Published in final edited form as:

J Mol Biol. 2015 September 25; 427(19): 3042–3055. doi:10.1016/j.jmb.2015.08.006.

Characterization of the Catalytic Domain of Human APOBEC3B and the Critical Structural Role for a Conserved Methionine

Sachini U. Siriwardena^{1,*}, Thisari A. Guruge¹, and Ashok S. Bhagwat^{1,2}

¹Department of Chemistry, Wayne State University, Detroit, MI 48202, U.S.A

²Department of Immunology and Microbiology, Wayne State University, Detroit, MI 48202, U.S.A

Abstract

Human APOBEC3B deaminates cytosines in DNA and belongs to the AID/APOBEC family of enzymes. These proteins are involved in innate and adaptive immunity, and may cause mutations in a variety of cancers. To characterize its ability to convert cytosines to uracils, we tested several derivatives of APOBEC3B gene for their ability to cause mutations in *Escherichia coli*. Through this analysis, a methionine residue at the junction of amino- and carboxy-terminal domains was found to be essential for high mutagenicity. Properties of mutants with substitutions at this position, examination of existing molecular structures of APOBEC3 family members and molecular modeling suggest that this residue is essential for the structural stability of this family of proteins. The APOBEC3B carboxy-terminal domain (CTD) with the highest mutational activity was purified to homogeneity and its kinetic parameters were determined. Size exclusion chromatography of the CTD monomer showed that it is in equilibrium with its dimeric form and MALDI-TOF analysis of the protein suggested that the dimer may be quite stable. The partially purified amino-terminal domain did not show intrinsic deamination activity and did not enhance the activity of the CTD in biochemical assays. Finally, APOBEC3B was at least 10-fold less efficient at mutating 5-methylcytosine (5mC) to thymine than APOBEC3A in a genetic assay, and was at least 10-fold less efficient at deaminating 5mC compared to C in biochemical assays. These results shed light on the structural organization of APOBEC3B catalytic domain, its substrate specificity and its possible role in causing genome-wide mutations.

Graphical abstract

*Corresponding author: Department of Chemistry, Wayne State University, Detroit, MI 48202, U.S.A.; Tel.: 1-313-577-2549; Fax: 1-313-577-8822.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

breast cancer; cancer mutational signatures; kataegis; APOBEC3B multimerization; demethylation

Introduction

APOBEC3B (A3B) belongs to the AID/APOBEC family of DNA-cytosine deaminases that has 11 members in *Homo sapiens* and other primates [1, 2]. A3B converts cytosines in single-stranded DNA to uracils and is part of the innate immunity response of the organism. It mutates and restricts viruses such as HIV-1 and hepatitis B [3, 4] and also inhibits retrotransposition of endogenous retroelements [5–7]. The seven human APOBEC3s can be divided into two groups based on whether they contain one or two zinc-binding domains, and A3B contains two such domains. Based on the properties of point mutations in the putative zinc-binding residues, the carboxy-terminal domain (CTD) is thought to be responsible for the catalytic activity of this enzyme [8, 9]. However, the amino-terminal domain (NTD) of A3B does contribute to its antiviral activity [8] suggesting that both the zinc-binding domains play a role in the immune function of the protein.

A remarkable recent finding is that many of the mutations found in a variety of cancer genomes bear the signature of mutations caused by members of the AID/APOBEC family [10–13]. The most compelling evidence for an AID/APOBEC member to be the causative agent for such mutations is found in breast cancers where ~50% of the tumor samples and cell lines have elevated levels of APOBEC3B. Furthermore, Burns *et al* [11] found that two of the breast cancer cell lines, MDA-MB-453 and HCC1569, have elevated levels of genomic uracils that could be attributed to APOBEC3B. This raises the possibility that many of the “spontaneous” mutations acquired by cancer cells- and perhaps even by normal cells- may be caused by A3B and other members of the AID/APOBEC family.

Several mutational studies of AID, A3G and some other members of this family have defined the residues important for catalysis, DNA or RNA binding and their interactions with HIV-coded protein Vif (reviewed in [1, 14–17]). Additionally, A3A, A3F and A3G proteins have been purified and studied using biochemical tools. Several X-ray crystallographic or NMR-based structures of the catalytic domain of A3G are available [18–22]. NMR-based structures of A3A [23], and X-ray crystal structures of A3F [24] and APOBEC2 [25] provide insights into the structure-function relationships in this family of proteins.

In contrast, A3B has not been characterized at the biochemical level. The mutational spectrum of A3B in virus genomes suggested that the enzyme deaminates cytosines in 5'-YC sequence context (Y is pyrimidine) with a preference for TC dinucleotides [4]. This observation was confirmed in biochemical studies of protein containing residues Pro-195 through C-terminal end of A3B partially purified from mammalian cells [11, 26]. However, full-length A3B, its NTD or CTD have not been purified to homogeneity and characterized biochemically. To better understand the catalytic and structural properties of A3B, we initiated a study to define the catalytic domain of this enzyme. It was purified from *E. coli* to homogeneity and was found to require residues upstream of Pro-195 for optimal activity. We also determined kinetic parameters of the catalytic domain and tested the protein for ability to deaminate 5-methylcytosines.

Results

Defining the catalytic domain of APOBEC3B

Although structure-based analysis of APOBEC3 proteins suggests that the A3B CTD begins at residue number 193 ([21, 27] and Supplementary Fig. S1), it has been reported that a shorter version that starts at residue number 195 is catalytically active [11, 26]. Consequently, we first cloned the gene fragment containing codons 195 to the C-terminus (195-CTD; Fig. 1A) in *E. coli* plasmid pSU24 and tested its ability to deaminate cytosines using previously described kanamycin-resistance reversion assay [28, 29]. In this assay transition mutation at the second cytosine in an ACCA sequence in a plasmid-borne kanamycin-sensitive allele restores wild-type amino acid sequence (phenotype- kanamycin-resistant; Kan^R) and the mutant bacteria are selected on kanamycin plates. The results are shown in Fig. 1B.

The 195-CTD was not significantly more mutagenic than the empty vector (Fig. 1B). The cloned A3B gene fragment is downstream of both the T7 and β -galactosidase promoters in pSU24 [30] and these experiments were performed using a strain that carries the T7 RNA polymerase gene in its genome. To increase expression of 195-CTD, the cells were grown in the presence of the inducer IPTG and reversion assays were repeated. This resulted in higher revertant frequencies, but the 195-CTD was still not highly mutagenic (Fig. 1C). We wondered whether making the CTD longer at its amino terminus would increase its activity and we prepared two additional constructs, 187-CTD and 193-CTD, to test this possibility (Fig. 1A).

