR2 carry out this reaction in dsRNA[6,7]. Reactions of ADARs lead to A to G transition mutations in mRNA, due to similarities of inosine to guanosine[6,7]. The consequences of these editing can lead to recoding events and secondary structural changes that regulate immune pathways[1-8,11].

1.2. RNA Editing by the Human ADAR Family.

Among these modifications, adenosine to inosine (A-to-I) deamination and cytosine to uridine (C-to-U) deamination present two unique examples in which the base pairing capabilities of the original nucleobase is altered[3,5]. The conversion of adenosine (A) to inosine (I), acting as a functional mimic of guanosine, is primarily catalyzed by Adenosine Deaminase Acting on RNA (ADARs) within their preferred double-stranded RNA substrates[6,7]. These enzymes were first discovered due to their role in unwinding double-stranded RNA resulting from A-to-I editing events in oocytes of *Xenopus laevis*[7]. ADARs perform an S_NAR (Substitution-Nucleophilic Aromatic-Reaction) addition-elimination reaction. During this process, a zinc-bound hydroxyl group initiates an attack on the C6 position of an adenosine. This attack leads to the removal of an ammonia molecule and the subsequent conversion of adenosine into inosine[6,7]. Notably, the hydrogen bonding face of inosine closely resembles that of guanosine, allowing it to preferentially base pair with cytosine. Consequently, ADAR-mediated editing results in A-to-G transitions at the RNA level. In humans, two key ADAR enzymes, ADAR1 and ADAR2, play central roles in this process. These enzymes exhibit a modular organizational structure. A conserved catalytic domain is located at their C-terminal region, while double-stranded RNA binding domains (dsRBDs) are positioned within the N-terminal region (**Figure 1.1**)[6,7]. ADAR1 exists in two distinct isoforms: the constitutively expressed short form, ADAR1 p110, and the interferon (IFN)-inducible form, p150. Importantly, p150 includes an additional Z-DNA/RNA (Zα) binding domain, which is crucial for substrate specificity and the editing of Z-RNAs[8,9]. Both ADAR1 isoforms can shuttle between the nucleus and the cytoplasm; however, ADAR1 p110 is predominantly localized in the nucleus, while p150 is primarily found in the cytoplasm[8,9]. Majority of the editing occurs in non-

coding sequences which include: 5' and 3' untranslated regions (UTRs), repeated elements such as Alu and long interspersed elements (LINEs)[8,9]. The significance of the repetitive RNAs and changes through RNA editing remains unclear. Furthermore, A-to-I editing can also change codon meaning and consequently lead to introduction of novel mutations. The recoding of glutamate inotropic receptor AMPA type subunit 2 (GluA2) was the first example of the A-to-I editing in the coding region. The Q/R site in GluA2 pre-mRNA regulates Ca^{2+} influx and is essential for healthy brain function[10]. In the Q/R site editing, a glutamine(Q) codon (CAG) is converted to an arginine(R) codon (CGG) by hADAR2 and this prevents calcium ion permeability in the ion channel. Hypo editing of this site results in neuronal death, epileptic seizures and is associated with the motor neurodegenerative disorder amyotrophic lateral sclerosis(ALS)[10,11,12,13]. Additionally, RNA editing plays a pivotal role in fine-tuning neurotransmission and receptor function. One prominent example is the serotonin receptor subtype 5-HT_{2C}, a G-protein-coupled receptor integral to the regulation of emotions, appetite, and motor behavior. Within a span of 13 nucleotides in its mRNA, five distinct edit sites (A to E) exist[14]. Complex, site-selective RNA editing events mediated by ADAR1 and ADAR2 enzymes give rise to multiple isoforms with altered amino acid sequences. This editing exerts a significant influence on G-protein coupling, subsequently affecting downstream signaling pathways and behavioral patterns[11]. Furthermore, the editing of the Gabra-3 transcript is believed to play a critical role in brain development and synapse formation. These examples highlight the recoding capabilities of ADARs and their role in human diseases[2,11].

1.3. ADAR1 and Innate Immunity.

Aside from the consequences mentioned above, ADAR editing is known to induce secondary structural perturbations. These changes in RNA lead to protection of dsRNA from key host-cell immune recognition receptors such as melanoma differentiation antigen 5 (MDA5) and protein kinase R (PKR)[15,16]. MDA5 is recognized for its role in inducing type I interferon production, while PKR activation leads to translation shutdown[17]. Therefore, the editing events by ADAR1 are a necessary component in distinguishing dsRNA as "self" or "non-self"[15,16,17]. By introducing structural modifications to RNA, ADAR1-mediated editing prevents the recognition of cellular RNA as foreign, thereby avoiding unnecessary immune responses[16]. While the roles of both ADAR enzymes are crucial, it's noteworthy that knockout of ADAR2 in mice leads to a susceptibility to seizures after 12 days, whereas ADAR1 knockout results in embryonic lethality[15,16]. However, this lethality can be circumvented by the deletion of the innate immune sensor, MDA5. Hence, emphasizing the primary responsibility of ADAR1, particularly its p150 isoform, in the regulation of cytoplasmic innate immunity at the level of dsRNA (Figure 2)[15,16,17].

Accordingly, loss-of-function mutations in the cytoplasmic ADAR1 p150 isoform are associated with the fatal immune disorder Aicardi Goutières Syndrome (AGS), and the non-fatal skin disorder Dyschromatosis Symmetrica Hereditarian (DSH). Both AGS and DSH are linked to specific mutations within the ADAR1 gene[16,18]. In the case of AGS, eight mutations have been identified, with seven of them mapping to the deaminase domain of the protein. Interestingly, only one of these mutations is found in the Z-DNA (Zα) binding domain of the protein[16,17,18]. Conversely, DSH is attributed to approximately 130 amino acid alterations. Both diseases share the common mutations G1007R and R892H, which are found as heterozygous mutations[18].

Structural analyses, based on the hADAR2 protein, suggest that these mutations could directly alter base-flipping and, consequently, catalytic activity[16,17,18,19]. Recent studies have also revealed that the AGS mutants can activate the MDA5 signaling pathway in both an editing-dependent and editing-independent manner[20,21,22]. Currently, the exact mechanisms underlying the heightened immune responses and clinical manifestations seen in these disorders are still under intensive investigation.[20,21,22]

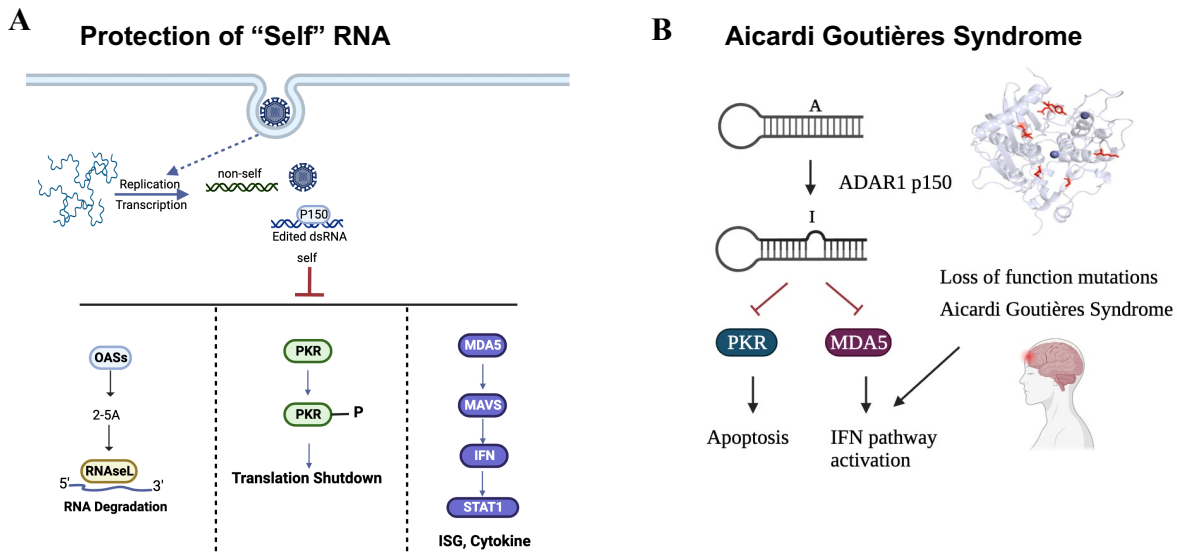


Figure 1.2. (A) Editing of dsRNA by ADAR1 prevents various immune activation pathways that lead to RNA degradation, translation shutdown, and immune infiltration. **(B)** Mutations within the ADAR1 p150 isoform leads to the AGS disease. These mutations cause MDA5 dependent immune activation.

1.4. ADAR1 and Cancer.

Loss-of-function mutations in ADAR1 are typically associated with immune disorders, whereas hyper-editing events have been linked to pro-oncogenic roles and are observed in various types of cancer [23]. Tumor environments often display elevated levels of interferon-stimulated

genes and type I interferon production, resulting in increased expression of IFN-inducible ADAR1 p150 and heightened editing of double-stranded RNA[23,24]. As previously discussed, these secondary structural changes enable dsRNA to evade detection by immune sensors like MDA5 and PKR[17]. For example, A-to-I RNA alterations in antizyme inhibitor 1 (AZIN1) is associated various cancer types, including hepatocellular carcinoma, non-small cell lung cancer, and colorectal cancer types. Hyper-editing of AZIN1 results in a gain-of-function phenotype associated with more aggressive tumors[11]. Another example is the editing of the R/G amino acid of human glioma-associated oncogene 1 (hGli1), where increased editing is linked to multiple myeloma[11]. Consequently, inhibiting ADAR1 is a compelling therapeutic strategy for cancer and an active area of research[25]. In fact, several studies have demonstrated that the deletion of ADAR1 can enhance sensitivity to immunotherapy. In contrast to ADAR1-associated immune disorders, the MDA5 pathway is not essential for ADAR dependency in cancer; instead, the predominant pathway involves the sensing of dsRNA by PKR[26]. The growth arrest phenotype in tumors is dependent on the recognition of dsRNA by PKR[23,26]. Companies such as Covant therapeutics and Exelixis are currently looking to discover inhibitors of ADAR1 that is intended to use in combination with other therapies.

1.5. Guided Site Directed RNA Editing by ADARs.

Genetic disorders resulting from single-point mutations account for a substantial 60% of all known disorders[11]. Emerging corrective technologies offer the promise of repairing these mutations at both the DNA and RNA levels[27]. Two particularly promising strategies revolve around gene therapy for genetic material replacement and gene editing for mutation reversal. In recent years, the remarkable adaptability of Cas9 RNA-programmed nucleases, coupled with

innovative fusions of inert Cas9 with refined base editing systems, has gained recognition[27,28]. While this technology presents enormous potential, it also raises concerns about permanent genomic alterations. Additionally, it relies on DNA repair mechanisms that can be inefficient and context-dependent[29]. This is compounded by the challenge of delivering large protein constructs alongside guide RNAs[29]. In response, RNA-level modifications offer a potentially safer, finely adjustable, and transient alternative. ADARs, with their widespread tissue expression and ability to accurately identify double-stranded RNA, are valuable for modifying specific adenosine residues within various transcripts[26,27]. This innovative approach is becoming increasingly popular as a compelling alternative to Cas9 genome editing and DNA base editors, effectively addressing delivery challenges and the inadvertent induction of DNA mutations[30]. Notably, G-to-A mutations constitute 28% of all pathogenic single-nucleotide variants, highlighting the significant potential of ADARs to directly reverse these specific point mutations[26,27]. Moreover, A-to-G conversions can result in 17 distinct amino acid substitutions, demonstrating how ADARs can play a pivotal role in restoring proper protein function in diseases characterized by missense mutations[11,26,27]. However, several challenges must be overcome before ADARs can be effectively applied in clinical settings. These challenges encompass the design of guide RNAs, ensuring precise delivery to target tissues, enhancing RNA stability, and optimizing the recruitment of ADARs to target sites, among other critical considerations[26,27]. Subsequent chapters will offer a comprehensive overview of the ongoing efforts dedicated to unraveling the intricacies of ADAR biology and refining guide RNA design, particularly for challenging target sites[26,27].

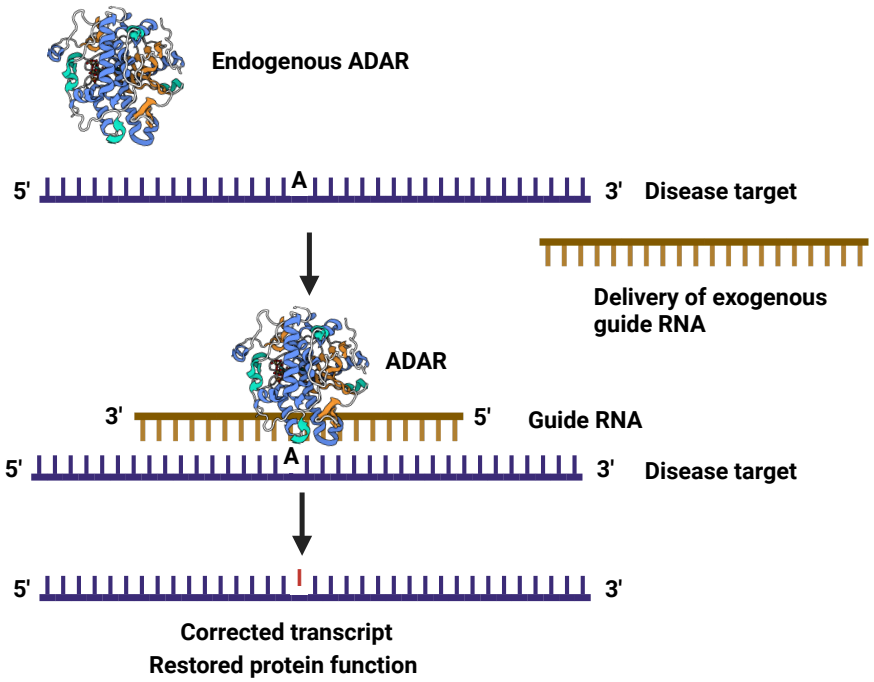


Figure 1.3. Recruitment of endogenous ADARs through delivery of exogenous guide RNA to revert G to A point mutations.

1.6. Structure of ADARs.

The effective design of guide RNAs relies on a comprehensive understanding of ADAR-RNA interactions. Three significant advancements have contributed to this understanding. The first breakthrough occurred in 2005 when the Bass lab successfully solved the structure of hADAR2's deaminase domain (ADAR2d)[31]. This structural revelation unveiled a zinc ion coordinated with a water molecule, accompanied by ligands H394, C451, and C516. In the catalytic process, E396 deprotonates the coordinated water molecule, facilitating the hydrolysis at the C6 position of the target adenosine. Furthermore, this structure confirmed the presence of an inositol hexakisphosphate (IP6) co-factor, playing a pivotal role in the folding and activity of

ADARs[31]. The second milestone in advancing our comprehension of ADAR-RNA interactions was achieved through the resolution of structures involving hADAR2's deaminase domain bound to short 23-base-pair RNA duplexes, a feat accomplished by our laboratory[32]. To effectively immobilize the protein-RNA complex, we utilized an adenosine analog called 8-azanebularine (8-azaN), where a hydrogen atom replaces the -NH_2 group[32,33]. Due to the absence of a suitable leaving group, the resulting hydrated 8-azaN product efficiently trapped the protein-RNA complex[33]. This methodology resulted in the diffraction of four distinct protein-RNA complexes[32].

The structures of ADAR2 bound to dsRNA revealed a 20-base pair segment of the duplex RNA that interacts with the deaminase domain primarily through the phosphodiester backbone[32]. This provided direct evidence that ADAR enzymes facilitate deamination by flipping the target adenosine onto the active site. For this process, ADAR approaches the RNA duplex from the minor groove, utilizing the conserved flipping loop of the protein comprising amino acid residues 487-489, including Gly487, E488, and G489[32,34,35]. To accommodate this intercalation, the duplex RNA deviates from the typical A form helix to a DNA like B form near the edit site. This event expands the major groove introducing a slight kink in the unedited strand inducing a 2' endo sugar pucker in the position across the nucleotide adjacent to edited A[32].

The conserved glycine residues in the flipping loop allow for flexibility and correct positioning of E488 in the space vacated by the flipped-out base. The confirmation of the flipped base is stabilized by a hydrogen bonding contact between E488 and the orphaned base[32,36,37]. This contact directly explains ADAR's sequence preferences, and the hyperactivity associated with its mutation to glutamine (E488Q)[14]. ADARs react preferentially with adenosines in A-C mismatches and A-U pairs[38]. A purine as a base pairing partner with adenosine would directly clash with the

E488 residue of the flipping loop. At the 3’ nearest neighbor nucleotide, ADARs favor a G. This preference is due to the hydrogen bonding contact of the carbonyl oxygen of S486 in ADAR2 with the 2-amino group of the edited A 3’ neighbor[32]. Only a 3’guanosine nucleobase would provide this stabilizing contact[32]. The 5’ nearest neighbor nucleotide preferences for ADARs show a strong bias against reacting with adenosines that have a guanosine or cytosine 5' to the target adenosine (5'-GA and 5’-CA sites). Modeling of a 5’G or a 5’C in the structures predict a detrimental clash between the 2-amino group of 5’G and the base pairing partner of 5’C with the Gly489 residue[32]. Indeed, replacement of a 5’U-A pair with 5’U-2aminopurine (2AP) results in 80% reduction in activity. The substitution of adenosine with 2AP introduces the amino group in the minor groove of the structure and causes destabilizing clash with G489[32]. Apart from the flipping loop, ADARs' substrate selectivity is influenced by their respective 5’ binding loops, which span residues 454-479. Despite sharing significant homology in the deaminase domain, ADAR1 and ADAR2 exhibit distinct substrate preferences[39,40]. These distinctions can be attributed to substantial differences in the sequences of the 5’ binding loops between ADAR1 and ADAR2. This is further substantiated by domain swapping experiments, where ADAR1's selectivity shifted to recognize ADAR2's preferred substrates upon altering its 5’ binding loop to match that of ADAR2[39,40].

The final advancement in understanding protein-RNA interactions comes from our lab’s recent structure of hADAR2 R2D bound to a 32 base pair duplex[36]. While ADARs are known to form both homo and hetero dimers, this study marked the first detailed characterization of the dimerization interface and its role in ADAR substrate recognition. Dimerization enables one domain to engage in catalysis, while the other positions its dsRBD to interact with the RNA duplex 3’ to the editing site[36]. Disrupting dimerization, by mutagenesis of the conserved contact interface,

resulted in decreased editing of certain dimer-dependent substrates[36]. Furthermore, the structure unveiled critical contacts at both the 5' and 3' ends of the edited site, essential for catalytic activity. Consequently, this research significantly informs the design of guide RNAs for therapeutic ADAR editing[37,41,42,43]. Unfortunately, complete structures for ADAR2 and high-resolution ADAR1 structures remain elusive. Nevertheless, our lab's high-throughput mutagenesis studies and homology modeling of ADAR1 have led to the discovery of a structural zinc binding site in ADAR1, suggesting its importance for protein folding and catalysis[40]. To comprehensively unravel ADAR1 substrate recognition, further structural biology efforts are essential.

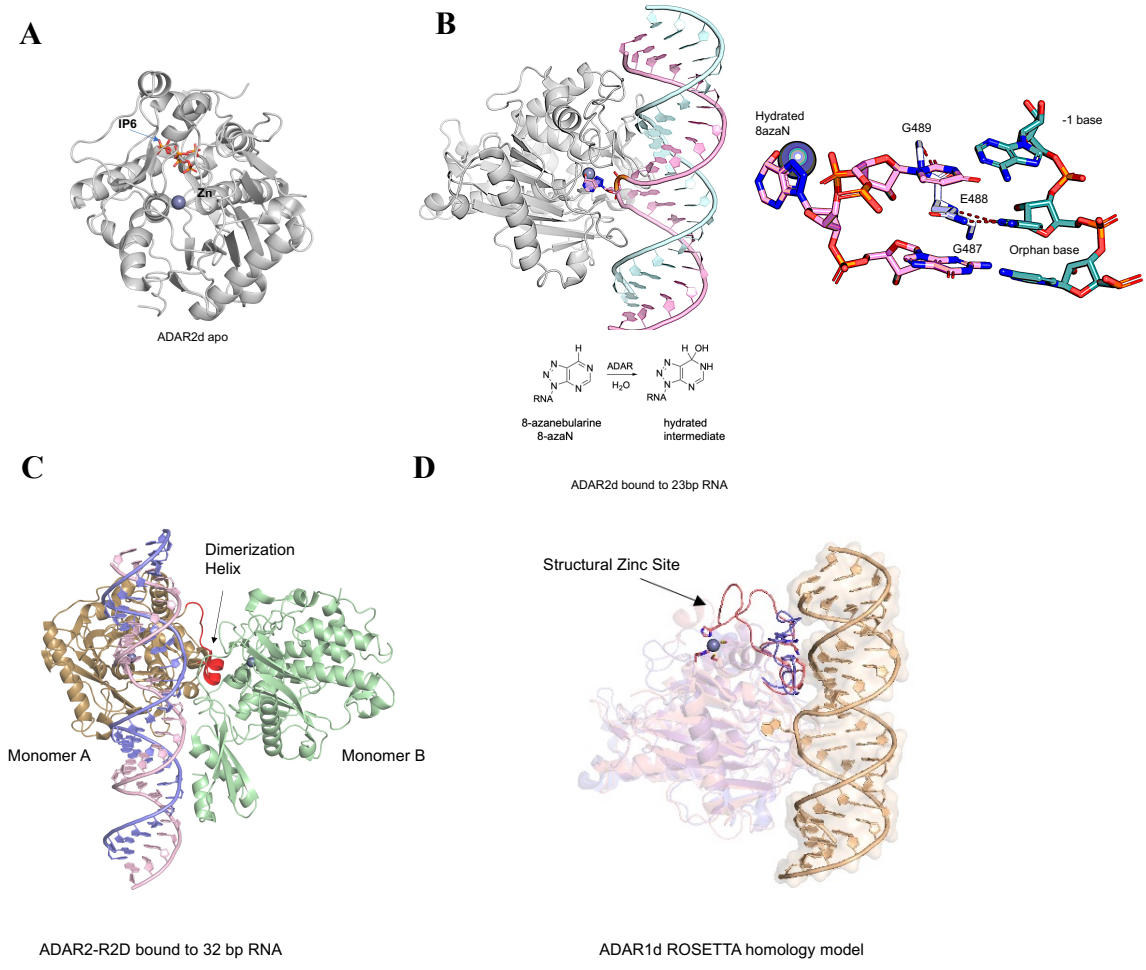


Figure 1.4. (A) The structure of ADAR2 deaminase domain (PDB:IZYF)[31]. **(B)** First published structure of ADAR2d bound a short dsRNA substrate, confirming its base-flipping mechanism and stabilizing contacts with its flipping loop (PDB: 5ED2)[32]. **(C)** The structure of ADAR2 R2D bound to a 32 bp RNA duplex revealing ADAR2 dimerization interface (PDB: 6VFF)[36]. (D) Predicted Rosetta-based homology model of ADAR1 based on the structure of hADAR2d. The structure displays the predicted second zinc binding site of ADAR1[40].

1.7. Summary and dissertation overview.

This dissertation seeks to address existing gaps in our comprehension of ADAR-RNA interactions and, more broadly, the involvement of ADAR1 in innate immunity and immune-related disorders. Chapter 2 is dedicated to exploring the impact of Aicardi Goutières Syndrome (AGS)-causing mutations on ADAR1's biochemical properties. This project involved refining purification techniques, conducting binding experiments, and developing truncated ADAR1 constructs. Using this method, the consequences of AGS mutations on RNA editing, substrate engagement, and thermal stability was determined. Chapter 3 provides a comprehensive overview of ADAR1 purification procedures. Currently, RNA-guided ADAR editing is constrained by its preferences for specific sequences located 5' to the editing site, as previously discussed. Chapters 4 and 5 present strategies to enhance editing efficiency in these regions by modifying the base-pairing partner of the 5' nucleobase adjacent to the edited A. Additionally, these chapters elucidate the rationale behind these guide designs and their effectiveness through the presentation of two high-resolution structures. Chapter 4 presents two high-resolution structures of an RNA duplex bound to hADAR2. These structures incorporate modifications that boost editing at the 5'GA sequence context. The methods employed for purification and crystallization have also been extended to study sugar modifications, enhancing dsRBD engagement and editing efficiency.

Kristen Campbell in the Beal lab spearheaded this work. Furthermore, we obtained an 8-angstrom resolution structure of RNA duplex containing unique sequence motifs that enable selective editing at 5'GA sites. In chapter 5, I discuss a high-throughput approach utilized to determine these sequence motifs, a project led by a Beal lab member, Prince Salvador. This chapter also covers efforts to determine guide RNA modifications that enable efficient editing at 5'CA sites. Overall, this dissertation contributes to our understanding of ADAR-RNA interactions and presents strategies to improve RNA editing, advancing our ability to develop therapeutic interventions.

1.8. References.

(1) Begik, O.; Mattick, J. S.; Novoa, E. M. Exploring the Epitranscriptome by Native RNA Sequencing. *RNA* **2022**, *28* (11), 1430–1439.
(2) Christofi, T.; Zaravinos, A. RNA Editing in the Forefront of Epitranscriptomics and Human Health. *J. Transl. Med.* **2019**, *17* (1), 319.
(3) Xiong, X.; Yi, C.; Peng, J. Epitranscriptomics: Toward A Better Understanding of RNA Modifications. *Genomics Proteomics Bioinformatics* **2017**, *15* (3), 147–153.
(4) Karikó, K.; Muramatsu, H.; Welsh, F. A.; Ludwig, J.; Kato, H.; Akira, S.; Weissman, D. Incorporation of Pseudouridine into MRNA Yields Superior Nonimmunogenic Vector With Increased Translational Capacity and Biological Stability. *Mol. Ther.* **2008**, *16* (11), 1833–1840.
(5) Schaefer, M.; Kapoor, U.; Jantsch, M. F. Understanding RNA Modifications: The Promises and Technological Bottlenecks of the 'Epitranscriptome.' *Open Biol.* **2017**, *7* (5), 170077.
(6) Wang, Y.; Zheng, Y.; Beal, P. A. Adenosine Deaminases That Act on RNA (ADARs). **2017**, 215–268.
(7) Bass, B. L. RNA Editing by Adenosine Deaminases That Act on RNA. *Annu. Rev. Biochem.* **2002**, *71* (1), 817–846.
(8) Nie, Y.; Zhao, Q.; Su, Y.; Yang, J.-H. Subcellular Distribution of ADAR1 Isoforms Is Synergistically Determined by Three Nuclear Discrimination Signals and a Regulatory Motif. *J. Biol. Chem.* **2004**, *279* (13), 13249–13255.
(9) Kleinova, R.; Rajendra, V.; Leuchtenberger, A. F.; Lo Giudice, C.; Vesely, C.; Kapoor, U.; Tanzer, A.; Derdak, S.; Picardi, E.; Jantsch, M. F. The ADAR1 Editome Reveals Drivers of Editing-Specificity for ADAR1-Isoforms. *Nucleic Acids Res.* **2023**, *51* (9), 4191–4207.
(10) Tan, B. Z.; Huang, H.; Lam, R.; Soong, T. W. Dynamic Regulation of RNA Editing of Ion Channels and Receptors in the Mammalian Nervous System. *Mol. Brain* **2009**, *2* (1), 13.

(11) Goldeck, M.; Gopal, A.; Jantsch, M. F.; Mansouri Khosravi, H. R.; Rajendra, V.; Vesely, C. How RNA Editing Keeps an I on Physiology. *Am. J. Physiol.-Cell Physiol.* **2022**, *323* (5), C1496–C1511. https://doi.org/10.1152/ajpcell.00191.2022.
(12) Heraud-Farlow, J. E.; Walkley, C. R. What Do Editors Do? Understanding the Physiological Functions of A-to-I RNA Editing by Adenosine Deaminase Acting on RNAs. *Open Biol.* **2020**, *10* (7).
(13) Slotkin, W.; Nishikura, K. Adenosine-to-Inosine RNA Editing and Human Disease. *Genome Med.* **2013**, *5* (11), 105.
(14) Kuttan, A.; Bass, B. L. Mechanistic Insights into Editing-Site Specificity of ADARs. *Proc. Natl. Acad. Sci.* **2012**, *109* (48).
(15) Eisenberg, E.; Levanon, E. Y. A-to-I RNA Editing — Immune Protector and Transcriptome Diversifier. *Nat. Rev. Genet.* **2018**, *19* (8), 473–490.
(16) Lamers, M. M.; van den Hoogen, B. G.; Haagmans, B. L. ADAR1: "Editor-in-Chief" of Cytoplasmic Innate Immunity. *Front. Immunol.* **2019**, *10*.
(17) 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.
(18) Rice, G. I.; Kasher, P. R.; Forte, G. M. A.; Mannion, N. M.; Greenwood, S. M.; Szynkiewicz, M.; Dickerson, J. E.; Bhaskar, S. S.; Zampini, M.; Briggs, T. A.; Jenkinson, E. M.; Bacino, C. A.; Battini, R.; Bertini, E.; Brogan, P. A.; Brueton, L. A.; Carpanelli, M.; De Laet, C.; De Lonlay, P.; Del Toro, M.; Desguerre, I.; Fazzi, E.; Garcia-Cazorla, À.; Heiberg, A.; Kawaguchi, M.; Kumar, R.; Lin, J.-P. S.-M.; Lourenco, C. M.; Male, A. M.; Marques, W.; Mignot, C.; Olivieri, I.; Orcesi, S.; Prabhakar, P.; Rasmussen, M.; Robinson, R. A.; Rozenberg, F.; Schmidt, J. L.; Steindl, K.; Tan, T. Y.; Van Der Merwe, W. G.; Vanderver, A.; Vassallo, G.; Wakeling, E. L.; Wassmer, E.; Whittaker, E.; Livingston, J. H.; Lebon, P.; Suzuki, T.; McLaughlin, P. J.; Keegan, L. P.; O'Connell, M. A.; Lovell, S. C.; Crow, Y. J. Mutations in ADAR1 Cause Aicardi-Goutières Syndrome Associated with a Type I Interferon Signature. *Nat. Genet.* **2012**, *44* (11), 1243–1248.
(19) Wang, P.; Yu, S.; Liu, J.; Zhang, D.; Kang, X. Seven Novel Mutations of *ADAR* in Multi-ethnic Pedigrees with Dyschromatosis Symmetrica Hereditaria in China. *Mol. Genet. Genomic Med.* **2019**, *7* (10), e00905.
(20) de Reuver, R.; Dierick, E.; Wiernicki, B.; Staes, K.; Seys, L.; De Meester, E.; Muyldermans, T.; Botzki, A.; Lambrecht, B. N.; Van Nieuwerburgh, F.; Vandenabeele, P.; Maelfait, J. ADAR1 Interaction with Z-RNA Promotes Editing of Endogenous Double-Stranded RNA and Prevents MDA5-Dependent Immune Activation. *Cell Rep.* **2021**, *36* (6), 109500.
(21) Guo, X.; Steinman, R. A.; Sheng, Y.; Cao, G.; Wiley, C. A.; Wang, Q. An AGS-Associated Mutation in ADAR1 Catalytic Domain Results in Early-Onset and MDA5-Dependent Encephalopathy with IFN Pathway Activation in the Brain. *J. Neuroinflammation* **2022**, *19* (1), 285.
(22) Guo, X.; Wiley, C. A.; Steinman, R. A.; Sheng, Y.; Ji, B.; Wang, J.; Zhang, L.; Wang, T.; Zenatai, M.; Billiar, T. R.; Wang, Q. Aicardi-Goutières Syndrome-Associated Mutation at

ADAR1 Gene Locus Activates Innate Immune Response in Mouse Brain. *J. Neuroinflammation* **2021**, *18* (1), 169.
(23) Gannon, H. S.; Zou, T.; Kiessling, M. K.; Gao, G. F.; Cai, D.; Choi, P. S.; Ivan, A. P.; Buchumenski, I.; Berger, A. C.; Goldstein, J. T.; Cherniack, A. D.; Vazquez, F.; Tsherniak, A.; Levanon, E. Y.; Hahn, W. C.; Meyerson, M. Identification of ADAR1 Adenosine Deaminase Dependency in a Subset of Cancer Cells. *Nat. Commun.* **2018**, *9* (1), 5450.
(24) Liu, J.; Wang, F.; Zhang, Y.; Liu, J.; Zhao, B. ADAR1-Mediated RNA Editing and Its Role in Cancer. *Front. Cell Dev. Biol.* **2022**, *10*.
(25) Kung, C.-P.; Cottrell, K. A.; Ryu, S.; Bramel, E. R.; Kladney, R. D.; Bao, E. A.; Freeman, E. C.; Sabloak, T.; Maggi, L.; Weber, J. D. Evaluating the Therapeutic Potential of ADAR1 Inhibition for Triple-Negative Breast Cancer. *Oncogene* **2021**, *40* (1), 189–202.
(26) 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.; Buchumenski, I.; Comstock, D. E.; Brown, F. D.; Ayer, A.; Kohnle, I. C.; Pope, H. W.; Zimmer, M. D.; Sen, D. R.; Lane-Reticker, S. K.; Robitschek, E. J.; Griffin, G. K.; Collins, N. B.; Long, A. H.; Doench, J. G.; Kozono, D.; Levanon, E. Y.; Haining, W. N. Loss of ADAR1 in Tumours Overcomes Resistance to Immune Checkpoint Blockade. *Nature* **2019**, *565* (7737), 43–48.
(27) Booth, B. J.; Nourreddine, S.; Katrekar, D.; Savva, Y.; Bose, D.; Long, T. J.; Huss, D. J.; Mali, P. RNA Editing: Expanding the Potential of RNA Therapeutics. *Mol. Ther.* **2023**, *31* (6), 1533–1549.
(28) Khosravi, H. M.; Jantsch, M. F. Site-Directed RNA Editing: Recent Advances and Open Challenges. *RNA Biol.* **2021**, *18* (sup1), 41–50.
(29) Liu, W.; Li, L.; Jiang, J.; Wu, M.; Lin, P. Applications and Challenges of CRISPR-Cas Gene-Editing to Disease Treatment in Clinics. *Precis. Clin. Med.* **2021**, *4* (3), 179–191.
(30) Jeong, Y. K.; Song, B.; Bae, S. Current Status and Challenges of DNA Base Editing Tools. *Mol. Ther.* **2020**, *28* (9), 1938–1952.
(31) 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.
(32) Matthews, M. M.; Thomas, J. M.; Zheng, Y.; Tran, K.; Phelps, K. J.; Scott, A. I.; Havel, J.; Fisher, A. J.; Beal, P. A. Structures of Human ADAR2 Bound to DsRNA Reveal Base-Flipping Mechanism and Basis for Site Selectivity. *Nat. Struct. Mol. Biol.* **2016**, *23* (5), 426–433.
(33) Haudenschild, B. L.; Maydanovych, O.; Véliz, E. A.; Macbeth, M. R.; Bass, B. L.; Beal, P. A. A Transition State Analogue for an RNA-Editing Reaction. *J. Am. Chem. Soc.* **2004**, *126* (36), 11213–11219.
(34) Wong, S. K.; Sato, S.; Lazinski, D. W. Substrate Recognition by ADAR1 and ADAR2. *RNA* **2001**, *7* (6), 846–858.
(35) Thomas, J. M.; Beal, P. A. How Do ADARs Bind RNA? New Protein-RNA Structures Illuminate Substrate Recognition by the RNA Editing ADARs. *BioEssays* **2017**, *39* (4), 1600187.

(36) 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 Res.* **2020**, *48* (14), 7958–7972.
(37) Doherty, E. E.; Wilcox, X. E.; van Sint Fiet, L.; Kemmel, C.; Turunen, J. J.; Klein, B.; Tantillo, D. J.; Fisher, A. J.; Beal, P. A. Rational Design of RNA Editing Guide Strands: Cytidine Analogs at the Orphan Position. *J. Am. Chem. Soc.* **2021**, *143* (18), 6865–6876.
(38) Eggington, J. M.; Greene, T.; Bass, B. L. Predicting Sites of ADAR Editing in Double-Stranded RNA. *Nat. Commun.* **2011**, *2* (1), 319.
(39) Wang, Y.; Park, S.; Beal, P. A. Selective Recognition of RNA Substrates by ADAR Deaminase Domains. *Biochemistry* **2018**, *57* (10), 1640–1651.
(40) Park, S.; Doherty, E. E.; Xie, Y.; Padyana, A. K.; Fang, F.; Zhang, Y.; Karki, A.; Lebrilla, C. B.; Siegel, J. B.; Beal, P. A. High-Throughput Mutagenesis Reveals Unique Structural Features of Human ADAR1. *Nat. Commun.* **2020**, *11* (1), 5130.
(41) Doherty, E. E.; Karki, A.; Wilcox, X. E.; Mendoza, H. G.; Manjunath, A.; Matos, V. J.; Fisher, A. J.; Beal, P. A. ADAR Activation by Inducing a Syn Conformation at Guanosine Adjacent to an Editing Site. *Nucleic Acids Res.* **2022**, *50* (19), 10857–10868.
(42) Latifi, N.; Mack, A. M.; Tellioglu, I.; Di Giorgio, S.; Stafforst, T. Precise and Efficient C-to-U RNA Base Editing with SNAP-CDAR-S. *Nucleic Acids Res.* **2023**.
(43) Monian, P.; Shivalila, C.; Lu, G.; Shimizu, M.; Boulay, D.; Bussow, K.; Byrne, M.; Bezigian, A.; Chatterjee, A.; Chew, D.; Desai, J.; Favaloro, F.; Godfrey, J.; Hoss, A.; Iwamoto, N.; Kawamoto, T.; Kumarasamy, J.; Lamattina, A.; Lindsey, A.; Liu, F.; Looby, R.; Marappan, S.; Metterville, J.; Murphy, R.; Rossi, J.; Pu, T.; Bhattarai, B.; Standley, S.; Tripathi, S.; Yang, H.; Yin, Y.; Yu, H.; Zhou, C.; Apponi, L. H.; Kandasamy, P.; Vargeese, C. Endogenous ADAR-Mediated RNA Editing in Non-Human Primates Using Stereopure Chemically Modified Oligonucleotides. *Nat. Biotechnol.* **2022**, *40* (7), 1093–1102.

Chapter 2

Effects of Aicardi Goutières Syndrome-causing Mutations on Catalysis of Adenosine Deamination by ADAR1

This work has been submitted to Biochemistry and done in collaboration with Sukanya Mozumder, and Kristen M. Campbell.

2.1. Introduction

RNA editing is an essential post-transcriptional modification process that regulates gene expression and generates protein diversity in eukaryotic systems. It involves the precise insertion, deletion, or modification of nucleotides within RNA molecules.[1] Among the various types of RNA editing, the most prevalent form in metazoans is the hydrolytic deamination of adenosine (A) to inosine (I) catalyzed by the Adenosine Deaminases Acting on double-stranded RNA (ADARs).[1,2,3] Inosine exhibits similar base pairing properties to that of guanosine, resulting in A to G transitions that alter RNA structure and function. Two active ADAR enzymes in humans (ADAR1 and ADAR2) exhibit a modular organization characterized by a conserved catalytic domain located at the C-terminal region and double-stranded RNA binding domains (dsRBDs) positioned within the N-terminal region.[2] ADAR1 exists as two distinct isoforms: the constitutively expressed short form, ADAR1 p110, and the interferon (IFN)-inducible form, p150 which consists of an additional Z-DNA/RNA (Zα) binding domain. The Zα domain of ADAR1 is important for substrate specificity and editing of Z-RNAs.[4,5,6] While both forms can shuttle between the nucleus and the cytoplasm, ADAR1 p110 is primarily localized in the nucleus, whereas p150 is predominantly found in the cytoplasm.[7] Despite their high degree of similarity, ADAR1 and ADAR2 exhibit divergent substrate selectivity, which can be attributed to differences in their RNA binding domains and the distinct features of their catalytic domains.[8] Notably, they possess dissimilar

loops in their catalytic domains that bind RNA on the 5' side of targeted adenosines (i.e. 5' binding loops), and the ADAR1 catalytic domain additionally harbors a structural zinc site absent in ADAR2.[9]

Proper regulation of ADAR-mediated A to I editing is crucial for maintaining normal cellular function.[10,11] Abnormal expression and excessive editing by ADAR1 have been observed in various cancer types, indicating its involvement in oncogenesis.[12] Interestingly, the knockout of ADAR1 in tumor cells leads to increased sensitivity to immunotherapy, highlighting its potential as a therapeutic target.[13,14] ADAR1 also plays a critical role in regulating cytoplasmic innate immunity at the level of double-stranded RNA (dsRNA).[15,16] By introducing structural modifications to RNA, ADAR1-mediated editing prevents the recognition of cellular RNA as foreign, thereby avoiding unnecessary immune responses. Consequently, loss-of-function mutations in ADAR1 have been linked to a severe childhood immune disorder called Aicardi Goutières Syndrome (AGS).[17,18,19] Multiple AGS-associated mutations have been identified in the ADAR1 gene, with seven occurring in the deaminase domain (A870T, I872T, R892H, K999N, G1007R, Y1112F, D1113H,) and one (P193A) in the Zα domain (**Figure 2.1**).[15,16,17] Notably, among the AGS mutations found in the deaminase domain, only two (G1007R and R892H) are predicted to involve residues that directly contact RNA given the available substrate-bound ADAR2 structures (PDBID 5ED2, 5ED1, 5HP2, 5HP3, and 6VFF).[20,21]

Structures of ADAR2 bound to dsRNA revealed a 20-base pair segment of the duplex RNA that interacts with the deaminase domain primarily through the phosphodiester backbone.[20,21] ADARs access the target nucleotide in duplex RNA by inducing a base flipping conformational change that positions the adenosine in the active site.[20,21] A loop of the protein whose sequence is conserved within the ADAR family inserts into the minor groove of the duplex. This flipping loop bears the residue E488 in ADAR2 (corresponding to E1008 in ADAR1) that plays a crucial role by contacting the "orphan base" when the reactive adenosine flips into the active site. Once in the active site, the adenosine reacts with a hydroxide ion bound to zinc. The resulting high energy tetrahedral intermediate loses ammonia forming the inosine product. Within the flipping loop, the E488 residue is flanked by two conserved glycine residues, G487 (G1007, ADAR1) and G489 (G1009, ADAR1), which allow for correct positioning of E488 in the space vacated by the flipped out base and hydrogen bonding with the orphaned base.[20,22]

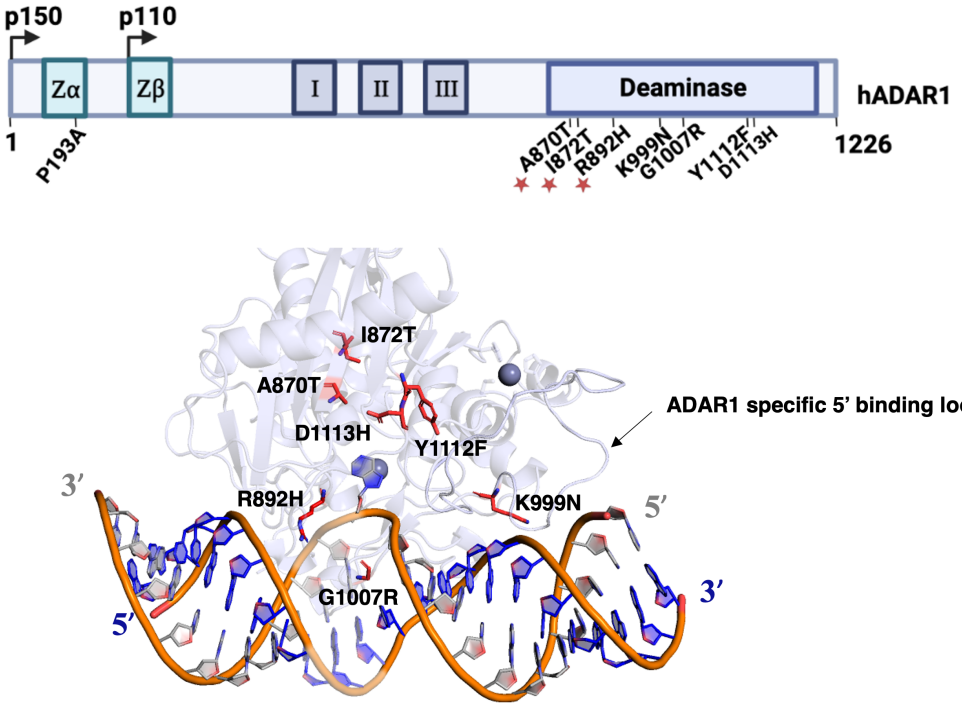


Figure 2.1. Mapping AGS mutations onto a model of ADAR1 deaminase domain bound to RNA. **(A)** Domain map of human ADAR1 p110 and p150. AGS mutations identified within the Zα domain and the catalytic domain of ADAR1 p150 are labeled. The mutations A870T, I872T, and R892H occur as double mutants with P193A and are denoted by asterisks. **(B)** Previously reported Rosetta model of ADAR1 deaminase domain bound to a dsRNA substrate.[9] AGS mutations within the deaminase domain of ADAR1 are highlighted in red. The ADAR1 specific 5' binding loop is also depicted. The active site zinc and secondary zinc are represented as dark grey spheres.

