Comparative analysis of the gene-inactivating potential of retroviral restriction factors APOBEC3F and APOBEC3G
Kasandra Bélinger and Marc-André Langlois

Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada

APOBEC3 (A3) proteins are host-encoded restriction factors that inhibit retrovirus infection by mutagenic deamination of cytosines in minus-strand DNA replication intermediates. APOBEC3F (A3F) and APOBEC3G (A3G) are two of the most potent A3 enzymes in humans with each having a different target DNA specificity. A3G prefers to deaminate cytosines preceded by a cytosine (5'-CC), whereas A3F preferentially targets cytosines preceded by a thymine (5'-TC). Here we performed a detailed comparative analysis of retrovirus-encoded gene sequences edited by A3F and A3G, with the aim of correlating the context and intensity of the mutations with their effects on gene function. Our results revealed that, when there are few (TGG) tryptophan codons in the sequence, both enzymes alter gene function with a similar efficiency when given equal opportunities to deaminate in their preferred target DNA context. In contrast, tryptophan-rich genes are efficiently inactivated in the presence of a low mutational burden, through termination codon generation by A3G but not A3F. Overall, our results clearly demonstrated that the target DNA specificity of an A3 enzyme along with the intensity of the mutational burden and the tryptophan content of the gene being targeted are the factors that have the most forceful influence on whether A3-induced mutations will favour either terminal inactivation or genetic diversification of a retrovirus.

INTRODUCTION

The human APOBEC3 (A3) family is composed of seven proteins (A3A, A3B, A3C, A3D, A3F, A3G and A3H) with DNA cytosine deaminase activity. Together they play an important role in preventing the replication and spread of a wide range of viruses, retroviruses and retroelements (reviewed by Desimmie et al., 2014; Harris & Dudley, 2015; Koito & Ikeda, 2013).

A3F and A3G are potent and well-characterized members of the A3 family that are mostly recognized for their roles in mutating and restricting the infection of human immunodeficiency virus type 1 (HIV-1). However, this virus has evolved effective strategies to dodge the effects of these host intrinsic proteins; one of them is through the expression of the viral infectivity factor (Vif) (Desimmie et al., 2014). The main role of Vif is to protect HIV-1 from the deleterious actions of A3G and A3F (and some other members of the A3 family) by linking these proteins to the E3 ubiquitin ligase complex, which then targets them for degradation in the proteasome (Kobayashi et al., 2005; Liu et al., 2004; Meleh et al., 2004; Yu et al., 2003). In the absence of Vif early in the infection or when Vif levels are low, A3 proteins are readily packaged into nascent viral particles that are released from virus-infected cells. Upon virus entry into new susceptible target cells, stowaway A3 proteins then deaminate cytosines (C) into uracils (U) in the minus-strand viral cDNA during reverse transcription (Desimmie et al., 2014; Harris & Dudley, 2015; Koito & Ikeda, 2013). This results in the accumulation of abundant G-to-A mutations, also called hypermutation, on the viral plus-strand that can lead to gene inactivation in progeny viruses. Some A3 proteins display important differences in their selection of the cytosines that they prefer to deaminate. The bases immediately upstream to the deaminated C (underlined) influence this selection. A3G has a strong preference to deaminate a C preceded by a C (5'-CC), while all other human A3 proteins, including A3F, prefer to deaminate C preceded by a thymine (5'-TC) (Bogerd et al., 2007; Chelico et al., 2006; Hultquist et al., 2011; Langlois et al., 2005; Love et al., 2012; Rausch et al., 2009). Considering that A3-induced deamination events occur predominantly on the minus-strand viral DNA, deamination in a 5'-CC context opposite to a tryptophan (Trp) codon (TGG) during reverse transcription produces a TAG termination codon, and also possibly TGA and TAA non-sense codons if the TGG codon is followed.
by a G (i.e. TGGG) (Ara et al., 2014; Yu et al., 2004; Zennou & Bieniasz, 2006). Equally, deamination in a 5′-TC context can also generate a TGA termination codon, but only when a Trp codon is immediately followed by an A (i.e. TGGA).

