Title

Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity.

Journal

Nature structural & molecular biology, 23(5)

ISSN

1545-9993

Authors

Matthews, Melissa M
Thomas, Justin M
Zheng, Yuxuan
et al.

Publication Date

2016-05-01

DOI

10.1038/nsmb.3203

Peer reviewed

nature structural & molecular biology

Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity

Melissa M Matthews[1,3], Justin M Thomas[1,3], Yuxuan Zheng[1], Kiet Tran[1], Kelly J Phelps[1], Anna I Scott[2], Jocelyn Havel[1], Andrew J Fisher[1,2] & Peter A Beal[1]

Adenosine deaminases acting on RNA (ADARs) are editing enzymes that convert adenosine to inosine in duplex RNA, a modification reaction with wide-ranging consequences in RNA function. Understanding of the ADAR reaction mechanism, the origin of editing-site selectivity, and the effect of mutations is limited by the lack of high-resolution structural data for complexes of ADARs bound to substrate RNAs. Here we describe four crystal structures of the human ADAR2 deaminase domain bound to RNA duplexes bearing a mimic of the deamination reaction intermediate. These structures, together with structure-guided mutagenesis and RNA-modification experiments, explain the basis of the ADAR deaminase domain's dsRNA specificity, its base-flipping mechanism, and its nearest-neighbor preferences. In addition, we identified an ADAR2-specific RNA-binding loop near the enzyme active site, thus rationalizing differences in selectivity observed between different ADARs. Finally, our results provide a structural framework for understanding the effects of ADAR mutations associated with human disease.

RNA-editing reactions alter a transcript's genomically encoded sequence by inserting, deleting, or modifying nucleotides[1]. Deamination of A, the most common form of RNA editing in humans, generates inosine (I) at the corresponding nucleotide position. Because I base-pairs with C, it functions similarly to G in cellular processes such as splicing, translation, and reverse transcription[2,3]. A-to-I editing has wide-ranging consequences in RNA function, including altering microRNA-recognition sites, redirecting splicing, and changing the meaning of specific codons[4–6]. Two different enzymes carry out A-to-I editing in humans: ADAR1 and ADAR2 (ref. 7). ADAR activity is required for nervous-system function, and altered editing has been linked to neurological disorders such as epilepsy and Prader–Willi syndrome[8–10]. In addition, mutations in the gene encoding ADAR1 cause the autoimmune disease Aicardi–Goutières syndrome (AGS) and the skin disorder dyschromatosis symmetrica hereditaria (DSH)[11–13]. Hyperediting has been observed at certain sites in cancer cells, such as in the *AZIN1* mRNA (for antizyme inhibitor 1)[14,15]. However, hypoediting also occurs in cancer-derived cell lines, as exemplified by the reduced editing observed in *GLI1* mRNA (for glioma-associated oncogene 1)[16].

The ADAR proteins have a modular structure with double-stranded RNA (dsRNA)-binding domains (dsRBDs) and a C-terminal deaminase domain[17] (human (h) ADAR2 domains in **Fig. 1a**). ADARs efficiently deaminate specific adenosines in duplex RNA while leaving most adenosines unmodified[18]. The mechanism of adenosine deamination requires ADAR to flip the reactive base out of an RNA double helix to access its active site[17]. How an enzyme can accomplish this task with a duplex RNA substrate is not known. Furthermore, how the ADAR deaminase domain contributes to editing-site selectivity is also unknown; to our knowledge, no structures of ADAR deaminase domain–RNA complexes have been reported. To address these knowledge gaps, we set out to trap the human ADAR2 deaminase domain (amino acids (aa) 299–701, hADAR2d) bound to different duplex RNAs and to solve structures of the resulting complexes with X-ray crystallography. We then evaluated the importance of protein-RNA contacts by using structure-guided mutagenesis and RNA-modification experiments coupled with adenosine-deamination kinetics.

RESULTS

Trapping the flipped conformation

The ADAR reaction involves the formation of a hydrated intermediate that loses ammonia and thereby generates the inosine-containing product RNA[17] (reaction scheme in **Fig. 1b**). The covalent hydrate of the nucleoside analog 8-azanebularine mimics the proposed high-energy intermediate[19] (reaction scheme in **Fig. 1b**). For trapping hADAR2d bound to RNA for crystallography, we incorporated 8-azanebularine into the edited site of duplex RNAs that have recently been shown to be excellent substrates for deamination by hADAR2d[20] (duplex sequence in **Fig. 1c**; characterization of protein–RNA complex in **Supplementary Fig. 1**). In addition, for one of these duplexes (*Bdf2*), we positioned the 8-azanebularine opposite either U or C to mimic an A-U pair or A-C mismatch at the editing site, creating a total of three different RNA substrates for structural studies (**Fig. 1c**). The hADAR2d protein (without RNA bound) has previously been crystallized and structurally characterized, thus revealing features of the active site, including the presence of zinc[21]. In addition, an inositol

[1]Department of Chemistry, University of California, Davis, Davis, California, USA. [2]Department of Molecular and Cellular Biology, University of California, Davis, Davis, California, USA. [3]These authors contributed equally to this work. Correspondence should be addressed to P.A.B. (pabeal@ucdavis.edu) or A.J.F. (ajfisher@ucdavis.edu).

Received 1 February; accepted 11 March; published online 11 April 2016; doi:10.1038/nsmb.3203

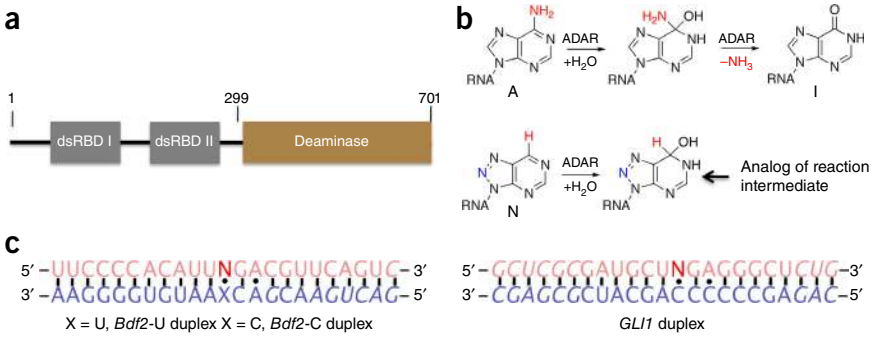


Figure 1 Human ADAR2 and modified RNAs for crystallography. (**a**) Domain map for human ADAR2. (**b**) ADAR reaction showing intermediate and 8-azanebularine (N) hydrate that mimics this structure. (**c**) Duplex RNAs used for crystallization. The *Bdf2* duplex sequence is derived from an editing site found in *Saccharomyces cerevisiae Bdf2* mRNA[31], and *GLI1* duplex has sequence surrounding the human *GLI1*-mRNA editing site[16]. Italics indicate nucleotides added for duplex stability.

hexakisphosphate (IHP) molecule has been found to be buried in the core of the protein hydrogen-bonded to numerous conserved polar residues. For crystallization of hADAR2d–RNA complexes, we used both the wild type (WT) deaminase domain and a mutant (E488Q) with enhanced catalytic activity[20,22] (descriptions of the crystallization conditions, X-ray diffraction data collection and solution of the structures in Online Methods).

Four protein–RNA combinations generated diffracting crystals that resulted in high-resolution structures (hADAR2d WT–*Bdf2*-U, hADAR2d WT–*Bdf2*-C, hADAR2d E488Q–*Bdf2*-C, and hADAR2d E488Q–*GLI1*) (**Table 1**). In each of these complexes, the protein binds the RNA on one face of the duplex, over ~20 bp, by using a positively charged surface near the zinc-containing active site (**Fig. 2** and **Supplementary Fig. 2a**). The large binding site (1,493 Å^2 RNA surface area and 1,277 Å^2 protein surface area buried) observed for hADAR2d is consistent with the results from recent footprinting studies[20]. Both strands of the RNA contact the protein, and the majority of these interactions are mediated through the phosphodiester-ribose backbone near the editing site (**Fig. 2c** and **Supplementary Fig. 2b–d**).

The structures show a large deviation from an A-form RNA conformation at the editing site (**Figs. 2** and **3** and **Supplementary Video 1**). The 8-azanebularine is flipped out of the helix and bound in the active site as its covalent hydrate, where it interacts with several amino acids including V351, T375, K376, E396, and R455 (**Fig. 3a** and **Supplementary Fig. 3a**). The side chain of E396 hydrogen-bonds to purine N1 and O6. This interaction was expected, given the proposed role of E396 in mediating proton transfer to and from N1 of the substrate adenosine[17]. The 2′-hydroxyl of 8-azanebularine hydrogen-bonds to the backbone carbonyl of T375 while the T375 side chain contacts its 3′-phosphodiester. R455 and K376 help position

Table 1 Data collection and refinement statistics

	ADAR2-D E488Q–*BDF2*-C 23-mer (PDB 5ED1)	ADAR2-D E488Q–*GLI1* 23-mer (PDB 5ED2)	ADAR2-D WT–*BDF2*-U 23-mer (PDB 5HP2)	ADAR2-D WT–*BDF2*-C 23-mer (PDB 5HP3)
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, *b*, *c*, (Å)	82.36, 107.50, 121.10	79.13, 81.61, 256.62	81.32, 106.68, 120.49	81.51, 107.21, 120.62
Resolution (Å)	100–2.75 (2.82–2.75)	50.0–2.95 (3.03–2.95)	100–2.98 (3.06–2.98)	100–3.09 (3.17–3.09)
R_{merge} (%)	7.0 (68.5)	9.6 (135.1)	14.4 (86.3)	11.6 (68.7)
$CC_{1/2}$	99.6 (66.3)	99.7 (47.6)	99.1 (75.3)	99.3 (77.1)
$I/\sigma(I)$	11.57 (1.52)	12.27 (1.16)	10.00 (1.88)	10.56 (1.75)
Completeness (%)	96.5 (98.8)	98.1 (98.9)	97.3 (90.0)	96.8 (89.1)
Redundancy	2.93 (3.00)	5.19 (5.13)	4.79 (4.56)	3.31 (2.84)
Refinement				
Resolution (Å)	2.75	2.95	2.98	3.09
No. of reflections ($F > 0$)	27,153	35,727	21,376	19,325
R_{work} / R_{free}	16.27 / 22.34	18.79 / 20.75	16.67 / 24.67	16.29 / 23.79
No. of atoms				
Protein	6,197	6,038	6,168	6,157
RNA	973	1,950	973	973
Inositol Hexakisphosphate	72	72	72	72
Zn	2	2	2	2
Water	33	–	1	1
B factors				
Protein	68.46	90.65	63.63	67.92
RNA	88.24	108.8	69.70	77.49
Inositol Hexakisphosphate	47.10	65.57	44.23	43.77
Zn	48.47	64.38	38.25	49.04
Water	48.11	–	43.66	50.20
R.m.s. deviations				
Bond lengths (Å)	0.010	0.007	0.009	0.010
Bond angles (°)	1.342	0.885	1.359	1.365

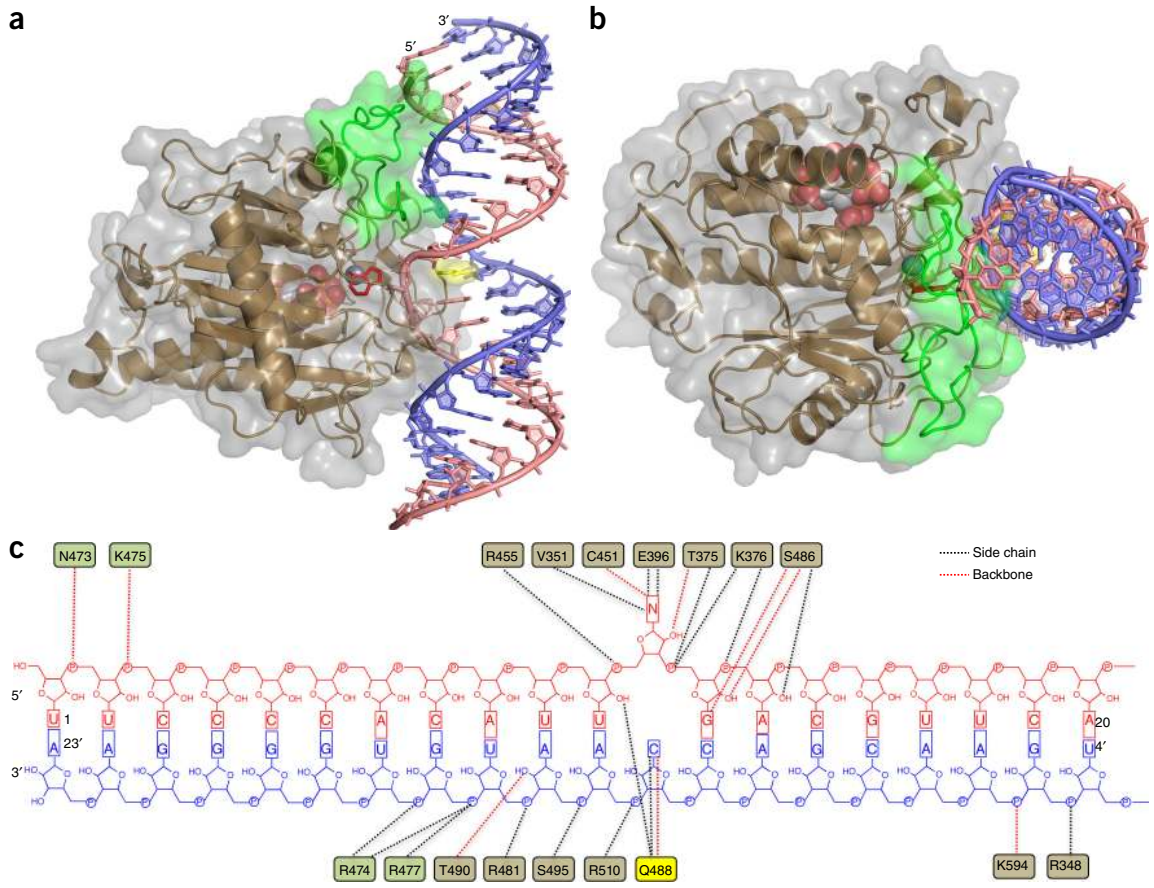


Figure 2 Structure of hADAR2d E488Q bound to the *Bdf2*-C RNA duplex at 2.75-Å resolution. (**a**) View of structure perpendicular to the dsRNA helical axis. Colors correspond to those in **Figure 1a**,**c**. Red, flipped-out base N; gray space-filling sphere, zinc; yellow, Q488; green, previously disordered aa 454–477 loop; space filling, IHP. A transparent surface is shown for the hADAR2d protein. (**b**) View of structure along the dsRNA helical axis. (**c**) Summary of the contacts between hADAR2d E488Q and the *Bdf2*-C RNA duplex.

the flipped nucleotide in the active site by fastening the phosphate backbone flanking the editing site. The R455 side chain ion-pairs with the 5′-phosphodiester of 8-azanebularine while the K376 side chain contacts its 3′-phosphodiester. Last, the side chain of V351 provides a hydrophobic surface for interaction with the nucleobase of the edited nucleotide. RNA binding does not alter IHP binding or the hydrogen-bonding network linking IHP to the active site[21].

ADARs use a unique mechanism to modify duplex RNA

The ADAR2 base-flipping loop, bearing residue 488, approaches the RNA duplex from the minor-groove side at the editing site. The side chain of this amino acid penetrates the helix, where it occupies the space vacated by the flipped-out base and hydrogen-bonds to the complementary-strand orphaned base and to the 2′-hydroxyl of the nucleotide immediately 5′ to the editing site (**Fig. 3b**,**c**). In the four structures reported here, we observed three different combinations of helix-penetrating residue and orphan base (i.e., E488 + U, E488 + C and Q488 + C), and all three combinations show the same side chain and base positions (**Fig. 3b**,**c** and overlay of all three in **Supplementary Fig. 4a**). For instance, in the complex with hADAR2d E488Q and the *Bdf2*-C duplex, the protein recognizes an orphaned C by donating hydrogen bonds from Nε2 to cytosine N3 and from its backbone NH to cytosine O2 (**Fig. 3b**). In the complex of hADAR2d WT and the *Bdf2*-U duplex, we observed a similar interaction, in which the E488 backbone NH hydrogen-bonds to the uracil O2, and the E488 side chain hydrogen-bonds to the uracil N3H (**Fig. 3c**). Interestingly, the E488Q mutant was initially discovered in a screen for highly active ADAR2 mutants, and this residue has been suggested to be involved in base-flipping, given its effect on editing substrates with a fluorescent nucleobase at the editing site[22]. ADARs react preferentially with adenosines in A•C mismatches and A-U pairs over A•A and A•G mismatches[23]. A purine at the orphan base position (in its *anti* conformation) would clash with residue 488, thus explaining the preference for pyrimidines observed here.

