Double-Stranded RNA Adenosine Deaminases ADAR1 and ADAR2 Have Overlapping Specificities†

Katrina A. Lehmann and Brenda L. Bass*

Department of Biochemistry and HHMI, University of Utah, 50 North Medical Drive, Room 211, Salt Lake City, Utah 84132

Received June 16, 2000; Revised Manuscript Received August 11, 2000

ABSTRACT: Adenosine deaminases that act on RNA (ADARs) deaminate adenosines to produce inosines within RNAs that are largely double-stranded (ds). Like most dsRNA binding proteins, the enzymes will bind to any dsRNA without apparent sequence specificity. However, once bound, ADARs deaminate certain adenosines more efficiently than others. Most of what is known about the intrinsic deamination specificity of ADARs derives from analyses of *Xenopus* ADAR1. In addition to ADAR1, mammalian cells have a second ADAR, named ADAR2; the deamination specificity of this enzyme has not been rigorously studied. Here we directly compare the specificity of human ADAR1 and ADAR2. We find that, like ADAR1, ADAR2 has a 5′ neighbor preference (A ≈ U > C = G), but, unlike ADAR1, also has a 3′ neighbor preference (U = G > C = A). Simultaneous analysis of both neighbor *preferences* reveals that ADAR2 prefers certain trinucleotide sequences (U<u>A</u>U, A<u>A</u>G, U<u>A</u>G, A<u>A</u>U). In addition to characterizing ADAR2 *preferences*, we analyzed the fraction of adenosines deaminated in a given RNA at complete reaction, or the enzyme's *selectivity*. We find that ADAR1 and ADAR2 deaminate a given RNA with the same *selectivity*, and this appears to be dictated by features of the RNA substrate. Finally, we observed that *Xenopus* and human ADAR1 deaminate the same adenosines on all RNAs tested, emphasizing the similarity of ADAR1 in these two species. Our data add substantially to the understanding of ADAR2 specificity, and aid in efforts to predict which ADAR deaminates a given editing site adenosine in vivo.

Adenosine deaminases that act on RNA (ADARs)[1] deaminate adenosines to produce inosines within a diverse set of RNAs, including cellular pre-mRNAs, viral RNAs, and noncoding RNAs of unknown function (*1, 2*). In some of these RNAs, the function of deamination is unclear, but in others ADARs clearly target codons so that multiple protein isoforms can be synthesized from a single RNA. Given the abundance of inosine within poly-A+ RNA (*3*), it is certain that many ADAR substrates remain to be discovered, and that the impact of ADAR activity on cellular processes is only beginning to be revealed.

In mammalian cells, two ADARs have been identified: ADAR1 and ADAR2. ADAR1 was the first identified ADAR, and was discovered in *Xenopus laevis* (*4, 5*). Both ADAR1 and ADAR2 contain multiple copies of the dsRNA binding motif (dsRBM), an ~70 amino acid sequence shared by many dsRNA binding proteins (dsRBPs), including RNase III and PKR (*6, 7*). Like all dsRBPs, both ADARs will bind to any dsRNA regardless of sequence. Accordingly, discrimination of a target adenosine from other adenosines is thought to occur after binding to the RNA has occurred.

Despite the identification of ADAR1 (*8, 9*) and ADAR2 (*10, 11*) in mammals, most of what is known about the intrinsic specificity of ADARs is based upon in vitro analyses of *Xenopus* ADAR1 (xADAR1). Studies of this enzyme show that deamination specificity has at least two determinants: deamination *preferences* and deamination *selectivity* (*12*). *Preferences* refer to the preferred sequence context of an adenosine, and the preferred location of an adenosine within a helix. For example, xADAR1 preferentially deaminates adenosines that have a 5′ nearest neighbor of A or U, and are more than 8 nt from a 3′ end (*12*). *Selectivity* refers to the fraction of adenosines within a helix that are deaminated at reaction completion, and studies of ADAR1 show that this is related to the effective length and stability of the substrate helix (*12, 13*). Short, unstable dsRNAs are selectively deaminated at less than 10% of their adenosines, while long, stable substrates are promiscuously, or nonselectively, deaminated at 50−60% of their adenosines. It has also been shown that ADAR1 recognizes certain internal loops as helix termini. Thus, insertion of internal loops into a long, promiscuously deaminated dsRNA increases *selectivity* by creating a series of short helices that are each deaminated at only a few sites (*14*).

Although discovered in 1996 (*10*), little is known about the specificity of ADAR2. In vitro, ADAR1 and ADAR2 can each deaminate certain biological editing sites within

† This work was supported by funds to B.L.B. from the National Institute of General Medical Sciences (GM44073). Oligonucleotides were synthesized by the Howard Hughes Medical Institute oligonucleotide synthesis facility at the University of Utah supported by the National Cancer Institute (Grant 42014) and HHMI. B.L.B. is an HHMI Associate Investigator.

* To whom correspondence should be addressed. E-mail: bbass@howard.genetics.utah.edu, phone: 801.581.3824, fax: 801. 581.5379.

[1] Abbreviations: ADAR, adenosine deaminase that acts on RNA; dsRNA, double-stranded RNA; dsRBM, dsRNA binding motif; dsRBP, double-stranded RNA binding protein; HDV, hepatitis delta virus; 5-HT, 5-hydroxytryptamine; NMP, nucleoside monophosphate; AMP, adenosine monophosphate; IMP, inosine monophosphate; ATP, adenosine triphosphate; TLC, thin-layer chromatography; nt, nucleotide; bp, base pair.

10.1021/bi001383g CCC: $19.00
Published on Web 09/28/2000

the glutamate receptor (gluR) and the 5-HT_{2C} serotonin receptor pre-mRNAs. For instance, either ADAR1 or ADAR2 can edit the gluR-B R/G site (*10*), and the serotonin A and C editing sites (*15*). In contrast, in vitro, the serotonin B site is deaminated only by ADAR1, while the gluR-B Q/R and the serotonin D sites are deaminated only by ADAR2 (*10, 15, 16*). Based on these observations, ADAR1 and ADAR2 are thought to have overlapping specificities, but ADAR2 deamination specificity has not been systematically investigated.

Here we investigate the specificity of ADAR2. We find that ADAR2 has a 5′ neighbor preference similar to that of ADAR1, but, unlike ADAR1, also has a 3′ neighbor preference. Interestingly, for a given RNA, the fraction of adenosines deaminated at complete reaction (the *selectivity*) is the same regardless of which enzyme is used, although which adenosines are deaminated differs. We compare the specificities of ADAR1 and ADAR2, and relate this to the context of biological editing sites. Our comparisons of ADAR1 and ADAR2 specificities suggest a model regarding how ADARs achieve efficient deamination at specific editing sites.