A3-induced mutations do not always lead to gene inactivation and can potentially be beneficial for a retrovirus given the right conditions. Low levels of non-deleterious G-to-A mutations are believed to help retroviruses evade and evade immune defences (Ara et al., 2014; Armitage et al., 2014; Kim et al., 2014; Monajemi et al., 2012, 2014). This phenomenon may be especially important soon after a new cell infection. At this time, Vif levels are increasing but have not yet reached a threshold capable of completely preventing the packaging of A3 proteins into HIV-1 virions (Streb, 2013). Importantly, sublethal mutations have been associated with the emergence of anti-retroviral drug resistance and reduced cytotoxic T lymphocyte recognition (Fourati et al., 2012a, b; Haché et al., 2006; Jern et al., 2009; Monajemi et al., 2014; Mulder et al., 2008; Neogi et al., 2013; Sadler et al., 2010; Simon et al., 2005). It is therefore of significant interest to better understand the conditions that lead A3-induced mutations towards either viral gene inactivation or towards sequence evolution that ultimately benefits the virus.

Tangentially, reporter cells such as CEM-GFP or TZM-bl, that express a reporter protein under the control of an integrated HIV-1 promoter, are tools commonly used to measure the impact of A3 deamination on HIV infection. These cells express a reporter gene, often luciferase, β-galactosidase or EGFP, upon the translation of a functional Tat and/or Rev protein contributed by an infecting HIV-1 provirus (Derdeyn et al., 2000; Gervaix et al., 1997; Platt et al., 1998). Similarly, indicator retroviruses that express an EGFP reporter protein once integrated in a target cell are also frequently used to assess the retroviral restriction potency of the various A3 proteins. In light of the fact that A3 proteins can display different target DNA specificities, a bias could be introduced in these reporter assays that is related to the composition of the specific gene sequences being mutated and their differential susceptibility to be deaminated by any one A3 protein. Furthermore, the mutation burden intensity required to alter or inactivate the function of these proteins has not been formally assessed experimentally.

To address some of these issues, we performed a detailed analysis of how A3-induced mutations and amino acid substitutions affect the fluorescence intensity of EGFP expressed in two types of reporter retroviruses: HIV-1 ΔVif and the Moloney murine leukaemia virus (M-MLV). We investigated the efficiency of A3F and A3G to reduce the fluorescence of either the EGFP reporter protein alone, or as an EGFP fusion protein with a viral envelope glycoprotein (Env). Our results showed that there is an incongruent relationship between mutation intensity and gene inactivation by A3F and A3G, which is dependent on the nature of the sequence being mutated. When target sequences have no or few Trp codons and nearly equal numbers of 5′-TC and 5′-CC deamination sites, A3F and A3G display similar abilities to mutate and diminish reporter gene fluorescence, with A3G being slightly more efficient at completely inactivating gene function. However, A3G potently inactivates gene function by high-frequency stop codon generation in target sequences containing numerous Trp codons. In contrast, A3F mainly inactivates through intense hypermutation, mostly in absence of stop codons. Additionally, our study reveals that mutations located outside the EGFP region of the Env–EGFP fusion protein can also alter fluorescence intensity. This indicates that A3 mutations could have an impact on various other parameters affecting protein function and gene expression such as mRNA and protein stability.

### RESULTS

A3F deamination contributes to sequence diversification and rarely creates termination codons

While EGFP has but a single Trp codon (TGG), the envelope glycoprotein gene of M-MLV (Env–EGFP) has 13 additional Trp codons in the N-terminal region of the protein (N-ter.), and five in its C terminus (Table 1). Of these, there are four Trp codons followed by adenine (A) (i.e. TGGA), which can result in a TGA termination codon when targeted by A3F in a 5′-TC context on the minus strand. The number of 5′-TC and 5′-CC deamination target sites were nearly identical in the sequences coding for EGFP and for the C terminus of Env, but there were 72 5′-CC A3G target sites in the N-terminal sequence of Env compared with 45 5′-TC target sites for A3F (Table 1).

