Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations

Anjum Sohail, Joanna Klapacz, Mala Samaranayake, Asad Ullah and Ashok S. Bhagwat*

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

ABSTRACT

Activation-induced cytidine deaminase (AID) is required for the maturation of antibodies in higher vertebrates, where it promotes somatic hypermutation (SHM), class switch recombination and gene conversion. While it is known that SHM requires high levels of transcription of the target genes, it is unclear whether this is because AID targets transcribed genes. We show here that the human AID protein promotes C to T mutations in Escherichia coli which are stimulated by transcription. The mutations are strand-biased and occur preferentially in the non-transcribed strand of the target gene. Human AID purified from E. coli is active without prior treatment with a ribonuclease and deaminates cytosines in plasmid DNA in vitro. Further, the action of this enzyme is greatly stimulated by the transcription of the target gene in a strand-dependent fashion. These results confirm the prediction that AID may act directly on DNA and show that it can act on transcribing DNA in the absence of specialized DNA structures such as R-loops. It suggests that AID may be recruited to variable genes through transcription without the assistance of other proteins and that the strand bias in SHM may be caused by the preference of AID for the non-transcribed strand.

INTRODUCTION

Activation-induced cytidine deaminase (AID) was first described as a protein that is required for somatic hypermutations (SHM) and class switch recombination (CSR) in B lymphocytes (1). Subsequently, it was also found to be required for the third mutational process involved in antibody maturation, gene conversion (2,3). AID is a homolog of a known RNA-cytosine deaminase, Apobec-1 (1), and the human AID gene causes C to T mutations in Escherichia coli (4). AID protein purified from insect cells and pre-treated with a ribonuclease was recently shown to deaminate cytosines to uracil in DNA (5), providing an explanation for its mutator phenotype.

It is not clear how AID causes base substitutions other than C to T that are observed during SHM, or how it creates the double-strand breaks that are thought to be necessary for CSR (6). Additionally, it is known that high transcription is a requirement for SHM in B cells as well as non-B cells (7–10) and that there is a strand bias in mutations introduced during SHM (11). It is unclear whether these properties of antibody maturation can be explained by biochemical properties of AID. We show here that the properties of human AID expressed in E. coli, and of the protein purified from E. coli are consistent with the dependence of antibody maturation on transcription and that its mutagenic action is strand-biased.

MATERIALS AND METHODS

Escherichia coli strains and plasmids

Escherichia coli strains BH156 (relevant genotype ung-1) and BH158 (=BH156 mgl::Tn10) have been described previously (12). BH214 is BH158 containing DE3 prophage, which contains the T7 RNA polymerase gene under the control of the lac promoter and was kindly provided by W. Franklin (Albert Einstein College of Medicine, New York, NY). BL21 (®, ompT, hsdS gal) containing DE3 is from our collection.

Plasmid pUP21 was described previously (13), but was referred to as pUP21-op75. PAUP21 was constructed from pUP21 by deleting an EcoRI–EcoO109I restriction fragment containing the UP-tac promoter from pUP21. To create pUP27, pUP21 was modified to introduce an XbaI site immediately upstream from the P kan and an NheI site downstream from the kan-ble cassette. The XbaI–NheI fragment in the resulting plasmid, pUP25, was inverted to create pUP27. The plasmids pAB7 and pAB8 were constructed by cloning a restriction fragment containing the P kan and kanS-94D elements into pET11d (Novagen Corp.). The orientation of the kanS-94D allele is the opposite with respect to the P T7 promoter in the two plasmids.