The interaction of residue 488 with the orphaned base is reminiscent of an interaction observed for the HhaI DNA methyltransfersase (MTase), a duplex-DNA-modifying enzyme that also uses a base-flipping mechanism to access 2′-dC for methylation[24,25]. For that enzyme, Q237 hydrogen-bonds to an orphaned dG while it fills the void left by the flipped-out dC[25] (**Supplementary Fig. 4b**). In addition, two glycine residues flank Q237, thus allowing the loop to adopt the conformation necessary for penetration into the helix[24]. The flipping loop in ADAR2 (i.e., aa 487–489) also has the helix-penetrating residue flanked by glycines. However, unlike the case of the DNA MTase, which approaches the DNA from the major groove, the ADAR2 loop approaches the duplex from the minor-groove side. Such an approach requires deeper penetration of the intercalating residue to access the hydrogen-bonding sites on the orphaned base, thus necessitating an additional conformational change in the RNA duplex. This change includes shifting of the base pairs immediately

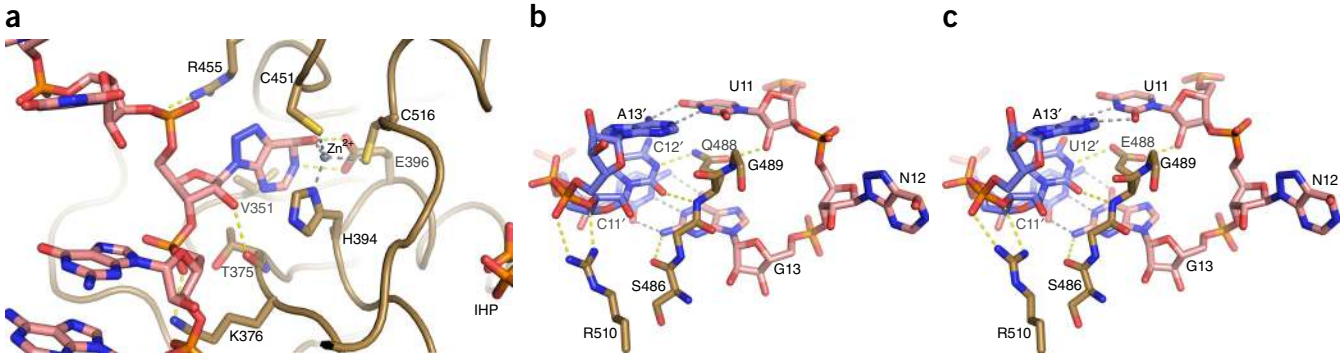


Figure 3 ADAR recognition of the flipped-out and orphaned nucleotides. (**a**) Contacts to the editing-site nucleotide (N) in the active site. Colors correspond to those in **Figures 1** and **2**. (**b**) Orphan-nucleotide recognition in the hADAR2d E488Q–*Bdf2*-C complex. (**c**) Orphan-nucleotide recognition in the hADAR2d WT–*Bdf2*-U complex.

5′ to the editing site toward the helical axis and a widening of the major groove opposite the editing site (**Fig. 4a**,**b** and **Supplementary Video 1**). In the case of the hADAR2d WT–*Bdf2*-U RNA, this shift is accompanied by a shearing of the U11-A13′ base pair with U11 shifted further in the direction of the major groove, thus creating an unusual U-A 'wobble' interaction with adenine N6 and N1, within hydrogen-bonding distance to uracil N3H and O2, respectively (**Fig. 4c** and **Supplementary Fig. 3b**). This type of wobble pair has been observed before and requires either the imino tautomer of adenine or the enol tautomer of uracil[26]. The ADAR-induced distortion in RNA conformation results in a kink in the RNA strand opposite to the editing site (**Fig. 4b**). This kink is stabilized by interactions of the side chains of R510 and S495 with phosphodiesters in the RNA backbone of the unedited strand (**Fig. 4a**). Interestingly, ADAR2's flipping-loop approach from the minor-groove side is similar to that seen with certain DNA-repair glycosylases (for example, UDG[27], HOGG1 (ref. 28), and AAG[29]) that also project intercalating residues from loops bound in the minor groove (**Supplementary Fig. 5a**). However, these enzymes typically bend the DNA duplex at the site of modification, thereby allowing for penetration of intercalating residues and damage recognition[30]. Whereas hADAR2d clearly alters the duplex conformation to gain access to the modification site from the minor groove, it does not bend the RNA duplex (**Figs. 2a**,**b** and **4b**). Furthermore, ADARs do not modify duplex DNA. The DNA B-form helix has groove widths and depths that would prevent productive interactions with ADAR. For instance, ADAR can readily penetrate an A-form helix from the minor-groove side and can place the helix-penetrating residue in the space occupied by the editing-site base (**Supplementary Fig. 6**). However, this residue cannot penetrate the minor groove enough to occupy the base position in a B-form helix (**Supplementary Fig. 6**). Furthermore, DNA lacks the 2′-hydroxyls that are used by ADAR for substrate recognition (**Fig. 2c**). Indeed, in each of the four complexes reported here, the protein contacts at least five ribose 2′-hydroxyl groups (**Fig. 2c** and **Supplementary Fig. 2b–d**). Thus, hADAR2d uses a substrate-recognition and base-flipping mechanism that bears similarities to other known nucleic acid–modifying enzymes but is uniquely suited for reaction with A in the context of duplex RNA.

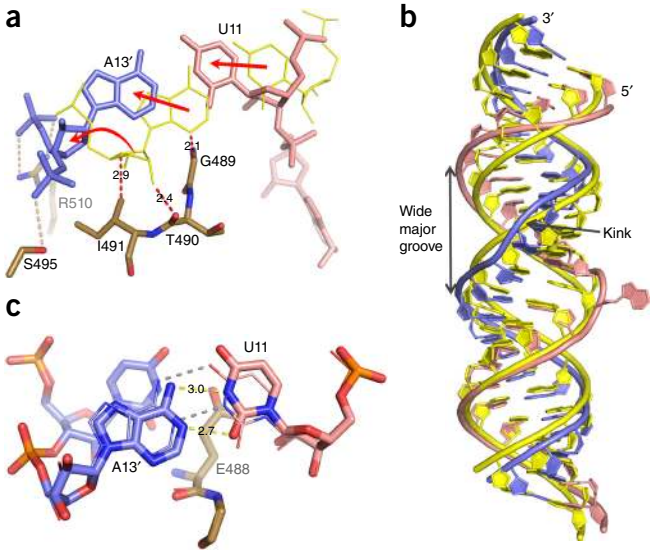


Figure 4 Other ADAR-induced changes in RNA conformation. (**a**) hADAR2d-induced shift in the position of the U11-A13′ base pair from ideal A-form RNA helix (yellow). (**b**) Overlay of *Bdf2* duplex RNA and an idealized A-form duplex of same sequence (yellow), illustrating a kink in the strand and the widening of the major groove opposite the editing site induced by hADAR2d. (**c**) Unusual wobble A13′-U11 interaction in the hADAR2d WT–*Bdf2*-U complex, shown in sticks with hydrogen bonds indicated with yellow dashes and distances shown in Å. The position of this base pair in the hADAR2d E488Q–*Bdf2*-C duplex is shown in wire with hydrogen bonds shown with gray dashes.

Structures explain nearest-neighbor preferences

ADARs have a preference for editing A nucleotides with 5′ nearest-neighbor U (or A) and 3′ nearest-neighbor G[18,31,32]. The ADAR2 flipping loop occupies the minor groove spanning the three base pairs that include the nearest-neighbor nucleotides flanking the edited base (**Fig. 3b**,**c**). As described above, the base pair including the 5′ nearest-neighbor U (U11-A13′ in the *Bdf2* duplex) is shifted from the position that it would occupy in a typical A-form helix to accommodate the loop (**Fig. 4a**). In addition, the minor-groove edge of this pair is juxtaposed with the protein backbone at G489. Modeling a G-C or C-G pair at this position (i.e., 5′-G or 5′-C) suggests that a 2-amino group in the minor groove would clash with the protein at G489 (**Fig. 5a** and **Supplementary Fig. 7c**). Indeed, replacing the U-A pair adjacent to the editing site with a C-G pair in the *GLI1* duplex substrate resulted in an 80% reduction in the rate of hADAR2d-catalyzed deamination (**Fig. 5b**,**c**). To determine whether this effect arises from an increase in local duplex stability from the substitution of C-G for U-A or from the presence of the 2-amino group, we replaced the U-A pair with a U-2-aminopurine (2AP) pair. 2AP is an adenosine analog that forms

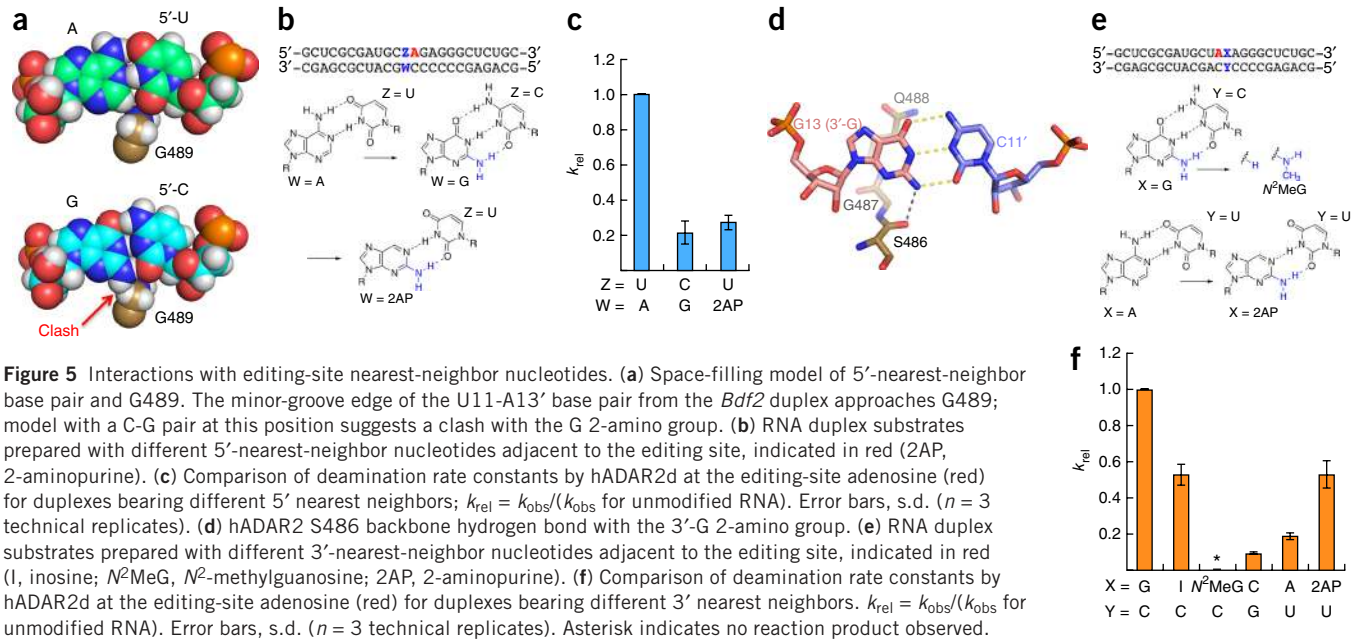


Figure 5 Interactions with editing-site nearest-neighbor nucleotides. (**a**) Space-filling model of 5′-nearest-neighbor base pair and G489. The minor-groove edge of the U11-A13′ base pair from the *Bdf2* duplex approaches G489; model with a C-G pair at this position suggests a clash with the G 2-amino group. (**b**) RNA duplex substrates prepared with different 5′-nearest-neighbor nucleotides adjacent to the editing site, indicated in red (2AP, 2-aminopurine). (**c**) Comparison of deamination rate constants by hADAR2d at the editing-site adenosine (red) for duplexes bearing different 5′ nearest neighbors; $k_{rel} = k_{obs}/(k_{obs}$ for unmodified RNA). Error bars, s.d. ($n = 3$ technical replicates). (**d**) hADAR2 S486 backbone hydrogen bond with the 3′-G 2-amino group. (**e**) RNA duplex substrates prepared with different 3′-nearest-neighbor nucleotides adjacent to the editing site, indicated in red (I, inosine; N^2MeG, N^2-methylguanosine; 2AP, 2-aminopurine). (**f**) Comparison of deamination rate constants by hADAR2d at the editing-site adenosine (red) for duplexes bearing different 3′ nearest neighbors. $k_{rel} = k_{obs}/(k_{obs}$ for unmodified RNA). Error bars, s.d. ($n = 3$ technical replicates). Asterisk indicates no reaction product observed.

a base pair with U, and has similar stability to a U-A pair[33], but places an amino group in the minor groove (**Fig. 5b**). Importantly, this substitution also resulted in an 80% reduction in rate, thus illustrating the detrimental effect of the amino group in the minor groove at this location. These observations suggest that hADAR2's 5′-nearest-neighbor preference for U (or A) is due to a destabilizing clash with the protein backbone at G489 that results from the presence of an amino group in the minor groove at this location for sequences with 5′-nearest-neighbor G or C. However, the observed clash is not severe, and the enzyme would be able to accommodate G or C 5′ nearest neighbors by slight structural perturbations, thus explaining why this sequence preference is not an absolute requirement.

In each of the hADAR2d–RNA structures reported here, the backbone carbonyl oxygen at S486 accepts a hydrogen bond from the 2-amino group of the G on the 3′ side of the edited nucleotide (**Fig. 5d**). G is the only common nucleobase that presents a hydrogen-bond donor in the RNA minor groove, thus suggesting that other nucleotides in this position would reduce editing efficiency. Indeed, mutation of this base to A, C or U, while maintaining base-pairing at this position, reduced the rate of deamination by hADAR2d in *GLI1*-mRNA model substrates (**Supplementary Fig. 7a**,**b**). To test the importance of the amino group on the 3′-G in the hADAR2d reaction, we prepared RNA duplex substrates with purine analogs on the 3′ side of the edited A (**Fig. 5e**). We tested a G analog lacking the 2-amino group (I) and one that blocks access to this amino group (N^2-methylguanosine (N^2MeG). In addition, we compared a 3′-A to a 3′-2AP, because 2AP might form the hydrogen-bonding interaction observed with S486. We found that the substrate with a 3′-N^2MeG was unreactive to hADAR2d-catalyzed deamination, thus confirming the importance of the observed close approach by the protein to the 3′-G 2-amino group (**Fig. 5f**). In addition, the substrate with a 3′-I displayed a decreased deamination rate compared with the substrate with a 3′-G, thus suggesting that the observed hydrogen bond to the 2-amino group contributes to the 3′-nearest-neighbor selectivity (**Fig. 5f**). This conclusion was further supported by the observation that deamination in the substrate with a 3′-2AP was faster than that in the substrate with a 3′-A (**Fig. 5f**).

RNA-binding loops of the ADAR catalytic domain

The structures reported here identify RNA-binding loops of the ADAR catalytic domain and suggest roles for several amino acids not previously known to be important for editing, either substrate binding or catalysis (**Fig. 6**). The side chain of R510 ion-pairs with the 3′-phosphodiester of the orphaned nucleotide (**Fig. 3a**,**c**). This residue is conserved in ADAR2s and ADAR1s but is glutamine in the editing-inactive ADAR3s (**Supplementary Table 1**). Mutation of hADAR2d at this site to either glutamine (R510Q) or alanine (R510A) reduced the measured deamination rate constant by approximately an order of magnitude (**Fig. 6c**). In addition, the contact point near the 5′ end of the unedited strand involved G593, K594, and R348, residues completely conserved in the family of ADAR2s (**Fig. 2c**, **Supplementary Table 1**). Mutation of any of these residues to alanine (G593A, K594A, or R348A) substantially reduced editing activity (**Fig. 6c**). In addition, mutation of G593 to glutamate (G593E) resulted in a rate reduction of nearly two orders of magnitude, a result consistent with the proximity of this residue to the negatively charged phosphodiester backbone of the RNA (**Fig. 6c**).

RNA binding leads to an ordering of the 454–477 loop, which was disordered in the RNA-free hADAR2d structure[21] (**Fig. 2a**,**b** and **Supplementary Video 2**). This loop binds the RNA duplex contacting the minor groove near the editing site and inserting into the adjacent major groove (**Fig. 6e**). This loop sequence is conserved in ADAR2s but is different in the family of ADAR1s (**Fig. 6d**). The substantial difference in sequence between the ADARs in this RNA-binding loop suggests that differences in editing-site selectivity between the two ADARs arise at least partially from differences in how this loop binds RNA substrates.

