

Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G

Priyanga Wijesinghe and Ashok S. Bhagwat*

Department of Chemistry, Wayne State University, Detroit, MI 48202, USA

Received May 1, 2012; Revised and Accepted June 20, 2012

ABSTRACT

The AID/APOBEC family of enzymes in higher vertebrates converts cytosines in DNA or RNA to uracil. They play a role in antibody maturation and innate immunity against viruses, and have also been implicated in the demethylation of DNA during early embryogenesis. This is based in part on reported ability of activation-induced deaminase (AID) to deaminate 5-methylcytosines (5mC) to thymine. We have reexamined this possibility for AID and two members of human APOBEC3 family using a novel genetic system in *Escherichia coli*. Our results show that while all three genes show strong ability to convert C to U, only APOBEC3A is an efficient deaminator of 5mC. To confirm this, APOBEC3A was purified partially and used in an *in vitro* deamination assay. We found that APOBEC3A can deaminate 5mC efficiently and this activity is comparable to its C to U deamination activity. When the DNA-binding segment of AID was replaced with the corresponding segment from APOBEC3A, the resulting hybrid had much higher ability to convert 5mC to T in the genetic assay. These and other results suggest that the human AID deaminates 5mC's only weakly because the 5-methyl group fits poorly in its DNA-binding pocket.

INTRODUCTION

Deamination of cytosines in DNA to uracil has emerged as a major mechanism by which higher vertebrates protect themselves against infections. In vertebrates the enzymes that can perform this reaction is the AID/APOBEC family and one member of this family, activation-induced deaminase (AID), has an essential role in the maturation

of antibodies. It diversifies the antibody repertoire by causing heavy mutagenesis of the variable segment of the rearranged antibody gene (called somatic hypermutation, SHM) or by promoting gene conversion between the variable segment and a pseudo-V segment. Additionally, AID is required for class-switch recombination (CSR) which replaces the μ constant segment of the immunoglobulin gene with other constant segments [Reviewed in (1–3)]. AID is also required for the translocation of *c-myc* gene to the immunoglobulin locus (4) and is implicated in the development of many cancers (5).

Among the APOBECs (APOBEC1 through APOBEC4), only APOBEC3 appears to have a protective immunity function. The APOBEC3s from a number of animals have been shown to protect cells against a number of viruses and to inhibit retrotransposition of chromosomal retroelements. APOBEC3s accomplish this through multiple mechanisms that include high level mutagenesis, strand breakage, inhibition of reverse transcription and packaging of the viral genomes. The human genome codes for seven paralogs of APOBEC3 (APOBEC3A through APOBEC3H) which contain either one or two zinc-binding motifs and in all cases the motif near the carboxy-terminus of the protein has cytosine deamination activity. In contrast non-primates contain a single APOBEC3 gene (6–8).

Morgan *et al.* (9) reported that purified human AID and rat APOBEC1 had the ability to deaminate 5-methylcytosines (5mC) in DNA oligomers and expression of AID in *Escherichia coli* also expressing the SssI methyltransferase (MTase) increased C to T mutations at a methylated cytosine in the *rpoB* gene. Additionally, they reported detection of expression of AID (and to a lesser extent APOBEC1) in oocytes, embryonic stem (ES) cells and other pluripotent tissues. Based on these results Morgan *et al.* (9) proposed that AID plays a role in epigenetic reprogramming in non-lymphoid tissues such as fertilized eggs and ES cells by causing demethylation of DNA. In its simplest form this would occur through

*To whom correspondence should be addressed. Tel: +313 577 2549; Fax: +313 577 8822; Email: axb@chem.wayne.edu

deamination of 5mC to T by AID followed by repair of the resulting T•G mispair by base-excision repair (BER) to C:G (10,11).

In subsequent studies AID/APOBEC genes were transfected into cells and the effect of their expression on DNA demethylation was studied. In Zebra fish embryos, introduction of AID, APOBEC2a or APOBEC2b resulted in DNA demethylation, and the DNA glycosylase MBD4 enhanced this effect (12). Bhutani *et al.* (13) showed that during the reprogramming of human cells to induced pluripotency, siRNA-mediated inhibition of AID expression resulted in the remethylation of OCT4 and NANOG gene promoters and loss of expression of the genes. Popp *et al.* (14) found that primordial germ cells have significantly higher methylation at many genomic loci in AID^{-/-} mice than in wild-type mice. Another suggested pathway for DNA demethylation involves conversion of 5mC to 5-hydroxymethylcytosine (5hmC) by Tet (Ten–eleven translocation) proteins (15), deamination of 5hmC by AID and the repair of subsequent 5-hydroxymethyluracil (5hmU)-guanine mispair by MBD4 or thymine–DNA glycosylase [TDG; (16)]. Guo *et al.* (17) showed that transfection of human embryonic kidney cells expressing Tet1 with AID gene reduced the level of genomic 5hmC. They also found that transfection of murine APOBEC1, human APOBEC2, APOBEC3A, APOBEC3C or APOBEC3E, but not APOBEC3B or APOBEC3G also resulted in significant reduction in genomic 5hmC (17). These results have led to the hypothesis that one or more of the AID/APOBEC family of proteins may participate in the demethylation pathways via deamination of 5mC and 5hmC (10,11,18).

Although many of these experiments are based on the assumption that AID will efficiently deaminate 5mC or 5hmC, there is significant evidence that shows that AID is not a good deaminator of 5mC. When AID tagged with glutathione-S-transferase (GST) was expressed and purified from insect cells and used *in vitro* in deamination reactions, 5mC was a substantially poorer substrate than C (19,20). In one study, the rate of deamination of C in DNA oligomers was found to be ten times the rate of deamination of 5mC (20). Additionally, Kohli *et al.* (21) reported that when oligomers with both C's and 5mC's were treated with AID tagged with maltose-binding protein (MBP), C's were deaminated much more efficiently than 5mC's. However, the principal difficulty in making biochemically valid statements about AID is that the protein is 'sticky' and aggregates quickly after purification. Consequently, no studies have reported k_{cat}/K_m for AID with any substrate and quantitative assessment of substrate preferences of AID is quite difficult.

We decided to take a fresh approach to this problem and developed a simple genetic system in *E. coli* that can quantify deamination of 5mC or C in the same sequence context in genomic DNA. When human AID and two APOBEC3 genes were tested using this system, we found that while APOBEC3A was a strong deaminator of both C and 5mC, AID and APOBEC3G were much weaker in their ability to deaminate this modified base.

MATERIALS AND METHODS

Bacterial strains and plasmids

The *kan* alleles were introduced in the *E. coli* K-12 strain BH143 genome [$\Delta(mrr-hsdRMS-mcrBC)$ *mcrA* Φ 80*dlacZ* Δ M15 Δ *lacX74* *deoR* *endA1* *araD139* Δ (*ara, leu*)7697 *galU* *galK* *rpsL* *nupG* Δ (*dcm-vsr*)] through recombineering using the Red/ET recombination system from Gene Bridges (Heidelberg, Germany). The *kan* alleles in the plasmids pUP31, pUP41 and pUP44 (22) were amplified using the 70-nt primer pairs (Supplementary Table S1) each of which contained 50 nt identical to the *manX* gene in the chromosome. The amplification products contained the wild-type bleomycin-resistance gene in addition to the *kan* alleles and the recombinants were selected using zeocin (20 μ g/ml) in plates. The recombinants were confirmed by DNA sequencing and the three new strains were named BH400 (from pUP31), BH300 (pUP41) and BH500 (pUP44). *Escherichia coli* B strain BL21DE3 was used for protein expression and purification.

