Enzyme cycling contributes to efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B

Madison B. Adolph, Robin P. Love, Yuqing Feng and Linda Chelico*

Department of Microbiology and Immunology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Received July 12, 2017; Revised September 07, 2017; Editorial Decision September 08, 2017; Accepted September 11, 2017

ABSTRACT

The single-stranded DNA cytidine deaminases APOBEC3B, APOBEC3H haplotype I, and APOBEC3A can contribute to cancer through deamination of cytosine to form promutagenic uracil in genomic DNA. The enzymes must access single-stranded DNA during the dynamic processes of DNA replication or transcription, but the enzymatic mechanisms enabling this activity are not known. To study this, we developed a method to purify full length APOBEC3B and characterized it in comparison to APOBEC3A and APOBEC3H on substrates relevant to cancer mutagenesis. We found that the ability of an APOBEC3 to cycle between DNA substrates determined whether it was able to efficiently deaminate single-stranded DNA produced by replication and single-stranded DNA bound by replication protein A (RPA). APOBEC3 deaminase activity during transcription had a size limitation that inhibited APOBEC3B tetramers, but not APOBEC3A monomers or APOBEC3H dimers. Altogether, the data support a model in which the availability of single-stranded DNA is necessary, but alone not sufficient for APOBEC3-induced mutagenesis in cells because there is also a dependence on the inherent biochemical properties of the enzymes. The biochemical properties identified in this study can be used to measure the mutagenic potential of other APOBEC enzymes in the genome.

INTRODUCTION

The APOBEC family of enzymes in humans has 12 members and is comprised of RNA and DNA cytidine deaminases (1). The enzymes are named after APOBEC1, the first family member discovered, that edits the mRNA of apolipoprotein B in the intestine by converting cytosine to uracil and forming a stop codon (1). Other family members with demonstrated deaminase activity have roles in immunity. Activation induced cytidine deaminase (AID) converts cytosine to uracil in single-stranded (ss) DNA created during transcription of immunoglobulin genes in activated B cells (2). These uracils promote antibody maturation and class switching (2). The APOBEC3 enzyme family has seven members (A–H, excluding E) that are normally expressed in germ cells, CD4+ T cells, or myeloid derived cells and are able to convert cytosine to uracil in ssDNA of invading viruses, endogenous retroviruses, and retrotransposons (3). These uracils result in mutagenesis and functional inactivation of the virus (3). APOBEC3A can also edit mRNAs in monocytes and macrophages (4). In the APOBEC3 family some of the enzymes have only one Zn-coordinating deaminase domain (A, C, H) and others have two (B, D, F, G). There are also other members such as APOBEC2, APOBEC4, and APOBEC5 that have no in vitro catalytic activity, although APOBEC2 appears to be involved in muscle development (5–8). Despite the diverse beneficial functions of this family, there is one unifying characteristic; if the enzymes are expressed at the wrong time or in the wrong cell type and have access to the nucleus they can deaminate cytosines in genomic DNA and this can contribute to cellular transformation, cancer, and the ongoing mutagenic processes in tumors (9–11). Thus far, three APOBEC3 (A3) enzyme family members have been characterized to contribute to cancer mutagenesis, APOBEC3A (A3A), APOBEC3B (A3B) and APOBEC3H haplotype I (A3H hap I) (12–16).

A3B serves as a good example for the role of A3s in cancer since it has been most extensively studied in this context. A3B is localized to the nucleus and has cytidine deaminase activity on ssDNA (17,18). When A3B expression occurs in the wrong context, A3B promotes genomic instability at the nucleotide level that results in contributions to the mutations that arise in breast cancer and leads to resistance to the chemotherapeutic drug tamoxifen (12,19). A3B can also contribute to mutations that arise in a number of other cancers (13,20–27). A3B recognizes 5′TC motifs that are preferably preceded by a G or A, and this sequence context of mutations has been found in cancer genomes (12,13,28). It is thought that during DNA replication, the cytosines in
the exposed ssDNA on the lagging strand are susceptible to deamination by A3B (29–33). If left unrepaired, these uracils will template addition of the incorrect base leading to C/G to T/A transition mutations (11). The uracils can also induce the formation of double-strand (ds) DNA breaks or be repaired in an error-free or error-prone manner (11).

Despite this understanding of A3B activity at a cellular level, biochemical characterizations have centered on using truncated forms of A3B that contain only the catalytically active C-terminal domain (CTD) (12,13,28,34–39). A biochemical analysis of full-length (fl) wild-type A3B on substrates relevant to cancer mutagenesis is lacking. Owing to poor solubility of A3B in heterologous Escherichia coli expression systems and its mutagenic activity, purification of fl A3B has posed difficulties (35,38,40,41). Although the N-terminal domain (NTD) of A3B is not catalytically active, it does contribute to activity and the CTD alone is 10-fold less active than fl A3B as measured by a mutator assay where A3B is expressed in E. coli or the in vitro activity of maltose binding protein tagged A3B (35,39). We have little knowledge of why the full length enzyme is more active than the CTD, how fl A3B can access ssDNA in the genome, if it requires cofactors, or how it competes with other ssDNA binding proteins such as replication protein A (RPA) (34,36–38). Although A3A and A3H have been previously characterized biochemically, there is similarly little information on how at a biochemical level these enzymes would access ssDNA in the genome (16,42–45).

To better understand the biochemical mechanisms underlying the ability of A3B, A3H hap I, and A3A to induce genomic mutagenesis we undertook an analysis of these enzymes on ssDNA and used in vitro phage based systems of transcription and replication to determine if these enzymes alone can deaminate dynamically produced ssDNA. Importantly, to complete this research, we purified fl A3B and for the first time provide a characterization of the enzyme, which shows that it is different than the commonly used A3B CTD, suggesting that the A3B CTD is not a functional substitute. We also find that deamination activity during DNA synthesis and when ssDNA is bound by RPA requires an enzyme that can rapidly cycle between DNA substrates. For deamination during transcription, we found that the larger oligomeric state of A3B inhibited its activity. These biochemical studies provide evidence that the intrinsic characteristics of the enzymes have a contribution to the mutagenic process, rather than the availability of ssDNA alone.

The inherent biochemical characteristics common to A3A, A3B, and A3H hap I can be used to measure the mutagenic potential of other APOBEC enzymes in the genome.

**MATERIALS AND METHODS**

**Cloning and site directed mutagenesis**

The fl A3B (a.a. 1–382), A3B 187 CTD (a.a. 187–382), A3B 193 CTD (a.a. 193–382), and A3H hap I proxy (A3H hap I*, G105R) sequences were cloned into a baculovirus transfer vector (pFAST-bac1) containing an N-terminal GST tag as described previously (45). The previously described pAcG2T GST-A3G construct was also used (46). All constructed plasmids were verified by DNA sequencing.