DISCUSSION

Base-flipping is a well-characterized mechanism by which nucleic acid–modifying enzymes gain access to sites of reaction that are otherwise buried in base-paired structures[34]. DNA methylases, DNA-repair glycosylases, and RNA-loop-modifying enzymes that flip a nucleotide out of a base pair are known[24,35–37]. However, none of the structurally characterized base-flipping enzymes access their reactive

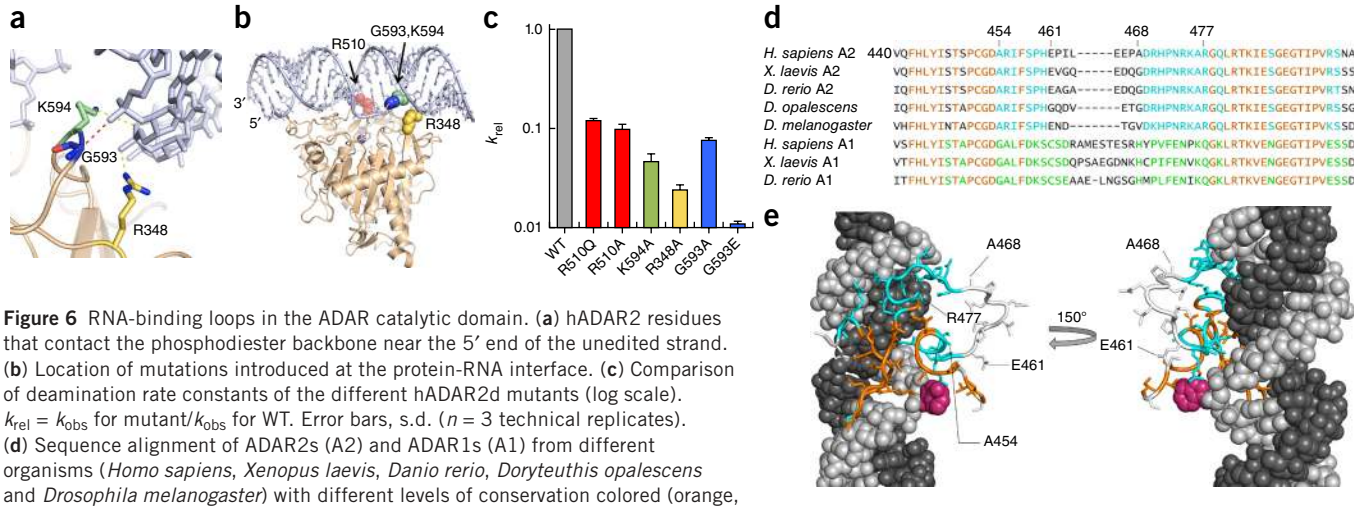


Figure 6 RNA-binding loops in the ADAR catalytic domain. (**a**) hADAR2 residues that contact the phosphodiester backbone near the 5′ end of the unedited strand. (**b**) Location of mutations introduced at the protein-RNA interface. (**c**) Comparison of deamination rate constants of the different hADAR2d mutants (log scale). $k_{rel} = k_{obs}$ for mutant/k_{obs} for WT. Error bars, s.d. ($n = 3$ technical replicates). (**d**) Sequence alignment of ADAR2s (A2) and ADAR1s (A1) from different organisms (*Homo sapiens*, *Xenopus laevis*, *Danio rerio*, *Doryteuthis opalescens* and *Drosophila melanogaster*) with different levels of conservation colored (orange, conserved in all ADAR1s and ADAR2s; cyan, conserved in ADAR2s; green, conserved in ADAR1s. (**e**) Interaction of the ADAR-specific RNA-binding loop near the 5′ end of the edited strand. Colors are as in **d**; white, not conserved; pink, flipped-out base.

sites from within a normal base-paired RNA duplex. We are aware of one other protein-induced nucleotide-flipping mechanism from an RNA duplex region[38]. Bacterial initiation factor 1 (IF1) binds the 30S ribosomal subunit at helix 44 of 16S RNA, with A1492 and A1493 flipped out of the helix and bound in protein pockets (**Supplementary Fig. 5b**). However, these nucleotides are located in a highly distorted and dynamic duplex region containing several mismatches and are predisposed to undergo this conformational change[39]. Thus, this system is not illustrative of base-flipping from a normal duplex and does not involve an enzyme that must carry out a chemical reaction on the flipped-out nucleotide. Other RNA-modification enzymes are known that flip nucleotides out of loops, even from base pairs in loop regions (e.g., pseudoU synthetase[35], tRNA transglycosylase[40], and restrictocin bound to the sarcin-ricin loop of 28S rRNA[37]) (**Supplementary Fig. 5b**). Because the modification sites are not flanked on both sides by normal duplex, these enzymes do not experience the same limits in approach to the substrate that ADARs experience. The requirement that ADARs must induce flipping from a normal duplex has implications on its preference for adenosines flanked by certain base pairs, a phenomenon that was not previously well understood.

In our structures, the flipped-out 8-azanebularine is hydrated, mimicking the tetrahedral intermediate predicted for deamination of adenosine (**Figs. 1b** and **3a** and **Supplementary Fig. 3a**). Our use of 8-azanebularine, with its high propensity to form a covalent hydrate[41], allowed us to capture a true mimic of the tetrahedral intermediate and to reveal the interactions between the deaminase active site and the reactive nucleotide. In addition, 8-azanebularine adopted a 2′ *endo* sugar pucker with its 2′-hydroxyl hydrogen-bonded to the protein backbone carbonyl at T375. The 2′ *endo* conformation appears to facilitate access of the nucleobase to the zinc-bound water for nucleophilic attack at C6.

Several other base-flipping enzymes stabilize the altered nucleic acid conformation by intercalation of an amino acid side chain into the space vacated by the flipped-out base[27,30,35]. For hADAR2, E488 serves this role. In the two structures with wild-type hADAR2, the E488 residue and orphan base are in nearly identical positions (overlay in **Supplementary Fig. 4a**). Thus, the E488 side chain directly contacts each orphan base, probably by accepting a hydrogen bond from uracil N3H or by donating a hydrogen bond to cytidine N3. The latter interaction requires E488 to be protonated. The pK_a of E488 in the ADAR–RNA complex has not been measured, but proximity to hydrogen-bond acceptors, such as the N3 of C, and insertion between stacked nucleobases, would undoubtedly elevate this value and might lead to a substantial fraction in the protonated state at physiologically relevant pH. Because the glutamine side chain is fully protonated under physiologically relevant conditions, a rate enhancement for the E488Q mutant would be expected if the reaction requires E488 protonation.

The interactions of hADAR2d with base pairs adjacent to the editing-site adenosine explain the known 5′ and 3′-nearest-neighbor preferences (**Fig. 5**). Although these studies indicate the ADAR2 catalytic domain makes an important contact to the 3′-nearest-neighbor G, Stefl *et al.* have suggested the 3′-G preference arises from dsRBD binding selectivity for ADAR2 (ref. 42). These authors have reported a model for ADAR2's dsRBDs bound to an editing substrate, based on NMR data from the isolated dsRBDs (lacking the deaminase domain) and short RNA fragments derived from the *GluR-B* (official symbol *Gria2*) R/G-site RNA[42]. They have described an interaction wherein the 3′-G 2-amino group hydrogen-bonds to the backbone carbonyl of S258 found in the β1–β2 loop of ADAR2's dsRBD II. It is not possible for the S486–3′-G interaction that we describe here and the S258–3′-G interaction reported by Stefl *et al.* to exist in the same complex, because both involve protein loops bound in the RNA minor groove at the same location. Because our structures capture the edited nucleotide in the conformation required to access the active site, the interactions observed here are highly likely to occur during the deamination reaction at the editing site. However, because dsRBDs are known to bind promiscuously with duplex RNA, it is possible that the S258–3′-G interaction found in a complex lacking the deaminase domain may not be relevant to catalysis at the editing site[43]. It is also possible that the ADAR dsRBD and catalytic domain binding are sequential, such that release of the dsRBD from the RNA takes place before catalytic-domain engagement and base-flipping.

AGS and DSH are human diseases caused by mutations in hADAR1, and several disease-associated mutations have been found in the deaminase domain[11,12]. Given the conservation in RNA-binding surface and active site residues, we expect the hADAR1 catalytic domain to bind RNA with an orientation of the helix similar to that found in our hADAR2d–RNA structures. When the locations of the AGS-associated mutations are mapped onto the hADAR2d–RNA complex, two mutations involve residues in proximity to the RNA (<4 Å)

(**Supplementary Fig. 8a**). G487 of hADAR2 is located in the flipping loop near the RNA (**Fig. 3b**). The sequence in this loop is highly conserved among ADARs and corresponds to G1007 in hADAR1 (**Supplementary Table 2**). An arginine at this position would preclude close approach of the flipping loop to the RNA, thus preventing E1008 insertion and base-flipping into the active site (**Supplementary Fig. 8b**). This scenario is consistent with the observation that the G1007R mutation in hADAR1 inhibits RNA-editing activity[11,44]. In addition, K376 forms salt bridges with both the 5′- and 3′-phosphodiesters of the guanosine on the 3′ side of the editing site (**Figs. 2c** and **3a**). The corresponding residue in hADAR1 (R892) may form similar contacts, and the R892H mutation would probably alter this interaction.

In summary, the structures described here establish human ADAR2 as a base-flipping enzyme that uses a unique mechanism well suited for modifying duplex RNA. In addition, this work provides a basis for understanding the role of the ADAR catalytic domain in determining the editing-site selectivity and additional structural context to evaluate the effects of ADAR mutations associated with human disease.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes PDB 5ED1 (hADAR2d E488Q–*Bdf2*-C RNA complex), PDB 5ED2 (hADAR2d E488Q–*GLI1* RNA complex), PDB 5HP2 (hADAR2d WT–*Bdf2*-U RNA complex), and PDB 5HP3 (hADAR2d WT–*Bdf2*-C RNA complex).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors acknowledge funding from the US National Institutes of Health (NIH) grant R01GM061115 (P.A.B.). A.I.S. was supported by NIH training grant T32 GM007377. C. Palumbo is acknowledged for technical assistance. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the US Department of Energy, Office of Biological and Environmental Research, and by the NIH, US National Institute of General Medical Sciences (including P41GM103393). Part of this work is also based on research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the NIH (P41 GM103403). The Pilatus 6M detector on the 24-ID-C beamline is funded by an NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a DOE Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

AUTHOR CONTRIBUTIONS

J.M.T., M.M.M., A.I.S., and Y.Z. purified protein. K.J.P. and J.M.T. designed and purified RNA for crystallography and characterized protein-RNA binding. M.M.M. and A.I.S. conducted crystallization trials. M.M.M. and A.J.F. collected diffraction data and solved and refined the crystal structures. J.M.T., Y.Z., and J.H. measured enzyme reaction rates. K.T. synthesized 8-azanebularane phosphoramidite. J.M.T. and A.I.S. conducted mutagenesis. J.M.T., M.M.M., P.A.B. and A.J.F. analyzed the structures. P.A.B. wrote the initial manuscript draft. J.M.T., M.M.M., P.A.B., and A.J.F. edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Grosjean, H. *Fine-Tuning of RNA Functions by Modification and Editing* (Springer, 2005).
2. Bass, B.L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).
3. Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **79**, 321–349 (2010).
4. Wang, Q. *et al.* ADAR1 regulates ARHGAP26 gene expression through RNA editing by disrupting miR-30b-3p and miR-573 binding. *RNA* **19**, 1525–1536 (2013).
5. Rueter, S.M., Dawson, T.R. & Emeson, R.B. Regulation of alternative splicing by RNA editing. *Nature* **399**, 75–80 (1999).
6. Yeo, J., Goodman, R.A., Schirle, N.T., David, S.S. & Beal, P.A. RNA editing changes the lesion specificity for the DNA repair enzyme NEIL1. *Proc. Natl. Acad. Sci. USA* **107**, 20715–20719 (2010).
7. Bass, B.L. *et al.* A standardized nomenclature for adenosine deaminases that act on RNA. *RNA* **3**, 947–949 (1997).
8. Maas, S., Kawahara, Y., Tamburro, K.M. & Nishikura, K. A-to-I RNA editing and human disease. *RNA Biol.* **3**, 1–9 (2006).
9. Slotkin, W. & Nishikura, K. Adenosine-to-inosine RNA editing and human disease. *Genome Med.* **5**, 105 (2013).
10. Morabito, M.V. *et al.* Mice with altered serotonin 2C receptor RNA editing display characteristics of Prader-Willi syndrome. *Neurobiol. Dis.* **39**, 169–180 (2010).
11. Rice, G.I. *et al.* Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nat. Genet.* **44**, 1243–1248 (2012).
12. Miyamura, Y. *et al.* Mutations of the RNA-specific adenosine deaminase gene (DSRAD) are involved in dyschromatosis symmetrica hereditaria. *Am. J. Hum. Genet.* **73**, 693–699 (2003).
13. Zhang, X.J. *et al.* Seven novel mutations of the ADAR gene in Chinese families and sporadic patients with dyschromatosis symmetrica hereditaria (DSH). *Hum. Mutat.* **23**, 629–630 (2004).
14. Chen, L. *et al.* Recoding RNA editing of *AZIN1* predisposes to hepatocellular carcinoma. *Nat. Med.* **19**, 209–216 (2013).
15. Gallo, A. RNA editing enters the limelight in cancer. *Nat. Med.* **19**, 130–131 (2013).
16. Shimokawa, T. *et al.* RNA editing of the GLI1 transcription factor modulates the output of Hedgehog signaling. *RNA Biol.* **10**, 321–333 (2013).
17. Goodman, R.A., Macbeth, M.R. & Beal, P.A. ADAR proteins: structure and catalytic mechanism. *Curr. Top. Microbiol. Immunol.* **353**, 1–33 (2012).
18. Li, J.B. *et al.* Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213 (2009).
19. Haudenschild, B.L. *et al.* A transition state analogue for an RNA-editing reaction. *J. Am. Chem. Soc.* **126**, 11213–11219 (2004).
20. Phelps, K.J. *et al.* Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. *Nucleic Acids Res.* **43**, 1123–1132 (2015).
21. Macbeth, M.R. *et al.* Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **309**, 1534–1539 (2005).
22. Kuttan, A. & Bass, B.L. Mechanistic insights into editing-site specificity of ADARs. *Proc. Natl. Acad. Sci. USA* **109**, E3295–E3304 (2012).
23. Wong, S.K., Sato, S. & Lazinski, D.W. Substrate recognition by ADAR1 and ADAR2. *RNA* **7**, 846–858 (2001).
24. Klimasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357–369 (1994).
25. Daujotyte, D. *et al.* HhaI DNA methyltransferase uses the protruding Gln237 for active flipping of its target cytosine. *Structure* **12**, 1047–1055 (2004).
26. Thiyagarajan, S., Rajan, S.S. & Gautham, N. Cobalt hexammine induced tautomeric shift in Z-DNA: the structure of d(CGCGCA)*d(TGCGCG) in two crystal forms. *Nucleic Acids Res.* **32**, 5945–5953 (2004).
27. Slupphaug, G. *et al.* A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87–92 (1996).
28. Bruner, S.D., Norman, D.P. & Verdine, G.L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859–866 (2000).
29. Lau, A.Y., Schärer, O.D., Samson, L., Verdine, G.L. & Ellenberger, T. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249–258 (1998).
30. Brooks, S.C., Adhikary, S., Rubinson, E.H. & Eichman, B.F. Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta* **1834**, 247–271 (2013).
31. Eifler, T., Pokharel, S. & Beal, P.A. RNA-Seq analysis identifies a novel set of editing substrates for human ADAR2 present in *Saccharomyces cerevisiae*. *Biochemistry* **52**, 7857–7869 (2013).
32. Eggington, J.M., Greene, T. & Bass, B.L. Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.* **2**, 319 (2011).
33. Peacock, H., Maydanovych, O. & Beal, P.A. N^2-Modified 2-aminopurine ribonucleosides as minor-groove-modulating adenosine replacements in duplex RNA. *Org. Lett.* **12**, 1044–1047 (2010).
34. Roberts, R.J. & Cheng, X. Base flipping. *Annu. Rev. Biochem.* **67**, 181–198 (1998).
35. Hoang, C. & Ferré-D'Amaré, A.R. Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107**, 929–939 (2001).
36. Parikh, S.S. *et al.* Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **17**, 5214–5226 (1998).