Among the identified AGS mutations, G1007R is predicted to have the greatest effect on adenosine deamination given its proximity to E1008, an essential residue in the flipping loop, and the substantial change in the side chain structure arising from the arginine for glycine substitution.[20,22] R892 corresponds to K376 in ADAR2, which contacts phosphodiesters adjacent to the edited adenosine, stabilizing the flipped out conformation. A similar contact is predicted for R892 and the mutation to histidine could alter the enzyme's ability to stabilize that essential conformation for catalysis.[20,22] In the available homology model of ADAR1, the methylene groups of K999 are predicted to make hydrophobic contacts that stabilize the 5'-binding loop.[9] Destabilization of the loop structure could result in inefficient engagement for certain RNA substrates that rely on 5' binding loop contacts. Finally, the Y1112F mutant, categorized as a second-layer mutation given its distance from the RNA-binding interface, may also influence the stability of the adjacent ADAR1-specific 5' binding loop .[9,22]

Crystal structures of ADAR2 bound to dsRNA provide insights into the possible consequences of AGS mutations. However, limitations arise from structural differences between ADAR1 and ADAR2 and the absence of a published high-resolution ADAR1 structure. Several studies have been conducted to investigate the impact of AGS mutants on ADAR1 function in cells and in animal models. Overexpression of AGS mutants of ADAR1 in HEK293

cells have provided insight into their effects on RNA editing.[17,18] Mouse studies, specifically examining AGS mutants P193A, K999N, and D1113H, have demonstrated distinct effects on ADAR1 substrates and tissue-specific immune activation.[23,24] Notably, the deletion of MDA5 has been shown to restore normal immune response in these mutants.[16,18,24] However, despite these investigations, a comprehensive biochemical understanding of how each mutation affects ADAR catalysis of adenosine deamination is lacking. Therefore, the objective of this study was to bridge this knowledge gap by conducting biochemical characterization, including kinetic, binding, and thermal analyses of selected AGS mutants.

2.2. Results

Effects of AGS mutations on ADAR1 p110. Earlier work demonstrated the effects on editing of a model RNA substrate when AGS mutations were introduced into the p110 and p150 isoforms of ADAR1 overexpressed in HEK293 cells.[17,18] Among the mutations tested, G1007R showed the largest effect on editing activity for both isoforms. These experiments were an important demonstration of the effect of the different AGS mutations on RNA editing with ADAR1 overexpressed in human cells. However, changes in RNA editing activity observed in these cell-based experiments could arise from several sources including changes in the protein's catalytic activity, RNA binding affinity, protein-protein interactions with ADAR1 binding partners, sensitivity to cellular proteases, etc. To gain additional understanding of the consequences of AGS-causing mutations on specific biochemical properties of ADAR1, we aimed to characterize their effects using purified proteins. Therefore, I introduced AGS mutations into ADAR1 p110 and two truncated forms of the protein. One of these shorter forms contains

dsRBDIII and the deaminase domain (ADAR1 R3D aa711-1226) while the other constitutes only the deaminase domain (ADAR1d, aa886-1226). We hypothesized that the effect of certain mutations may also depend on the substrate RNA used. Therefore, I tested the effect of AGS mutations using two well-established ADAR1 substrates from the human transcriptome: the Glioma-associated Oncogene Homolog 1 (hGli1) mRNA[25], and 5-hydroxytryptamine receptor 2C (5-HT_{2c}R) pre-mRNA.[26] By examining the impact of these mutations on ADAR1 catalysis with different protein constructs and different substrate RNAs, we aimed to uncover specific consequences of each AGS mutation.

I initiated the analysis by introducing mutations into the C-terminal histidine-tagged p110 protein. Expression and purification of the wild-type and mutant variants (G1007R, K999N, R892H, and Y1112F) enabled investigation of the impact of these mutations within and outside the RNA binding interface. Initially, the functional consequences of these variants on catalysis were assessed by measuring deamination rate constants under single turnover conditions (**Figure 2.2)**. Consistent with the findings from cellular studies, the G1007R mutant exhibited the slowest reaction rate, showing a 30-fold decrease compared to the wild-type enzyme with the hGli1 substrate RNA (**Figure 2.2B**; **Table 2.1**).[17,18] Interestingly, the R892H mutation also resulted in a 15-fold reduction in reaction rate, while the Y1112F and K999N mutations caused only 2-3-fold decreases. Notably, while rates and reaction end points measured on the more slowly edited 5-HT_{2c}R substrate were slightly reduced for Y1112F and K999N mutants, we could not detect any editing activity on this substrate with the G1007R and R892H p110 mutants (**Figure 2.2C**). Thus, the deamination rate measurements for p110 mutants on the two RNA substrates analyzed showed

that two AGS mutations (G1007R and R892H) have severe defects in catalysis while the other two mutations tested (K999N and Y1112F) resulted in much smaller effect on catalysis.

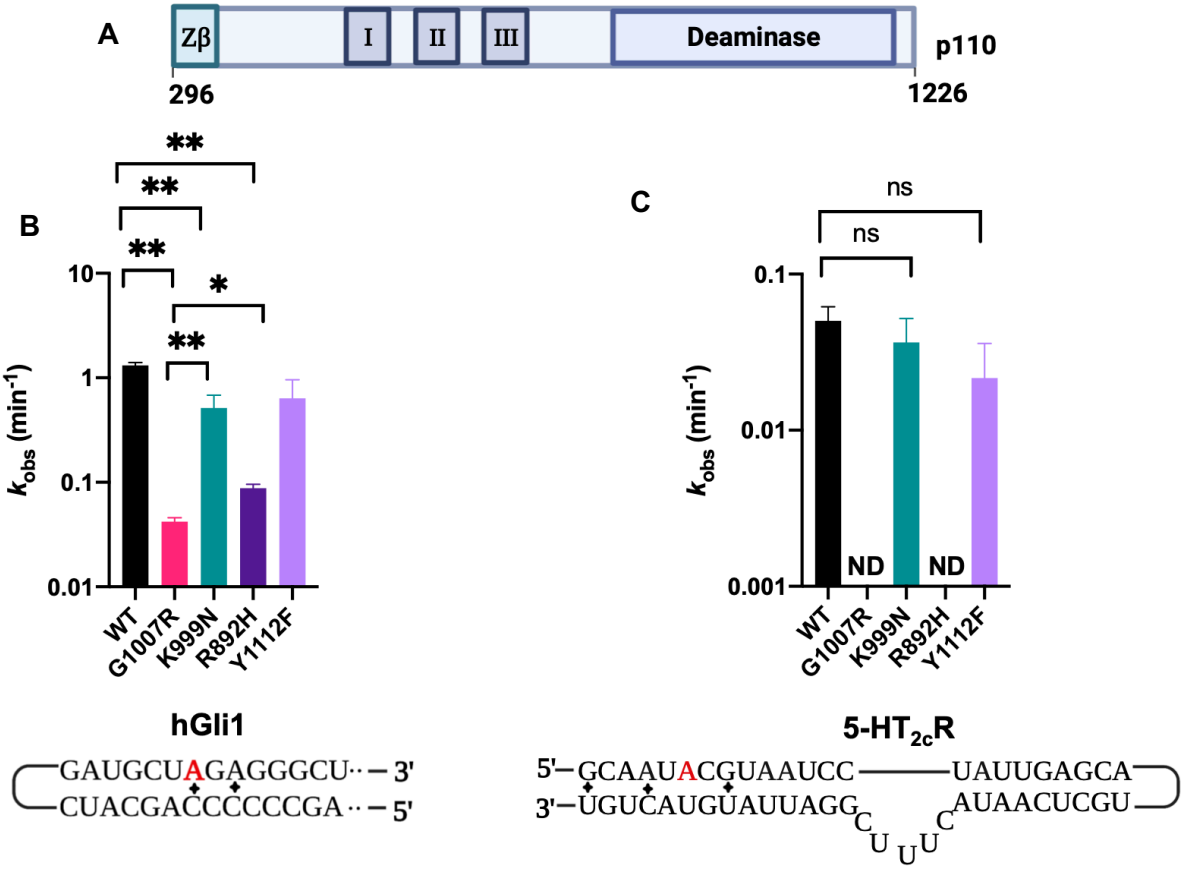


Figure 2.2. Adenosine deamination observed rate constants measured with different AGS mutants in the context of ADAR1 p110. (A) Comparison of observed rate constants for reaction with 10 nM hGli1 RNA substrate and 100 nM ADAR1 p110. (B) Comparison of observed rate constants for reaction with 10 nM 5-HT$_{2c}$R RNA (B site) and 100 nM ADAR1 p110. Editing site monitored for each substrate is highlighted in red. ND: no product detected. Mismatches in the substrates are represented by diamonds.

Table 2.1. Adenosine deamination rate constants measured under single-turnover conditions for ADAR1 p110 and AGS mutants on hGli1 and 5-HT_{2c}R B site RNAs [a].

Substrate	Enzyme	$k_{obs}(min^{-1})$ [b]	k_{rel} [c]
Gli1	ADAR1 p110	1.3 ± 0.1	1
	ADAR1 p110, G1007R	0.042 ± 0.003	0.03
	ADAR1 p110, K999N	0.4 ± 0.1	0.3
	ADAR1 p110, R892H	0.09 ± 0.01	0.1
	ADAR1 p110, Y1112F	0.6 ± 0.3	0.5
5-HT_{2c}	ADAR1 p110	0.05 ± 0.01	1.0
	ADAR1 p110, G1007R	-	-
	ADAR1 p110, K999N	0.04 ± 0.01	0.7
	ADAR1 p110, R892H	-	-
	ADAR1 p110, Y1112F	0.02 ± 0.01	0.5

[a]hGli1 and 5-HT_{2c}R pre-mRNA substrate sequences are shown in Supplementary Table 1. ADAR1 p110 reactions were carried out with 15 mM Tris–HCl pH 7.5, 26 mM KCl, 40 mM potassium glutamate, 1.5 mM EDTA, 0.003% (v/v) NP-40, 4% (v/v) glycerol, 0.5 mM DTT, 1 μg/mL yeast tRNA, and 0.16 U/μL RNAse inhibitor. [b]k_{obs} was calculated by fitting product formed at different time points to the equation: $[P]_t = \alpha[1 - e^{-kobst}]$ where $[P]_t$ is percent edited, α is the end point fitted to 95%, and k_{obs} is the observed rate constant. [c]$k_{rel} = k_{obs}$ for mutant/k_{obs} for ADAR1 p110. "-" indicates rate constants that could not be measured due to no editing detected.

We have previously shown that introduction of the nucleoside analog 8-azanebularine (8-azaN) into ADAR substrate RNAs allows for the mechanistic trapping of ADAR-RNA complexes for binding and structural studies.[20,27,28] The 8-azaN nucleobase is hydrated by ADAR to form a product that is a structural mimic of the ADAR reaction intermediate.[27,28] Since the 8-azaN hydrate lacks a good leaving group, the 8-azaN reaction cannot proceed forward, preventing catalytic turnover. The resulting complexes can then be analyzed by various techniques, including electrophoretic mobility shift assays (EMSAs).[20,21]Dr. Sukanaya Mozumder from the Beal lab, tested the effects of different AGS mutations on the formation of ADAR1-RNA complexes stabilized by 8-azaN-modification of the RNA. She carried out EMSAs using a 61-bp RNA duplex

with sequence derived from hGli1 bearing 8-azaN near the center, a 3' Cy5 tag for visualization in the gels and varying concentrations of ADAR1 p110 and AGS mutants (**Figure 2.3**). Interestingly, the G1007R mutant required a significantly higher protein concentration to observe a stable, shifted complex than the wild-type p110 or other AGS mutants. Whereas the wild-type p110, K999N, Y1112F, and R892H each showed > 30% complex formation at 10 nM enzyme, 100 nM G1007R was required to observe a shifted band in these experiments (**Figure 2.3B, Figure 2.3C**). Unfortunately, due to the high RNA concentrations required to visualize the protein-RNA complexes using fluorescently labeled RNA (20 nM), we were unable to quantify the apparent dissociation constants using this approach.

The inability of the G1007R mutant to form a stable complex with 8-azaN RNA was somewhat surprising given its three unaltered dsRBDs and the previous report of this mutant's ability to inhibit the activity of the wild type protein in HEK293 cells.[17] To test directly the ability of the AGS mutants to influence the reaction of wild type ADAR1, I performed a competition experiment using the wild-type enzyme to edit the 5-HT_{2c}R substrate RNA in the presence of varying concentrations of added AGS mutant p110 proteins. Interestingly, despite being unable to form a stable complex with 8-azaN-modified RNA, the G1007R mutant strongly inhibited wild-type activity on the 5-HT_{2c}R substrate (**Figure 2.4**). At a concentration of 200 nM G1007R p110, editing of the wild-type protein was completely abolished on this substrate. Also, R892H p110 and, to a lesser degree, K999N p110 inhibited the activity of the wild-type protein. No inhibition was observed in this competition experiment with the Y1112F mutant. We suggest the inhibition observed is caused by competitive binding to the substrate RNA by the deaminase–defective

mutant enzyme with the most potent inhibition observed for the mutant least capable of catalyzing adenosine deamination itself on the RNA substrate (i.e. G1007R).

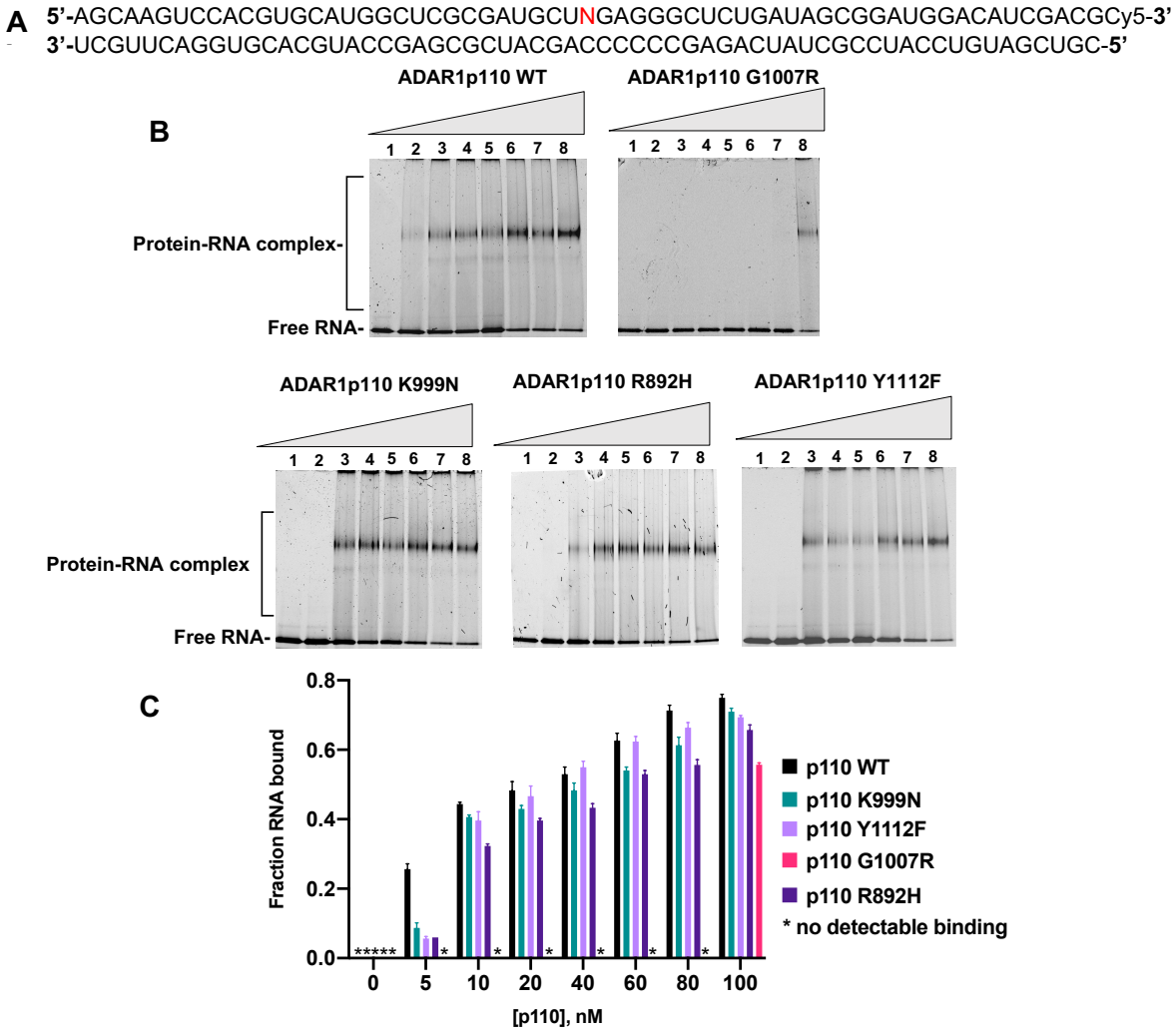


Figure 2.3. Binding of AGS mutants of ADAR1 p110 to a 61 bp duplex bearing 8-azanebularine (N) analyzed by EMSA. **(A)** Sequence of 61 bp RNA duplex with the adenosine analog 8-azanebularine (N). N allows for trapping of the protein-RNA complex in the base-flipped conformation[18,19]. The sequence is labeled with Cy5 at the 3' end. **(B)** Representative EMSA gels of p110 WT and AGS mutants (G1007R, K999N, R892H, Y1112F) with the 61 bp duplex. **(C)** Quantification of fraction RNA bound at various concentrations (lanes 1-8: 0, 5, 10, 20, 40, 60, 80 and 100 nM) of WT or mutant (G1007R, K999N, R892H and Y1112F) p110 and 20 nM 61 bp duplex.

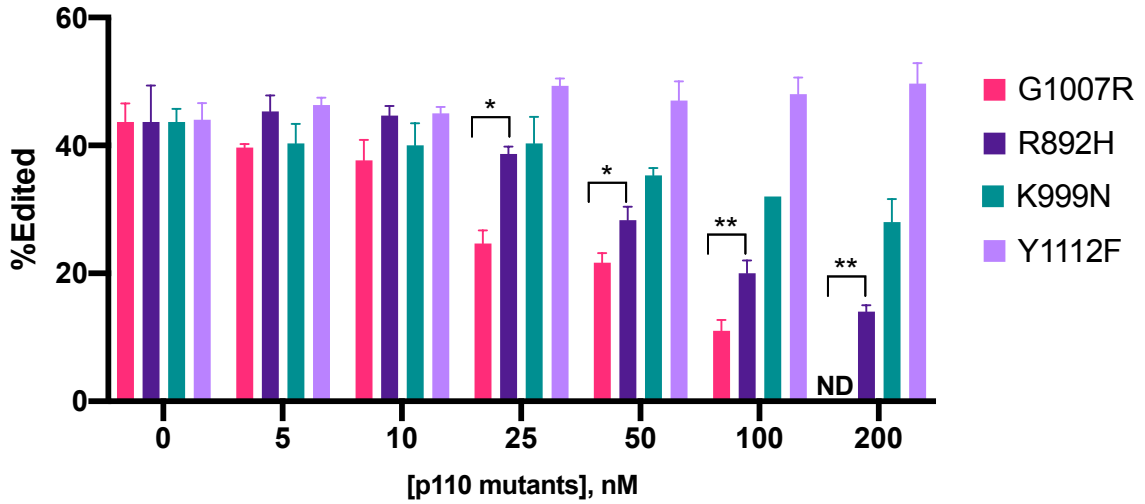


Figure 2.4. G1007R, K999N, and R892H are antagonists of ADAR1 p110 wild-type. Inhibition of ADAR1 p110 with select AGS mutants (G1007R, R892H, K999N and Y1112F) tested on the 5-HT_{2c}R substrate. In vitro deamination was performed with 10nM substrate, 50 nM wild-type and 0, 5, 10, 25, 50, 100 and 200 nM p110 AGS mutants at 30°C. Reaction was initiated by the addition of substrate and quenched after 30 min. Error bars represent n ≥ 3 technical replicates.

Effects of AGS mutations on the N-terminal deletion mutant ADAR1 R3D. Subsequently, I investigated the effects of the AGS mutants in the ADAR1 R3D protein (**Figure 2.5**). Overall, the trends in editing were similar to those observed in the p110 isoform. Notably, the G1007R mutation had a substantially larger impact on the deamination rate than the other mutations with a 28-fold decrease in editing activity with the hGli1 substrate (**Figure 2.5**; **Table 2.2**). The other mutations reduced the deamination rate from 1.3-5.6-fold with this substrate (**Table 2.2**). It is interesting to note that impact of the K999N mutation appears to be greater in ADAR1 R3D than in p110 since editing of the hGli1 substrate is slowed by 4.2-fold compared to wild type (vs 3.3-fold compared to wild type with p110) (**Table 2.1** and **Table 2.2**). This effect is much more pronounced with the 5-HT_{2c}R substrate. With this RNA, no editing could be detected for the

K999N mutant (or with G1007R and R892H). Wild type ADAR1 R3D and the Y1112F mutant deaminated the 5-HT2cR substrate with nearly identical rates.

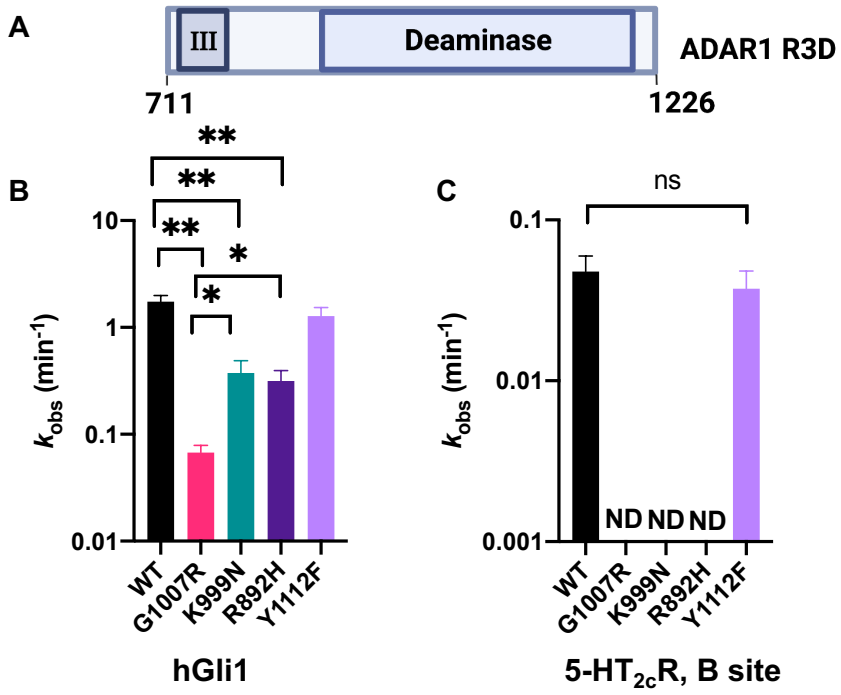


Figure 1.5. Adenosine deamination observed rate constants measured with different AGS mutants in the context of ADAR1 R3D. **(A)** Domain map for ADAR1 R3D. **(B)** Comparison of the observed rate constants of reactions with 10 nM hGli1 substrate and 100 nM ADAR1 R3D. **(C)** Comparison of the observed rate constants of reactions with 10 nM 5HT$_{2c}$R (B site) and 100 nM ADAR1 R3D. Plotted values are the means of three technical replicates ± standard deviation. ND: no product detected.

Table 2.2. Adenosine deamination rate constants measured under single-turnover conditions for ADAR1 R3D and AGS mutants on hGli1 and 5-HT_{2c}R B site RNAs [a].

Substrate	Enzyme	$k_{obs}(min^{-1})$ [b]	k_{rel} [c]
Gli1	ADAR1 R3D	1.7 ± 0.2	1
	ADAR1 R3D, G1007R	0.07 ± 0.01	0.04
	ADAR1 R3D, K999N	0.4 ± 0.1	0.2
	ADAR1 R3D, R892H	0.3 ± 0.1	0.2
	ADAR1 R3D, Y1112F	1.3 ± 0.2	0.7
5-HT_{2c}	ADAR1 R3D	0.05 ± 0.01	1.0
	ADAR1 R3D, G1007R	-	-
	ADAR1 R3D, K999N	-	-
	ADAR1 R3D, R892H	-	-
	ADAR1 R3D, Y1112F	0.04 ± 0.01	0.8

[a]hGli1 and 5-HT_{2c}R pre-mRNA substrate sequences are shown in Supplementary Table 1. ADAR1 R3D reactions were carried out with 15 mM Tris–HCl pH 7.5, 26 mM KCl, 40 mM potassium glutamate, 1.5 mM EDTA, 0.003% (v/v) NP-40, 4% (v/v) glycerol, 0.5 mM DTT, 1 μg/mL yeast tRNA, and 0.16 U/μL RNAse inhibitor. [b]k_{obs} was calculated by fitting fitting product formed at different time points to the equation: $[P]_t = \alpha[1 - e^{-kobst}]$ where $[P]_t$ is percent edited, α is the end point fitted to 95%, and k_{obs} is the observed rate constant. [c]$k_{rel} = k_{obs}$ for mutant/k_{obs} for ADAR1 R3D. "-" indicates rate constants that could not be measured due to no editing detected.

The truncated ADAR1 R3D construct studied in conjunction with a ^{32}P-labeled 32 bp duplex derived from hGli1, facilitated quantification of the effect of these mutations on 8-azaN RNA binding by EMSA (**Figure 2.6**). We had previously used this duplex and a similar truncated variant of human ADAR2 to measure K_d values.[21] Consistent with our observations for the p110 protein, the G1007R mutation within ADAR1 R3D showed a detrimental effect on 8-azaN RNA binding with a K_d value of 49 ± 8 nM compared to 8 ± 2 nM for wild type R3D (6.3-fold difference) (**Figure 2.6B**, **Figure 2.6C**). Importantly, we noticed that the complex formed with the G1007R mutant moved to a position in the gel consistent with the binding of a protein monomer [21], which subsequently transitioned to a higher-order complex at elevated concentrations (**Figure 2.6B**).

This suggested the possibility that this mutation influenced ADAR1 dimerization in addition to adenosine deamination. The measured dissociation constants for each of the other mutations tested (R892H, K999N, Y1112F) were approximately 3-fold higher than for wild type ADAR1 R3D. Considering results from the qualitative gel shift analysis with the p110 mutants and the quantitative binding analysis using the ADAR1 R3D, it appears the G1007R mutation is the most detrimental to 8-azaN RNA binding of all of the AGS mutations tested here.

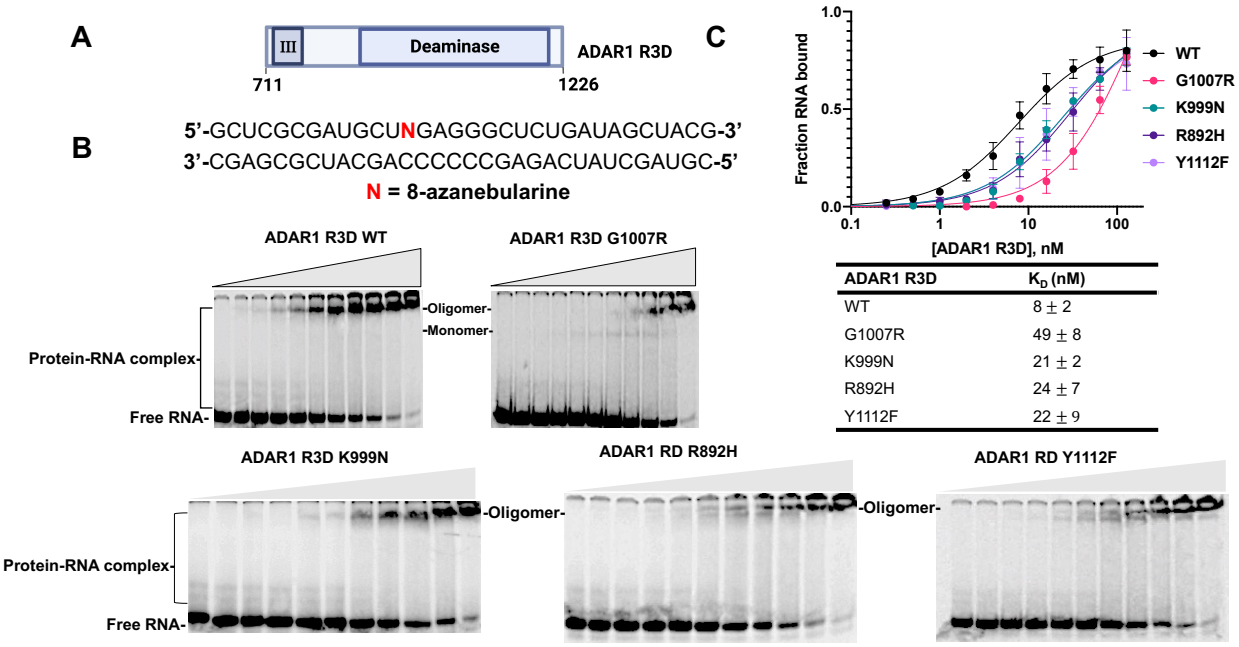


Figure 2.6. Binding of AGS mutants of ADAR1 R3D to a 32 bp duplex bearing 8-azanebularine (N) analyzed by quantitative EMSA. **(A)** Domain map of ADAR1 R3D. **(B)** Sequence of 32 bp duplex bearing 8-azanebularine (N) and representative EMSA gels for ADAR1 R3D WT and AGS mutants (G1007R,K999N, R892H, A870T, Y1112F). Reaction was performed at 1.1 nM 32 bp duplex and protein concentration was varied from 0 to 128 nM. **(C)** Quantitative analysis of the gel shift data. The binding data obtained from the gel shifts were plotted using the equation: $y = A \times [x/ (K_D + x)]$, where y represents the fraction of RNA bound, x corresponds to the concentration of ADAR1 R3D, A denotes the binding endpoint, and K_D represents the dissociation constant. The dissociation constant values were calculated from the plot and reported as the means of three technical replicates ± standard deviation.

Effects of AGS mutants on the ADAR1 deaminase domain. To explore the functional consequences of the AGS mutations further, the protein was truncated to solely the deaminase domain (ADAR1d). In addition to the AGS-causing mutations, a hyperactive mutation, E1008Q was introduced(**Figure 2.7**). The corresponding mutation in ADAR2 is known to enhance base flipping and increase deamination rate.[29] We evaluated the impact of each mutation by measuring deamination rates for editing the hGli1 substrate RNA. The purification and analysis of all deaminase domain mutations were performed in collaboration with the Beal lab member, Kristen Campbell. As previously observed, the G1007R mutation had the greatest effect on editing as no product was observed when this mutation was introduced into the ADAR1 deaminase domain, even with the hyperactive E1008Q variant also present. In addition, upon removal of the RNA binding domains, a large decrease in the rate of editing (100-fold) was observed for the R892H mutation (k_{obs} = 0.02 min-1, **Table 2.3**). Consistent with our findings using p110, the Y1112F and K999N mutations led to small reductions in rate for the hyperactive mutant of the deaminase domain. EMSAs performed with an 8-azaN-bearing 23 bp duplex derived from HER1 mRNA, showed the G1007R double mutant exhibited weaker binding to 8-azaN RNA and primarily existed as a monomer, while the E1008Q mutant shifted to a higher order complex at significantly lower concentrations (**Figure 2.7**). This provides further evidence supporting the hypothesis that G1007R disrupts ADAR1 dimerization in addition to adenosine deamination. Finally, we examined the impact of each mutation on the thermal stability of the ADAR1 deaminase domain. Most mutations did not affect protein stability, including the E1008Q mutation, as indicated by comparable T_M values to the wild-type deaminase domain. However, the G1007R double mutant

exhibited a notable deviation with a T_M value lower by 5 °C compared to the protein with the E1008Q mutation alone (**Table 2.4**)

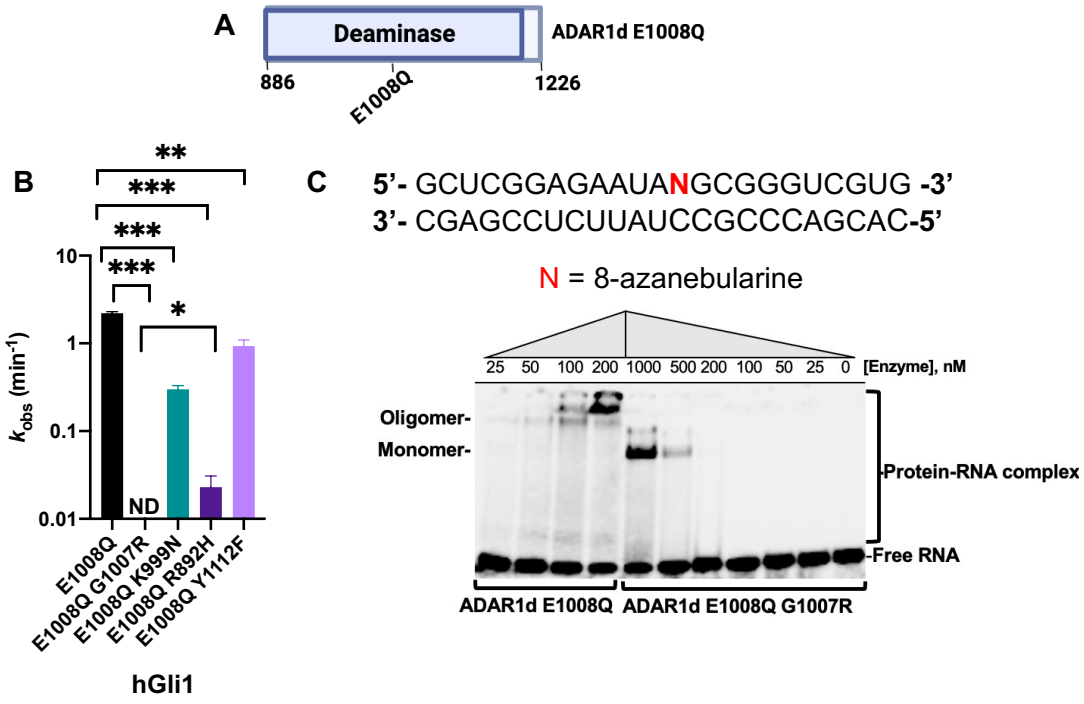


Figure 2.7. Adenosine deamination rate constants measured with different AGS mutants in the context of ADAR1d E1008Q. **(A)** Domain map of ADAR1d E1008Q. **(B)** Comparison of the observed rate constants of 10 nM hGli1 substrate and 100 nM ADAR1d E1008Q. Plotted values are the means of three technical replicates ± standard deviation**.** ND: no product detected. **(C)** EMSA gel with ADAR1d E1008Q and ADAR1d E1008Q G1007R. Reaction was performed at 5 nM 8-azaN-containing 23 bp duplex and protein concentrations [ADAR1d E1008Q]: 25 nM, 50 nM, 100 nM, 200 nM and [ADAR1d E1008Q G1007R]: 1000 nM, 500nM, 200 nM, 100nM, 50 nM and 25 nM.

Table 2.3. Adenosine deamination rate constants measured under single-turnover conditions for ADAR1d E1008Q and AGS mutants on hGli1 RNA [a]. "-" indicates rate constants that could not be measured due to no editing detected.

Substrate	Enzyme	$k_{obs}(min^{-1})$ [b]	k_{rel} [c]
Gli1	ADAR1d E1008Q	2.0 ± 0.1	1
	ADAR1d E1008Q, G1007R	-	-
	ADAR1d E1008Q, K999N	0.30 ± 0.03	0.1
	ADAR1d E1008Q, R892H	0.02 ± 0.01	0.01
	ADAR1d E1008Q, Y1112F	0.4 ± 0.1	0.4

[a]Gli1 substrate sequence is shown in Supplementary Table 1. ADAR1d E1008Q reactions were carried out with 15 mM Tris–HCl pH 7.5, 26 mM KCl, 40 mM potassium glutamate, 1.5 mM EDTA, 0.003% (v/v) NP-40, 4% (v/v) glycerol, 0.5 mM DTT, 1 μg/mL yeast tRNA, and 0.16 U/μL RNAse inhibitor. [b]k_{obs} was calculated by fitting fitting product formed at different time points to the equation: $[P]_t = \alpha[1 - e^{-kobst}]$ where $[P]_t$ is percent edited, α is the end point fitted to 95%, and k_{obs} is the observed rate constant.[c]$k_{rel} = k_{obs}$ for mutant/k_{obs} for ADAR1d E1008Q.

Table 2.4. Experimental thermal melting temperature for ADAR1d E1008Q and AGS mutants [a].

Enzyme	T_M(°C)[b] *
ADAR1d WT	50 ± 1
ADAR1d E1008Q	51 ± 1
ADAR1d E1008Q G1007R	46 ± 0
ADAR1d E1008Q K999N	49 ± 0
ADAR1d E1008Q R892H	52 ± 0
ADAR1d E1008Q Y1112F	50 ± 2

* Values reported are the average of three independent measurements ± standard deviation.

2.4. Discussion

Recent studies have highlighted the regulatory roles of RNA editing by ADAR1 in preventing recognition of self RNA as viral or non-self.[10,16,18,31] Given its role in immune regulation, loss of function mutations within the ADAR1 gene are associated with several immune disorders, including AGS.[17,18,31] There are a few reports in the field investigating the consequences of these mutations on RNA editing by ADAR1 in mammalian cell lines and animal models. 6,[17,18,23,24] However, there no published reports of the effects of these mutations on the biochemical

properties of ADAR1 evaluated under well-defined solution conditions with purified proteins. Here we measured rates of adenosine deamination in two different RNA substrates and binding to 8-azaN RNAs for four different AGS-causing mutants of ADAR1. In addition, we tested the effect of these mutations in N-terminal deletion variants that lack RNA-binding domains. The results of these experiments have highlighted differences among these mutations suggesting the following trend of impact on adenosine deamination: G1007R >> R892H > K999N > Y1112F.

G1007R is highly detrimental to catalysis and binding to 8-azaN RNA. G1007R emerged as the most potent mutation studied here. Interestingly, it is also consistently observed in a heterozygous form among all AGS patients carrying this mutation.[17,32] Given the impact of this mutation on adenosine deamination by ADAR1, a G1007R homozygote would likely be similar to an ADAR1 null mutation, which is known to be embryonic lethal.[17,18,32] Positioned within the GEG triplet in the protein's flipping loop, G1007 plays a crucial role in facilitating the intercalation of E1008 into duplex RNA substrate, thereby ensuring the stability of the flipped-out adenosine conformation for editing. Perturbation of the adjacent glycine residues would be expected to profoundly disrupt the enzyme's ability to induce base flipping.[20,21] Our kinetic analyses, employing both full-length and truncated ADAR1 proteins, illustrates the significant impairment of catalysis caused by the G1007R mutation. In addition, in EMSAs with 8-azaN RNAs using various ADAR1 constructs, we consistently observed the highly detrimental effect of the G1007R mutation. It should be noted here that tight binding of an ADAR protein to 8-azaN RNA requires the ability to flip the 8-azaN into the ADAR active site [20,21,27,28] so a mutation that substantially alters the flipping loop, like G1007R does, would be expected to impair binding to 8-azaN RNA.

However, the G1007R mutation does not prevent the mutant protein from blocking the activity of the wild- type enzyme as indicated in the competition experiments with p110, suggesting a mode of RNA binding not captured in our EMSAs with 8-azaN RNA (**Figure 2.3, Figure 2.4**). This is consistent with previous reports that described G1007R can function as a dominant-negative mutant capable of inhibiting the activity of the wild-type enzyme.[17]

Our results also suggest that G1007R may disrupt ADAR1 dimerization, as indicated by EMSAs with ADAR1 R3D and ADAR1d constructs (**Figure 2.5B**, **Figure 2.7**). The recently reported crystal structure of an ADAR2 dimer bound to RNA identified a dimer interface including an alpha helix composed of a conserved sequence shared between ADAR1 and ADAR2.[21] In one monomer of the ADAR2 dimer, G487 is positioned near the dimerization interface and mutation of this residue to arginine would lead to steric clashes with neighboring residues (**Figure 2.8A**). We suggest that these clashes may disrupt dimerization in the G1007R mutant of ADAR1. Indeed, the AlphaFold model of the ADAR1 dimer also shows clashes with neighboring residues for the G1007R mutant (**Figure 2.8B**). Therefore, it is plausible that the inhibitory effects of G1007R are exerted by preventing dimerization, while still retaining the ability to bind substrates at elevated concentrations. Finally, G1007R exhibits noticeable deviations in thermal stability compared to both the wild-type and other mutants (**Table 2.4)**.