To quantify the gene-inactivating potential of A3 proteins, we produced reporter viruses expressing either EGFP alone under the control of an internal Tat/Rev-independent promoter (HIV-1 ΔVif), or EGFP embedded in-frame within

<table>
<thead>
<tr>
<th>Sequence motifs</th>
<th>EGFP†</th>
<th>N-terminal Env†</th>
<th>C-terminal Env†</th>
<th>Env-EGFP (complete sequence)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>5′-TC*</td>
<td>52</td>
<td>45</td>
<td>70</td>
<td>167</td>
</tr>
<tr>
<td>5′-CC*</td>
<td>51</td>
<td>72</td>
<td>70</td>
<td>193</td>
</tr>
<tr>
<td>TGG (Trp) codons</td>
<td>1</td>
<td>13</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>Trp followed by a G (TGG.G)</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Trp followed by an A (TGG.A)</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

*Cytosines in bold are putative deamination target sites.
†Target motifs tabulated on the retroviral minus DNA strand.
Env–EGFP of M-MLV, in cells co-transfected with either A3F or A3G (Fig. 1). The M-MLV and HIV-1 ΔVif viruses produced were then used to infect target cells from which integrated proviral DNA was amplified by PCR. Flow cytometry analysis was used to monitor virus restriction by the A3 proteins. Both A3F and A3G potently restricted the viruses, with A3G causing near complete restriction (Fig. 1a, b). Human APOBEC2 (A2) was used as a negative control. A3 protein expression in the virus producer cells was similar in all cases (Fig. 1c, d).

To assess the effect of A3-induced mutations on reporter protein fluorescence, PCR amplicons containing the complete coding sequence for EGFP or Env–EGFP were digested and ligated into a cytomegalovirus (CMV) expression plasmid and then transformed into E. coli to establish a library of A3F and A3G mutated sequences for each virus (Fig. 2a). Assays were performed on 20 randomly selected mutated clones and one randomly selected unmutated clone from each experimental condition (Table 2). Amongst mutated clones of EGFP, A3F displayed an average mutation rate of 8.2 G-to-A mutations per kb and A3G, 12.4 mutations per kb. These results show that A3G has greater mutagenic potential than A3F when provided with equal numbers of optimal consensus sites and when expressed at similar levels in the virus producing cells (Fig. 1c, d). The rationale for using two different reporter viruses was to see if the profiles, positions and intensities of the mutations in EGFP differed, but we found no clear evidence of this (data not shown).

To visualize the distribution of putative mutations in the Env–EGFP gene, all 5′-TC and 5′-CC dinucleotides were first located and their positions were then mapped to the corresponding codon that would be affected by the deamination of the targeted cytosine (Fig. 2b). In total, there are 163 codons that can be affected by at least one deamination event occurring in a 5′-TC context, and 167 in a 5′-CC context; some codons are affected by mutations in both contexts (e.g. GAGGNN). The sequence of the adjacent codons was also therefore taken into consideration for these calculations. We next pooled the mutations observed in all EGFP and Env–EGFP clones and analysed them separately for A3F (Fig. 2c) and A3G (Fig. 2d). We then plotted every codon containing a mutation occurring in either a 5′-TC or 5′-CC context (indicated as Mutated Total). We also plotted the location of the codons that were not deaminated (indicated as Non-mutated). Analysis of the local DNA sequence context flanking these non-mutated codons did not reveal a common motif (data not shown).