37. Yang, X., Gérczei, T., Glover, L.T. & Correll, C.C. Crystal structures of restrictocin-inhibitor complexes with implications for RNA recognition and base flipping. *Nat. Struct. Biol.* **8**, 968–973 (2001).
38. Carter, A.P. *et al.* Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* **291**, 498–501 (2001).
39. Réblová, K. *et al.* Structure, dynamics, and elasticity of free 16s rRNA helix 44 studied by molecular dynamics simulations. *Biopolymers* **82**, 504–520 (2006).
40. Xie, W., Liu, X. & Huang, R.H. Chemical trapping and crystal structure of a catalytic tRNA guanine transglycosylase covalent intermediate. *Nat. Struct. Biol.* **10**, 781–788 (2003).
41. Erion, M.D. & Reddy, M.R. Calculation of relative hydration free energy differences for heteroaromatic compounds: use in the design of adenosine deaminase and cytidine deaminase inhibitors. *J. Am. Chem. Soc.* **120**, 3295–3304 (1998).
42. Stefl, R. *et al.* The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **143**, 225–237 (2010).
43. Fierro-Monti, I. & Mathews, M.B. Proteins binding to duplexed RNA: one motif, multiple functions. *Trends Biochem. Sci.* **25**, 241–246 (2000).
44. Mannion, N.M. *et al.* The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.* **9**, 1482–1494 (2014).

ONLINE METHODS

Unless otherwise stated, reagents were purchased from Fisher Scientific, Sigma-Aldrich, or Life Technologies. T4 polynucleotide kinase, T4 DNA ligase, molecular-biology-grade bovine serum albumin (BSA), and RNase inhibitor were purchased from New England BioLabs. [γ-^{32}P]ATP was purchased from PerkinElmer Life Sciences. The avian myeloblastosis virus (AMV) reverse transcriptase, dNTP mix and RQ1 RNase-free DNase were purchased from Promega. Pfu Ultra II was purchased from Stratagene. Dpn 1 was purchased from Invitrogen. The QuikChange XL II mutagenesis kit was purchased from Agilent Technologies. RNA oligonucleotides were synthesized at the University of Utah DNA/Peptide Core Facility or purchased from GE Healthcare Dharmacon or Sigma Aldrich. DNA oligonucleotides were purchased from Integrated DNA Technologies. Storage phosphorimaging plates from Molecular Dynamics were imaged with a Molecular Dynamics 9400 Typhoon phosphorimager. Data were analyzed with Molecular Dynamics ImageQuant 5.2 software. Electrospray Ionization (ESI) mass spectrometry of oligonucleotide samples was performed at the Campus Mass Spectrometry Facilities, UC Davis. Oligonucleotide masses were determined with Mongo Oligo Mass Calculator v2.06.

Expression and purification of hADAR2 deaminase domain (hADAR2d) for crystallography. Protein expression and purification were carried out by modifying a previously reported protocol[45]. In brief, *S. cerevisiae* BCY123 cells were transformed with a pSc-ADAR construct encoding either hADAR2d WT or hADAR2d E488Q (corresponding to the deaminase domain; residues 299–701). Cells were streaked on yeast minimal medium minus uracil (CM–ura) plates. A single colony was used to inoculate a 15-mL CM–ura starter culture. After cultures were shaken at 300 r.p.m. and 30 °C overnight, 10 mL of starter culture was used to inoculate each liter of yeast growth medium. After 24 h, cells were induced with the addition of 110 mL of sterile 30% galactose per liter, and protein was expressed for 5 h. Cells were collected by centrifugation and stored at −80 °C. Cells were lysed 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 750 mM NaCl with a microfluidizer, and cell lysate was clarified by centrifugation (39,000*g* for 25 min). Lysate was passed over a 5-mL Ni–NTA column, which was then washed in three steps with 20–50 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 the target protein were pooled and further purified on a 2-mL GE Healthcare Lifesciences Hi-Trap Heparin HP column in the absence of BME. The His_{10} fusion protein was cleaved with an optimized ratio of 1 mg of TEV protease per 1 mg of protein. Cleavage was carried out for 1–2 h before the product was passed over another Ni–NTA column at 0.5 mL/min. The flow through and wash were collected; dialyzed against 20 mM Tris, pH 8.0, 200 mM NaCl, 5% glycerol, and 1 mM BME; and concentrated to just under 1 mL for gel filtration on a GE Healthcare HiLoad 16/600 Superdex 200 PG column. Fractions containing purified protein were pooled and concentrated to 5–7 mg/mL for crystallography trials.

Purification of RNAs for crystallography. The 8-azanebularine (N) phosphoramidite was synthesized as previously described[19], and RNAs were synthesized as previously described[46]. Single-stranded RNAs (sequences in **Supplementary Table 2**) were purified by denaturing PAGE and visualized with UV shadowing. Bands were excised from the gel, crushed, and soaked overnight at 4 °C in 0.5 M NH_4OAc, 0.1% sodium dodecyl sulfate (SDS), and 0.1 mM EDTA. Polyacrylamide fragments were removed with a 0.2-µm filter, and this was followed by desalting on a C18 Sep-Pak column. The RNA solutions were lyophilized to dryness, resuspended in nuclease-free water, quantified by absorbance at 260 nm, and stored at −70 °C. Oligonucleotide mass was confirmed by electrospray ionization mass spectrometry. Unmodified RNA stands were purchased from GE Healthcare Dharmacon and purified as described above. Duplex RNA was hybridized in a 1:1 ratio by heating to 95 °C for 5 min and slow cooling to 30 °C.

Crystallization of the hADAR2d–RNA complex. Crystals of the hADAR2d E488Q–*Bdf2*-C RNA complex were grown at room temperature by the sitting-drop vapor-diffusion method. A solution of 0.5-µL volume containing 4.5 mg/mL protein and 70 µM of *Bdf2*-C 23-mer RNA (1:0.7 ADAR2/RNA molar ratio) was mixed with 0.5 µL of 0.1 M MES-NaOH, pH 6.5, 9% (w/v) PEG 3350, 13% glycerol, and 0.015 M NAD, which was added to improve crystal growth. Crystals took several weeks to grow. A single cube-shaped crystal of approximately 120 µm in size was soaked briefly in a solution of mother liquor plus 30% glycerol before flash cooling in liquid nitrogen. Data were collected via fine-phi slicing with 0.2° oscillations on beamline 24-ID-C at the Advanced Photon Source at the Argonne National Laboratories in Chicago. To obtain crystals of the hADAR2d WT–*Bdf2*-C RNA, an identical procedure was used as above; however, the crystallization conditions had slightly different concentrations of reagents (10% PEG 3350, 15% glycerol, 0.1 M MES-NaOH, pH 6.5, and no NAD). For the hADAR2d WT–*Bdf2*-U construct, hanging-drop vapor diffusion with 200 nL of a mixture containing 4.5 mg/mL protein and 70 µM of *Bdf2*-U (1:0.7 molar ratio) and 200 nL of mother liquor (0.1 M ammonium acetate, 0.1 M Bis-Tris, pH 5.5, and 17% PEG 10,000) yielded several crystals with a morphology similar to that described above. All wild-type crystals were soaked briefly in a solution of mother liquor plus 30% glycerol before flash cooling in liquid nitrogen. Data were collected via fine-phi slicing with 0.2° oscillations on beamline 12-2 at the Stanford Synchrotron Radiation Lightsource. Crystals of the hADAR2d E488Q–*GLI1* RNA complex were grown through hanging-drop vapor diffusion. A solution of volume 200 nL containing 4.5 mg/mL protein and 100 µM of *GLI1* 23-mer RNA (1:1 ADAR2/RNA molar ratio) was mixed with 200 nL of 0.1 M MES-NaOH, pH 6.5, and 12% PEG 20,000. At room temperature, a single diamond-shaped crystal approximately 150 µm long and 50 µm wide was observed approximately a week later. This crystal was soaked briefly in a solution of mother liquor plus 30% glycerol before flash cooling in liquid nitrogen. Data were collected on beamline 12-2 at the Stanford Synchrotron Radiation Lightsource with the fine-phi splicing described above.

Processing and refinement of crystallographic data. Data for the E488Q *Bdf2*-C–bound and *GLI1*-bound structures were processed with XDS[47] and scaled with Aimless (CCP4). Diffraction data for hADAR2d wild-type structures were processed with XDS and scaled with XSCALE[47]. The RNA-free hADAR2d crystal structure (PDB 1ZY7)[21] was used as a model for molecular replacement with PHENIX[48]. The structures were refined with PHENIX[49] including TLS parameters and zinc coordination restraints. Ideal zinc-ligand distances were determined with average distances found for similar coordination models in the PDB database. **Table 1** shows the statistics in data processing and model refinement. The asymmetric unit for *GLI1*-bound hADAR2d E488Q included two complexes of protein–RNA. In each of these complexes, the first 17 residues of the deaminase domain (residues 299–316) as well as a C-terminal proline (P701) were disordered and were therefore not included in the model. However, although the RNA-free ADAR2 structure (PDB 1ZY7) lacked electron density for residues 457–475, we observed density for the backbone atoms of these residues. These residues were initially modeled as polyalanine. After several rounds of refinement, electron density revealed the location of some side chains. Residues whose basic side chains interact with the RNA backbone were clearly defined in the final density map. Although some non-RNA-binding side chains showed only weak density, the backbone density was strong. As observed in the original hADAR2d RNA-free structure, IHP was buried in the enzyme core[21]. The asymmetric units for *Bdf2*-bound ADARs contained one ADAR2d–RNA complex (protein chain A) and one RNA-free ADAR2d monomer (chain D). The N terminus of the *Bdf2*-bound structures included more residues than did the *GLI1*-bound structure, beginning at P305 in chain A and T304 for chain D in the mutant structure, and beginning at R307 in chain A and T304 or P305 in chain D in the wild-type structures. The first few residues (in structures in which the specified residues were modeled) had weak side chain density, including residues 305 and 307 in chain A, and residues 304–307 in chain D, and were modeled in the structure as alanine. The last residue of E488Q–*Bdf2*-C, P701, had very weak electron density for both protein subunits in the asymmetric unit. Unlike the E488Q–*GLI1* structure, electron density was defined better in the originally disordered loop (residues 457–475) for most residues in the *Bdf2*-bound structures. With the exception of E466 in the wild-type structures, we were able to model-build in main chain and side chain atoms for all residues of this loop in the ADAR subunit in complex with the *Bdf2* RNA duplexes. In the RNA-free subunit (chain D) of E488Q–*Bdf2*-C, a crystal contact stabilized this flexible loop so that we were able to model in the backbone for residues 457–475, but residues 465–475 were modeled as alanine because of poorly defined side chain density. An identical crystal contact was observed in the wild-type structures. In the WT–*Bdf2*-C complex, the density

for residues 467–470 was strong enough for side chains to be included in these structures; however, side chain density was not strong for residues 465, 466, 471, 473–475, and 477. Therefore, these side chains were not included in the model. In WT–*Bdf2*-U, the density for side chains 465, 466, 470, 471, and 473–475 was too weak to model. IHP was observed in all ADAR deaminase domains in the asymmetric unit. To model the hydrated 8-azanebularine nucleotide in all RNAs, a CCP4 dictionary file for A was modified to replace the 6-amino group with hydrogen, to change atom 8 to nitrogen and to include an additional hydroxyl group off carbon 6. Additionally, an energy-minimization-calculated idealized structure was used to determine ideal bond angles and distances for the modified base of the hydrated 8-azanebularine (D. Tantillo, unpublished data). The refinement-restraint dictionary file was edited to match these parameters.

Expression and purification of hADAR2d for *in vitro* deamination kinetics. Histidine-tagged human ADAR2 deaminase domain (hADAR2d) and hADAR2d mutant proteins were expressed in *S. cerevisiae* strain BCY123 and purified as described above with the following modifications. Cell lysate was purified with a 0.45-μm filter after centrifugation and loaded three times through 5-mL Ni–NTA Superflow (Qiagen) at 3 mL/min. Washes of 50 ml with buffers 1, 2 and 3 at 4 mL/min were followed by elution with a 35-mL gradient from buffer 3 to elution buffer. Selected elution fractions from the Ni–NTA column were pooled and loaded at 0.5 mL/min on a 1-mL HiTrap Heparin HP column from GE. The column was washed with 10 mL of heparin 1 buffer at 0.5 mL/min and eluted with a 12 mL gradient from heparin 1 to heparin 2 buffer. Selected elution fractions from the heparin column were pooled and concentrated to <300 μL in a 10,000 MWCO Amicon Ultra 4 centrifugal filter at 6,500 RCF and 4 °C. TEV protease cleavage and gel-filtration steps were omitted. Buffer exchange was accomplished via three rounds of concentration to <300 μL and subsequent addition of 3 mL of storage buffer. After final concentration, protein concentrations were determined with BSA standards, as visualized by SYPRO Orange staining on SDS–PAGE gels, and the purified proteins were stored at −70 °C.

Site-directed mutagenesis. Mutagenesis of the hADAR2 catalytic domain was carried out via PCR site-directed mutagenesis with the primers listed in **Supplementary Table 2**. All primers were purchased from IDT and were PAGE purified as described above but were desalted by phenol-chloroform extraction, ethanol precipitation and 70% ethanol wash instead of C18 Sep-Pack. Sequences for mutant plasmids were confirmed by Sanger sequencing.

Preparation of h*GLI1* splint-ligated RNA. Oligonucleotides were purified as described above but were desalted by phenol-chloroform extraction, ethanol precipitation and 70% ethanol wash. The 3′ *GLI1* top-strand 12-mer RNAs were radiolabeled with γ-^{32}P at the 5′ end with T4 PNK, as described previously[50]. Labeled 3′ *GLI1* top-strand 12-mer RNAs were ligated as previously described to produce internally labeled RNA. The splint ligation products were PAGE purified as described above. Labeled RNAs were hybridized with the complementary *GLI1* bottom-strand 24-mer RNA (Y was chosen according to the identity of X; **Fig. 3b**) in 10 mM Tris-HCl, 0.1 mM EDTA, pH 7.5, and 100 mM NaCl. RNA sequences are shown in **Supplementary Table 2**.

***In vitro* deamination kinetics with internally ^{32}P-labeled substrates.** Deamination kinetics of analog-containing RNAs was determined as previously described[45] but with the following modifications. The final reaction volume was 10 μL. The final enzyme concentration was 300 nM. The final RNA concentration was 10 nM. The final reaction conditions were 16 mM Tris HCl, pH 7.4, 3.3% glycerol, 1.6 mM EDTA, 0.003% NP-40, 60 mM KCl, 7.1 mM NaCl, 0.5 mM DTT, 160 units/mL RNasin, and 1 μg/mL yeast tRNA. Reactions were quenched by addition of 190 μL 95 °C nuclease-free water followed by incubation at 95 °C for 5 min or by 10 μL 0.5% SDS at 95 °C followed by incubation at 95 °C for 5 min. Each experiment was carried out in triplicate, and the rate constants reported in the text are average values ± s.d. Sequences of RNAs used to prepare internally labeled substrates are shown in **Supplementary Table 2**. For comparison of hADAR2-D mutants, deamination kinetics was determined as described above with the following modifications. The final reaction conditions were 300 nM hADAR2d, 10 nM RNA, 16 mM Tris HCl, pH 7.4, 3.6% glycerol, 1.6 mM EDTA, 0.003% NP-40, 60 mM KCl, 8.6 mM NaCl, 0.5 mM DTT, 160 units/mL RNasin, and 1 μg/mL yeast tRNA.

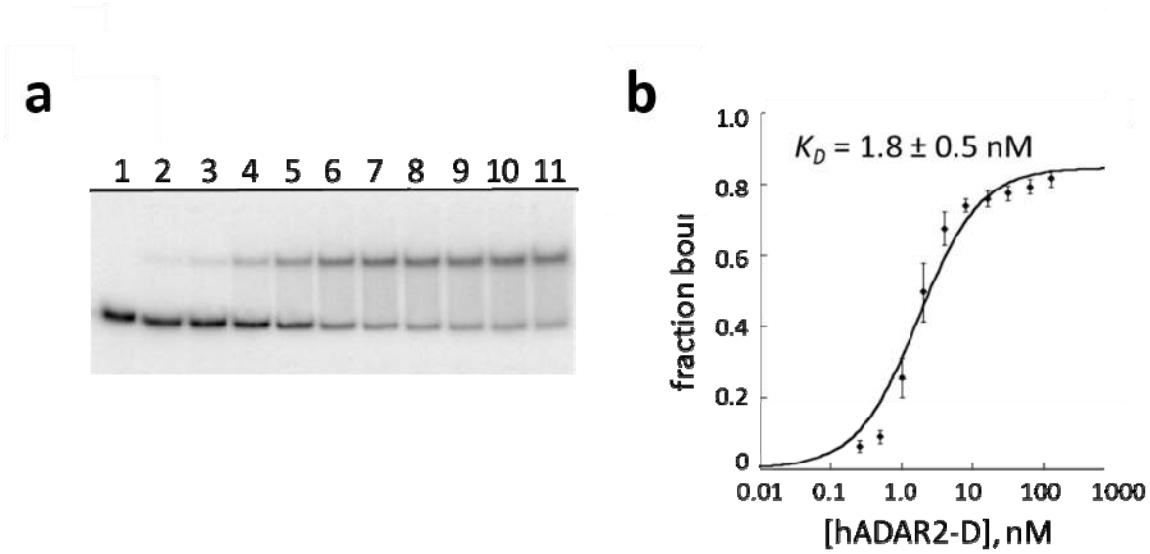
EMSA analysis of radioactively labeled RNA. Duplex ^{32}P-labeled RNAs containing 8-azanebularine were prepared as previously described[20]. Samples containing 0.25 nM RNA and different concentrations of hADAR2d E488Q (128, 64, 32, 16, 8, 4, 2, 1, 0.5, 0.25 and 0 nM) were equilibrated in 20 mM Tris-HCl, pH 7, 6% glycerol, 0.5 mM DTT, 60 mM KCl, 20 mM NaCl, 0.1 mM BME, 1.5 mM EDTA, 0.003% NP-40, 160 units/ml RNasin, 100 μg/ml BSA, and 1.0 μg/ml yeast tRNA for 30 min at 30 °C. Assays and data analysis were carried out as previously described[20]. RNA sequences are shown in **Supplementary Table 2**.