Both 187-CTD and 193-CTD constructs were much more mutagenic than 195-CTD even without IPTG in the growth media (Fig. 1D). The 193-CTD was slightly less mutagenic than 187-CTD and both the CTD constructs were somewhat less mutagenic than the full-length A3B (Fig. 1D). However, in a different genetic background where the kanamycin-sensitive allele is located in the bacterial chromosome, both the CTD constructs were more mutagenic than full-length A3B and 193-CTD was significantly more mutagenic than 187-CTD (Supplementary Fig. S2). We also constructed two additional CTD versions that began at residue numbers 170 and 181 respectively (Supplementary Fig. S1) and tested their ability to increase Kan^R revertant frequency. However, they were less mutagenic than the 187-CTD and 193-CTD (data not shown). Based on these results we conclude that both 187-CTD and

193-CTD are substantially more mutagenic than the 195-CTD, and are likely to have enzymatic activity which is comparable to the full-length A3B.

The importance of Met-193 for APOBEC3B enzyme activity

The sequence at the amino terminal end of 193-CTD is MDPDTF, while the 195-CTD has MPDTF (Fig. 1A, the underlined Met residue is the initiator methionine). We were surprised by the results described above which implied that an insertion of a single Asp after the initiator methionine in the 195-CTD construct had a strong effect on its mutagenicity. To investigate whether this aspartate plays a critical role in the mutagenicity of A3B, a D194A mutant of full-length A3B was constructed. However, it was as mutagenic as the wild-type A3B (Fig. 2A). The same mutation within the 193-CTD also did not reduce mutagenicity of this gene fragment (Supplementary Fig. S3). This shows that the Asp-194 residue is not essential for the mutagenic activity of the CTD.

In contrast, when Met-193 in the full-length A3B was changed to Ala, the Kan^R revertant frequency decreased ~28-fold and changing the residue to Glu decreased the mutagenicity further by a factor of ~4 (Fig. 2B). Changing Met-193 to Phe had a much more modest effect on mutagenicity (Fig. 2B), suggesting that both the size and the hydrophobicity of the side chain of the residue at this position may be important for the catalytic function of A3B CTD. To confirm that the lower mutagenicity of Met-193 mutants in the Kan^R assay was not due to changes in sequence selectivity of the enzyme, we performed rifampicin-resistance (Rif^R) assay. Rifampicin-resistance is acquired principally as a result of mutations in the *rpoB* gene in the chromosome and C to T mutations in many different sequence contexts can create the Rif^R phenotype [31]. This assay also showed that the M193A and M193E mutants of full-length A3B were highly defective in their ability to cause mutations (Fig. 2C). We also made an additional construct in which a glycine codon was inserted between the Met-193 and Asp-194 codons in 193-CTD (N-terminal sequence MGD¹⁹³DTF) and this variant also had substantially reduced mutagenic activity (data not shown). For these and other reasons we conclude that Met-193 is required for full catalytic activity of A3B and its position relative to other amino acids within the protein is also important for its function.

Purification and characterization of APOBEC3B CTD

A carboxy terminal poly-His tag was added to both the 193-CTD and 187-CTD fragments of A3B gene, the tagged proteins were expressed in *E. coli* and purified over two successive chromatography columns. The proteins appear nearly homogenous in Coomassie Blue stained gels (Fig. 3A and Supplementary Fig. S4A). However, the yield of the 193-CTD (1.75 mg/L) was about three times higher than that of 187-CTD. In contrast to these two constructs, 195-CTD expressed poorly in *E. coli* (data not shown) and hence no attempt was made to purify it to homogeneity. Due to the lower yield of 187-CTD and somewhat higher catalytic activity of the 193-CTD (Supplementary Fig. S4B), biochemical studies of only 193-CTD are described below.

The purified 193-CTD was analyzed using time-of-flight matrix-assisted laser desorption ionization (MALDI-TOF) spectrometry to determine its molecular mass. The protein showed one strong peak close to its expected molecular mass (23,473.28 Da; Fig. 3B),

confirming that the protein preparation was quite pure. The size of the purified 193-CTD is consistent with the initiator methionine being present in the protein. Similarly MALDI-TOF analysis of purified 187-CTD showed a single major peak at the expected molecular mass (Supplementary Fig. S4C).

Interestingly, the mass spectrum for 193-CTD repeatedly showed a second minor peak at approximately twice the molecular mass of the main peak (Fig. 3B). The original spectrometer trace is shown in Supplementary Fig. S5. Although the intensity of this peak was very low (about 1% of the intensity of the major peak; Fig. 3B, inset), this suggested that 193-CTD dimerizes and at least a fraction of the dimers remain intact during the laser-induced vaporization of the protein.

To confirm the dimerization of 193-CTD, the purified protein was rechromatographed on a pre-calibrated Superdex-75 column. Two distinct peaks were seen in the absorbance profile of the eluted protein and the elution volumes of the peaks matched roughly with the predicted monomer and dimer species of 193-CTD (Fig 3C). SDS-PAGE analysis of fractions from both the peaks showed that they contained the same protein and deamination assays showed that both peaks contained catalytically active protein (data not shown). Together these results show that the purified 193-CTD monomer is in thermodynamic equilibrium with its dimeric form. Further experiments are needed to determine whether the monomer or the dimer is catalytically more active.

Cytosine deamination by APOBEC3B CTD

Purified 193-CTD was further characterized biochemically starting with its stability in solution. The protein was pre-incubated at 37 °C in the reaction buffer for various lengths of time and then a DNA substrate containing a single cytosine was added to each reaction. The incubations were continued further for 15 min before they were terminated, the DNA was cleaved at the uracils and the products were separated on a denaturing gel as described previously [32]. The results from two independent experiments showed that the enzyme was stable for at least the first 15 minutes of pre-incubation (Fig. 4A). Additional preliminary experiments also showed that the enzyme was able to create more product than its molar amount- i.e. “turn over” (Fig. 4B).

Based on these results we designed experiments to determine the kinetic parameters of the enzyme using three different oligomeric substrates (Supplementary Table 1). The $v_{\max}$ and K_M for the three DNA substrates were roughly similar. However, the apparent K_M of the enzyme (37 to 74 μM) was about the same or higher than the highest substrate concentration used in biochemical assays (40 μM). Additionally, as the protein exists in both monomeric and dimeric forms, the concentration of the catalytically active form of the enzyme is unknown. Consequently the catalytic efficiency of the enzyme is reported only as apparent $v_{\max}/K_M$ in Table 1. This number was similar for the three substrates and is quite low (Table 1). These results suggest that the 193-CTD purified from *E. coli* is a slow enzyme and the nucleotide 3' to the target cytosine does not have a strong effect on enzyme activity.

As shown above, the full-length A3B has higher mutagenicity than the CTD in at least one genetic background (Fig. 1D) and hence we wondered whether the NTD enhances the

catalytic activity of the CTD. When we attempted to test this by cloning the full-length gene in an expression plasmid, all the clones were found to contain missense mutations. Fourteen independent clones were examined and they contained base substitutions, insertion/deletions or both. In each case the mutation(s) changed the amino acid sequence of A3B CTD. We conclude that full-length A3B is toxic to *E. coli* due to its high mutagenicity (Fig. 1D) and this leads to selection of mutant variants. For this reason, the full-length protein was not purified and instead we examined the mutagenicity of A3B NTD in *E. coli* and the ability of A3B NTD protein to deaminate cytosines on its own and in cooperation with 193-CTD.