The pBR322-based plasmids carrying genes for M.HpaII [pM.HpaII, (23)], Dcm [pDCM21, (24)] and M.MspI [pQ8, (25)] have been described before. The plasmid pDCM22 contains *dcm*⁺ and *vsr*⁺ genes and has also been described (24). Human AID (26), APOBEC3A(A3A) and APOBEC3G (A3G) genes were cloned into pACYC184-based plasmid pSU24 creating respectively pSUAID, pSUA3A and pSUA3G. The clones for A3A cDNA were kindly provided by Reuben Harris (University of Minnesota) and A3G cDNA was obtained from ATCC (Manassas, VA). The primers used for the cloning are listed in Supplementary Table S1. The gene for UGI was amplified from a plasmid kindly provided by Umesh Varshney (Indian Institute of Science, Bangalore, India) and inserted at an EcoRI site in pSUAID to create pSUAID-UGI. Catalytic mutants of AID (AID-E58A) and APOBEC3A (A3A-E72A) were constructed using a whole plasmid PCR mutagenesis strategy [(27), Supplementary Table S1]. For the purification of A3A protein, both the A3A gene and the catalytic mutant A3A-E72A were amplified (Supplementary Table S1) and cloned into pET28a (+) as EcoRI-XhoI fragments. The hybrid gene AID-A3AR2 was synthesized by DNA 2.0 (Menlo Park, CA) and cloned into the pSU24 vector.

Kanamycin-resistance reversion assay

The reversion assay has been described previously (23,28). To quantify 5mC to T deaminations, AID, AID-E58A, AID-A3AR2, A3A, A3A-E72A or A3G and one of the MTase genes were coexpressed from compatible plasmids and kanamycin-resistant revertants (50 μ g/ml, phenotype-Kan^R) were scored. Kan^R revertant frequency is the ratio (Number of Kan^R revertants/Total number of viable cells). To quantify C to U deaminations, either the UGI gene was expressed from the same plasmid as AID or one of the APOBECs, or an *ung* strain was used as host. To study repair of T•G mispairs, the plasmid pDCM21 (*dcm*⁺) or pDCM22 (*dcm*⁺ *vsr*⁺) was introduced in

BH400 and the Kan^R revertant frequencies were determined.

To determine the base change in the *kan* gene, 12 independent revertants were amplified, the products purified using PCR purification kit (Epoch-Life Science) and sequenced (Supplementary Table S1). The sequences of revertants were aligned with the sequences of the *kan* alleles using MacVector software (MacVector, Inc., Cary, NC) and the mutations were identified.

Uracil quantification assay

The uracils in the genomic DNA were quantified as described previously (29). Briefly, the genomic DNA was incubated with methoxyamine to block any pre-existing abasic sites and was treated with *E. coli* UDG (New England Biolabs) and aldehyde-reactive probe (Dojindo Molecular Technologies, Rockville, MD). The DNA was transferred to a nylon membrane (EMD-Millipore, Billerica, MA) and cross linked to the membrane using a UV light. The membrane was incubated with 5×10^{-4} mg/ml of streptavidin-Cy5 (GE Healthcare), it was scanned using a Typhoon 9210 phosphorimager (GE Healthcare) and the images were analyzed with ImageQuant software. The uracil standard was a 75-mer duplex containing a single U•G mispair.

Purification of A3A and its mutant

BL21DE3 cells expressing A3A or A3A-E72A gene in pET28a (+) were grown to mid-log phase and transcription was induced with the addition of IPTG. The cells were harvested after 5 h and broken using the French Pressure Cell Press (Thermo Spectronic) and the cell-free lysate was cleared by centrifugation. The lysate was passed over a Ni-NTA column (Novagen, Madison, WI) and the bound proteins were eluted with 250 mM imidazole. The eluted proteins were dialyzed and concentrated using Amicon Ultra Centrifugal devices (EMD-Millipore, Billerica, MA). The concentrated proteins were equilibrated in the storage buffer (20 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol).

Biochemical assay for C and 5mC deamination

The substrates for the study of C and 5mC deamination, A3A-C and A3A-5mC respectively, are listed in Supplementary Table S2. Six picomoles of oligomer A3A-C was incubated at 37°C with 140 ng of partially purified A3A enzyme in a 10 µl volume in the reaction buffer [40 mM Tris-HCl (pH 7.5), 5 mM EDTA, 1 mM DTT, 40 mM NaCl]. The reaction was terminated by the addition of 1,10-phenanthroline (Sigma-Aldrich) to 5 mM. Two units of *E. coli* UDG (New England Biolabs, Ipswich, MA) were added to the reaction and incubation was continued at 37°C for 1 h. The reactions were stopped by adding NaOH to 0.1 M and heating to 95°C for 7 min. The products were separated on a 15% denaturing acrylamide gel and the gel was scanned using Typhoon 9210 phosphorimager. ImageQuant software was used to quantify the intensities of the substrates and the deaminated products. For 5mC deamination 2 pmol of oligomer A3A-5mC was incubated with 140 ng of

partially purified protein at 37°C for 1 h. The oligomer complement added to the reaction at 3-fold molar excess to create a T•G mismatch. The duplex was incubated with 1.5 units of thermostable thymine DNA glycosylase (Trevigen, Gaithersburg, MD) for 1 h at 47°C and the products were processed and analyzed in a manner similar to C-deamination products.

RESULTS

Construction and validation of genetic system for 5mC deamination

We described previously two defective alleles of the kanamycin-resistance gene (*kan*) that can be methylated in *E. coli* by different MTases and used to quantify 5mC to thymine conversions (23). In this plasmid-based system, the *kan*⁻ alleles revert to *kan*⁺ (phenotype-Kan^R) through 5mC to T deamination increasing the Kan^R frequency by at least an order of magnitude (23). We have expanded and adapted this system for the study of 5mC to T conversion by AID/Apobec family proteins.

To reduce the number of plasmids maintained in the cells, we constructed three strains of *E. coli*, each with a different *kan* allele (22,23) inserted into the *manX* gene using homologous recombination (Supplementary Figure S1). Three different cytosine MTase genes were introduced into these *E. coli* strains (BH300, BH400 and BH500) on plasmids to create five different sequence contexts for cytosine methylation. These included two in which the 5mC was in a CpG sequence context while in one it was in the WRC context (W is A or T, R is purine) preferred by AID (30). The sequence context of 5mCs in the different strains is shown in Figure 1A.

We first confirmed that the chromosomal DNA in these strains was appropriately methylated. The genomic DNA from cells expressing Dcm, M.HpaII or M.MspI MTases was digested respectively with EcoRII, HpaII or MspI and was found to be resistant to the endonucleases (Supplementary Figure S2). We also compared the frequency of spontaneous Kan^R revertants in the presence and absence of MTases. In each case, the frequency of Kan^R revertants increased by a factor of ~100 (Figure 1B) and this result is consistent with the previously reported increases in Kan^R revertant frequencies due to methylation by Dcm and M.HpaII (23,28). The large magnitude of the increase in revertant frequency is due partly to the fact that 5mC deaminates at approximately four times the rate of deamination of C (31) and partly because of the inability of these cells to repair T•G mispairs created by 5mC deamination. *E. coli* lacks DNA glycosylases like MBD4 and TDG that excise thymines from a T•G mispair and the gene for the T•G-specific endonuclease, Vsr, has been deleted from the genomes of the cells used (see 'Materials and Methods' section, and below). We also sequenced 12 independent Kan^R revertants from cells that expressed different MTases and nearly all the revertants had the methylated cytosine changed to thymine (Supplementary Table S3). The near complete absence of mutations at unmethylated cytosines is probably because the