EXPERIMENTAL PROCEDURES

Expression and Purification of Human ADAR2. The vector pSc[hA2a-H6] used for hADAR2a overexpression in *Saccharomyces cerevisiae* strain BCY123 was constructed from vector pJEL167 and modified to provide expression from an inducible GAL1 promoter (Ley, H. L., III, Lingam, A., and Bass, B. L., unpublished). pJEL167 is a derivative of YEpTOP2PGAL1 (*17*) in which the *Pvu*I site immediately downstream of the TOP2 gene has been replaced with a unique *Xho*I site. pSc[hA2a-H6] included human ADAR2a with a C-terminal $(his)_6$ epitope tag and a 3′ end forming signal for processing in yeast (*18*). BCY123 colonies containing pSc[hA2a-H6] were screened for ADAR activity prior to choosing one clone for overexpression.

hADAR2a expressed in *S. cerevisiae* was purified using a method previously developed for recombinant ADARs in *Pichia pastoris* (*19*) with modifications that include a heparin Sepharose column as a third chromatographic step (Ley, H. L., III, and Bass, B. L., personal communication). Briefly, 3.6 g of BCY123 cells overexpressing hADAR2a was disrupted using a French Pressure cell in a buffer (*19*) that included elevated KCl (160 mM). The cell lysate was clarified by centrifugation at 100000*g* for 90 min at 4 °C prior to loading onto a 50 mL Macro-Prep High Q column (BioRad). The Macro-Prep column was developed in a KCl gradient as described (*19*), and the eluted fractions exhibiting ADAR activity were identified by the altered mobility of an ~800 bp duplex on a (29:1) 6% native polyacrylamide gel as described (*4*) except that the RNA was not capped. Selected fractions were pooled and batch-bound to Ni^{2+}-NTA resin and eluted as described (*19*) except Superflow resin (Qiagen) was used and 200 mM NaCl was included in the starting buffers. Fractions containing hADAR2 were evident by the presence of a band of the correct migration by 4−15% SDS−PAGE (BioRad) following SYPRO Red (Molecular Dynamics) staining and visualization by red fluorescence on a Molecular Dynamics Storm PhosphorImager. The Ni^{2+}-NTA pool was further purified on a 1 mL Hi-Trap heparin column (Pharmacia) developed in a KCl gradient using a BioCAD SPRINT Perfusion Chromotography System (PerSeptive Biosystems). Fractions containing hADAR2 were identified by SYPRO Red staining of SDS−polyacrylamide gels as above, and pooled prior to dialysis (10 000 MWCO) into a final storage buffer as described (*19*) except additional protease inhibitors were not added and glycerol was 20% (v/v). Aliquots were stored at −80 °C. Human ADAR1 was purified using the first two columns.

The final hADAR2 pool was of single band purity by SYPRO Red staining as above. These gels were also used to determine hADAR2 concentration, which was quantified on a Molecular Dynamics Storm PhosphorImager by comparing the intensity of the ADAR2 band to the intensity of BSA standards of known concentration (data not shown). The stock hADAR2 concentration was determined to be 70.5 nM. The stock hADAR1 concentration was determined to be 884 nM (Ley, H. L., III, and Bass, B. L., unpublished data). By SDS−PAGE, the hADAR1 band comprises approximately 40% of the SYPRO Red staining material (data not shown).

Nucleic Acid Preparation. RNAs annealed to create the 61mer, 36mer, inv-36mer, and 48mer duplexes were transcribed from partially single-stranded DNA templates (*20*) using the method described for 36mer (*14*). 102mer, previously referred to as 100mer (*7*), was created by bi-directional runoff transcription of a 139 bp *Bss*HII fragment from plasmid JP11-11 as described (*7*). All transcripts were purified on a denaturing gel, and then annealed to their complementary strands in 10 mM Tris (pH 8.0) prior to isolating the RNA duplexes from a 10% native gel as described (*14*). Although not shown in Figure 2c, the sequence of the 102mer 3′ overhangs has been given previously (*7*).

For nuclease mapping applications, the RNA strand to be analyzed was gel-purified, labeled at its 5′ end using 6000 Ci/mmol [γ-^{32}P]ATP and T4 PNK kinase as described (*12*), purified from a second denaturing gel, and then annealed to its complementary strand. dsRNAs prepared for TLC applications were internally labeled at adenosines during transcription with 0.2 mCi of 3000 Ci/mmol [α-^{32}P]ATP (Amersham) as described (*14*). RNAs to be analyzed by primer extension were labeled with 0.1 mCi of [5,6-^{3}H]UTP during transcription for quantitation purposes.

Deamination Reactions. All deamination reactions were performed in a standard assay buffer (1 × AB) as described (*21*) with modifications (*14*). Complete deamination of RNAs by hADAR2 was confirmed by time course experiments and enzyme titrations; control experiments verified that the enzyme was fully active throughout the course of the reaction (3 h). Unless stated otherwise, complete deamination by ADAR2 was achieved in 50 μL reactions containing 0.3 nM RNA, 7.1 nM hADAR2, and 1 × AB, after incubation for 3 h at 30 °C. Deamination reactions were stopped and RNAs purified as described (*12*) prior to analysis of inosine content.

Analysis of Deaminated RNAs. TLC was performed as described (*14*) using RNA internally labeled at adenosines. Incubation was for 3 h at 30 °C for hADAR2 or at 25 °C for xADAR1 under the same conditions that give complete deamination of RNAs used to map the location of inosines.

Nuclease mapping was performed as described (*12*). Briefly, 10 000 cpm of ^{32}P-5′-end-labeled RNA was treated

with 0.66 unit of ribonuclease U2 (RNase U2) or 0.05 unit of (E46Q) RNase T1 (*22*) in a 10 μL reaction volume in the presence of 20 mM sodium acetate (pH 4.5) and 1 μg of torula RNA. Alkaline hydrolysis reactions were performed as described (*12*).

Primer extension reactions were performed as described (*23*) with modifications (*14*). Oligonucleotides used as primers were prepared as described (*14*), and are as follows: TDRpe (5′ GATCTCGTGCGTCAGG 3′) and TDYpe (5′ GATCCGTGGTGTAAGG 3′) anneal to the R and Y strands of 61mer, respectively; 102T7pe (5′ CGCGCAATTAACCCTCACTAAAGGGAAC 3′) and 102 T7pe54 (5′ AGCGGAGGCGCAAG 3′) anneal to the T7 strand of 102mer; 102T3pe (5′ CGCGCGTAATACGACTCACTATAGGGCG 3′) and 102T3pe56 (5′ GCAAAGCGAGGCA) anneal to the T3 strand of 102mer.

Nuclease mapping and primer extension reaction products were heated (95 °C, 3 min) in 1 × formamide loading buffer (*14*) and loaded onto a 0.5 mm thick, 8 M urea, 12% polyacrylamide (19:1), 1 × TBE, gel. Electrophoresis was at 50 W. Gels were dried and subjected to autoradiography and PhosphorImager analysis.

Determination of Percent Deamination in Mapping Experiments. Nuclease mapping experiments were quantified using RNase U2 cleavage data. After RNase U2 lanes were normalized for loading and RNase digestion differences by comparing RNase U2 cleavage at guanosines (*12*), deamination efficiency was evaluated by calculating the percent loss of U2 cleavage at adenosines upon ADAR treatment, and was used to determine the average percent deamination of adenosines in a given context:

$$\%\ \text{deamination} = \frac{\sum_{\text{all As in context}}\left(\left(\frac{\text{U2}(-\text{ADAR}) - \text{U2}(+\text{ADAR})}{\text{U2}(-\text{ADAR})}\right) \times 100\right)}{\text{total As in context}}$$

Here U2 is the intensity of the RNase U2 cleavage band, −ADAR means a mock reaction that did not include ADAR, and +ADAR means the ADAR-treated sample. The "% deamination" value is the average deamination efficiency of adenosines in a given context, and is reported in Tables 1−3. The RNase U2 data carried a standard deviation of ∼12%, and variations were largely due to differences in the extent of partial U2 digestion and the influence of the 3′ neighbor preferences of U2 on digestion efficiency [see (*12*)].