***In vitro* transcription of RNA.** A truncation of h*GLI1* mRNA incorporating 81 nucleotides upstream and 65 nucleotides downstream of the edited site was transcribed and purified as previously described[31]. The 3′-nearest-neighbor mutants of h*GLI1* RNA were generated by site-directed mutagenesis to generate G-to-A, G-to-C and G-to-U nearest-neighbor mutants. A second site −32 bases from the editing site was mutated to maintain the original secondary structure of the RNA. Primers used for mutagenesis are shown in **Supplementary Table 2**.

Deamination kinetics of transcribed h*GLI1* RNAs. Deamination kinetics of transcribed RNAs was determined as previously described[31], but with the following modifications. The final reaction volume was 20 μL, the final enzyme concentration was 10 nM, and the final RNA concentration was 2 nM. The final reaction conditions were 17 mM Tris HCl, pH 7.4, 5.0% glycerol, 1.6 mM EDTA, 0.003% NP-40, 60 mM KCl, 15.6 mM NaCl, 0.5 mM DTT, 160 units/mL RNasin, and 1 μg/mL yeast tRNA. Reactions were quenched by addition of 10 μL 0.5% SDS at 95 °C and incubation at 95 °C for 5 min. cDNA was generated from RNA via RT–PCR, Sanger sequenced and quantified with SeqScanner 2 software from Applied Biosystems. The k_{obs} (min^{-1}) of each assay was calculated as described previously[31].

45. Macbeth, M.R. & Bass, B.L. Large-scale overexpression and purification of ADARs from *Saccharomyces cerevisiae* for biophysical and biochemical studies. *Methods Enzymol.* **424**, 319–331 (2007).
46. Pokharel, S. *et al.* Matching active site and substrate structures for an RNA editing reaction. *J. Am. Chem. Soc.* **131**, 11882–11891 (2009).
47. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
48. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
49. Afonine, P.V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
50. Phelps, K.J., Ibarra-Soza, J.M., Tran, K., Fisher, A.J. & Beal, P.A. Click modification of RNA at adenosine: structure and reactivity of 7-ethynyl- and 7-triazolyl-8-aza-7-deazaadenosine in RNA. *ACS Chem. Biol.* **9**, 1780–1787 (2014).

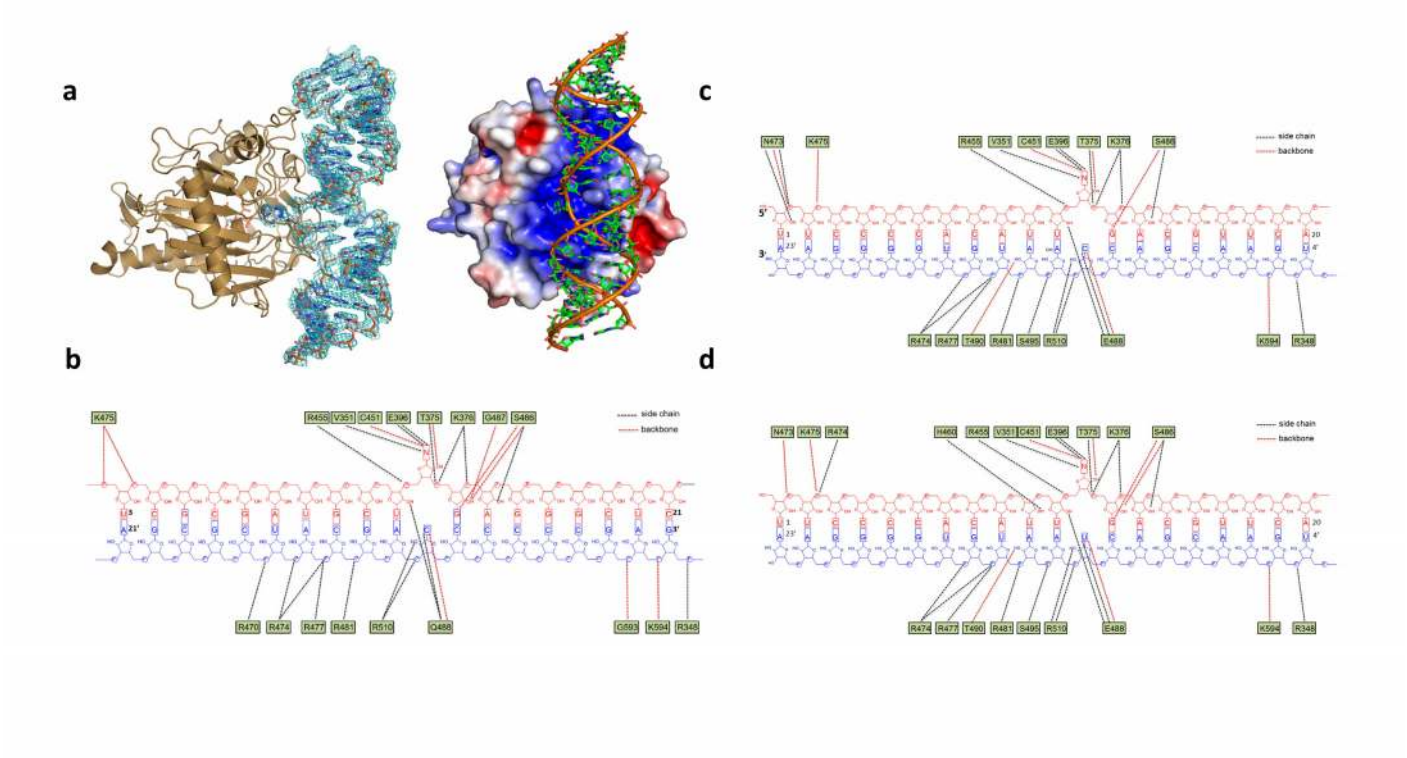
doi:10.1038/nsmb.3203



Supplementary Figure 1

8-Azanebularine-containing duplex RNAs form a tight and specific complex with hADAR2d.

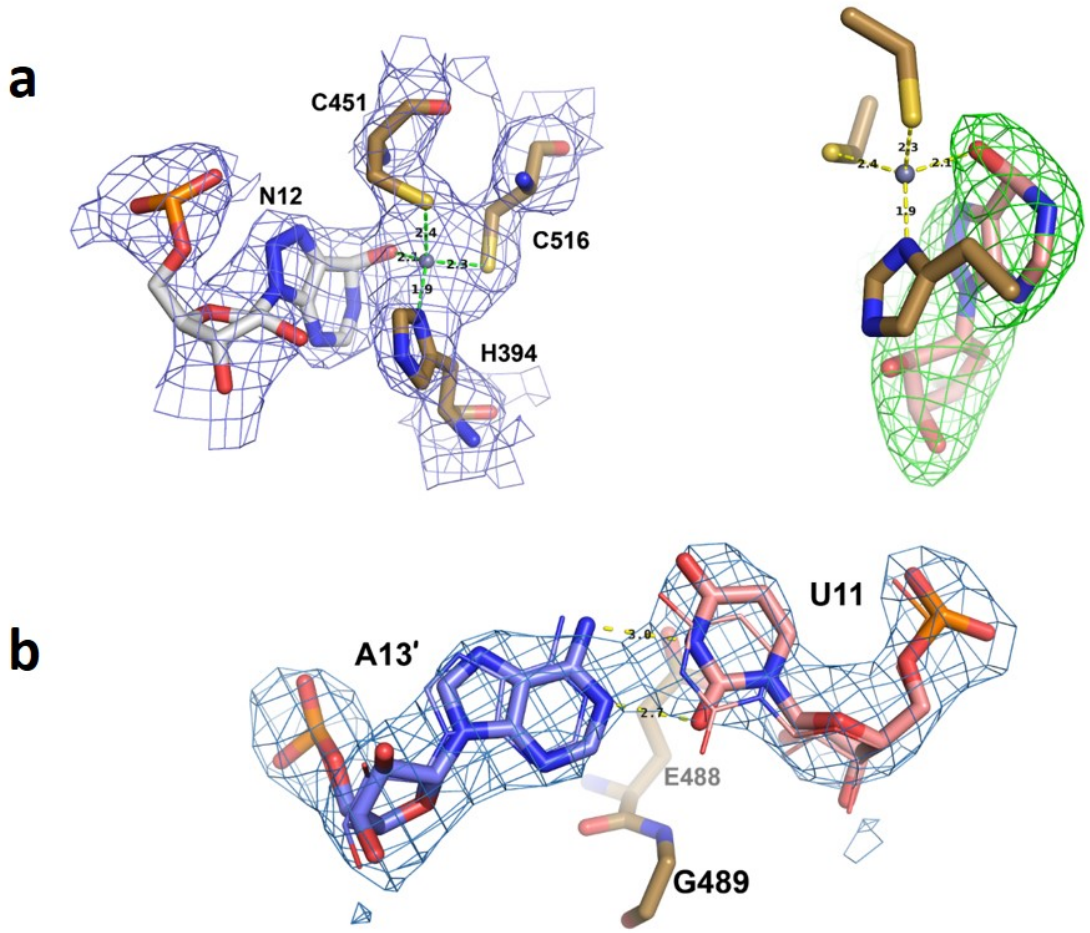
(a) Representative electrophoretic mobility shift assay (EMSA) gel displaying tight and specific binding of hADARd E488Q mutant and N-containing hGLI1 24mer duplex. Lane 1: no protein added; Lanes 2-11: 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, and 128 nM hADAR2d E488Q. **(b)** Fitted plot of fraction RNA bound vs. hADAR2d E488Q concentration.



Supplementary Figure 2

Overall structure of the hADAR2d–RNA complex.

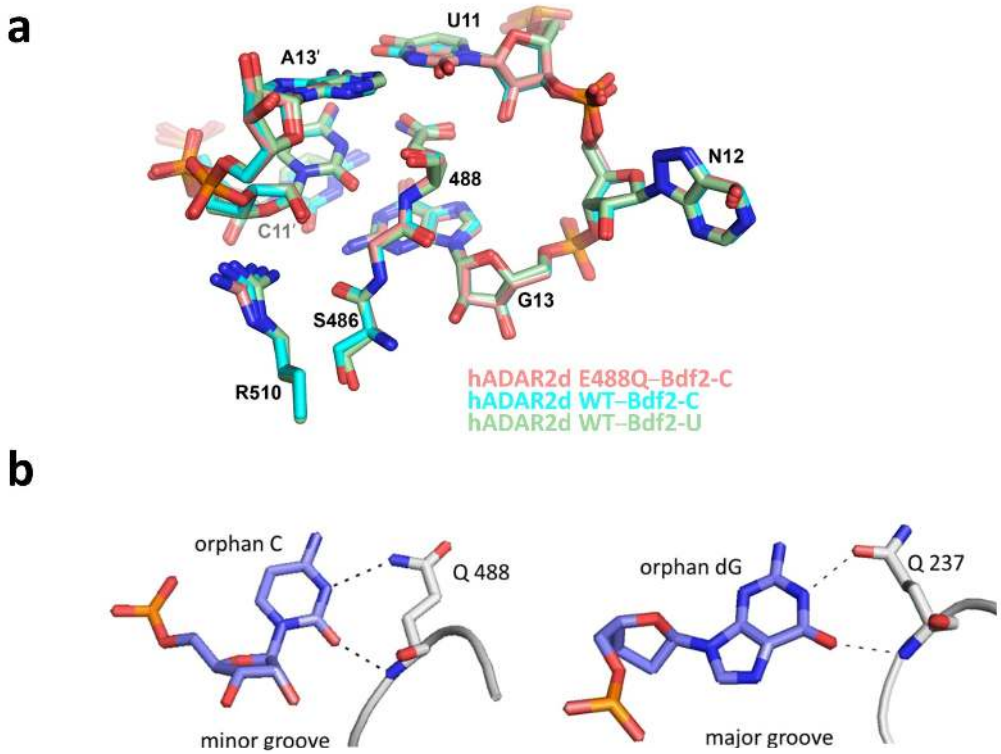
(a) **(Left)** A simulated-annealed composite omit electron density map calculated in PHENIX (Afonine, P.V. et al. Acta Crystallogr D Biol Crystallogr **71**, 646-666, 2015) is shown in blue mesh. The map is contoured at 1σ at 2.75Å resolution. The inositol hexakisphosphate is shown in white-colored-carbon sticks, and the active site zinc atom as a gray sphere near the flipped out base. **(Right) View of hADAR2d E488Q-Bdf2 complex showing protein surface electrostatic potential.** The electrostatic potential was calculated use the Adaptive Poisson-Boltzmann Solver (APBS) plugin in PyMol (Baker, N.A et al. Proc Natl Acad Sci USA **98**, 10037-10041, 2001). Electrostatic potential was calculated using the default values, and displayed on the protein surface with blue and red representing the positive and negative potentials respectively. **(b) Summary of protein-RNA contacts observed in the Gli1-hADAR2d E488Q complex. (c) Summary of protein-RNA contacts observed in the hADAR2d WT-Bdf2-C complex. (d) Summary of protein-RNA contacts observed in the hADAR2d WT-Bdf2-U complex**



Supplementary Figure 3

Close-up views of the active site and 5´-nearest-neighbor nucleotides with electron density map.