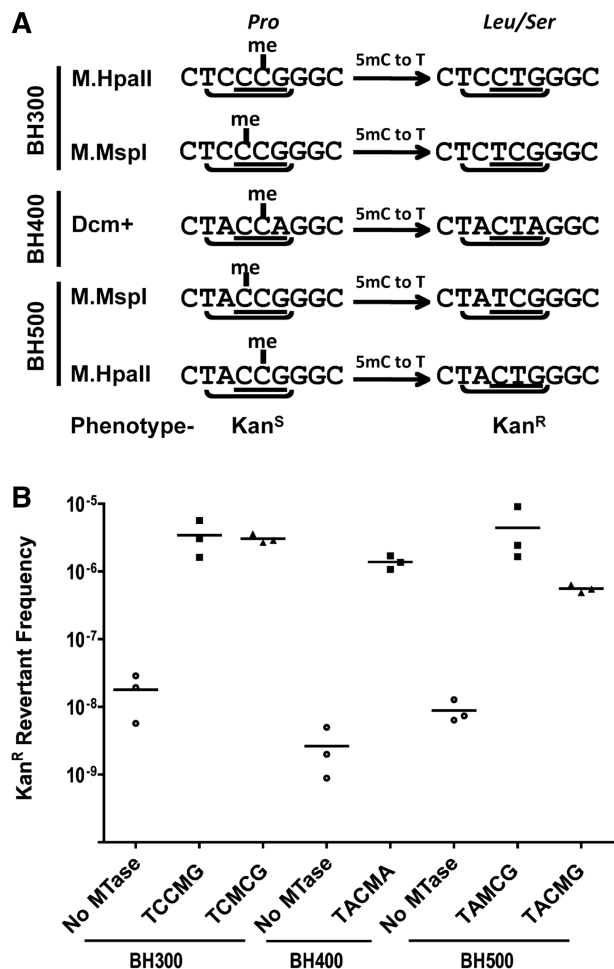


Figure 1. (A) Sequence context of 5mC in *kan* alleles. The five different methylation contexts created in the *kan* gene through changes in base sequence or methylation are shown. Three different *kan* alleles are present in *E. coli* strains BH300, BH400 and BH500. The gene for the MTase M.HpaII, M.MspI or Dcm was introduced in these strains to methylate one of the cytosines in a proline codon (underlined in the figure). The methylated cytosine is indicated by 'me' above the C. The five bases unique to each sequence context are indicated by a bracket below the sequence. 5mC to thymine mutation changes the proline codon to either leucine or serine and changes the cellular phenotype from kanamycin-sensitive (Kan^S) to kanamycin-resistant (Kan^R). (B) Effects of MTases on Kan^R revertant frequency. The frequency of revertants with or without the presence of a MTase is shown. The bacterial host used (BH300, BH400 or BH500) and the relevant methylated sequences are shown. The genes for M.HpaII, M.MspI or Dcm were introduced in the cells to methylate the DNA. 'M' is 5-mC and the horizontal line within the data points is the mean value.

E. coli strains BH300, BH400 and BH500 are proficient in the repair of U•G mismatches created by the deamination of C. These results show that in this genetic system the frequency of Kan^R revertants is an accurate measure of the 5mC to T conversions.

Finally, we also confirmed that a majority of the revertants arose through the conversion of a 5mC:G pair to a T•G mismatch in the case of at least the Dcm MTase. We did this by introducing the gene *vsr*⁺ in the strain BH400 in addition to Dcm. *Vsr* is a T•G-specific endonuclease that hydrolyzes the phosphodiester linkage immediately

upstream of the mismatched T (32) initiating very short-patch (VSP) repair pathway that replaces the T with a C [Figure 2A; (33)]. As expected, when *dcm*⁺ gene alone was introduced in *E. coli*, Kan^R frequency increased by a factor of ~55 compared to the vector control, but when both *dcm*⁺ and *vsr*⁺ genes were introduced in the cells, the increase was reduced to only ~4-fold over the control (Figure 2B). Together these results show that the genetic reversion system used here can be used to quantify 5mC to T deaminations.

Modest deamination of 5mC by human AID

When human AID gene was introduced into BH500 lacking DNA methylation, ~1 in 10⁶ cells became resistant to kanamycin (Figure 3A). However, BH500 is proficient in the repair of U•G and hence bulk of the uracils generated by cytosine deamination are expected to be repaired. To quantify the full extent of cytosine deaminations caused by AID, we coexpressed Ung inhibitor UGI (34) in the cells. Expression of both AID and UGI increased the frequency of Kan^R revertants ~200-fold to >10⁻⁴ (Figure 3A). This result is consistent with previous reports that human AID is a strong deaminator of cytosines in the *E. coli* genome (35).

In contrast, when M.MspI was present in BH500 along with AID, only a modest increase in 5mC deamination was scored by the Kan^R assay. The median frequency of revertants was 10⁻⁵, which was only 1.9-fold higher than the frequency due to empty vector and only slightly higher (1.5-fold) than a catalytically inactive mutant of AID (Figure 3B). It should be noted that the 5mC in this genetic assay was in the WRC sequence context preferred by AID (30) and the overall Kan^R frequency was highest in this sequence context compared to other contexts (Supplementary Figure S3). Consistent with the results in the WRC context, only modest increases in the Kan^R frequency were seen due to AID in these other contexts including two where the methylated base was in a CpG dinucleotide. In no case was the increase in revertant frequency >2-fold compared to the vector control (Supplementary Figure S3A–D).

We wanted to make sure that in cells where AID was a poor deaminator of 5mC, it was still a strong deaminator of cytosines. To establish this, we used a biochemical assay for the quantification of genomic uracils. The assay introduces a fluorescent Cy5 tag at sites in DNA that have uracils, and uses Cy5 fluorescence to quantify the uracils and has been described previously (36–38). The results show that while presence of AID caused 5mC to T conversion to increase <2-fold (Figure 3B), the amount of uracils in the genome of same cells increased ~10-fold (Figure 3C). These results show that while AID acts as a strong deaminator of cytosines, it is poor at deaminating 5mC.

Efficient 5mC deaminations by APOBEC3A, but not APOBEC3G

APOBEC3A (A3A) and APOBEC3G (A3G) are two members of the human APOBEC3 family and are sequence homologs of AID. First, we expressed the

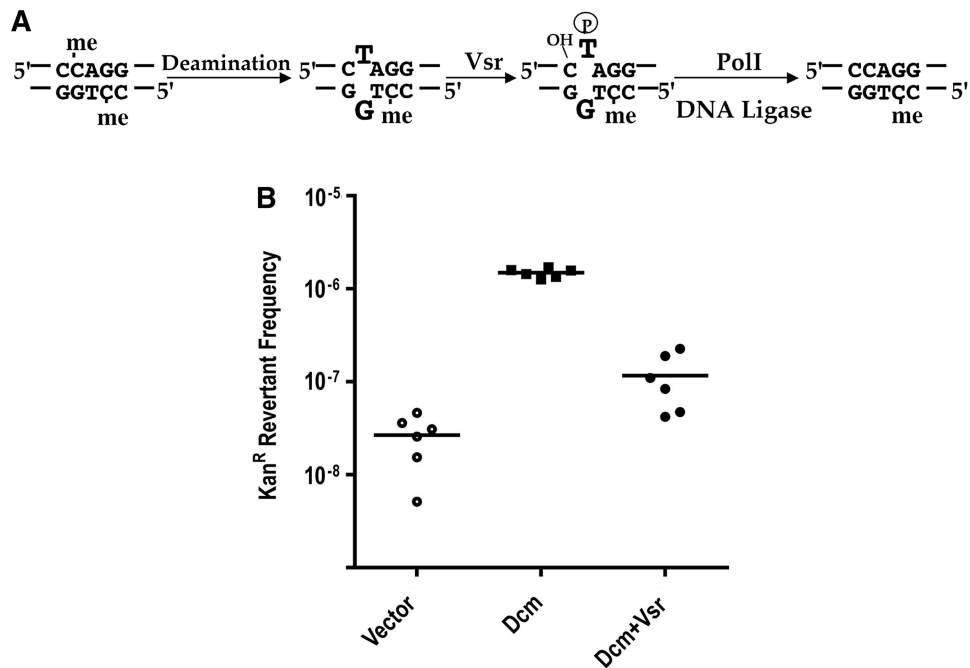


Figure 2. (A) Very short-patch repair of T•G mispairs in *E. coli*. The pathway (33) by which *E. coli* VSP repair replaces T•G mispairs created by the deamination of 5mCs in the Dcm sequence context with C•G is shown. Deamination of one of the 5mCs in a Dcm sequence context creates a T•G mispair. The Vsr endonuclease hydrolyzes the phosphodiester linkage 5' of the mispaired T and DNA polymerase I (PolI) and DNA ligase complete the reaction. (B) Kan^R revertant frequency with and without VSP repair. The revertant frequencies in the presence of Dcm alone or with both Dcm and Vsr in the cells (BH400) are shown and are compared with the frequency in the absence of either enzyme. The horizontal line within the data points is the median value.