Since the intensity of primer extension stopping at adenosines is potentially affected by the processivity and fidelity of the reverse transcriptase enzyme, the deamination efficiency at a given adenosine was approximated by repetitions of the experiment and comparisons with nuclease mapping data as described (*14*). A minor gain in ddCTP stop with no change in ddTTP stop indicated a minor deamination site (i) with an estimated value of 25% inosine. Sites with ∼50% drop in ddTTP stop and a corresponding ∼50% gain of ddCTP were given a value of 50%. Sites with a complete loss of ddTTP stop and a significant gain in ddCTP stop were considered major sites with an estimated value of 75%, which agreed well with highly deaminated sites assayed by nuclease mapping. The latter two categories were designated as major deamination sites, and are symbolized as I in Figure 2c sequences of 61mer and 102mer.

RESULTS

As mentioned, in vitro studies comparing the ability of ADAR1 and ADAR2 to deaminate editing sites in biological substrates suggest that the enzymes have overlapping specificities (*10*, *15*, *16*). Although the deamination specificity of *Xenopus* ADAR1 has been extensively studied (*12*−*14*), the intrinsic specificity of ADAR2 has not been investigated. Understanding the specificity of both enzymes is central to predicting which ADAR is responsible for deaminating a specific editing site in vivo. Of related interest, we wanted to know if the specificity of human ADAR1 differed from that of the well-characterized *Xenopus* ADAR1 enzyme; these enzymes show 50% amino acid identity overall. We reasoned that if xADAR1 and hADAR1 have the same specificity, we could use our knowledge of xADAR1 specificity for comparisons with hADAR2, and to studies of human ADAR substrates.

To compare the deamination specificity of ADAR1 and ADAR2, we analyzed the deamination patterns produced by purified xADAR1, hADAR1, or hADAR2 on the model RNAs used previously to study xADAR1 specificity (*12*). One of these RNAs, 36mer, is a dsRNA 36 bp in length that contains a purine-rich (R) and pyrimidine-rich (Y) strand, and places adenosines within numerous sequence contexts. 36mer was reacted to completion with each ADAR, and inosines were mapped using a mutant (E46Q) RNase T1 [(*12*, *22*); Figure 1]. When the 36mer was reacted to completion with xADAR1 at 25 or 30 °C, or hADAR1 at 30 °C, we observed the same deamination pattern, emphasizing the functional similarity of these enzymes (Figure 1a). In contrast, when 36mer was reacted to completion with hADAR2 at 30 °C, a distinct but overlapping deamination pattern was observed. A comparison of the lanes shown in Figure 1a shows that some adenosines in 36mer are deaminated by both ADAR1 and ADAR2 (e.g., A5, A6, and A10), while others are unique to either ADAR1 (e.g., A9, A14, and A15) or ADAR2 (e.g., A16, A19, A22, and A27). Like the ADAR1 and ADAR2 deamination patterns on 36mer (Figure 1b), ADAR1 and ADAR2 deamination patterns also overlap on 48mer (Figure 1c), and inv-36mer (not shown), RNAs related in sequence to 36mer (see Figure 2c, arrows). These data indicate that ADAR1 and ADAR2 have overlapping specificities, and that *Xenopus* and human ADAR1 have a highly similar, or identical, deamination specificity.

Although *preferences* and *selectivity* have been characterized for ADAR1, little is known about ADAR2 specificity. We were curious about the specific differences between the specificities of ADAR1 and ADAR2 that led to the production of overlapping deamination patterns. Thus, we analyzed ADAR2 deamination patterns on five dsRNAs (Figure 2). The RNAs were deaminated with hADAR2 to completion, and then purified and analyzed by nuclease mapping and primer extension methods. Nuclease mapping experiments were done with (E46Q) RNase T1 as in Figure 1 and also included RNase U2, which cleaves 3′ of adenosines. Deamination sites showed a decrease in RNase U2 cleavage, and an increase in (E46Q) RNase T1 cleavage (Figure 2a; e.g., compare T1 and U2 lanes for A22, R strand). The primer extension method allows detection of inosines through the loss of a reverse transcriptase ddTTP stop and the corresponding gain of a ddCTP stop (Figure 2b; e.g., compare