Analysis of the clones derived from mutated M-MLV and HIV-1 ΔVif proviral sequences revealed that 144 different codons were targeted by A3F: 63% of mutations occurred in a 5′-TC context, 13% were in 5′-CC and the remaining 24% were in a 5′-RC context, where R is either A or G (Fig. 2c). Only four of these 40 clones contained at least one stop codon, indicated by red lines in Fig. 2(c) (Tables 3, S1 and S2, available in the online Supplementary Material). In contrast to A3F, A3G deaminated 187 different codons, 65% of which in a 5′-CC context, 19% in 5′-TC and 16% in a 5′-RC context (Fig. 2d).
The most striking difference between A3F and A3G is the large number of stop codons that were generated by A3G. A total of 152 stop codons in 40 clones were generated by A3G that converted 17 out of the possible 19 Trp codons into termination codons (Tables 3, S3 and S4). Of interest, only one of the 80 clones analysed (Env-A3G-20) contained a stop codon in the EGFP coding sequence (Fig. 2d and Table S8). In terms of generating non-synonymous (NS) amino acid substitutions in the EGFP gene, A3G was only slightly more efficient than A3F (45% compared with 32%) (Tables 4, S5 and S6). As for the full-length Env–EGFP gene sequence, A3F produced overall 6% more NS substitutions than A3G (71% compared with 65%) (Tables 4, S7 and S8).

**Fig. 2.** Codons in the gene coding for Env-EGFP that are affected by A3F and A3G deamination. (a) Schematic representation of the EGFP or Env–EGFP genes cloned into the pcDNA3.1 expression vector. The green triangle represents the start codon (ATG), and the red square represents the termination codon (TAG). (b) Graphical representation of all the codons in Env–EGFP affected by a deamination in a 5′-TC (top) or 5′-CC (bottom) dinucleotide context. The total number (n) of affected codons is represented to the right of the graph. (c) Positions of all the codons targeted by A3F deamination (top) and codons in a 5′-TC context that were not mutated (bottom). Compiled mutations for all A3F and Env-A3F clones are represented. Red bars indicate a deamination event causing a termination codon. The total number (n) of mutated and non-mutated codons is indicated to the right. The proportion (%) of codons mutated in a 5′-TC or 5′-CC dinucleotide context is also indicated. (d) Similar analysis as in (c) but for viruses produced in presence of A3G.
Disparity between reporter fluorescence intensity and mutation burden

We next wanted to correlate A3-induced mutation intensity with alterations in reporter gene fluorescence. The main question we wanted to address is whether there is a minimal mutational threshold that would reliably predict gene inactivation. Mutated EGFP and Env–EGFP clones were individually transfected into 293T cells in three independent experiments and analysed for EGFP expression by flow cytometry (Tables S1–S4). The relative mean fluorescence, the total number of G-to-A mutations and the number of NS substitutions in EGFP are graphically presented for EGFP in Fig. 3(a) and the N-ter. of Env–EGFP clones in Fig. 3(b). Clones containing at least one termination codon are indicated by a red hexagon with the total number of stop codons contained within. Our results with A3F-mutated HIV D env-derived EGFP clones show that total G-to-A mutations per sequence ranged from one to 21, with NS substitutions ranging from zero to six per sequence (Fig. 3a, Tables S1 and S5). A3G-mutated EGFP clones contained between two and 26 G-to-A mutations, and between one and ten NS substitutions (Fig. 3a, Table S2 and S6). The mutation burden could not reliably predict a reduction in fluorescence intensity as some heavily mutated sequences still exhibited high levels of fluorescence (Fig. 3a). However, no detectable fluorescence was observed in clones containing 11 or more G-to-A mutations caused by either A3F or A3G.

Interestingly, several EGFP clones displayed very low fluorescence when few mutations were present such as clones A3F-3 and A3F-5. Both these clones share a common substitution, E214K, but the defect in fluorescence appears to be partially rescued in presence of mutation S209N, as seen in clone A3F-15 that contains a dual mutation (Fig. 3a and Table S5). This specifically illustrates the epistatic-like effect that mutations can have on gene function. Conversely, several intensely mutated A3F clones remained strongly fluorescent (e.g. A3F-18 and A3F-19). This is in contrast to the more heavily mutated A3G-derived clones that all exhibit reduced fluorescence, but these generally also contained slightly higher numbers of NS substitutions than their equivalent A3F counterparts.