full-length A3A and A3G in *ung*⁻ *E. coli* and compared their ability to deaminate cytosine to that of AID. As expected, A3G was significantly better at deaminating cytosines in its preferred sequence context (a run of C's) than AID (Figure 4A). However, when A3G was tested for its ability to deaminate 5mC in the preferred sequence, no increase in Kan^R revertant frequency was detected (Figure 4B). In this sequence context, AID is also inefficient at deaminating 5mC (Figure 4B and Supplementary Figure S3A). In these and other experiments we have been unable to detect any deamination of 5mC by A3G (data not shown).

In contrast, A3A was not only a much stronger mutator than AID when the genetic system was designed to score C to U deaminations (Figure 5A), but also when 5mC to T conversions were scored (Figure 5B). The methylated cytosine in this experiment was in the CpG context and a cytosine flanked the CpG on the 5'-side. We chose this sequence context for experiments with A3A because this enzyme is known to prefer a T or a C on the 5'-side of the target cytosine (39). We also found A3A to increase substantially the revertant frequency when 5mC was in CpH context (H is not G; data not shown). Changing a conserved glutamate residue in A3A expected to be required for catalysis to alanine completely eliminated 5mC to T mutations (Figure 5C). When independent revertants obtained in experiments where cells expressed an MTase along with AID or A3A were sequenced, an overwhelming majority of the mutations were at the methylated cytosines (Supplementary Table S4). This shows that the mutations resulted from a 5mC to T

change. The presence of WT A3A in cells increased the revertant frequency by at least 15-fold compared to empty vector and in some experiments it was much higher (Figure 5C and data not shown). These data suggest that A3A is much more efficient at deaminating 5mC than AID.

5mC to thymine deaminations by APOBEC3A *in vitro*

To confirm this deamination activity biochemically, the human A3A gene was modified at its 3'-end with six His codons and the tagged protein was purified partially over a nickel affinity column. The purified protein was active and was able to completely convert a cytosine in a synthetic oligomer to uracil (Figure 6A, lane 4). Based on the purity of the enzyme (Supplementary Figure S4), we conclude that ~2 pmol of the enzyme completely converted 6 pmol of C's to U's in about one hour. These data suggest that the enzyme acts slowly on the substrate used, but it does turn over.

The partially purified A3A also converted 5mC to T efficiently. To demonstrate this, an oligomer with a single 5mC was treated with this enzyme and an excess of the complementary strand was added to create T•G mispairs at the deaminated 5mCs. This duplex was successively treated with the TDG and NaOH to create strand breaks. Using this procedure we found that under similar reaction conditions A3A treatment converted ~78% of 5mC containing substrate to product (Figure 6A, lane 8), while converting ~99% of C containing substrate to product (Figure 6A, lane 4). As expected, A3A mutant

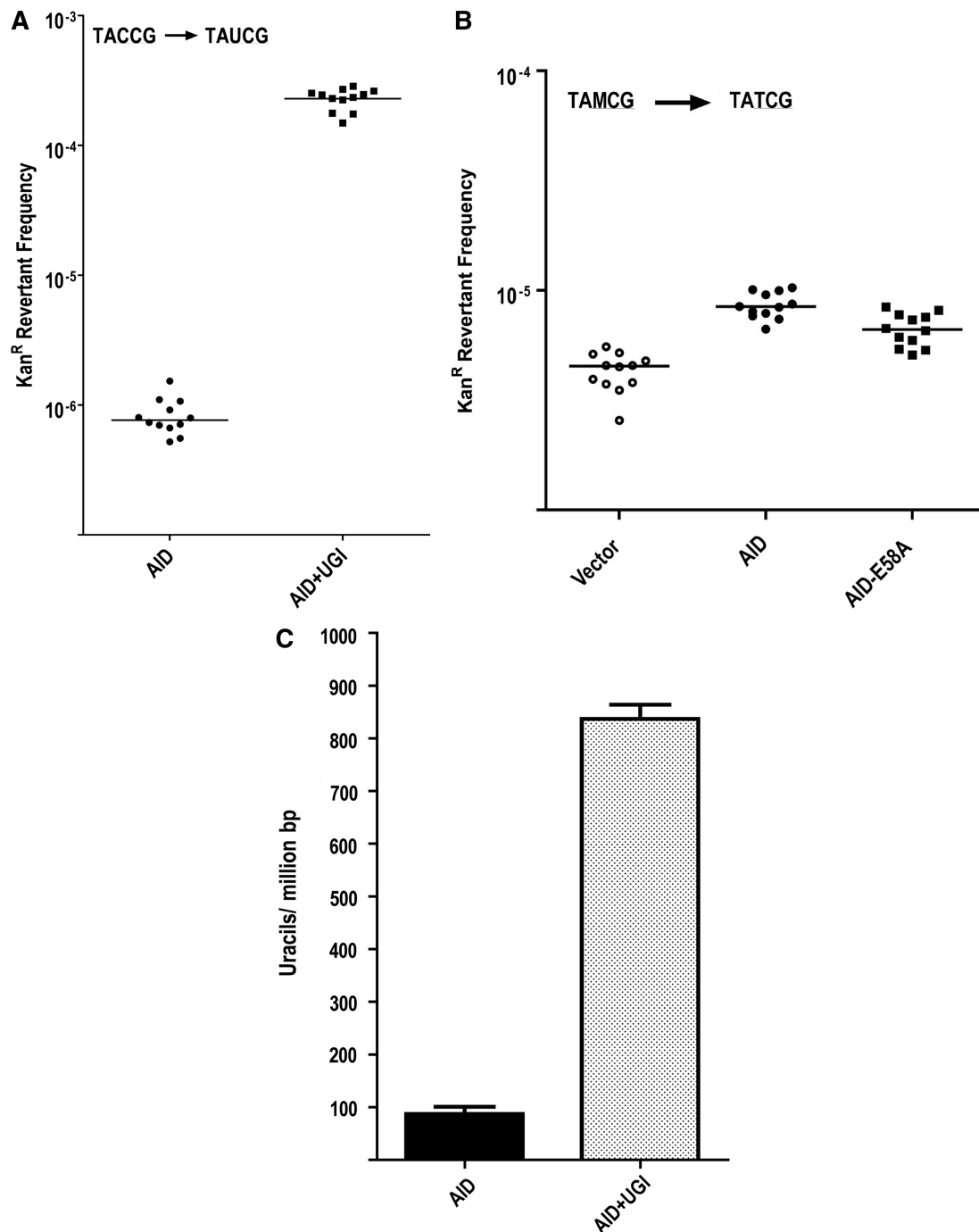


Figure 3. (A) Cytosine deamination by AID. Kan^R revertant frequencies in BH500 cells expressing AID alone or AID and UGI. The horizontal line within the data points is the median value. (B) 5mC deamination by AID. Kan^R revertant frequencies in BH500 cells expressing AID, E58A mutant of AID or empty vector are shown. The horizontal line within the data points is the median value. (C) Quantification of genomic uracils. The amount of genomic uracil created by AID alone or AID and UGI is shown. The error bars indicate the standard deviation.

with E72A mutation was unable to convert 5mC to T (Figure 6A, lane 10). When the time-dependence of deamination by A3A was studied, significant conversion of 5mC could be detected even at the earliest time points (Figure 6B and C). For technical reasons, there was some set-to-set variation in the data (see 'Discussion' section), but in some experiments >90% of 5mC was converted to T in 60–120 min (Supplementary Figure S5, Set 2 and data

not shown). Thus, both genetic and biochemical data show that A3A is an efficient deaminase of both C and 5mC in DNA.