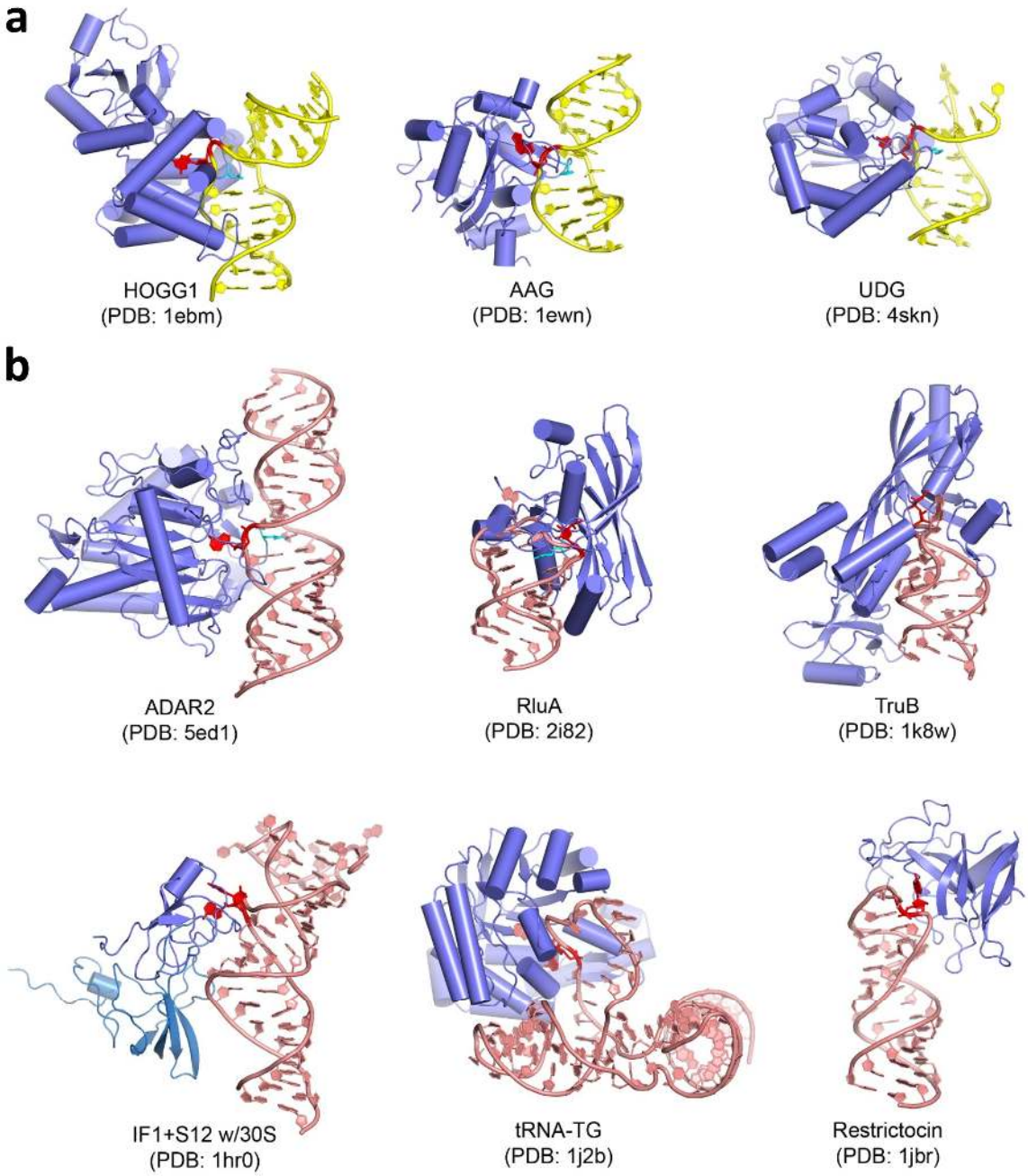
(a) (Left) Close-up view of the flipped out base with electron density. The flipped out 8-azanebularine nucleotide (hADAR2d E488Q + Bdf2-C) is shown as sticks with white-colored carbon atoms. Residues that ligate the active site zinc are shown with brown-colored carbon atoms. Simulated annealed composite omit map shown in blue mesh contoured at 1σ. Numbers next to yellow dashed lines represent ligation distances in Å. **(Right) Simulated annealed omit map of the flipped out base.** The 8-azanebularine nucleotide was omitted from the structure and a simulated annealed omit map calculated in PHENIX (Adams, P.D et al. Acta Crystallogr D Biol Crystallogr **66**, 213-221, 2010). Positive electron density contoured at 3σ is displayed in green. This clearly shows the presence of oxygen O6 (extending from carbon 6) ligating to the active site zinc atom. This definitively reveals the hydrated intermediate is being trapped in the crystal. **(b) Close up view of 5' UA pair from (hADARd WT + Bdf2-U) shown with composite omit-map electron density at 1σ.** The geometry of the 5' nearest neighbor base pair from hADAR2d WT+ Bdf2-U (sticks) and electron density overlaid with structure of hADAR2d WT+ Bdf2-C (lines) suggests a non-canonical base pair between the A and U.



Supplementary Figure 4

Intercalating residue and adjacent RNA.

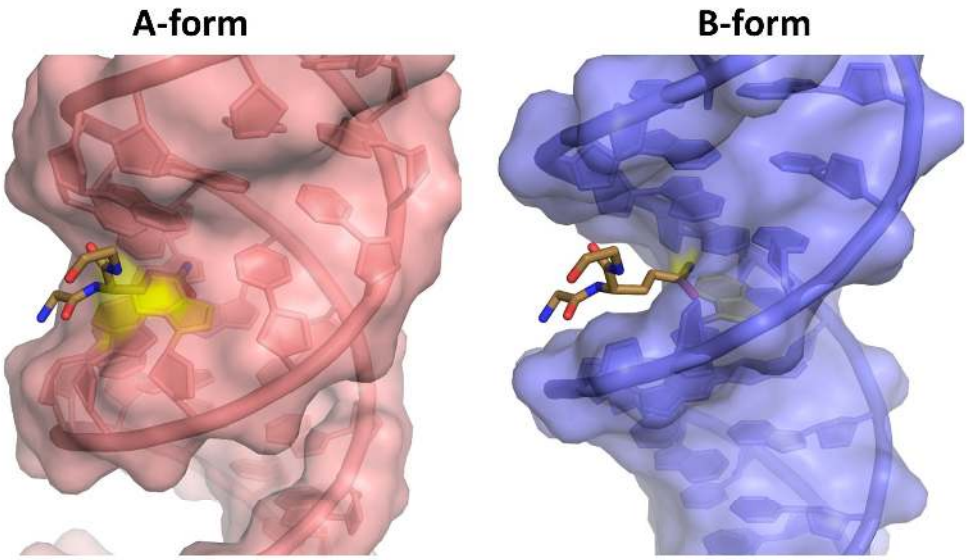
(a) Overlay of intercalating residue and RNA from three structures. Overlay of structures of wild type and E488Q hADAR2d with Bdf2-C RNA and wild type hADAR2d with Bdf2-U RNA show high degree of similarity in both RNA and enzyme conformations **(b) Comparison of ADAR2 and HhaI Mtase orphan base recognition.** Shown is the interaction between intercalating residue side chain and orphan nucleotide in ADAR2 **(Left)** and HhaI cytidine methyltransferase **(Right)**. For ADAR2 the side chain approaches the nucleotide from the minor groove of the A form helix while for HhaI Mtase the approach is made from the major groove of the B form DNA (Klimasauskas, S. et al. Cell, **76**, 357-369 1994).



Supplementary Figure 5

Comparison of protein–nucleic acid complexes with flipped nucleotides.

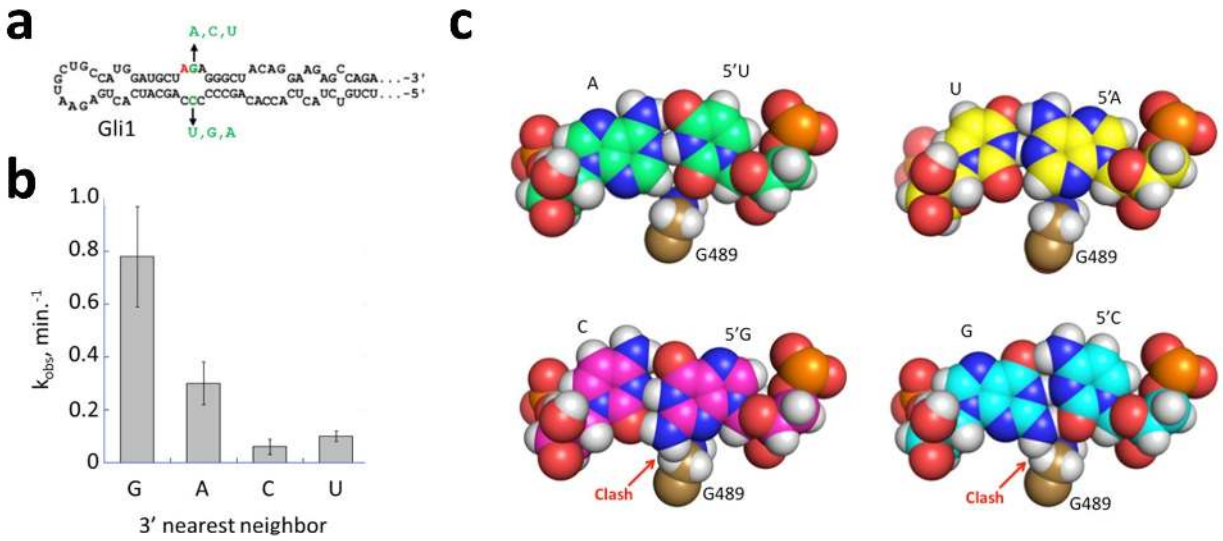
(a) Structures of DNA modifying enzymes with base flipping dependent mechanisms approaching from the minor groove. Protein in blue, DNA in yellow, which is bent, flipped nucleotide in red. **(b) Structures of protein–RNA complexes with flipped nucleotides.** Protein in blue, RN in pink, flipped nucleotide in red.



Supplementary Figure 6

Access of residue 488 into the minor grooves of A-form and B-form helices.

Overlay of intercalating residue 488 with idealized A form RNA **(Left)** shows that the geometry of the minor groove in an A form helix allows the side chain to occupy the position of the adenosine base (yellow). Overlay of residue 488 with idealized B form DNA **(Right)** shows that the greater depth of the minor groove in a B form helix prevents the side chain from intercalating into the space occupied by an adenosine base (yellow). Idealized A-RNA and B-DNA of Bdf2 sequence were superimposed onto the RNA duplex observed in the hADAR2d complexed with RNA (as shown in Figure 4B).

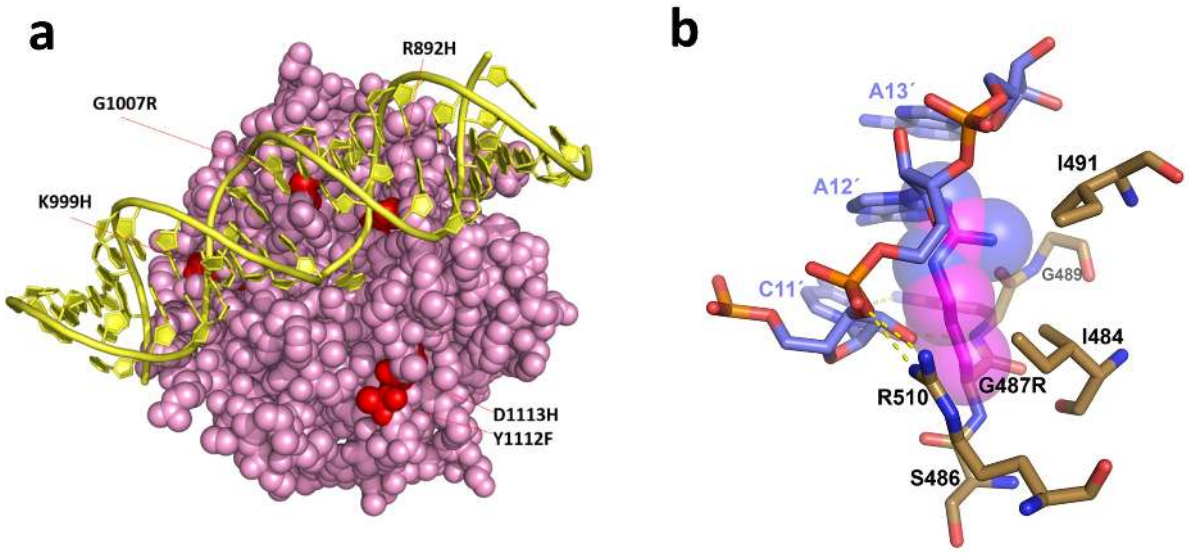


Supplementary Figure 7

Nearest-neighbor effects of hADAR2d.

(a) Predicted secondary structure of GLI1 mRNA editing site showing 3' nearest neighbor and complementary mutations. Edited A is highlighted in red. **(b) Comparison of deamination rates in 3' nearest neighbor mutants in hGLI1 mRNA.** Observed in vitro deamination rate constants for deamination of hGLI1 mRNA by hADAR2d and 3' nearest neighbor mutants.

(c)Space filling models of 5' nearest neighbor base pair and hADAR2d G489. Minor groove edge of U11-A13' base pair is in close proximity to the protein backbone at G489 in the Bdf2-hADAR2d complex (top left). U11 is located on the 5' side of the editing site (i.e. 5' U). This site appears to accommodate a 5' A (top right) but when either a C-G (bottom left) or a G-C (bottom right) base pair is modeled into this site, a clash is apparent between the G 2-amino group and the alpha carbon of G489.



Supplementary Figure 8

hADAR1 residues associated with AGS, mapped to analogous positions in hADAR2d.

(a) Space filling model of hADAR2d E488Q + Bdf2-C RNA highlighting the locations corresponding to residues in hADAR1 where mutations cause Aicardi–Goutières syndrome. Modeling positions of surface residues of hADAR1 with mutations previously associated with Aicardi–Goutières syndrome (Rice, G.I. et al. Nat Genetics, **44**, 1243-1248, 2012) provides insight into the basis for the mutations' effect on editing. hADAR2 residues corresponding to hADAR1 R892 and G1007 (K376 and G487 respectively) are found at the protein/RNA binding interface. K999 maps to ADAR2's 5' major groove binding loop (Q479). While Y1112 and D1113 are not directly involved in the protein/RNA interface, they appear to be in position to stabilize an RNA binding loop. E588 in hADAR2 (corresponding to D1113 in hADAR1) is involved in an intramolecular salt bridge with highly conserved R349 located adjacent to RNA contact residue R348. **(b) hADAR2 G487 is located at the protein-RNA interface.** Mutation to arginine (corresponding to the G1007R mutation in hADAR1 associated with the human diseases AGS and DHS) would cause a severe clash with the RNA, especially with orphan C11' and A12'. The G487R mutation is shown with transparent CPK spheres with magenta-colored carbons.