**Protein expression and purification**

Recombinant baculovirus production for expression of proteins in S9 cells was carried out using the pACG2T or pFast-bac1 transfer vector as previously described (45,47). S9 cells were infected with recombinant GST-A3 virus at an MOI of 1 (A3B 187 CTD and A3B 193 CTD) or an MOI of 10 (fl A3B). The infection conditions for A3A, A3G, and A3H have been previously described (42,45,48). Recombinant baculovirus infected S9 cells were harvested after 72 h of infection. Cells lysates treated with RNaseA were incubated with glutathione-Sepharose 4B resin (GE Healthcare) at 4°C and were subjected to a series of salt washes, as previously described (48). For all the enzymes, except the fl A3B, on-column cleavage from the GST tag with thrombin (GE Healthcare) was performed at 21°C for 18 h in thrombin digestion buffer (20 mm HEPES, pH 7.5, 150 mm NaCl, 10% glycerol, and 1 mm DTT). Due to the lower yield of full-length A3B, the enzyme was eluted with the GST tag in elution buffer (100 mM Tris, pH 8.8, 150 mM NaCl, 10% (v/v) glycerol, and 50 mM reduced glutathione). A3B was then cleaved from the GST tag in solution at 21°C for 4 h with Thrombin (GE Healthcare) before being dialyzed against 100 mM Tris pH 7.5, 250 mM NaCl, 10% (v/v) glycerol, and 1 mM DTT. To purify A3B from the free GST and thrombin, the enzyme stock was diluted to achieve a solution of 50 mM Tris pH 7.5, 50 mM NaCl, 10% (v/v) glycerol and 1 mM DTT for loading onto a DEAE FF column (GE Healthcare). A3B was eluted with a linear gradient of NaCl. The A3B eluted at approximately 300 mM NaCl. Enzymes were assessed as >90% pure by SDS-PAGE (Supplementary Figure S1). The RPA was generously provided by Dr John Turchi (Indiana University School of Medicine).

**In vitro deamination assays**

All ssDNA substrates were obtained from Tri-Link Biotechnologies and have been previously described or are in Supplementary Table S1 (45,49). Substrates with complementary oligonucleotides annealed were formed by mixing the oligonucleotides in buffer containing 50 mM Tris, pH 7.5 and 100 mM NaCl before heating at 95°C for 10 min and then slowly cooling (1°C per minute). For all deamination assays, except the R-loop, the substrates contained specific deamination motifs for each A3 enzyme which were: 5′ TTC (A3A), 5′ ATC (A3B), 5′ CTC (A3H hap I*), or 5′ CCC (A3G). The R-loop substrates used only 5′ TTC (A3A, A3B, and A3H hap I*) or 5′ CCC (A3G). Unless otherwise indicated, reactions were conducted at 37°C in RT buffer (50 mM Tris, pH 7.5, 40 mM KCl, 10 mM MgCl2 and 1 mM DTT).

Proccessivity reactions were carried out under single-hit conditions (i.e. <15% substrate usage) to ensure a single enzyme-substrate encounter. Under these conditions, a processivity factor can be determined by comparing the quantified total amount of deaminations occurring at two sites on the same ssDNA with a calculated theoretical value of deaminations at these two sites if the deamination events were uncorrelated (not processive). Since the processivity factor is a ratio, a value of 1.0 means that the enzyme is not processive. Alternatively, a non-processive enzyme may not have a visible amount of deamination at two sites under the
single-hit conditions of the reaction. An ssDNA substrate containing two deamination motifs (100 nM) was incubated with 50 nM of enzyme for 2.5–20 min. The reaction time was varied on each ssDNA according to the specific activity of the enzymes to ensure ~10% substrate usage. Reactions were started by the addition of the ssDNA substrate. The specific activity was calculated from these single-hit condition reactions by determining the picomoles of substrate used per minute for a microgram of enzyme.

For the time course of A3B in comparison to the CTD mutants, 100 nM of a 118 nt ssDNA was incubated with 50 nM (A3B) or 2 μM of the CTD mutants for 5–60 min (A3B) or 30–120 min (CTD mutants) in order to compare their activities. Reactions were started by the addition of the ssDNA substrate.

For intersegmental transfer assays, the A3:ssDNA ratio was kept the same, but increasing concentrations of enzyme and substrate was titrated in (118 nt ssDNA: 100–500 nM, A3: 50–250 nM). Reactions were started by the addition of the ssDNA substrate.

For the detection of deaminations on R-loop substrates 100 nM of an R-loop substrate was incubated with 100 nM A3 enzyme for 5–20 min. Reactions were started by the addition of the ssDNA substrate.

For cycling assays, 100 nM of A3A, A3B, A3H hap I*, or A3G was incubated for 3 min at 21°C with 0 nM (0×), 500 nM (5×) or 5000 nM (50×) of an unlabeled 69 nt ssDNA. To start the reaction, 100 nM of a labeled 85 nt ssDNA was added and the reaction was allowed to proceed at 37°C for 10 min.

For deamination in the presence of RPA, saturating amounts of RPA (300 nM) was preincubated with 100 nM of ssDNA for 5 min before the addition of 50 nM A3 enzyme to initiate the reaction. Reactions proceeded for 10–30 min.

All A3-catalyzed deaminations were stopped using a Phenol: chloroform extraction and cleaned using two additional chloroform extractions. The deaminations were detected by treating the substrates with uracil DNA glycosylase (New England Biolabs) and heating under alkaline conditions before resolving the fluorescein-labeled ssDNA on 10, 16 or 20% (v/v) denaturing polyacrylamide gels, depending on the sizes of the ssDNA fragments. Gel photos were obtained using a Typhoon Trio multipurpose scanner (GE Healthcare) and integrated gel band intensities were analyzed using ImageQuant (GE Healthcare).

**Electrophoretic mobility shift assay**

A fluorescein labeled 70 nt ssDNA (100 nM) was incubated with 300 nM of RPA for 5 min at 21°C in the presence of RT buffer and 40% glycerol to facilitate gel loading. The samples were run on a 5% (v/v) Native-PAGE gel at 4°C. Gel images were obtained using a Typhoon Trio multipurpose scanner (GE Healthcare).

**Size exclusion chromatography**

The oligomerization states of the enzymes were determined by loading 10 μg of purified enzyme on a 10 mL Superdex 200 (GE Healthcare) size exclusion column. The column was prepared by pouring the resin bed in a column with 16 cm height and 0.5 cm diameter. The running buffer contained 50 mM Tris pH 8.0, 200 mM NaCl and 1 mM DTT. The Bio-Rad standard set was used to generate a standard curve from which molecular weight and oligomerization states of the enzymes were determined.