An eight amino acid segment of A3A confers AID 5mC deamination ability

We and others showed previously that when the putative DNA-binding domain (DBD) of AID is replaced with

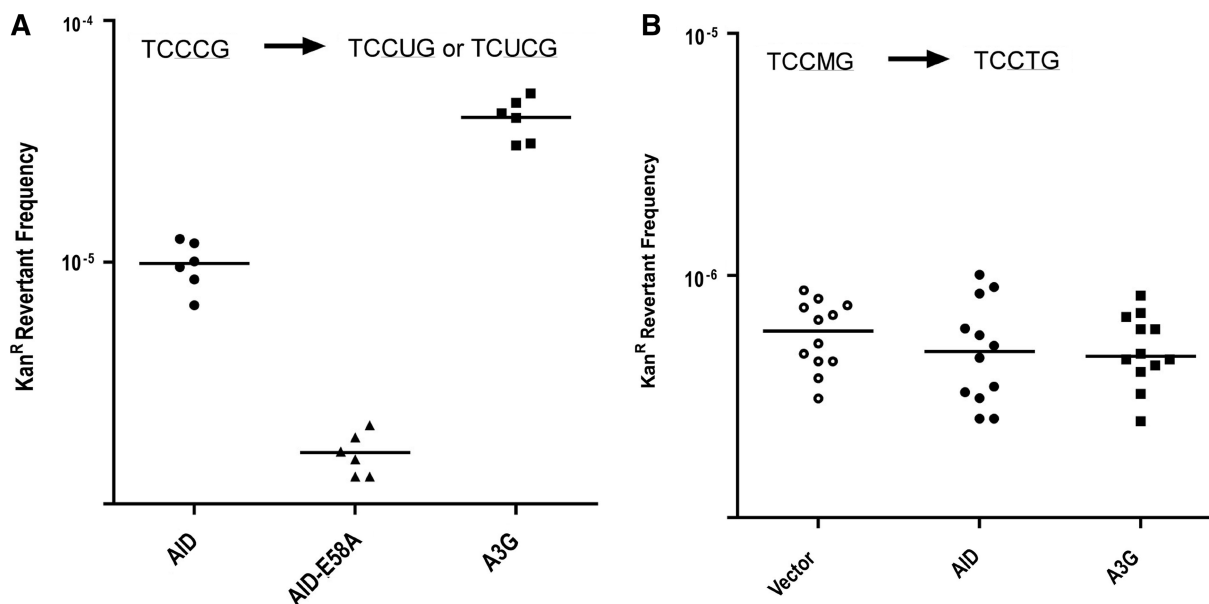


Figure 4. (A) Comparison of cytosine deamination by AID and A3G. Kan^R revertant frequencies in BH300 cells expressing AID, E58A mutant of AID or A3G are shown. The horizontal line within the data points is the median value. (B) Comparison of 5mC deamination by AID and A3G. Kan^R revertant frequencies in BH300 cells expressing vector, AID or A3G along with M.HpaII are shown. The horizontal line within the data points is the median value.

segment from either A3G or A3F carboxy-terminal domain, the sequence preference of AID is altered to that of the latter enzymes (21,40,41). We hypothesized that A3A may contain a DBD that can accommodate 5mC and replacement of the DBD of AID with this domain may allow AID to deaminate 5mCs. To test this, the DBD of AID was replaced with the corresponding sequences from A3A and the resulting hybrid (AID-A3AR2; Figure 7A) was tested for 5mC deamination ability in the genetic assay. The hybrid gene caused much higher Kan^R revertant frequency than AID and was almost as efficient as A3A at deaminating 5mC in this assay (Figure 7B). Thus, an eight amino acid segment of A3A was able to confer upon AID the ability to deaminate 5mC in DNA.

DISCUSSION

We showed here that while human AID and A3G proteins were quite proficient at converting C to U in *E. coli* genomic DNA, they had little or no activity when 5mC was the substrate. Furthermore, the strong cytosine deamination activity by AID (Figure 3A and C) in the same cells in which little 5mC to T conversion could be detected (Figure 3B and Supplementary Figure S3A–D), strongly argues that the poor 5mC deamination activity was not due to low expression level of the protein or its stability in *E. coli*. The genetic system used here scored 5mC to T mutations in the WRC sequence context preferred by AID, as well as in the CpG context where the bulk of 5mC is found in mammalian cells. In no case was the increase in 5mC to T mutations >2-fold above the vector control (Figure 3B and Supplementary Figure S3A–D). These results are inconsistent with the

conclusions of the largely qualitative studies reported by Morgan *et al.* (9) and models of DNA demethylation that depend upon deamination of 5mC by AID (10,11).

We were concerned that the low 5mC deamination abilities of AID and A3G scored by the Kan^R reversion assay reflected some inherent weakness of the genetic system. To dispel such criticisms, we tested two additional APOBEC3 family members, A3A and APOBEC3C (A3C) with this assay. The expression of A3C was toxic to *E. coli* and this gene did not give consistent results in the Kan^R assay (data not shown). However, A3A was more mutagenic than AID in all sequence contexts including a WRC sequence in which a C to U deamination was being scored (Figure 5A). Previous studies using the *E. coli* rifampicin-resistance assay showed that A3G is somewhat more mutagenic than AID (42) and other studies using the same assay have shown A3A to be more mutagenic than A3G (43,44). While these earlier results when taken together suggested that A3A was a more potent mutator than AID, the present study is the first one to compare directly these two deaminases. Importantly, A3A also scored much better at deaminating 5mC than AID (Figure 5C). When M.HpaII is expressed in *E. coli*, the genome contains ~64 cytosines for every 5mC and hence the high frequency deamination of one 5mC in the proline codon in the *kan* gene in a sea of C's suggests that A3A does not have a strong preference for C over 5mC.

We confirmed this biochemically by purifying partially A3A from *E. coli* and testing it *in vitro* for C and 5mC deamination using single-stranded DNA substrates. A3A was able to deaminate both the bases with only a moderate preference for C over 5mC. It should be noted that the efficiency of the TDG reaction varied from