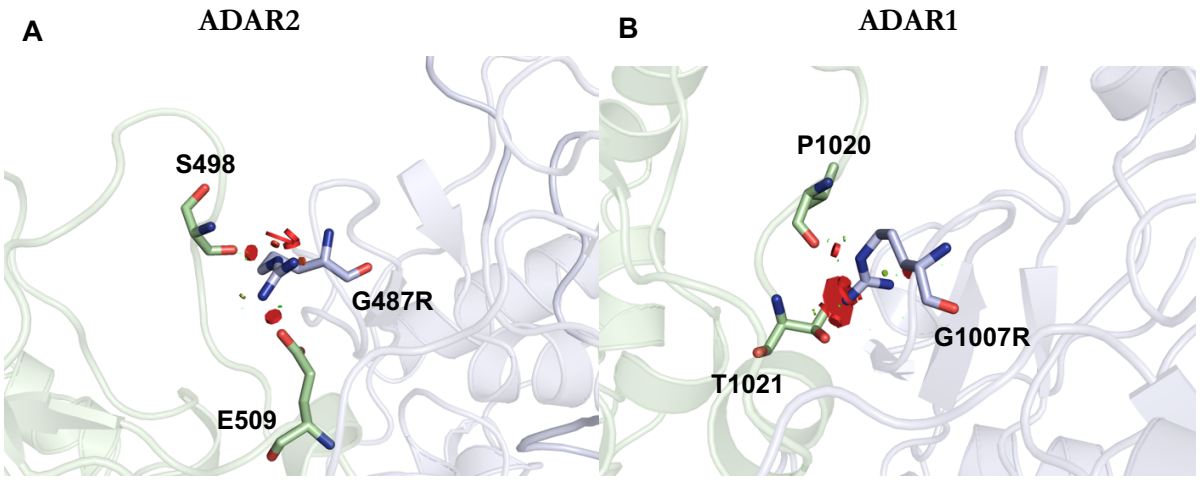


Figure 2.8. G1007R effect on the ADAR dimer interface. **(A)** Close up view mapping the location of G487R (analogous to G1007R mutation in ADAR1) on the non-catalytic monomer (light blue) of hADAR2 R2D dimer (PDBID 6VFF). The red discs represent a significant Van der Waals overlap, suggesting a clash between atoms. At the dimer interface, G487R clashes with E509 and S498. **(B)** An AlphaFold model of the ADAR1 dimer interface. In the model, G1007R clashes with T1021 and P1020.

R892H inhibits adenosine deamination by ADAR1. ADAR sequence alignments suggest R892 of human ADAR1 plays the same role as K376 in human ADAR2.[21] In structures of ADAR2 bound to RNA, K376 is shown stabilizing the base-flipped conformation by hydrogen bonding to both the 5' and 3' phosphodiesters of the nucleotide immediately adjacent to the editing site on the 3' side (**Figure 2.9**).[20,21] The guanidinium group of R892 could also simultaneously hydrogen bond to both 5' and 3' phosphodiesters of this nucleotide in the substrate RNA. Mutation of R892 to histidine would disrupt this contact. Our observation of dramatic reductions in catalytic rate for

R892H ADAR1 p110 and R892H ADAR1d (E1008Q) with both RNA substrates and R892H ADAR1 R3D with the 5-HT_{2c}R substrate is consistent with this prediction. In addition, R892H was a strong inhibitor of ADAR1 p110 wild-type activity with the 5-HT_{2c}R substrate. The R892H mutation is not as detrimental to catalysis of adenosine deamination as is the G1007R mutation as indicated by a comparison of reaction rates for R892H ADAR1 R3D and R892H ADAR1d (E1008Q) with the hGli1 RNA substrate and the corresponding G1007R mutants (**Table 2.2**, **Table 2.3**). This is understood when one considers that a histidine at position 892 could maintain a hydrogen bonding contact and, in its protonated form, even a salt bridge with at least one of the adjacent phosphodiesters in the substrate RNA. It is noteworthy that the R892H mutant was identified in conjunction with the Z domain mutation P193A[17,33]. It is possible that the presence of both the P193A mutation and a catalytic mutant is necessary to observe a disease phenotype. However, homozygous mice carrying the P193A mutation alone exhibit a normal phenotype.[6,33,34] Since our analysis revealed that the R892H mutation alone can substantially disrupt catalysis of adenosine deamination by ADAR1, it is possible that the R892H mutation is the key change leading to disease for the P193A, R892H double mutant found in the AGS patient population.

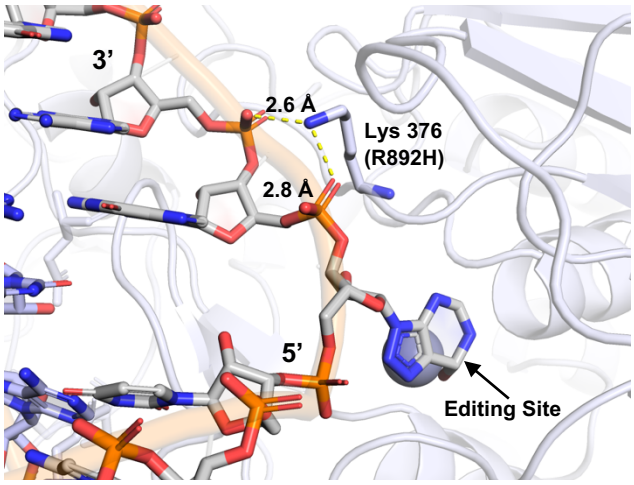


Figure 2.9. Close up view mapping the location of R892H mutation onto hADAR2 R2D structure (PDBID 6VFF). Residue Lys 376 in ADAR2 makes a contact with phosphates of G14 and A15 3' to the edited base.

K999N and Y1112F exert mild effects on ADAR1. Overall, the K999N and Y1112F mutations had more mild effects on the ADAR1 properties evaluated in this study than did G1007R or R892H. The K999N and Y1112F mutations, located adjacent to an important RNA binding loop in the catalytic domain, are observed as homozygous mutations in AGS patients consistent with a less severe impact on catalytic activity.[17,18,23,24] In the context of p110, we measured only a slight reduction in deamination rate with K999N and Y1112F variants (1-3-fold changes in rate). We did note that the K999N mutation had a greater effect in ADAR1 R3D, particularly with the 5-HT_{2c}R substrate, and with ADAR1d (E1008Q) and the hGli1 substrate, highlighting the context-dependent effect of this mutation. Interestingly, a recent mouse model study by Guo et al. also observed a substrate-dependent effect of the ADAR1 p150 K999N mutant.[23,24] They found a significant reduction in editing of the 5-HT_{2c}R substrate at sites A, B, and C, while site D, which is a preferred ADAR2 site, showed an increase. In contrast, editing of glutamate ionotropic receptor AMPA type subunit 2 and 3 (GRIA2 and 3) RNAs remained comparable to the wild-type.

These changes in editing patterns were sufficient to hinder the growth of mice, elevate levels of several immune signature genes, and increase inflammatory cytokine levels in the brain.[24] In contrast to K999N, our studies showed only small effects for the Y1112F mutation in each of our assays. Given its location in the ADAR1 catalytic domain, Y1112F is predicted to indirectly effect substrate recognition by interrupting interactions with the 5' binding loop. Like K999N, this mutation may exert substrate-dependent effects that were not observed in our studies. Clearly more studies with ADAR1 Y1112F will be necessary to identify properties of the protein that are substantially altered by this mutation.

2.5. Conclusions

While the association between mutations within ADAR1 and AGS is established, the underlying mechanisms through which they contribute to disease phenotypes are still being unraveled. Our biochemical studies provide mechanistic insights into select mutants and their interference with catalysis of adenosine deamination. Further investigations using appropriate model systems will be important to determine if specific substrates are affected by each mutation. Unraveling the effects of ADAR1 mutations deepens our understanding of RNA editing.

2.3 Material and methods

Synthesis of oligonucleotides. RNA oligonucleotides containing 8-azanebularine (8-azaN) were synthesized using an in-house ABI 394 synthesizer at 0.2 μmol scale. The 8-azaN phosphoramidite was purchased from Berry & Associates. All other phosphoramidites were purchased from Glen research. Upon completion of oligonucleotide synthesis, columns were dried

under reduced pressure overnight. Once dried, the oligonucleotides were cleaved from the solid support with the treatment of 1:3 ethanol/ 30% NH_4OH at 55 °C overnight. The supernatant was collected in a screw-cap tube and dried under reduced pressure. Removal of silyl groups was performed by suspending the pellets in anhydrous DMSO and treating it with 55% (v/v) Et_3N-3HF at room temperature overnight. The oligonucleotides were precipitated in solution of 65% butanol at -70 °C for 2 h. The pellets were obtained after centrifugation at 13200 x g for 20 min, washing twice with cold 70% ethanol. The RNA pellets were then desalted using a 3000 MWCO Amicon Ultra 0.5 mL centrifugal filter and purified as described below (see Table S1 for sequences). Oligonucleotide masses were confirmed by MALDI-TOF MS (Table S2).

In vitro transcription of RNA. Substrates for deamination studies (5-HT_{2c}R and hGlil) were transcribed from a DNA template using NEB HiScribe T7 RNA synthesis kit. The sequence of each substrate RNA can be found in Supplementary Table S1.

Purification of RNA oligonucleotides. Single-stranded transcribed and synthesized oligonucleotides were purified using denaturing urea-polyacrylamide gels and visualized by UV shadowing. Bands corresponding to desired products were excised from gels, crushed, and soaked overnight at 4 °C in 500 mM NH_4OAc and 1 mM EDTA. Polyacrylamide fragments were removed with a 0.2 μm cellulose acetate filter. RNA was precipitated from the supernatant in a solution of 75% EtOH at -70 °C for 2 h. The supernatant was centrifuged 17000 x g for 20 min and supernatant was removed and washed with 70% cold ethanol. The pellet was lyophilized to dryness, dissolved in nuclease free water, and quantified by absorbance at 260 nm. Synthesized

oligonucleotides were desalted using 3000 MWCO Amicon Ultra 0.5 mL centrifugal filter, and masses for synthetic oligonucleotides were confirmed by MALDI-TOF MS (Appendix Table 2.2).

Protein overexpression and purification. *ADAR1 p110.* Human ADAR1 p110 (UniProtKB P55265-5), consisting of a C-terminal His_{10}-tag was overexpressed in *Saccharomyces cerevisiae* BCY123 as previously described.[30] Cells were lysed using a microfluidizer in lysis buffer containing 20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 1 M KCl, 30 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine-HCl (TCEP-HCl), 0.05% (v/v) Triton X-100, and 50 μM $ZnCl_2$. The lysate was centrifuged at 39000 x g, 4 °C for 1 h and filtered using 0.45 μm filter. The clarified lysate was then passed over a 5 ml Ni-NTA column at a flow rate of 2 ml min^{-1} using an ÄKTA pure 25 FPLC system. The column was washed first with ten column volumes (10 CVs) of lysis buffer, followed with 10 CVs of wash I buffer (20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 500 mM KCl, 30 mM imidazole, 1mM TCEP-HCl, and 50 μM $ZnCl_2$). The protein was eluted with wash I buffer with a gradient of imidazole (30 mM to 400 mM) for 10 CVs. Fractions containing the target protein were pooled, concentrated, and dialyzed against a storage buffer containing 50 mM Tris-HCl pH 8.0, 10% (v/v) glycerol, 400 mM KCl, 50 mM imidazole, 1 mM TCEP-HCl, and 0.01% (v/v) Nonidet P-40 (NP-40). Final protein concentrations were determined by running the samples alongside bovine serum albumin (BSA) standards in an SDS-PAGE gel and visualized by staining with SYPRO Orange dye (Invitrogen). *ADAR1 R3D.* The ADAR1 R3D protein consisting of C-terminal self-cleaving intein sequence, and a chitin binding domain (CBD) was overexpressed in *Saccharomyces cerevisiae* BCY123 as described above. Lysis was carried out in buffer containing 20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 750 mM NaCl, 50 mM imidazole, 1 mM

tris(2-carboxyethyl)phosphine-HCl (TCEP-HCl), 0.05% (v/v) Triton X-100, and 50 μM $ZnCl_2$. The clarified lysate was passed over a chitin binding column using gravity flow. The bound protein was washed with 10 CVs of lysis buffer, followed by 3 CVs of cleavage buffer (20 mM Tris-HCl pH 8.0, 350 mM NaCl, 50 mM imidazole, 5% glycerol, 50 mM DTT). To ensure cleavage of the CBD domain, resin was incubated in 2 CVs of cleavage buffer overnight at 4 °C. The protein was eluted by washing columns with additional two CVs of cleavage buffer and analyzed by running on 4-13% SDS page gel. The fractions corresponding to cleaved protein were collected, diluted 5-fold with heparin equilibration buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 50 mM imidazole, 5% glycerol, 50 mM DTT, 1mM TCEP, 50 μM $ZnCl_2$) and the protein was further purified through a 5 ml GE Healthcare Lifesciences Hi-Trap Heparin HP column at flow rate of 0.5 ml/min. The bound protein was washed with 10 CVs of equilibration buffer and eluted with a gradient of NaCl (200 mM – 1M NaCl) over 5 column volumes followed with step elution for 5 CVs at flow rate of 1 ml/min. Fractions were analyzed by SDS PAGE, concentrated to ~ 1 mg/ml, and dialyzed in 20 mM Tris-HCl pH 8.0, 350 mM KCl, 50 mM imidazole, 10% glycerol, 50 mM DTT, 1mM TCEP, 50 μM $ZnCl_2$. The final protein concentration was determined as described above. *ADAR1 deaminase domain.* ADAR1 E1008Q deaminase domain containing an N-terminal histidine tag was purified similarly to p110 except using the following buffers: (1) lysis buffer (20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 750 mM NaCl, 30 mM imidazole, 1 mM TCEP-HCl, 0.05% (v/v) Triton X-100, and 50 μM $ZnCl_2$. (2) wash buffer (20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 350 mM NaCl, 30 mM imidazole, 1 mM TCEP-HCl, 0.05% (v/v) Triton X-100, and 50 μM $ZnCl_2$); and (3) storage buffer (50 mM Tris-HCl pH 8.0, 10% (v/v) glycerol, 350 mM KCl, 30 mM imidazole, 1 mM TCEP-HCl, 0.01% (v/v) NP-40).

Preparation of RNA substrates for deamination assays. Transcribed RNA was allowed to fold in hybridization buffer (180 nM transcribed RNA target, 1X TE buffer, 100 mM NaCl), heated to 95 °C for 5 min, and slowly cooled to room temperature.

Preparation of duplex substrates for EMSA with Cy5-labeled RNA. A 38 nucleotide (nt) 8-azaN RNA was 5' phosphorylated using T4 Polynucleotide Kinase (NEB) following manufacturer's protocol. Ligation of the phosphorylated 38 nt RNA to a 23 nt oligonucleotide was carried out by mixing 6 nmol of each RNA and hybridized to a DNA splint at 1:1 ratio by heating at 95 °C for 5 min and slow cooling to ≤ 30 °C. T4 DNA ligase (NEB) was used to catalyze the ligation of the two RNAs at 16 °C for 16 h. The ligated 61 nt RNA was phenol:chloroform extracted, ethanol precipitated, and further purified by denaturing PAGE gel as described above. Subsequently, the ligated 61 nt 8-azaN RNA was 3'-end-labeled with fluorescent dye using pCp-Cy5 (Jena Bioscience) and T4 RNA Ligase I (NEB). The mixture was incubated at 16 °C for overnight under light-shielded conditions. Excess Cy5 dye was removed by passing the reaction mixture through a Sephadex G-25 column and the labeled RNA was purified with the phenol:chloroform extraction and ethanol precipitation. The final labeled oligonucleotide was hybridized at a 1:3 ratio to its complement in 10 mM Tris−HCl, pH 7.5, 1 mM EDTA, and 100 mM NaCl by heating at 95 °C for 5 min and slow cooling to ≤ 30 °C. The nucleotides surrounding the 8azaN are derived from the Glioma-associated Oncogene Homolog 1 (hGli1) mRNA.[25] Sequences of all oligonucleotides are found in the Supplementary Information (Appendix Table 2.1).

Preparation of duplex substrates for EMSA with ^{32}P-Labeled RNA. The 5' end of either a 16 nt or 32 nt 8-azaN-containing RNA strand was labeled with γ-[^{32}P]ATP (6000 Ci/mmol) using NEB polynucleotide kinase. The labeled reaction was passed through a G-25 column to remove excess ATP and further purified using a 19% denaturing PAGE gel. The labeled products were visualized using storage phosphor autoradiography. Gel bands containing labeled RNA were excised, crushed, and soaked and worked up as described for other gel-purified oligonucleotides. The dried pellet was resuspended in nuclease free water to a stock solution of approximately 300 nM and hybridized to its compliment at 1:3 ratio in 1X TE buffer, pH 7.5 and 200 mM NaCl by heating at 95 °C for 5 min and slowly cooling to 30 °C to a final concentration of approximately 50 nM. The nucleotide sequences flanking a 16 nt 8-azaN site were derived from the mRNA of Human Epidermal Growth Factor Receptor 1 (HER1).[7] In addition, nucleotide sequences surrounding a 32 nt 8-azaN site were derived from hGli1 mRNA. Sequences of all oligonucleotides are found in the Supplementary Information (Appendix Table 2.1).

Site-directed mutagenesis. Mutagenesis of ADAR1 p110, ADAR1 R3D, and ADAR1d E1008Q was carried out using PCR site-directed mutagenesis using a QuikChange XL Site-Directed Mutagensis Kit (Agilent) with the primers listed in Supplementary (Appendix Table 2.4). All primers were purchased from IDT and purified as described above for other oligonucleotides. Sequences for mutant plasmids were confirmed by Sanger sequencing. See Supplementary Information (Appendix Table 2.5) for the primer sequences.

Deamination assays. Deamination assays were performed under single-turnover conditions in 15 mM Tris–HCl pH 7.5, 26 mM KCl, 40 mM potassium glutamate, 1.5 mM EDTA,

0.003% (v/v) NP-40, 4% glycerol, 0.5 mM DTT, 1 μg/mL yeast tRNA, 0.16 U/μL RNAse inhibitor, 10 nM RNA, and 100 nM protein at 30 °C. Reactions were quenched at 1, 5, 15, 30, 60 and 90 min for hGli1 RNA and 5, 15, 30, 60, 90, and 120 min for 5-HT_{2C}R with 20 ul water at 95°C and heating at 95°C for 5 min. Reaction products were used to generate cDNA using RT-PCR (Promega Access RT-PCR System). The DNA product was purified with DNA Clean & Concentrator kit (Zymo) and subjected to Sanger sequencing via GeneWiz (Azenta). Sequencing peak heights at the edit site was quantified using 4Peaks (Nucleobytes). Data were fit to the equation $[P]t = [P]_f[1-e^{(-kobs \cdot t)}]$ where $[P]t$ is percent edited at time t, $[P]f$ is the final endpoint of editing, and k_{obs} is the observed rate constant. All statistical analyses and nonlinear fits were conducted in Microsoft Excel and GraphPad Prism.

Competition Assay. Samples containing 50 nM wild type ADAR1 p110, and 5, 10, 25, 50, 100 or 200 nM of mutant p110 proteins were incubated in 15 mM Tris–HCl pH 7.5, 26 mM KCl, 40 mM potassium glutamate, 1.5 mM EDTA, 0.003% (v/v) NP-40, 4% glycerol, 0.5 mM DTT, 1 μg/mL yeast tRNA, 0.16 U/μL RNAse inhibitor for 5 min. Reaction was initiated with addition of 10 nM of substrate and quenched after 30 min with 100 ul of water at 95°C and heating at 95°C for 5 min. Reaction products were used to generate cDNA and worked up as described above.

Qualitative EMSA using Cy5-labeled duplex RNA and ADAR1 p110. Samples containing 20 nM Cy5 labeled duplex RNA and 0 to 100 nM enzyme (ADAR1 p110 WT, K999N, R892H and Y1112F) and 0 to 120 nM enzyme (ADAR1p110 G1007R) were incubated respectively in 20 mM Tris−HCl pH 7.4, 140 mM KCl, 10 mM NaCl, 1mM $MgCl_2$, 0.5 mM EDTA and 0.003% (v/v) NP-40 at room temperature for 30 min. Samples were loaded onto a 4 to 16%

Bis-Tris, NativePAGE mini protein gel (Invitrogen) and electrophoresed under nondenaturing conditions in 1× Native PAGE running Buffer (Invitrogen) at 4 °C for 1.5 h under light-shielded conditions. The gels were scanned using a Bio-Rad GelDoc Imaging system.

Quantitative EMSA using ^{32}P duplex RNA and ADAR1 R3D. Samples containing ≤ 1 nM 32 base pair (bp) RNA duplex and varying concentrations of the ADAR1 R3D proteins (0, 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128 nM) were incubated together in 20 mM Tris-HCl, pH 7.0, 3.5% glycerol, 0.5 mM DTT, 60 mM KCl, 20 mM NaCl, 0.1 mM β-mercaptoethanol, 1.5 mM EDTA, 0.003% Nonidet P-40, 0.16 U/μL RNAse inhibitor , 0.2 mg/mL BSA, and 1 μg/mL yeast tRNA for 30 min at 30 °C. Samples were loaded onto a 6% gel, electrophoresed in nondenaturing polyacrylamide gel (79:1 acrylamide:bisacrylamide) in 1× TBE buffer at 4 °C for 90 min. The gels were dried on a Bio-Rad gel dryer for 90 min at 80 °C under vacuum followed by exposure to storage phosphor imaging plates (Kodak) for 24 h in the dark. After exposure, the gels were removed, and the phosphor imaging plates were scanned by Typhoon Trio Variable Mode Imager (GE Healthcare). Dissociation constants were measured by calculating the fraction of RNA bound by the protein and using the equation fraction bound = A*([protein]/[protein]+K_d), where the K_d is the fitted apparent disassociation constant and A is the fitted maximum fraction of RNA bound.[14,16] The sequences for the 32 bp duplex used in the assay are described in Appendix Table 2.1. EMSA for ADAR1d E1008Q and mutants with the 8-azaN-modified 23 bp duplex was performed as described above for ADAR1 R3D except with ≤ 5 nM RNA and in buffer containing 15 mM Tris–HCl pH 7.5, 26 mM KCl, 40 mM potassium glutamate, 1.5 mM EDTA, 0.003% (v/v)

NP-40, 4% glycerol, 0.5 mM DTT, 1 µg/mL yeast tRNA, 0.2 mg/ml BSA and 0.16 U/µL RNAse inhibitor.

Protein melting temperature analysis. Solutions containing 2X SYPRO orange dye, 5 nM ADAR1d E1008Q or mutants were mixed under the following conditions: 50 mM Tris-HCl, pH 8.0, 10% (v/v) glycerol, 350 mM KCl, 30 mM imidazole, 1 mM TCEP-HCl, 0.01% (v/v) NP-40, and 50 µM $ZnCl_2$. To a 96 well plate, 20 ul of each solution was added, wells were sealed with PCR plate sealing film. Fluorescence was measured as the solutions were heated from 5 °C to 90 °C at a rate of 2 °C/min. Spectra were obtained using a Bio-Rad CFX Connect Real-Time PCR Detection System. The derivative of fluorescence signal as a function of temperature ($-dF/dT$) was exported, and the background values of the buffered solution without protein was subtracted from each sample. Melting temperature was determined as the temperature where the derivative of fluorescence signal was at a minimum. Measurements were performed in triplicate. Melting temperature values reported are the average of each replicate ± standard deviation (SD).

Appendix Chapter 2.

Appendix Table 2.1. Oligonucleotide sequences used in this work. All bases are ribonucleotides unless specified. **N** is 8-azanebularine.

Oligonucleotide Name	**Sequence**
32 nt 8-azanebularine top	5'- GCUCGCGAUGCU**N**GAGGGCUCUGAUAGCUACG -3'
32 nt bottom	5'- CGUAGCUAUCAGAGCCCCCCAGCAUCGCGAGC-3'
(3' end Cy5 label) 61 mer 8-azanebularine top	5'- AGCAAGUCCACGUGCAUGGCUCGCGAUGCU**N**GAGGGCUC UGAUAGCGGAUGGACAUCGACG**Cy5** -3'
61 mer bottom	5'- CGUCGAUGUCCAUCCGCUAUCAGAGCCCCCCAGCAUCGC GAGCCAUGCACGUGGACUUGCU -3'

23 nt 8-azanebularine top	5’- GCUCGGAGAAUU**N**GCGGGUCGUG -3’
23 nt bottom	5’- CACGACCCGCCAAUUCUCCGAGC -3’
5HT$_{2C}$R 332 nt	5’- UGGGUACGAAUUCCCACUUACGUACAAGCUUACCUAGAU AUUUGUGCCCCGUCUGGAUUUCUUUAGAUGUUUUAUUUUC AACAGCGUCCAUCAUGCACCUCUGCGCUAUAUCGCUGGAUC GGUAUGUAGCAAU**A**CGUAAUCCUAUUGAGCAUAGCCGUUU CAAUUCGCGGACUAAGGCCAUCAUGAAGAUUGCUAUUGUU UGGGCAAUUUCUAUAGGUAAAUAAAACUUUUUGGCCAUAA GAAUUGCAGCGGCUAUGCUCAAUACUUUCGGAUUAUGUAC UGUGAACAACGUACAGACGUCGACUGGUAACAUUUGCGUU UGAUCGGGUUCU -3’
hGli1 155 nt	5’- CAGAACUUUGAUCCUUACCUCCCAACCUCUGUCUCUCUG UCUACUCACCACAGCCCCCCAGCAUCACUGAGAAUGCUGCC AUGGAUGCU**A**GAGGGCUACAGGAAGAGCCAGAAGUUGGGA CCUCCAUGGUGGGCAGUGGUCUGAACCCCUAUAUG -3’

Appendix Table 2.2. Observed and calculated masses of oligonucleotides.

Oligonucleotide Name	**Observed Mass (m/z)**	**Expected mass (a.m.u.)**
32 nt 8-azanebularine top	10329	10328
32 nt bottom	10203	10206
(For 61 bp, top 5’ fragment) 38 nt 8-azanebularine	12303	12304
(For 61 bp, top 3’ fragment) 23 nt	6994	6991
23 bp 8-azanebularine top	7456	7446
23 bp bottom	7250	7241

Appendix Table 2.3. RT-PCR and sequencing DNA primers used for in vitro deamination in this work.

Oligonucleotide Primers	**Sequence**
5HT$_{2C}$R RT Forward	5’- TGGGTACGAATTCCCACTTACGTACAAGCTT -3’
5HT$_{2C}$R Reverse	5’- AGAACCCGATCAAACGCAAATGTTAC -3’
hGli1 RT Forward	5’- TAATACGACTCACTATAGGGCAGAACTTTGATCCTTACCTC -3’

*hGli1*Reverse	5’- CATATAGGGGTTCAGACCACTG -3’

Appendix Table 2.4. Primers used for site-directed mutagenesis of ADAR1 R3D, ADAR1d, ADAR1 p110.

Mutant	**Sequence**
Y1112F Forward	5’- GGTTGGTAGAGTTTCTATCTTCGACTCTAAGAGACAATCTGG -3’
Y1112F Reverse	5’- CCAGATTGTCTCTTAGAGTCGAAGATAGAAACTCTACCAACC -3’
G1007R Forward	5’- GTTGAGAACTAAGGTTGAAAACAGAGAAGGTACTATCCCAGTTGAATC -3’
G1007R Reverse	5’- GATTCAACTGGGATAGTACCTTCTCTGTTTTCAACCTTAGTTCTCAAC -3’
K999N Forward	5’- CGAAAACCCAAAGCAAGGTAACTTGAGAACTAAGGTTGAAAAC -3’
K999N Reverse	5’- GTTTTCAACCTTAGTTCTCAAGTTACCTTGCTTTGGGTTTTCG -3’
R892H Forward	5’- GTTTCTTTGGGTACTGGTAACCACTGTGTTAAGGGTGACTCTTTGTC -3’
R892H Reverse	5’- GACAAAGAGTCACCCTTAACACAGTGGTTACCAGTACCCAAAGAAAC -3’

Appendix Table 2.5. Sequencing DNA primers for ADAR1 p110, ADAR1 R3D, and ADAR1d constructs.

Oligonucleotide Primers	**Sequence**
Galpromoterfor	5’- ATGTAAAGAGCCCCATTATCTTAGCC -3’
Zdomainfor	5’- CTACACCACCAATTTGGCATTTGACTG -3’
ADAR1 R3D for	5’- GAGCTATTATGGAAATGCCATCCTTC-3’
ADAR1 R2D for	5’- CAAACTCCAACACCATCTGCTAC -3’
ADAR1 R1D for	5’-CACTAATCCAGTTGGTGGTTTATTGG-3’
ADAR1d for	5’-TTGCCATTGACTGGTTCTACTTTCCAC-3’
ADAR1d for 2	5’- GGGTGAAAGATTGAGAACTATGTC-3’
Intein CBD for	5’- CTTCTACTTGTGTCCAGTTGCAATG -3’
Vector rev	5’- TTTTCTCGGGCAGATCTTTGTC -3’

2.6. References

(1) Eisenberg, E.; Levanon, E. Y. A-to-I RNA Editing — Immune Protector and Transcriptome Diversifier. *Nat Rev Genet* **2018**, *19* (8), 473–490.

(2) Bass, B. L. RNA Editing by Adenosine Deaminases That Act on RNA. *Annu Rev Biochem* **2002**, *71* (1), 817–846.

(3) Wang, Y.; Zheng, Y.; Beal, P. A. Adenosine Deaminases That Act on RNA (ADARs). *Enzymes.* 2017; pp 215–268.

(4) Nakahama, T.; Kato, Y.; Shibuya, T.; Inoue, M.; Kim, J. I.; Vongpipatana, T.; Todo, H.; Xing, Y.; Kawahara, Y. Mutations in the Adenosine Deaminase ADAR1 That Prevent Endogenous Z-RNA Binding Induce Aicardi-Goutières-Syndrome-like Encephalopathy. *Immunity* **2021**, *54* (9), 1976-1988.e7.

(5) de Reuver, R.; Dierick, E.; Wiernicki, B.; Staes, K.; Seys, L.; De Meester, E.; Muyldermans, T.; Botzki, A.; Lambrecht, B. N.; Van Nieuwerburgh, F.; Vandenabeele, P.; Maelfait, J. ADAR1 Interaction with Z-RNA Promotes Editing of Endogenous Double-Stranded RNA and Prevents MDA5-Dependent Immune Activation. *Cell Rep* **2021**, *36* (6), 109500.

(6) Guo, X.; Liu, S.; Sheng, Y.; Zenati, M.; Billiar, T.; Herbert, A.; Wang, Q. ADAR1 Zα Domain P195A Mutation Activates the MDA5-Dependent RNA-Sensing Signaling Pathway in Brain without Decreasing Overall RNA Editing. *Cell Rep* **2023**, *42* (7), 112733.

(7) Nie, Y.; Zhao, Q.; Su, Y.; Yang, J.-H. Subcellular Distribution of ADAR1 Isoforms Is Synergistically Determined by Three Nuclear Discrimination Signals and a Regulatory Motif. *J Biol Chem.* **2004**, *279* (13), 13249–13255.

(8) Wang, Y.; Park, S.; Beal, P. A. Selective Recognition of RNA Substrates by ADAR Deaminase Domains. *Biochemistry* **2018**, *57* (10), 1640–1651.

(9) Park, S.; Doherty, E. E.; Xie, Y.; Padyana, A. K.; Fang, F.; Zhang, Y.; Karki, A.; Lebrilla, C. B.; Siegel, J. B.; Beal, P. A. High-Throughput Mutagenesis Reveals Unique Structural Features of Human ADAR1, *Nat Commun* **2020**, *11*(1), 5130.

(10) Nakahama, T.; Kawahara, Y. The RNA-Editing Enzyme ADAR1: A Regulatory Hub That Tunes Multiple DsRNA-Sensing Pathways. *Int Immunol* **2023**, *35* (3), 123–133.

(11) Song, B.; Shiromoto, Y.; Minakuchi, M.; Nishikura, K. The Role of RNA Editing Enzyme ADAR1 in Human Disease. *WIREs RNA* **2022**, *13* (1).

(12) Liu, J.; Wang, F.; Zhang, Y.; Liu, J.; Zhao, B. ADAR1-Mediated RNA Editing and Its Role in Cancer. *Front Cell Dev Biol* **2022**, *10*.

(13) Gannon, H. S.; Zou, T.; Kiessling, M. K.; Gao, G. F.; Cai, D.; Choi, P. S.; Ivan, A. P.; Buchumenski, I.; Berger, A. C.; Goldstein, J. T.; Cherniack, A. D.; Vazquez, F.; Tsherniak, A.; Levanon, E. Y.; Hahn, W. C.; Meyerson, M. Identification of ADAR1 Adenosine Deaminase Dependency in a Subset of Cancer Cells. *Nat Commun* **2018**, *9* (1), 5450.

(14) Bhate, A.; Sun, T.; Li, J. B. ADAR1: A New Target for Immuno-Oncology Therapy. *Mol Cell* **2019**, *73* (5), 866–868.

(15) Heraud-Farlow, J. E.; Walkley, C. R. What Do Editors Do? Understanding the Physiological Functions of A-to-I RNA Editing by Adenosine Deaminase Acting on RNAs. *Open Biol* **2020**, *10* (7).

(16) 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 (1979)* **2015**, *349* (6252), 1115–1120.

(17) Rice, G. I.; Kasher, P. R.; Forte, G. M. A.; Mannion, N. M.; Greenwood, S. M.; Szynkiewicz, M.; Dickerson, J. E.; Bhaskar, S. S.; Zampini, M.; Briggs, T. A.; Jenkinson, E. M.; Bacino, C. A.; Battini, R.; Bertini, E.; Brogan, P. A.; Brueton, L. A.; Carpanelli, M.; De Laet, C.; de Lonlay, P.; del Toro, M.; Desguerre, I.; Fazzi, E.; Garcia-Cazorla, À.; Heiberg, A.; Kawaguchi, M.; Kumar, R.; Lin, J.-P. S.-M.; Lourenco, C. M.; Male, A. M.; Marques, W.; Mignot, C.; Olivieri, I.; Orcesi, S.; Prabhakar, P.; Rasmussen, M.; Robinson, R. A.; Rozenberg, F.; Schmidt, J. L.; Steindl, K.; Tan, T. Y.; van der Merwe, W. G.; Vanderver, A.; Vassallo, G.; Wakeling, E. L.; Wassmer, E.; Whittaker, E.; Livingston, J. H.; Lebon, P.; Suzuki, T.; McLaughlin, P. J.; Keegan, L. P.; O'Connell, M. A.; Lovell, S. C.; Crow, Y. J. Mutations in ADAR1 Cause Aicardi-Goutières Syndrome Associated with a Type I Interferon Signature. *Nat Genet* **2012**, *44* (11), 1243–1248.

(18) Lamers, M. M.; van den Hoogen, B. G.; Haagmans, B. L. ADAR1: "Editor-in-Chief" of Cytoplasmic Innate Immunity. *Front Immunol* **2019**, *10*.

(19) Crow, Y. J.; Chase, D. S.; Lowenstein Schmidt, J.; Szynkiewicz, M.; Forte, G. M. A.; Gornall, H. L.; Oojageer, A.; Anderson, B.; Pizzino, A.; Helman, G.; Abdel-Hamid, M. S.; Abdel-Salam, G. M.; Ackroyd, S.; Aeby, A.; Agosta, G.; Albin, C.; Allon-Shalev, S.; Arellano, M.; Ariaudo, G.; Aswani, V.; Babul-Hirji, R.; Baildam, E. M.; Bahi-Buisson, N.; Bailey, K. M.; Barnerias, C.; Barth, M.; Battini, R.; Beresford, M. W.; Bernard, G.; Bianchi, M.; Billette de Villemeur, T.; Blair, E. M.; Bloom, M.; Burlina, A. B.; Luisa Carpanelli, M.; Carvalho, D. R.; Castro-Gago, M.; Cavallini, A.; Cereda, C.; Chandler, K. E.; Chitayat, D. A.; Collins, A. E.; Sierra Corcoles, C.; Cordeiro, N. J. V.; Crichiutti, G.; Dabydeen, L.; Dale, R. C.; D′Arrigo, S.; De Goede, C. G. E. L.; De Laet, C.; De Waele, L. M. H.; Denzler, I.; Desguerre, I.; Devriendt, K.; Di Rocco, M.; Fahey, M. C.; Fazzi, E.; Ferrie, C. D.; Figueiredo, A.; Gener, B.; Goizet, C.; Gowrinathan, N. R.; Gowrishankar, K.; Hanrahan, D.; Isidor, B.; Kara, B.; Khan, N.; King, M. D.; Kirk, E. P.; Kumar, R.; Lagae, L.; Landrieu, P.; Lauffer, H.; Laugel, V.; Piana, R. La; Lim, M. J.; Lin, J.-P. S.-M.; Linnankivi, T.; Mackay, M. T.; Marom, D. R.; Marques Lourenço, C.; McKee, S. A.; Moroni, I.; Morton, J. E. V.; Moutard, M.-L.; Murray, K.; Nabbout, R.; Nampoothiri, S.; Nunez-Enamorado, N.; Oades, P. J.; Olivieri, I.; Ostergaard, J. R.; Pérez-Dueñas, B.; Prendiville, J. S.; Ramesh, V.; Rasmussen, M.; Régal, L.; Ricci, F.; Rio, M.; Rodriguez, D.; Roubertie, A.; Salvatici, E.; Segers, K. A.; Sinha, G. P.; Soler, D.; Spiegel, R.; Stödberg, T. I.; Straussberg, R.; Swoboda, K. J.; Suri, M.; Tacke, U.; Tan, T. Y.; te Water Naude, J.; Wee Teik, K.; Mary Thomas, M.; Till, M.; Tonduti, D.; Maria Valente, E.; Noel Van Coster, R.; van der Knaap, M. S.; Vassallo, G.; Vijzelaar, R.; Vogt, J.; Wallace, G. B.; Wassmer, E.; Webb, H. J.; Whitehouse, W. P.; Whitney, R. N.; Zaki, M. S.; Zuberi, S. M.; Livingston, J. H.; Rozenberg, F.; Lebon, P.; Vanderver, A.; Orcesi, S.; Rice, G. I. Characterization of Human Disease Phenotypes Associated with Mutations in *TREX1* , *RNASEH2A* , *RNASEH2B* , *RNASEH2C* , *SAMHD1* , *ADAR* , and *IFIH1*. *Am J Med Genet A* **2015**, *167* (2), 296–312.

(20) Matthews, M. M.; Thomas, J. M.; Zheng, Y.; Tran, K.; Phelps, K. J.; Scott, A. I.; Havel, J.; Fisher, A. J.; Beal, P. A. Structures of Human ADAR2 Bound to DsRNA Reveal Base-Flipping Mechanism and Basis for Site Selectivity. *Nat Struct Mol Biol* **2016**, *23* (5), 426–433.

(21) 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 Res* **2020**, *48* (14), 7958–7972.
(22) Fisher, A. J.; Beal, P. A. Effects of Aicardi-Goutières Syndrome Mutations Predicted from ADAR-RNA Structures. *RNA Biol* **2017**, *14* (2), 164–170.
(23) Guo, X.; Wiley, C. A.; Steinman, R. A.; Sheng, Y.; Ji, B.; Wang, J.; Zhang, L.; Wang, T.; Zenatai, M.; Billiar, T. R.; Wang, Q. Aicardi-Goutières Syndrome-Associated Mutation at ADAR1 Gene Locus Activates Innate Immune Response in Mouse Brain. *J Neuroinflammation* **2021**, *18* (1), 169.
(24) Guo, X.; Steinman, R. A.; Sheng, Y.; Cao, G.; Wiley, C. A.; Wang, Q. An AGS-Associated Mutation in ADAR1 Catalytic Domain Results in Early-Onset and MDA5-Dependent Encephalopathy with IFN Pathway Activation in the Brain. *J Neuroinflammation* **2022**, *19* (1), 285.
(25) Mendoza, H. G.; Matos, V. J.; Park, S.; Pham, K. M.; Beal, P. A. Selective Inhibition of ADAR1 Using 8-Azanebularine-Modified RNA Duplexes. *Biochemistry* **2023**, *62* (8), 1376–1387.
(26) Shimokawa, T.; Rahman, M. F.-U.; Tostar, U.; Sonkoly, E.; Ståhle, M.; Pivarcsi, A.; Palaniswamy, R.; Zaphiropoulos, P. G. RNA Editing of the GLI1 Transcription Factor Modulates the Output of Hedgehog Signaling. *RNA Biol* **2013**, *10* (2), 321–333.
(27) Eggington, J. M.; Greene, T.; Bass, B. L. Predicting Sites of ADAR Editing in Double-Stranded RNA. *Nat Commun* **2011**, *2* (1), 319.
(28) Haudenschild, B. L.; Maydanovych, O.; Véliz, E. A.; Macbeth, M. R.; Bass, B. L.; Beal, P. A. A Transition State Analogue for an RNA-Editing Reaction. *J Am Chem Soc* **2004**, *126* (36), 11213–11219.
(29) 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 Res* **2015**, *43* (2), 1123–1132.
(30) Kuttan, A.; Bass, B. L. Mechanistic Insights into Editing-Site Specificity of ADARs. *Proc Natl Acad Sci U S A.* **2012**, *109* (48).
(31) Kleinova, R.; Rajendra, V.; Leuchtenberger, A. F.; Lo Giudice, C.; Vesely, C.; Kapoor, U.; Tanzer, A.; Derdak, S.; Picardi, E.; Jantsch, M. F. The ADAR1 Editome Reveals Drivers of Editing-Specificity for ADAR1-Isoforms. *Nucleic Acids Res* **2023**, *51* (9), 4191–4207.
(32) Rice, G.; Kitabayashi, N.; Barth, M.; Briggs, T.; Burton, A.; Carpanelli, M.; Cerisola, A.; Colson, C.; Dale, R.; Danti, F.; Darin, N.; De Azua, B.; De Giorgis, V.; De Goede, C.; Desguerre, I.; De Laet, C.; Eslahi, A.; Fahey, M.; Fallon, P.; Fay, A.; Fazzi, E.; Gorman, M.; Gowrinathan, N.; Hully, M.; Kurian, M.; Leboucq, N.; Lin, J.-P.; Lines, M.; Mar, S.; Maroofian, R.; Martí-Sanchez, L.; McCullagh, G.; Mojarrad, M.; Narayanan, V.; Orcesi, S.; Ortigoza-Escobar, J.; Pérez-Dueñas, B.; Petit, F.; Ramsey, K.; Rasmussen, M.; Rivier, F.; Rodríguez-Pombo, P.; Roubertie, A.; Stödberg, T.; Toosi, M.; Toutain, A.; Uettwiller, F.; Ulrick, N.; Vanderver, A.; Waldman, A.; Livingston, J.; Crow, Y. Genetic, Phenotypic, and Interferon Biomarker Status in ADAR1-Related Neurological Disease. *Neuropediatrics* **2017**, *48* (03), 166–184.
(33) Liang, Z.; Chalk, A. M.; Taylor, S.; Goradia, A.; Heraud-Farlow, J. E.; Walkley, C. R. The Phenotype of the Most Common Human ADAR1p150 Zα Mutation P193A in Mice Is Partially Penetrant. *EMBO Rep* **2023**, *24* (5).

Chapter 3

Optimization of Purification Methods and Biophysical Characterization of ADAR1

This chapter outlines the efforts to improve purification methods to enhance solubility of ADAR1. Parts of the work presented in this chapter, specifically discovery of second zinc binding site of ADAR1 was led by Dr. SeHee Park and published in Nature Communications in 2020.