**In vitro transcription assay**

Transcription-dependent deamination reactions were performed as described previously (47,50). The dsDNA substrates contain a T7 RNA polymerase promoter and a single 5′TTC (A3A, A3B), 5′CTC (A3H hap I*), or 5′CCC (A3G, A3G FW) deamination motif on the non-transcribed strand. Substrates were previously reported, except for the 5′CTC substrate, but the DNA sequences are identical other than the deamination motif (42). The dsDNA substrates (30 nM) were reacted with 90 nM A3A, A3B, A3H hap I*, A3G or A3G FW in the presence of ribonucleotide triphosphates (500 μM), T7 polymerase (1 unit, Promega), and DNase-free RNase A (5 ng/μl, Roche Applied Science) in transcription buffer (50 mM Tris, pH 7.4, 10 mM MgCl2, 1 mM DTT) at 37°C. Control reactions either contained no T7 Polymerase or no A3 enzyme to ensure that deamination was transcription dependent. Sequencing of the non-transcribed strand for detection of deamination was performed using Thermo Sequenase (Affymetrix), as described previously (50).

**In vitro DNA replication assay**

Deamination of ssDNA by A3 enzymes during DNA synthesis was assessed using a circular template (pUC19). The pUC19 (0.05 nM) was heated at 95°C for 1 min in the presence of 200 nM random hexamer primers, 4 μg BSA, 200 μM dNTPs, and Phi29 buffer (Thermo Fisher), and then cooled to 4°C. The primed pUC19 was then incubated with Φ29 DNA polymerase that initiates rolling circle replication and strand displacement synthesis on subsequent DNA synthesis rounds, in the presence or absence of 200 nM A3 enzyme. The reaction was incubated at 30°C for 6 hr and inactivated at 65°C for 10 min before treatment with DpnI (New England Biolabs) for 1 h at 37°C to remove contaminating plasmid DNA. A 345 nt region of the plasmid (nucleotides 473–818) was PCR amplified using Pfu Ultra Turbo Hotstart (Agilent Technologies) that can use uracils as a template with high fidelity. These amplicons were then cloned with the CloneJET PCR cloning kit (Thermo Fisher). At least 25 clones were sequenced with primers specific to pUC19 at the National Research Council of Canada (Saskatoon, Canada).

**Steady state rotational anisotropy**

Steady state fluorescence depolarization (rotational anisotropy) was used to measure the binding affinity of the enzymes to fluorescein-labeled ssDNA, DNA/RNA, R-loop, and dsDNA. Reactions were 60 μl and contained fluorescein-labeled DNA substrate (10 nM) in RT buffer and the enzyme was titrated into the solution until saturation. A QuantaMaster QM-4 spectrofluorometer (Photon...
Technology International) with a dual emission channel was used to collect data and calculate anisotropy. Samples were excited with vertically polarized light at 495 nm (6 nm band pass) and vertical and horizontal emissions were measured at 520 nm (6 nm band pass). Apparent dissociation constants ($K_d$) were obtained by fitting to a rectangular hyperbola or sigmoidal curve using Sigma Plot 11.2 software.

RESULTS

A3B is a processive enzyme that forms higher order oligomers

The purification of A3B has posed difficulties because even the basal expression levels of A3B from eukaryotic vectors when being amplified in *E. coli* results in the purification of mutated A3B vector sequences (38,40,41). Some labs have overcome this by inserting an intron in the A3B (38,41). However, since both A3 enzyme yield and activity are highest when expressed from recombinant baculovirus infected S9 cells (51), and this system is not amenable to splicing of intron containing A3B, we were unable to overcome expression difficulties in this manner (data not shown) (52). Instead, we made a GST-tagged version of A3B that inactivated the enzyme and amplified this GST-A3B clone in a baculovirus transfer vector for subsequent generation of recombinant baculovirus. We confirmed that the GST-A3B becomes inserted in the baculovirus genome with no mutations (data not shown). The GST-A3B is then expressed and the GST tag is cleaved during purification to impart A3B activity.

The purified fl A3B has a specific activity on ssDNA of 1.60 pmol/μg/min, which is similar to A3G (Figure 1A and Table 1). In comparison to the literature, the S9 insect cell produced fl A3B is 32-fold more active than the Maltose Binding Protein tagged fl A3B produced from *E. coli* (39). We also produced in S9 cells the two commonly used A3B CTD enzymes that contain amino acids 187–392 (187 CTD) or 193–392 (193 CTD) of A3B (34,35). The fl A3B is still 40-fold more active than the A3B 193 CTD that was reported to have increased activity over the more commonly used A3B 187 CTD (Figure 1A, Table 1, and Supplementary Figure S2) (34,35). We did not detect in vitro catalytic activity for the A3B 187 CTD (Figure 1A, Table 1, and Supplementary Figure S2). Thus, despite the A3B NTD not having catalytic activity, it appears to be able to facilitate enzyme activity. This is common to other A3 enzymes that contain two Zn$^{2+}$ coordinating domains, such as A3G and A3F (49,53). The NTD in these enzymes is a processivity domain and mediates oligomerization (49,53,54). The fl A3B may also oligomerize since it bound to ssDNA cooperatively as demonstrated by a best fit of the binding data to a sigmoidal binding curve (Table 1 and Supplementary Figure S3A). To determine if fl A3B oligomerization occurred in solution and if it was different between fl A3B and the A3B CTDs, we used size exclusion chromatography (SEC). The SEC showed that fl A3B (46 kDa) formed predominantly tetramers (184 kDa) and a small proportion existed as dimers (92 kDa) (Figure 1B and Supplementary Figure S4A-B). In contrast, the A3B 187 CTD was monomeric and the A3B 193 CTD was primarily monomeric with a small proportion of dimers, consistent with an earlier report (Figure 1C-D, Supplementary Figure S4C-D) (35). However, both A3B CTDs bound ssDNA non-cooperatively with an apparent dissociation constant ($K_d$) that was ~4-fold higher than the fl A3B (Table 1, and Supplementary Figure S3B, best fit to a rectangular hyperbola). This suggests that the CTD is mainly a catalytic center and that the NTD promotes ssDNA interaction and self-interaction. Altogether, the data strongly support that the A3B CTD is not a suitable replacement for the full-length enzyme and we continued our study with fl A3B only (hereafter referred to as A3B).

Since A3B is an ssDNA binding enzyme that deaminates only in a specific recognition motif, it must first find the motifs by searching the ssDNA non-specifically through a DNA scanning process called facilitated diffusion (55–58). In facilitated diffusion, electrostatic interactions of the enzyme with the DNA facilitate a search by enabling the enzyme to remain bound longer to the DNA than the time it spends in the bulk solution not bound to a substrate. The ssDNA scanning mechanism enables the enzyme to be processive and deaminate more than one cytosine in a single enzyme-substrate encounter. Other A3 enzymes have been found to use facilitated diffusion (45,46,49,59). Facilitated diffusion is also used by restriction enzymes and for DNA repair proteins, such as Uracil DNA glycosylase (55,60).