Genetic and biochemical characterization of the APOBEC3B NTD

The ability of A3B NTD (residues 1 through 192) to deaminate cytosines was judged using both genetic and biochemical assays. In the Kan^R reversion assay, the NTD increased the revertant frequency above the frequency obtained with empty vector by a small amount, but the increase was not statistically significant (Fig. 5A). These results are consistent with previous reports of low mutational activity of the NTD in rifampicin-resistance assay [33]. This suggests that the A3B NTD is at best a weak deaminase on its own and we confirmed this by characterizing partially purified NTD.

An amino-terminal poly-His tag was attached to the NTD (residues 1 through 192), the tagged protein was expressed in *E. coli* and purified over a Ni²⁺ affinity column. Although its expression was not as high as 187-CTD or 193-CTD (data not shown), Coomassie Blue staining showed a prominent band at the expected size (Supplementary Fig. S6). The concentration of the NTD in the preparation was estimated based on BSA standards and used for *in vitro* deamination assays.

When TCGA-40 DNA substrate was incubated with excess A3B NTD, no deamination product was detected (lanes 4, 6 and 8, respectively; Fig. 5B). Furthermore, the NTD did not enhance the deamination activity of 193-CTD when the DNA was co-incubated with the two proteins (lanes 5, 7 and 9; Fig. 5B). These results suggest that A3B NTD has little intrinsic deamination activity, and it does not enhance the catalytic activity of the CTD in *trans*.

5-Methylcytosine deamination activity of APOBEC3B CTD

The ability of 187-CTD and 193-CTD constructs to deaminate 5-methylcytosine (5mC) was tested using the Kan^R reversion assay in the *E. coli* strain BH400 which also contained the gene for a cytosine methylase, Dcm, on a compatible plasmid [32]. In this strain, Uracil-DNA glycosylase (Ung) repairs all the uracils in DNA and only a 5mC to T deamination creates Kan^R revertants [32]. A3A, which is an efficient deaminase of 5mC [32, 34, 35], served as the positive control in these experiments and the data are presented in Fig. 6A.

The full-length A3B, 187-CTD and 193-CTD caused a significant increase in the Kan^R revertant frequency in this genetic background and the CTD constructs displayed higher mutagenicity than the full-length A3B (Fig. 6A). However, the revertant frequencies due to the A3B constructs were about ten-fold lower than those due to A3A (Fig. 6A). This suggests that although A3B is capable of 5mC deamination, it may not be as efficient as A3A in this reaction.

The ability of the purified 193-CTD protein to deaminate 5mC was tested using two different DNA substrates (Supplementary Table 1) and this was compared with its ability to deaminate cytosine in the same sequence context. The 193-CTD did convert some 5mC in each substrate to T, but this was much less efficient than its ability to convert C to U (Fig. 6B). When the maximum reaction velocity was calculated for TMAG and TMGG substrates (M is 5mC), the specific activity of 193-CTD was respectively about 25-fold and 11-fold (average 16-fold) lower for the 5mC-containing oligomers than the corresponding C-containing substrates. For comparison, A3A protein was partially purified and was used in deamination assays with TCGG and TMGG oligomers as substrates. The enzyme was about five times more efficient at deaminating C than 5mC (Fig. 6C). The more modest preference of A3A for the unmodified cytosine has been noted before [32], and the data presented above suggest that A3B has a stronger preference for C over 5mC than A3A.

Discussion

Using deletion constructs and point mutations, we defined the catalytic domain of the human APOBEC3B gene. The maximally active domain spans from Met-193 to the carboxy-terminal end of the protein (193-CTD). This domain was somewhat less mutagenic than the full-length A3B. The purified 193-CTD protein was active *in vitro* and was used to determine the kinetic parameters for deamination of cytosines and 5-methylcytosines in single-stranded DNA. These studies show that the bacterially purified protein is a slow enzyme with a strong preference for deaminating C compared to 5mC. We also analyzed the purified 193-CTD by MALDI-TOF and size exclusion chromatography and found that the carboxy-terminal domain dimerizes. This raises the possibility that this interaction may also occur in full-length A3B resulting in its dimerization. These studies also show that Met-193 residue is required for the optimal mutational activity of the full-length A3B.

The previously characterized version of A3B CTD that starts at residue 195 [11, 26] is only weakly active (Fig. 1B and 1C) and 193-CTD is more mutagenic than 195-CTD by at least a factor of ten (Fig. 1B and 1D). Additionally, we found that replacing Asp-194 with Ala did not decrease its mutagenicity (Fig. 2A). In contrast, while Met to Glu substitution at position 193 reduced mutagenicity of full-length A3B by two orders of magnitude, a Met to Phe substitution had a much more modest effect (Fig. 2B and 2C).

These mutational studies led us to examine the role of this methionine in A3B structure and whether it is conserved in the AID/APOBEC family of proteins. We first aligned A3B CTD sequence with the sequences of these enzymes using Clustal Omega software [36] and manually adjusted it to conserve secondary structure elements found within 3D structures of CTDs of two double-domain APOBEC3 proteins (A3F and A3G) and complete structures of two single-domain APOBEC3 proteins (A3A and A3C) [18, 19, 23, 24, 37, 38]. The alignment suggests that Met-193 is conserved among these cytosine deaminases (Fig. 7). We note that while some investigators have previously suggested the same alignment for Met-193 of A3B shown in Fig. 7 [27, 39, 40], others have aligned Met-193 with different residues in other APOBEC3s [22, 41].

Mutational studies of other members of this family are consistent with this alignment. In particular, alanine-scanning mutagenesis of the NTD-CTD junction in A3G [21] found that among the 20 residues mutated only mutations at Met-197, Phe-202, Phe-206 and Asn-208 eliminated mutator phenotype of A3G in *E. coli* [21]. Furthermore, larger hydrophobic substitutions at Met-197 position such as Cys and Phe did not eliminate mutagenicity [21]. Mutational analysis of APOBEC3A has shown that while the first methionine in APOBEC3A is dispensable for its activity, the methionine aligned with Met-193 of A3B (Fig. 7; Met-13) is essential for the deamination activity of the protein [42]. Similar mutational analysis of AID has also shown that the methionine that aligns with Met-197 of A3G, Met-6, is essential for its activity [43].