3.1. Introduction

RNA editing is a pivotal post-transcriptional modification occurring in eukaryotic systems. It involves the manipulation of nucleobases through insertion, deletion, or modification.[1] The outcomes encompass alterations in protein function, splice site deletions or incorporations, and changes to RNA secondary structures.[2] A prevalent form of RNA editing is the site-selective hydrolytic deamination of adenosine to inosine within double-stranded RNA (dsRNA) substrates. Inosine mimics guanosine's hydrogen bonding properties and pairs with cytosine.[3,4] This conversion is facilitated by a class of enzymes known as adenosine deaminases acting on RNA (ADARs). Humans possess two active variants: ADAR1 and ADAR2. [3,5] The controlled regulation of ADARs' A-to-I editing activity is crucial for normal cellular functions. Dysregulation of ADAR-mediated editing is associated with neurological disorders such as schizophrenia and epilepsy, as well as with cancer.[5] In cancer, hyper-editing by ADAR1 can lead to the evasion of cell lethality mediated by PKR.[6] The loss of ADAR1 function is also linked to enhanced tumor sensitivity to immunotherapy and the activation of anti-tumor cytokines.[6,7,8] Consequently, targeting ADAR1 for inhibition is a significant therapeutic strategy for ADAR1-dependent cancer cell lines.[8] However, ADAR loss of function is also implicated in Aicardi Goutières syndrome (AGS), a

severe autoimmune disorder in children caused by mutations in the ADAR1 catalytic domain.[9] This underscores the importance of investigating the regulation of ADAR1 activity and its ability to distinguish between self and non-self RNAs.[9] This understanding is essential for unraveling the immune response pathways controlled by ADAR1.

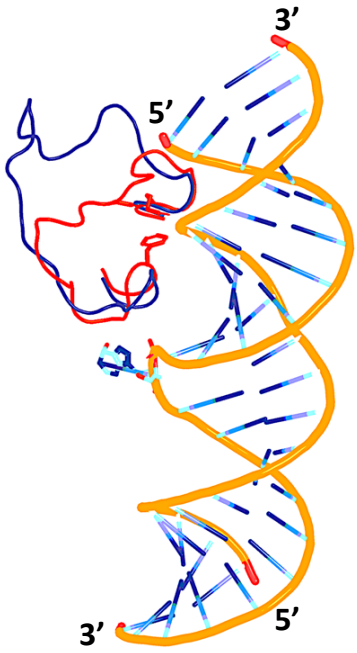


Figure 3.1 Homology model of ADAR1 based on the structure of hADAR2d bound to dsRNA (PDB 5ED2). Modeled are 5'binding loop of ADAR1 (blue), compared to 5' binding loop of ADAR2 (red), responsible for differences in substrate selectivity of their deaminase domain.

The ADAR proteins have a modular structure with double-stranded RNA-binding domains (dsRBDs) and a C-terminal deaminase domain.[3] ADARs require duplex secondary structure in their substrate RNAs and use a base flipping mechanism to place the reactive adenosine into a zinc-containing active site.[10,11] Indeed, surface loops present on the ADAR deaminase domain have been identified that bind RNA on the 5′ side of an editing site (i.e. 5′ binding loop), and on

the 3′ side of the editing site (i.e., 3′ binding loop), both of which are directly involved in base flipping (i.e., flipping loop).[12, 13] Selective editing of adenosines in RNA molecules is a hallmark of ADAR activity, with ADAR1 and ADAR2 displaying overlapping yet distinct preferences.[14] While domain swapping experiments have highlighted differences in selectivity arising from variations in their deaminase domains, the exact basis for ADAR-specific selectivity remains a subject of study.[13,14] Notably, our earlier investigations have underscored the significance of differences in the sequence of the 5' binding loops of the deaminase domains for ADAR-specific selectivity (**Figure 3.1**).[13] The structure and RNA recognition properties of the deaminase domain of human ADAR2 (hADAR2d) have been extensively studied.[15,16] Less is known about the ADAR1 deaminase domain (hADAR1d). This is, in part, due to the lack of structural information for the ADAR1 deaminase domain alone or in complex with RNA. Thus, studies that advance our understanding of structural features of the ADAR1 deaminase domain, particularly those that are unique to ADAR1, are important. Biochemical studies of AGS mutants are limited due to low solubility of ADAR1 even in its truncated forms. Although, ADAR1 and ADAR2 deaminase domains are 39% identical and contain 59% sequence similarity, they differ in their substrate selectivity and solubility. These differences are not revealed by the crystal structure of ADAR2 bound to dsRNA, motivating a need for a high-resolution structure. Additionally, a new method of purification was required to generate the ADAR1 AGS mutations for their biochemical characterization. Therefore, in this chapter, I highlight all efforts to improve ADAR1 solubility for structural studies, and the discovery of a structural zinc binding site critical for its function and editing of certain substrates.

3.2. Results

A N-terminal MBP tag enhances ADAR1 solubility. The previously used histidine tag was replaced with yeast codon optimized maltose binding protein (MBP). MBP is frequently used as a fusion partner to improve protein solubility for crystallography.[17] The resulting protein was purified using an amylose resin. To compare solubility, the protein was concentrated to ~70 μM, a concentration at which his-tag ADAR1d would precipitate. To ensure that MBP ADAR1d protein retains activity, a deamination assay was performed with a human glioma-associated oncogene 1 (hGLI1) RNA (**Figure 3.2**). The protein was found to be active but required a higher concentration than the histidine tagged protein for similar levels of editing to be observed. MBP is a large fusion tag, potentially blocking binding sites of RNA in the deaminase domain. Although less active, this construct significantly improved the protein's solubility allowing a new method to generate ADAR1 mutants that may otherwise misfold or precipitate. Despite, improvement in solubility, the efforts to crystallize ADAR1d remained unsuccessful. On the other hand, this fusion protein construct enabled generation of cystine mutants that play a crucial role in its functional and structural properties.[18]

ADAR1 consists of a structural zinc binding site.[18] In addition to featuring a flexible 5' binding loop, ADAR1 also incorporates 11 cysteines.[13,18] To identify surface cysteines responsible for ADAR1 insolubility, a previous lab member, Dr. SeHee Park, employed the Sat-FACS-Seq screening method. This technique generated a cysteine library, exploring all possible combinations of amino acids and their impact on ADAR1's function. The study highlighted C1081 and C1082,

which displayed a strong preference for cysteines. Importantly, C1082 tolerated histidine, glutamic acid, and aspartic acid – amino acids well-suited for coordinating metal co-factors.[18]

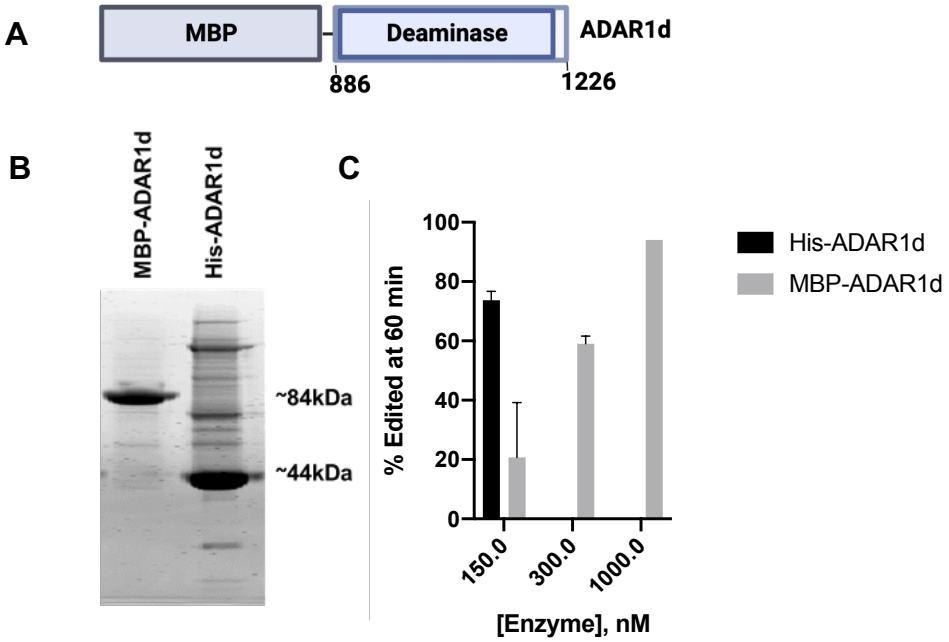


Figure 3.2 MBP fused ADAR1d is active. **(A)** Domain map of MBP-ADAR1. (**B)** Representative page-gel of MBP-ADAR1d in comparison to His-ADAR1d**. (C)** Activity profile of MBP-ADAR1d in varying concentrations with hGli1.

Due to the adjacency of cysteine residues at positions 1081 and 1082, both of which exhibited a preference for amino acids commonly associated with metal binding, it was speculated that an additional metal-binding site other than the catalytic zinc-binding site could exist in hADAR1d. This notion was supported by a homology model of the ADAR1 deaminase domain created by another member of the Beal lab, Dr. Erin Doherty. This Rosetta-based homology model utilized the available structure of hADAR2d bound to dsRNA as well as biochemical data as constraints for accurate modeling. The model suggested that H1103 was in proximity to C1081

and C1082, aligning with a potential metal-binding site **(Figure 3.3).**[18] The conservation of C1081, C1082, and H1103 among ADAR1 proteins from various organisms implies the functional importance of these residues. To verify if these cysteines were indeed involved in metal binding, MBP-ADAR1 wild-type and cysteine mutants were subjected to ICP-MS metal analysis. Interestingly, the putative metal-binding residues, C1082 or H1103, to either Asp or Glu caused a substantial loss of zinc. In addition, mutating a different cysteine in the protein not predicted to be involved in the second zinc site (C893) to Asp minimally affected the measured zinc metal content compared to that of WT MBP-hADAR1d fusion. In addition, kinetic analysis in vitro and in HEK293T revealed mutation of C1081 to aspartic acid or glutamic acid, significantly reduced editing revealing the importance of zinc in cellular ADAR editing.[18]

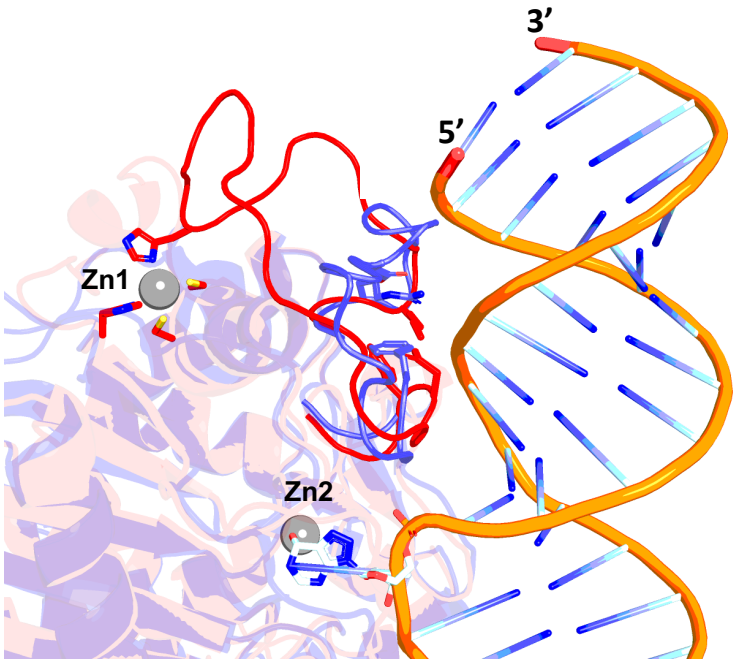


Figure 3.3. Representative Rosetta based homology model of ADAR1, based on the structure of hADAR2d bound to dsRNA (PDB 5ED2). The model displays four zinc ligands, coordinating the second zinc binding site. Catalytic zinc binding site is also shown.

Given the solubility enhancement of MBP fused ADAR1d, I wondered if the protein would remain soluble once the tag is cleaved. I hypothesized, that MBP should encourage proper folding of the deaminase domain and remain well folded even without the tag. To cleave the MBP tag, a small-scale study was first performed to determine optimal TEV and ADAR1 ratio for cleavage. The proteins were incubated at varying ratio for 16 h at 4 °C. Cleavage at these conditions was not efficient even at 1:6 ratio of ADAR1 and TEV (**Figure 3.4A**). A literature search revealed that TEV protease is most active in temperature 30-34 °C.[19] To optimize cleavage conditions of ADAR1, another small-scale study was performed at 1:6 ratio of ADAR1:TEV at room temperature for 4 h, and 4 ºC for 12 h (**Figure 3.4 B**). The conditions allowed for full removal of the MBP tag. The cleaved ADAR1d was separated from MBP using a heparin column. This resulted in successful separation of the proteins. Next, the activity of the cleaved ADAR1d was assessed against the hGli1 substrate. The cleaved protein achieved 100% editing at 150nM, while the MBP-ADAR1d only retained ~40% editing at this concentration (**Figure 3.4C**). Unfortunately, the cleaved ADAR1d did not retain similar solubility, it suffered severe precipitation during TEV protease cleavage step and the concentration steps after the heparin column. Therefore, this strategy could not be utilized for ADAR1 crystallography.

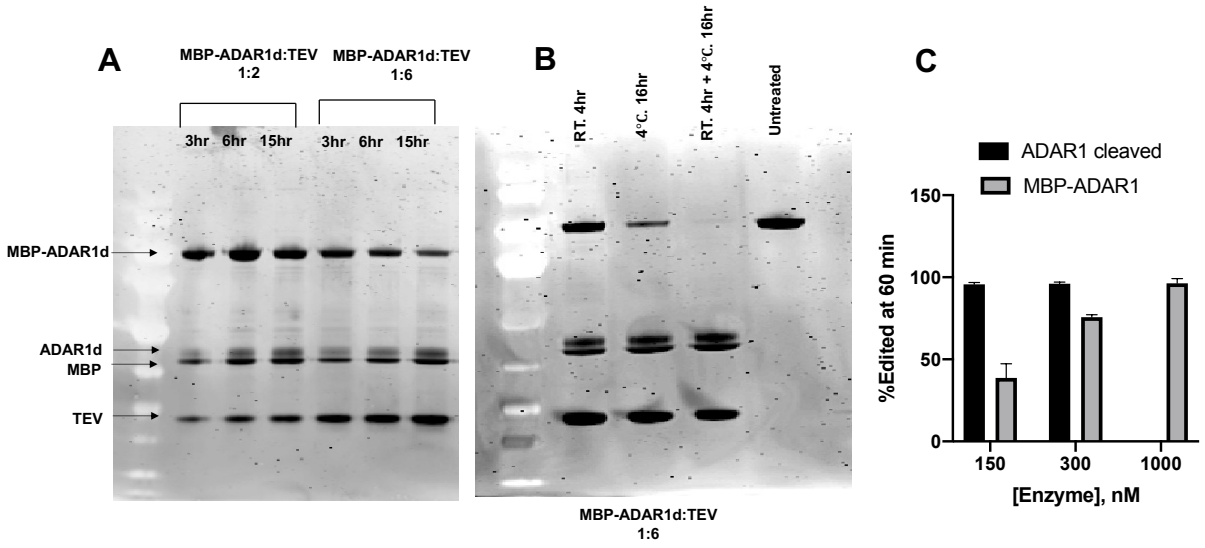


Figure 3.4. **(A)** TEV protease cleavage of MBP-ADAR1d at varying 1:2 or 1:6 ratio at varying incubation times. **(B)** TEV protease cleavage of MBP-ADAR1d at 1:6 ratio at varying temperatures and incubation time. **(C)** Activity of cleaved ADAR1d in comparison to un-cleaved MBP-ADAR1d with hGli1 mRNA.

MBP-ADAR1b R3D has distinct aggregation states. With some advancements in purification techniques and the identification of the structural zinc binding site, my focus shifted towards creating an extended ADAR1 construct that better mirror the full-length protein. This was intended to facilitate biochemical investigations of loss of function ADAR1 mutants involved in AGS disease. Additionally, we came across a shorter ADAR1 isoform lacking 20 hydrophobic residues (**Figure 3.5 A, B**).[20] This sparked our curiosity about whether the solubility of longer ADAR1 constructs could be improved by eliminating these hydrophobic residues. To ensure proper expression of AGS mutants, the N-terminal MBP tag was incorporated. Notably, the MBP-ADAR1 R3D construct exhibited heightened editing efficiency compared to the deaminase domain

alone, particularly evident in its interaction with the hGli1 substrate (**Figure 3.5 C**). Incorporating a dsRBD domain appeared to enhance substrate engagement.

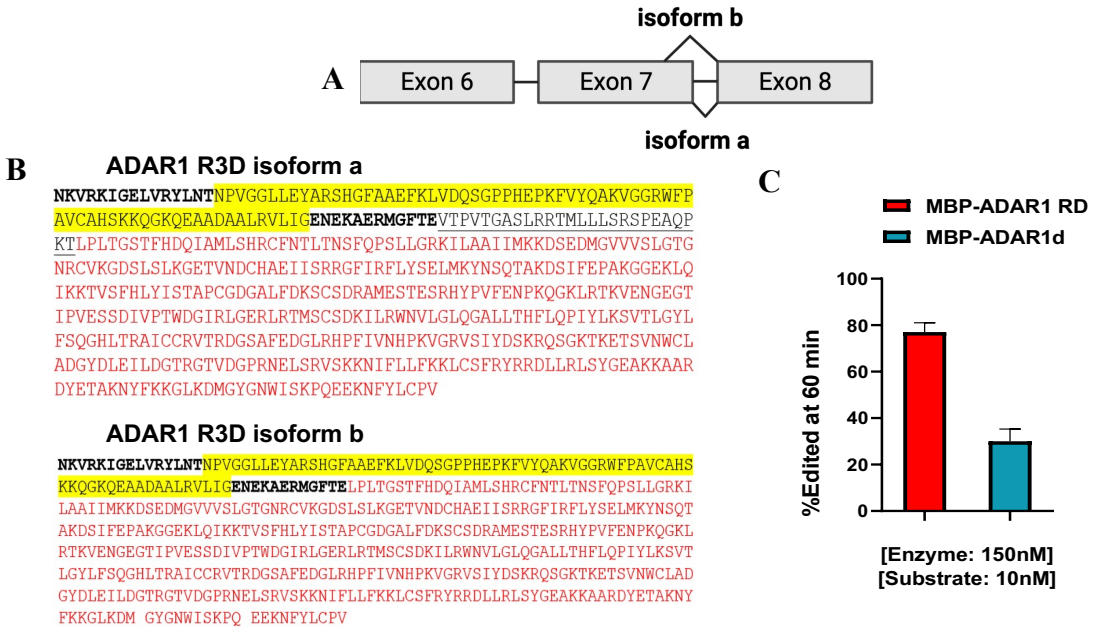
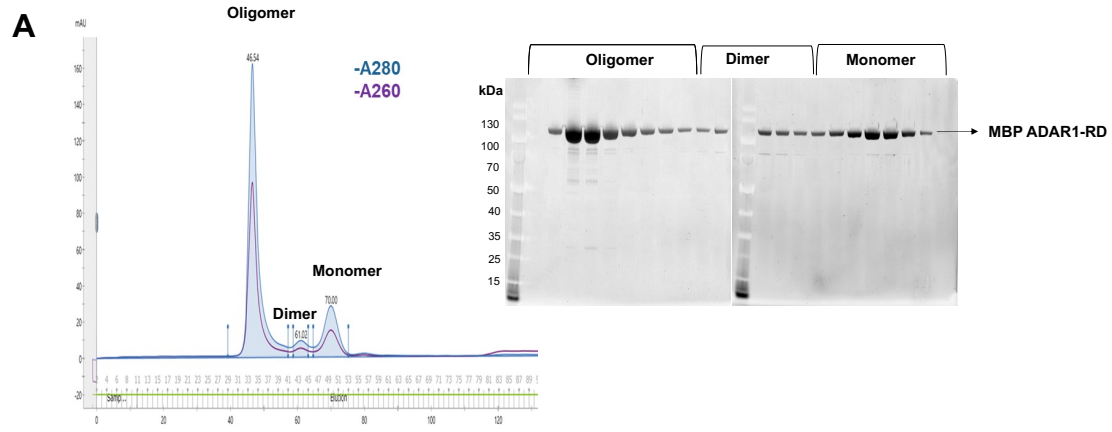


Figure 3.5. **(A)** Exon splicing of ADAR1 resulting in full-length isoform a, and another spliced isoform (-26 amino acid) resulting in truncated protein isoform b. **(B)** Amino acid composition of dsRBD3(yellow), linker (black), and deaminase domain (red). Truncation of 26 hydrophobic amino acid in the linker between dsRBD3 and deaminase domain results in isoform b. **(C)** Comparison of editing activity of MBP-ADAR1b-R3D, and ADAR1d against hGli1 substrate.

In our pursuit to assess the potential of the MBP tag to enhance the solubility of misfolded ADAR1, we performed an analytical size exclusion chromatography. The eluate pattern in the gel filtration revealed an intriguing behavior of the protein, characterized by the separation into three distinct states: monomer, dimer, and oligomer (**Figure 3.6A**). After analyzing these samples through a reducing page gel, these peaks corresponded to the same protein. This result indicated that ADAR1b R3D exists in different oligomeric states which may be in an equilibrium. To gain

deeper insights into this phenomenon, the peaks corresponding to the monomer and oligomer was subjected back into the size exclusion column (SEC). The intent was to examine if they would equilibrate into the distinctive peaks observed earlier. This investigation revealed that all the distinct peaks seen in the size exclusion analysis eventually equilibrated into monomer, dimer, and oligomer states. Notably, the monomeric species exhibited a higher fraction of the protein in its monomeric state (**Figure 3.6B**).



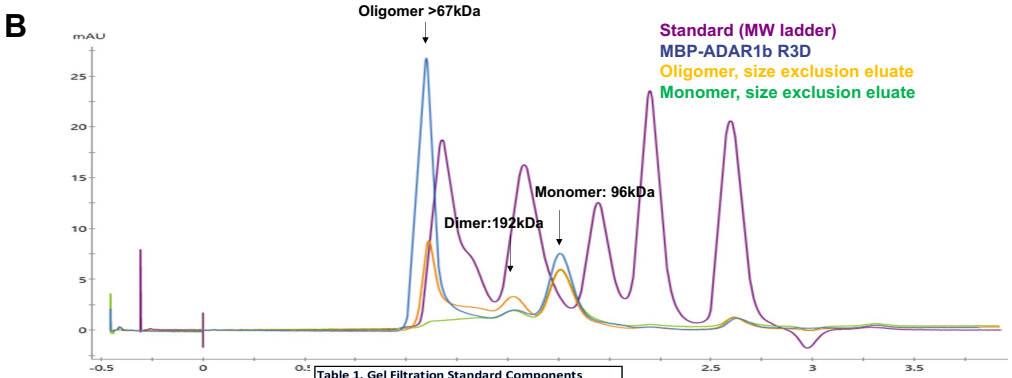
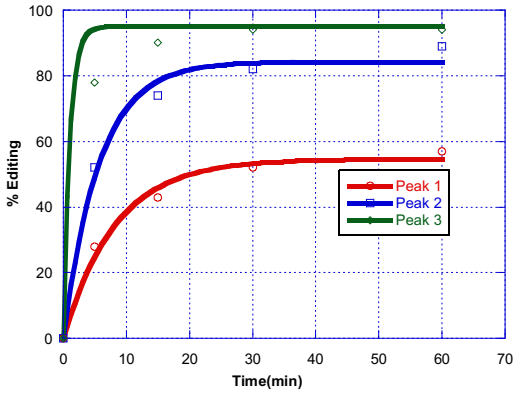


Table 1. Gel Filtration Standard Components	
Component	**Molecular Weight(kDa)**
Thyroglobulin(bovine)	670,000
γ-globulin(bovine)	158,000
Ovalbumin(chicken)	44,000
Myoglobin(horse)	17,000
Vitamin B12	1,350
MBP-ADAR1 dsRBD	96,000

Monomer

Figure 3.6. **(A)** Size exclusion analysis of MBP-ADAR1b R3D reveals its distinct separation into a monomer, dimer, and oligomer. **(B)** Comparative analysis of MBP-ADAR1b R3D, Monomer, and Oligomer against a molecular weight standard.

Subsequently, the functional activity of these distinct states was tested using the hGli1 substrate. Our findings indicated that the oligomeric states of the protein impede the rate and extent of editing (**Figure 3.7**). This analysis captured the complex behavior of ADAR1, and the ability of MBP to solubilize the differing aggregated states. Given the significance of the structural zinc binding site, I inquired whether the presence or absence of zinc in the purification buffers affected these aggregated states. To address this question, I evaluated the activity of the oligomeric state of the hyperactive MBP-ADAR1b R3D E1008Q, purified with and without zinc, using the hGli1 substrate. Indeed, we discovered the oligomer demonstrated greater aggregation, and reduced reaction rate when purified without zinc. Furthermore, concentration dependent aggregation was observed with hyperactive mutant E1008Q, and the wild-type protein as demonstrated by the reduced activity at higher concentrations. (**Figure 3.8**)



Enzyme	k_{obs} min^{-1}
Oligomer	0.12
Dimer	0.18
Monomer	≥ 1.00

Figure 3.7. Comparison of the activity of MBP-ADAR1b R3D, monomer, dimer, and oligomer against the hGli1 substrate. Reactions were performed under single turn-over conditions, with 10 nM hGli1, 75 nM MBP- ADAR1b R3D.

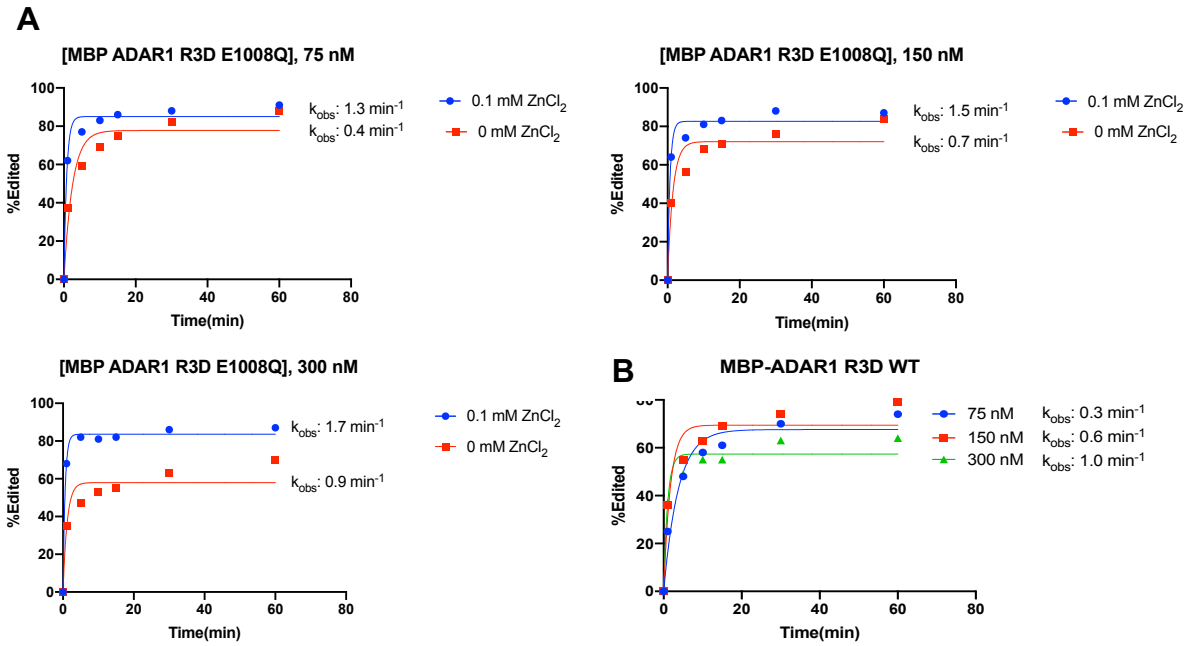
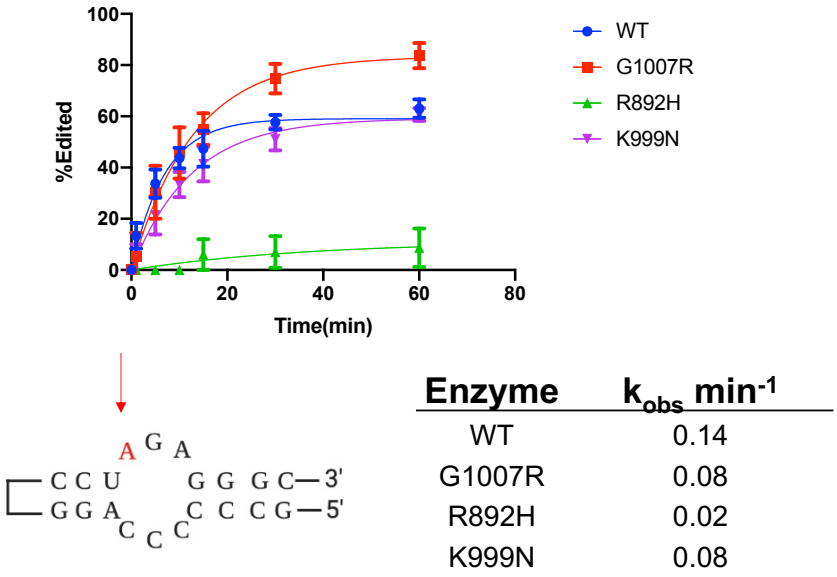


Figure 3.8. (**A**) Comparison of the activity of MBP-ADAR1b R3D E1008Q purified with and without zinc against the hGli1 substrate at varying concentrations (75 nM, 150 nM, 300 nM). (**B**) Comparison of MBP-ADAR1b R3D wild-type at 75 nM, 150 nM, and 300 nM.

Observing the rate enhancement of the oligomeric state in buffers containing zinc, we extended our investigation to MBP-ADAR1b R3D wild-type and specific AGS mutants (G1007R, K999N, and R892H). These variants were purified in buffers containing 0.1mM $ZnCl_2$. The samples were maintained as a mixture without separation into monomeric, dimeric, or oligomeric species. The kinetics of both the wild-type, and AGS mutants were subsequently assessed against the hGli1 substrate. Unexpectedly, the G1007R mutant, predicted to hinder base flipping, exhibited

robust activity.[21] Furthermore, the wild-type variant reached an editing extent of only 60%, underscoring the intricate behavior of the MBP-fused ADAR1 construct.

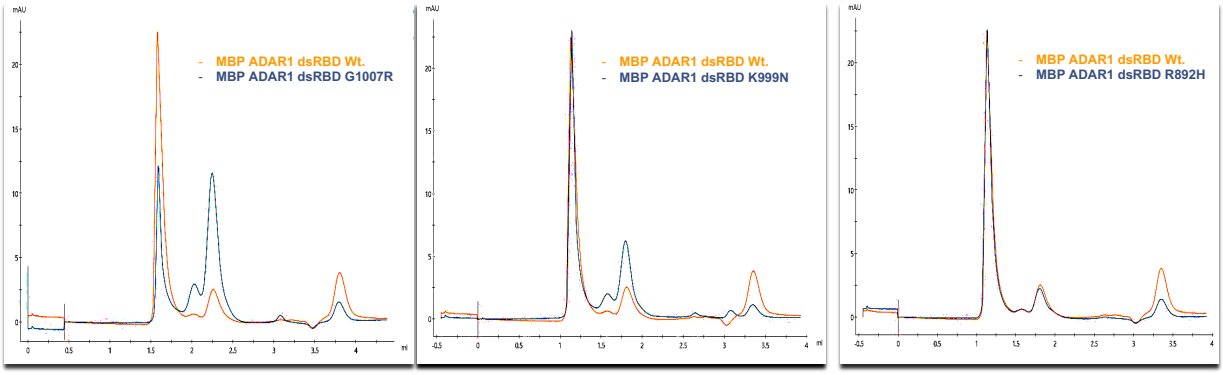


Enzyme	k_{obs} min^{-1}
WT	0.14
G1007R	0.08
R892H	0.02
K999N	0.08

Figure 3.9. Comparison of the activity of MBP-ADAR1b R3D WT and AGS mutants against the hGli1 substrate. Summarized is the observed rate constants of reactions with 10 nM hGli1 substrate and 100 nM MBP-ADAR1 R3D

To investigate whether the kinetics of the wild-type and AGS mutants were influenced by their oligomeric states, a size exclusion analysis was performed (**Figure 3.10**). The distribution of peaks in the size exclusion chromatograms revealed distinct patterns. The wild-type and the R892H mutant were predominantly present as oligomers. In contrast, the K999N mutant exhibited a higher proportion in the monomeric state (32%) compared to the wild-type (17%). Notably, the G1007R mutant primarily distributed across the monomeric (55%) and dimeric (13%) states, with only 32% in the oligomeric state. Interestingly, the kinetic profile of the wild-type and AGS mutants exhibited a correlation with the proportion of the oligomeric state in the sample. While

the solubility enhancement from MBP was initially promising, in longer constructs, it solubilized misfolded or aggregated states. Consequently, this construct did not appropriately represent the true consequences of AGS mutants in ADAR1.



Sample	%Oligomer	%Dimer	%Monomer	Monomer/Dimer	Monomer/Oligomer
ADAR1 MBP dsRBD Wt.	78.2 ± 1.2	4.7± 0.2	17.1 ± 1.2	3.6 ± 0.3	0.2
ADAR1 MBP dsRBD G1007R	32.1 ± 0.7	13.0 ± 0.3	54.9 ± 1.0	4.2 ± 0.2	1.7 ± 0.1
ADAR1 MBP dsRBD K999N	58.6 ± 0.4	9.1 ± 0.6	32.3 ± 0.5	3.6 ± 0.3	0.6
ADAR1 MBP dsRBD R892H	77.6 ± 2.7	5.3 ± 0.9	17.2 ± 1.9	3.3 ± 9.2	0.2

Figure 3.10. Size-exclusion profile of wild-type and three AGS mutations G1007R, K999N, and R892H. Summarized is the percent of monomer, dimer and oligomer in each chromatogram.

Impact of zinc chloride in ADAR1 p110 C-his purification. In pursuit of a more representative construct, I began efforts in improving purification methods of the his-tag ADAR1 p110 isoform a. A former lab member, Dr. Vanda Gaded, had devised a c-terminal his-tag construct for p110 and formulated the initial purification protocol. Furthermore, Dr. SeHee Park had fine-tuned the purification process for the ADAR1d construct in the context of crystallography and had noted the destabilizing impact of $ZnCl_2$ along with the solubilizing effect of 50 mM imidazole. Building on these insights, I confirmed that the absence of zinc chloride led to enhanced

recovery of ADAR1 p110, evident from robust protein bands in a 2 L purification, in contrast to faint bands observed when using 50 uM $ZnCl_2$ (**Figure 3.11**).

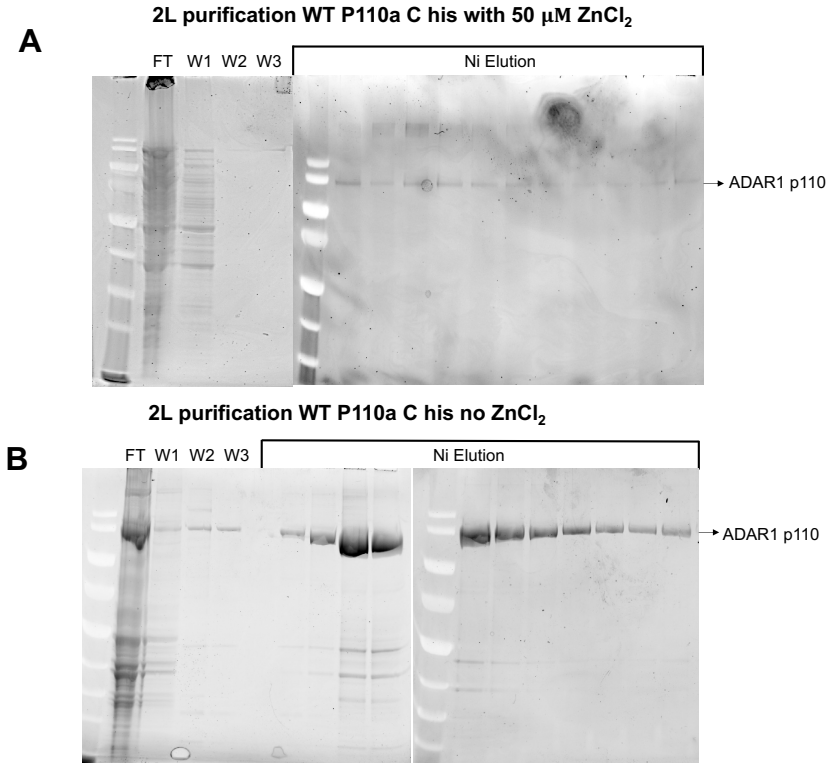


Figure 3.11. (**A**) Representative page-gel of p110a C-his purified in buffer containing 50 uM $ZnCl_2$. (**B**) Representative page-gel of p110a C-his purified without $ZnCl_2$.

Nevertheless, the presence of ZnCl2 proved crucial for attaining active protein, as evidenced by lower editing levels observed in samples lacking $ZnCl_2$ (**Figure 3.12**). However, it's worth noting that these samples also contained 2-fold higher levels of imidazole in the final buffer composition, therefore multiple factors could have influenced reaction kinetics. To investigate the possibility of saturating a second zinc site during later purification stages, a deamination assay was performed by varying zinc concentrations in the reaction conditions. Yet, the addition of zinc

chloride during the assay failed to enhance reaction rates or the endpoint indicating the need of $ZnCl_2$ supplementation in early stages of the purification. Additionally, when concentrating the protein in buffer containing 50 uM $ZnCl_2$, I observed a higher propensity for precipitation in samples lacking 50 mM imidazole supplementation (**Figure 3.13 A**). Guided by these findings, the refined purification process for p110 ADAR1 consist of a buffer comprising 50 uM $ZnCl_2$ and 50 mM imidazole, except in storage conditions, where zinc chloride was omitted. This finding was discovered in collaboration with Herra Grajo Mendoza. These purification conditions led to well-behaved ADAR1 p110 with robust editing of hGli1 and the 5-HT_{2c} substrate thus enabling studies of AGS mutants in ADAR1 (**Figure 3.13 B**).[22]

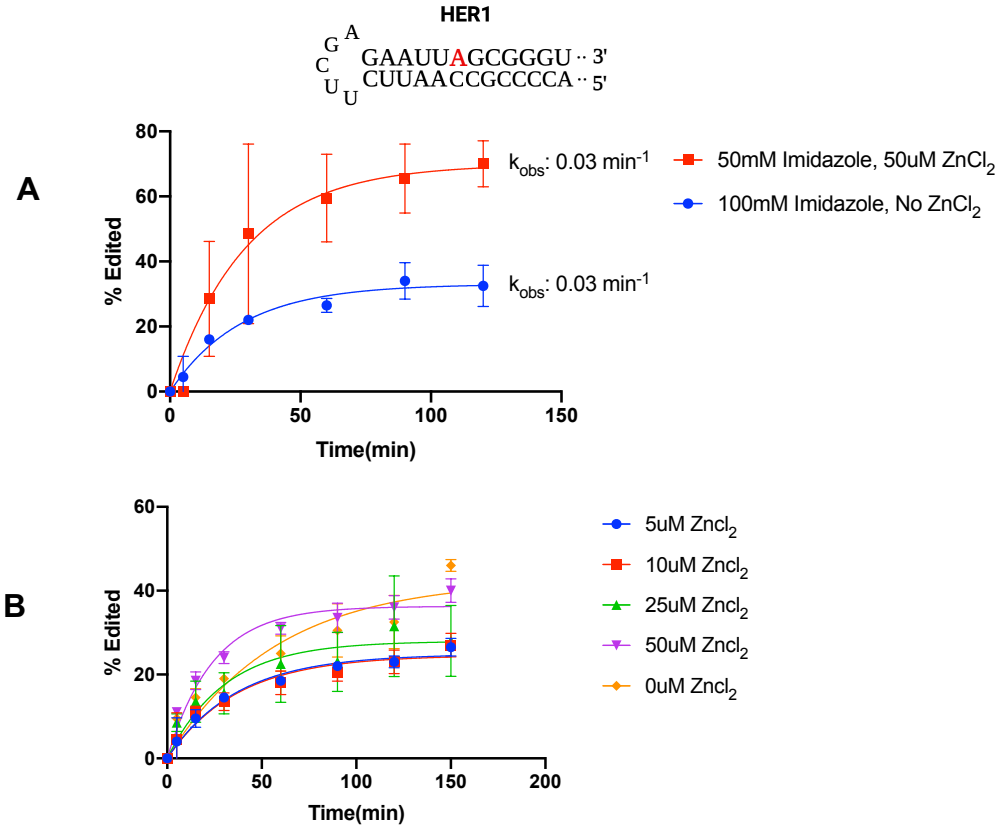


Figure 3.12. (**A**) Comparison of the activity of p110a C-his purified in buffer containing 0 or 50 uM $ZnCl_2$ against the HER1 substrate . Reaction was performed under single-turn over conditions with 10 nM HER1 and 100 nM p110. (**B**) Comparison of p110 activity against the HER1 substrate in varying concentrations of $ZnCl_2$. Reaction was performed under single-turn over conditions with 10 nM HER1 and 100 nM p110.

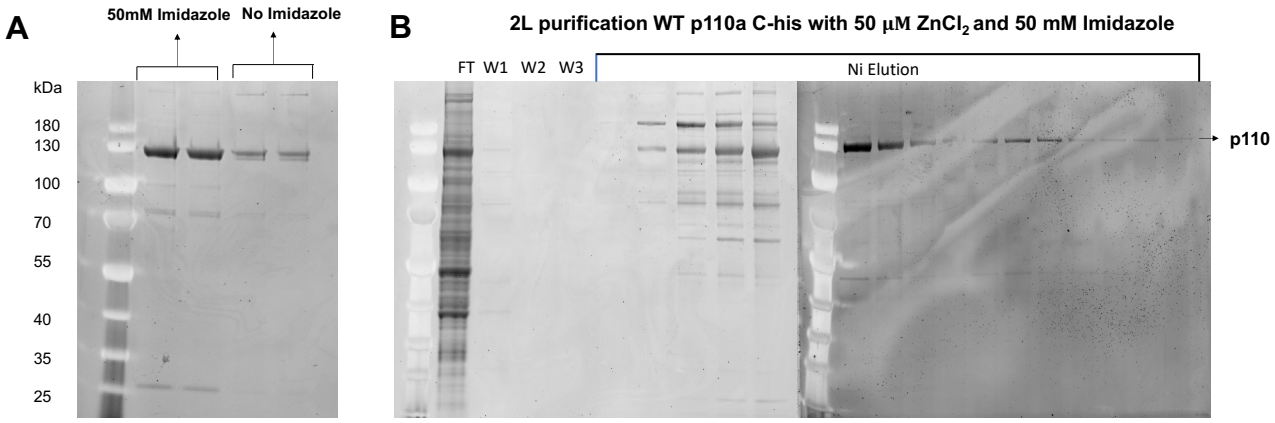
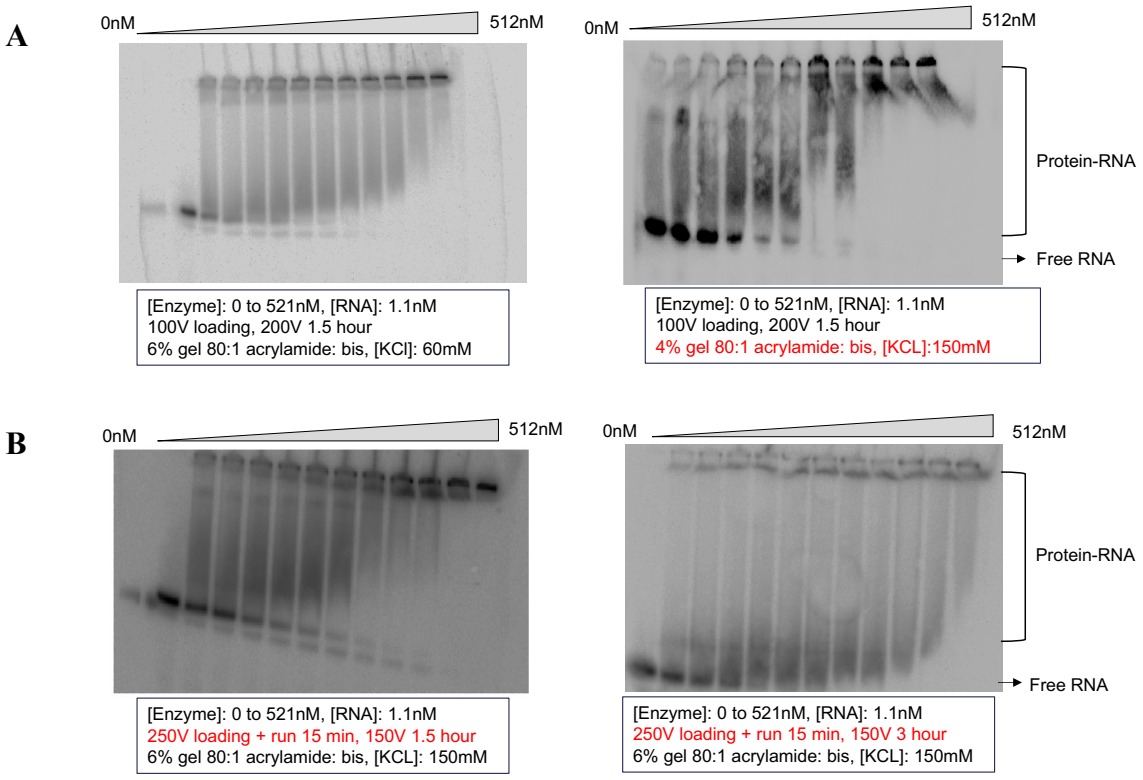


Figure 3.13. (**A**) Representative page-gel of p110a C-his concentrated with or without 50 mM imidazole supplementation. (**B**) Representative page-gel of p110a C-his purified in optimized conditions, with buffers containing 50 uM $ZnCl_2$ and ≥ 50 mM imidazole. The final protein was stored in (50 mM Tris-HCL pH 8.0, 10 % glycerol, 0.1% (v/v) NP-40, 400 mM KCl, 50 mM imidazole).