Using an in vitro deamination assay on a synthetic ssDNA substrate containing two A3B 5′ATC deamination motifs we can measure processive deaminations that occurred by facilitated diffusion (46). Essential to this essay is that the reactions are carried out under single-hit conditions ($<15\%$ substrate usage) to ensure that each ssDNA substrate was acted upon by only one enzyme during the course of the reaction (61). If under single-hit conditions we observe deamination of both 5′ATC motifs we can conclude that the enzyme is processive and deaminated both motifs in a single enzyme-substrate encounter. To decrease the chance that cooperative binding of different A3B molecules at different times on the ssDNA resulted in an apparent processivity, the reactions took place with an excess of ssDNA to promote interaction of single molecules of A3B with the ssDNA, which may be either dimer or tetramer molecules from solution (Figure 1B). We also calculated a processivity factor which is a ratio of the observed deaminations at both 5′ATC sites to the calculated theoretical number of deaminations that would occur independently at both 5′ATC motifs if the enzyme were non-processive (see Materials and Methods) (46). Facilitated diffusion encompasses a range of movements termed sliding, jumping or hopping, and intersegmental transfer (3,55,57,62). The sliding occurs along the phosphate backbone and for A3 enzymes moving on ssDNA is limited to a scanning length of 20 nt or less (49,63). The jumping, hopping, and intersegmental transfer are long range movements where the enzyme can use a 3D search to sample distal DNA regions either by microscopic dissociations and reassociations along the DNA (jumping or hopping) or through a doubly bound state (intersegmental transfer) (3,55,57,62).

To determine if A3B could slide along the ssDNA phosphate backbone, we conducted the deamination experiment with 5′ATC motifs that were separated by only 5- or 14-nt. A3B is able to slide and had a processivity factor of 4 to
Figure 1. A3B NTD mediates enzyme activity and oligomerization. (A) Time course of A3B, A3B 193 CTD, and A3B 187 CTD on a 118 nt fluorescently labeled ssDNA with two 5′ATC deamination motifs spaced 63 nt apart. Reactions were performed with 100 nM substrate DNA and 50 nM (fl A3B) or 2 μM (187 CTD, 193 CTD) for the indicated amount of time (5–60 min fl A3B or 30–120 min CTD mutants). (B-D) Size exclusion chromatography profiles of 10 μg (B) fl A3B, (C) 193 CTD, and (D) 187 CTD from a 10 ml Superdex 200 column was used to calculate the oligomerization state of the enzyme from a standard calibration curve. An ‘M’ denotes a monomer fraction, a ‘D’ denotes a dimer fraction, and a ‘T’ indicates a tetramer fraction. (A) fl A3B formed tetramers (apparent molecular weight 184 kDa) and dimers (apparent molecular weight 92 kDa). (C and D) 193 CTD and 187 CTD resolved as monomers (apparent molecular weight 23 kDa). The chromatograms were constructed by analyzing the integrated gel band intensities of each protein in each fraction after resolution by SDS-PAGE (Supplementary Figure S4).

Table 1. Deamination activity and binding of APOBEC3 enzymes on oligonucleotide substrates

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Specific activity (pmol/μg/min)</th>
<th>Apparent K_d, ssDNA (Hill coefficient)</th>
<th>Specific activity, R-loop (pmol/μg/min)</th>
<th>Apparent K_d, R-loop</th>
<th>Apparent K_d, DNA/RNA (Hill coefficient)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fl A3B</td>
<td>1.60 ± 0.05</td>
<td>0.35 ± 0.02 μM (2.8)</td>
<td>0.54 ± 0.09</td>
<td>0.90 ± 0.05 μM</td>
<td>0.31 ± 0.03 μM</td>
</tr>
<tr>
<td>A3B 187 CTD</td>
<td>None detected</td>
<td>1.51 ± 0.50 μM</td>
<td>Not determined</td>
<td>Not determined</td>
<td>Not determined</td>
</tr>
<tr>
<td>A3B 193 CTD</td>
<td>0.04 ± 0.01</td>
<td>1.30 ± 0.34 μM</td>
<td>Not determined</td>
<td>Not determined</td>
<td>Not determined</td>
</tr>
<tr>
<td>A3A</td>
<td>1.10 ± 0.10</td>
<td>9.10 ± 2.50 μM (1.7)</td>
<td>0.25 ± 0.07</td>
<td>&gt;16 μM²</td>
<td>&gt;10 μM²</td>
</tr>
<tr>
<td>A3H hap I*</td>
<td>0.60 ± 0.02</td>
<td>0.29 ± 0.01 μM (1.6)</td>
<td>0.35 ± 0.08</td>
<td>&gt;7 μM²</td>
<td>1.20 ± 0.16 μM</td>
</tr>
<tr>
<td>A3G</td>
<td>2.50 ± 0.15</td>
<td>0.09 ± 0.02 μM (1.9)</td>
<td>0.48 ± 0.10</td>
<td>1.30 ± 0.15 μM</td>
<td>0.16 ± 0.01 μM (2.5)</td>
</tr>
</tbody>
</table>

The Hill coefficient indicates that the enzyme binding curve best fit to a sigmoidal (cooperative) regression, rather than a rectangular hyperbola, by least squares analysis.

²For these experiments, a 118 nt ssDNA with two deamination motifs spaced 63 nt apart was used.

³The apparent K_d was estimated because the oligonucleotide could not be saturated with the enzyme.

6, depending on the substrate. The processivity factor indicates that A3B was at least 4-fold more likely to processively deaminate both 5′ATC motifs than to make a non-processive deamination of only one 5′ATC motif (Figure 2A-B). To determine if A3B could also move by 3D diffusion we tested deamination on a substrate where the 5′ATC motifs were separated by 63 nt. The A3B had a processivity factor of 4, similar to the processivity exhibited on the substrates with closer spaced motifs (Figure 2C). Due to

the distance between the motifs alone the data suggest that A3B can move 3-dimensionally by jumping or intersegmental transfer. However, to confirm that A3B could move 3-dimensionally, we annealed a double-stranded (ds) DNA between the two 5′ATC motifs. Since A3B binds dsDNA 13-fold less well than ssDNA (Supplementary Figure S3A), the dsDNA acts to block A3B sliding in between the two 5′ATC motifs. Thus, to deaminate both A3B would have to traverse the dsDNA by 3D diffusion. Consistent with A3B not
Figure 2. A3B is a processive enzyme that both slides and moves 3-dimensionally on ssDNA. Processivity of A3B was tested on ssDNA substrates that contained a fluorescein-labeled deoxythymidine (yellow star) between two 5′ATC deamination motifs separated by different distances. (A) Deamination of a 60 nt ssDNA substrate with deamination motifs spaced 5 nt apart. Single deaminations of the 5′C and 3′C are detected as the appearance of labeled 42- and 23-nt fragments, respectively; double deamination of both C residues on the same molecule results in a 5 nt labeled fragment. (B) Deamination of a 69 nt ssDNA substrate with deamination motifs spaced 14 nt apart. Single deaminations of the 5′C and 3′C are detected as the appearance of labeled 51- and 32-nt fragments, respectively; double deamination of both C residues on the same molecule results in a 14 nt labeled fragment. (C) Deamination of a 118 nt ssDNA substrate with deaminated cytosines spaced 63 nt apart. Single deaminations of the 5′C and 3′C are detected as the appearance of labeled 100- and 81-nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment. (D) Deamination of a 118 nt ssDNA substrate as in (C) but with a 30 nt complementary DNA annealed between the deamination motifs. The measurements of processivity factor (P.F.) and the standard deviation (S.D.) from three independent experiments are shown below the gel.