Purification of human AID

Human AID cDNA (I.M.A.G.E. Consortium ID 4054915) was obtained from the American Type Culture Collection (Rockville, MD) and the gene was amplified using the primers...
Human AID is a DNA cytosine deaminase

Human AID was expressed in E.coli as a fusion with GST and purified over an affinity column. A 56 bp DNA duplex containing a 5 nt bubble (Fig. 1A) was sequentially treated with AID, UDG and NaOH. If AID deaminates cytosines in this DNA to uracil, UDG would excise these uracils, and cleavage at the resulting abasic sites by NaOH would create shorter DNA fragments. Treatment of this substrate with AID resulted in the conversion of 17% of labeled DNA into a single, shorter oligonucleotide (Fig. 1B, lane 3). This oligomer is 27 nt in length and its length is consistent with the conversion of one of the two cytosines in the bubble to uracil (Fig. 1B). Remarkably, under the reaction conditions used, the other cytosine in the bubble and the 12 cytosines in the rest of the DNA strand appeared untouched by AID.

However, the other cytosine in the bubble and some of the other cytosines in the duplex become susceptible to AID under different reaction conditions (Fig. 1C) or if the bottom DNA strand is used as a substrate without its 5' end with 32P. The cytosine that is the target of AID is indicated by a solid arrow. (B and C) 32P-labeled substrate was incubated with the reagents indicated above each lane and then heated in the presence of NaOH. The buffer contained 50 mM Tris–HCl (B) or 20 mM Tris–HCl (C). Φ is 1,10-phenanthroline.

RESULTS

Human AID was expressed in E.coli as a fusion with GST and purified over an affinity column. A 56 bp DNA duplex containing a 5 nt bubble (Fig. 1A) was sequentially treated with AID, UDG and NaOH. If AID deaminates cytosines in this DNA to uracil, UDG would excise these uracils, and cleavage at the resulting abasic sites by NaOH would create shorter DNA fragments. Treatment of this substrate with AID resulted in the conversion of 17% of labeled DNA into a single, shorter oligonucleotide (Fig. 1B, lane 3). This oligomer is 27 nt in length and its length is consistent with the conversion of one of the two cytosines in the bubble to uracil (Fig. 1B). Remarkably, under the reaction conditions used, the other cytosine in the bubble and the 12 cytosines in the rest of the DNA strand appeared untouched by AID.

However, the other cytosine in the bubble and some of the other cytosines in the duplex become susceptible to AID under different reaction conditions (Fig. 1C) or if the bottom DNA strand is used as a substrate without its sequence complement (not shown). We also found that AID action is inhibited by 1,10-phenanthroline, a strong chelator, but not by the weaker chelator EDTA (Fig. 1B, lanes 4 and 5). These results show that AID has a strong preference for unpaired cytosines and requires a tightly bound metal ion for its action.

Figure 1. Deamination of cytosines in a DNA bubble by AID. (A) The sequence of the duplex used as substrate for the AID reaction is shown. The bottom strand is labeled at the 5‘ end with 32P. The cytosine that is the target of AID is indicated by a solid arrow. (B and C) 32P-labeled substrate was incubated with the reagents indicated above each lane and then heated in the presence of NaOH. The buffer contained 50 mM Tris–HCl (B) or 20 mM Tris–HCl (C). Φ is 1,10-phenanthroline.
Cytosine deamination by AID in *E. coli* is transcription-dependent

We used an *E. coli* plasmid-based genetic reversion assay to study the transcription-dependence of AID mutagenesis. An allele of the *kan* gene, *kanS-94D*, reverts only through C to T changes in cells lacking UDG (genotype *ung*), conferring Kan\(^R\) phenotype upon its host (14). If AID deaminates cytosines in DNA to uracil, there would be an increase in Kan\(^R\) revertants and would be scored as such. In one of the plasmids used, pAUP21, *kan* is transcribed from a weak constitutive promoter (P\(_{kan}\)), while in the other plasmids *kan* is either transcribed from an inducible UP-tac promoter or a combination of UP-tac and P\(_{kan}\) (15) (Fig. 2A). Previous studies have shown that transcription from an induced UP-tac promoter dominates over transcription from P\(_{kan}\) (15,16); thus different DNA strands of the gene are predominantly transcribed in pUP21 and pUP27 when IPTG is present in the growth media.