EMSA binding studies with 8-azaN containing dsRNA to trap ADAR1 p110–dsRNA complex. Encouraged by the improved purification approach, we investigated the interaction between wild-type p110 and a 32-base pair (bp) hGli1 RNA containing adenosine analog 8-Azanebularine (8-azaN). Due to lack of a good leaving group, the hydrated 8-azaN product effectively traps the protein-RNA complex.[16] Our aim was to understand the potential impact of AGS mutants on RNA binding. Initially, we conducted experiments where 1.1 nM or 5 nM dsRNA was incubated with varying concentrations of ADAR1 p110 (ranging from 0 to 521 nM) under

different buffer conditions. (**Figure 3.14 A-B**) Unfortunately, the resulting complex exhibited unfavorable behavior in a gel shift assay, showing intense aggregation, low migration, and gel smearing. Distinguishing whether these shifts resulted from complex formation or co-aggregated species proved challenging. Despite extensive testing of factors such as pH, salt concentrations and solubilizing agents, the complex did not migrate effectively in the gel. A summary of all tested conditions is provided below (**Figure 3.14 B**).

]



5'-GCUCGCGAUGCU**N**GAGGGCUCUGAUAGCUACG**-3'**
3'-CGAGCGCUACGACCCCCCGAGACUAUCGAUGC**-5'**
N = 8-azanebularine

Buffer	[Salt]	BSA	Yeast tRNA	%Glycerol	Gel% + Acrylmide: Bis	Loading Voltage	Run Voltage	Run time	[RNA]	Incubation temp/time	Aggregation?
15mM Tris-HCL pH 7.5	60mM KCL	100ug/ml	1ug/ml	3%	6% 80:1	100V, 20min	200V	1.5h	1.1nM	30C, 30 min	Yes
15mM Tris-HCL pH 7.5	150mM KCL	100ug/ml	2ug/ml	3%	4% 80:1	100V, 20min	200V	1.5h	1.1nM	30C, 30 min	Yes
15mM Tris-HCL pH 7.5	150mM KCL	100ug/ml	1ug/ml	3%	6% 80:1	250V, 15min	150V	1.5h	1.1nM	30C, 30 min	Yes
15mM Tris-HCL pH 7.5	150mM KCL	100ug/ml	1ug/ml	3%	6% 80:1	250V, 15min	150V	3h	1.1nM	30C, 30 min	Yes
15mM Bis-Tris pH 6.7	150mM KCL	100ug/ml	1ug/ml	3%	6% 80:1	250V, 15min	150V	1.5h	1.1nM	30C, 30 min	Yes
15mM Tris-HCL pH 7.5	150mM KCL	100ug/ml	1ug/ml	10%	6% 80:1	250V, 15min	150V	1.5h	1.1nM	30C, 30 min	Yes
15mM Tris-HCL pH 7.5	150mM KCL	100ug/ml	1ug/ml	3%	6% 80:1	250V, 15min	150V	1.5h	1.1nM	4C, 30min	Yes
10mM Tris-HCL pH 8.0	100mM NaCl	0ug/ml	0ug/ml	10%	4% 40:1	250V, 15min	150V	1.5h	1.1nM	30C, 30 min	Severe
15mM Tris-HCL pH 7.5	150mM KCL	100ug/ml	1ug/ml	3%	6% 80:1	250V, 15min	150V	1.5h	3nM	30C, 30 min	Yes
15mM Tris-HCL pH 7.5	150mM KCL	100ug/ml	1ug/ml	3%	6% 80:1	250V, 15min	150V	1.5h	3nM	30C, 30 min	Yes

Figure 3.14. (**A**) Representative gel shift results of p110 wild-type with a 32 bp 8-azaN bearing duplex. Protein concentrations were varied from 0-512 nM (0, 0.5, 1, 2, 4, 8, 16, 32, 65, 130, 260, and 521 nM) and RNA concentration was kept at either 1.1 nM or 5 nM. (**B**) Summary of conditions varied to optimize the gel shifts of ADAR1 p110 and hGli1 32 bp dulex.

Following iterative refinements, we finally identified a 61 bp hGli1 duplex bearing 8-azaN that exhibited well-behaved complex formation. This final optimization was led by Dr. Sukanya Mozumder, and the outcomes for both wild-type and AGS mutants are outlined in chapter 2. Collectively, our findings revealed that incubation of the ADAR1 p110-61 bp dsRNA complex at 30°C resulted in pronounced aggregation, rendering the complex immobile on the gel. Conversely, incubation at 4°C led to diminished RNA binding. The optimal protein-RNA complex formation was achieved through incubation at room temperature, utilizing a buffer comprising 20 mM Tris−HCl at pH 7.4, 140 mM KCl, 10 mM NaCl, 1 mM MgCl2, 0.5 mM EDTA, and 0.003% (v/v) NP-40 (**Figure 3.15**). These results underscore the susceptibility of ADAR1 to elevated temperatures and salt compositions. Consequently, future investigations involving ADAR1 in gel

shift analyses may gain from employing buffers with elevated salt content, experimenting with construct appropriate RNA lengths, and incubating at room temperature.

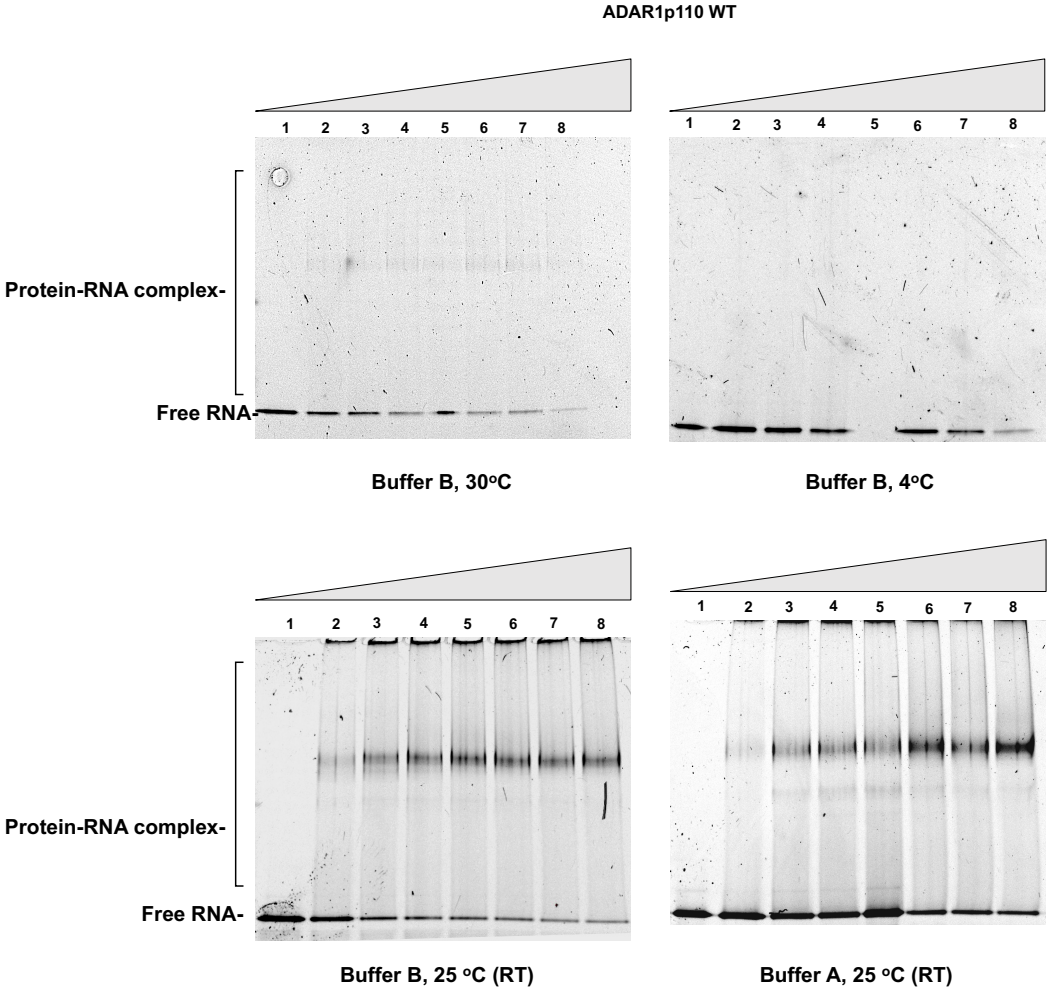


Figure 3.15. Representative gel shift results of p110 wild-type with a 61 bp 8-azaN bearing duplex. Protein concentrations were varied from lanes 1-8: 0, 5, 10, 20, 40, 60, 80 and 100 nM and RNA concentration was kept at 20 nM. The EMSA was performed at varying temperatures (4°C, 25 °C, and 30 °C) in two different buffer conditions: buffer B (15 mM Tris–HCl pH 7.5, 26 mM KCl, 40 mM potassium glutamate, 1.5 mM EDTA, 0.003% (v/v) NP-40, 4% glycerol, 0.5 mM DTT, 1 µg/mL yeast tRNA, 0.16 U/µL RNAse inhibitor) or buffer A (20 mM Tris−HCl at pH 7.4, 140 mM KCl, 10 mM NaCl, 1 mM MgCl2, 0.5 mM EDTA, and 0.003% (v/v) NP-40). The best gel shift was observed with buffer A and room temperature incubation for 30 min.

3.3. Discussion.

Despite sharing homology with the deaminase domain of hADAR2, ADAR1 is prone to significant aggregation and precipitation, impeding biochemical investigations.[18] The introduction of an N-terminal MBP fusion tag has proved instrumental in enhancing the solubility of the deaminase domain, facilitating the study of cystine mutants that would otherwise suffer from low expression, instability, or precipitation. SAT-FACS-seq analysis of multiple cystine mutants unveiled a structural zinc binding site within ADAR1. Notably, investigations involving the deaminase domain and other constructs emphasized the crucial role of the structural zinc site in maintaining the active conformation of the protein. This was evident from the absence of editing when C1081 was mutated to aspartic acid or glutamic acid, as well as the limited editing observed in protein samples purified without zinc supplementation.[18]

While the incorporation of the MBP-fusion tag improved the solubility of hADAR1d, its application to longer ADAR1 constructs led to the solubilization of improperly folded protein fractions, complicating the analysis of disease-associated mutations. The intriguing relationship between the structural zinc binding site and ADAR1's solubility became apparent. Preparing protein samples without zinc supplementation resulted in elevated protein yields but diminished protein activity. Conversely, supplementation with 50 uM zinc chloride led to reduced protein recovery but yielded active protein fractions. Furthermore, an optimal purification strategy for hADAR1 p110 was achieved through the supplementation of 50 mM imidazole alongside 50 uM $ZnCl_2$. The mechanism behind imidazole enhanced solubility, complete metyllation of second zinc site or its role in protein-protein interaction remains incompletely understood. Despite these improvements in purification, gel shift studies of hADAR p110 proved to be challenging. Severe

aggregation, co-precipitation of RNA, and limited migration of the protein-RNA complex on the gel were observed. Optimization of reaction conditions unveiled the susceptibility of the ADAR1-RNA complex to aggregation at elevated temperatures and low salt conditions. The optimal gel shift conditions were achieved using a high salt buffer, a 61 bp duplex, and incubation at room temperature (**Figure 3.15**).

The unresolved solubility challenges discussed in this chapter continue to be a subject of ongoing research in our laboratory. Although the use of imidazole and zinc supplementation has shown promising results in achieving robust protein purification, its success relies on the condition of BCY123 yeast cells, effective transformation, and efficient protein expression. Further biochemical investigations are vital to delve into the role of the second zinc binding site in ADAR1 kinetics and oligomerization. Additionally, exploring different expression systems, examining post-translational modifications, or considering the use of small self-cleavable soluble fusion tags may lead to a dependable purification approach for both short constructs and full-length hADAR1. It is possible an RNA is required for stabilization of ADAR1; therefore, the lab is currently exploring purification or ADAR1-RNA complexes to isolate well-folded active proteins fractions. This approach involves the use of biotinylated RNA, which can be captured using an avidin column. The protein can subsequently be separated by disrupting the protein-RNA interaction through high salt conditions. Achieving properly folded ADAR1 for comprehensive biochemical and structural studies will necessitate the implementation of various strategies and optimization procedures.

3.4. Methods.

Synthesis of oligonucleotides. RNA oligonucleotides containing 8-azanebularine (8-azaN) were synthesized using an in-house ABI 394 synthesizer at 0.2 μmol scale. The 8-azaN phosphoramidite was purchased from Berry & Associates. All other phosphoramidites were purchased from Glen research. Upon completion of oligonucleotide synthesis, columns were dried under reduced pressure overnight. Once dried, the oligonucleotides were cleaved from the solid support with the treatment of 1:3 ethanol/ 30% NH_4OH at 55 °C overnight. The supernatant was collected in a screw-cap tube and dried under reduced pressure. Removal of silyl groups was performed by suspending the pellets in anhydrous DMSO and treating it with 55% (v/v) Et_3N-3HF at room temperature overnight. The oligonucleotides were precipitated in solution of 65% butanol at -70 °C for 2 h. The pellets were obtained after centrifugation at 13200 x g for 20 min, washing twice with cold 70% ethanol. The RNA pellets were then desalted using a 3000 MWCO Amicon Ultra 0.5 mL centrifugal filter and purified as described below (see Table S1 for sequences). Oligonucleotide masses were confirmed by MALDI-TOF MS (Table S2).

In vitro transcription of RNA. Substrates for deamination studies (5-HT_{2c}R and hGlil) were transcribed from a DNA template using NEB HiScribe T7 RNA synthesis kit. The sequence of each substrate RNA can be found in Appendix Table 3.1.

Purification of RNA oligonucleotides. Single-stranded transcribed and synthesized oligonucleotides were purified using denaturing urea-polyacrylamide gels and visualized by UV shadowing. Bands corresponding to desired products were excised from gels, crushed, and soaked

overnight at 4 °C in 500 mM NH_4OAc and 1 mM EDTA. Polyacrylamide fragments were removed with a 0.2 μm cellulose acetate filter. RNA was precipitated from the supernatant in a solution of 75% EtOH at -70 °C for 2 h. The supernatant was centrifuged 17000 x g for 20 min and supernatant was removed and washed with 70% cold ethanol. The pellet was lyophilized to dryness, dissolved in nuclease free water, and quantified by absorbance at 260 nm. Synthesized oligonucleotides were desalted using 3000 MWCO Amicon Ultra 0.5 mL centrifugal filter, and masses for synthetic oligonucleotides were confirmed by MALDI-TOF MS Appendix Table 3.2.

Protein overexpression and purification. *MBP-ADAR1d and MBP-ADAR1b R3D.* A codon-optimized gene fragment of MBP for optimal expression in S. cerevisiae was ordered from GeneArt (Thermo Fisher Scientific). Cloning of MBP-tagged hADAR1d and ADAR1b R3D was carried out using Gibson master mix (New England Biolabs) following the manufacturer's protocol. Protein was expressed in S. cerevisiae BCY123 cells as previously described. Lysate obtained as described above was passed through 3 ml Amylose column (New England Biolabs) and washed with 50 ml of each wash buffers: wash I buffer (20 mM Tris-HCl, pH 8.0, 5% glycerol, 1 M NaCl, 0.1 mM $ZnCl_2$, 1 mM BME), wash II buffer (20 mM Tris-HCl, pH 7.5, 5% glycerol, 500 mM NaCl, 0.1 mM $ZnCl_2$, 1 mM BME), wash III buffer (20 mM Bis-Tris, pH 7.0, 5% glycerol, 75 mM NaCl, 0.1 mM $ZnCl_2$, 1 mM BME). The target protein was eluted with a 10 mM maltose in wash III buffer and further purified on a 5 ml HiTrap Heparin HP column (GE Healthcare Lifesciences) by washing with 50 ml of heparin wash buffer (20 mM Bis-Tris, pH 7.0, 5% glycerol, 75 mM NaCl, 0.1 mM $ZnCl_2$, 1 mM BME) and eluting with a linear gradient of 75 mM – 1 M NaCl in heparin wash buffer. Samples purified for structural studies were passed

through size exclusion column in SEC buffer (20 mM Tris-HCl, pH 8.0, 5% glycerol, 350 mM NaCl, 0.1 mM $ZnCl_2$, 1 mM BME). The purified protein was pooled and dialyzed against 50 mM Tris-HCl, pH 8.0, 10% glycerol, 200 mM KCl, 0.5 mM EDTA, 0.01% NP-40 substitute, 1 mM DTT for storage. Final protein concentrations were determined by running the samples alongside bovine serum albumin (BSA) standards in an SDS-PAGE gel and visualized by staining with SYPRO Orange dye (Invitrogen).

hADAR1 p110a C-his. 3L Cells were lysed using a microfluidizer in lysis buffer containing 20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 1 M KCl, 30 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine-HCl (TCEP-HCl), 0.05% (v/v) Triton X-100. The lysate was centrifuged at 39000 x g, 4 °C for 1 h and filtered using 0.45 μm filter. The clarified lysate was then passed over a 5 ml Ni-NTA column at a flow rate of 2 ml min^{-1} using an ÄKTA pure 25 FPLC system. The column was washed first with ten column volumes (10 CVs) of lysis buffer, followed with 10 CVs of wash I buffer (50 mM Tris-HCl pH 8.0, 5% glycerol ,5 mM BME,1000 mM KCL, 30 mM Imidazole, 0.01% (v/v) Nonidet P-40 (NP-40)). The protein was eluted with wash I buffer with a gradient of imidazole (30 mM to 300 mM) for 10 CVs. Fractions containing the target protein were pooled, concentrated, and dialyzed against a storage buffer containing 50 mM Tris-HCl pH 8.0, 10% (v/v) glycerol, 400 mM KCl, 1 mM TCEP-HCl, and 0.02% (v/v) Nonidet P-40 (NP-40). All buffers either contained 50 uM $ZnCl_2$ or no $ZnCl_2$. The final storage buffer either contained 50-, 100- or 0-mM Imidazole. Final protein concentrations were determined by running the samples alongside bovine serum albumin (BSA) standards in an SDS-PAGE gel and visualized by staining with SYPRO Orange dye (Invitrogen).

hADAR1 p110a C-his optimized. Cells were lysed using a microfluidizer in lysis buffer containing 20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 1 M KCl, 30 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine-HCl (TCEP-HCl), 0.05% (v/v) Triton X-100, and 50 μM $ZnCl_2$. The lysate was centrifuged at 39000 x g, 4 °C for 1 h and filtered using 0.45 μm filter. The clarified lysate was then passed over a 5 ml Ni-NTA column at a flow rate of 2 ml min^{-1} using an ÄKTA pure 25 FPLC system. The column was washed first with ten column volumes (10 CVs) of lysis buffer, followed with 10 CVs of wash I buffer (20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 500 mM KCl, 30 mM imidazole, 1mM TCEP-HCl, and 50 μM $ZnCl_2$). The protein was eluted with wash I buffer with a gradient of imidazole (30 mM to 400 mM) for 10 CVs. Fractions containing the target protein were pooled, concentrated, and dialyzed against a storage buffer containing 50 mM Tris-HCl pH 8.0, 10% (v/v) glycerol, 400 mM KCl, 50 mM imidazole, 1 mM TCEP-HCl, and 0.01% (v/v) Nonidet P-40 (NP-40).

Analytical size exclusion analysis. For analytical gel filtration analysis, 25 μl of 10 μM MBP-ADAR1b R3D, was injected onto a Superdex increase 5/150 GL gel filtration column using an AKTA FPLC system (GE healthcare). The protein was eluted with 20 mM Tris–HCl pH 8.0, 350 mM NaCl, 5% glycerol and 1 mM BME. The elution profile was analyzed by the evaluation tools included in the Unicorn 7.4 software package. To assess relative size of the eluted fractions, gel filtration column using molecular weight standards (Biorad and Sigma).

Deamination assays. Deamination assays were performed under single-turnover conditions in 15 mM Tris–HCl pH 7.5, 26 mM KCl, 40 mM potassium glutamate, 1.5 mM EDTA, 0.003% (v/v) NP-40, 4% glycerol, 0.5 mM DTT, 1 μg/mL yeast tRNA, 0.16 U/μL RNAse

inhibitor, 10 nM RNA, and 100 nM protein at 30 °C. Reactions were quenched at 1, 5, 10, 15, 30, 60 min with 20 ul water at 95°C and heating at 95°C for 5 min. Reaction products were used to generate cDNA using RT-PCR (Promega Access RT-PCR System). The DNA product was purified with DNA Clean & Concentrator kit (Zymo) and subjected to Sanger sequencing via GeneWiz (Azenta). Sequencing peak heights at the edit site was quantified using 4Peaks (Nucleobytes). Data were fit to the equation $[P]t = [P]_f[1-e^{(-kobs \cdot t)}]$ where $[P]t$ is percent edited at time t, $[P]f$ is the final endpoint of editing, and k_{obs} is the observed rate constant. All statistical analyses and nonlinear fits were conducted in Microsoft Excel and GraphPad Prism.

Preparation of duplex substrates for EMSA with ^{32}P-Labeled RNA. The 5' end of either of a 32 nt 8-azaN-containing RNA derived from hGli1 substrate was labeled with γ-[^{32}P]ATP (6000 Ci/mmol) using NEB polynucleotide kinase. The labeled reaction was passed through a G-25 column to remove excess ATP and further purified using a 19% denaturing PAGE gel. The labeled products were visualized using storage phosphor autoradiography. Gel bands containing labeled RNA were excised, crushed, and soaked and worked up as described for other gel-purified oligonucleotides. The dried pellet was resuspended in nuclease free water to a stock solution of approximately 300 nM and hybridized to its compliment at 1:3 ratio in 1X TE buffer, pH 7.5 and 200 mM NaCl by heating at 95 °C for 5 min and slowly cooling to 30 °C to a final concentration of approximately 50 nM. Sequence of all oligonucleotides are found in the Appendix table 3.1.

Quantitative EMSA using ^{32}P duplex RNA and ADAR1 p110. Samples containing ≤ 1 nM or ≤ 5nM 32 base pair (bp) RNA duplex and varying concentrations of the ADAR1 p110 (0,

, 0.5, 1, 2, 4, 8, 16, 32, 65, 130, 260, and 521 nM) were incubated together in 20 mM Tris-HCl, pH 7.0, 3.5% glycerol, 0.5 mM DTT, 60 mM KCl, 20 mM NaCl, 0.1 mM β-mercaptoethanol, 1.5 mM EDTA, 0.003% Nonidet P-40, 0.16 U/μL RNAse inhibitor , 0.2 mg/mL BSA, and 1 μg/mL yeast tRNA for 30 min at 30 °C. Samples were loaded onto a 6% gel, electrophoresed in nondenaturing polyacrylamide gel (79:1 acrylamide:bisacrylamide) in 1× TBE buffer at 4 °C for 90 min. The gels were dried on a Bio-Rad gel dryer for 90 min at 80 °C under vacuum followed by exposure to storage phosphor imaging plates (Kodak) for 24 h in the dark. After exposure, the gels were removed, and the phosphor imaging plates were scanned by Typhoon Trio Variable Mode Imager (GE Healthcare). The sequences for the 32 bp duplex used in the assay are described in Appendix Table 3.1.

Qualitative EMSA using Cy5-labeled duplex RNA and ADAR1 p110. Samples containing 20 nM Cy5 labeled duplex RNA and 0 to 100 nM enzyme (ADAR1 p110 WT, K999N, R892H and Y1112F) and 0 to 120 nM enzyme (ADAR1p110 G1007R) were incubated respectively in 20 mM Tris−HCl pH 7.4, 140 mM KCl, 10 mM NaCl, 1mM $MgCl_2$, 0.5 mM EDTA and 0.003% (v/v) NP-40 at room temperature for 30 min. Samples were loaded onto a 4 to 16% Bis-Tris, NativePAGE mini protein gel (Invitrogen) and electrophoresed under nondenaturing conditions in 1× Native PAGE running Buffer (Invitrogen) at 4 °C for 1.5 h under light-shielded conditions. The gels were scanned using a Bio-Rad GelDoc Imaging system.

Appendix Chapter 3.

Appendix Table 3.1. Oligonucleotide sequences used in this work. All bases are ribonucleotides unless specified. N is 8-azanebularine.

Oligonucleotide Name	Sequence
32 nt 8-azanebularine top	5'- GCUCGCGAUGCUNGAGGGCUCUGAUAGCUACG -3'
32 nt bottom	5'- CGUAGCUAUCAGAGCCCCCCAGCAUCGCGAGC-3'
(3' end Cy5 label) 61 mer 8-azanebularine top	5'- AGCAAGUCCACGUGCAUGGCUCGCGAUGCUNGAGGGCUC UGAUAGCGGAUGGACAUCGACGCy5 -3'
61 mer bottom	5'- CGUCGAUGUCCAUCCGCUAUCAGAGCCCCCCAGCAUCGC GAGCCAUGCACGUGGACUUGCU -3'
23 nt 8-azanebularine top	5'- GCUCGGAGAAUUNGCGGGUCGUG -3'
23 nt bottom	5'- CACGACCCGCCAAUUCUCCGAGC -3'
HER1	5'AGAAGAGGAAGGTGAATTACATAAATGGGGTGTGTTTGTAC ACACCAAAGCTCAAGACATAACTCTGGGCTTCCTTCATCCG CCAATTCTTCGAGAATTAGCGGGTCGTTAACACCAGATAGCA GTGTCGCCGGAGGAAAGAAAGGGGAATCCTCTCGCACATCT GGTACAAGGCCAAAAATCTTACCAAAGATTCCTACCG-3'
5HT$_{2C}$R 332 nt	5'- UGGGUACGAAUUCCCACUUACGUACAAGCUUACCUAGAU AUUUGUGCCCCGUCUGGAUUUCUUUAGAUGUUUUAUUUUC AACAGCGUCCAUCAUGCACCUCUGCGCUAUAUCGCUGGAUC GGUAUGUAGCAAUACGUAAUCCUAUUGAGCAUAGCCGUUU CAAUUCGCGGACUAAGGCCAUCAUGAAGAUUGCUAUUGUU UGGGCAAUUUCUAUAGGUAAAUAAAACUUUUUGGCCAUAA GAAUUGCAGCGGCUAUGCUCAAUACUUUCGGAUUAUGUAC UGUGAACAACGUACAGACGUCGACUGGUAACAUUUGCGUU UGAUCGGGUUCU -3'
hGli1 155 nt	5'- CAGAACUUUGAUCCUUACCUCCCAACCUCUGUCUCUCUG UCUACUCACCACAGCCCCCCAGCAUCACUGAGAAUGCUGCC AUGGAUGCUAGAGGGCUACAGGAAGAGCCAGAAGUUGGGA CCUCCAUGGUGGGCAGUGGUCUGAACCCCUAUAUG -3'

Appendix Table 3.2. Observed and calculated masses of oligonucleotides.

Oligonucleotide Name	**Observed Mass (m/z)**	**Expected mass (a.m.u.)**
32 nt 8-azanebularine top	10329	10328
32 nt bottom	10203	10206
(For 61 bp, top 5' fragment) 38 nt 8-azanebularine	12303	12304
(For 61 bp, top 3' fragment) 23 nt	6994	6991
23 bp 8-azanebularine top	7456	7446
23 bp bottom	7250	7241

Appendix Table 3.3. RT-PCR and sequencing DNA primers used for in vitro deamination in this work.

Oligonucleotide Primers	Sequence
HER1 RT Forward	5'-AGAAGAGGAAGGTGAATTACATAAATGG-3'
HER RT Reverse	5'-CGGTAGGAATCTTTGGTAAGATTTTTG-3'
5HT$_{2C}$R RT Forward	5'- TGGGTACGAATTCCCACTTACGTACAAGCTT -3'
5HT$_{2C}$R Reverse	5'- AGAACCCGATCAAACGCAAATGTTAC -3'
hGli1 RT Forward	5'TAATACGACTCACTATAGGGCAGAACTTTGATCCTTACCTC 3'
*hGli1*Reverse	5'- CATATAGGGGTTCAGACCACTG -3'

3.4. References

(1) Bass, B. L. RNA Editing by Adenosine Deaminases That Act on RNA. *Annu. Rev. Biochem.* **2002**, *71* (1), 817–846.
(2) Xiong, X.; Yi, C.; Peng, J. Epitranscriptomics: Toward A Better Understanding of RNA Modifications. *Genomics Proteomics Bioinformatics* **2017**, *15* (3), 147–153.
(3) Wang, Y.; Zheng, Y.; Beal, P. A. Adenosine Deaminases That Act on RNA (ADARs). **2017**, 215–268.
(4) Eisenberg, E.; Levanon, E. Y. A-to-I RNA Editing — Immune Protector and Transcriptome Diversifier. *Nat. Rev. Genet.* **2018**, *19* (8), 473–490.
(5) Booth, B. J.; Nourreddine, S.; Katrekar, D.; Savva, Y.; Bose, D.; Long, T. J.; Huss, D. J.; Mali, P. RNA Editing: Expanding the Potential of RNA Therapeutics. *Mol. Ther.* **2023**, *31* (6), 1533–1549.
(6) Gannon, H. S.; Zou, T.; Kiessling, M. K.; Gao, G. F.; Cai, D.; Choi, P. S.; Ivan, A. P.; Buchumenski, I.; Berger, A. C.; Goldstein, J. T.; Cherniack, A. D.; Vazquez, F.; Tsherniak, A.; Levanon, E. Y.; Hahn, W. C.; Meyerson, M. Identification of ADAR1 Adenosine Deaminase Dependency in a Subset of Cancer Cells. *Nat. Commun.* **2018**, *9* (1), 5450.
(7) Liu, J.; Wang, F.; Zhang, Y.; Liu, J.; Zhao, B. ADAR1-Mediated RNA Editing and Its Role in Cancer. *Front. Cell Dev. Biol.* **2022**, *10*.
(8) Bhate, A.; Sun, T.; Li, J. B. ADAR1: A New Target for Immuno-Oncology Therapy. *Mol. Cell* **2019**, *73* (5), 866–868.
(9) Nakahama, T.; Kawahara, Y. The RNA-Editing Enzyme ADAR1: A Regulatory Hub That Tunes Multiple DsRNA-Sensing Pathways. *Int. Immunol.* **2023**, *35* (3), 123–133.
(10) Bass, B. L. RNA Editing by Adenosine Deaminases That Act on RNA. *Annu. Rev. Biochem.* **2002**, *71* (1), 817–846.
(11) Haudenschild, B. L.; Maydanovych, O.; Véliz, E. A.; Macbeth, M. R.; Bass, B. L.; Beal, P. A. A Transition State Analogue for an RNA-Editing Reaction. *J. Am. Chem. Soc.* **2004**, *126* (36), 11213–11219.

(12) 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 Res.* **2015**, *43* (2), 1123–1132.
(13) Wang, Y.; Park, S.; Beal, P. A. Selective Recognition of RNA Substrates by ADAR Deaminase Domains. *Biochemistry* **2018**, *57* (10), 1640–1651.
(14) Wong, S. K.; Sato, S.; Lazinski, D. W. Substrate Recognition by ADAR1 and ADAR2. *RNA* **2001**, *7* (6), 846–858.
(15) Matthews, M. M.; Thomas, J. M.; Zheng, Y.; Tran, K.; Phelps, K. J.; Scott, A. I.; Havel, J.; Fisher, A. J.; Beal, P. A. Structures of Human ADAR2 Bound to DsRNA Reveal Base-Flipping Mechanism and Basis for Site Selectivity. *Nat. Struct. Mol. Biol.* **2016**, *23* (5), 426–433.
(16) 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 Res.* **2020**, *48* (14), 7958–7972.
(17) Costa, S.; Almeida, A.; Castro, A.; Domingues, L. Fusion Tags for Protein Solubility, Purification and Immunogenicity in Escherichia Coli: The Novel Fh8 System. *Front. Microbiol.* **2014**, *5*.
(18) Park, S.; Doherty, E. E.; Xie, Y.; Padyana, A. K.; Fang, F.; Zhang, Y.; Karki, A.; Lebrilla, C. B.; Siegel, J. B.; Beal, P. A. High-Throughput Mutagenesis Reveals Unique Structural Features of Human ADAR1. *Nat. Commun.* **2020**, *11* (1), 5130.
(19) Tropea, J. E.; Cherry, S.; Waugh, D. S. Expression and Purification of Soluble His6-Tagged TEV Protease. In *High Throughput Protein Expression and Purification*; Doyle, S. A., Ed.; Walker, J. M., Series Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2009; Vol. 498, pp 297–307.
(20) Yang, J.-H.; Nie, Y.; Zhao, Q.; Su, Y.; Pypaert, M.; Su, H.; Rabinovici, R. Intracellular Localization of Differentially Regulated RNA-Specific Adenosine Deaminase Isoforms in Inflammation. *J. Biol. Chem.* **2003**, *278* (46), 45833–45842.
(21) Fisher, A. J.; Beal, P. A. Effects of Aicardi-Goutières Syndrome Mutations Predicted from ADAR-RNA Structures. *RNA Biol.* **2017**, *14* (2), 164–170.
(22) Mendoza, H. G.; Matos, V. J.; Park, S.; Pham, K. M.; Beal, P. A. Selective Inhibition of ADAR1 Using 8-Azanebularine-Modified RNA Duplexes. *Biochemistry* **2023**, *62* (8), 1376–1387.

Chapter 4

RNA Editing at Disfavored Sites: ADAR Activation by a G_{syn}:$Purine_{anti}$ Pair.

This work was a highly collaborative effort with Dr. Erin Doherty (Beal lab alumni 2022) who spearheaded the project through her discovery that a purine-purine mismatch enables efficient editing at disfavored 5'G sites. She performed all kinetic analysis presented in this work. I solved all structures presented in collaboration with Dr. Xander Wilcox (Fisher lab alumni 2023), and Herra Mendoza. This work was published in NAR on September 27, 2022.

4.1. Introduction

RNA molecules are frequently decorated by post-transcriptional modifications. Among the modifications found in an mRNA, Inosine presents itself to be most fascinating. This nucleobase is one of the few examples, in which the base pairing capabilities of the original nucleobase is completely altered. This enzymatic modification is carried out by ADARs (adenosine deaminases that act on RNA). [1,2] Since inosine is decoded like guanosine, the modification introduced by ADARs can change codon meaning [3,4]. ADARs are multidomain proteins with N-terminal double stranded RNA binding domains (dsRBDs) and C-terminal deaminase domains. There are two catalytically active ADARs in humans (ADAR1 and ADAR2) with ADAR1 expressed as two different protein isoforms (p110 and p150) bearing different N-terminal structures.[5] In a RNA duplex, ADARs access the reactive adenosine from the minor groove of the protein, using a base flipping mechanism.[6] Due to their ubiquitous expression in various tissues and their ability to precisely identify double-stranded RNA, ADARs can be harnessed to modify specific adenosine

residues within various transcripts.[7] This is achieved by employing guide strands that are complementary to these target locations.[7,8] This innovative approach is being pursued as an alternative to Cas-9 genome editing and DNA base editors, addressing challenges related to delivery and the inadvertent induction of DNA mutations.[7-11]

While this approach is promising, ADARs have sequence preferences that make certain adenosines disfavored for editing, limiting the current scope of this approach. The nearest neighbor nucleotide preferences for ADARs show a strong bias against reaction at adenosines with a guanosine 5′ to the target adenosine (5′-GA sites).[12,13] This preference is explained by our structural studies of ADAR2 bound to transition state analog-containing RNA that suggest a clash between the 2-amino group of the 5′ G and Gly489 of the ADAR2 loop involved in stabilizing the flipped out conformation required for the adenosine deamination reaction (**Figure 4.1**).[13] This clash would disrupt the base-flipped conformation, consequently reducing the editing efficiency. Notably, the Stafforst lab had previously shown that fusion proteins carrying ADAR deaminase domains enhance editing efficiency at 5′-GA sites when paired with G:A or G:G as the nearest neighbor.[14] However, it was unclear how this effect translates to full-length ADARs with their native dsRBDs. Following up on this work, Dr. Erin Doherty demonstrated G:A and G:G pairs on the 5′ side of an editing site improve editing efficiency compared to a 5′ G:C pair for full length ADAR2 and ADAR1 p110. This finding gave rise to two compelling hypotheses. One possibility is that a purine-purine mismatch may facilitate a crucial conformational change, which is necessary to prevent clashes with G489. Alternatively, these mismatches might be forming non-canonical but stable hydrogen bonding pairs, as previously observed in the literature.[15]

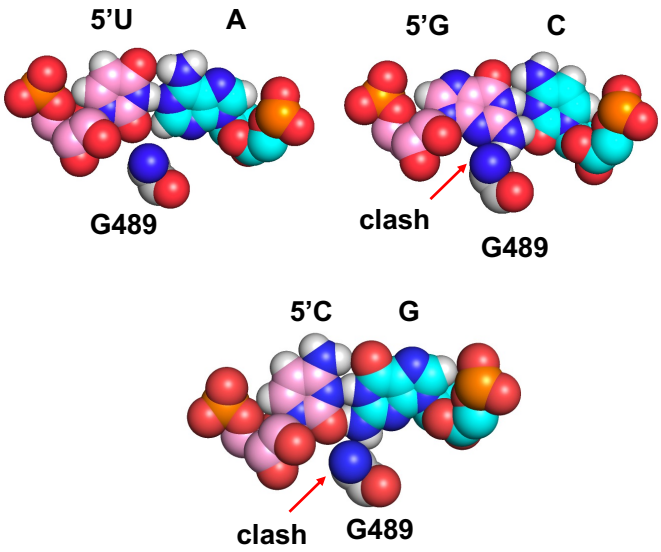


Figure 4.1. Model of ADARs sequence preference based on the structural analysis of hADAR2d in complex with dsRNA (PDB: 5ED2). In cases where there is a 5′G or a 5′C as the nearest neighbor, a clash occurs between the amino group of guanosines and the G489 residue. This clash ultimately leads to the destabilization of the base-flipped conformation.

This proposed mechanism involves a 5' G being flipped into a syn conformation, exposing its Hoogsteen face for base pairing and positioning the detrimental 2-amino group away from the G489 residue. This structural rearrangement could explain the observed enhancement in editing efficiency. To delve deeper into this, X-ray crystallography was employed to determine the structure of an active fragment of human ADAR2 bound to duplex RNA containing a G:G pair adjacent to an editing site. The results showed that the two guanosines indeed formed a hydrogen-bonded Gsyn:Ganti pair (**Figure 4.2**). I solved this structure in collaboration with Dr. Xander Wilcox from the Fisher lab. To elucidate the significance of this pairing, a comparative analysis was conducted with similar structures featuring U:A pairs adjacent to ADAR editing site.