being able to slide across the dsDNA, we find that the processivity decreases 2-fold in the presence of the complementary dsDNA, but is not diminished completely (Figure 2D). The 2-fold decrease results from failed attempts to slide over the dsDNA, which results in enzyme dissociation (49). This supports that A3B can move over the dsDNA to complete processive deaminations by either jumping or intersegmental transfer. In contrast, the A3B CTD is not processive on any ssDNA substrate tested (Supplementary Figure S5).

Having characterized the basic biochemical features of A3B we wanted to make a comparison to the other deaminases implicated in cancer mutagenesis to determine if they shared specific biochemical abilities. A3A has been previously characterized by multiple labs and consistent with these past studies we report that A3A is not processive and binds ssDNA in the high micromolar range (∼9 μM) (42–44) (Figure 3A, Table 1, and Supplementary Figure S3C). A3H hap I is not amenable for biochemical studies due to it being thermodynamically unstable (16,64). However, it is known that this is due to a Gly at position 105 and mutagenesis of this position to form a G105R mutant stabilizes the protein (64). This A3H form is found in humans as haplotype VII (65). For the purposes of our study the A3H hap I G105R (haplotype VII) was used as an A3H hap I proxy since the two enzymes are otherwise identical in their amino acid sequences. We refer to it as A3H hap I*. In contrast to A3A, the A3H hap I* is processive and binds ssDNA in the nanomolar range, similar to other A3H haplotypes previously characterized and A3B (Figure 3B, Table 1, and Supplementary Figure S3D) (45).

Processivity is not required for deamination during transcription

Deamination of genomic DNA undergoing transcription has not been identified as a major mechanism by which these enzymes access ssDNA in the nucleus (29–33). We wanted to test whether this was due to inherent characteristics of the enzymes. To determine if A3B, A3A, and A3H hap I* could deaminate during active transcription, we used an in vitro transcription system driven by the phage
Figure 3. Deamination during transcription is mediated by size and not by processivity. (A and B) Processivity of (A) A3A and (B) A3H hap I* was tested on a 118 nt ssDNA substrate with deamination motifs (5′TTC, A3A; 5′CTC, A3H hap I*) spaced 63 nt apart. Single deaminations of the 5′C and 3′C are detected as the appearance of labeled 100- and 81-nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment. A processive enzyme has a processivity factor >1 (see Materials and Methods). (C) Deamination activity of A3B, A3A, A3H hap I*, A3G, and A3G F126A/W127A (FW) during T7 RNA polymerase transcription of a 66 nt dsDNA. The nontranscribed strand contained a single deamination motif for A3A (5′TTC), A3B (5′TTC), A3H hap I* (5′CTC), or A3G (5′CCC). Percent deamination was calculated from the integrated intensity of the product band resolved by PAGE (Supplementary Figure S6). (D) Deamination activity of A3B, A3A, A3H hap I*, and A3G on a 70 nt R-loop substrate with a fluorescein label at the 5′ end and a single deamination motif within the R-loop region. The motifs were 5′TTC (A3A, A3B, A3H hap I*) and 5′CCC (A3G). Each A3 enzyme (100 nM) was incubated with 100 nM of R-loop substrate for 5–20 min. Percent deamination was calculated from the integrated intensity of the product band resolved by PAGE (Supplementary Figure S7). (E-G) Size exclusion chromatography profiles of 10 μg (E) A3A, (F) A3H hap I* and (G) A3G from a 10 mL Superdex 200 column was used to calculate the oligomerization state of the enzyme from a standard calibration curve. An ‘M’ denotes a monomer fraction, a ‘D’ denotes a dimer fraction, and a ‘T’ indicates a tetramer fraction. (E) A3A resolved as a monomer (apparent molecular weight 23 kDa). (F) A3H hap I* formed a combination of tetramers (apparent molecular weight 92 kDa), dimers (apparent molecular weight 46 kDa), and monomers (apparent molecular weight 23 kDa). (G) A3G formed both dimers (apparent molecular weight 46 kDa) and monomers (apparent molecular weight 23 kDa). The chromatograms were constructed by analyzing the integrated gel-band intensities of each protein in each fraction after resolution by SDS-PAGE (Supplementary Figure S4).
T7 RNA polymerase. This is a minimal system where transcription requires only the RNA polymerase, dsDNA substrate and rNTPs. In order to quantify enzyme activity as deamination over time, we used an oligonucleotide dsDNA substrate, rather than a plasmid based T7 system. Each dsDNA substrate used the same transcribed strand, but had a unique non-transcribed strand with a single deamination motif on the non-transcribed strand specific for each A3. We let the reaction proceed in the presence of an A3 and T7 RNA polymerase and then the non-transcribed strand was sequenced using a poisoned primer assay to detect deaminations (50). Although not absolute, deamination of the non-transcribed strand would be consistent with what has been observed for APOBEC enzymes in yeast model systems (29,66,67), but for the purpose of this experiment we sought to quantify deamination rather than test the propensity for an A3 to deaminate a specific strand in the transcription bubble. The data show that A3B is unable to deaminate within the transcription bubble, but A3A and A3H hap I* are able (Figure 3C and Supplementary Figure S6). To further investigate why A3B was unable to deaminate during transcription, we used an R-loop substrate to determine if A3B could deaminate in a ‘static’ transcription bubble. A3B could deaminate this substrate, but the activity was 3-fold less than on ssDNA (Figure 3D, Table 1 and Supplementary Figure S7A). Even A3A, which could deaminate the ssDNA created during transcription, was 4-fold less active on the R-loop substrate compared to fully ssDNA (Figure 3D, Table 1 and Supplementary Figure S7B). The A3H hap I* had 1.7-fold less deamination activity on the R-loop substrate than ssDNA (Figure 3D, Table 1, and Supplementary Figure S7C). Altogether, A3B did not appear to be deficient in deamination of an in vitro R-loop in comparison to A3A and A3H hap I*, suggesting that A3B activity was not limited by the small segment of ssDNA available (~8–20 nt) (Figure 3C and D). Although A3B could bind a DNA/RNA hybrid, the apparent Kd value was similar to that for ssDNA suggesting that A3B was not being sequestered away from the ssDNA through preferential binding to a DNA/RNA hybrid (Table 1). Notably, we could not reach saturation of the R-loop substrate in steady state binding experiments with A3A or A3H hap I*, despite demonstrated deamination activity, suggesting that the enzymes bound the R-loop with short residence times (Figure 3C and D, Table 1, and Supplementary Figure S3C and D). This was also observed for A3A with the DNA/RNA hybrid (Table 1 and Supplementary Figure S3C).