Supplementary table 1, Multiple sequence alignment of hADAR1, hADAR2, hADAR3 and homologs. Multiple sequence alignment of catalytic domains of ADAR proteins. Alignment was performed using Clustal Omega software available from the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/clustalo/). Homologs of ADAR1, 2 and 3 are denoted in parenthesis.

```
(2)H.sapiens          AAIFNLHLDQTPSRQ---------------PI-----------PSEGLQLHLPQVLADAV     327
(2)D.opalescens       AQIFNMELF-PPGLQ---------------PV-----------PKDGEET-SSKVLADII     292
(2)R.norvegicus       ATVFNLHLDQTPSRQ---------------PV-----------LSEGLQLHLPQVLADAV     327
(2)M.musculus         ATVFNLHLDQTPSRQ---------------PV-----------LSEGLQLHLPQVLADAV     327
(2)X.laevis           ASLFNMHLDQTPSRQ---------------PI-----------PSEGLQLHLPQVLADAV     326
(2)O.hannah           ASLFNMQLDQTPSHQ---------------PI-----------PSEGLQLHLPQVLADAV     332
(2)G.gallus           ASLFNMQLDQTPSRQ---------------PI-----------PSEGLQLHLPQVLADAV     328
(2)C.anna             ASLFNMQLDQTPSRQ---------------PI-----------PSEGLQLHLPQVLADAV     326
(2)D.rerio            SGLFNMQLDQTPSRQ---------------PI-----------PREGLQLHLPQVLADAV     346
(2)T.rubripes         SALFNMQLDQAPSRQ---------------PI-----------PREGLQLHLPQVLADAV     319
D.melanogaster        ASLCNISYSPMVVPQK----------NVPLPI-----------DDKSSSMELPQIHADTI     299
A.mellifera           AKLRNVHSSSFCIPLP----------VRVLPNF----PSS---GHWQDQMSLPQMLADKI     255
(1)O.hannah           ------------------------------------------------------------     435
(1)T.rubripes         RVLIGEAERAARTGELI----------------------------PAELPVTGSTLHDQI     801
(1)G.gallus           RVLIGETEKAERMGGMSVA--------------------------ELPVRGSTLHDQM     469
(1)C.anna             RVLIRETEHVEGMEIPEVKPFPWKLVLFGSPCLSSAFQAGRKTPLRSTLRSFGSTFHDQL     666
(1)D.rerio            RVLIGEAEKAARTGELM----------------------------PELPVSGSTIHDQI     962
(1)X.laevis           RVLIGEAEKAAREGDIM----------------------------AELPVSGSTFHDQI     744
(1)H.sapiens          RVLIGENEKAERMGFTEVTPVTGASLRRTMLLL-----SRSPEAQPKTLPLTGSTFHDQI     844
(1)M.musculus         RVLIGESEKAEQLGFAE----------------------------LPLSGSTFHDQI     767
(1)R.norvegicus       RVLIGESEKAEQLGFAE----------------------------LPLTGSTFHDQI     249
(3)D.labrax           QALFNIRTAPEGRTG---------------LK-----------ASRKSCPHLPQDFADSI     349
(3)C.anna             QALFNIRLP-----S---------------HI-----------PSRSKCHHLPQDFADSI     238
(3)H.sapien           QELFDIQMP-----G---------------HA-----------PGRARRTPMPQEFADSI     365
(3)M.musculus         QALFDIRLP-----G---------------HI-----------PSRSKSNLLPQDFADSV     371
(3)R.norvegicus       QALFDIRLP-----G---------------HI-----------PSRSKSNLLPQDFADSV     372


(2)H.sapiens          SRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTDVKDAKVISVSTGTKCINGEYMSDRG     387
(2)D.opalescens       AKVVHEKFSELTDNYTSQYSRRKVLAGIVMTHGPDGQGSQVIGIGTGTKCINGEYMNDRG     352
(2)R.norvegicus       SRLVLGKFSDLTDNFSSPHARRKVLSGVVMTTGTDVKDAKVISVSTGTKCINGEYMSDRG     387
(2)M.musculus         SRLVLGKFSDLTDNFSSPHARRKVLSGVVMTTGTDVKDAKVISVSTGTKCINGEYMSDRG     387
(2)X.laevis           ARLVVDKFSDLTDNFTSPHARRKVLAGVVMTTGTDVKDAQVICVSTGTKCINGEYMSDRG     386
(2)O.hannah           ARLVVEKFSDLTDNFTSPHARRKVLAGVVMTTGTDVKDAQVISVSTGTKCINGEYMSDRG     392
(2)G.gallus           ARLVVDKFSDLTENFTSPHARRKVLAGIVMTTGTDVKDALVISVSTGTKCINGEYMSDRG     388
(2)C.anna             ARLVVDKFSDLTENFTSPHARRKVLAGVVMTTGTDVKDALVISVSTGTKCINGEYMSDRG     386
(2)D.rerio            SRLVVDKFSELTDNFSSPHARRKVLAGVVMTTGKDVMDAQVICVTTGTKCINGEYMSDRG     406
(2)T.rubripes         SHLVVDKFSELTDNFTSPHARRKVLAGVVKTTGTDVKEAEVICVSSGTKCINGEYMSDRG     379
D.melanogaster        GRLVLEKFMEVIKG-QEAYSRRKVLAGIVMTENMNFCEAKVISVSTGTKCVSGEHMSVNG     358
A.mellifera           GKMVNQKFSELIQS-KPQHARRKVLAGIVQTKG---SDAELICVTTGTKCVSGEHLSVSG     311
(1)O.hannah           -----RPHPAQPAGAQDXLLGRKILAAIIMRRGQEGL-GVVVSIGTGNRCVKGEELSLKG     489
(1)T.rubripes         AMLSHQRFNALTTRIQHSLLGRKILATIIMRKG-DGL-GTVVSLGTGNRCVKGEELSLKG     859
(1)G.gallus           AMLSHQRFNVLTACIQHSLLGRKILAAIIMRKGSQDL-GVVVSIGTGNRCVKGEELSLKG     528
(1)C.anna             AMLSHQRFNSLTARIQHSLLGRKILAAIVMQKGKEGL-GVVVSIGTGNRCVKGEELSLKG     725
(1)D.rerio            AMLSHQRFNALTARIQHSLLGRKILATIVMRNGTDSL-GKVVSLGTGNRCVKGEELSLRG     1021
(1)X.laevis           AMLSHQKFNSLTARIQNSLLGRKILAAIIMKKSSDDL-GTVVSIGTGNRCVKGEELSLRG     803
(1)H.sapiens          AMLSHRCFNTLTNSFQPSLLGRKILAAIIMKKDSEDM-GVVVSLGTGNRCVKGDSLSLKG     903
(1)M.musculus         AMLSHRCFNALTNSFQPSLLGRKILAAIIMKRDPEDM-GVVVSLGTGNRCVKGDSLSLKG     826
(1)R.norvegicus       AMLSHRCFNALTNSFQPSLLGRKILAAIIMKRDPEDM-GVVVSLGTGNRCVKGDSLSLKG     308
(3)D.labrax           FHLVREKYRSLVGGCSPAHARHKSLAGIVMTRGLDLRHAQVVALSTGTKCINGEYLSDQG     409
(3)C.anna             YQMVAQKFQELTDNFTSMHARHKTLAGIVMTKGLDIRQAQVIVLSSGTKCINGEYINDQG     298
(3)H.sapien           SQLVTQKFREVTTDLTPMHARHKALAGIVMTKGLDARQAQVVALSSGTKCISGEHLSDQG     425
(3)M.musculus         SQLVTQKFRELTVGLTSVYARHKTLAGIVMTKGLDTKQAQVIVLSSGTKCISGEHISDQG     431
(3)R.norvegicus       SQLVTQKFRELTVGLTSVYARHKTLAGIVMTKGLDTKQAQVIVLSSGTKCISGEHISDQG     432
                             .            :* *: ::         . :: : :*.:*:.*: :.  *
```

```
(2)H.sapiens         LALNDCHAEIISRRSLLRFLYTQLELYLNNKDD-QKESIFQKSE--RGGFRLKENVQFHL     444
(2)D.opalescens      FAVNDCHAEIIARRCFLRFIYDQLELHLSDNAEVRGQSIFELRD--AGGYKLKPNIQFHL     410
(2)R.norvegicus      LALNDCHAEIISRRSLLRFLYAQLELYLNNKED-QKKSIFQKSE--RGGFRLKDTVQFHL     444
(2)M.musculus        LALNDCHAEIISRRSLLRFLYAQLELYLNNKED-QKKSIFQKSE--RGGFRLKDTVQFHL     444
(2)X.laevis          LALNDCHAEIISRRSLVRFLYTQLELFLSTKED-QQKSIFMKSE--RGGYKLKDNVQFHL     443
(2)O.hannah          LALNDCHAEIIARRCLLKFLYTQLELHLSNKED-QQKSIFIKSE--QGGFKLDENVQFHL     449
(2)G.gallus          LALNDCHAEIISRRCLLKFLYTQLELYLSNKDD-QEKSIFIKSE--RGGFKLKENVQFHL     445
(2)C.anna            LALNDCHAEIVSRRCLLKFLYTQLELYLSNKDD-QQKSIFIRSE--RGGFKLKENVQFHL     443
(2)D.rerio           LALNDCHAEIIARRSLIRYLYNQLEYFLSDSTEEHEKSIFRWCS--EHGYRLKDDIQFHL     464
(2)T.rubripes        LALNDCHAEIVARRSLIRYLYSQLEHFLSNHEEEHHKSMFTRCE-KKQGFRLKENVQFHL     438
D.melanogaster       AVLNDSHAEIVSRRCLLKYLYAQLDLQCNQATA--YQSIFVRNTDGQYPYKLKSGVHFHL     416
A.mellifera          GALNDCHAEVVARRCLCEYLYKQLELHTEDRAA--ES-IL---EPAKKGFKLKQGIQFHL     365
(1)O.hannah          ETVNDCHAEIISRRGFLRFLYNELMAFDPL---ATEESIFQLAE--DGKLKIKDDVSFHL     544
(1)T.rubripes        DTVNDCHAEIISRRGFVRFVYSEILKYQDG----NDDCIFEPAE--NNKLQVKPDTTFHL     913
(1)G.gallus          ETVNDCHAEIISRRGFVRFLYSELMKYDPCKPS-AEESIFEPAG--GRRLKIKSSITFHL     585
(1)C.anna            ETVNDCHAEIISRRGFVRFLYSELLKYDPSNPSSAEDSIFEPAG--GNRLKIKSSVTFHL     783
(1)D.rerio           DTVNDCHAEIISRRGFIRFLYSELMKHWES---PGDETIFELAG--DGKLKIKSDITFHL    1076
(1)X.laevis          ETVNDCHAEIVSRRGFIRFLYSQLMKYNPD---VPGDSIFEEAD--GDLLRVRPGVTFHL     858
(1)H.sapiens         ETVNDCHAEIISRRGFIRFLYSELMKYNSQ---TAKDSIFEPAK-GGEKLQIKKTVSFHL     959
(1)M.musculus        ETVNDCHAEIISRRGFIRFLYSELMKYNHH---TAKNSIFELAR-GGEKLQIKKTVSFHL     882
(1)R.norvegicus      ETVNDCHAEIISRRGFIRFLYSELMKYNHH---TAKNSIFELAR-GGEKLQIKKTVSFHL     364
(3)D.labrax          QVVNDCHAEVTARRALLRFLYSQLELFLSKRPEDWEESIFVRHK--ECGYRLRDNVHFHM     467
(3)C.anna            LAVNDCHAEIVARRAFVHFLYSQLELHLSKRREDWERSIFVRLK--EGGYRLRENILFHL     356
(3)H.sapien          LVVNDCHAEVVARRAFLHFLYTQLELHLSKRREDSERSIFVRLK--EGGYRLRENILFHL     483
(3)M.musculus        LVVNDCHAEIVARRAFLHFLYSQLELHLSKHQEDPERSIFIRLK--EGGYRLRENILFHL     489
(3)R.norvegicus      LVVNDCHAEIVARRAFLHFLYTQLELHLSKHQEDPERSIFIRVK--EGGYRLRENILFHL     490
                      .:**.***: :** : .::* ::              ::          ::     **:

(2)H.sapiens         YISTSPCGDARIFSPHEPI-----LEEPADRHPNRKARGQLRTKIESGEGTIPVRSNASI     499
(2)D.opalescens      YISTAPCGDARIFSPHGQD------VETGDRHPNRKARGQLRTKIESGEGTIPVRSSGFI     464
(2)R.norvegicus      YISTSPCGDARIFSPHEPV-----LEEPADRHPNRKARGQLRTKIESGEGTIPVRSNASI     499
(2)M.musculus        YISTSPCGDARIFSPHEPV-----LEEPADRHPNRKARGQLRTKIESGEGTIPVRSNASI     499
(2)X.laevis          YISTSPCGDARIFSPHEVG-----QEDQGDRHPNRKARGQLRTKIESGEGTIPVRSSSTI     498
(2)O.hannah          YISTSPCGDARIFSPHEAA-----QEDQGDRHPNRKARGQLRTKIESGEGTIPVRSTTTI     504
(2)G.gallus          YISTSPCGDARIFSPHEAA-----QEDQGDRHPNRKARGQLRTKIESGEGTIPVRSTTTI     500
(2)C.anna            YISTSPCGDARIFSPHEAA-----QEDQGDRHPNRKARGQLRTKIESGEGTIPVRSTTTI     498
(2)D.rerio           YISTSPCGDARIFSPHEAG-----AEDQGDRHPNRKARGQLRTKIESGEGTIPVRTSNTI     519
(2)T.rubripes        YISTSPCGDARIFSPHEAG-----VEDQGDRHPNRKARGQLRTKIESGEGTIPVRSSNTI     493
D.melanogaster       YINTAPCGDARIFSPHEND-------TGVDKHPNRKARGQLRTKIESGEGTIPVKSSDGI     469
A.mellifera          YINTAPCGDARIFSPHEEN-------ESVDKHPNRRARGQLRTKIESGEGTIPVKSSEGI     418
(1)O.hannah          YISTAPCGDGALFDKSCSEQANTAGEDQHQPLFENPKQGKLRTKVENGEGTIPVESSDIV     604
(1)T.rubripes        YISTAPCGDGALFDKSCSETGDEI--KGHQPLFENVKQGKLRTKVENGEGTIPVESSAIV     971
(1)G.gallus          YISTAPCGDGALFDKSCSDQASVVGPAEHQPLFENSKQGKLRTKVENGEGTIPVESSDIV     645
(1)C.anna            YVSTAPCGDGALFDKSCSDQASVVGQTQHQPLFENPKQGKLRTKVENGEGTIPVESSDIM     843
(1)D.rerio           YISTAPCGDGALFDKSCSEAAE-LNGSGHMPLFENIKQGKLRTKVENGEGTIPVESSDIV    1135
(1)X.laevis          YISTAPCGDGALFDKSCSDQPSAEGDNKHCPIFENVKQGKLRTKVENGEGTIPVESSDIV     918
(1)H.sapiens         YISTAPCGDGALFDKSCSDRAMESTESRHYPVFENPKQGKLRTKVENGEGTIPVESSDIV    1019
(1)M.musculus        YISTAPCGDGALFDKSCSDRAVESTESRHYPVFENPKQGKLRTKVENGEGTIPVESSDIV     942
(1)R.norvegicus      YISTAPCGDGALFDKSCSDRAVESTESRHYPVFENPKQGKLRTKVENGEGTIPVESSDIV     424
(3)D.labrax          YISTSPCGDGRLNSPYEIT--TDLHSSRHL---MRKHRSHLRTKIESGEGTVPVRGRGSV     522
(3)C.anna            YVSTSPCGDARLNSPYEIT--TDLNSSKHI---VKKYRGHLRTKIESGEGTIPVRCHNAA     411
(3)H.sapien          YVSTSPCGDARLHSPYEIT--TDLHSSKHL---VRKFRGHLRTKIESGEGTVPVRGPSAV     538
(3)M.musculus        YVSTSPCGDARVNSPYEIT--TDLNSSKHI---VRKFRGHLRTKIESGEGTVPVRGPSAV     544
(3)R.norvegicus      YVSTSPCGDARLNSPYEIT--TDLNSSKHI---VRKFRGHLRTKIESGEGTVPVRGPSAV     545
                     *:.*:****. : .             *     .  :.:****:*.****:**.
```

```
(2)H.sapiens          QTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFSSIILGSLYHGDHLSRA     559
(2)D.opalescens       QTWDGVLEGERLLTMSCSDKIARWNVLGVQGALLCHFMHPIYLESIILGSLYHSDHLSRA     524
(2)R.norvegicus       QTWDGVLQGERLLTMSCSDKIARWNVVGIQGALLSIFVEPIYFSSIILGSLYHGDHLSRA     559
(2)M.musculus         QTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEPIYFSSIILGGLYHGDHLSRA     559
(2)X.laevis           QTWDGVLQGERLLTMSCSDKLARWNVVGIQGSLLSLFVEPIYLSSIILGSLYHGDHLSRA     558
(2)O.hannah           QTWDGVLQGERLLTMSCSDKIARWNVLGIQGALLSLFVEPIYLSSIILGSLYHGDHLSRA     564
(2)G.gallus           QTWDGVLQGERLLTMSCSDKIARWNVLGIQGALLSLFVEPIYFSSIILGSLYHGDHLSRA     560
(2)C.anna             QTWDGVLQGERLLTMSCSDKIARWNVLGIQGALLSLFVEPIYFSSIILGSLYHGDHLSRA     558
(2)D.rerio            QTWDGVLQGERLLTMSCSDKIARWNVIGVQGSLLSYFTEPIYFSSLILGSLYHADHLSRA     579
(2)T.rubripes         QTWDGVLQGERLLTMSCSDKIARWNVVGFQGSLMSYFTEPIYFSSVILGSLYHADHLSRA     553
D.melanogaster        QTWDGVLQGQRLLTMSCSDKIARWNIVGIQGSLLSSIIEPVYLHSIVLGSLLHPEHMYRA     529
A.mellifera           QTWDGVLMGQRLLTMSCSDKIARWNVLGVQGALLSYFIEPIYFHSIVLGSLLNPSHMYRA     478
(1)O.hannah           PTWDGIQHGERLRTMSCSDKILRWNV---------------------------------     630
(1)T.rubripes         PTWDGIQHGERLRTMSCSDKILRWNVLGLQGALLSHFINPIYLKSITLGYLYSHGHLTRA     1031
(1)G.gallus           PTWDGIQHGERLRTMSCSDKILRWNVLGLQGALLSHFIEPVYLSSVTLGYLYSQGHLTRA     705
(1)C.anna             PTWDGIQHGERLRTMSCSDKILRWNVLGLQGALLSHFMEPVYLSSVTLGYLYSQGHLTRA     903
(1)D.rerio            PTWDGIQHGERLRTMSCSDKILRWNVLGLQGALLTHFIHPIYLHSITLGYLYSHGHLTRA     1195
(1)X.laevis           PTWDGIQHGERLRTMSCSDKILRWNVLGLQGGLLTHFMEPVYLSSLTLGYLFSKGHLTRA     978
(1)H.sapiens          PTWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFLQPIYLKSVTLGYLFSQGHLTRA     1079
(1)M.musculus         PTWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFLQPVYLKSVTLGYLFSQGHLTRA     1002
(1)R.norvegicus       PTWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFLQPVYLKSVTLGYLFSQGHLTRA     484
(3)D.labrax           QTWDGVLQGEQLITMSCTDKITRWNILGLQGALLSHFVEPVYLHSVTVGSLRHTGHLGRV     582
(3)C.anna             QTWDGVLLGEQLITMSCTDKIARWNILGLQGALLSSFIEPMYLHSIIVGSLYHTGHLSRV     471
(3)H.sapien           QTWDGVLLGEQLITMSCTDKIARWNVLGLQGALLSHFVEPVYLQSIVVGSLHHTGHLARV     598
(3)M.musculus         QTWDGILLGEQLITMSCTDKIASWNVLGLQGALLCHFIEPVYLHSIIVGSLHHTGHLARV     604
(3)R.norvegicus       QTWDGILLGEQLVTMSCTDKIASWNVLGLQGALLCHFIEPVYLHSIIVGSLHHTGHLARV     605
                       ****:  *::* ****:**:  **:


(2)H.sapiens          MYQRISN-----IEDLPPLYTLNKPLLSGISNA-EARQPGKAPNFSVNWTVGDS-AIEVI     612
(2)D.opalescens       VYCRIAA-----IENLPDLFRLNRPFLSGISSP-ESRQPGKAPNFGINWRRNDD-SFEVI     577
(2)R.norvegicus       MYQRISN-----IEDLPPLYTLNKPLLSGISNA-EARQPGKAPNFSVNWTVGDT-AIEVI     612
(2)M.musculus         MYQRISN-----IEDLPPLYTLNKPLLSGISNA-EARQPGKAPNFSVNWTVGDA-TIEVI     612
(2)X.laevis           VYQRISD-----IEDLPILYALNKPLLSGISNA-EARQPGKAPSFSVNWTVGDL-SLEVI     611
(2)O.hannah           AYQRIAE-----IEDLPSLYVLNRPLLSGISNA-EARQPGKAPNFSVNWTVGDA-GLEVI     617
(2)G.gallus           VYQRIAE-----IEDLPPLYTLNRPLLSGISNA-EARQPGKAPNFSVNWTVGDT-GLEVI     613
(2)C.anna             VYQRIAE-----IEDLPPLYTLNRPLLSGISNA-EARQPGKAPNFSVNWTVGDT-GLEVI     610
(2)D.rerio            MYQRIAD-----MDDIPPPFRINRPLLSGISNT-EARQPGKAPNFSVNWTVGDQ-GLEII     632
(2)T.rubripes         MYQRITD-----IEDLPQSFSLNRPLLSGISNA-EARQPGKAPNFRVNWTVGDQ-GLKVI     606
D.melanogaster        VCGRIEKS----IQGLPPPYHLNKPRLALVTSA-EPRNQAKAPNFGINWTIGDT-ELEVV     583
A.mellifera           VCGRIENT----IQGLPPPYRLNKPLMSLITSS-EVRQPGKAPNYSVNWTIGQL-EAEVI     532
(1)O.hannah           ---------------------------GRVSVYDSARQTGKTKESSINWCLPDGTSVEIL     663
(1)T.rubripes         VCCRMATNGQEFAQSLSAPFMLNHPEVGRVSVYDSTRHTGKTKESSVNWSFPDQHSVEVL     1091
(1)G.gallus           ICCRLVRDGDVLQKRLQAPYQINHPEVGRVSVYDSARQTGKTKESSVNWCLADGSKVEVL     765
(1)C.anna             ICCRVARERNVLQAKLQAPYHINHPEVGRVSVYDSARQTGKTKESSVNWCLADESEVEVL     963
(1)D.rerio            VCCRLSRDGDTFKSSLPANFTLNHPEVGRVSVYDSTRHTSKTKESSVNWSQPDQYSVEVL     1255
(1)X.laevis           ICCRMSRDGDAFQNQLPDLYLVNHPEVGRVSVYDSTRQTGKTKESSVNWCLADE-EAEVL     1037
(1)H.sapiens          ICCRVTRDGSAFEDGLRHPFIVNHPKVGRVSIYDSKRQSGKTKETSVNWCLADGYDLEIL     1139
(1)M.musculus         ICCRVTRDGKAFEDGLRYPFIVNHPKVGRVSVYDSKRQSGKTKETSVNWCMADGYDLEIL     1062
(1)R.norvegicus       ICCRVTRDGNAFEDGLRYPFIVNHPKVGRVSVYDSKRQSGKTKETSVNWCLADGYDLEIL     544
(3)D.labrax           LNQRLER-----LGPLPATHRRNQPLLSGLSSA-EYQQPGKASCVSVNWTLGDT-QLEVV     635
(3)C.anna             MSHRIED-----IGQLPASYRRNQLLLSGVSHA-DARQPGKSPGFSVNWIVGNT-DLEVI     524
(3)H.sapien           MSHRMEG-----VGQLPASYRHNRPLLSGVSDA-EARQPGKSPPFSMNWVVGSA-DLEII     651
(3)M.musculus         MSHRMEG-----IGQLPASYRQNRPLLSGVSHA-EARQPGKSPHFSANWVVGSA-DLEII     657
(3)R.norvegicus       MSHRMEG-----IGQLPASYRQNRPLLSGVSNA-EARQPGKSPHFSANWVVGSA-DLEII     658
                                              . ::   . :. .*:     **   .    :::
```

```
(2)H.sapiens          NATTGKD-ELG--RASRLCKHALYCRWMRVHGKVPSH-LLRSKITKPNVYHESKLAAKEY     668
(2)D.opalescens       NAMTGRV-EGG--SMSRICKQALFDRFMNLYGKLSSL-TGQSVTTRPALYSEAKATVMEY     633
(2)R.norvegicus       NATTGKD-ELG--RPSRLCKHALYCRWMRVHGKVPPH-LLRTKITKPTTYHESKLAAKEY     668
(2)M.musculus         NATTGKD-ELG--RPSRLCKHALYCRWMRVHGKVPPH-LLRTKITKPTTYHESKLAAREY     668
(2)X.laevis           NATTGKD-EMG--RASRLCKHALYSRWMRIHAKLSSS--VRCKFGKPNVYHETKQSAVEY     666
(2)O.hannah           NATTGKD-ELG--RASRLCKHALYTRWMRVYTKLPAS--LHSKVNKPNIYHETKQVAAEY     672
(2)G.gallus           NATTGKD-EMG--RASRLCKHALYSRWMRIHAKLSSS--LRLKIFKPNLYHDTKQGATEY     668
(2)C.anna             NATTGKD-EMG--RASRLCKHAFYSRWMRIHAKLSSS--LRSKILKPNLYHETKQGAVEY     665
(2)D.rerio            NATTGKH-DLG--RQSQLCKHALYSRWVCLHAKLSET--LRIRGSRPGSYHEAKQGAVEY     687
(2)T.rubripes         NATTGKD-DLG--RPSRLCKHALYGRWMRLHSKLSPS--LRIRTVRPSSYHEAKQAAVDY     661
D.melanogaster        NSLTGRT-IGG--QVSRITKQAFFVKYGFLMANLPGI--LV--RKVTTDYGQTKANVKDY     636
A.mellifera           NCTTGKD-ELG--KPSRISKQGLFRRFYNLLGKLSTI--EDADKNQCRHYLDAKSSVQNY     587
(1)O.hannah           DGTKGKV-DGPKLDISRVSKQSLFQLFRMLCTKM-----VRKDLKNFVVYSEAKESATDY     717
(1)T.rubripes         DGTTGKL-DGNKLSVSRVTKSNLFALFRAVCQRC-----GRTDLLSLHSYSQAKMAALSF    1145
(1)G.gallus           DGTKGKV-DGPKLEVSRVSKRKTFALFQQLCAKS-----DCKELQKLSVYSEAKKAAVQY     819
(1)C.anna             DGTKGKV-DGPKLEVSRVSKRKMFALFQQLCAKS-----NRGDLQSLSVYSDAKEAATAY    1017
(1)D.rerio            DGTKGKL-DSPKMEVSRVSKSNLFRLFHALCQRA-----GRADLLALQSYAHAKMAATSF    1309
(1)X.laevis           DGTKGKV-EGAKLEISRVSKLHMFTLFQELCLLR-----GRHDLLALSSYSDVKATVGTY    1091
(1)H.sapiens          DGTRGTV-DGPRNELSRVSKKNIFLLFKKLCSFR-----YRRDLLRLS-YGEAKKAARDY    1192
(1)M.musculus         DGTRGTV-DGPGKELSRVSKKNIFLQFKKLCSFR-----ARRDLLQLS-YGEAKKAARDY    1115
(1)R.norvegicus       DGTRGTV-DGPGKELSRVSKKNIFLQFKKLCSFR-----ARRDLLQLS-YGEAKKAARDY     597
(3)D.labrax           NTATGRRRESG--TPSRLCKHALFTRWNRLYRKVRLGIHVSSSADRQLMYCEAKMAARPY     693
(3)C.anna             NAMTGKR-TCG--SPSRLCKHMFFTRWAKLHGKLS--TRTPSHGEMPSVYSEAKLVAQTY     579
(3)H.sapien           NATTGRR-SCG--GPSRLCKHVLSARWARLYGRLS--TRTPSPGDTPSMYCEAKLGAHTY     706
(3)M.musculus         NATTGKR-SCG--GSSRLCKHVFSAWWARLHGRLS--TRIPSHGDTPSMYCEAKQGAHTY     712
(3)R.norvegicus       NATTGKR-SCG--GSSRLCKHVFSARWARLHGRLS--TRIPGHGDTPSMYCEAKRGAHTY     713
                      :   *          *:: *      :  :                   * . *  .  :

(2)H.sapiens          QAAKARLFTAFIKAGLGAWVEKPTEQDQFSLTP----------------------------     701
(2)D.opalescens       QLAKQCVFQAFQKAGLGNWVQKPIEQDQFEMSLDAPAIQLKAAETETANTETSA------     687
(2)R.norvegicus       QAAKARLFTAFIKAGLGAWVEKPTEQDQFSFTP----------------------------     701
(2)M.musculus         QAAKARLFTAFIKAGLGAWVEKPTEQDQFSFTP----------------------------     701
(2)X.laevis           QSAKECVFKVFQKTGLGAWVKKPIEQDQFSLNV----------------------------     699
(2)O.hannah           QTAKECLFKAFLKAGLGAWVEKPIEQDQFSLVV----------------------------     705
(2)G.gallus           QTAKECLFKAFLKAGLGAWVEKPIEQDQFSLTV----------------------------     701
(2)C.anna             QTAKECLFKAFLKAGLGAWVEKPIEQDQFSLTV----------------------------     698
(2)D.rerio            HEAKQTLFKAFYKAGLGAWVEKPIEQDQFSLIP----------------------------     720
(2)T.rubripes         HSAKQTLFKAFQKSGLGAWVKKPIEQDQFSVTT----------------------------     694
D.melanogaster        QIAKLELFSAFKREDLGSWLKKPIEQDEFGLAE----------------------------     669
A.mellifera           SLAKHQLKEAFVKAHLGSWVKKPIEQDMFEVDI----------------------------     620
(1)O.hannah           QSAKQQFFWGLQEMGYGSWICKPQEEEAFVLPEPAAPPFL---------------------     757
(1)T.rubripes         QLAKQQFFQALTAHGYGTWIGKPLEEKSFEAGESSWNNEASVP------CGRRQQ-----    1194
(1)G.gallus           QKAKQCFFSALEEMGYGSWICKPQEEDNFSIPDA---------------------------     853
(1)C.anna             QEAKQSFFSALEELGYGSWIQKPQEEENFSL------------------------------    1048
(1)D.rerio            QEAKRLFFLALSQHGYGAWIGKPLEEKSFEGEAKSFEPMGGSGDVQASGCVVDPNSFLSS    1369
(1)X.laevis           QTARGQFFRALEQMGYGNWISKPQEEKCFSLSI----------------------------    1124
(1)H.sapiens          ETAKNYFKKGLKDMGYGNWISKPQEEKNFYLCPV---------------------------    1226
(1)M.musculus         DLAKNYFKKSLRDMGYGNWISKPQEEKNFYLCPVPND------------------------    1152
(1)R.norvegicus       DLAKNYFKKSLRDMGYGNWISKPQEEKNFYLCPVPND------------------------     634
(3)D.labrax           QTVKQQWFRSLQETGLGTWVKKPPEQEQFLLTV----------------------------     726
(3)C.anna             QSVKQQLFKAFQKAGLGTWVKKPPEQDQFLLTV----------------------------     612
(3)H.sapien           QSVKQQLFKAFQKAGLGTWVRKPPEQQQFLLTL----------------------------     739
(3)M.musculus         QSVKQQLFKAFQKAGLGTWVRKPPEQDQFLLSL----------------------------     745
(3)R.norvegicus       QSVKQQLFKAFQKAGLGTWVRKPPEQDQFLLSL----------------------------     746
                      .:      :     * *: ** *:. *
```

```
(2)H.sapiens          ------------     701
(2)D.opalescens       ------------     687
(2)R.norvegicus       ------------     701
(2)M.musculus         ------------     701
(2)X.laevis           ------------     699
(2)O.hannah           ------------     705
(2)G.gallus           ------------     701
(2)C.anna             ------------     698
(2)D.rerio            ------------     720
(2)T.rubripes         ------------     694
D.melanogaster        ------------     669
A.mellifera           ------------     620
(1)O.hannah           ------------     757
(1)T.rubripes         ------------     1194
(1)G.gallus           ------------     853
(1)C.anna             ------------     1048
(1)D.rerio            ENNNFLSAAQGF     1381
(1)X.laevis           ------------     1124
(1)H.sapiens          ------------     1226
(1)M.musculus         ------------     1152
(1)R.norvegicus       ------------     634
(3)D.labrax           ------------     726
(3)C.anna             ------------     612
(3)H.sapien           ------------     739
(3)M.musculus         ------------     745
(3)R.norvegicus       ------------     746
```