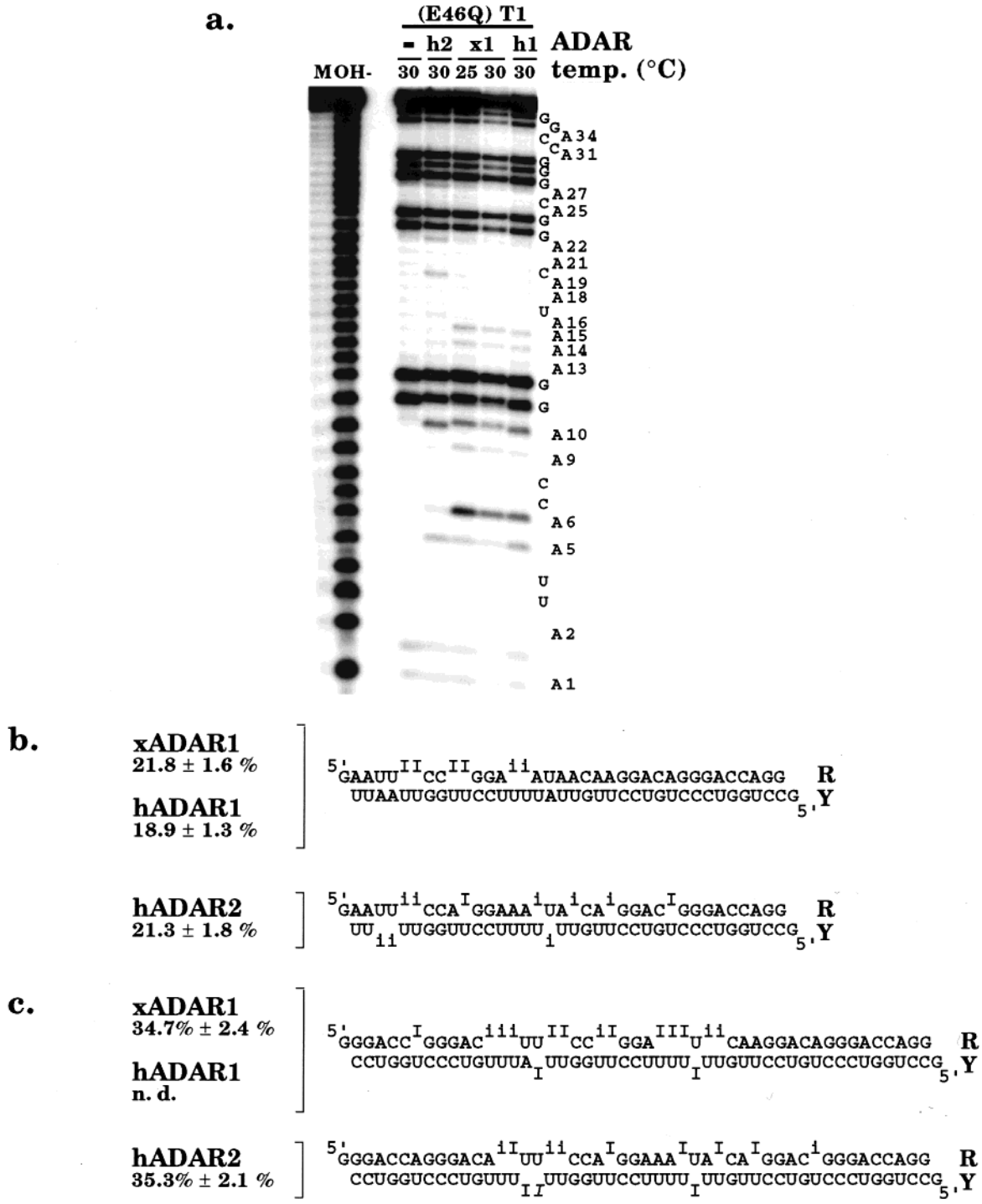

FIGURE 1: Comparison of ADAR1 and ADAR2 deamination patterns. (a) The autoradiogram shows an example of the nuclease mapping data used to determine deamination patterns. In the absence of ADAR (−), (E46Q) RNase T1 cleavages show the positions of guanosines. New cleavages upon ADAR treatment (h2, x1, h1) indicate the position of inosines. Reactions containing ADAR and 0.1 nM 36mer (R strand 5′ end labeled with ^{32}P) were incubated at the specified temperature (25 or 30 °C) for 3 h. To obtain completely reacted RNAs using ADAR stocks of different specific activity or purity, human ADAR1 was used at 35.4 nM, and native *Xenopus* ADAR1 AF-Blue-650 M pool at 0.5 nM (see Experimental Procedures). h1, hADAR1; h2, hADAR2; x1, xADAR1; M, mock incubated; OH-, alkaline hydrolysis ladder. The locations of inosines within 36mer (b) and 48mer (c) deaminated and nuclease-mapped as described in (a) are shown. The percentage of adenosines deaminated under the complete reaction conditions used for nuclease mapping was determined by TLC analysis, and is given beneath the enzyme name. Data shown in (b) and (c) are derived from reactions performed at 25 °C (xADAR1) or 30 °C (hADAR1 and 2). Minor sites (<50% deaminated) are indicated with i, and major sites (≥50% deaminated) are shown with I, and were determined using data from multiple experiments (b, $n = 4$; c, $n = 2$). Quantification of mapping data is described in the legend to Figure 2 and under Experimental Procedures. n. d., not determined.

ddTTP and ddCTP lanes for A50, R strand), and was useful in mapping inosines within longer RNAs. The locations of inosines within the five RNAs as determined from the mapping experiments are shown in Figure 2c.

The context and deamination efficiency of each of the 145 adenosines in these five dsRNAs were used to determine hADAR2 deamination *preferences*. Deamination efficiency, or the percent of the population deaminated at a given site,

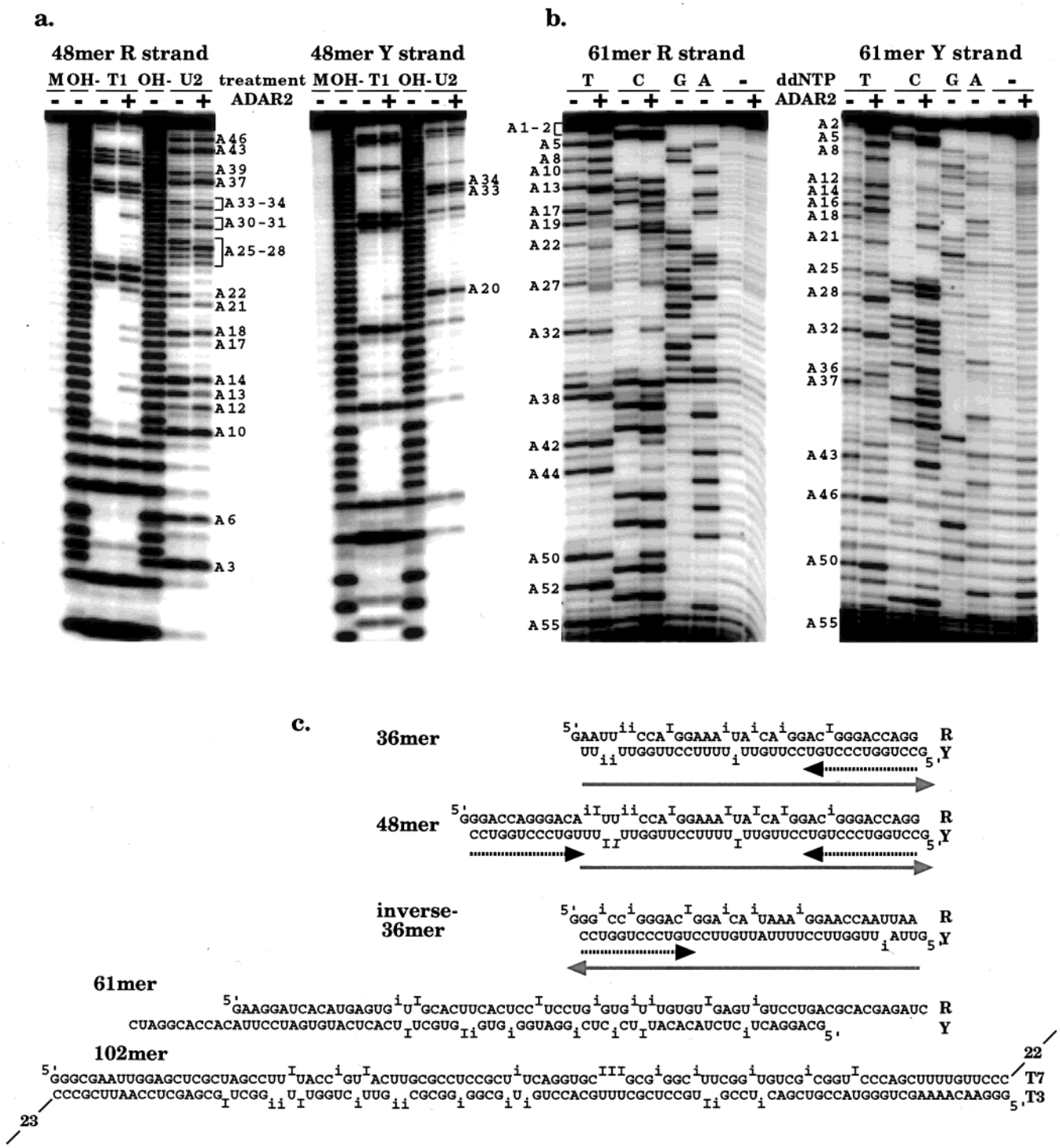

FIGURE 2: Mapping the position of inosines within RNA deaminated to completion with hADAR2. R, purine strand; Y, pyrimidine strand. (a) For nuclease mapping, RNA was reacted with (+) or without (−) hADAR2, and then incubated with (E46Q) RNase T1 (T1) as in Figure 1a, or with RNase U2 (U2). OH-, alkaline hydrolysis ladder; M, mock incubated. (b) For primer extension mapping, RNA was reacted with (+) or without (−) hADAR2, and then incubated with reverse transcriptase (RT), a DNA primer labeled at its 5′ end with ^{32}P, all four dNTPs, and, as indicated, one of the four ddNTPs. A decrease in the intensity of a ddTTP stop and the corresponding increase of a ddCTP stop indicate the position of an inosine. RT products generated from deaminated RNAs have a faster gel mobility due to their altered nucleotide composition (cytidines instead of thymidines). (c) Mapping experiments as shown in (a) and (b) were repeated multiple times (n = 3−5, except for Y strands of 36mer and inverse-36mer where n = 2), quantified (Experimental Procedures), and used to determine the deamination patterns of five RNAs. Minor sites (<50% deaminated) are indicated with i, and major sites (≥50% deaminated) are shown with I (see Experimental Procedures). The italicized *I* in the 48mer Y strand shows an adenosine with a strong and reproducible ADAR2-dependent T1 cleavage but no corresponding drop in U2 cleavage. Lines extending from the T7 and T3 strands of 102mer indicate 3′ overhangs of the indicated length. Arrows highlight regions of 36mer, inverse-36mer, and 48mer that are related in sequence. Solid arrows indicate the directionality of the 36mer sequence, and stippled arrows show a 12 bp GC-rich region that is duplicated in 48mer.