To understand why A3B did not efficiently deaminate during the dynamic transcription assay in contrast to A3A and A3H hap I* we first conducted a comparative analysis. A3B is made up of four individual enzyme units (Figure 1B). In contrast, A3A is a monomer and A3H hap I* is primarily a dimer (Figure 3E and F). Thus, the ability of A3B to stabilize itself on ssDNA may be diminished when the transcription bubble is moving since it has to account for eight nucleic acid binding domains, whereas A3A or A3H hap I* have only one or two to four, respectively (Figures 1B and 3E and F). Although A3B bound ssDNA cooperatively, it did not bind the R-loop cooperatively suggesting that A3B bound the ssDNA during transcription at the oligomeric state observed in solution (Table 1 and Supplementary Figure S3A). Nevertheless, the tetramer of A3B may become more readily destabilized by a moving substrate than an enzyme with less ssDNA binding sites.

To test this line of reasoning we used another processive and double deaminase domain A3 that has been extensively studied with regards to processivity and oligomerization (48,54,68,69). A3G forms monomers and dimers in solution and further oligomerizes on ssDNA by binding cooperatively (Figure 3G, Table 1, and Supplementary Figure S3E) (48,70,71). Similar to A3B, A3G was unable to deaminate ssDNA undergoing transcription, but was able bind and deaminate an in vitro R-loop (Figure 3C and D, Table 1, and Supplementary Figure S6, and Supplementary Figure S7D). A3G also bound ssDNA with more affinity than the DNA/RNA hybrid (Table 1 and Supplementary Figure S3E). We then used a previously characterized monomeric form of A3G, an A3G F126A/W127A mutant (A3G FW) to determine if an A3G with less binding domains could deaminate during transcription (48). The A3G FW was able to deaminate during transcription at an equivalent level to A3H hap I* (Figure 3C and Supplementary Figure S6). The in vitro data support the conclusion that in a dynamic transcription system, the size of the enzyme is a determinant in activity and not substrate binding affinity or processivity. However, since R-loops or ssDNA generated during transcription have not been identified as a major source of mutations from whole genome sequencing data, there are other cellular factors not considered here that must protect the ssDNA at a genome level (29–33).

**Enzyme cycling is required for efficient deamination during DNA replication**

The deamination of the lagging strand of genomic DNA undergoing replication has been identified as a major source of A3 catalyzed cytosine deaminations (29–33). This is especially true in cancer cells with a high amount of replication stress where ssDNA can accumulate from incomplete lagging strand synthesis, stalled replication forks, or excessive replication fork firing (21,32). As a result, there is a view that the ssDNA is simply left available for A3 enzymes to catalyze deamination of cytosine (28,72,73). To test this experimentally we used another processive and double deaminase domain A3 that has been extensively studied with regards to processivity and oligomerization (48,54,68,69). A3G forms monomers and dimers in solution and further oligomerizes on ssDNA by binding cooperatively (Figure 3G, Table 1, and Supplementary Figure S3E) (48,70,71). Similar to A3B, A3G was unable to deaminate ssDNA undergoing transcription, but was able bind and deaminate an in vitro R-loop (Figure 3C and D, Table 1, and Supplementary Figure S6, and Supplementary Figure S7D). A3G also bound ssDNA with more affinity than the DNA/RNA hybrid (Table 1 and Supplementary Figure S3E). We then used a previously characterized monomeric form of A3G, an A3G F126A/W127A mutant (A3G FW) to determine if an A3G with less binding domains could deaminate during transcription (48). The A3G FW was able to deaminate during transcription at an equivalent level to A3H hap I* (Figure 3C and Supplementary Figure S6). The in vitro data support the conclusion that in a dynamic transcription system, the size of the enzyme is a determinant in activity and not substrate binding affinity or processivity. However, since R-loops or ssDNA generated during transcription have not been identified as a major source of mutations from whole genome sequencing data, there are other cellular factors not considered here that must protect the ssDNA at a genome level (29–33).
Figure 4. A3-mediated mutagenesis during DNA replication requires enzyme cycling. (A) A minimal phage replication system was used where the DNA polymerase of phage λ29 can initiate rolling circle replication and strand displacement synthesis in the presence of random hexamer primers (sketch). The bar graph depicts the number of mutations induced by each A3 per kb sequenced. (B) The ability of an A3 enzyme to cycle through different ssDNA substrates was measured by adding increasing amounts of unlabeled ssDNA relative to the labeled ssDNA (0x, 5x, or 50x). The labeled ssDNA had two deamination motifs separated by 30 nt for A3A (5′TTC), A3B (5′ATC), A3H hap I* (5′CTC), or A3G (5′CCC). (C–E) Intersegmental transfer ability of A3H hap I*, A3B, and A3G were determined by keeping an A3:ssDNA ratio constant, but increasing the total reaction components. The ssDNA substrate contained a fluorescein-labeled deoxythymidine (yellow star) between two deamination motifs for A3A (5′TTC), A3B (5′ATC), A3H hap I* (5′CTC), or A3G (5′CCC) separated by 63 nt. The measurements of processivity factor (P.F.), standard deviation (S.D.), and reaction rate (%) from three independent experiments are shown below the gel.

lower than A3B and suggested that the A3s able to induce mutations during cancer did share a similar biochemical characteristic that was not common to A3G (Figure 4A).