The human AID gene was cloned into pSU24 (17) (pSU-AID) under the control of an inducible T7 promoter and introduced into *E. coli*. In cells containing pAUP21, AID expression results in increased Kan\(^R\) revertants in the cultures. The Lea–Coulson method of the median (18) was used to determine the number of revertants per culture, m, and AID expression increased m by ~12-fold compared to the uninduced cultures (310 compared to 25; Table 1). In control cultures lacking AID, there was little increase in m upon addition of IPTG to the growth media (Table 1).

Similar experiments were also performed using cells containing pUP21. When transcription was induced from the strong promoter in front of *kan* in cells lacking AID, m increased ~4-fold (Table 1). This increase in mutations is due to spontaneous cytosine deaminations in the non-transcribed strand during transcription from UP-tac, and has been described previously (16,19). When parallel experiments were performed using cells containing pUP21 and pSU-AID, induction of transcription increased m by ~50-fold (Table 1). The number of reversions per culture obtained when transcription of both AID and *kan* genes was induced (4545) is in vast excess of the sum of revertants obtained as a result of overexpression of AID in *E. coli* (310) and transcription of *kan* (158). Thus, the extent of transcription of the target gene directly correlates with increased AID mutagenesis.

Further, this increase in mutations was dependent on which of the two DNA strands was being copied by the RNA polymerase. When the experiments were repeated using pUP27 instead of pUP21 (Fig. 2A), the increase in mutations caused by AID was modest (m = 289) and was comparable to the increase observed for *kan* in pAUP21 (310; Table 1). The target cytosine is in opposite strands with respect to UP-tac in the two plasmids and m is ~16-fold lower with pUP27 compared to pUP21 (Table 1). These data suggest that AID acts preferentially on cytosines in the non-transcribed strand of *kan*.

Sixty-two independent Kan\(^R\) revertants from different experiments were sequenced and all contained a C to T change in codon 94 of *kan* (not shown). Our observations are consistent with the spectra of rifampicin-resistance mutations promoted by AID in *E. coli* (4). AID increased C to T mutations in the non-transcribed strand of *rpoB* from 30 to 63%, while G to A transitions in this strand increased from 1.4 to only 16% (4).

Table 1. Kan\(^R\) revertants promoted by AID *in vivo*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No IPTG</td>
<td>32 (123)</td>
<td>25 (119)</td>
<td>41 (167)</td>
<td>91 (372)</td>
<td>25 (96)</td>
<td>21 (84)</td>
</tr>
<tr>
<td>With IPTG</td>
<td>55 (168)</td>
<td>310 (1683)</td>
<td>158 (555)</td>
<td>4545 (30 455)</td>
<td>30 (79)</td>
<td>289 (1467)</td>
</tr>
</tbody>
</table>

*aGenetic reversion assays were performed as described previously (13). The host was BH214 (relevant genotype *ung*); the results from eight or more independent cultures are shown. m is the number of mutants per culture (18) and M is the median number of mutations per culture normalized to 10⁸ viable cells. The mutation rate is directly related to m (18).*
In vitro cytosine deaminations by AID are also transcription-dependent

Plasmid pAB7 containing kanS-94D (Fig. 2B) was incubated with GST-AID under a variety of conditions and the DNA was introduced into ung E.coli to score KanR revertants. Incubation of the plasmid DNA with GST-AID increased the frequency of KanR revertants and the magnitude of this increase was dependent on the reaction conditions. While the increase was ~5-fold in the absence of transcription, it was ~1600-fold when the gene was actively transcribed by T7 RNA polymerase (Fig. 3A). The increase in KanR frequency was almost completely eliminated when E.coli UDG was included in the reaction suggesting that AID had converted the target cytosine to uracil. The residual increase in KanR frequency by AID in the presence of UDG is most likely due to some uracils escaping repair. If non-transcribing plasmid DNA is treated with GST-AID in the presence of UDG, no increase in KanR revertants is observed compared to the control reaction (not shown).