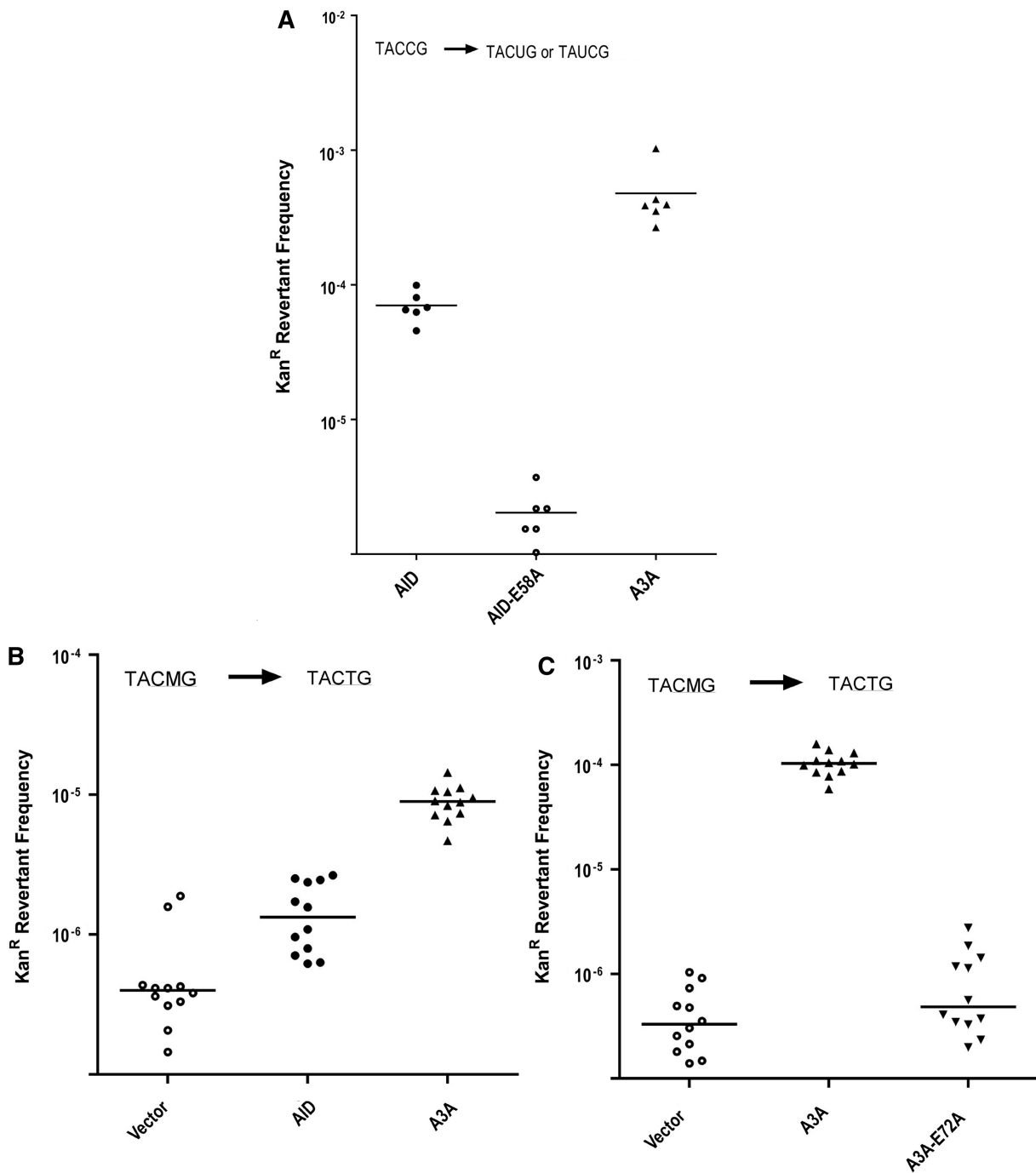


Figure 5. (A) Comparison of cytosine deamination by AID and A3A. Kan^R revertant frequencies in BH500 cells expressing AID, E58A mutant of AID or A3A are shown. The horizontal line within the data points is the median value. (B) Comparison of 5mC deamination by AID and A3A. Kan^R revertant frequencies in BH500 cells expressing AID, E58A mutant of AID or A3A along with M.HpaII are shown. (C) 5mC deamination by A3A. Kan^R revertant frequencies in BH500 cells expressing A3A, E72A mutant of A3A or empty vector are shown. The horizontal line within the data points is the median value.

experiment to experiment partly because the DNA duplex was unstable at the recommended reaction temperature (47°C). Consequently, not all the T•G mismatches created by A3A are processed by TDG and hence there is a lower than actual appearance of 5mC to T conversion in Figure 6B and Supplementary Figure S5. Regardless, these data confirm the conclusion from our genetic

studies that A3A is an efficient deaminase of 5mC. The only quantitative biochemical study of AID using DNAs with C and 5mC concluded that the former base is ~10-fold better substrate than latter (20). Our data show (Figure 6 and Supplementary Figure S5) that cytosine is preferred over 5mC by A3A by a factor of only two to three. It should be noted that there has been no previous

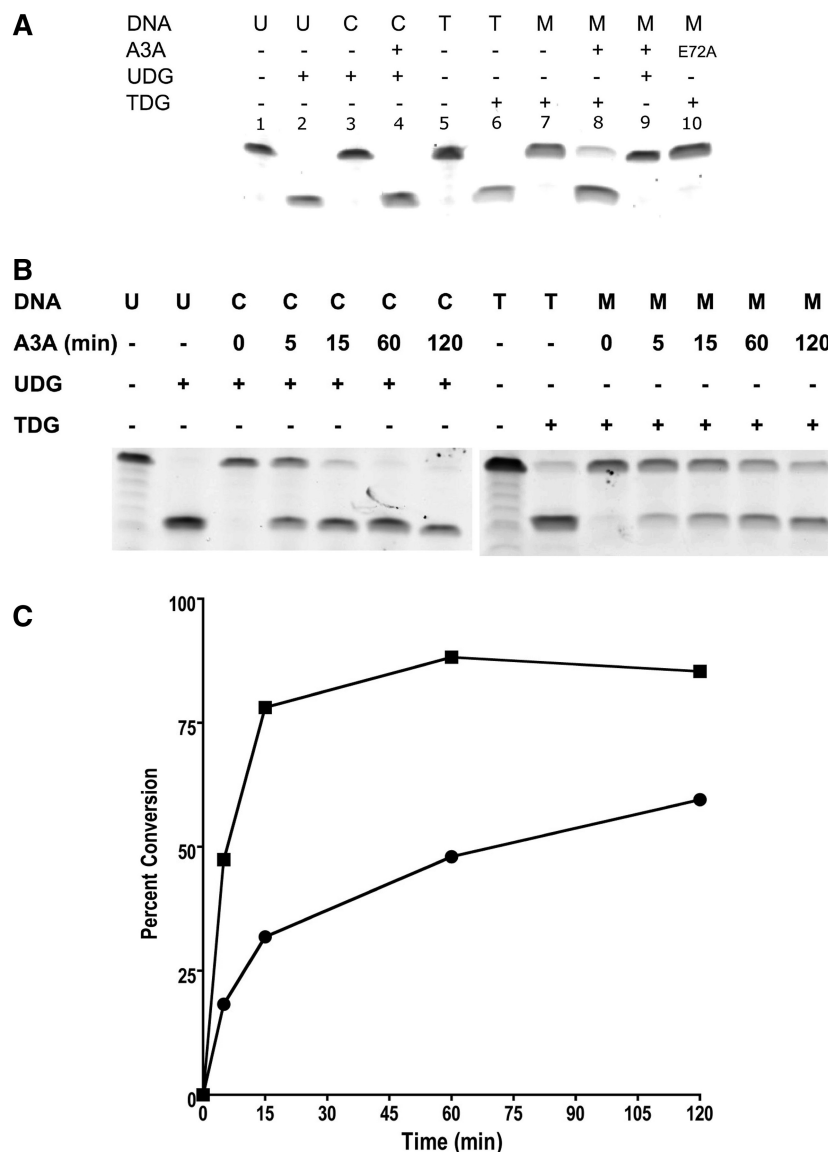


Figure 6. (A) Cytosine and 5mC deamination by A3A. Fluorescently labeled DNA oligomers with a single C or 5mC were treated with A3A or A3A-E72A. The C-containing oligomer was treated with UDG and cleaved while the 5mC-containing oligomer was hybridized to its complement, treated with TDG and cleaved. (B) Kinetics of cytosine and 5mC deamination by A3A. Fluorescently labeled DNA oligomers with a single C or 5mC were treated with A3A and the reactions were stopped at indicated times. The complementary strand was hybridized to the oligomer and UDG or TDG was added to cleave the DNA strand. DNA duplex containing U•G or T•G mispairs respectively serve as controls for the efficiency of reactions with UDG and TDG. (C) Quantification of data in part B above. The data are normalized for the efficiency of UDG and TDG reactions, and the signal at zero time was subtracted from all data points.

report of any enzyme from any organism that efficiently convert 5mC in DNA to T.