To identify the difference between A3G and the other A3s tested, we used a more defined in vitro system with oligonucleotide ssDNA substrates. We first focused on A3A and A3G since we have previously conducted a side-by-side characterization of these enzymes (42). The major differences between the two enzymes are that A3A is not processive and A3G is processive, A3A has a $K_d$ in the micromolar range and A3G has a $K_d$ in the nanomolar range, and A3A cycles through substrates more frequently than A3G. The processivity did not appear to be a differentiating factor, since A3B and A3H hap I* are processive and are similar to A3A in this assay (Figures 2, 3B and 4A). However, A3G has a long half-life on ssDNA, ranging from 3 to 5 min in contrast to A3A that cycles on and off ssDNA rapidly (42,48,70,71,75,76). This difference between A3A and A3G suggested that cycling on and off ssDNA is a determinant in the efficiency of deamination during DNA replication. To test this we interrogated the ability of the enzymes to cycle through different amounts of unlabeled ssDNA substrate.
to find and deaminate motifs on a labeled ssDNA substrate (Figure 4B). First the enzymes were incubated with 5× or 50× excess unlabeled DNA substrate or no unlabeled DNA substrate (0×), in comparison to the amount of labeled substrate that would be added to the reaction. After an incubation period, the labeled substrate was added to monitor deamination activity. Consistent with A3A cycling frequently, the 5× or 50× of unlabeled ssDNA did not diminish the deamination activity (Figure 4B and Supplementary Figure S8). Consistent with increased concentrations of ssDNA making molecular collisions more favorable, the A3A activity increased up to ~1.5-fold at 50× concentration of unlabeled substrate DNA, although this value was highly variable (Figure 4B). In contrast, A3G deamination activity was increasingly inhibited by the A3G getting ‘trapped’ in the pool of unlabeled ssDNA substrate and not cycling often enough to find and deaminate the labeled ssDNA substrate before the end of the reaction (Figure 4B and Supplementary Figure S8). In this cycling assay A3B was similar to A3A and only had a 20% decrease in relative deamination activity even in the presence of 50× unlabeled ssDNA (Figure 4B and Supplementary Figure S8). A3H hap I* maintained its activity at a 5× level of unlabeled ssDNA, but could not cycle often enough to deaminate the labeled ssDNA efficiently in the presence of 50× unlabeled ssDNA (Figure 4B and Supplementary Figure S8).

These data support the hypothesis that frequent enzyme cycling improves deamination activity during DNA replication. However, it seemed counter intuitive that both A3B and A3H that were processive enzymes, were cycling frequently, since this could preclude the ability to search effectively on each ssDNA. Further, these enzymes had steady state binding constants for ssDNA in the nanomolar range (Table 1), this appears to be a distinct feature required for efficient mutagenesis of ssDNA created during DNA replication.

Enzyme cycling enables A3 enzymes to compete with RPA for ssDNA

In cells, the ssDNA exposed during replication stress would be coated with RPA (80). It has been assumed that A3 enzymes cannot displace RPA and that they are limited to deaminating ssDNA at gaps in the RPA coating (73). However, RPA is displaced to enable other recombination proteins to access the ssDNA, such as Rad51 (81,82). To determine if RPA would influence A3 deamination activity we added RPA to an oligonucleotide ssDNA substrate at saturating concentrations and then tested A3 deamination activity and processivity. It has been shown experimentally that RPA can be competed off ssDNA by other proteins in excess through a mechanism known as facilitated dissociation, which can be caused by Rad 51 or excess RPA itself (81,82). Notably, the facilitated dissociation occurs when RPA is in excess and it is not displaced as readily when amounts are limited. Thus, we used an in vitro system where the amount of oligonucleotide ssDNA was constant, rather than the Φ29 replication system where the DNA concentration would be continually increasing and precluding achieving RPA saturation on ssDNA. For this experiment
with oligonucleotides, RPA was pre-incubated with ssDNA and saturation was confirmed by an electrophoretic mobility shift assay (EMSA), in which the ssDNA was completely bound by RPA (Supplementary Figure S9). An A3 enzyme was then added to determine if RPA could be displaced. Displacement of RPA was interpreted based on the A3 deamination activity and processivity in comparison to naked ssDNA. We found that each A3 enzyme tested was able to displace RPA from the ssDNA in order to catalyze deaminations on the substrate. However, both the specific activity and processivity of the enzymes was decreased in the presence of RPA. For the A3B, A3H hap I*, and A3G, the processivity decreased 2-fold (Figure 5A-C). Interestingly, this is similar to when dsDNA segment is present between the deamination motifs (Figure 2D and Refs (45,53)) indicating that RPA may be acting as a roadblock to the 1-dimensional sliding motion of the enzymes. The A3A remained nonprocessive, but active (Figure 5D). Notably, we determined that there was a correlation between the cycling ability of the A3 enzyme and the specific activity in the presence of RPA. A3A, which cycles the most efficiently, had no change in specific activity in the absence and presence of RPA (Figure 4B and Table 2). The presence of RPA decreased the specific activities of A3B and A3H hap I* 2- to 3-fold (Table 2). However, A3G, an enzyme that does not efficiently cycle between substrates, due to a long residence time on ssDNA, had a 10-fold decrease in specific activity in the presence of RPA (Table 2) (48,70,71). These data indicate that the ability to sample multiple substrates through enzyme cycling is important in the presence of the ssDNA binding protein RPA.

**DISCUSSION**

The deoxycytidine deaminase activity of A3B, A3A, and A3H hap I has been implicated in having a driving role in cancer and tumor evolution by providing the cells with a diverse pool of mutations (11,73). Beyond analysis of mutational signatures and mRNA levels, there has been a paucity of biochemical data characterizing how these enzymes act on substrates relevant to deamination of genomic DNA (42). Here we provide the first biochemical characterization of A3B and compare properties of A3B, A3A, and A3H hap I* on substrates relevant to catalyzing cytosine deaminations in genomic DNA. Specific to A3B, the data demonstrate that using an A3B CTD is not a suitable replacement for the fl A3B (Figure 1). In regards to how A3 enzymes access ssDNA in the genome during transcription or replication, we find that deamination during transcription is more selective than during replication (Figures 3C and 4A). A3B was unable to deaminate during transcription and by comparison to other A3 enzymes we demonstrated that this is due to its propensity to form tetramers. Although all enzymes tested could deaminate during DNA replication or on ssDNA bound by RPA, the amount of deamination correlated with the enzyme’s ability to cycle between ssDNA substrates (Figures 4B, 5, and Table 2). These data support a model in which the inherent properties of the A3 enzyme determine the amount of A3-induced mutagenesis, not solely the availability of ssDNA.