Treatment of pAB8 with GST-AID also increased KanR frequency, but this effect was not strongly dependent on transcription (Fig. 3B). In contrast to pAB7, transcription caused only a slight increase in KanR revertants; the maximum revertant frequency obtained with pAB8 was ~50-fold lower than what was observed with pAB7 (Fig. 3). Total RNA was quantified from representative reactions and all the reactions were found to contain comparable amounts of RNA (not shown). Consequently, the differences in the mutagenicity of AID for pAB7 and pAB8 are not due to poor transcription of the latter plasmid and must be related to the transcriptional orientation of kan within these plasmids.

DISCUSSION

We have shown that the human AID protein is an active DNA cytosine deaminase both in vivo and in vitro, and AID mutagenesis is stimulated by transcription. Further, the protein targets cytosines in the non-transcribed strand. After this work was completed, similar results were reported from other laboratories (5,20,21). Additionally, purified Apobec-1 was shown to deaminate cytosines in single-stranded DNA in vitro (22). While the results presented here are generally consistent with the results of those studies, they clarify the biochemical activity and substrate specificity of AID in important ways.

It was reported that GST-AID purified from an insect expression system was inactive unless treated with a ribonuclease (5). The sequence of the RNA bound to GST-AID is unknown, but it was speculated that it may have a regulatory role in AID action (5). In contrast to those results, we did not find it necessary to include a ribonuclease in AID reactions to detect its biochemical activity. To assess whether a significant fraction of GST-AID purified from E.coli was bound to RNA, the strand cleavage assay was repeated with the inclusion of ribonuclease A in the reaction. Ribonuclease treatment did not improve the activity of AID (Fig. 1C) and may have decreased it slightly. These data show that unlike the GST-AID purified from insect cells, the protein purified from E.coli does not contain RNA and is active. Further, if AID does bind RNA (5), it is likely to bind specific RNA sequences. It should also be noted that Chaudhari et al. (20) did not treat the partially purified AID from B-cells with ribonucleases to demonstrate its cytosine deaminase activity. Consequently, whether AID binds RNA and whether such binding has biological significance remains unclear.

It has been argued that the segment of DNA where CSR occurs (S region) may contain RNA–DNA hybrids (R-loops), and that such structures are essential for recombination (23–25). Further, Chaudhuri et al. (20) showed that partially purified AID targets the unpaired DNA strand when such R-loops are formed as a result of transcription in vitro. We have shown here that R-loop formation is not essential for AID action (Fig. 3). Normal transcription of DNA, which presumably creates transient transcription bubbles, may be all that is necessary for AID action.

Finally, Bransteitter et al. (5) have argued that the propensity by AID to cause mutations within WRCY sequences (W = A or T; R = purine and Y = pyrimidine) is the result of preferential targeting of AID at these sites. The data presented here show that such preferences may depend on specific reaction conditions (compare Fig. 1B with C). Only one of the cytosines in the DNA bubble is within a WRCY sequence (product 27 nt), but both cytosines are converted to uracil at about the same frequency under one reaction condition (Fig. 1C). Hence, conclusions regarding the specificity of AID must await a more thorough biochemical analysis than has been presented so far.

It has been argued that the dependence of SHM on the transcription of immunoglobulin genes (8,26) is the result of the action of a mutator protein that associates with the transcription elongation complex (9). The results presented here suggest that AID is likely to be this mutator. It is also unknown whether the targeting of the transcribing immunoglobulin genes by this mutator requires additional proteins (26,27). Our observation that AID targets a highly transcribed gene in vitro and in E.coli suggests that the targeting of AID may not require additional mammalian proteins. Once AID is brought to the region of the chromatin containing immunoglobulin genes, it may find the actively transcribed variable gene and help mutate it. The strand bias in SHM has variously been interpreted as evidence for transcription-dependent damage to DNA (28,29), or the involvement of
transcription-coupled repair in eliminating DNA damage (9). While it is clear that downstream events may process uracils generated by AID and influence the mutation spectrum of SHM, it is tempting to suggest that the strand selectivity of AID with respect to transcription is the source of the mutational strand bias.

ACKNOWLEDGEMENTS

We would like to thank T. Roy and J. McLellan for assistance with initial experiments and P. Gearhart (National Institute of Aging) for useful comments on the manuscript.

REFERENCES