Examination of the known APOBEC3 family 3D structures points to a structural role for this conserved methionine. All the structures show that the conserved methionine side chain is inserted into a hydrophobic pocket formed mainly by phenylalanines and cysteines (Figs. 7 and 8). For example, A3G CTD structure (PDB: 3IR2) shows that this pocket is created by residues from helices $\alpha 1$ and $\alpha 5$, and strands $\beta 1$, $\beta 2'$ and $\beta 3$ (Figs. 7 and 8A). In A3G CTD the Met-197 residue is sandwiched between walls formed by Phe-202 and Phe-206 on one side, and Phe-241 and the phenolic ring of Tyr-219 on the other side. Phe-350 and Cys-346 form the floor of this pocket (Fig. 8A). Representation of this region of A3G based on hydrophobicity shows that the hydrophobic pocket is close to the protein surface and is strongly hydrophobic on one side (orange) and less so on the other side. The terminal methyl group in Met-197 is inserted into the pocket and points the terminal methyl group towards Phe-282 and Cys-221 (Fig. 8B). As noted above, Harjes *et al* [21] found that substitution of Phe-202 or Phe-206 with Ala eliminated mutagenicity of A3G. This shows that the high hydrophobicity of this pocket is required for enzyme activity. The pocket is quite large and hence the replacement of Met-197 with the larger Phe side chain is conservative. However, the pocket is too large for the smaller Ala side chain explaining the lack of activity for the M197A mutant of A3G [21].

The structures of other APOBEC3 family members confirm the existence of this deep hydrophobic pocket and show that a majority of the residues lining the pocket are conserved among the APOBEC3s (Fig. 7 and 8C). We created a 3D structure for A3B 193-CTD using homology modeling, and this structure also showed Met-193 to be similarly inserted in a hydrophobic pocket (Supplementary Fig. S7). Together these results show that the conserved methionine at the junction of NTD and CTD domains of two domain APOBEC3s and near the amino terminus of single domain APOBEC3s (Fig. 7) acts as the “glue” that holds together two α helices and three β strands and is essential for the structural stability of these enzymes.

Another structural insight from this work is that 193-CTD dimerizes and the dimers are stable enough that a fraction of them survive laser-induced vaporization (Fig. 3B and 3C). It is likely that full-length A3B also dimerizes using the same interactions and this may explain the *in vivo* multimerization of highly expressed A3B reported recently [44]. It is well-known that A3G oligomerizes [22, 45–48] and this ability is related to its biological function [47, 49–52]. Head-to-head and head-to-tail dimers as well as higher order oligomers of A3G have been proposed based on physical studies and it will be interesting to

determine the type of multimeric structures A3B can form and the relationship of such structures with the biological function of A3B.

An important conclusion from this study is that A3B is much less efficient at deaminating 5mC compared to C. In particular, full-length A3B as well as the 187-CTD and 193-CTD constructs were much less mutagenic at a methylated cytosine in DNA in an *E. coli* genetic assay than A3A (Fig. 6A). Furthermore, biochemical assays showed that the purified 193-CTD was on average 16-fold weaker at 5mC deamination in different DNA substrates compared to C (Fig. 6B). Other members of AID/APOBEC family also show a preference for unmethylated cytosines. While we find that A3A has only a four [32] to five-fold preference (this study) for the unmethylated substrate over 5mC, another group reported a 10-fold selectivity for C [34]. AID also prefers C over 5mC by a factor of about 10 [32, 53–55]. Thus, A3B appears to disfavor 5mC over C at least as much as the other characterized members of the AID/APOBEC family and may be more biased against 5mC bases in this regard than A3A.

The poor ability of A3B at deaminating 5mC has implications to the possible role of A3B in causing DNA demethylation and genomic mutations. The AID/APOBEC family of enzymes have been widely proposed to play a role in DNA demethylation [56–59]. According to this model, the first step in demethylation pathway is the deamination of 5mC to thymine (T) by AID/APOBEC enzymes followed by removal of T through base excision repair. Recent work has also shown that many of the mutations found in a variety of cancer genomes bear the signature of mutations caused by members of the AID/APOBEC family [10–12, 60]. While the signatures 2 and 13 which contain C to T and C to G mutations in TCX context (X is any base) have been suggested to be due to deamination by the APOBEC family members [10] other signatures may also be linked to these enzymes. In principle, most commonly found mutations, C to T substitutions at XpCpG (signatures, 1A and 1B) could be attributed to deamination of 5mC residues in CpGs by APOBEC family members. Additionally, signatures 6, 7, 11, 15 and 19 also predominantly contain C to T mutations but they are in both CpG and non-CpG contexts [10]. Based on our results using bacterial genetic assays and using protein purified from *E. coli*, we conclude that A3B is much more efficient at deaminating C compared to 5mC and this makes it less likely that it plays a major role in DNA demethylation or causing mutations at methylated CpG sites in the human genome.

Our results also suggest that the full-length A3B is more mutagenic than A3B CTD and is toxic to *E. coli*. Although these genetic results did not clearly establish how the A3B NTD affects catalytic activity of A3B CTD, they did suggest the A3B NTD may affect the function of the CTD. However, genetic studies of the NTD did not show a mutator phenotype and biochemical studies with partially purified NTD showed that the NTD had no detectable catalytic activity and did not significantly enhance or inhibit the ability of the 193-CTD to deaminate cytosines in *trans* (Fig. 5B). It is still possible that the NTD can enhance the catalytic activity of the CTD in *cis*; i.e. when the two domains are covalently attached to each other.

In summary, we have used genetic and biochemical tools to define the catalytic domain of human A3B protein and identified a novel structural role of a methionine at the junction of

its NTD and CTD. We have determined the kinetic parameters of this catalytic domain and also showed that it is a poor deaminase of 5-methylcytosines. We have further shown that the A3B CTD self-dimerizes, and that A3B NTD has little intrinsic ability to deaminate cytosines. These observations clarify the structural organization of the APOBEC3 family of mutators and suggest that A3B is unlikely to play a role as a major mutator of methylated cytosines or as the first enzyme in a demethylation pathway.

Materials and Methods

Bacterial Strains and Plasmids

E. coli K-12 strains BH214 and BH400 have been described previously [32]. *E. coli* B strain BL21(DE3) was used for protein expression and purification. The plasmid pUP31 contains an inactive kanamycin resistance allele [61] and the plasmid pDCM21 carries the *E. coli* cytosine methylase gene [62].

The full-length of APOBEC3B was amplified from the human APOBEC3B cDNA purchased from OriGene Technologies (Rockville, MD) and cloned into vector plasmid pSU24 as HindIII-KpnI fragments. The primers used for the cloning are listed in Supplementary Table 2. The resulting clone, pSUA3B, was validated by sequencing (DNA sequencing core, University of Michigan). The gene for the Ung inhibitor, UGI, was amplified from a plasmid kindly provided by Umesh Varshney (Indian Institute of Science, Bangalore, India) and inserted at the KpnI site in pSU24 or pSUA3B to respectively create pSU24-UGI or pSUA3B-UGI.