was quantified in nuclease mapping experiments from the change in RNase U2 cleavage, and in primer extension experiments by comparison of ddTTP and ddCTP reverse transcriptase stops (see Experimental Procedures for details on quantitation). Once we determined the deamination efficiency at each adenosine, we grouped adenosines ac-

Table 1: Independent Nearest Neighbor *Preferences*[a]

5′ neighbor	total[b]	% deaminated[c]
A	42	27
U	32	32
G	35[d]	9
C	36	13
3′ neighbor	**total[b]**	**% deaminated[c]**
A	42	11
U	34	27
G	37	29
C	32	13

[a] Since ADAR2 has both 5′ and 3′ neighbor *preferences*, the apparent preference for each neighbor is influenced by the preference for the other neighbor. To ensure that the calculated percent deamination of, for example, 5′ U was not biased due to an unequal sampling of 3′ U, 3′ G, 3′ C, and 3′ A, the percent deamination was averaged from the four constituent triplets (i.e., UAU, UAG, UAC, and UAA from Table 2). Calculation without averaging gave similar values. [b] The total number of adenosines in the given context. [c] The average efficiency of deamination of adenosines in the given context as determined from the mapping data. [d] The total number of GAA triplets includes the 5′-terminal adenosine of 61mer R strand which has a 5′ overhanging guanosine.

Table 2: Triplet *Preferences*[a]

triplet	total[b]	% deaminated[c]	type
UAU	9	47	I
AAG	11	44	
UAG	6	42	
AAU	12	30	
CAG	12	23	II
AAC	8	23	
UAC	6	21	
UAA	11	17	
CAU	6	17	
GAU	7	14	
GAG	8	13	
AAA	11	11	
CAA	11	9	III
GAA	9[d]	6	
GAC	11	5	
CAC	7	4	

[a] The observed percent deamination for the AAG, AAU, CAG, and GAA triplets may be slightly lowered due to proximity to helix ends in some RNAs; exclusion of adenosines near helix ends in our calculations gave similar values: AAG, 49%; AAU, 36%; CAG, 25%; GAA, 7%. [b−d]Defined as in Table 1.

Table 3: Proximity to Helix Termini *Preferences*

distance from 5′ end (nt)	total[a]	% deaminated[a]	distance from 3′ end (nt)	total[a]	% deaminated[a]
1	2	0	1	1	0
2	3	0	2	1	0
3	3	8	3	3	7
4	2	10	4	1	19
5	4	10	5	1	0
6	4	16	6	4	0
7	2	0	7	1	25
8	3	8	8	1	0
9	2	0	9	2	0
10	5	17	10	4	25
11	0	0	11	1	0
12	3	20	12	3	17
13	4	12	13	2	17
14	2	33	14	1	0
15	3	0	15	5	19

[a] Defined as in Table 1.

cording to factors such as sequence context or location within the helix. The average deamination efficiency of adenosines within the same sequence context is given as "% deamination" within Table 1. The percent deamination of adenosines within one context relative to the percent deamination of adenosines within other contexts reveals ADAR2 *preferences*. For example, a high value for "% deamination" in Table 1 indicates a preferred ADAR2 context.

We first grouped our data according to the 5′ and 3′ neighbor of each adenosine (Table 1). We found that hADAR2 has a 5′ neighbor preference of U ≈ A > C = G, which is similar to the 5′ neighbor preference of ADAR1 [U = A > C > G; (*12*)]. Unlike ADAR1, however, ADAR2 also has a 3′ neighbor preference (U = G > C = A). These similar, but distinct, neighbor *preferences* are consistent with the overlapping deamination patterns observed for these enzymes (Figure 1). Since ADAR2 has both a 5′ and a 3′ neighbor preference, we analyzed ADAR2 deamination efficiency for the 16 possible trinucleotide sequences (triplets) that include a central adenosine (NAX; Table 2). The most preferred triplets for ADAR2 deamination are UAU, AAG, UAG, and AAU (referred to as "Type I"). Since the triplet *preferences* reflect the combined preference for the 5′ and 3′ neighbors (compare Tables 1 and 2), our data suggest that the nearest neighbors equally influence whether an adenosine will be preferentially deaminated by ADAR2.

ADAR1 does not efficiently deaminate adenosines within 8 nt of a 3′ helix end nor within 3 nt of a 5′ helix end (*12*). To determine if ADAR2 similarly disfavors adenosines located near helix termini, we calculated the efficiency of ADAR2 deamination near 5′ and 3′ termini (Table 3). In the five RNAs tested, ADAR2 did not deaminate adenosines located 1−2 nt of either helix end. Like ADAR1, ADAR2 disfavored adenosines that were near either blunt ends or helix ends that had single-stranded overhangs, suggesting that ADAR2 is sensitive to the end of the double-stranded region rather than the end of an RNA strand. In many cases, these adenosines were in preferred triplets (Table 2), suggesting that they were disfavored for deamination due to their proximity to a helix end rather than their sequence context. Central positions that show 0% deamination consist of adenosines in Type II or Type III triplets, with the following exceptions: 7 nt from the 5′ end, and 5 or 8 nt from the 3′ end (Table 3). Since adenosines that are closer to helix termini than these three adenosines are deaminated by ADAR2, we believe these positions are not deaminated because the reaction had reached completion before these adenosines were targeted, and not due to their proximity to a helix end. Although our data suggest that adenosines near helix termini are disfavored for deamination, the data set is small. Thus, we cannot exclude the possibility that other positions are similiarly disfavored for ADAR2 deamination, or that ADAR2 may be able to deaminate terminal adenosines in certain cases.

For ADAR1, the fraction of adenosines deaminated within a helix is related to the stability, effective length, and structure of the RNA duplex, suggesting that characteristics of the RNA substrate determine deamination *selectivity* (*12*, *14*). Accordingly, we predicted that ADAR1 and ADAR2 would deaminate the same fraction of adenosines within an RNA, even though the choice of adenosines would differ

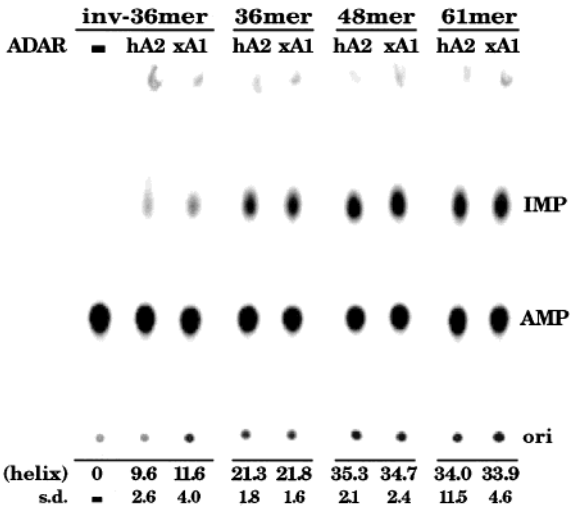

FIGURE 3: ADAR1 and ADAR2 deaminate RNAs with the same *selectivity*. All RNAs were deaminated with ADAR1 or ADAR2 to reaction completion as described (Figure 1a legend). The TLC plate shows the relative amount of 5′ AMP and 5′ IMP at the 3 h time point for each of the four RNAs, with the percent inosine: [IMP/(IMP + AMP + ori)] × 100 ± standard deviation (s.d.) shown beneath the plate ($n \geq 3$). Since ADARs specifically target adenosines within double-stranded regions of RNA, our calculations of % IMP in 61mer exclude adenosines within the single-stranded overhangs; we thus refer to the % IMP (helix) in the figure to indicate this. Calculations that include all adenosines give a value of 28% deamination for both enzymes. hA2, hADAR2; xA1, xADAR1; ori, origin.