It can be argued that the reason why AID behaves as a weak deaminator of 5mC in our *E. coli* assays is because the protein lacks a key modification or partner. While it is not possible to disprove this, it is unlikely for several reasons. First, studies have shown that AID is a strong mutator at unmethylated cytosines in *E. coli* and our genetic system reproduces the dependence of this activity on transcription of the target gene in the same way as in B lymphocytes (3). Second, human AID purified from insect cells contains the only modification of the protein [Ser-38 phosphorylation; (45)] that is thought to affect its

substrate specificity (46). This phosphorylated protein was 10-fold less active against 5mCs' than Cs' (19,20). Third, we have shown here that AID can be changed into a more active 5mC deaminase by replacing its putative DBD with the corresponding domain of A3A (Figure 7B). This suggests that while A3A is able to accommodate the 5-methyl group on the target cytosine in its active site, AID may not be able to do so.

We attempted to dock 5-methyl-dC nucleotide into the active site of A3G using Autodock Vina (<http://autodock.scripps.edu/>) and found that Tyr-315 clashed with the methyl group (M. Carpenter and A.S. Bhagwat, unpublished results). Homology modeling of AID and A3A

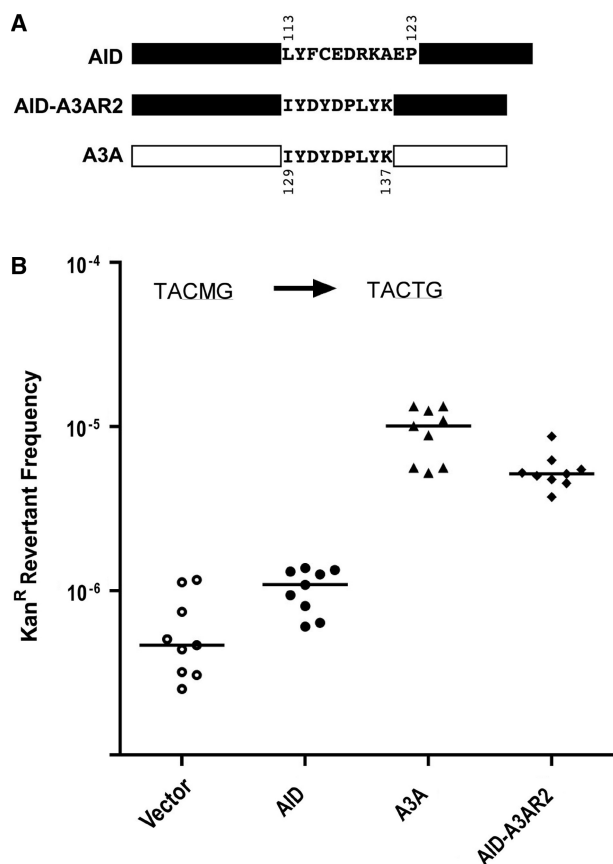


Figure 7. (A) Domain swap between AID and A3A. The sequence of the putative DBDs of AID (21,40,41) and A3A are shown schematically. The DBD of A3A was identified by aligning the sequence of this protein with the sequence of AID and of the carboxy-terminal domain of A3G. AID-A3AR2 contains all of AID except its DBD which is replaced with eight amino acid DBD from A3A. The numbers above and below the sequences are amino acid residue numbers. (B) Comparison of 5mC deamination by AID, A3A and AID-A3AR2. Kan^R revertant frequencies in BH500 cells expressing the different proteins are shown. The horizontal line within the data points is the median value.

based on the published structures of A3G suggest that while this conserved tyrosine is similarly positioned in AID, it is much further away from the cytosine-binding pocket in A3A [(47) and data not shown]. Thus, the position of this tyrosine in the active sites of AID and A3G may be the principal reason for their poor ability to deaminate 5mC.

Our demonstration that A3A is a strong 5mC deaminase does not necessarily mean that this enzyme plays a significant role in DNA demethylation during early embryogenesis. While A3A is imported into the nucleus (6), its expression has not been reported in germ cells or embryonic cells. It has been shown to play a role in restricting human viruses and retrotransposons (39,48), and in degrading foreign DNA (49). There is some evidence that it may act on 5mCs in DNA, but the evidence is indirect and comes from somatic cells. In experiments in which A3A gene was transfected into human cell lines, several C to T or G to A mutations were detected in CpG sequences within *c-myc* and *p53* genes suggesting 5mC to T deaminations (50). However,

this study did not determine the state of methylation of CpGs in these genes and hence it is possible that these were merely C to T changes.

There are several biochemical arguments against AID, A3A or other cytosine deaminases playing a major role in the demethylation of DNA in pluripotent cells. First, cytosine is the preferred substrate for all AID/APOBEC family deaminases for which this has been studied. As there is ~30-fold excess of unmethylated cytosines in mammalian genomic DNA over 5mC, it is difficult to visualize how most 5mCs could be deaminated without also deaminating all cytosines. Second, all AID/APOBEC family deaminases prefer single-stranded DNA as substrate. It is difficult to see how all the genomic DNA can be presented to these enzymes in this form in the paternal genome where demethylation is thought to occur prior to first round of replication (51). Third, a replacement of all 5mCs in the genome with cytosines through BER would require several million separate repair events per cell. The bulk of BER uses DNA polymerase β which has an error frequency of ~ 1 in 10^4 bases synthesized (52). This would lead to an unacceptably high mutational load on the embryonic genome. In summary, the work presented here suggests that the ability of human A3A to deaminate 5mC may be biological relevant, however much more work is needed before a specific biological role can be ascribed to this enzymatic activity.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–4 and Supplementary Figures 1–5.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Susan Lovett (Brandeis University) for useful advice on recombineering *E. coli* genome and Dr Reuben Harris (University of Minnesota) for providing clone of human APOBEC3A. Some of the preliminary work on this project was done by Michael Carpenter and Erandi Rajagurubandara and the authors would like to thank Sophia Shalhout for assistance with the uracil quantification assay.

FUNDING

National Institutes of Health (NIH) [GM 57200 and CA 153936]. Funding for open access charge: NIH GM57200.

Conflict of interest statement. None declared.

REFERENCES

- Conticello, S.G. (2008) The AID/APOBEC family of nucleic acid mutators. *Genome Biol.*, **9**, 229.
- Okazaki, I., Yoshikawa, K., Kinoshita, K., Muramatsu, M., Nagaoka, H. and Honjo, T. (2003) Activation-induced cytidine deaminase links class switch recombination and somatic hypermutation. *Ann. N Y Acad. Sci.*, **987**, 1–8.