4.2. Results

ADAR2 binds a duplex RNA substrate with a $G_{(syn)}$:$G_{(anti)}$ pair adjacent to the editing site. To provide the structural basis for efficient editing of ADAR with a G:G pair adjacent to an editing site, we turned to X-ray crystallography. First a 32 bp hGli1 derived duplex bearing 8-azanebularine (azaN) and a 5' G:G pair was prepared. (Figure 4.2A).When azaN is properly positioned within an ADAR substrate RNA, ADARs promote its covalent hydration to form a structure that mimics the adenosine deamination intermediate.[13,16] We introduced a G:G pair adjacent to the azaN and evaluated its interaction with ADAR2-R2D E488Q. The gel shift analysis revealed a well-defined complex with this with $K_d = 6 \pm 2$ nM (**Figure 4.2B**.). A similar duplex bearing the ideal nearest neighbor (5′-U:A) bound this protein with a $K_d = 0.9 \pm 0.5$ nM (**Figure 4.3**). In addition, the gel shifts formed a complex indicative of a monomer-RNA complex, which further shifted into a dimer-RNA complex at elevated concentrations. These patterns are comparable to the favored 5'U:A pair and supported by the available structure.[23]

Following our binding studies, protein–RNA crystals of the 32 bp azaN-containing duplex with the G:G pairing and ADAR2-R2D E488Q formed readily, and diffracted X-rays beyond 2.7 Å resolution (Table 3) (Figure 4.2C). As we have seen with the related protein–RNA combination[23], the protein bound the RNA as an asymmetric homodimer with the deaminase domain of one monomer (monomer A) involved in direct RNA binding to the flipped-out azaN nucleoside (Figure 4.2C). Importantly, the G:G pair adjacent to the editing site is well resolved with electron density that best fits a G_{syn}:G_{anti} pair with the guanosine on the 5′ side of the azaN in a *syn* conformation with its Hoogsteen face accepting two hydrogen bonds from the Watson–Crick

face of the guanosine on the opposing strand (**Figures 4.2D and 4.2A**). The guanosine of this guide strand is in an *anti*-conformation. The G:G pairing involves N1 to O6 and 2-amino to N7 hydrogen bonding seen in other G_{syn}:G_{anti} pairs in RNA.[25,26]

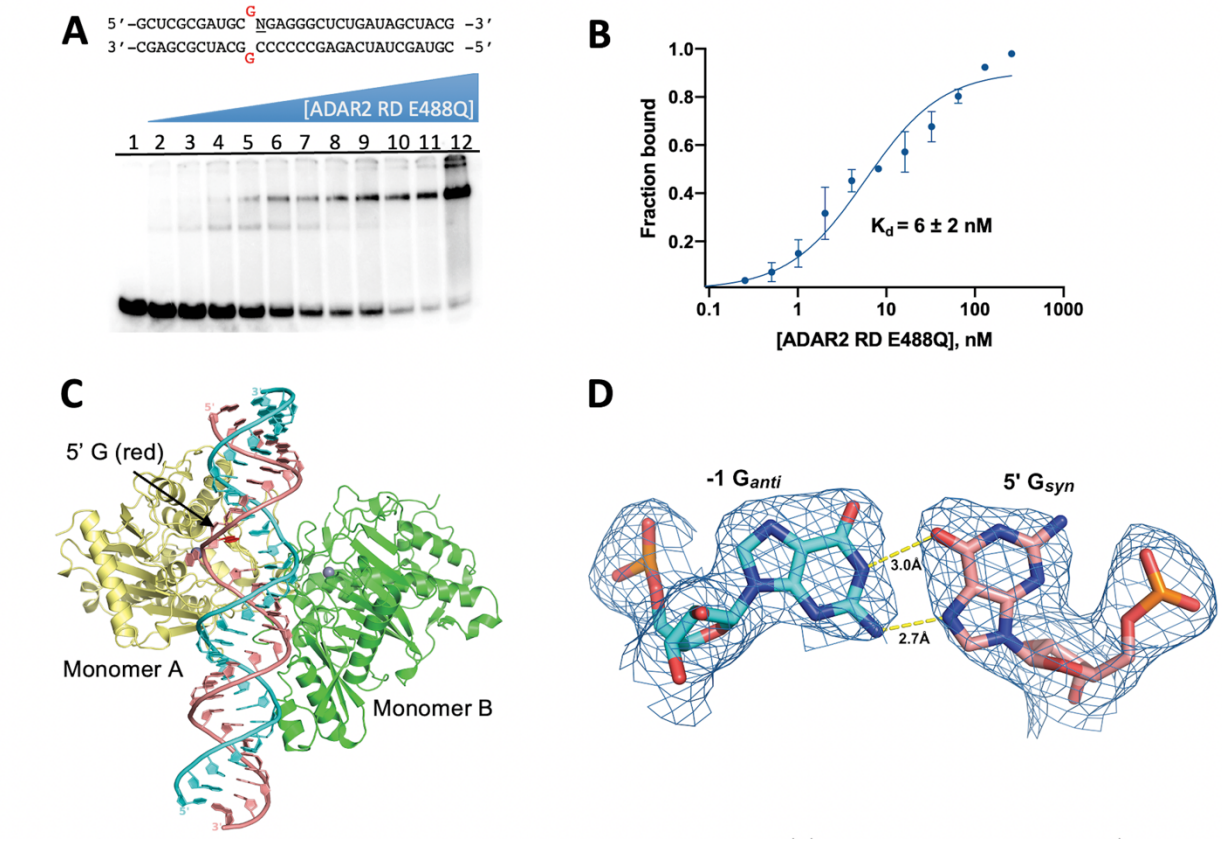


Figure 4.2. Characterization of complex formed between ADAR2-R2D E488Q and a 32 bp 8-azanebularine (N) containing duplex with G:G pair (32 bp GG RNA) adjacent to N. (**A**) (Top) Sequence of 32 bp duplex used for crystallization. (Bottom) EMSA gel of hADAR2-R2D E488Q with this duplex. Protein concentrations are as follows: lane 1: no protein, lanes 2–12: 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256 nM. (**B**) Quantification of protein binding by EMSA. (**C**) Ribbon diagram depicting the structure of hADAR2-RD E488Q dimer bound to 32 bp G:G pair RNA at 2.7 Å resolution. The edited strand (with azaN flipped into the active site of monomer A is colored salmon, while the unedited guide strand is colored cyan. The 5′G in the *syn* conformation is show with a red base. (**D**) Fit of a G_{syn}:G_{anti} base pair in the $2F_o – F_c$ electron density map contoured at 1σ. 5′ G indicates guanosine linked to azaN on its 5′ side. –1 G refers to the guanosine in the opposite strand on the 3′ side of the orphan base paired with the 5′ G.

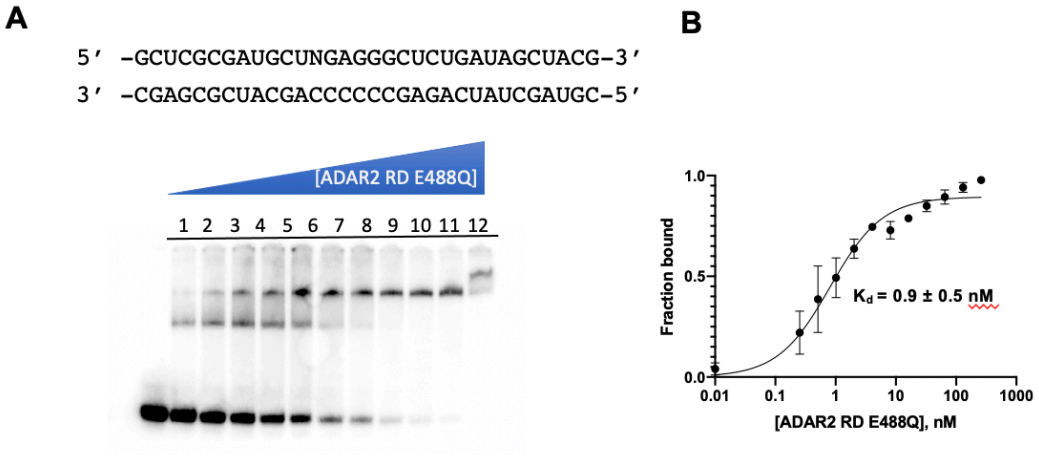


Figure 4.3. (A) Representative electrophoretic mobility shift assay (EMSA) gel displaying tightly bound complex of ADAR2 R2D E488Q to 8-azanebularine (N)-containing hGLI1 32 bp (5'U) duplex. Lane 1: no protein added; lanes 2–12: 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256 nM. RNA concentration was kept constant at ≥ 1.1 nM **(B)** Fitted plot of fraction RNA bound vs. ADAR2 RD concentration.

Previous structural analyses of ADAR2 bound to RNA revealed the molecular basis for ADAR2's preference for a 5′-U nearest neighbor over a 5′-G.[13] This preference arises from potential clashes between the 2-amino group and the base-flipping loop residue Gly489, as depicted in **Figure 4.4 A-C**. Notably, this minor groove clash can be resolved when a 5′-G nearest neighbor adopts a syn conformation, positioning the 2-amino group in the major groove (**Figure 4.4 A,C**). The hypothesis is evidenced by G:G pair showing enhanced editing efficiency and structural verification of the 5′-G syn conformation. However, the G_{syn}:G_{anti} pair still exposes a 2-amino group to the minor groove from the guide strand G_{anti}. This results in a slight shift in the base-flipping loop compared to the previous ADAR2-R2D – GLI1 structure (**Figure 4.4 D**). Specifically, the α-carbon of Gly489 shifts approximately 1.0 Å toward the edited strand.

Importantly, this shift can be accommodated due to the increased ribose-ribose spacing inherent in a G_{syn}:G_{anti} pair, a characteristic observed in other structural studies. Comparing our findings to the previous ADAR2-R2D – GLI1 structure, which contains an A:U pair, the C3′ of the G_{syn} nucleotide also shifts by 0.9 Å relative to C3′ of the native GLI1 RNA 5′ U (**Figure 4.4D**).

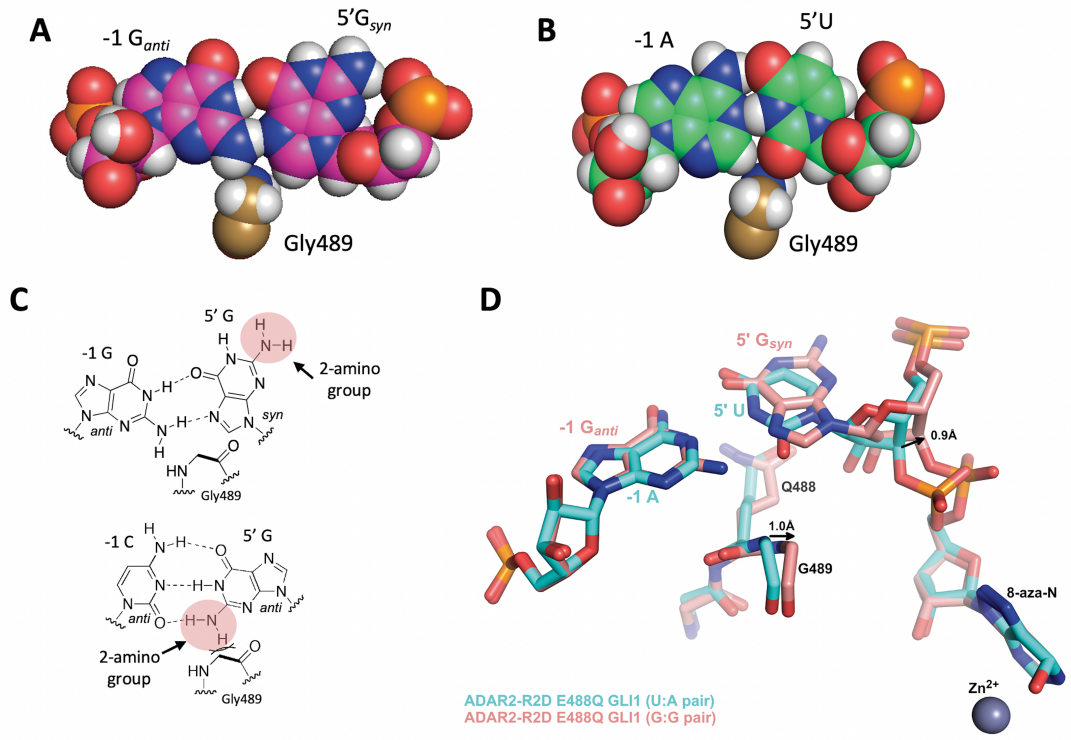


Figure 4.4. The G_{syn}:G_{anti} pair accommodates G489 in the minor groove. (**A**) Space filling representation of the G_{syn}:G_{anti} pair and its location relative to G489 in the complex. (**B**) Space filling representation of U:A pair and its location relative to G489 in a complex with ADAR2 for the ideal 5′ nearest neighbor base pair.[13] (**C**) (Top) Chemical structure of G_{syn}:G_{anti} pair highlighting the location of the 2-amino group of the 5′ G relative to G489. (Bottom) Chemical structure of G_{anti}:C_{anti} pair highlighting the location of the 2-amino group of the 5′ G relative to G489. (**D**) Overlay of ADAR2 R2D E488Q structures with RNA bearing either 5′ U paired with A (cyan) or 5′ G paired with G (salmon). Covalent hydration of 8-azanebularine, as seen by the out of plane oxygen present at the C6 position, enables trapping of these structures which mimic the adenosine deamination intermediate as confirmed in previous structures.[13]

This shift additionally causes the phosphate between 5′-G_{syn} and azaN to rotate by approximately 1.8 Å. Furthermore, sugar puckers in both nucleotides of the G_{syn}:G_{anti} pair as well as the corresponding A:U pair in native *GLI1* all reside in the 2′-endo conformation. As evidenced in previous structures, the protein's flipping loop approaches the adenosine from the minor groove, resulting in a substantial kink in the unedited strand.[13] This structural alteration causes a deviation from the typical A-form to a B-form conformation in this region, consistent with previous observations.[13]

Another notable feature is a small difference in how the dsRNA binding domain of monomer B interacts with the RNA. In the original ADAR2-R2D–*GLI1* RNA structure, His259 hydrogen bonds to the 2′OH of C18 of the unedited guide strand, while in this structure, His259 swings over and interacts with the 2′OH of A15 in the edited strand (**Figure 4.5**). Additionally, Ser258 hydrogen bonds with the 2-amino group of G16 of the edited strand of the previous structure, while in the structure presented here, Ser258 hydrogen bonds with His258 side chain and forms a weak (3.7 Å) interaction with the 2′OH of C19 of the guide strand. A similar conformation is also seen for the G:3-deaza dA pair structure (**Figure 4.6**). These structural differences together with the weaker electron density of the dsRBDs in all three structures, suggest weak, non-specific interactions between the dsRBDs and the dsRNA.

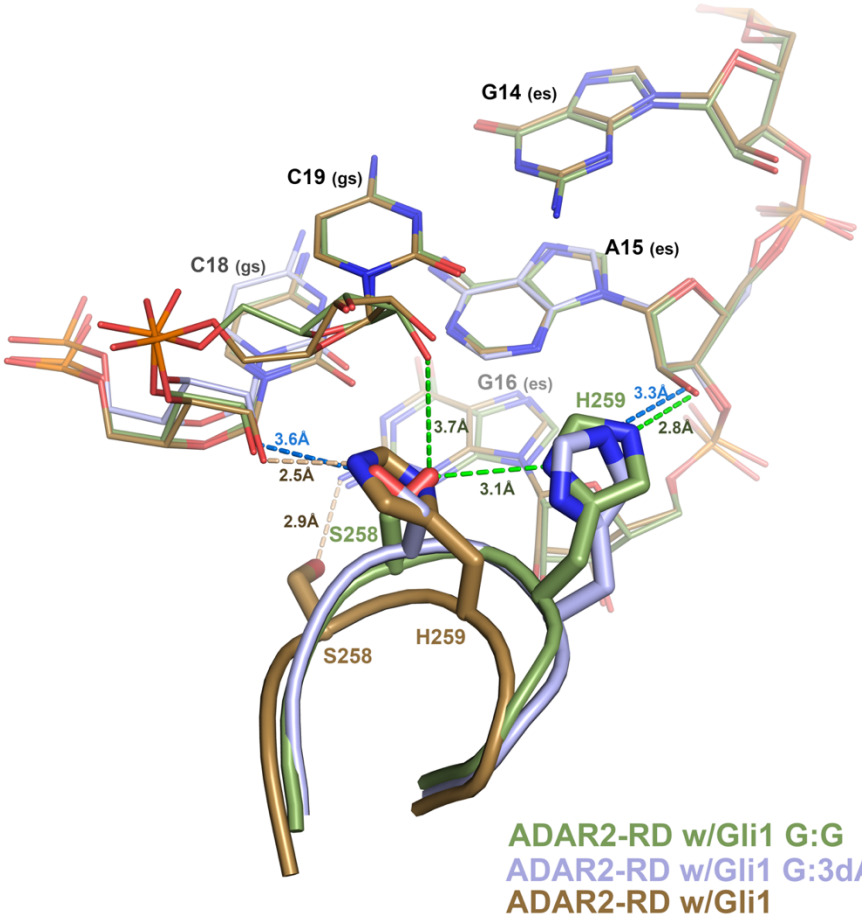


Figure 4.5. Comparison of ADAR2's dsRBD2 interactions with RNA. Shown is the superposition of ADAR2-R2D E488Q – Gli1 G:G pair (green) and ADAR2-R2D E488Q – Gli1 G:3-deaza-dA pair (light blue), with our previously reported structure of ADAR2-R2D E488Q – Gli1 (PDBID 6vff) (in sand color). Note the different RNA contacts involving residues Ser258 and His259 between the three structures. For clarity, only the A15(es) and C18(gs) bases for the G:3-deaza-dA pair are shown. es = edited strand, gs = guide strand.

Table 4.1. Data Processing and Refinement Statistics for ADAR2-R2D E488Q bound to dsRNA substrates		
dsRNA Substrate	GLI1 G:G w/azaN 32mer	GLI1 G:3deaza-dA w/azaN 32mer
PDBID	8e0f	8e4x
Synchrotron (Beamline)	APS (24 ID-E)	APS (24 ID-C)
Wavelength (Å)	0.97918	0.97918
Space Group	*C*2	*C*2
Unit Cell Parameters	*a* = 171.52Å, *b* = 63.39Å, *c* = 142.13Å, β = 117.69°	*a* = 169.91Å, *b* = 63.24Å, *c* = 142.65Å, β = 118.07°
Resolution Range (Å)	125 – 2.70 (2.82 – 2.70)	125 – 2.80 (2.95 – 2.80)
No. observed reflections	112,747 (13,563)	99,409 (15,009)
No. unique reflections	36,652 (4,467)	31,664 (4,741)
Completeness (%)	97.8 (98.6)	95.3 (98.1)
I/σ (I)	9.6 (1.7)	12.0 (1.2)
R_{merge}[a] (%)	7.5 (72.8)	6.4 (117.5)
$CC_{1/2}$	0.995 (0.616)	0.998 (0.524)
Refinement Statistics		
R_{factor}[b] (%)	19.33	19.50
R_{free}[b] (%)	23.05	23.94
RMS bond length (Å)	0.007	0.011
RMS bond angle (°)	1.307	1.422
Ramachandran Plot Statistics[c]		
Favored (%)	96.3	94.5
Allowed (%)	3.0	5.4
Outliers (%)	0.7	0.1
No. of atoms		
Protein	6,558	6,558
RNA	1,361	1,359
Inositol Hexakisphosphate (IHP)	72	72
Zn	2	2
Waters	41	42

[a]$R_{\text{merge}} = [\sum h \sum i |Ih - Ihi| / \sum h \sum i Ihi]$ where Ih is the mean of Ihi observations of reflection h. Numbers in parenthesis represent highest resolution shell.
[b]R_{factor} and $R_{\text{free}} = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}| \times 100$ for 95% of recorded data (R_{factor}) or 5% data (R_{free}).
[c]Ramachandran plot statistics from MolProbity.[38]

ADAR2 binds a duplex RNA substrate with a $G_{(syn)}$:$AH^{+}_{(anti)}$ pair adjacent to the editing site. The structure of G_{syn}-G_{anti} pair inspired investigations of nucleoside analogs capable of inducing a syn confirmation of 5'G to further enhance editing. The kinetic analysis of ADAR1 p110 initially suggested an adenosine paired with a 5'G may induce similar syn confirmation.

However, N-1 proton of an adenosine has a pKa of 3.7. Therefore, Dr. Erin Doherty tested series of analogs that probe this hypothesis by modulating the hydrogen bonding face of the -1 nucleobase and enchancing the pKa of the N-1 proton. Interestingly, she found adenosine analog 3-deaza dA (N-1 pKa 6.7) paired with a 5'G significantly boosted editing with both ADAR2 and ADAR1 p110. The efficient reactions with RNA substrates bearing 3-deaza dA paired with a 5′ G for both ADAR2 and ADAR1 p110 stimulated us to characterize this ADAR-RNA interaction further. Therefore, in collaboration with a Beal lab member, Herra Grajo Mendoza, I prepared a 32 bp duplex bearing azaN adjacent to a 5′ G paired with 3-deaza dA for X-ray crystallography with ADAR2-R2D E488Q (**Figure 6**). Crystals of this complex diffracted beyond 2.8 Å (**Table 3**). The overall structure of the asymmetric homodimeric protein bound to RNA is very similar to the G:G structure described above (rmsd of 0.201 Å for 698 equivalent α-carbons) with well-resolved electron density for the purine:purine pair adjacent to the azaN (**Figure 4.8 A,B**). The G on the 5′ side of the azaN is in a *syn* conformation with its Hoogsteen face directed toward the Watson–Crick face of the 3-deaza dA on the opposing strand. The position of the substrate strand is shifted slightly from that of the structure described above such that, in this structure, the base pair hydrogen bonding involves 3-deaza dA N1 to G N7 (2.8 Å) and 3-deaza dA 6-amino to G O6 (3.0 Å) (Figure 4.8C). This interaction causes the base of 3-deaza dA to shift slightly towards the minor groove, while also pushing the 5′G_{syn} out towards the major groove (**Figure 4.8 B**). This orientation suggests the 3-deaza dA N1 is protonated to donate a hydrogen bond to N7 forming a G_{syn}:AH^{+}_{anti} pair [28] (**Figure 4.8C**). Interestingly, the shift in substrate RNA strand position observed in this G_{syn}:AH^{+}_{anti} structure allows it to adopt a phosphodiester backbone conformation that is nearly identical to that seen in complexes with RNA bearing ADARs' preferred 5′ nearest

neighbor (5′ U paired with A) (**Figure 4.6 E**). Notably, the phosphate group between the 5′G and azaN swings back towards the guanidino group of Arg455 as seen in all other ADAR2-RNA structures. In the G_{syn}:G_{anti} pair structure, only the *pro*-S_P non-bridging oxygen interacts with Arg455, while in previous ADAR2–RNA complex structures, both non-bridging oxygens interact with the guanidino group of Arg455 (**Figure 4.6 C, E**). The standard conformation of this phosphate in the G:3-deaza dA pair might contribute to the observed higher deamination rate relative to the G:G pair.

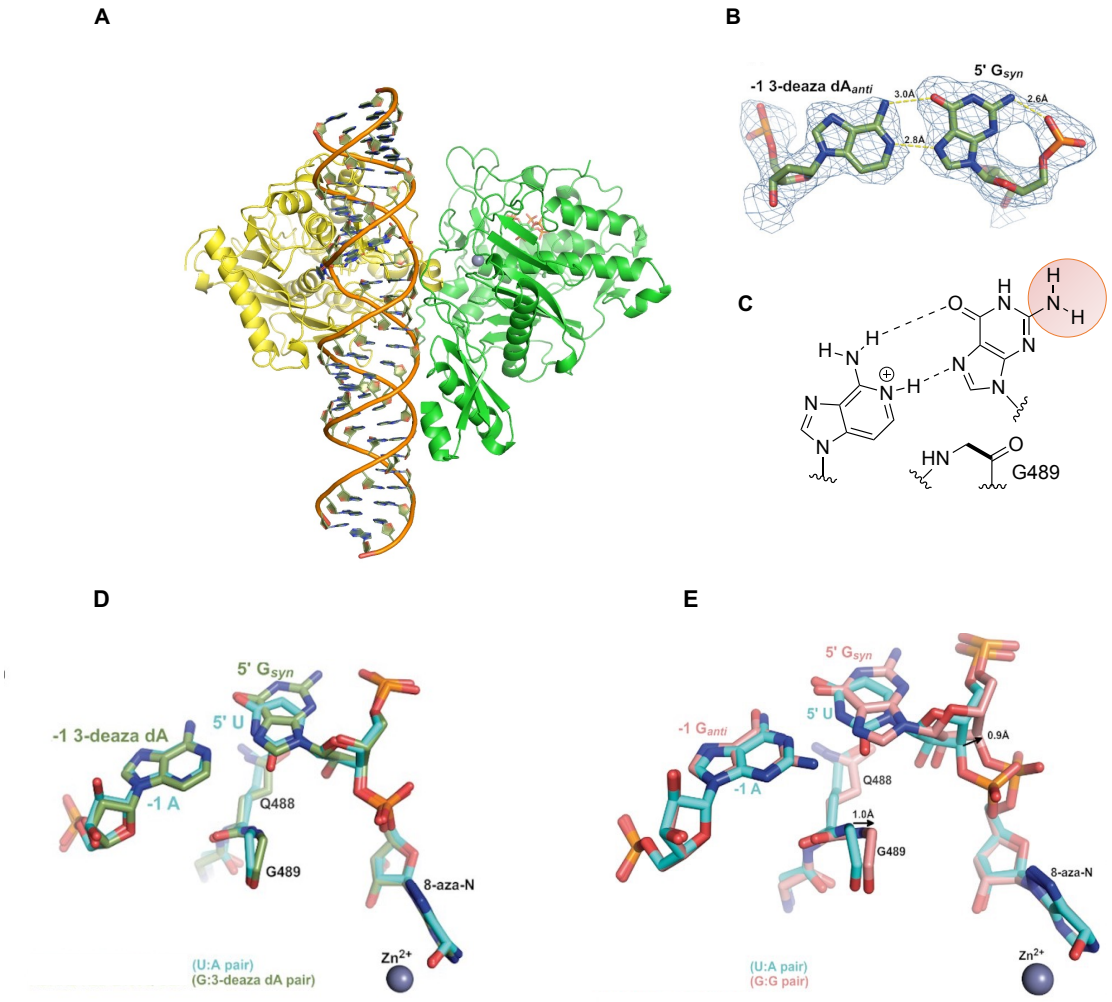


Figure 4.6. X-ray crystal structure of a complex formed between ADAR2-R2D E488Q and a 32 bp 8-azanebularine (azaN) containing duplex with G:3-deaza dA pair (32 bp G3A RNA) adjacent to azaN. (**A**) Fit of a G_{syn}:3-deaza dA_{anti} base pair in the $2F_o - F_c$ electron density map contoured at 1σ. (**B**) Overlay of ADAR2 R2D E488Q structures with RNA bearing either 5′ G paired with G (salmon colored carbons) or 5′ G paired with 3-deaza dA (green colored carbons). Arg455 in both structures is identical and shown with white-colored carbons. (**C**) The G_{syn}:3-deaza dAH^+_{anti} pair. (**D**) Overlay of ADAR2 R2D E488Q structures with RNA bearing either 5′ U paired with A (cyan colred carbons) or 5′ G paired with 3-deaza dA (green colored carbons).

4.3. Discussion

ADARs exhibit a strong inclination to edit adenosines located adjacent to a 5′ U (or A) while disfavoring those with a 5′ G neighbor. This preference poses challenges for therapeutic RNA editing applications targeting adenosines at a 5′ G, like those associated with Rett Syndrome mutations such as R168X, R255X, and R270X in the MECP2 gene.[29] In these cases, therapeutic editing by ADARs is essential to restore protein translation and functionality, typically by converting the nonsense mutation to a tryptophan codon. In the previously published structures of ADAR2 bound to RNA bearing a 5′ nearest neighbor U, we identified a loop of the protein involved in stabilizing the flipped out conformation (i.e. Ser486-Gly489) that occupied the RNA minor groove spanning three base pairs that included the nearest neighbor nucleotides and the edited base.[13,23,31,32] The minor groove edge of the base pair that includes the 5′ nearest neighbor base was juxtaposed to the protein backbone at Gly489. Modeling a G:C pair at this position (i.e. 5′ G) suggested the guanine 2-amino group in the minor groove would clash with the protein at Gly489. The Gsyn:Ganti pair resolves this apparent clash by inducing a change in the glycosidic bond angle at the 5′ G from anti to syn. This repositions the 2-amino group into the major groove, avoiding clashes with ADAR. This explanation for the G:G pair's effect on the ADAR reaction appears more plausible than the simple destabilization of the duplex due to a purine:purine

mismatch, especially considering that the structure of the purine paired with the 5′ G significantly influences the rate enhancement. For instance, a substrate with 3-deaza dA paired with a 5′ G reacts over 40-fold faster with ADAR2 than a substrate with 7-deaza dG paired with the 5′ G. This phenomenon extends beyond the G:G pair and can be generalized to other purines capable of inducing a syn conformation in the 5′ G. For example, A paired with 5′ G increases the ADAR rate compared to G:C and G:U. When protonated at N1, AH+ can donate two hydrogen bonds to the Hoogsteen face of G, forming the Gsyn:AH+anti pair. The observed increase in deamination rate with 3-deaza dA compared to dA paired with the 5′ G supports the formation of the Gsyn:AH+anti pair. Protonation of the adenine ring at N1H-N7 hydrogen bond requires a higher pKa for 3-deazaadenosine (6.8) than for adenosine (3.7)[33-35], making this protonation more likely in the 3-deaza system under the ADAR reaction conditions (**Figure 4.7**).

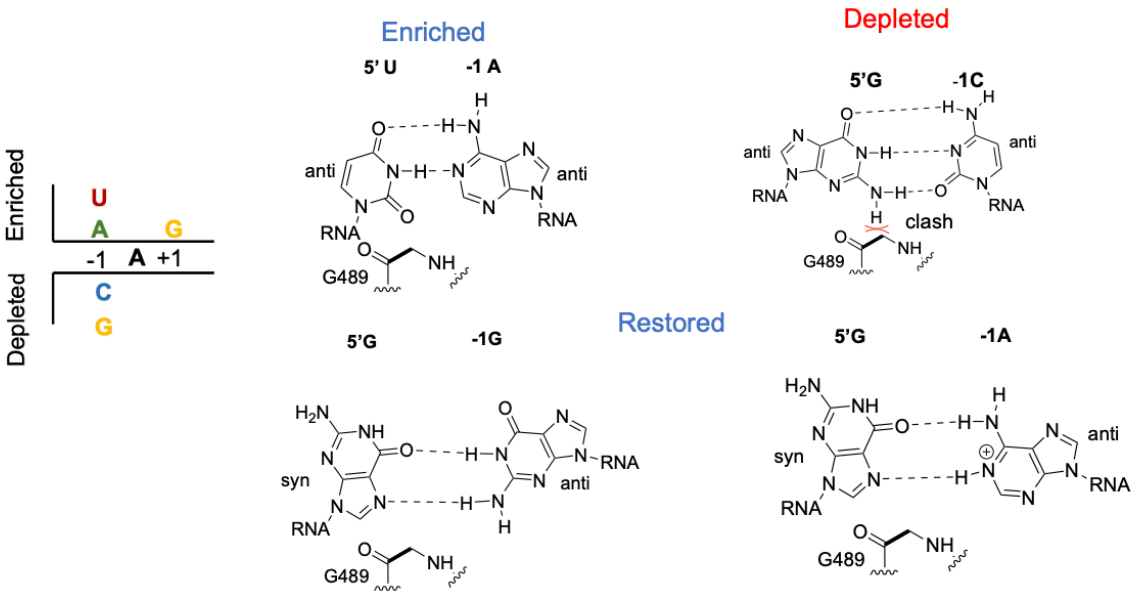


Figure 4.7. Summary of ADAR's sequence preferences. ADAR editing is enriched with a 5' U-A pair and depleted with a 5'G-C pair. Editing at 5'GA sequence context can be restored by nucleotides and analogs that induce a syn conformation.

In conclusion, the combination of deamination kinetics and structural studies described here identified an approach to facilitate ADAR editing at challenging 5′-GA sites. The use of nucleosides capable of hydrogen bonding to the Hoogsteen face of the 5′ G and inducing a *syn* conformation at this location in the RNA without also introducing additional sterically demanding groups into the minor groove enables efficient editing at these sites (Figure 4.8).

4.4. Materials and methods

General biochemical procedures. Molecular-biology-grade bovine serum albumin (BSA) and RNase inhibitor (RNAsin) were purchased from New England BioLabs. SDS-polyacrylamide gels were visualized with a Molecular Dynamics 9400 Typhoon phosphorimager. Data were analyzed with Molecular Dynamics ImageQuant 5.2 software. All Matrix Assisted Laser Desorption/Ionization (MALDI) analyses were performed at the University of California, Davis Mass Spectrometry Facilities using a Bruker UltraFlextreme MALDI TOF/TOF mass spectrometer. Oligonucleotide masses were determined with Mongo Oligo Calculator v2.06. Oligonucleotides for sequencing and PCR were purchased from Integrated DNA Technologies or Dharmacon. All other oligonucleotides were synthesized as described below.

Synthesis of Oligonucleotides. Chemical synthesis for all oligonucleotides was performed using an ABI 394 synthesizer. All protected phosphoramidites were purchased from Glen Research except the 8-azanebularine (azaN) phosphoramidite which was purchased from Berry & Associates or synthesized as previously described.[15] Nucleosides were incorporated during the

appropriate cycle on a 0.2 or 1.0 μmol scale**.** Appendix Table 1-5 show sequences of all oligonucleotides used in this study. Upon completion of the synthesis, columns were evaporated under reduced pressure for 12 h. All oligonucleotides were cleaved from the solid support by treatment with 1.5 ml 1:3 ethanol/30% NH4OH at 55◦C for 12 h. The supernatant was transferred to a new screw-cap tube and evaporated under reduced pressure. For all oligonucleotides except the azaN-modified strand, desilylation was performed by treating the pellets with 250 μl of 1M TBAF–THF at room temperature overnight. For the azaN strand, desilylation was carried out in TEA•3HF as previously described.[15] To each reaction was added 75 mM sodium acetate in butanol. The oligonucleotides were then precipitated from a solution of 65% butanol at –70◦C for 2 h. The solution was centrifuged at 17 000 × g for 20 min, supernatant was removed, and the pellet was washed twice with cold 95% ethanol. The RNA pellets were then desalted using a Sephadex G-25 column and purified as described below. Purification of oligonucleotides Single-stranded RNA oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis and visualized by UV shadowing. Bands were excised from the gel, crushed and soaked overnight at 4°C in 0.5 M NaOAc, 0.1% sodium dodecyl sulfate (SDS), and 0.1 mM EDTA. Polyacrylamide fragments were removed with a 0.2 m filter, and the RNAs were precipitated from a solution of 75% EtOH at –70°C for 12 h. The solution was centrifuged 17 000 × g for 20 min and supernatant was removed. The RNA solutions were lyophilized to dryness, resuspended in nuclease-free water and quantified by absorbance at 260 nm. Oligonucleotide mass was confirmed by MALDI-TOF. Mass spectrometry data can be found in Appendix Table 6.

Preparation of duplex RNA substrates for crystallography. For crystallography, the unmodified RNA guide strand was purchased from Horizon Dharmacon and purified as described above. As in previous structures, the edited strand contained the adenosine analog azaN at the editing site. Duplex RNA was hybridized in water in a 1:1 ratio by heating to 95°C for 5 min and slow cooling to 30°C.

In vitro transcription of editing target RNAs. Target RNAs for deamination kinetic analyses were transcribed from DNA templates with the MEGAScript T7 Kit (ThermoFisher). DNA digestion was performed using RQ1 RNase-free DNase (Promega). DNase-treated RNA product was purified as described above.

Preparation of duplex substrates for ADAR deamination kinetics. Purified guide and transcribed RNA were added in a 10:1 ratio to hybridization buffer (180 nM transcribed RNA target, 1.8 μM guide, 1X TE Buffer, 100 mM NaCl), heated to 95°C for 5 min, and slowly cooled to room temperature.

Expression and purification of human ADAR2 constructs for deamination kinetics. Full length human ADAR2 (hADAR2) was overexpressed in Saccharomyces cerevisiae as previously described.[16] Purification of hADAR2 was carried out by lysing cells in buffer containing 20 mM Tris–HCl, pH 8.0, 5% glycerol, 1 mM -mercaptoethanol (BME), 750 mM NaCl, 35 mM imidazole and 0.01% Nonidet P-40 (NP-40) using a French press. Cell lysate was clarified by centrifugation (39 000 × g for 1 h). Lysate was passed over a 3 ml Ni-NTA column, which was then washed in three steps with 20 ml lysis buffer, wash I buffer (20 mM Tris–HCl, pH 8.0, 5% glycerol, 1 mM

BME, 750 mM NaCl, 35 mM imidazole, 0.01% NP-40), wash II buffer (20 mM Tris–HCl, pH 8.0, 5% glycerol, 1 mM BME, 35 mM imidazole, 500 mM NaCl), and eluted with 20 mM Tris–HCl, pH 8.0, 5% glycerol, 1 mM BME, 400 mM imidazole, 100 mM NaCl. Fractions containing the target protein were pooled and concentrated to 30–80 μM for use in biochemical assays. Protein concentrations were determined using BSA standards visualized by SYPRO orange staining of SDS-polyacrylamide gels. Purified hADAR2 WT was stored in 20 mM Tris– HCl pH 8.0, 100 mM NaCl, 20% glycerol and 1 mM BME at –70◦C.

Expression and purification of ADAR1 p110 for deamination kinetics. MBP-tagged human ADAR1 p110 construct was cloned into a pSc vector using standard PCR techniques. The generated construct (yeast codon optimized) consisted of an N-terminal MBP-tag, a tobacco etch virus (TEV) protease cleavage site followed by the human ADAR1 p110 gene. S. cerevisiae BCY123 cells were transformed with this plasmid and the fusion protein was overexpressed as described previously.[17] Purification was carried out by lysing cells in lysis/binding buffer containing 50 mM Tris–HCl, pH 8.0, 5% glycerol, 5 mM BME, 1000 mM KCl, 0.05% NP40 and 50 μM ZnCl2 using a microfluidizer. Cell lysate was clarified by centrifugation (39 000 × g for 50 min). Lysate was passed over a 2 ml NEB amylose column (preequilibrated with binding buffer), which was then washed in two steps with 50 ml binding buffer followed by 100 ml wash buffer (50 mM Tris–HCl, pH 8.0, 5% glycerol, 5 mM BME, 500 mM KCl, 0.01% NP-40 and 50 μM $ZnCl_2$) and eluted with buffer containing 50 mM Tris–HCl, pH 8.0, 10% glycerol, 5 mM BME, 500 mM KCl, 0.01% NP-40, 50 μM $ZnCl_2$ and 20 mM maltose. Fractions containing the target protein were pooled and dialyzed against a storage buffer containing 50 mM Tris–HCl, pH 8.0,

400 mM KCl, 0.5 mM EDTA, 0.01% NP-40, 10% glycerol and 1 mM tris(2-carboxyethyl)phosphine. Dialyzed protein was concentrated to 2–50 M and stored as aliquots at –70◦C until further use in biochemical assays. Protein concentrations were determined using BSA standards visualized by SYPRO orange staining of SDS-polyacrylamide gels.

Deamination assays with ADAR2 and ADAR1 p110. Deamination assays were performed under single-turnover conditions in 15 mM Tris–HCl pH 7.5 3% glycerol, 60 mM KCl, 1.5 mM EDTA, 0.003% NP-40, 3 mM MgCl2, 160 U/ml RNAsin, 1.0 g/ml yeast tRNA, 10 nM RNA, and 75 nM human ADAR2. Each reaction solution was incubated at 30°C for 30 min before the addition of enzyme. Reactions were then incubated at 30°C for varying times prior to quenching with 95°C water and heating at 95°C for 5 min. Reaction products were used to generate cDNA using RT-PCR (Promega Access RT-PCR System). DNA was purified using a DNA Clean & Concentrator kit (Zymo) and subjected to Sanger Sequencing via GeneWiz (Azenta). The sequencing peak heights were quantified in SnapGene (Domatics). Data were fit to the equation $[P]t = [P]\ f\ [1 - e(-kobs \cdot t)\ ]$ for ADAR2 where [P]t is percent edited at time t, [P]f is the final endpoint of editing, and kobs is the observed rate constant. Because of the slower reactions for ADAR1 p110 and lower reaction end point, data were fit to the equation $[P]t = 0.4 \cdot [1 - e(-kobs \cdot t)\ ]$. Each experiment was carried out in triplicate where the k_{obs} reported is the average of each replicate ± standard deviation (SD). Statistical significance between groups was determined by one-way Analysis of Variance (ANOVA) with Tukey's multiple comparisons test using Prism software (GraphPad). For the ADAR1 p110 enzyme, deamination reactions were performed as above with the following modifications: The final reaction solution for ADAR1 p110 contained 15 mM Tris–HCl, pH 7.0

4% glycerol, 26 mM KCl, 40 mM potassium glutamate, 1.5 mM EDTA, 0.003% NP-40, 160 U/ml RNAsin, 1.0 g/ml yeast tRNA, 10 nM RNA and 250 nM ADAR1 p110.

Expression and purification of hADAR2 double stranded RNA binding domain and deaminase domain (hADAR2- R2D) for crystallography. Protein expression and purification were carried out by modifying a previously reported protocol (18). S. cerevisiae BCY123 cells were transformed with a pSc-ADAR construct encoding hADAR2-R2D E488Q (corresponding to residues 214–701). Cells were streaked on yeast minimal media minus uracil (CM-ura) plates. A single colony was used to inoculate a 15 ml CM-ura starter culture. After cultures were shaken at 300 rpm and 30°C overnight, 10 ml of starter culture was used to inoculate each liter of yeast growth medium. After cells reached an OD600 of 1.5 (~20– 24 h) cells were induced with 110 ml of sterile 30% galactose per liter and protein was expressed for 6 h. Cells were collected by centrifugation at 5000 × g for 10 min and stored at –80°C. Cells were lysed in 750 mM NaCl in buffer A (20 mM Tris–HCl, pH 8.0, 5% glycerol, 35 mM imidazole, 1 mM BME, and 0.01% Triton X-100) with a microfluidizer. Cell lysate was clarified by centrifugation (39 000 × g for 25 min). Lysate was passed over a 5 ml Ni-NTA column equilibrated with buffer A with 750 mM NaCl, which was then washed in three steps with 50 ml of lysis buffer, wash I buffer (buffer A + 300 mM NaCl), and wash II buffer (buffer A + 100 mM NaCl). Protein was eluted with a 35– 300 mM imidazole gradient in wash II buffer over 80 min at a flow rate of 1 ml/min. Fractions containing target protein were pooled and further purified on a 2 ml GE Healthcare Lifesciences Hi-Trap Heparin HP column in wash II buffer without BME. The His10 fusion protein was washed with 50 ml of wash II buffer without BME and eluted with a 100–1000 mM NaCl gradient over

60 min at a flow rate of 0.8 ml/min. Fractions containing target protein were pooled and cleaved with an optimized ratio of 1 mg of Histagged TEV protease per 1 mg of protein. The final salt concentration was targeted to be 300mM NaCl, high salt can slow down TEV reaction. Cleavage was carried out for 4 h at room temperature without agitation, followed by overnight cleavage at 4°C before the product was passed over another Ni-NTA column at a flow rate of 0.5 ml/min. The flow through and wash were collected and passed through another Ni-NTA column to remove remaining uncleaved protein. It is important to note, high glycerol content after TEV cleavage must be dialyzed away or diluted to final concentration of 5% glycerol. The glycerol content can affect proper separation of TEV and ADAR2. In addition, TEV and ADAR proteins can associate with each other, and a second column is sometimes required. The flow through and wash were collected, dialyzed against 20 mM Tris, pH 8.0, 200 mM NaCl, 5% glycerol and 1 mM BME, followed by concentration to just under 1ml for gel filtration on a GE Healthcare HiLoad 16/600 Superdex 200 PG column. Fractions containing purified protein were pooled and concentrated to 7–9 mg/ml for crystallization trials.

Crystallization of the hADAR2-R2D E488Q-RNA complex. Crystals of the hADAR2-R2D E488Q-GLI1 (G:G pair) RNA complex were grown at room temperature by the sitting-drop vapor-diffusion method. A solution of 1.0 μl volume containing 5.6 mg/ml protein (95 μM) and 47.5 μM GLI1-GG RNA was mixed with 1.0 μl of 50 mM MOPS pH 7.0, 200 mM NaCl, 17% PEG 4000. Crystals took about 10 days to grow. A cluster of crystals was broken apart and a single cuboid-shaped crystal approximately 100 μm in size was soaked briefly in a solution of mother liquor plus 30% ethylene glycol before flash cooling in liquid nitrogen. Crystals of the hADAR2-

R2D E488Q-GLI1 (G:3-deazadA pair) RNA complex were grown at room temperature by the sitting-drop vapor-diffusion method. A solution of 1.0 μl volume containing 100 M protein and 50 M GLI1- G3dA RNA was mixed with 1.0 μl of 50 mM MOPS pH 7.0, 100 mM NaCl, and 13% PEG 4000. Flat rhomboid shape crystals took about 8 days to grow to 100 μm. A single crystal was soaked briefly in a solution of mother liquor plus 30% glycerol before flash cooling in liquid nitrogen.