Mutants of pSUA3B (E255A, D194A, M193A, M193F, M193E) were constructed using the Quickchange mutagenesis protocol (Agilent Technologies). The mutagenesis primers are listed in Supplementary Table 2. The A3B CTD gene fragments starting at amino acids 170, 180, 187, 193 and 195 were amplified from pSUA3B (Supplementary Table 2), cloned into pSU24 and used in genetic reversion assays. The gene fragments starting at 187, 193 and 195 were also cloned into pET28a(+) vector (Novagen) as a XbaI-XhoI fragment to create a carboxy terminal poly-His tagged versions of the proteins. The A3B NTD gene fragment (residues 1 through 192) was amplified and cloned into pET28a(+) as a NdeI-XhoI fragment to create an amino terminal poly-His tag version of the protein.

Kanamycin-resistance Reversion Assay

The kanamycin-resistance (Kan^R) reversion assay has been described previously [29, 63]. To quantify C to U deaminations, BH214 was cotransformed with different A3B constructs along with pUP31. In some experiments A3B constructs carrying UGI gene were transformed into BH400. To quantify 5mC to T deaminations, BH400 was cotransformed with a A3B construct and pDCM21 and kanamycin-resistant revertants were scored. In all the experiments Kan^R revertants were selected on LB plates with 50µg/mL of the antibiotic and Kan^R revertant frequency was calculated as the ratio of number of Kan^R revertants/total number of viable cells.

Rifampicin-resistance assay

Full length A3B or its mutants (E255A, M193A and M193E) were transformed into the strain BH214 and transformants were selected on LB plates containing 25 µg/mL chloramphenicol. Independent transformants were picked and grown overnight in LB media containing chloramphenicol. The following day, the cultures were diluted 1000-fold and were grown at 37 °C until OD₆₀₀ reached ~0.5 – 0.6. Appropriate dilutions of each culture were spread on LB plates containing chloramphenicol to determine the total number of viable cells. The cultures were also spread on LB plates containing 100 µg/mL rifampicin (Sigma-Aldrich) to determine the number of rifampicin -resistant (Rif^R) colonies. Mutant frequency was obtained by dividing the number of Rif^R colonies by the number of viable bacteria present in the culture.

Purification of A3B 193-CTD, 187-CTD and A3B NTD

BL21(DE3) cells expressing 187-CTD, 193-CTD, 195-CTD or NTD were grown to mid-log phase and transcription was induced with the addition of IPTG to 0.4 mM. The cells were grown overnight at 18°C and harvested by centrifugation. The cells were broken using the French Pressure Cell Press (Thermo Spectronic) and the cell-free lysate was cleared by centrifugation. The lysate was passed over Ni-NTA agarose resin (Qiagen) and the bound proteins were eluted with 250 mM imidazole. The eluted proteins were dialyzed and concentrated using Amicon Ultra Centrifugal devices (EMD-Milipore, Billerica, MA). The concentrated proteins were equilibrated in the storage buffer [20mM Tris-HCl (pH 8), 50mM NaCl, 1mM EDTA, 1mM DTT, 10% glycerol].

The proteins 187-CTD and 193-CTD from the affinity column were loaded onto a Superdex-75 column (GE Healthcare), the column was equilibrated in buffer [20mM Tris-HCl (pH 8), 50mM NaCl, 1mM EDTA, 1mM DTT] and the proteins were eluted in same buffer. Fractions of the eluate were analyzed by SDS-PAGE and the fractions containing protein of the expected size were pooled and concentrated. Glycerol was added to 10% v/v and the proteins were stored at –80°C in 50 µL aliquots.

MALDI-TOF analysis of proteins

Isolated proteins were desalted using ZipTip C18 columns (Millipore Corp.) and 1 µL samples were deposited on a polished steel MALDI target plate in 50% acetonitrile/0.1% TFA. The solvent was allowed to evaporate at room temperature and the samples were overlaid with 1 µL of matrix solution of saturated α-cyano-4 hydroxy-cinnamic acid (CHCA) in 50% acetonitrile/0.1% TFA and were air dried at room temperature. The samples were analyzed using Ultraflex MALDI-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) in LUMIGEN instrument center (Wayne State University). The spectra were recorded in the manual linear positive ion mode.

Size exclusion chromatography of the purified 193-CTD

The Superdex-75 column (GE Healthcare) was calibrated using the proteins Aprotinin (MW 6.5 kDa), Cytochrome c (12.4 kDa), Carbonic Anhydrase (29 kDa) and Albumin (66 kDa) from a kit (MWGF70; Sigma Aldrich). The column void volume (V₀) was determined using Blue Dextran (MW 2000 kDa). A calibration curve was produced by plotting the logarithms

of the known molecular masses of protein standards versus their respective V_e/V_0 (V_e is the elution volume). The 193-CTD protein purified to homogeneity over Ni-affinity and Superdex-75 columns was rechromatographed over the pre-calibrated Superdex-75 column. The protein elution was monitored using absorbance at 280 nm and the molecular weights of the two peaks were calculated from their V_e/V_0 values using the calibration curve.

Biochemical assay for cytosine deamination

The substrates for the study of C and 5mC deamination are listed in Supplementary Table 1. The deamination assays with 187-CTD and 193-CTD were performed in the reaction buffer [20mM Tris-HCl (pH 8.0), 1 mM EDTA, 1mM DTT] for the indicated lengths of time. The reactions were terminated by heating them to 95 °C for 10 min. Following incubation of C-containing oligomers with the protein, 2 units of *E. coli* Ung (New England Biolabs, Ipswich, MA) were added to the reaction and incubation was continued at 37 °C for 30 min. The reactions were stopped by adding NaOH to 0.1 M and heating to 95 °C for 10 minutes. The reaction products were separated on a 15% denaturing acrylamide gel and visualized using Typhoon 9210 phosphorimager (GE Healthcare). ImageQuant software was used to quantify the intensities of the substrates and the deamination products.

The kinetic parameters for cytosine deamination were determined using 193-CTD and three different DNA oligomer substrates. For the TCGA-40 substrate 2 μ M of A3B-193 was incubated with increasing concentration of substrate (0.5 – 12 μ M) for increasing lengths of incubation (5, 10 and 20 min) at 37 °C. For substrates TCAG-30 and TCGG-30 1.3 μ M of A3B-193 was incubated with increasing concentration of substrate (2 – 40 μ M) for increasing lengths of incubation (3, 5 and 10 min) at 37 °C. The products were separated on a gel and analyzed as described above. The data were fitted to a linear regression and analysis was done using GraphPad Prism to calculate the V_{max} and K_M values.

To study *in vitro* complementation of 193-CTD with A3B-NTD, 10 pmol of 193-CTD was incubated with 2 pmol of TCGA-40 substrate along with different amounts of A3B-NTD (0 to 20 pmol). The amount of A3B NTD in the partially purified protein was determined by comparing it to BSA standards (Supplementary Fig. S6). The reactions were performed at 37°C for 30 min and the products were processed and analyzed as described above.