3. Samaranyake, M., Bujnicki, J.M., Carpenter, M. and Bhagwat, A.S. (2006) Evaluation of molecular models for the affinity maturation of antibodies: roles of cytosine deamination by AID and DNA repair. *Chem. Rev.*, **106**, 700–719.
4. Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A. and Nussenzweig, M.C. (2004) AID is required for c-myc/IgH chromosome translocations in vivo. *Cell*, **118**, 431–438.
5. Okazaki, I.M., Kotani, A. and Honjo, T. (2007) Role of AID in tumorigenesis. *Adv. Immunol.*, **94**, 245–273.
6. Chiu, Y.L. and Greene, W.C. (2008) The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.*, **26**, 317–353.
7. Goila-Gaur, R. and Strebel, K. (2008) HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology*, **5**, 51.
8. Rosenberg, B.R. and Papavasiliou, F.N. (2007) Beyond SHM and CSR: AID and related cytidine deaminases in the host response to viral infection. *Adv. Immunol.*, **94**, 215–244.
9. Morgan, H.D., Dean, W., Coker, H.A., Reik, W. and Petersen-Mahrt, S.K. (2004) Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.*, **279**, 52353–52360.
10. Zhu, J.K. (2009) Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.*, **43**, 143–166.
11. Teperek-Tkacz, M., Pasque, V., Gentsch, G. and Ferguson-Smith, A.C. (2011) Epigenetic reprogramming: is deamination key to active DNA demethylation? *Reproduction*, **142**, 621–632.
12. Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A. and Cairns, B.R. (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell*, **135**, 1201–1212.
13. Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y. and Blau, H.M. (2010) Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature*, **463**, 1042–1047.
14. Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E. and Reik, W. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*, **463**, 1101–1105.
15. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
16. Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D. et al. (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, **146**, 67–79.
17. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. and Song, H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, **145**, 423–434.
18. Chen, Z.X. and Riggs, A.D. (2011) DNA methylation and demethylation in mammals. *J. Biol. Chem.*, **286**, 18347–18353.
19. Larijani, M., Frieder, D., Sonbuchner, T.M., Bransteitter, R., Goodman, M.F., Bouhassira, E.E., Scharff, M.D. and Martin, A. (2005) Methylation protects cytidines from AID-mediated deamination. *Mol. Immunol.*, **42**, 599–604.
20. Bransteitter, R., Pham, P., Scharff, M.D. and Goodman, M.F. (2003) Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl Acad. Sci. USA*, **100**, 4102–4107.
21. Kohli, R.M., Abrams, S.R., Gajula, K.S., Maul, R.W., Gearhart, P.J. and Stivers, J.T. (2009) A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *J. Biol. Chem.*, **284**, 22898–22904.
22. Carpenter, M., Divvela, P., Pingoud, V., Bujnicki, J. and Bhagwat, A.S. (2006) Sequence-dependent enhancement of hydrolytic deamination of cytosines in DNA by the restriction enzyme PspGI. *Nucleic Acids Res.*, **34**, 3762–3770.
23. Bandaru, B., Wyszynski, M. and Bhagwat, A.S. (1995) HpaII methyltransferase is mutagenic in *Escherichia coli*. *J. Bacteriol.*, **177**, 2950–2952.
24. Sohail, A., Lieb, M., Dar, M. and Bhagwat, A.S. (1990) A gene required for very short patch repair in *Escherichia coli* is adjacent to the DNA cytosine methylase gene. *J. Bacteriol.*, **172**, 4214–4221.
25. Dubey, A.K., Mollet, B. and Roberts, R.J. (1992) Purification and characterization of the MspI DNA methyltransferase cloned and overexpressed in *E. coli*. *Nucleic Acids Res.*, **20**, 1579–1585.
26. Sohail, A., Klapacz, J., Samaranyake, M., Ullah, A. and Bhagwat, A.S. (2003) Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res.*, **31**, 2990–2994.
27. Weiner, M.P. and Costa, G.L. (1994) Rapid PCR site-directed mutagenesis. *PCR Methods Appl.*, **4**, S131–S136.
28. Wyszynski, M., Gabbara, S. and Bhagwat, A.S. (1994) Cytosine deaminations catalyzed by DNA cytosine methyltransferases are unlikely to be the major cause of mutational hot spots at sites of cytosine methylation in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **91**, 1574–1578.
29. Parisien, R. and Bhagwat, A.S. (2009) In: Grosjean, H. (ed.), *DNA and RNA Modification Enzymes: Comparative Structure, Mechanism, Functions, Cellular Interactions and Evolution*.
30. Pham, P., Bransteitter, R., Petruska, J. and Goodman, M.F. (2003) Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*, **424**, 103–107.
31. Lindahl, T. and Nyberg, B. (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry*, **13**, 3405–3410.
32. Hennecke, F., Kolmar, H., Brundl, K. and Fritz, H.J. (1991) The *vsr* gene product of *E. coli* K-12 is a strand- and sequence-specific DNA mismatch endonuclease. *Nature*, **353**, 776–778.
33. Lieb, M. and Bhagwat, A.S. (1996) Very short patch repair: reducing the cost of cytosine methylation. *Mol. Microbiol.*, **20**, 467–473.
34. Wang, Z. and Mosbaugh, D.W. (1988) Uracil-DNA glycosylase inhibitor of bacteriophage PBS2: cloning and effects of expression of the inhibitor gene in *Escherichia coli*. *J. Bacteriol.*, **170**, 1082–1091.
35. Petersen-Mahrt, S.K., Harris, R.S. and Neuberger, M.S. (2002) AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*, **418**, 99–103.
36. Cabelof, D.C., Nakamura, J. and Heydari, A.R. (2006) A sensitive biochemical assay for the detection of uracil. *Environ. Mol. Mutagen.*, **47**, 31–37.
37. Lari, S.U., Chen, C.Y., Vertessy, B.G., Morre, J. and Bennett, S.E. (2006) Quantitative determination of uracil residues in *Escherichia coli* DNA: contribution of *ung*, *dug*, and *dut* genes to uracil avoidance. *DNA Repair*, **5**, 1407–1420.
38. Parisien, R. and Bhagwat, A.S. (2009) In: Grosjean, H. (ed.), *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Landes Bioscience, Austin, TX.
39. Chen, H., Lilley, C.E., Yu, Q., Lee, D.V., Chou, J., Narvaiza, I., Landau, N.R. and Weitzman, M.D. (2006) APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.*, **16**, 480–485.
40. Carpenter, M.A., Rajagurubandara, E., Wijesinghe, P. and Bhagwat, A.S. (2010) Determinants of sequence-specificity within human AID and APOBEC3G. *DNA Repair*, **9**, 579–587.
41. Wang, M., Rada, C. and Neuberger, M.S. (2010) Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *J. Exp. Med.*, **207**, 141–153.
42. Harris, R.S., Petersen-Mahrt, S.K. and Neuberger, M.S. (2002) RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell*, **10**, 1247–1253.
43. Bogerd, H.P., Tallmadge, R.L., Oaks, J.L., Carpenter, S. and Cullen, B.R. (2008) Equine infectious anemia virus resists the antiretroviral activity of equine APOBEC3 proteins through a packaging-independent mechanism. *J. Virol.*, **82**, 11889–11901.
44. Bogerd, H.P., Wiegand, H.L., Doehle, B.P. and Cullen, B.R. (2007) The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains. *Virology*, **364**, 486–493.

45. Pham,P., Smolka,M.B., Calabrese,P., Landolph,A., Zhang,K., Zhou,H. and Goodman,M.F. (2008) Impact of phosphorylation and phosphorylation-null mutants on the activity and deamination specificity of activation-induced cytidine deaminase. *J. Biol. Chem.*, **283**, 17428–17439.
46. Basu,U., Chaudhuri,J., Alpert,C., Dutt,S., Ranganath,S., Li,G., Schrum,J.P., Manis,J.P. and Alt,F.W. (2005) The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature*, **438**, 508–511.
47. Bulliard,Y., Narvaiza,I., Bertero,A., Peddi,S., Rohrig,U.F., Ortiz,M., Zoete,V., Castro-Diaz,N., Turelli,P., Telenti,A. *et al.* (2011) Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities. *J. Virol.*, **85**, 1765–1776.
48. Bogerd,H.P., Wiegand,H.L., Doehle,B.P., Lueders,K.K. and Cullen,B.R. (2006) APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res.*, **34**, 89–95.
49. Stenglein,M.D., Burns,M.B., Li,M., Lengyel,J. and Harris,R.S. (2010) APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat. Struct. Mol. Biol.*, **17**, 222–229.
50. Suspene,R., Aynaud,M.M., Guetard,D., Henry,M., Eckhoff,G., Marchio,A., Pineau,P., Dejean,A., Vartanian,J.P. and Wain-Hobson,S. (2011) Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc. Natl Acad. Sci. USA*, **108**, 4858–4863.
51. Feil,R. (2009) Epigenetic asymmetry in the zygote and mammalian development. *Int. J. Dev. Biol.*, **53**, 191–201.
52. Chan,K.K., Zhang,Q.M. and Dianov,G.L. (2006) Base excision repair fidelity in normal and cancer cells. *Mutagenesis*, **21**, 173–178.