Processing and refinement of crystallographic data. X ray diffraction data for both structures were collected at the Advanced Photon Source (APS). Diffraction data for the ADAR2-R2D E488Q GLI1 (G:G pair) complex were collected on beamline 24-ID-E to 2.7Å resolution while the ADAR2-R2D E488Q GLI1 (G:3-deaza dA pair) data were collected on the 24-ID-C beamline to 2.8Å resolution. Both data sets were processed with XDS9[19] and scaled with AIMLESS.[20] The structures were determined by molecular replacement using PHENIX.[21] The previous ADAR2-R2D E488Q GLI1-bound crystal structure (PDBID: 6vff) was used as the phasing model. The structures were refined with PHENIX including non-crystallographic symmetry (NCS) restraints at the start but relaxed at the final stages resulting in a lower R-free. Additionally, RNA base-stacking and base-pair restraints, when appropriate, were also imposed in refinement. Table 3 shows statistics in data processing and model refinement. As observed in the previous ADAR2-R2D-Gli1 structure[22], the asymmetric unit for both structures include two protein monomers assembled as an asymmetric homodimer complexed with RNA 32 bp duplex. Both complexes displayed similar overall structures. The double stranded RNA binding domains (residues 215–315) of monomer A were disordered and therefore not included in the model. The

first 20 residues (215–234) and part of the 5 RNA binding loop, residues 464–475, of monomer B were disordered and not included in the model. The dsRNA binding domain (dsRBD) of monomer B interacts with the 5' end of the dsRNA relative to the 8-azanebularine (azaN).

Appendix Chapter 4.

Appendix Table 4.1. Sequences for *in vitro* kinetics of the *IDUA* target (nucleotides in brackets are 2'-deoxy. All others are ribonucleotides). All PCR primers are 2'-deoxynucleotides. (3dA) indicates 3-deaza-2'-deoxyadenosine.

IDUA guide strand -1 A	5'-UUUGAGACCUCUGUC[C]AGAGUUGUUCUCC-3'
IDUA guide strand -1 G	5'-UUUGAGACCUCUGUC[C]GGAGUUGUUCUCC-3'
IDUA guide strand -1 C	5'-UUUGAGACCUCUGUC[C]CGAGUUGUUCUCC-3'
IDUA guide strand -1 U	5'-UUUGAGACCUCUGUC[C]UGAGUUGUUCUCC-3'
IDUA guide strand -1 2'-deoxy G	5'-UUUGAGACCUCUGUC[CG]GAGUUGUUCUCC-3'
IDUA guide strand -1 2'-deoxy A	5'-UUUGAGACCUCUGUC[CA]GAGUUGUUCUCC-3'
IDUA guide strand -1 2'-deoxy-3-deazaadenosine	5'-UUUGAGACCUCUGUC[C(3dA)]GAGUUGUUCUCC-3'
Idua RT-PCR forward and sequencing primer	5'-GCTCCTCCCATCCTGTGGGCTGAACAGT-3'
Idua RT-PCR reverse primer	5'-CGGGGTGTGCGTGGGTGTCATCACT-3'

DNA template sequence for *in vitro* kinetics of the *IDUA* 5'-UA target. Grey indicates the T7 promoter, underline is the region complementary to the guide strands, and the red A is the target adenosine.

TAATACGACTCACTATAGGGCTCCTCCCATCCTGTGGGCTGAACAGTATAACAGACT
CCCAGTATACAAATGGTGGGAGCTAGATATTAGGGTAGGAAGCCAGATGCTAGGTA
TGAGAGAGCCAACAGCCTCAGCCCTCTGCTTGGCTTATAGATGGAGAACAACTCTAG
GCAGAGGTCTCAAAGGCTGGGGCTGTGTTGGACAGCAATCATACAGTGGGTGTCCT
GGCCAGCACCCATCACCCTGAAGGCTCCGCAGCGGCCTGGAGTACCACAGTCCTCA
TCTACACTAGTGATGACACCCACGCACACCCCGGATCC

DNA template sequence for *in vitro* kinetics of the *IDUA* 5'-GA target. Grey indicates the T7 promoter, underline is the region complementary to the guide strands, and the red A is the target adenosine.

TAATACGACTCACTATAGGGCTCCTCCCATCCTGTGGGCTGAACAGTATAACAGACT
CCCAGTATACAAATGGTGGGAGCTAGATATTAGGGTAGGAAGCCAGATGCTAGGTA
TGAGAGAGCCAACAGCCTCAGCCCTCTGCTTGGCTTATAGATGGAGAACAACTCGA
GGCAGAGGTCTCAAAGGCTGGGGCTGTGTTGGACAGCAATCATACAGTGGGTGTCC
TGGCCAGCACCCATCACCCTGAAGGCTCCGCAGCGGCCTGGAGTACCACAGTCCTC
ATCTACACTAGTGATGACACCCACGCACACCCCGGATCC

Appendix Table 4.2. Sequences for *in vitro* kinetics of the *MECP2* R255 target. All guides are ribonucleotides. All PCR primers are 2'-deoxynucleotides.

MECP2 R255 guide strand -1 C	5'-GUCGGCCUCAGCUUUCCGCUUCCUGCCGG-3'
MECP2 R255 guide strand -1 G	5'-GUCGGCCUCAGCUUUCGGCUUCCUGCCGG-3'
MECP2 R255 RT-PCR forward and sequencing primer	5'-GTGCAGGTGAAAAGGGTC-3'
MECP2 R255 RT-PCR reverse primer	5'-TACGGTCTCCTGCACAGATCG-3'

DNA template sequence for *in vitro* kinetics of the *MECP2* R255 target. Grey indicates the T7 promoter, underline is the region complementary to the guide strands, and the red A is the target adenosine.

TAATACGACTCACTATAGGGGTGCAGGTGAAAAGGGTCCTGGAGAAAAGTCCTGGG
AAGCTCCTTGTCAAGATGCCTTTTCAAACTTCGCCAGGGGGCAAGGCTGAGGGGGGT
GGGGCCACCACATCCACCCAGGTCATGGTGATCAAACGCCCCGGCAGGAAGCGAAA
AGCTGAGGCCGACCCTCAGGCCATTCCCAAGAAACGGGGCCGAAAGCCGGGGAGTG
TGGTGGCAGCCGCTGCCGCCGAGGCCAAAAAGAAAGCCGTGAAGGAGTCTTCTATC
CGATCTGTGCAGGAGACCGTA

Appendix Table 4.3. Sequences for *in vitro* kinetics of the *MECP2* R255X target. Nucleotides in brackets are 2'-deoxy. All others are ribonucleotides. All PCR primers are 2'-deoxynucleotides. (8BrG) is 8-bromo-2'-deoxyguanosine, (7dA) is 7-deazaadenosine, (3dA) is 3-deazaadenosine.

MECP2 R255X guide strand -1 C	5'-GUCGGCCUCAGCUUUCCACUUCCUGCCGG -3'
MECP2 R255X guide strand -1 G	5'-GUCGGCCUCAGCUUUCGACUUCCUGCCGG -3'
MECP2 R255X guide strand -1 A	5'-GUCGGCCUCAGCUUUCAACUUCCUGCCGG -3'
MECP2 R255X guide strand -1 2'-deoxy A	5'-GUCGGCCUCAGCUUUC[A]ACUUCCUGCCGG -3'
MECP2 R255X guide strand -1 2'-deoxy G	5'-GUCGGCCUCAGCUUUC[G]ACUUCCUGCCGG -3'

MECP2 R255X guide strand -1 2'-deoxy 8-bromoguanosine	5'-GUCGGCCUCAGCUUUC[(8BrG)]ACUUCCUGCCGG -3'
MECP2 R255X guide strand -1 2'-deoxy 7-deazaguanosine	5'-GUCGGCCUCAGCUUUC[(7dG)]ACUUCCUGCCGG -3'
MECP2 R255X guide strand -1 2'-deoxy 3-deazaadenosine	5'-GUCGGCCUCAGCUUUC[(3dA)]ACUUCCUGCCGG -3'
MECP2 R255X RT-PCR forward and sequencing primer	5'-GGGTGTGCAGGTGAAAAGG-3'
MECP2 R255X RT-PCR reverse primer	5'-TCTTGATGGGGAGTACGGTC-3'

DNA template sequence for *in vitro* kinetics of the *MECP2* R255X target. Grey indicates the T7 promoter, underline is the region complementary to the guide strands, and the red A is the target adenosine.

CACGATTAATACGACTCACTATAGGGTGTGCAGGTGAAAAGGGTCCTGGAGAAAAG
TCCTGGGAAGCTCCTTGTCAAGATGCCTTTTCAAACTTCGCCAGGGGGCAAGGCTGA
GGGGGGTGGGGCCACCACATCCACCCAGGTCATGGTGATCAAACGCCCCGGCAGGA
AGTGAAAAGCTGAGGCCGACCCTCAGGCCATTCCCAAGAAACGGGGCCGAAAGCCG
GGGAGTGTGGTGGCAGCCGCTGCCGCCGAGGCCAAAAAGAAAGCCGTGAAGGAGT
CTTCTATCCGATCTGTGCAGGAGACCGTACTCCCCATCAAGAA

Appendix Table 4.4. Sequences for crystallography. All bases are ribonucleotides. (N) is 8-azanebularine.

GLI1(GG, G3dA) 32mer top with 8-azanebularine	5'- GCUCGCGAUGCG(N)GAGGGCUCUGAUAGCUACG -3'
GLI1 (GG) 32mer bottom	5'- CGUAGCUAUCAGAGCCCCCCGGCAUCGCGAGC -3'
GLI1 (G3dA) 32mer bottom	5'- CGUAGCUAUCAGAGCCCCCC(3dA)GCAUCGCGAGC -3'

Appendix Table 4.5. Sequences for binding experiments. All bases are ribonucleotides. (N) is 8-azanebularine.

GLI1(GG) 32mer top with 8-azanebularine	5'- GCUCGCGAUGCG(N)GAGGGCUCUGAUAGCUACG -3'
GLI1 (GG) 32mer bottom	5'- CGUAGCUAUCAGAGCCCCCCGGCAUCGCGAGC -3'
GLI1(UA) 32mer top with 8-azanebularine	5'- GCUCGCGAUGCU(N)GAGGGCUCUGAUAGCUACG -3'
GLI1 (UA) 32mer bottom	5'- CGUAGCUAUCAGAGCCCCCCAGCAUCGCGAGC-3'

4.5. References

1. Bass, B.L., *RNA editing by adenosine deaminases that act on RNA.* Annu. Rev. Biochem., 2002. **71**: p. 817-846.
2. Wang, Y., Y. Zheng, and P.A. Beal, *Adenosine Deaminases that Act on RNA (ADARs).* Enzymes, 2017. **41**: p. 215-268.
3. Eisenberg, E., *Proteome Diversification by RNA Editing.* Methods Mol Biol, 2021. **2181**: p. 229-251.
4. Erdmann, E.A., et al., *To protect and modify double-stranded RNA-the critical roles of ADARs in development, immunity and oncogenesis.* Crit Rev Biochem Mol Biol, 2021. **56**(1): p. 54-87.
5. Heraud-Farlow, J.E. and C.R. Walkley, *What do editors do? Understanding the physiological functions of A-to-I RNA editing by adenosine deaminase acting on RNAs.* Open Biol, 2020. **10**(7): p. 200085.
6. Stephens, O.M., H.-Y. Yi-Brunozzi, and P.A. Beal, *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): p. 12243-12251.
7. Qu, L., et al., *Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs.* Nat. Biotechnol., 2019. **37**(9): p. 1059-1069.
8. Merkle, T., et al., *Precise RNA Editing by Recruiting Endogenous ADARs with Antisense Oligonucleotides.* Nat. Biotechnol. , 2019. **37**(2): p. 133-138.
9. Katrekar, D., et al., *In Vivo RNA Editing of Point Mutations via RNA-guided Adenosine Deaminases.* Nat. Methods, 2019. **16**(3): p. 239-242.
10. Monian, P., et al., *Endogenous ADAR-mediated RNA editing in non-human primates using stereopure chemically modified oligonucleotides.* Nat. Biotechnol. , 2022: p. doi: 10.1038/s41587-022-01225-1.
11. Khosravi, H.M. and M.F. Jantsch, *Site-directed RNA editing: recent advances and open challenges.* RNA Biol., 2021. **18**: p. 41-50.
12. Eggington, J.M., T. Greene, and B.L. Bass, *Predicting sites of ADAR editing in double stranded RNA.* Nat. Commun., 2011. **2**(319): p. DOI:10.1038/ncomms1324.
13. Matthews, M.M., et al., *Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity.* Nat. Struct. Mol. Biol., 2016. **23**(5): p. 426-433.
14. Schneider, M.F., et al., *Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans.* Nucleic Acids Res., 2014. **42**(10): p. e87.
15. Rypniewski, W., et al., *Noncanonical G(syn)-G(anti) base pairs stabilized by sulphate anions in two X-ray structures of the (GUGGUCUGAUGAGGCC) RNA duplex.* RNA, 2008. **14**(9): p. 1845-1851
16. Haudenschild, B.L., et al., *A transition state analogue for an RNA-editing reaction.* J. Am. Chem. Soc., 2004. **126**: p. 11213-11219.
17. Macbeth, M.R. and B.L. Bass, *Large-Scale Overexpression and Purification of ADARs from Saccharomyces cerevisiae for Biophysical and Biochemical Studies.* Methods Enzymol., 2007. **424**: p. 319-331.

18. Malik, T.N., et al., *Regulation of RNA Edtiting by Intracellular Acidification.* Nucleic Acids Res., 2021. **49**: p. 4020-4036.
19. Macbeth, M.R., A.T. Lingam, and B.L. Bass, *Evidence for auto-inhibition by the N-terminus of hADAR2 and activation by dsRNA binding.* RNA, 2004. **10**: p. 1563-1571.
20. Kabsch, W., *XDS.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(2): p. 125-132.
21. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. **69**(7): p. 1204-1214.
22. McCoy, A.J., et al., *Phaser crystallographic software.* J. Appl. Crystallogr., 2007. **40**(4): p. 658-674.
23. Thuy-Boun, A.S., et al., *Asymmetric Dimerization of Adenosine Deaminase acting on RNA Facilitates Substrate Recognition.* Nucleic Acids Res., 2020. **48**(14): p. 7958-7972.
24. Leontis, N.B. and E. Westhof, *Geometric nomenclature and classification of RNA base pairs.* RNA, 2001. **7**: p. 499-512.
25. Burkhard, M.E. and D.H. Turner, *NMR Structures of r(GCAGGCGUGC)2 and Determinants of Stability for a Single Guanosine-Guanosine Base Pairs.* Biochemistry, 2000. **39**: p. 11748-11762.
26. Jiang, F., et al., *Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex.* Nature, 1996. **382**: p. 183-186.
27. Krishnaraj, R., G. Ho, and J. Christodoulou, *RettBASE: Rett syndrome database update.* Hum. Mutat., 2017. **38**(8): p. 922-931.
28. Pan, B., S.N. Mitra, and M. Sundaralingam, *Crystal Structure of an RNA 16-mer Duplex R(GCAGAGUUAAAUCUGC)2 with Nonadjacent G(syn)•+(Anti) Mispairs.* Biochemistry, 1999. **38**: p. 2826-2831.
29. Li, J.B., et al., *Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing.* Science, 2009. **324**(5931): p. 1210-1213.
30. Eifler, T., S. Pokharel, and P.A. Beal, *RNA-Seq Analysis Identifies A Novel Set of Editing Substrates for Human ADAR2 Present in Saccharomyces cerevisiae.* Biochemistry, 2013. **52**(45): p. 7857-7869.
31. Monteleone, L.R., et al., *A Bump-Hole Approach for Directed RNA Editing.* Cell Chem. Biol., 2019. **26**: p. 269-277.
32. Doherty, E.E., et al., *Rational Design of RNA Editing Guide Strands: Cytidine Analogs at the Orphan Position.* J. Am. Chem. Soc., 2021. **143**: p. 6865-6876.
33. Krishnamurthy, R., *Role of pKa of Nucleobases in the Origins of Chemical Evolution.* Acc. Chem. Res. , 2012. **45**(12): p. 2035-2044.
34. Mlotkowski, A. J.; Schlegel, H. B.; Chow, C. S. Calculated PKa Values for a Series of Aza- and Deaza-Modified Nucleobases. *J Phys Chem A* **2023**, *127* (15), 3526–3534.
35. Ikehara, M., S. Uesugi, and K. Yoshida, *Studies of the Conformation of Purine Nucleosides and Their 5'-Phospates.* Biochemistry, 1972. **11**(5): p. 830-836.
36. Williams, C.J., et al., *MolProbity: More and better reference data for improved all-atom structure validation.* Protein Sci., 2018. **27**: p. 293-315.

Chapter 5

Enabling Selective ADAR Editing in Disfavored Sites

This chapter details ongoing efforts in the lab to enable selective editing of ADARs at disfavored target sites. In this chapter, I will highlight two approaches to get to this goal.

5.1. Introduction.

Epitranscriptomics, analogous to epigenetics, unveils post-transcriptional RNA modifications that regulate gene expression, stability, and localization[1,2]. Among these, inosine stands as a unique example where the modified base's pairing properties diverge from its original form.[1,2] This conversion of Adenosine (A) to Inosine (I), acting as a functional mimic of guanosine, is primarily catalyzed by Adenosine Deaminase Acting on RNA (ADARs) within their preferred double-stranded RNA substrates.[3] With ADAR enzymes specifically targeting duplex RNA and their ubiquitous expression in most tissues, they offer a potential avenue for correcting disease-causing G to A point mutations.[4] By utilizing tailored guide RNAs for specific targets, endogenous ADARs can be harnessed to rectify G to A mutations, highlighting the promise of these enzymes in precision RNA-based therapeutics. Worth noting is the prevalence of G to A point mutations, responsible for approximately 28% of missense and nonsense mutations.[5] ADARs, capable of effecting complete restoration or reversion to functional proteins, provide an alternative to Cas9-based genome editing.[56] While the latter has advanced significantly, concerns about off-target effects at the DNA level, and the need for efficient delivery of a large cargo persist. In contrast, endogenous ADARs offer a transient and reversible RNA-level intervention, further emphasizing their potential in addressing genetic anomalies. [56]

While the potential of utilizing endogenous ADAR editing as therapeutics is promising, several critical considerations must be addressed to bring this approach to fruition. One notable concern is the possibility of off-target editing occurring at the transcript level, leading to unintended consequences on splicing and translation.[5,6,7] Additionally, the challenge of effectively delivering guide RNAs to specific target tissues, and proper ADAR recruitment remains a significant hurdle.[5] Moreover, ADARs exhibit a preference for certain editing sites, adding another layer of complexity to the therapeutic strategy.[8,9,10] ADARs highly disfavor adenosines with a 5'G to target site and they display reduced efficiency when encountering a cytosine adjacent to the editing site. Previous structural studies of ADAR2 bound to transition state analog containing RNA suggest the exocyclic amine of guanosine suffers a steric clash with a critical glycine residue in the flipping loop of the protein when presented with a 5'G-C or a 5'C-G pairing (**Figure 5.1**).[11]

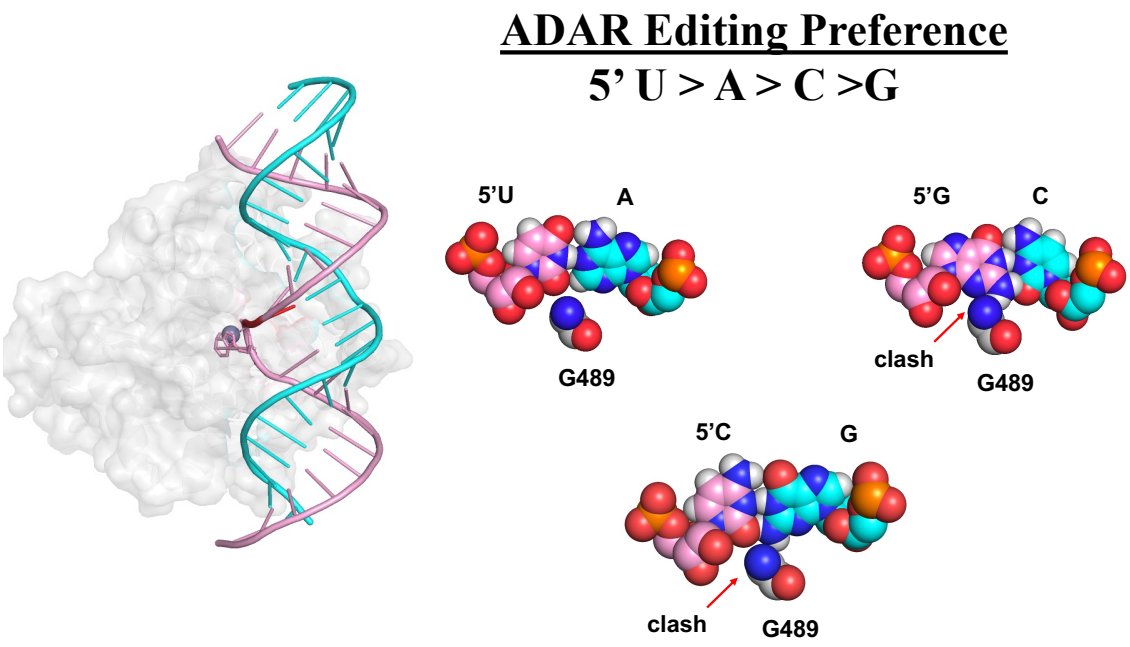


Figure 5.1. Structure of ADAR2 bound dsRNA showcasing ADAR flipping loop accommodating a 5' U-A pair next to edited adenosine. Modeled are 5'G-C and 5'C-G pairs presenting a clash with G489 in the minor groove of the protein.

In Chapter 4, I described our lab's recent efforts that demonstrated that certain purine analogs, capable of inducing a syn conformation in 5'G, facilitate efficient editing at 5'-GA-3' sites. Notably, we found a 5'G_{syn}-G_{anti} mismatch or 3-deaza-2'deoxyadenosine capable of pairing with a 5'G in its syn conformation induce editing.[8] Similarly, guanosine or adenosine analogs capable of pairing with a cytosine while removing the exocyclic amine should enable robust and selective editing at 5'-CA-3' sites.[12,13] Therefore, we sought to identify guide strands and offer mechanistic insights supporting enhanced reactivity (**Figure 5.2**). Furthermore, for a comprehensive understanding of the local secondary structures and motifs within guide RNAs enabling editing at 5'GA sites, we harnessed the Beal lab NGS screening method: En Masse Evaluation of Guide RNA (EMERGe).[14] The effort described in this chapter strives to offer a structural rationale for the motifs identified, shedding light on the mechanisms underpinning the observed editing enhancements.

5.2. Results.

2'-Deoxyinosine and 2'-deoxy-7-deazaadenosine enhances editing at 5'-CA-3'. To determine if the 5'C base pairing partner influences ADAR editing, we first designed guide RNAs consisting of either canonical bases (G,A,C,U) or the guanosine analog 2'-deoxyinosine (dI) at the -1 position of the guide RNA. Inosine lacks an exocyclic amine and should prevent a clash with G489. The guide RNAs were designed to target a G to A point mutation in the SerpinA1 gene, associated with alpha 1 antitrypsin deficiency (AATD).[4] The missense mutation E342K prompts

protein aggregation in the liver and impairs its transport to the lungs. As a result, both lung and liver diseases manifest. ADAR editing can lead to complete reversion of the mutation and therefore the disease.[4] Indeed, our analysis demonstrated that dI significantly and selectively enhanced editing at 5'-CA -3' sites in Serpin A1, providing 50-fold enhancement in editing rate. In addition, the adjacent adenosine, typically a preferred site for ADARs, exhibited lower editing efficiency using guide RNA with dI in comparison to the other canonical bases (**Figure 5.2**).

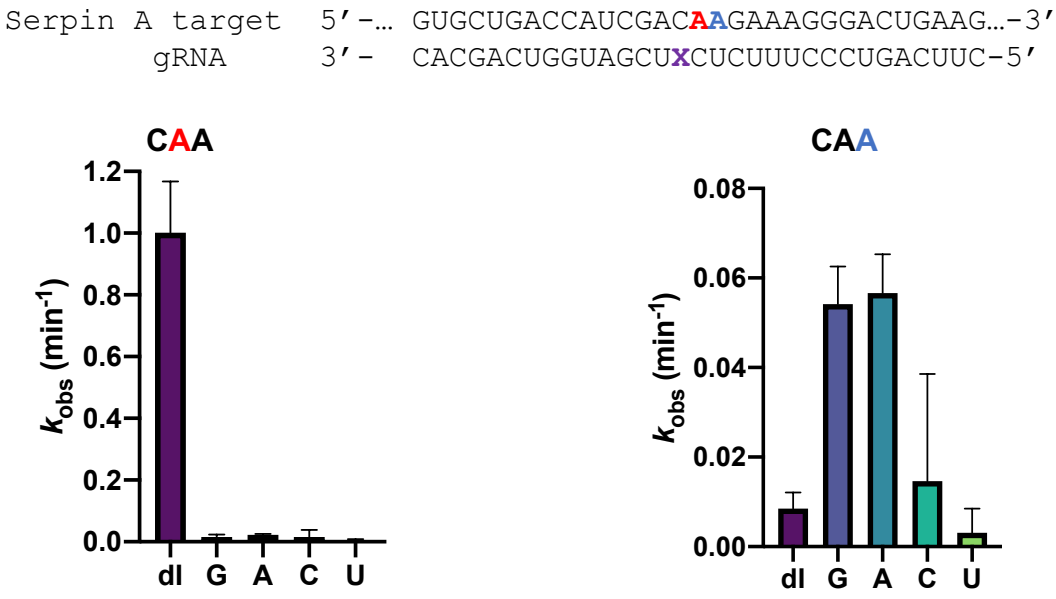


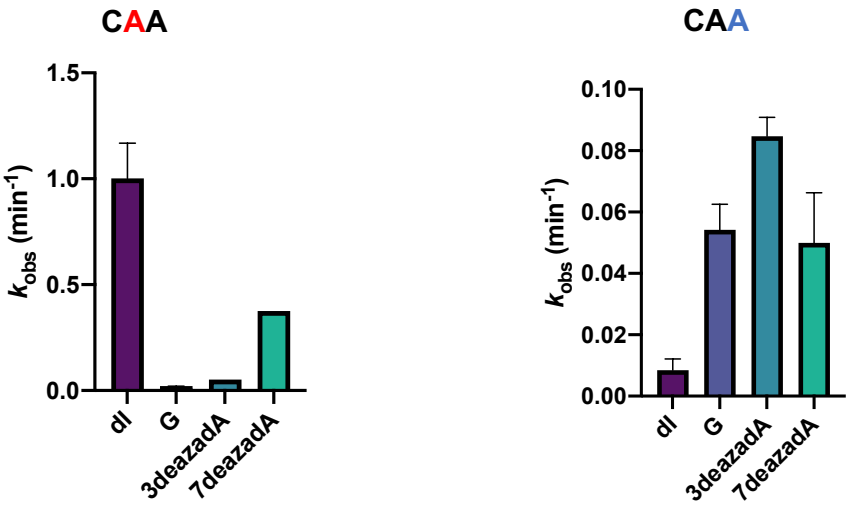
Figure 5.2. In vitro kinetics of hADAR2. Plotted are the deamination observed rate constants for the Serpin A target. The reaction was performed under single turn over conditions with 100 nM ADAR2 on 10 nM mutant SerpinA1 substrate. X in the guide RNA indicates the base pairing in the -1 position. [a] Data were fitted to the equation $[P]t = \alpha[1-\exp(-k_{obs} \cdot t)]$. [b] $k_{rel} = k_{obs}$.

We wondered if adenosine analogs capable of base pairing with cytosine could similarly enhance editing selectively at 5'-CA-3' sites. Therefore, we tested 2'-deoxy-3-deazaadenosine (3-deaza dA) and 2'-deoxy-7-deazaadenosine (7-deaza dA) at the -1 position of the guide RNA. Both

analogs would enhance the pKa of the N1 proton (3-deaza dA pKa 6.8, 7-deazadA pKa 5.3)[15], and therefore increase the likelihood of formation of an A-C base pair with two hydrogen bonds. Interestingly, we found only 7-deazadA boosted editing at the desired site by 20-fold, while editing of 3-deazadA was comparable to that of guanosine. In addition, 3-deazadA slightly enhanced editing at the off-target site (**Figure 5.3, Table 5.1**) . The effects observed though could be dependent on the ADAR isoform tested. Given ubiquitous expression of hADAR1, it is critical that guide designs promote editing with this isoform.[4,16] Therefore, we tested these guides against the ADAR1 p110 isoform in single turnover conditions. Due to slow reaction with p110 and the duplex tested, we could not report observed rate of reaction. Instead, we found the extent of editing at 120 min correlated with the rate observed for hADAR2. Generally, ADAR1 followed the same trend, it exhibited higher editing extent with dI followed by 7-deazadA. We did not observe editing at other off-target sites in the reaction conditions tested (**Figure 5.4**).

A

```
Serpin A target   5′-… GUGCUGACCAUCGACAAGAAAGGGACUGAAG…-3′
        gRNA      3′-  CACGACUGGUAGCUXCUCUUUCCCUGACUUC-5′
```



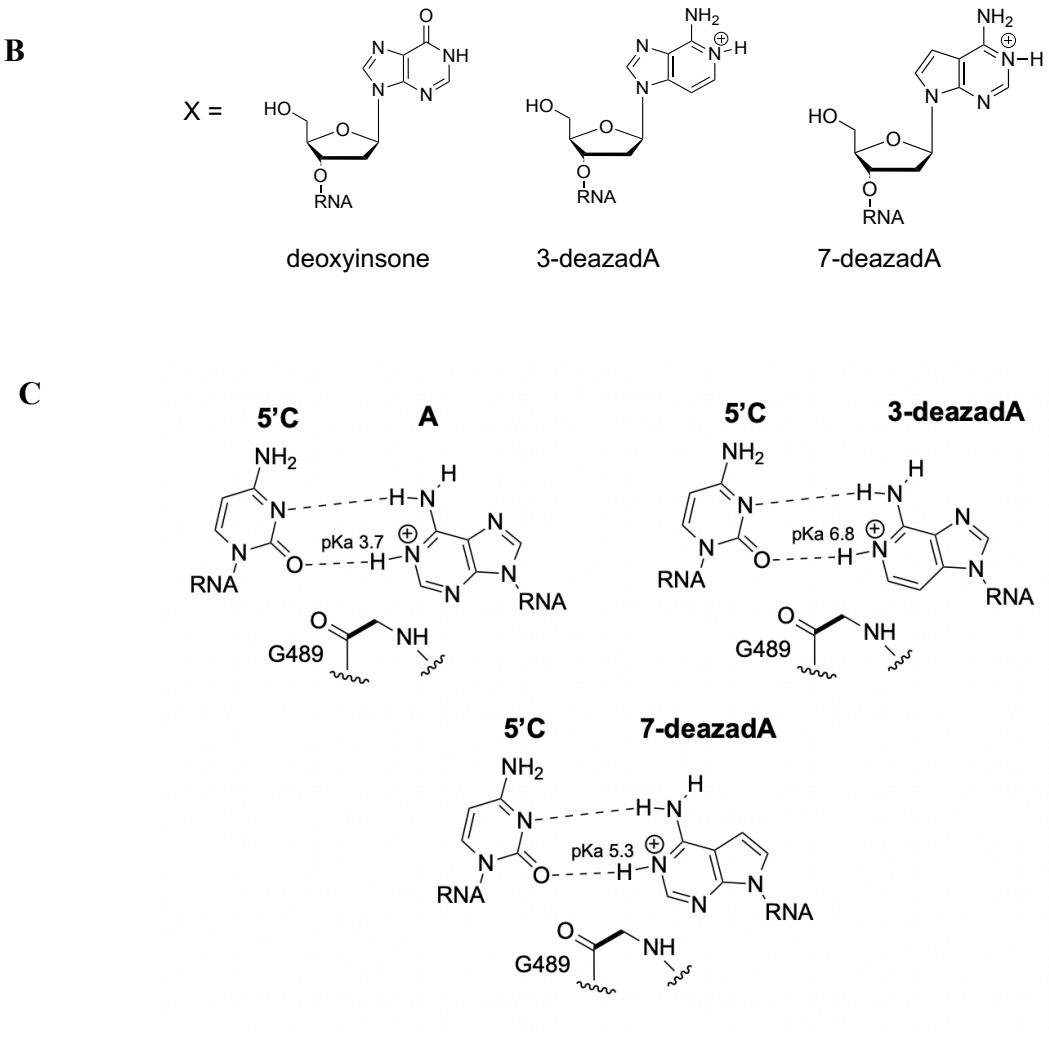


Figure 5.3. In vitro kinetics of hADAR2. Kinetic analysis of nucleoside analogs paired with a 5'C. (A) Sequence of the Serpin A target. (B) Adenosine deamination observed rate constants plotted for 100 nM hADAR2 on 10 nM mutant SerpinA1 substrate under single turnover conditions. X indicates the base pairing in the -1 position. [a] Data were fitted to the equation $[P]t = \alpha[1-\exp(-k_{obs} \cdot t)]$. [b] $k_{rel} = k_{obs}$ for analog / k_{obs}. (C) Predicted base-pairing event of 5'C with adenosine, 3-deaza dA and 7-deaza dA.

Table 5.1. **In vitro kinetics of hADAR2.** Adenosine deamination observed rate constants measured under single-turnover conditions for Serpin A1 target with hADAR2 under single turn over conditions. [a]Data were fitted to the equation: $[P]t = [P]f[1 - \exp(-k_{obs} \cdot t)]$. [b]$k_{rel} = k_{obs}$ for different nucleosides at X position/k_{obs} for X = G.

Enzyme	-1 Base(X)	$K_{obs}(min^{-1})$	k_{rel}
	G	0.02 ± 0.001	1
	A	0.03 ± 0.004	2
	C	-	-
	U	-	-
	dI	1.0 ± 0.14	53
	7-deazadA	0.4 ± 0.033	20
	3-deazadA	0.05 ± 0.01	3

```
Serpin A target  5′-… GUGCUGACCAUCGACAAGAAAGGGACUGAAG…-3′
        gRNA     3′-  CACGACUGGUAGCUXCUCUUUCCCUGACUUC-5′
```

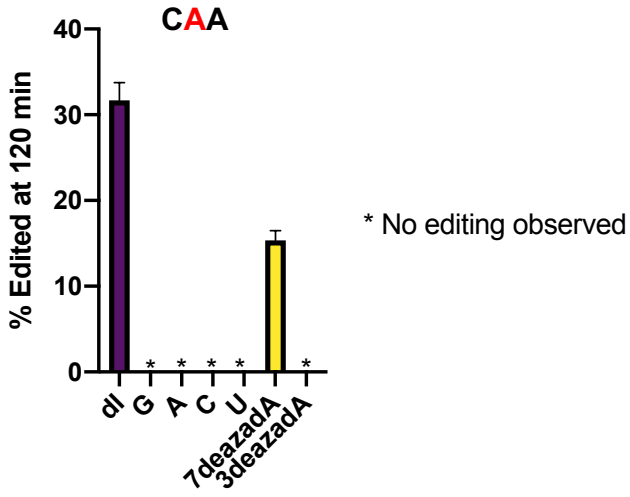
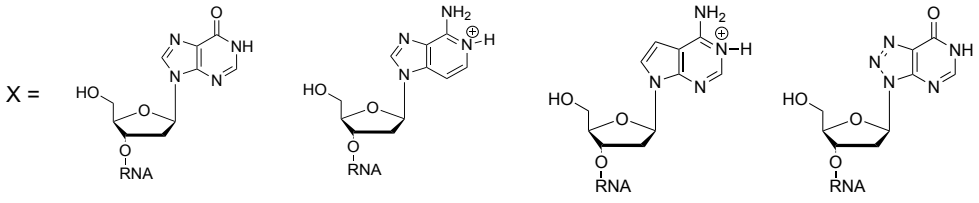


Figure 5.4. In vitro deamination endpoints of hADAR1. Adenosine deamination reaction end points plotted for 100 nM hADAR1 p110 on 10 nM mutant SerpinA1 substrate under single turnover conditions. X indicates the base pairing in the -1 position.

Subsequently, we sought to determine whether the observed editing effects were contingent on the substrate being used. Our assessment initiated with the design of a model substrate, consisting of a 5'-CA-3' sequence context derived from the SRC kinase mRNA.[17] This specific

target was originally conceived by Dr. Erin Doherty in the Beal lab. For our analysis, we created guide RNAs consisting of conventional bases (G, A, C, U), as well as analogs dI, 7-deazadA, 3-deazaA, and an additional analog, 8-aza-2'deoxyinosine (8-azadI) in the -1 position. The 8-azadI was synthesized by a lab member Randall Ouye. Successively, we evaluated these guide RNAs against full-length hADAR2. Consistent with our observations in the context of Serpin A1, we noted that for the SRC target, the most rapid reaction rate occurred with dI (7-fold increase), followed by 7-deazadA (5-fold increase) at the desired target site. Notably, 8-azadI, whose N1 pKa is lower (~7.8)[15,18] in comparison to dI (8.8)[18], exhibited a modest increase in editing, approximately 2-fold lower than dI (**Figure 5.5, Table 5.2**). It is worth noting, that SRC target was very robust substrate for ADAR2, and we observed substantial levels of editing even with a 5'C-G pairing (**Figure 5.5, Table 5.2**).



SRC Kinase Target 5'-…ACCAGGGUGGCCAUCAAAACCCUGAAGCCUG…-3'
gRNA 3'- UGGUCCCACCGGUAXCUUUGGGACUUCGGAC-5'

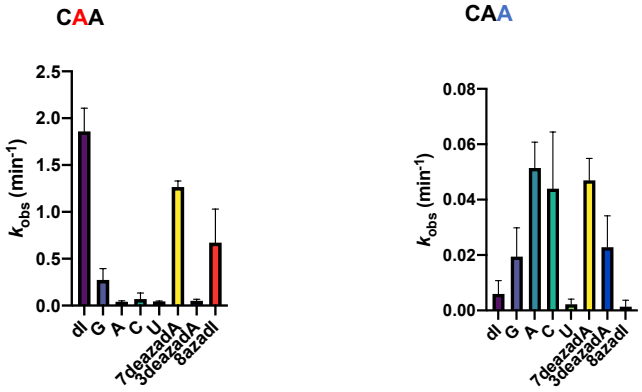


Figure 5.5. In vitro kinetics of hADAR2. Adenosine deamination observed rate constants plotted for 100 nM hADAR2 on 10 nM SRC substrate under single turnover conditions. X indicates the base pairing in the -1 position. [a] Data were fitted to the equation $[P]t = \alpha[1-\exp(-k_{obs}\cdot t)]$.

Table 5.2. **In vitro kinetics of hADAR2.** Adenosine deamination observed rate constants measured under single-turnover conditions for SRC target with hADAR2. [a]Data were fitted to the equation: $[P]t = [P]f[1 - \exp(-k_{obs}\cdot t)]$. [b]$k_{rel} = k_{obs}$ for different nucleosides at X position/k_{obs} for X = G.

Enzyme	-1 Base(X)	$K_{obs}(min^{-1})$	k_{rel}
	G	0.3 ± 0.1	1
	A	0.04 ± 0.01	0.2
	C	0.06 ± 0.05	0.2
	U	0.04 ± 0.005	0.2
	dI	1.8 ± 0.2	7
	7-deazadA	1.3 ± 0.05	5
	3-deazadA	0.05 ± 0.01	0.2
	8-azadI	2.5 ± 0.3	2.5

Lastly, we evaluated the reaction endpoint of the SRC target using hADAR1 p110. Parallel to our observations with SerpinA1 and consistent with the rates seen for hADAR2, we noted that 2'-deoxyinosine, 7-deaza-2'-deoxyadenosine, and 8-aza-2'-deoxyinosine (8-azadI) all contributed to an enhancement in editing (**Figure 5.6**). Notably, 2'-deoxyinosine exhibited a 2-fold increase in editing extent in comparison to 7-deazadA or 8-azadI . Importantly, no editing was detected at any off-target sites. This is encouraging due to the widespread expression of ADAR1, which is often the predominant isoform in various tissues.[19] Collectively, the enhancements brought about by 2'-deoxyinosine, 7-deazaadenosine, and 8-azaI underscore their potential in improving editing efficiency across distinct ADAR isoforms and disease targets consisting of 5'-CA-3' sites.

SRC Kinase Target 5′-…ACCAGGGUGGCCAUCAAAACCCUGAAGCCUG…-3′
gRNA 3′- UGGUCCCACCGGUAXCUUUGGGACUUCGGAC-5′

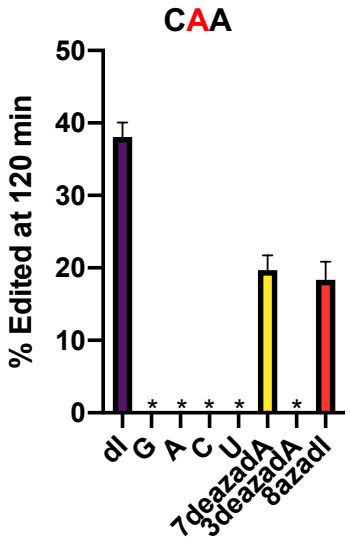


Figure 5.6. In vitro deamination endpoints of hADAR1. Adenosine deamination reaction endpoints plotted for 100 nM hADAR1 p110 on 10 nM SRC substrate under single turnover conditions. X indicates the base pairing in the -1 position.

Biochemical analysis of guide designs identified through EMERGe. In continuation of our work on enabling editing at 5'-GA-3' sites, we focused on discovering local secondary structures and motifs that promote on-target editing while mitigating off-target effects.[14] Despite the effectiveness of a 5'-G_{syn}-G_{anti} pair in facilitating on-target editing, this approach also heightened levels of off-target editing (**Figure 5.7**). Seeking to refine our approach, our laboratory has pioneered an innovative NGS–based EMERGe screening method aimed at identifying optimal guide RNAs.[14] This design incorporates a self-complementary hairpin structure, along with a 10-nucleotide variable region across the editing site. This strategic setup enables the identification of guide RNAs that facilitate editing within the confines of the same hairpin structure. Subsequently, the most promising guide RNAs undergo rigorous *in vitro* testing in a therapeutically relevant bi-

molecular system. Employing this approach, Beal lab graduate student Prince Salvador has successfully identified an effective guide strand noted as S5 that enables selective editing in R255X target implicated in Rett syndrome (**Figure 5.7**).[20] Notably, the editing of these targets by ADAR culminates in the restoration of a nonsense mutation to a functional MECP2 protein.[8] This is attributed to the A to I editing, which converts a stop codon to a tryptophan codon. When compared to previously identified G_{syn}-G_{anti} pair that enables editing at this target site, S5 demonstrated minimal off-target activity.