(Figure 1). To test this, we radiolabeled four dsRNAs (inv-36mer, 61mer, 36mer, and 48mer) at every adenosine, and then deaminated them to completion with either hADAR2 or xADAR1. The RNAs were purified and digested to 5′ NMPs using nuclease P1, and then the nucleotides were separated on a TLC plate (Figure 3). For each of the four dsRNAs, the percentage of 5′ AMP converted to 5′ IMP was the same regardless of which enzyme was used. Given the differences between xADAR1 and hADAR2, which are only 41% identical in their C-terminal catalytic domains, this result is rather remarkable and emphasizes that the RNA substrate, rather than the enzyme, dictates *selectivity*.

DISCUSSION

At present we know of two intrinsic properties of the ADARs that contribute to their specificity. First, ADARs have deamination *preferences*, and more frequently target adenosines with certain nearest neighbors, that are a certain distance from the helix termini. Second, ADARs exhibit *selectivity*, and stop reacting before all adenosines in preferred context are deaminated; the precise end-point of a reaction is dictated by features of the RNA substrate, such as its length and stability. We characterized the *preferences* and *selectivity* of ADAR2 by analyzing deamination patterns on five dsRNAs. Comparison of the specificity of ADAR2 to that of ADAR1 (*12*) revealed similarities as well as differences. We detected no difference in the specificities of the *Xenopus* and human ADAR1 enzymes, suggesting that these enzymes are true homologues.

ADAR2 Deamination Specificity. The "% deaminated" values for various triplet contexts (Table 2) reflect the efficiency of deamination, on average, of each adenosine appearing in that triplet sequence in the five RNAs studied here. We found that the ADAR2 preference for a given triplet sequence reflected its combined preference for the 5′ and 3′ nearest neighbors (Tables 1 and 2). For example, adenosines with conflicting neighbor *preferences* (Type II; e.g., U<u>A</u>C) were deaminated with an efficiency intermediate to that of adenosines having two preferred neighbors (Type I) and two disfavored neighbors (Type III). Our analysis of nearest neighbor *preferences* suggests that neither neighbor is more important than the other in determining ADAR2 deamination efficiency. Since previous analyses of xADAR1 reaction intermediates (*12*) indicate that adenosines in the most preferred sequence contexts are deaminated first, we predict that when reaction time or enzyme is limiting, deamination by ADAR2 will occur predominantly within the most preferred triplets.

Although we do not know the basis of the ADAR2 neighbor *preferences*, *preferences* for secondary residues often indicate that an enzyme has a subsite that functions to increase specificity. Both RNases U2 and T1 have 3′ neighbor preferences that are attributed to subsite specificity (*24, 25*). The preference of both ADAR1 and ADAR2 for a 5′ A or U may indicate an ADAR subsite interaction, possibly one that disfavors the N2 amino group of G, which is the only unique functional group in the narrow A-form minor groove (*26*). It has been proposed that ADARs access target adenosines by flipping them out of the helix (*12, 27*), which might be influenced by stacking interactions with 5′ and 3′ neighbors (*28*). We considered the possibility that if ADAR2 was less adept than ADAR1 at base-flipping, it might have greater restrictions on the nearest neighbors of a target adenosine. However, we found no obvious correlation between preferred triplets (Table 2) and a lower predicted thermodynamic stability of these sequences (*29*).

In the five dsRNAs tested, ADAR2 did not deaminate adenosines within 2 nt of either helix terminus (Table 3). However, work on minimal substrates of the gluR-B R/G hairpin in vitro indicates that ADAR2 can edit the R/G site when located at the 5′ end of the helix (*16, 30*). This discrepancy may reflect the fact that the number of termini sampled in our study was small and may not account for certain cases. Since the structures of the minimal substrates have not been analyzed, it is also possible that the RNA is sampling alternate conformers that promote editing. Surprisingly, approximately half of the region thought to comprise the initial ADAR2 binding site on the wild-type hairpin (*31*) was removed in the minimal substrate constructs. Clearly, further studies are required to resolve these discrepancies.

Specificities of ADAR1 and ADAR2, and Biological Editing Sites. The specificities of ADAR1 and ADAR2 are summarized in Table 4. Both ADARs have a 5′ neighbor preference, and ADAR2 also has a 3′ neighbor preference. Although both ADARs seem to disfavor adenosines near helix termini, ADAR1 disfavors adenosines within a larger region near 3′ termini than ADAR2. This difference could indicate that a larger area is required for ADAR1 binding, consistent with the fact that this enzyme has three dsRBMs, while ADAR2 has only two. Strikingly, even with the differences in their deamination *preferences*, both enzymes will continue to deaminate an RNA until the same fraction of adenosines is deaminated (Figure 3).

Both ADARs will preferentially deaminate an adenosine that has a 5′ A or U, and a 3′ U or G. Interestingly, this is

Table 4: Comparison of the Specificities of ADAR1 and ADAR2

specificity determinant	ADAR1[a]	ADAR2[b]
5′ nearest neighbor	U=A>C>G	U≈A>C=G
3′ nearest neighbor	no apparent preference	G=U>C=A
preferred triplets	no apparent preference	UAU, AAG, UAG, AAU (30−50% inosine)
proximity to helix termini	within 3 nt of 5′ end or 8 nt of 3′ end not preferred	within 2 nt of either end may not be preferred
deamination *selectivity*	related to the stability or effective length of duplex	the same as ADAR1 on all RNAs tested

[a] (*12, 14*). [b] This work.

Table 5: ADAR *Preferences* and Sequence Context of Editing Sites

editing site	triplet	ADAR1 preference[a]	ADAR2 preference[a]	previous prediction	ref
gluR-B R/G site	AAG	+++	+++	ADAR1, ADAR2	*b*
gluR-B Q/R site	CAG	++	++	ADAR2	*b*
HDV amber/W site	UAG	+++	+++	ADAR1, ADAR2	*c*
serotonin-2C A site	AAU	+++	+++	ADAR1, ADAR2	*d*
serotonin-2C B site	UAC	+++	++	ADAR1	*d*
serotonin-2C C site	AAU	+++	+++	ADAR1, ADAR2	*d*
serotonin-2C D site	UAU	+++	+++	ADAR2	*d*

[a] +++, highly preferred context (ADAR1, 5′ A or U; ADAR2, type I triplet); ++, preferred context (ADAR1, 5′ C; ADAR2, type II triplet). [b] (*10, 16*). [c] (*40*) and Ley, H. L., III, and Bass, B. L., personal communication. [d] (*15, 32*).

the sequence context of most of the mammalian editing sites identified thus far (Table 5). In some cases in vitro editing assays suggest the best candidate (see "previous prediction", Table 5), and these predictions have been substantiated by a recent report (*41*). The ADAR previously proposed to deaminate various editing sites correlates well, in general, with the deamination specificities of the two enzymes (Table 4). For instance, consistent with observations from in vitro studies (*15*), the serotonin receptor A and C sites are in the most preferred context for either ADAR (AAU triplet), whereas the B site is in a better context for ADAR1 (UAC triplet). However, it is unclear why the gluR-B Q/R site (CAG triplet) is efficiently deaminated by ADAR2 in vitro, but very inefficiently by ADAR1, since our studies predict this site would be only moderately deaminated by either ADAR (Table 5). Undoubtedly, there are still aspects of ADAR specificity that we do not understand. Possibly, ADAR2 preferentially binds to a discrete position near the Q/R site that favors efficient deamination of this site. Although ADARs will bind to many different sequences, preference for a specific site has been demonstrated for ADAR2 binding to the gluR-B R/G site helix (*31*). Finally, although ADAR expression levels have been studied in certain tissues (*3, 8, 10, 32*), the relative expression of ADARs is largely unknown, and could play a role in which ADAR deaminates a given RNA in vivo.