Biochemical assay for 5-methylcytosine deamination

The 399 construct containing the glycosylase domain of the murine MBD4 was purified using a N-terminal poly-His tag as described previously [64]. Briefly, the expression of this protein was induced in BL21(DE3) cells using 0.4 mM IPTG and cells were grown overnight at 22 °C. Following purification of the protein using a Ni^{2+} -affinity column it was purified further to near homogeneity using size exclusion chromatography over a Superdex-75 column. The protein was stored in a buffer containing 20 mM Tris-HCl (pH 8), 50 mM NaCl, 1 mM EDTA, 1 mM DTT and 50% v/v glycerol.

To study 5mC deamination, 80 nM oligomer TMAG or TMGG (M is 5mC) was incubated with indicated concentrations of purified 193-CTD at 37°C. All the reactions were terminated after 30 min incubation and three-fold molar excess of the complementary strand

(TMAG rev-com or TMGG rev-com) was added to the reaction to create a T•G mismatch. The duplex was incubated with 5 µl of purified MBD4 catalytic domain for 1 hour at 37 °C and the products were processed and analyzed in a manner similar to the C deamination reactions.

Purification of APOBEC3A and deamination of C and 5mC by APOBEC3A

Purification of human APOBEC3A and biochemical assays of this protein have been described previously [32]. Briefly, BL21(DE3) cells expressing A3A with a C-terminal poly-His tag were grown to mid-log phase and transcription of the A3A gene was induced with the addition of IPTG to 0.4 mM. The cells were grown overnight at 18°C and the protein was purified by Ni²⁺-affinity column. The partially purified protein was dialyzed against the storage buffer [20mM Tris-HCl (pH 8), 50mM NaCl, 1mM EDTA, 1mM DTT, 10% glycerol] and stored at -80°C. The amount of A3A in the preparation was determined by SDS-PAGE of the partially purified protein and its comparison with BSA standards. To study C and 5mC deamination by A3A, 80 nM oligomer TCGG or TMGG was incubated with indicated concentrations of A3A at 37°C for 30 min.

Analysis of protein structures and homology modeling

The full sequences of A3A and A3C, and the CTDs of A3B, A3F and A3G were aligned using Clustal-omega software [36] and the alignments were manually adjusted to reduce gaps. PDB structures 2M65 (A3A), 3VOW (A3C), 4IOU (A3F CTD) and 3IR2 (A3G CTD) were visualized and analyses were performed using the UCSF Chimera package [65]. The hydrophobic pocket surrounding the conserved methionine in each case was defined using the built-in hydrophobic surface module and all the relevant residues were labeled. The transparency of the methionine was set at 90%, while the transparency of the residues forming the hydrophobic pocket was set at 10%.

The structure of 193-CTD was generated using the homology modeling server SWISS-MODEL [66]. The QMEAN4 score for the modeled structure was 0.54. The structure was opened using UCSF Chimera, and analyzed and labeled in the same manner as the APOBEC3 members with known structures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work reported here was supported by NIH grant GM 57200 to A.S.B. and funds from Wayne State University. We would like to thank the following individuals at Wayne State University for assistance with Superdex-75 chromatography (Amit Kumar), MALDI-TOF analysis (Vimukthi Senevirathne) and oligonucleotide purification (Shanqiao Wei).

Literature Cited

1. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome biology*. 2008; 9:229. [PubMed: 18598372]

2. Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Molecular biology and evolution*. 2005; 22:367–77. [PubMed: 15496550]
3. Doehle BP, Schafer A, Cullen BR. Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. *Virology*. 2005; 339:281–8. [PubMed: 15993456]
4. Yu Q, Chen D, Konig R, Mariani R, Unutmaz D, Landau NR. APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. *The Journal of biological chemistry*. 2004; 279:53379–86. [PubMed: 15466872]
5. Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic acids research*. 2006; 34:89–95. [PubMed: 16407327]
6. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, et al. APOBEC3 proteins inhibit human LINE-1 retrotransposition. *The Journal of biological chemistry*. 2006; 281:22161–72. [PubMed: 16735504]
7. Stenglein MD, Harris RS. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *The Journal of biological chemistry*. 2006; 281:16837–41. [PubMed: 16648136]
8. Bonvin M, Greeve J. Effects of point mutations in the cytidine deaminase domains of APOBEC3B on replication and hypermutation of hepatitis B virus in vitro. *The Journal of general virology*. 2007; 88:3270–4. [PubMed: 18024895]
9. Hakata Y, Landau NR. Reversed functional organization of mouse and human APOBEC3 cytidine deaminase domains. *The Journal of biological chemistry*. 2006; 281:36624–31. [PubMed: 17020885]
10. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500:415–21. [PubMed: 23945592]
11. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*. 2013; 494:366–70. [PubMed: 23389445]
12. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012; 149:979–93. [PubMed: 22608084]
13. Jinks-Robertson S, Bhagwat AS. Transcription-associated mutagenesis. *Annual review of genetics*. 2014; 48:341–59.
14. Desimmie BA, Delviks-Frankenberry KA, Burdick RC, Qi D, Izumi T, Pathak VK. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *Journal of molecular biology*. 2014; 426:1220–45. [PubMed: 24189052]
15. Goila-Gaur R, Strebel K. HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology*. 2008; 5:51. [PubMed: 18577210]
16. Jaszczur M, Bertram JG, Pham P, Scharff MD, Goodman MF. AID and Apobec3G haphazard deamination and mutational diversity. *Cellular and molecular life sciences: CMLS*. 2013; 70:3089–108. [PubMed: 23178850]
17. Malim MH. APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 2009; 364:675–87. [PubMed: 19038776]
18. Chen KM, Harjes E, Gross PJ, Fahmy A, Lu Y, Shindo K, et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature*. 2008; 452:116–9. [PubMed: 18288108]
19. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, et al. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature*. 2008; 456:121–4. [PubMed: 18849968]
20. Furukawa A, Nagata T, Matsugami A, Habu Y, Sugiyama R, Hayashi F, et al. Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *The EMBO journal*. 2009; 28:440–51. [PubMed: 19153609]