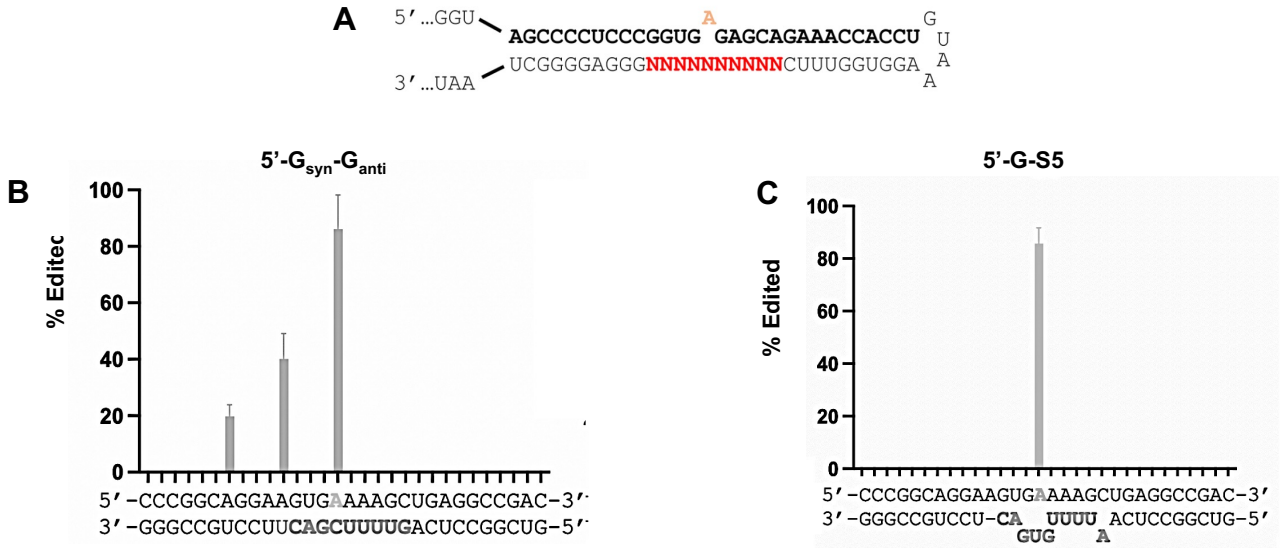


Figure 5.7. Adenosine deamination reaction endpoints plotted for 100 nM hADAR2 on 10 nM R255X (N10) hairpin substrate under single turnover conditions. N indicates the variable region for identification of unique guide strands. **(A)** Sequence of R255X target, consisting of 5'GAG-3' site and a variable N10 region. **(B)** Comparison of G_{syn}-G_{anti} to identified S5 motif for enabling editing at 5'-GAG-3'. Note: Figure generated by P. S. and edited by A.K.

To gain insights into the specific interactions of ADAR2 with the S5 motif that facilitate selective editing, we looked to solve a crystal structure of the S5 ADAR-RNA complex. To evaluate the feasibility of using this complex for crystallography, we synthesized a 32-base pair duplex consisting of S5 motif and hGli1 sequence. The guide design also contained the adenosine

analog (8-azanebularine). The hydration of 8-azanebularine triggers the formation of a tetrahedral intermediate lacking a suitable leaving group, effectively entrapping the protein-RNA complex in its bound state.[21] Subsequently, we conducted gel shift binding analysis on this duplex using ADAR R2D, a truncated form of hADAR2 (**Figure 5.8**). Furthermore, we compared the binding patterns to those of a duplex featuring the optimal ADAR codon 'UAG' (**Figure 5.8, Figure 5.9**).[9] Interestingly, the duplex containing the S5 motif exhibited an immediate shift to a dimer complex. Under similar conditions, the UAG duplex initially transitioned to a monomer complex, eventually forming a dimer at elevated concentrations of protein (**Figure 5.8, Figure 5.9**). Notably, the S5 duplex demonstrated significantly lower binding affinity (Kd = 74 nM), in comparison to UAG (Kd = 0.9 nM) (**Figure 5.8**). Consequently, our attempts to crystallize this complex using previously successful conditions for similar complexes, yielded no positive results.

Therefore, we embarked on structure-function studies and devised a modified version of the S5 motif, enhancing duplex stability with minimal effects on activity. In addition, we chose to use a disease relevant sequence R270X, another Rett syndrome associated nonsense mutation. Similarly, to R255X, A to I edit would result to a functional protein by conversion of a stop codon to a tryptophan.[20] This effort resulted in a high affinity protein-RNA complex. Additionally, we observed that ADAR2 R2D wild-type exhibited more effective RNA binding when compared to the hyperactive mutant (E488Q)[10], known for enhancing base flipping with typical RNA substrates (**Figure 5.10**). Our attempts to crystallize this complex resulted in an 8 Å resolution structure. The diffracted crystals grew under unique conditions consisting of 50mM HEPES pH 8.0, 50% PEG 200. Crystals were also observed in Pact Premier: B2, Natrix: F9, and Nuc Pro- F4.

Crystallization trials with optimized conditions varying salt, pH, and PEG concentrations are currently ongoing (**Table 5.4**).

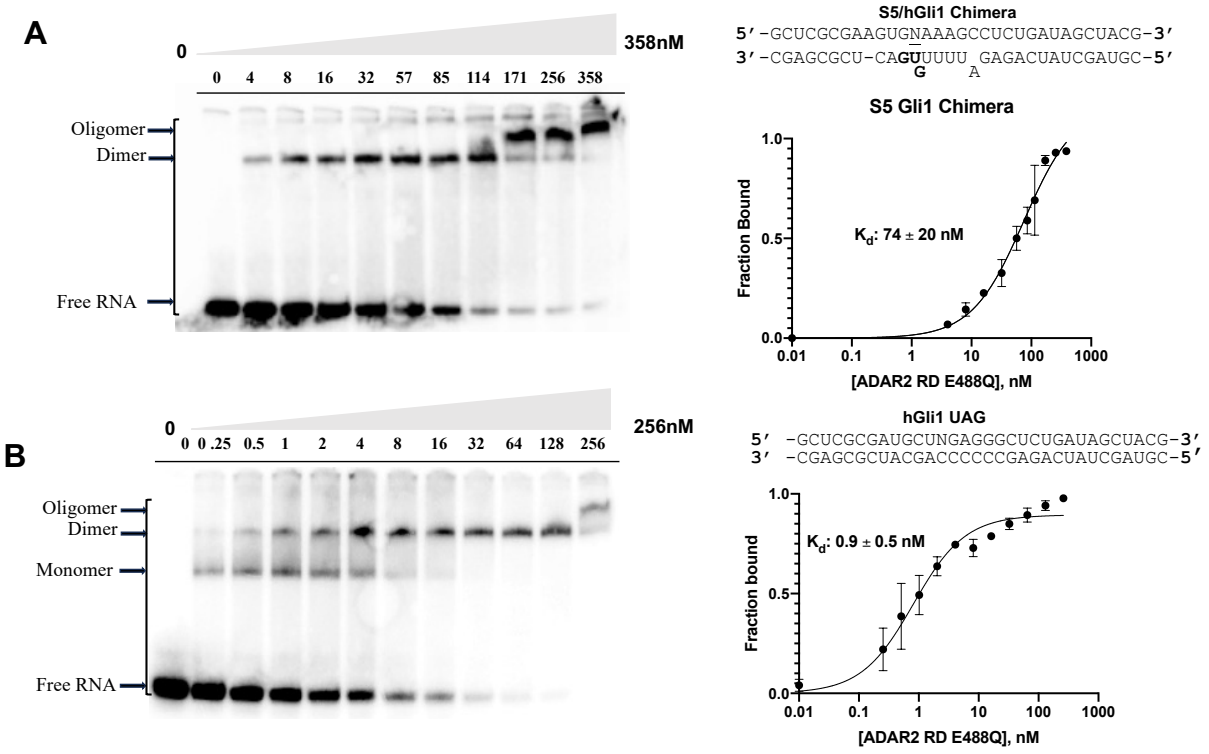


Figure. 5.8. Characterization of complex formed between ADAR2-R2D E488Q and a 32 bp 8-azanebularine (N) containing duplex with GUG motif and 8-azaN. (**A**) Sequence of S5 32 bp duplex used for crystallization and EMSA gel of hADAR-R2D E488Q with this duplex. Protein concentrations are as follows: lane 1: no protein, lanes 2-11: 4, 8, 16, 32, 57, 85, 114, 171, 256, 358 nM . Labeled are the bands for free RNA, and protein-RNA complex corresponding to a dimer, and oligomer. (**B**)Sequence of 32 bp duplex containing optimal 'UAG' sequence and EMSA gel of hADAR2-R2D E488Q with this duplex. Protein concentrations are as follows: lane 1: no protein, lanes 2-12: 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128, and 256 nM . (B) Quantification of protein binding by EMSA. Labeled are the bands for free RNA, and protein-RNA complex corresponding to a monomer, dimer, and oligomer.

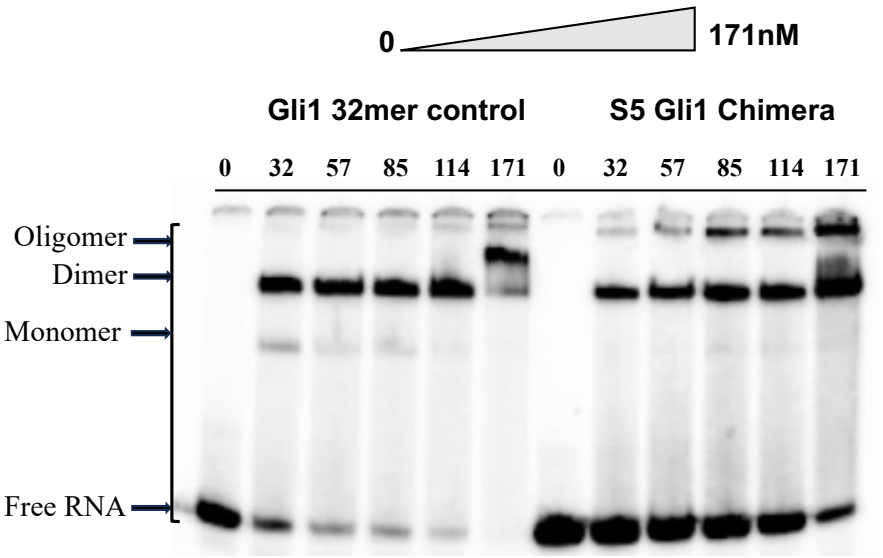
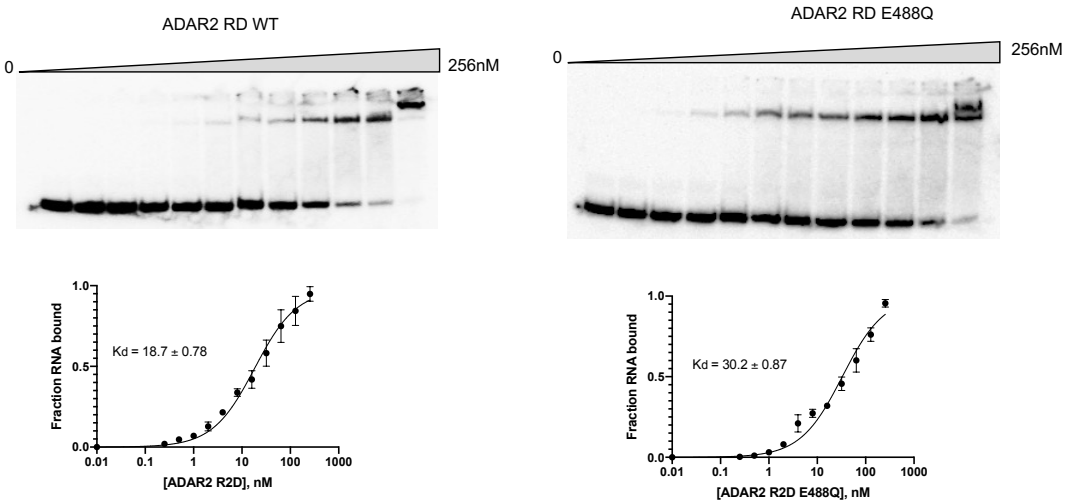


Figure. 5.9. Characterization of complex formed between ADAR2-R2D E488Q and a 32 bp 8-azanebularine (N) containing duplex with GUG motif or optimal sequence context UAG bearing duplex and 8-azaN. The gel displays both a monomer and dimer complex formation for UAG duplex, while the GUG motif immediately shifts to a dimer.

```
R270X Target:   5′-GAAACGGGGCUGNAAGCCGGGGAGUGUGGGUGG- 3′
   GUG Guide:   3′-CUUUGCCCCGAGUUUCGGCCCCUCACACCACC- 5′
                                G
```



ADAR2 R2D	K_D (nM)
WT	18.7 ± 0.78
E488Q	30.2 ± 0.87

Figure. 5.10. Characterization of complexes formed between wild-type ADAR2 RD or the hyper-active mutant E488Q with a 32 bp duplex containing 8-azanebularine (N), derived from the R270X sequence. The guide strand was designed to consist of the GUG motif. The gel-shift complexes show both ADAR2 R2D wild-type and E488Q form a protein-RNA complexes corresponding to a dimer and super shifted oligomer. In comparison, the wild-type protein displayed greater binding affinity.

5.3. Discussion.

Adenosine and guanosine analogs capable of pairing with 5'C enhances editing. ADARs exhibit a preference for editing nucleotides with a 5' nearest neighbor of U or A while avoiding those with 5' G or 5' C. This preference is attributed to the protein's flipping loop, which occupies the minor groove encompassing the edited base and its immediate 5' and 3' nearest neighbors. When modeling a 5' G-C or 5' C-G pair at this site, it is predicted to result in a clash with the 2-amino group in the minor groove and with G489, a key residue in the flipping loop (Figure 5.1). Previous research demonstrated that substituting the 5'U-A pair adjacent to the editing site with a 5'C-G pair led to an 80% reduction in reaction rate with hADAR2d. Similarly, using the adenosine analog 2-amino purine (2AP) as a base pairing partner resulted in an 80% rate reduction. These findings indicated that introducing an -NH2 group into the minor groove destabilizes the flipping loop and hinders the editing process (**Figure 5.11**). However, the implications of this effect on the full-length protein and in various sequence contexts remained unclear. Further kinetic analysis on the SerpinA1 target associated with AATD disease confirmed the adverse effect of a 5'C-G pair on editing. Consistent with previous findings, adenosine and guanosine analogs lacking a 2-amino group and capable of base pairing with a 5'C improved reaction rates. Notably, adenosine analog 7-deazadA and guanosine analog 2'-deoxyinosine significantly enhanced reaction rates by 20-fold and 50-fold, respectively (**Figure 5.3**). This enhanced editing effect extended to other sequence

contexts, as evidenced by heightened reaction rates with the same analogs on the SRC mRNA. Additionally, 8-azadI led to a slight 2.5-fold increase in the reaction rate (**Figure 5.5, Table 5.2**).

The improved editing effects of these analogs remained consistent across various ADAR isoforms and sequence contexts, indicating their potential utility in addressing disease targets characterized by a 5'-CA-3' sequence. However, the impact of the 5'CG pair was found to be sequence-dependent. While the SRC mRNA was a robust substrate for hADAR2, the most effective analogs yielded only a 5-7-fold improvement in reaction rates compared to the 5'CG pair. In contrast, this same substrate displayed limited reactivity with hADAR p110, with analogs 7-deazadA, 8-azadI, and 2'-deoxyinosine boosting editing extent (**Figure 5.4, Figure 5.6**). It's worth noting that all analogs included the 2'deoxy modification, while the other nucleobases retained a 2'OH, resulting in different sugar puckering at the -1 position. Previous structural analyses from our lab have indicated that the -1 nucleobase tends to favor a DNA-like 2' endo sugar pucker.[11] Therefore, it is imperative to investigate the impact of sugar puckering on the -1 base and editing within the 5'CA sequence contexts. To advance our understanding, future work should encompass a broader range of sequence contexts by varying the 3' nearest neighbor, explore cellular editing to comprehensively assess the efficacy of the tested analogs, and delve into the influence of sugar puckering within the 5'CA sequence contexts.

A GUG motif promotes selective editing at 5'G. The high throughput screening method EMERGe identified a distinctive 'GUG' motif that selectively enhanced editing within a 5'G<u>A</u> sequence context. Although the reaction rates were comparatively lower than those of a 5'G_{syn}-G_{anti} pair, minimal off-target editing was observed. To delve deeper into the specific protein-RNA interactions driving this selectivity, we pursued X-ray crystallography for this complex. The gel

shift analysis of this complex revealed a unique behavior: the GUG motif recruited a protein dimer even at low concentrations, with no observed monomer-RNA complexes (**Figure 5.8, Figure 5.9**). This effect could suggest the absence of off-target editing due to differing binding registers compared to the monomer-RNA complex. Crystallization trials with the GUG motif are ongoing at this time. In conclusion, these studies shed light on the structural factors influencing ADAR editing preferences. Nucleobase analogs and structural perturbations provide alternative strategies to enhance editing at disfavored sites 5'GA and 5'CA.

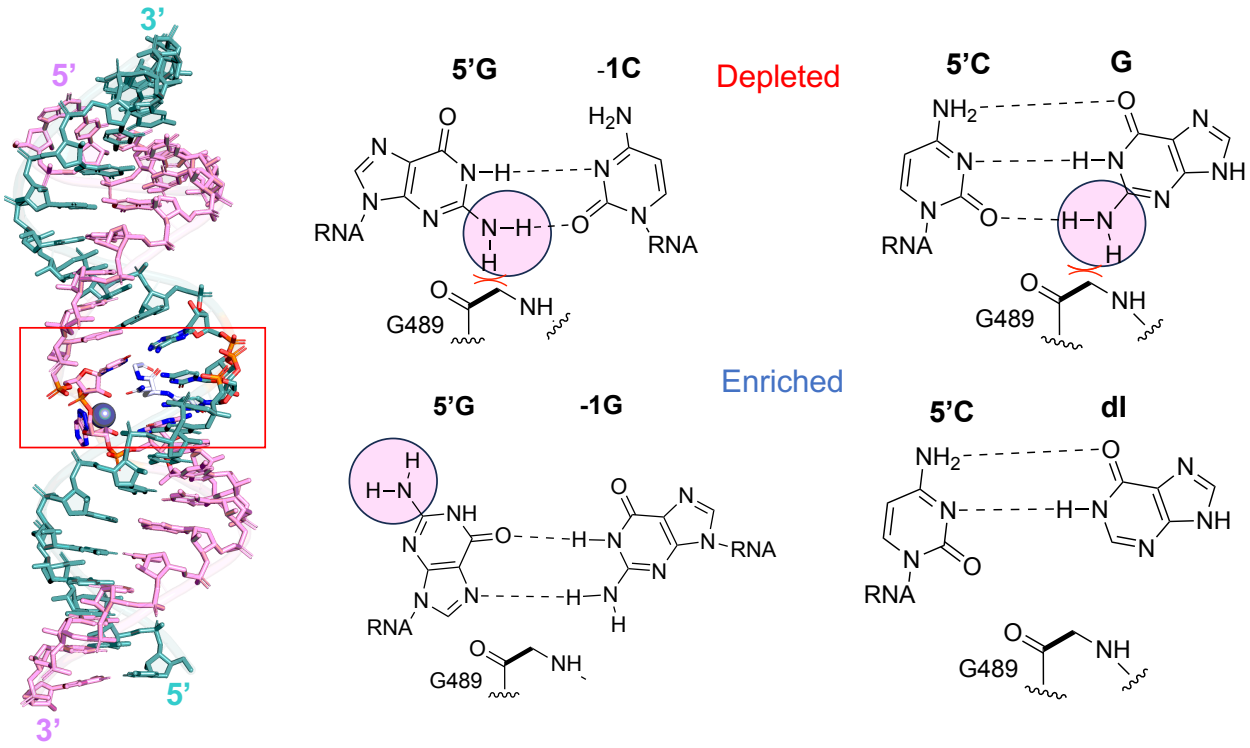


Figure 5.11. Impact of sequence context in ADAR editing. This figure summarizes ADAR editing preferences. Editing is reduced in 5' GA and 5' CA sequences but can be restored by modifying the guanosine's 2-amino group in 5' GA or positioning the base pairing partner at the -1 position of 5' CA away from G489, the key flipping loop residue.

5.4. Materials and methods

General biochemical procedures. Molecular-biology-grade bovine serum albumin (BSA) and RNase inhibitor (RNAsin) were purchased from New England BioLabs. SDS-polyacrylamide gels were visualized with a Molecular Dynamics 9400 Typhoon phosphorimager. Data were analyzed with Molecular Dynamics ImageQuant 5.2 software. All Matrix Assisted Laser Desorption/Ionization (MALDI) analyses were performed at the University of California, Davis Mass Spectrometry Facilities using a Bruker UltraFlextreme MALDI TOF/TOF mass spectrometer. Oligonucleotide masses were determined with Mongo Oligo Calculator v2.06. Oligonucleotides for sequencing and PCR were purchased from Integrated DNA Technologies or Dharmacon. All other oligonucleotides were synthesized as described below.

Synthesis of Oligonucleotides. Chemical synthesis for all oligonucleotides was performed using an ABI 394 synthesizer. All protected phosphoramidites were purchased from Glen Research except the 8-azanebularine (azaN) phosphoramidite which was purchased from Berry & Associates or synthesized as previously described.[15] Nucleosides were incorporated during the appropriate cycle on a 0.2 or 1.0 μmol scale**.** Appendix Table 5.1-5.3 show sequences of all oligonucleotides used in this study. Upon completion of the synthesis, columns were evaporated under reduced pressure for 12 h. All oligonucleotides were cleaved from the solid support by treatment with 1.5 ml 1:3 ethanol/30% NH_4OH at 55◦C for 12 h. The supernatant was transferred to a new screw-cap tube and evaporated under reduced pressure. For all oligonucleotides except the azaN-modified strand, desilylation was performed by treating the pellets with 250 l of 1M

TBAF–THF at room temperature overnight. For the azaN strand, desilylation was carried out in TEA•3HF as previously described.[15] To each reaction was added 75 mM sodium acetate in butanol. The oligonucleotides were then precipitated from a solution of 65% butanol at –70◦C for 2 h. The solution was centrifuged at 17 000 × g for 20 min, supernatant was removed, and the pellet was washed twice with cold 95% ethanol. The RNA pellets were then desalted using a Sephadex G-25 column and purified as described below. Purification of oligonucleotides Single-stranded RNA oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis and visualized by UV shadowing. Bands were excised from the gel, crushed and soaked overnight at 4°C in 0.5 M NaOAc, 0.1% sodium dodecyl sulfate (SDS), and 0.1 mM EDTA. Polyacrylamide fragments were removed with a 0.2 m filter, and the RNAs were precipitated from a solution of 75% EtOH at –70◦C for 12 h. The solution was centrifuged 17 000 × g for 20 min and supernatant was removed. The RNA solutions were lyophilized to dryness, resuspended in nuclease-free water and quantified by absorbance at 260 nm.

Preparation of duplex RNA substrates for crystallography. For crystallography, the unmodified RNA guide strand was purchased from Horizon Dharmacon and purified as described above. As in previous structures, the edited strand contained the adenosine analog azaN at the editing site. Duplex RNA was hybridized in water in a 1:1 ratio by heating to 95°C for 5 min and slow cooling to 30°C.

In vitro transcription of editing target RNAs. Target RNAs for deamination kinetic analyses were transcribed from DNA templates with the MEGAScript T7 Kit (ThermoFisher).

DNA digestion was performed using RQ1 RNase-free DNase (Promega). DNase-treated RNA product was purified as described above.

Preparation of duplex substrates for ADAR deamination kinetics. Purified guide and transcribed RNA were added in a 10:1 ratio to hybridization buffer (180 nM transcribed RNA target, 1.8 μM guide, 1X TE Buffer, 100 mM NaCl), heated to 95°C for 5 min, and slowly cooled to room temperature. Guide RNA sequences and target DNA sequences used for the experiments are listed in Appendix table 1 and 2.

Expression and purification of human ADAR2 constructs for deamination kinetics. Full length human ADAR2 (hADAR2) was overexpressed in Saccharomyces cerevisiae as previously described.[16] Purification of hADAR2 was carried out by lysing cells in buffer containing 20 mM Tris–HCl, pH 8.0, 5% glycerol, 1 mM -mercaptoethanol (BME), 750 mM NaCl, 35 mM imidazole and 0.01% Nonidet P-40 (NP-40) using a French press. Cell lysate was clarified by centrifugation (39 000 × g for 1 h). Lysate was passed over a 3 ml Ni-NTA column, which was then washed in three steps with 20 ml lysis buffer, wash I buffer (20 mM Tris–HCl, pH 8.0, 5% glycerol, 1 mM BME, 750 mM NaCl, 35 mM imidazole, 0.01% NP-40), wash II buffer (20 mM Tris–HCl, pH 8.0, 5% glycerol, 1 mM BME, 35 mM imidazole, 500 mM NaCl), and eluted with 20 mM Tris–HCl, pH 8.0, 5% glycerol, 1 mM BME, 400 mM imidazole, 100 mM NaCl. Fractions containing the target protein were pooled and concentrated to 30–80 M for use in biochemical assays. Protein concentrations were determined using BSA standards visualized by SYPRO orange staining of

SDS-polyacrylamide gels. Purified hADAR2 WT was stored in 20 mM Tris– HCl pH 8.0, 100 mM NaCl, 20% glycerol and 1 mM BME at –70◦C.

Expression and purification of ADAR1 p110 for deamination kinetics. Human ADAR1 p110 (UniProtKB P55265-5), consisting of a C-terminal His_{10}-tag was overexpressed in *Saccharomyces cerevisiae* BCY123 as previously described.[22] Cells were lysed using a microfluidizer in lysis buffer containing 20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 1 M KCl, 30 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine-HCl (TCEP-HCl), 0.05% (v/v) Triton X-100, and 50 μM $ZnCl_2$. The lysate was centrifuged at 39000 x g, 4 °C for 1 h and filtered using 0.45 μm filter. The clarified lysate was then passed over a 5 ml Ni-NTA column at a flow rate of 2 ml min^{-1} using an ÄKTA pure 25 FPLC system. The column was washed first with ten column volumes (10 CVs) of lysis buffer, followed with 10 CVs of wash I buffer (20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 500 mM KCl, 30 mM imidazole, 1mM TCEP-HCl, and 50 μM $ZnCl_2$). The protein was eluted with wash I buffer with a gradient of imidazole (30 mM to 400 mM) for 10 CVs. Fractions containing the target protein were pooled, concentrated, and dialyzed against a storage buffer containing 50 mM Tris-HCl pH 8.0, 10% (v/v) glycerol, 400 mM KCl, 50 mM imidazole, 1 mM TCEP-HCl, and 0.01% (v/v) Nonidet P-40 (NP-40). Final protein concentrations were determined by running the samples alongside bovine serum albumin (BSA) standards in an SDS-PAGE gel and visualized by staining with SYPRO Orange dye (Invitrogen).

Deamination assays with ADAR2 and ADAR1 p110. Deamination assays were performed under single-turnover conditions in 15 mM Tris–HCl pH 7.5 3% glycerol, 60 mM KCl, 1.5 mM

EDTA, 0.003% NP-40, 3 mM MgCl2, 160 U/ml RNAsin, 1.0 g/ml yeast tRNA, 10 nM RNA, and 75 nM human ADAR2. Each reaction solution was incubated at 30°C for 30 min before the addition of enzyme. Reactions were then incubated at 30°C for varying times prior to quenching with 95°C water and heating at 95°C for 5 min. Reaction products were used to generate cDNA using RT-PCR (Promega Access RT-PCR System). DNA was purified using a DNA Clean & Concentrator kit (Zymo) and subjected to Sanger Sequencing via GeneWiz (Azenta). The sequencing peak heights were quantified in SnapGene (Domatics). Data were fit to the equation $[P]t = [P]\,f\,[1 - e(-kobs \cdot t)\,]$ for ADAR2 where [P]t is percent edited at time t, [P]f is the final endpoint of editing, and kobs is the observed rate constant. Because of the slower reactions for ADAR1 p110 and lower reaction end point, data were fit to the equation $[P]t = 0.4 \cdot [1 - e(-kobs \cdot t)\,]$. Each experiment was carried out in triplicate where the k_{obs} reported is the average of each replicate ± standard deviation (SD). Statistical significance between groups was determined by one-way Analysis of Variance (ANOVA) with Tukey's multiple comparisons test using Prism software (GraphPad). For the ADAR1 p110 enzyme, deamination reactions were performed as above with the following modifications: The final reaction solution for ADAR1 p110 contained 15 mM Tris–HCl, pH 7.0 4% glycerol, 26 mM KCl, 40 mM potassium glutamate, 1.5 mM EDTA, 0.003% NP-40, 160 U/ml RNAsin, 1.0 g/ml yeast tRNA, 10 nM RNA and 250 nM ADAR1 p110.

Expression and purification of hADAR2 double stranded RNA binding domain and deaminase domain (hADAR2- R2D) for crystallography. Protein expression and purification were carried out by modifying a previously reported protocol (18). S. cerevisiae BCY123 cells were transformed with a pSc-ADAR construct encoding hADAR2-R2D (corresponding to

residues 214–701). Cells were streaked on yeast minimal media minus uracil (CM-ura) plates. A single colony was used to inoculate a 15 ml CM-ura starter culture. After cultures were shaken at 300 rpm and 30°C overnight, 10 ml of starter culture was used to inoculate each liter of yeast growth medium. After cells reached an OD600 of 1.5 (∼20– 24 h) cells were induced with 110 ml of sterile 30% galactose per liter and protein was expressed for 6 h. Cells were collected by centrifugation at 5000 × g for 10 min and stored at –80°C. Cells were lysed in 750 mM NaCl in buffer A (20 mM Tris–HCl, pH 8.0, 5% glycerol, 35 mM imidazole, 1 mM BME, and 0.01% Triton X-100) with a microfluidizer. Cell lysate was clarified by centrifugation (39 000 × g for 25 min). Lysate was passed over a 5 ml Ni-NTA column equilibrated with buffer A with 750 mM NaCl, which was then washed in three steps with 50 ml of lysis buffer, wash I buffer (buffer A + 300 mM NaCl), and wash II buffer (buffer A + 100 mM NaCl). Protein was eluted with a 35– 300 mM imidazole gradient in wash II buffer over 80 min at a flow rate of 1 ml/min. Fractions containing target protein were pooled and further purified on a 2 ml GE Healthcare Lifesciences Hi-Trap Heparin HP column in wash II buffer without BME. The His10 fusion protein was washed with 50 ml of wash II buffer without BME and eluted with a 100–1000 mM NaCl gradient over 60 min at a flow rate of 0.8 ml/min. Fractions containing target protein were pooled and cleaved with an optimized ratio of 1 mg of Histagged TEV protease per 1 mg of protein. The final salt concentration was targeted to be 300 mM NaCl, high salt can slow down TEV reaction. Cleavage was carried out for 4 h at room temperature without agitation, followed by overnight cleavage at 4°C before the product was passed over another Ni-NTA column at a flow rate of 0.5 ml/min. The flow through and wash were collected and passed through another Ni-NTA column to remove remaining uncleaved protein. It is important to note, high glycerol content after TEV cleavage

must be dialyzed away or diluted to final concentration of 5% glycerol. The glycerol content can affect proper separation of TEV and ADAR2. In addition, TEV and ADAR proteins can associate with each other, and a second column is sometimes required. The flow through and wash were collected, dialyzed against 20 mM Tris, pH 8.0, 200 mM NaCl, 5% glycerol and 1 mM BME, followed by concentration to just under 1ml for gel filtration on a GE Healthcare HiLoad 16/600 Superdex 200 PG column. Fractions containing purified protein were pooled and concentrated to 7–9 mg/ml for crystallization trials.

Crystallization of the hADAR2-R2D-RNA complex. Crystals of the hADAR2-R2D GUG RNA complex were grown at room temperature by the sitting-drop vapor-diffusion method. A solution of 1.0 μl volume containing 6.0 mg/ml protein (100 uM) and 50 uM RNA was mixed with 1.0 μl of varying concentrations of salt (50-100 mM NaCl), PEG 4000 (10-30%) and pH (50mM MOPS pH 6.5-8.0). These conditions were based on optimization trays that led to successful crystallization of ADAR2 R2D bound to 5'G_{syn}-G_{anti}, and 5'G-3deazadA_{anti} complexes. In addition, hanging drop methods were used to set up 96 well screens with MCSG-1, MCSG-2, and Nuc-pro HTS.

Appendix Chapter 5.

Appendix Table 5.1. Sequences for *in vitro* kinetics of the *SerpinA1* target. All PCR primers are 2'-deoxynucleotides.

SerpinA1 guide strand -1 A	5'-CUUCAGUCCCUUUCUC**A**UCGAUGGUCAGCAC-3'
SerpinA1 guide strand -1 G	5'-CUUCAGUCCCUUUCUC**G**UCGAUGGUCAGCAC-3'
SerpinA1 guide strand -1 C	5'-CUUCAGUCCCUUUCUC**C**UCGAUGGUCAGCAC-3'
SerpinA1 guide strand -1 U	5'-CUUCAGUCCCUUUCUC**U**UCGAUGGUCAGCAC-3'
SerpinA1 guide strand -1 2'-deoxy Inosine (dI)	5'-CUUCAGUCCCUUUCUC**dI**UCGAUGGUCAGCAC-3'
*SerpinA1*guide strand -1 3-deazadeoxyadenosine(3dA)	5'-CUUCAGUCCCUUUCUC**3dA**UCGAUGGUCAGCAC-3'
SerpinA1 guide strand -1 7-deazadeoxyadenosine(7dA)	5'-CUUCAGUCCCUUUCUC**7dA**UCGAUGGUCAGCAC-3'
SerpinA1 RT-PCR forward and sequencing primer	5'TAATACGACTCACTATAGGGTCCATTACTGGAACCTATGATCTG-3'
SerpinA1 RT-PCR reverse primer	5'-TGGGATTCACCACTTTTCCC-3'

DNA template sequence for *in vitro* kinetics of the SerpinA1 target. T7 sequence is shown in maroon. Editing site is highlighted in red.
TAATACGACTCACTATAGG**G**TCCATTACTGGAACCTATGATCTGAAGAGCGTCCTGG
GTCAACTGGGCATCACTAAGGTCTTCAGCAATGGGGCTGACCTCTCCGGGGTCACAG
AGGAGGCACCCCTGAAGCTCTCCAAGGCCGTGCATAAGGCTGTGCTGACCATCGAC
AAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATC
CCCCCCGAGGTCAAGTTCAACAAACCCTTTGTCTTCTTAATGATTGAACAAAATACC
AAGTCTCCCCTCTTCATGGGAAAAGTGGTGAATCCCA

Appendix Table 5.2. Sequences for *in vitro* kinetics of the *SRC* target. All PCR primers are 2'-deoxynucleotides.

SRC guide strand -1 A	5'-CAGGCUUCAGGGUUUC**A**AUGGCCACCCUGGU -3'
SRC guide strand -1 G	5'-CAGGCUUCAGGGUUUC**G**AUGGCCACCCUGGU -3'
SRC guide strand -1 C	5'-CAGGCUUCAGGGUUUC**C**AUGGCCACCCUGGU -3'
SRC guide strand -1 U	5'-CAGGCUUCAGGGUUUC**U**AUGGCCACCCUGGU -3'
SRC guide strand -1 2'-deoxy Inosine (dI)	5'-CAGGCUUCAGGGUUUC**dI**AUGGCCACCCUGGU -3'
SRC guide strand -1 3-deazadeoxyadenosine(3dA)	5'-CAGGCUUCAGGGUUUC**3dA**AUGGCCACCCUGGU -3'
SRC guide strand -1 7-deazadeoxyadenosine(7dA)	5'-CAGGCUUCAGGGUUUC**7dA**AUGGCCACCCUGGU -3'

SRC RT-PCR forward and sequencing primer	TAATACGACTCACTATAGGGTGCCCCACGTCCAAGCCG
SRC RT-PCR reverse primer	TCCCCTTGCTCATGTACTCCG

DNA template sequence for *in vitro* kinetics of the SRC target. T7 sequence is shown in maroon. Editing site is highlighted in red.
TAATACGACTCACTATAGGGTGCCCCACGTCCAAGCCGCAGACTCAGGGCCTGGCC
AAGGATGCCTGGGAGATCCCTCGGGAGTCGCTGCGGCTGGAGGTCAAGCTGGGCCA
GGGCTGCTTTGGCGAGGTGTGGATGGGGACCTGGAACGGTACCACCAGGGTGGCCA
TCAAAACCCTGAAGCCTGGCACGATGTCTCCAGAGGCCTTCCTGCAGGAGGCCCAG
GTCATGAAGAAGCTGAGGCATGAGAAGCTGGTGCAGTTGTATGCTGTGGTTTCAGA
GGAGCCCATTTACATCGTCACGGAGTACATGAGCAAGGGGA

Appendix Table 5.3. Sequences for binding studies of GUG motifs.

hGli1 top 8AN	5'-GCUCGCGAAGUGNAAAGCCUCUGAUAGCUACG -3'
hGli1 GUG bottom	5'-CGUAGCUAUCAGAGAUUUUGUGACUCGCGAGC -3'
R270X top 8AN	5'-GAAACGGGGCUGNAAGCCGGGGAGUGUGGUGG -3'
R270X bottom	5'-CCACCACACUCCCCGGCUUGUGAGCCCCGUUUC -3'

Appendix Table 5.4. Conditions for optimizations trays for R270X GUG and ADAR2 RD WT

100mM Tris-HCl	PEG 300	PEG 200	NaCl	MgSO4
pH 8	0%	30%, 40%, 50%, 60%	75, 100, 125, 150, 175, 200 mM	0mM
pH 8	0%	30%, 40%, 50%, 60%	75, 100, 125, 150, 175, 200 mM	50mM
pH 7, 7.5, 8	0%	50%	75, 100, 125, 150, 175, 200 mM	0mM
pH 8	30%, 40%, 50%, 60%	0%	75, 100, 125, 150, 175, 200 mM	0mM

5.5. References

(1) Xiong, X.; Yi, C.; Peng, J. Epitranscriptomics: Toward A Better Understanding of RNA Modifications. *Genomics Proteomics Bioinformatics* **2017**, *15* (3), 147–153.
(2) Eisenberg, E.; Levanon, E. Y. A-to-I RNA Editing — Immune Protector and Transcriptome Diversifier. *Nat Rev Genet* **2018**, *19* (8), 473–490.
(3) Bass, B. L. RNA Editing by Adenosine Deaminases That Act on RNA. *Annu Rev Biochem* **2002**, *71* (1), 817–846.
(4) Monian, P.; Shivalila, C.; Lu, G.; Shimizu, M.; Boulay, D.; Bussow, K.; Byrne, M.; Bezigian, A.; Chatterjee, A.; Chew, D.; Desai, J.; Favaloro, F.; Godfrey, J.; Hoss, A.; Iwamoto, N.; Kawamoto, T.; Kumarasamy, J.; Lamattina, A.; Lindsey, A.; Liu, F.; Looby, R.; Marappan, S.; Metterville, J.; Murphy, R.; Rossi, J.; Pu, T.; Bhattarai, B.; Standley, S.; Tripathi, S.; Yang, H.; Yin, Y.; Yu, H.; Zhou, C.; Apponi, L. H.; Kandasamy, P.; Vargeese, C. Endogenous ADAR-Mediated RNA Editing in Non-Human Primates Using Stereopure Chemically Modified Oligonucleotides. *Nat Biotechnol* **2022**, *40* (7), 1093–1102.
(5) Booth, B. J.; Nourreddine, S.; Katrekar, D.; Savva, Y.; Bose, D.; Long, T. J.; Huss, D. J.; Mali, P. RNA Editing: Expanding the Potential of RNA Therapeutics. *Molecular Therapy* **2023**, *31* (6), 1533–1549.
(6) Li, M.; Yan, C.; Jiao, Y.; Xu, Y.; Bai, C.; Miao, R.; Jiang, J.; Liu, J. Site-Directed RNA Editing by Harnessing ADARs: Advances and Challenges. *Funct Integr Genomics* **2022**, *22* (6), 1089–1103.
(7) Bellingrath, J.-S.; McClements, M. E.; Fischer, M. D.; MacLaren, R. E. Programmable RNA Editing with Endogenous ADAR Enzymes – a Feasible Option for the Treatment of Inherited Retinal Disease? *Front Mol Neurosci* **2023**, *16*.
(8) Doherty, E. E.; Karki, A.; Wilcox, X. E.; Mendoza, H. G.; Manjunath, A.; Matos, V. J.; Fisher, A. J.; Beal, P. A. ADAR Activation by Inducing a Syn Conformation at Guanosine Adjacent to an Editing Site. *Nucleic Acids Res* **2022**, *50* (19), 10857–10868.
(9) Eggington, J. M.; Greene, T.; Bass, B. L. Predicting Sites of ADAR Editing in Double-Stranded RNA. *Nat Commun* **2011**, *2* (1), 319.
(10) Kuttan, A.; Bass, B. L. Mechanistic Insights into Editing-Site Specificity of ADARs. *Proceedings of the National Academy of Sciences* **2012**, *109* (48).
(11) Matthews, M. M.; Thomas, J. M.; Zheng, Y.; Tran, K.; Phelps, K. J.; Scott, A. I.; Havel, J.; Fisher, A. J.; Beal, P. A. Structures of Human ADAR2 Bound to DsRNA Reveal Base-Flipping Mechanism and Basis for Site Selectivity. *Nat Struct Mol Biol* **2016**, *23* (5), 426–433.
(12) Schneider, M. F.; Wettengel, J.; Hoffmann, P. C.; Stafforst, T. Optimal GuideRNAs for Re-Directing Deaminase Activity of HADAR1 and HADAR2 in Trans. *Nucleic Acids Res* **2014**, *42* (10), e87–e87.
(13) Latifi, N.; Mack, A. M.; Tellioglu, I.; Di Giorgio, S.; Stafforst, T. Precise and Efficient C-to-U RNA Base Editing with SNAP-CDAR-S. *Nucleic Acids Res* **2023**.

(14) Jacobsen, C. S.; Salvador, P.; Yung, J. F.; Kragness, S.; Mendoza, H. G.; Mandel, G.; Beal, P. A. Library Screening Reveals Sequence Motifs That Enable ADAR2 Editing at Recalcitrant Sites. *ACS Chem Biol* **2023**. https://doi.org/10.1021/acschembio.3c00107.
(15) Mlotkowski, A. J.; Schlegel, H. B.; Chow, C. S. Calculated PKa Values for a Series of Aza- and Deaza-Modified Nucleobases. *J Phys Chem A* **2023**, *127* (15), 3526–3534.
(16) Kim, U.; Wang, Y.; Sanford, T.; Zeng, Y.; Nishikura, K. Molecular Cloning of CDNA for Double-Stranded RNA Adenosine Deaminase, a Candidate Enzyme for Nuclear RNA Editing. *Proceedings of the National Academy of Sciences* **1994**, *91* (24), 11457–11461.
(17) Kamps, M. P.; Sefton, B. M. Neither Arginine nor Histidine Can Carry out the Function of Lysine-295 in the ATP-Binding Site of P60src. *Mol Cell Biol* **1986**, *6* (3), 751–757.
(18) Jones, E. L.; Mlotkowski, A. J.; Hebert, S. P.; Schlegel, H. B.; Chow, C. S. Calculations of PKa Values for a Series of Naturally Occurring Modified Nucleobases. *J Phys Chem A* **2022**, *126* (9), 1518–1529.
(19) George, C. X.; Wagner, M. V.; Samuel, C. E. Expression of Interferon-Inducible RNA Adenosine Deaminase ADAR1 during Pathogen Infection and Mouse Embryo Development Involves Tissue-Selective Promoter Utilization and Alternative Splicing. *Journal of Biological Chemistry* **2005**, *280* (15), 15020–15028.
(20) Palmieri, M.; Pozzer, D.; Landsberger, N. Advanced Genetic Therapies for the Treatment of Rett Syndrome: State of the Art and Future Perspectives. *Front Neurosci* **2023**, *17*.
(21) Haudenschild, B. L.; Maydanovych, O.; Véliz, E. A.; Macbeth, M. R.; Bass, B. L.; Beal, P. A. A Transition State Analogue for an RNA-Editing Reaction. *J Am Chem Soc* **2004**, *126* (36), 11213–11219.
(22) Mendoza, H. G.; Matos, V. J.; Park, S.; Pham, K. M.; Beal, P. A. Selective Inhibition of ADAR1 Using 8-Azanebularine-Modified RNA Duplexes. *Biochemistry* **2023**, *62* (8), 1376–1387.