Model: Selectivity, ADAR Binding, and Deamination Efficiency. On 36mer, ADAR1 and ADAR2 reactions continue until 21−22% of the adenosines are deaminated (Figure 3), yet the deamination patterns are very different (Figure 1a,b). ADAR1 targets 6 adenosines, whereas ADAR2 targets 10 adenosines. In contrast, 48mer reacted with ADAR1 or ADAR2 contains ∼35% inosine (Figure 3), and is targeted at 12−15 sites by each enzyme (Figure 1c). Stated simply, ADAR2 creates a more heterogeneous 36mer population than ADAR1, but both enzymes create similar 48mer populations. The number of different adenosines targeted by ADAR1 or ADAR2 cannot be attributed solely to the *preferences* of these enzymes. ADAR2 deaminates, with varying efficiency, 70−80% of the adenosines in Type I and Type II triplets that are at preferred distances from termini, in both the 36mer and 48mer RNAs. In contrast, while ADAR1 targets 79% of adenosines with a 5′ A, U, or C in a preferred location in 48mer, this enzyme targets only 46% of its preferred adenosines in 36mer.

We believe these differences may be related to differences in the binding site size of ADAR1 versus ADAR2, and the limited number of binding sites that exist on 36mer compared to 48mer. Like all proteins that bind polymeric substrates without sequence specificity, an increase in the length of the polymer (i.e., dsRNA) results in a corresponding increase in the number of binding sites it contains, and the latter depends on the number of monomers (i.e., base pairs) with which each protein interacts (*33*). As shown in Figure 4, since 36mer is deaminated throughout its length by ADAR2, this enzyme likely binds 36mer in multiple ways that position preferred adenosines near its catalytic active site. Although 36mer contains many adenosines in apparently good context for ADAR1 deamination, only a subset are targeted by ADAR1, and these are clustered near the 5′ end of the 36mer R strand (Figure 1b). Our data suggest that ADAR1 is more limited than ADAR2 in the number of ways that it can bind 36mer, and this could mean that ADAR1 has a larger binding site.

Deamination of an adenosine to an inosine within dsRNA converts an AU pair to a mismatched IU pair, and this is thought to be the basis of *selectivity*. The insertion of IU pairs increases the single-strandedness of the RNA, and decreases its thermodynamic stability (*34, 35*). Thus, the reaction stops before all adenosines are deaminated because the substrate is no longer recognized as double-stranded. Since *selectivity* dictates that deamination continues until a certain number of IU pairs exists within each molecule, decreasing the number of ADAR binding sites would produce a population of reacted molecules that showed more efficient deamination at fewer sites (Figure 4). In 36mer, 4 of the 6 ADAR1 sites are ≥50% deaminated, whereas only 2 of the 10 ADAR2 sites are ≥50% deaminated (Figure 1b). The correlation between the number of ADAR binding sites and deamination efficiency at a particular adenosine has been observed on the gluR-B R/G site hairpin (*31*). The native R/G hairpin contains three mismatches, one of which contains

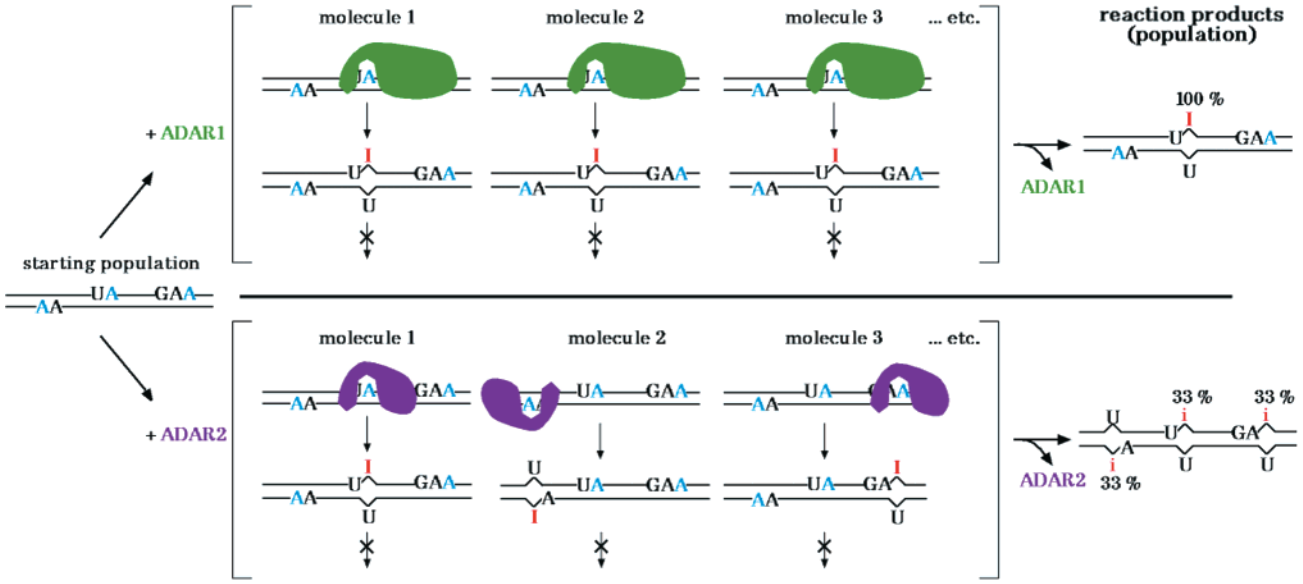

FIGURE 4: A model: how to get efficient and specific deamination. The model proposes that restriction of ADAR binding to a limited number of sites decreases the number of adenosines targeted for deamination, while increasing the efficiency of deamination at targeted sites. The starting population consists of an RNA substrate that contains three adenosines (blue) in preferred context for either ADAR; the continuing RNA strands are indicated with lines. In brackets, the substrate is shown reacting with ADAR1 (large, green) or ADAR2 (small, purple). For simplicity, the *selectivity* of deamination on this RNA is such that each dsRNA is deaminated at only one site before the reaction is over (arrow marked by X). Because of the large size of ADAR1, and the short length of the substrate, ADAR1 is shown with only one binding site that gives access to a preferred adenosine. Since one deamination event is required to stop the hypothetical reaction, this adenosine is deaminated by ADAR1 with 100% efficiency. In contrast, ADAR2 has more binding sites on this RNA and is able to deaminate each of the preferred adenosines equally well. Since the reaction stops after a single deamination event, all three reaction products are represented in a fraction of the population, i.e., with low efficiency (33%). Note that the *selectivity*, or the fraction of adenosines deaminated in the population at complete reaction, is the same for the ADAR1 (1 site at 100%) and ADAR2 (3 sites at 33%) reactions. Within the reaction product populations (right), "i" and "I" (red) indicate minor (<50%) and major (≥50%) deamination sites, respectively. The relative size of the ADARs is proportional to their molecular weights, and are shaped to reflect their deamination *preferences*.

the R/G site adenosine. At low concentrations of ADAR2, the mismatches restrict binding to one primary position that encompasses the R/G site, resulting in 80% R/G editing. Pairing of the two distal mismatches permits binding over the length of the hairpin stem, and reduces R/G editing by half. While mismatches restrict binding on the R/G hairpin, ADAR1 binding may be inherently more restricted than ADAR2 due to a larger binding site size. In fact, ADAR2 may require mismatches to restrict its binding on even short RNAs, and more stringent *preferences* to limit deamination to fewer sites. Interestingly, the crystal structure of a single dsRBM from Xlrbpa in complex with dsRNA shows binding across two minor grooves and one widened major groove (*36*). Although the widened major groove occurs at the interface of two coaxially stacked helices in the structure, mismatches could widen the major groove in a continuous helix to direct dsRBP binding to discrete positions.