Supplementary table 2. Sequences of DNA and RNA oligonucleotides used in methods.

a) RNA sequences used for crystallography

RNA name	RNA sequence	Calculated/Observed monoisotopic mass(ESI)
Bdf2 8-aza-N TOP	UUCCCCACAUU**N**GACGUUCAGUC	7186.93/7187.09
Bdf2-C BOTTOM	GACUGAACGACCAAUGUGGGGAA	7472.6/7472.1
Bdf2-U BOTTOM	GACUGAACGACCAAUGUGGGGAA	7473.6/7473.6
GLI1-23mer 8AN TOP	GCUCGCGAUGCU**N**GAGGGCUCUG	7376.98/7376.24
GLI1-23mer BOTTOM	CAGAGCCCCCCAGCAUCGCGAGC	7315.06/7306.18

b) Primers for site directed mutagenesis

Oligo name	Sequence
3' end of ligated Gli1 top strand RNA	5'-A**X**A GGG CUC UGC-3'
5' end of ligated Gli1 top strand RNA	5'-GCU CGC GAU GCU-3'
Gli1 bottom strand RNA	5'-GCA GAG CCC C**Y**C AGC AUC GCG AGC-3'
DNA splint	5'-GCA GAG CCC T**Y**T AGC ATC GCG AGC-3'

c) RNA sequences used for in vitro deamination kinetics

Mutant	Forward primer	Reverse primer
R510Q	GGTGCTGCAAGGGGAGCAACTGCTCACCATGTCC	GGACATGGTGAGCAGTTGCTCCCCTTGCAGCACC
R510A	GGTGCTGCAAGGGGAGGCTCTGCTCACCATGTCC	GGACATGGTGAGCAGAGCCTCCCCTTGCAGCACC
G593A	GCAGAAGCACGGCAGCCAGCTAAGGCCCCCAACTTCAGTG	CACTGAAGTTGGGGGCCTTAGCTGGCTGCCGTGCTTCTGC
G593E	GCAGAAGCACGGCAGCCAGAAAAGGCCCCCAACTTCAGTG	CACTGAAGTTGGGGGCCTTTTCTGGCTGCCGTGCTTCTGC
K594A	CAGAAGCACGGCAGCCAGGGGCTGCCCCCAACTTCAGTGTC	GACACTGAAGTTGGGGGCAGCCCCTGGCTGCCGTGCTTCTG
R348A	CTTCTCCTCCCCTCACGCTGCAAGAAAAGTGCTGGCTGG	CCAGCCAGCACTTTTCTTGCAGCGTGAGGGGAGGAGAAG

d) primers used for mutagenesis of hGLI1 in vitro transcription template

3' nearest neighbor mutant	Forward primer	Reverse primer
G to C	GTCTACTCACCACAGCCCCGCAGCATCACTGAGAATGCTGCCATGGATGCTACAGGGCTACAGGAAGAGCCAG	CTGGCTCTTCCTGTAGCCCTGTAGCATCCATGGCAGCATTCTCAGTGATGCTGCGGGGCTGTGGTGAGTAGAC
G to A	GTCTACTCACCACAGCCCCTCAGCATCACTGAGAATGCTGCCATGGATGCTAAAGGGCTACAGGAAGAGCCAG	CTGGCTCTTCCTGTAGCCCTTTAGCATCCATGGCAGCATTCTCAGTGATGCTGAGGGGCTGTGGTGAGTAGAC
G to U	GTCTACTCACCACAGCCCCACAGCATCACTGAGAATGCTGCCATGGATGCTATAGGGCTACAGGAAGAGCCAG	CTGGCTCTTCCTGTAGCCCTATAGCATCCATGGCAGCATTCTCAGTGATGCTGTGGGGCTGTGGTGAGTAGAC

e) RNA sequences used for EMSA binding experiments

RNA name	RNA sequence	Calculated/Observed monoisotopic mass(ESI)
GLI1-24mer 8AN TOP	GCUCGCGAUGCU**N**GAGGGCUCUGC	7682.0/7681.9
GLI1-24mer BOTTOM	GCAGAGCCCCCCAGCAUCGCGAGC	7663.7/7663.4