21. Harjes E, Gross PJ, Chen KM, Lu Y, Shindo K, Nowarski R, et al. An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *Journal of molecular biology*. 2009; 389:819–32. [PubMed: 19389408]
22. Shandilya SM, Nalam MN, Nalivaika EA, Gross PJ, Valesano JC, Shindo K, et al. Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure*. 2010; 18:28–38. [PubMed: 20152150]
23. Byeon IJ, Ahn J, Mitra M, Byeon CH, Hercik K, Hritz J, et al. NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nature communications*. 2013; 4:1890.
24. Bohn MF, Shandilya SM, Albin JS, Kouno T, Anderson BD, McDougale RM, et al. Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. *Structure*. 2013; 21:1042–50. [PubMed: 23685212]
25. Prochnow C, Bransteitter R, Klein MG, Goodman MF, Chen XS. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature*. 2007; 445:447–51. [PubMed: 17187054]
26. Leonard B, Hart SN, Burns MB, Carpenter MA, Temiz NA, Rathore A, et al. APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer research*. 2013; 73:7222–31. [PubMed: 24154874]
27. Aydin H, Taylor MW, Lee JE. Structure-guided analysis of the human APOBEC3-HIV restrictome. *Structure*. 2014; 22:668–84. [PubMed: 24657093]
28. Beletskii A, Bhagwat AS. Transcription-induced cytosine-to-thymine mutations are not dependent on sequence context of the target cytosine. *Journal of bacteriology*. 2001; 183:6491–3. [PubMed: 11591695]
29. Wyszynski M, Gabbara S, Bhagwat AS. 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*. *Proceedings of the National Academy of Sciences of the United States of America*. 1994; 91:1574–8. [PubMed: 8108447]
30. Bartolome B, Jubete Y, Martinez E, de la Cruz F. Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. *Gene*. 1991; 102:75–8. [PubMed: 1840539]
31. Garibyan L, Huang T, Kim M, Wolff E, Nguyen A, Nguyen T, et al. Use of the *rpoB* gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome. *DNA Repair (Amst)*. 2003; 2:593–608. [PubMed: 12713816]
32. Wijesinghe P, Bhagwat AS. Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic acids research*. 2012; 40:9206–17. [PubMed: 22798497]
33. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, et al. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:8780–5. [PubMed: 16728505]
34. Carpenter MA, Li M, Rathore A, Lackey L, Law EK, Land AM, et al. Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *The Journal of biological chemistry*. 2012; 287:34801–8. [PubMed: 22896697]
35. Suspene R, Aynaud MM, Vartanian JP, Wain-Hobson S. Efficient deamination of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A cytidine deaminase. *PloS one*. 2013; 8:e63461. [PubMed: 23840298]
36. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic acids research*. 2010; 38:W695–9. [PubMed: 20439314]
37. Kitamura S, Ode H, Nakashima M, Imahashi M, Naganawa Y, Kurosawa T, et al. The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nature structural & molecular biology*. 2012; 19:1005–10.
38. Siu KK, Sultana A, Azimi FC, Lee JE. Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. *Nature communications*. 2013; 4:2593.

39. Mitra M, Singer D, Mano Y, Hritz J, Nam G, Gorelick RJ, et al. Sequence and structural determinants of human APOBEC3H deaminase and anti-HIV-1 activities. *Retrovirology*. 2015; 12:3. [PubMed: 25614027]
40. Narvaiza I, Linfesty DC, Greener BN, Hakata Y, Pintel DJ, Logue E, et al. Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. *PLoS pathogens*. 2009; 5:e1000439. [PubMed: 19461882]
41. Dang Y, Wang X, Esselman WJ, Zheng YH. Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *Journal of virology*. 2006; 80:10522–33. [PubMed: 16920826]
42. Thielen BK, McNevin JP, McElrath MJ, Hunt BV, Klein KC, Lingappa JR. Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms. *The Journal of biological chemistry*. 2010; 285:27753–66. [PubMed: 20615867]
43. Mu Y, Prochnow C, Pham P, Chen XS, Goodman MF. A structural basis for the biochemical behavior of activation-induced deoxycytidine deaminase class-switch recombination-defective hyper-IgM-2 mutants. *The Journal of biological chemistry*. 2012; 287:28007–16. [PubMed: 22715099]
44. Li J, Chen Y, Li M, Carpenter MA, McDougale RM, Luengas EM, et al. APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. *Journal of molecular biology*. 2014; 426:1296–307. [PubMed: 24361275]
45. Bennett RP, Salter JD, Liu X, Wedekind JE, Smith HC. APOBEC3G subunits self-associate via the C-terminal deaminase domain. *The Journal of biological chemistry*. 2008; 283:33329–36. [PubMed: 18842592]
46. Shlyakhtenko LS, Lushnikov AY, Miyagi A, Li M, Harris RS, Lyubchenko YL. Atomic force microscopy studies of APOBEC3G oligomerization and dynamics. *Journal of structural biology*. 2013; 184:217–25. [PubMed: 24055458]
47. Wedekind JE, Gillilan R, Janda A, Krucinska J, Salter JD, Bennett RP, et al. Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits. *The Journal of biological chemistry*. 2006; 281:38122–6. [PubMed: 17079235]
48. Chelico L, Prochnow C, Erie DA, Chen XS, Goodman MF. Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. *The Journal of biological chemistry*. 2010; 285:16195–205. [PubMed: 20212048]
49. Burnett A, Spearman P. APOBEC3G multimers are recruited to the plasma membrane for packaging into human immunodeficiency virus type 1 virus-like particles in an RNA-dependent process requiring the NC basic linker. *Journal of virology*. 2007; 81:5000–13. [PubMed: 17344295]
50. Friew YN, Boyko V, Hu WS, Pathak VK. Intracellular interactions between APOBEC3G, RNA, and HIV-1 Gag: APOBEC3G multimerization is dependent on its association with RNA. *Retrovirology*. 2009; 6:56. [PubMed: 19497112]
51. Huthoff H, Autore F, Gallois-Montbrun S, Fraternali F, Malim MH. RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. *PLoS pathogens*. 2009; 5:e1000330. [PubMed: 19266078]
52. Koyama T, Arias JF, Iwabu Y, Yokoyama M, Fujita H, Sato H, et al. APOBEC3G oligomerization is associated with the inhibition of both Alu and LINE-1 retrotransposition. *PloS one*. 2013; 8:e84228. [PubMed: 24367644]
53. Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:4102–7. [PubMed: 12651944]
54. Larijani M, Frieder D, Sonbuchner TM, Bransteitter R, Goodman MF, Bouhassira EE, et al. Methylation protects cytidines from AID-mediated deamination. *Molecular immunology*. 2005; 42:599–604. [PubMed: 15607819]

55. Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, et al. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature chemical biology*. 2012; 8:751–8. [PubMed: 22772155]
56. Cadet J, Wagner JR. TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine. *Mutation research Genetic toxicology and environmental mutagenesis*. 2014; 764–765:18–35.
57. Schmitz KM, Petersen-Mahrt SK. AIDing the immune system-DIAbolic in cancer. *Seminars in immunology*. 2012; 24:241–5. [PubMed: 22841422]
58. Wallace SS. Base excision repair: a critical player in many games. *DNA Repair (Amst)*. 2014; 19:14–26. [PubMed: 24780558]
59. Wang KY, Chen CC, Shen CK. Active DNA demethylation of the vertebrate genomes by DNA methyltransferases: deaminase, dehydroxymethylase or demethylase? *Epigenomics*. 2014; 6:353–63. [PubMed: 25111488]
60. Roberts SA, Gordenin DA. Clustered and genome-wide transient mutagenesis in human cancers: Hypermutation without permanent mutators or loss of fitness. *BioEssays: news and reviews in molecular, cellular and developmental biology*. 2014
61. Beletskii A, Bhagwat AS. Correlation between transcription and C to T mutations in the non-transcribed DNA strand. *Biological chemistry*. 1998; 379:549–51. [PubMed: 9628351]
62. Sohail A, Lieb M, Dar M, Bhagwat AS. A gene required for very short patch repair in *Escherichia coli* is adjacent to the DNA cytosine methylase gene. *Journal of bacteriology*. 1990; 172:4214–21. [PubMed: 2198248]
63. Bandaru B, Wyszynski M, Bhagwat AS. HpaII methyltransferase is mutagenic in *Escherichia coli*. *Journal of bacteriology*. 1995; 177:2950–2. [PubMed: 7751315]
64. Wu P, Qiu C, Sohail A, Zhang X, Bhagwat AS, Cheng X. Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4. *The Journal of biological chemistry*. 2003; 278:5285–91. [PubMed: 12456671]
65. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*. 2004; 25:1605–12. [PubMed: 15264254]
66. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*. 2014; 42:W252–8. [PubMed: 24782522]