A3 enzymes have been most extensively studied for their role in HIV-1 restriction where their cytosine deamination activity can inactivate HIV-1 proviral DNA through lethal mutagenesis (3). In this scenario, the relevant A3 enzymes, A3D, A3F, A3G, and A3H (hap II, V, or VII), which are notably different than the ones that contribute to cancer mutagenesis, contend with a transient ssDNA substrate that is undergoing replication during the time when the enzymes must catalyze deaminations (3). Studies from our lab have shown that a primary determinant in the mutagenic efficiency of the enzyme is their ability to scan the ssDNA efficiently for the preferred deamination motif and to processively deaminate these motifs (49,53,59,83). Although both the HIV-1 replication and genomic DNA replication and transcription processes are dynamic, they did not require the same biochemical characteristics. Processivity was not detrimental to deamination during transcription, replication, or ssDNA bound by RPA, but it was also not required as exemplified by A3A (Figures 3A, C 4A and 5, and Table 2). This difference may be due to the HIV-1 reverse transcriptase replicating the viral DNA at a slower rate than the eukaryotic polymerases (84,85). During HIV-1 replication, an A3 enzyme can use processive ssDNA scanning to compete for ssDNA with the less processive reverse transcriptase and the rapidly cycling HIV-1 nucleic acid chaperone, nucleocapsid (84,86). However, in genomic DNA the RPA binds much more stably to ssDNA than nucleocapsid and the polymerases synthesize DNA faster and can use PCNA to increase processivity (80–82,85). As a result, the ability to more stoichiometrically interact with ssDNA by cycling to different substrates or regions of the same substrate could provide an advantage. Processively tracking along the ssDNA with the polymerase is not possible since A3 enzymes move on ssDNA without an energy source by facilitated diffusion (3). Rather, repetitive sampling of ssDNA regions in 3-dimensions by macroscopic cycling or intersegmental transfers may enable better access to the available ssDNA pool that would be constantly changing. There is evidence from A3A that this type of cycling can approximate the benefits of processivity (42). Nevertheless, this conclusion was unexpected considering the number of processive DNA repair and replication proteins (62). However, these other proteins that are normally functional in the genome usually bind and scan dsDNA, which is a key difference with A3 enzymes that scan ssDNA or, for A3H, scan ssDNA and DNA/RNA hybrids (3,45).

Previous work has shown that A3 enzymes primarily deaminate genomic DNA during replication and favor the lagging strand of genomic DNA during replication due to the greater abundance of ssDNA from discontinuous synthesis (29–33). However, in order to access this ssDNA, A3 enzymes would still have to displace RPA (80). In the APOBEC family there were two opposing studies on how these enzymes could deaminate ssDNA bound by RPA. For AID, a physical interaction with RPA occurred and promoted AID-catalyzed deaminations in immunoglobulin variable genes and switch regions (87,88). However, A3G was characterized to have decreased specific activity and processivity in the presence of RPA (89). The study with A3G agrees with our data (Figure 5C and Table 2) and we provide a mechanism for why A3G activity decreases in the
Figure 5. A3 enzymes can compete with RPA to deaminate ssDNA. The ability of A3 enzymes to deaminate cytosines on ssDNA in the presence of saturating amounts of RPA was examined on a fluorescently labeled (yellow star) 69 nt ssDNA with deamination motifs for A3A (5′TTC), A3B (5′ATC), A3H hap I* (5′CTC), or A3G (5′CCC) separated by 15 nt. To prepare the substrate, 100 nM of ssDNA was preincubated with 300 nM RPA (for a 3:1 ratio) and 50 nM of (A) A3B, (B) A3A, (C) A3H hap I*, or (D) A3G was added to initiate the reaction. Single deaminations of the 5′C & 3′C are detected as the appearance of labeled 51- and 32- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 14 nt labeled fragment. The measurements of processivity factor (P.F.) and standard deviation (S.D.) from three independent experiments are shown below the gel.

Table 2. Specific activity of A3 enzymes on ssDNA in the absence or presence of RPA.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Specific Activity, ssDNA (pmol/μg/min)</th>
<th>Specific Activity, ssDNA with RPA (pmol/μg/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3B</td>
<td>1.6 ± 0.7</td>
<td>0.9 ± 0.2</td>
</tr>
<tr>
<td>A3A</td>
<td>1.1 ± 0.1</td>
<td>1.3 ± 0.2</td>
</tr>
<tr>
<td>A3H hap I*</td>
<td>1.7 ± 0.2</td>
<td>0.7 ± 0.1</td>
</tr>
<tr>
<td>A3G</td>
<td>2.5 ± 0.2</td>
<td>0.3 ± 0.1</td>
</tr>
</tbody>
</table>

For these experiments, a 69 nt ssDNA with two deamination motifs spaced 14 nt apart was used.
ssDNA generated from transcription are limited (29). A3H hap I induced mutations in cells have not been analyzed in this manner. Altogether, these data indicate that the cellular conditions prevent excessive deamination during transcription. This may be because ssDNA generated during transcription is most accessible only during extremely high levels of transcription, RNA polymerase pausing, R-loop formation, or if specific targeting mechanisms exist as they do for related family member, AID (66,67,91,92). In support of this, our study confirms that deamination during transcription and transcription related structures such as R-loops is more selective than during replication (Figures 3C-D and 4). For example, a recent study in yeast found that although A3B is unable to deaminate during transcription of mRNA genes, it can deaminate during transcription of tRNA genes (93). The ssDNA generated during tRNA gene transcription is thought to be more prone to form R-loops, which consistent with our in vitro data, would be required for A3B to exhibit deaminase activity (Figure 3C-D) (94). However, the resulting uracils in tRNA regions are thought to be repaired faithfully and not contribute to cancer mutagenesis (93).

Although A3B is similar in oligomeric state and processivity to the other double deaminase domain enzymes A3G and A3F, it is uniquely able to cycle through ssDNA substrates in contrast to A3G and A3F that have long lifetimes on single ssDNA substrates (48,49) (Figure 4B, D, E). A3A is still unique among other A3 enzymes because despite not being processive, it has one of the highest specific activities and the highest genotoxicity when ectopically expressed in cell culture (14,15,42,95). Our studies demonstrate that this is due the ability of A3A to cycle through different substrates so that it is able to sample a larger amount of transient ssDNA during replication or transcription (Figure 4B). The A3H hap I* appears to have features of both A3B and A3A that enable it to also contribute to genomic mutagenesis.

Another aspect of A3 activity not directly addressed in our study is the active site accessibility and the molecular conformation of ssDNA substrate binding. A3A activity is not only high because of its ability to cycle through different substrates, but also because of its highly accessible catalytic pocket that is normally in an open conformation (37,96). This is in contrast to A3B CTD that has a closed catalytic pocket (34,37,96). While it has been proposed by others that the closed catalytic pocket state serves as protective measure to reduce off-target mutations by APOBEC family enzymes (96,97), these conclusions for A3B CTD may need to be reconsidered since the fl A3B is 40-fold more active than A3B CTD, if the NTD of A3B fl A3B activity is due to different catalytic pocket accessibility (Figure 1A). It remainstobedeterminedwhetherincreased affinity interaction with ssDNA presented here further our understanding of the dynamics of how A3 enzymes can so efficiently induce deaminations that are actually ‘off-target’ events and not part of their normal cellular functions.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

FUNDING
Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant [RGPIN-2016-04113 to L.C.]. Funding for open access charge: Natural Sciences and Engineering Research Council of Canada.

Conflict of interest statement. None declared.

REFERENCES