Perspectives. The interaction of a dsRBM with its dsRNA substrate occurs largely within the RNA minor groove (*36*), where the functional groups available for base pair discrimination are limited to the N2 amino group of guanosine (*26*, *37*). Although, in theory, this precludes many sequence-specific interactions of dsRBPs with their substrates, RNase III shows sequence preferences in binding (*38*), and ADARs exhibit sequence *preferences* in their catalytic step, that is, in which adenosines they target for deamination [(*1*, *12*, *39*) and Table 1]. Interestingly, all ADARs characterized to date demonstrate similar nearest neighbor *preferences* with regard to their deamination sites, suggesting that neighbor *preferences* derive from some intrinsic feature of the catalytic mechanism. For instance, a *Drosophila melanogaster* ADAR activity has a 5′ neighbor preference of A, a 3′ neighbor preference of G, and disfavors adenosines near helix termini (*39*). Analysis of in vivo edited *C. elegans* ADAR substrates indicates a 5′ neighbor preference of U > A > C > G (*1*).

Hopefully future studies will provide information on the mechanism and structure of ADARs that leads to an understanding of why ADARs exhibit *preferences* and *selectivity*. Regardless, the work described here facilitates predictions of which ADAR deaminates a given editing site in vivo and adds to our understanding of how efficient deamination of only a few adenosines, as occurs during RNA editing, is achieved by ADARs.

ACKNOWLEDGMENT

We thank members of our research group for helpful discussions, and D. Morse and M. Macbeth for critically reading the manuscript. Plasmid pJEL167 and yeast strain BCY123 were generous gifts from J. Lindsley and B. Cairns, respectively, both from the University of Utah. We thank A. Lingam for providing excellent technical assistance, and R. Hough and H. L. Ley III for the ADAR1 used in enzyme comparison studies, and their helpful advice and assistance.

REFERENCES

1. Morse, D. P., and Bass, B. L. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 6048−6053.
2. Bass, B. L. (1997) *Trends Biochem. Sci. 22*, 157−162.
3. Paul, M. S., and Bass, B. L. (1998) *EMBO J. 17*, 1120−1127.
4. Bass, B. L., and Weintraub, H. (1987) *Cell 48*, 607−613.
5. Rebagliati, M. R., and Melton, D. A. (1987) *Cell 48*, 599−605.

6. St. Johnston, D., Brown, N. H., Gall, J. G., and Jantsch, M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 10979−10983.
7. Bass, B. L., Hurst, S. R., and Singer, J. D. (1994) *Curr. Biol. 4*, 301−314.
8. Kim, U., Wang, Y., Sanford, T., Zeng, Y., and Nishikura, K. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11457−11461.
9. O'Connell, M. A., Krause, S., Higuchi, M., Hsuan, J. J., Totty, N. F., Jenny, A., and Keller, W. (1995) *Mol. Cell. Biol. 15*, 1389−1397.
10. Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P. H., and Higuchi, M. (1996) *Nature 379*, 460−464.
11. O'Connell, M. A., Gerber, A., and Keller, W. (1997) *J. Biol. Chem. 272*, 473−478.
12. Polson, A. G., and Bass, B. L. (1994) *EMBO J. 13*, 5701−5711.
13. Nishikura, K., Yoo, C., Kim, U., Murray, J. M., Estes, P. A., Cash, F. E., and Liebhaber, S. A. (1991) *EMBO J. 10*, 3523−3532.
14. Lehmann, K. A., and Bass, B. L. (1999) *J. Mol. Biol. 291*, 1−13.
15. Burns, C. M., Chu, H., Rueter, S. M., Hutchinson, L. K., Canton, H., Sanders-Bush, E., and Emeson, R. B. (1997) *Nature 387*, 303−308.
16. Yang, J.-H., Sklar, P., Axel, R., and Maniatis, T. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 4354−4359.
17. Giaever, G. N., Snyder, L., and Wang, J. C. (1988) *Biophys. Chem. 29*, 7−15.
18. Guo, Z., and Sherman, F. (1996) *Mol. Cell. Biol. 16*, 2772−2776.
19. Gerber, A., O'Connell, M. A., and Keller, W. (1997) *RNA 3*, 453−463.
20. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
21. Hough, R. F., and Bass, B. L. (1994) *J. Biol. Chem. 269*, 9933−9939.
22. Steyaert, J., Opsomer, C., Wyns, L., and Stanssens, P. (1991) *Biochemistry 30*, 494−499.
23. Inoue, T., and Cech, T. R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 648−652.
24. Yasuda, T., and Inoue, Y. (1982) *Biochemistry 21*, 364−369.
25. Steyaert, J., Wyns, L., and Stanssens, P. (1991) *Biochemistry 30*, 8661−8665.
26. Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804−808.
27. Hough, R. F., and Bass, B. L. (1997) *RNA 3*, 356−370.
28. Klimasauskas, S., and Roberts, R. J. (1995) *Nucleic Acids Res. 23*, 1388−1395.
29. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) *Biochemistry 37*, 14719−14735.
30. Yi-Brunozzi, H. Y., Easterwood, L. M., Kamilar, G. M., and Beal, P. A. (1999) *Nucleic Acids Res. 27*, 2912−2917.
31. Öhman, M., Källman, A. M., and Bass, B. L. (2000) *RNA 6*, 687−697.
32. Liu, Y., Emeson, R. B., and Samuel, C. E. (1999) *J. Biol. Chem. 274*, 18351−18358.
33. McGhee, J. D., and von Hippel, P. H. (1974) *J. Mol. Biol. 86*, 469−489.
34. Bass, B. L., and Weintraub, H. (1988) *Cell 55*, 1089−1098.
35. Strobel, S. A., Cech, T. R., Usman, N., and Beigelman, L. (1994) *Biochemistry 33*, 13824−13835.
36. Ryter, J. M., and Schultz, S. C. (1998) *EMBO J. 17*, 7505−7513.
37. Frugier, M., and Schimmel, P. (1997) *Proc. Natl. Acad. Sci. U.S.A 94*, 11291−11294.
38. Zhang, K., and Nicholson, A. W. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 13437−13441.
39. Petschek, J. P., Mermer, M. J., Scheckelhoff, M. R., Simone, A. A., and Vaughn, J. C. (1996) *J. Mol. Biol. 259*, 885−890.
40. Polson, A. G., Bass, B. L., and Casey, J. L. (1996) *Nature 380*, 454−456.
41. Higuchi, M., Maas, S., Single, F. N., Hartner, J., Rozov, A., Burnashev, N., Feldmeyer, D., Sprengel, R., and Seeburg, P. H. (2000) *Nature 406*, 78−81.

BI001383G