Highlights

- Defines the catalytic domain of human APOBEC3B (A3B) enzyme and characterizes it in terms of kinetic constants, substrate preference and multimerization.
- Identifies a structural role for a conserved methionine at the junction of amino terminal and carboxy terminal domains.
- Determines kinetic constants for the catalytic domain and shows that the amino terminal domain is inactive.
- Determines the selectivity of A3B for cytosine compared to 5-methylcytosine.
- These results suggest that A3B is unlikely to play a major role in mutating or demethylating 5-methylcytosines in the human genome.

Figure 1. Structure and activity of APOBEC3B constructs

(A) Alignment between parts of A3G and A3B sequences is shown. The secondary structural elements within A3G [21] are shown as boxes above the sequences. The amino terminal end of 193-CTD, 195-CTD and 187-CTD are indicated by arrows. (B) Results of Kan^R reversion assay for 195-CTD in the absence of IPTG in the host BH214. The horizontal lines within the data points represent the median values. (C) Kan^R revertants due to 195-CTD following induction of its transcription using 50 μM or 100 μM IPTG. ** represents P value of <0.01. (D) The results of Kan^R reversion assay for the full-length A3B and its 193-CTD and 187-CTD constructs in BH214 in the absence of IPTG.

Figure 2. Kan^R reversion and Rif^R mutation assay for APOBEC3B mutants

(A) Effect of D194A mutation on Kan^R revertant frequency in full length A3B. Revertant frequencies due to WT gene and its catalytically inactive mutant E255A are also shown. (B) Effect of Met-193 mutations on Kan^R revertant frequency. Revertant frequencies due to Ala, Glu or Phe mutants are compared to the full length A3B and the E255A. ** represents P value of <0.01 and *** represents P value of <0.001. (C) Effect of Met-193 mutations on Rif^R frequency in BH214. Rif^R frequencies due to Ala or Glu mutants are compared to the full length A3B and the E255A.

Figure 3. Properties of purified A3B 193-CTD protein

(A) Image of Coomassie Blue-stained gels for purified 193-CTD along with the size markers. (B) The MALDI-TOF mass spectra of purified 193-CTD. The intensity (arbitrary units) is plotted against the m/z ratio. The molecular weight of the strong peak is marked on the spectrum. Inset- Enlargement of the peak around 47,000 Da. (C) Chromatography pattern of purified 193-CTD on Superdex-75 column. The UV absorbance (arbitrary units) is plotted against the V_e/V₀ ratio. Positions of different size standards used to calibrate the column are shown by downward arrows. The observed peaks are indicated by upward arrows and their calculated molecular weights are shown.

Figure 4. Preliminary characterization of purified APOBEC3B 193-CTD

(A) Stability of 193-CTD during reaction incubations. The ability of 193-CTD protein to convert cytosine to uracil is quantified following pre-incubations for different lengths of time. (B) Ability of 193-CTD to turnover. The amount of product generated by the enzyme is shown under two different ratios of substrate to enzyme.

Figure 5. Inability of APOBEC3B NTD to deaminate cytosines

(A) Kan^R reversion assay for A3B NTD. The revertant frequencies due to the NTD are compared to full length A3B and the empty vector. (B) Lack of cytosine deamination activity for partially purified NTD. The TCGA-40 substrate (2 pmol) was incubated with different amounts of the NTD alone (lanes 4, 6 and 8 respectively) or together with 10 pmol of 193-CTD (lanes 5, 7 and 9).

Figure 6. Ability of APOBEC3B to deaminate 5-methylcytosines

(A) Comparison of Kan^R revertant frequencies due to conversion of 5mC to T. The abilities of 193-CTD, 187-CTD and full length A3B to deaminate 5mC are compared with empty vector and A3A. ** represents P value of <0.01 and *** represents P value of <0.001. (B) Comparison of kinetics of cytosine and 5mC deamination by 193-CTD. DNA oligomers containing a single C or 5mC in the indicated sequence context were treated with increasing concentrations of purified 193-CTD and the products were quantified. (C) Kinetics of cytosine and 5mC deamination by partially purified A3A.

Figure 7. Sequence alignment of APOBEC3B CTD and APOBEC3 proteins with known structures

Clustal Omega alignment of APOBEC3 single domain proteins (A3A and A3C), CTD of two domain proteins (A3F and A3G) with A3B CTD is shown. The structural motifs are shown according the structure of A3G CTD [21]. The conserved methionine at the junction of NTD and CTD is marked with a green arrow. All the conserved residues between the five proteins are shown in blue and marked by asterisks. The residues that line the hydrophobic pocket in which the conserved methionine residues are shown in red and are marked with a dot.

Figure 8. Structure of the hydrophobic pocket in different APOBEC3 proteins

(A) Structure of the hydrophobic pocket in which A3G Met-197 resides (based on PDB entry 3IR2). Met-197 is shown in purple and yellow (sulfur), other side chains are shown in grey (carbon) and red (oxygen). Active site zinc ion is labeled as Zn1 and is coordinated to two cysteines and one histidine. Helices $\alpha 1$ and $\alpha 5$, and beta strands $\beta 1$, $\beta 2'$ and $\beta 3$ are marked. (B) Representation of the pocket in which A3G Met-197 resides based on hydrophobicity. Met-197 side chain is shown as a grey stick with sulfur in yellow. The hydrophobicity scale goes from red-orange (hydrophobic) to Dodger blue (polar) with white in the middle. (C) Representation of the pocket in which the conserved methionine resides in A3A (PDB: 2M65), A3C (3VOW) and A3F (4IOU) structures based on hydrophobicity. Same color scheme as part B. For A3A structure all the NMR conformers are shown.

Table 1

Kinetic Parameters of APOBEC3B 193-CTD

Name	Oligomer size (nt)	Sequence context	Total protein concentration* (μM)	Apparent K_M (μM)	Apparent v_{max}/K_M (min^{-1})
TCGA-40	40	ATCGA	0.20	73.9	8.9×10^{-3}
TCAG-30	30	ATCAG	0.13	47.8	9.1×10^{-3}
TCGG-30	30	ATCGG	0.13	37.3	8.3×10^{-3}

* The protein is assumed to be monomeric.

nt- nucleotides.