A3G inactivates reporter function by termination codon generation when Trp codons are abundant

M-MLV encoding the Env–EGFP fusion offers 13 additional Trp codons upstream of EGFP that can be converted into stop codons by deamination in a 5’-CC context (Table 1). Of these, four Trp codons were followed by an A (i.e. TGG A), and can therefore also be potentially turned into TGA stop codons by deamination in a 5’-TC context. All Env–EGFP clones mutated by A3G contained

Table 2. Comparative analysis of the mutations induced by A3F and A3G

<table>
<thead>
<tr>
<th>A3</th>
<th>Deaminated sequence</th>
<th>Clones analysed</th>
<th>Mutated clones</th>
<th>Base pairs analysed in mutated clones</th>
<th>Total number of G-to-A mutations</th>
<th>Mutation rate (mutations/kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3F</td>
<td>EGFP*</td>
<td>20</td>
<td>20</td>
<td>14400</td>
<td>118</td>
<td>8.2</td>
</tr>
<tr>
<td>A3G</td>
<td>EGFP</td>
<td>20</td>
<td>20</td>
<td>14400</td>
<td>178</td>
<td>12.4</td>
</tr>
<tr>
<td>A3F</td>
<td>Env–EGFP†</td>
<td>20</td>
<td>20</td>
<td>54060</td>
<td>293</td>
<td>5.4</td>
</tr>
<tr>
<td>A3G</td>
<td>Env–EGFP</td>
<td>20</td>
<td>20</td>
<td>54060</td>
<td>970</td>
<td>18.0</td>
</tr>
</tbody>
</table>

*The EGFP sequence is 720 bp in length.
†Env-EGFP is 2703 bp.

Table 3. Gene-inactivating potential of A3F and A3G

<table>
<thead>
<tr>
<th>A3</th>
<th>Deaminated sequence</th>
<th>Mutated clones</th>
<th>Mutated clones with reduced fluorescence*</th>
<th>Mean relative fluorescence†</th>
<th>Mutated clones with no fluorescence‡</th>
<th>Clones with stop codons</th>
<th>Total number of stop codons</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3F</td>
<td>EGFP</td>
<td>20</td>
<td>7 (35 %)</td>
<td>47</td>
<td>6 (30 %)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A3G</td>
<td>EGFP</td>
<td>20</td>
<td>6 (30 %)</td>
<td>45</td>
<td>10 (50 %)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A3F</td>
<td>Env–EGFP</td>
<td>20</td>
<td>8 (40 %)</td>
<td>55</td>
<td>11 (55 %)</td>
<td>4 (20 %)</td>
<td>6</td>
</tr>
<tr>
<td>A3G</td>
<td>Env–EGFP</td>
<td>20</td>
<td>0</td>
<td>–</td>
<td>20 (100 %)</td>
<td>20 (100 %)</td>
<td>152</td>
</tr>
</tbody>
</table>

*Clones with reduced fluorescence are defined as clones displaying less than 85 % of the mean fluorescence of the unmutated control but more than background.
†Mean relative fluorescence was calculated for mutated clones with reduced fluorescence compared with the control.
at least two stop codons that completely abrogated EGFP fluorescence (Fig. 3b and Table S4). A3F, on the other hand, inhibited mainly through NS substitutions that resulted in reduced reporter fluorescence, although four clones contained a stop codon at the same position in the N-terminal region of Env (Fig. 3b and Table S2).

DISCUSSION

It is now well documented that human A3F and A3G are potent antiretroviral restriction factors that inhibit HIV-1 infection when the Vif protein is absent or expressed at low levels. It is also clear that when A3 proteins are expressed at near-physiological levels the main mechanism for the restriction is their ability to lethally mutate retroviral DNA replication intermediates (Albin et al., 2014; Browne et al., 2009; Kobayashi et al., 2014; Miyagi et al., 2007). The mutator activities of A3F and A3G can therefore be involved in shaping part of the early repertoire of HIV-1 viruses produced during a natural infection in humans when Vif levels are low enough to allow A3 virion incorporation (Feng et al., 2014). In addition, A3 mutation intensity is strongly influenced by the position of a gene within a retroviral genome, as twin 3' to 5'-mutational gradients have been described for HIV-1 and M-MLV (Belanger et al., 2014; Suspène et al., 2006; Yu et al., 2004). Here we were interested in comparing the gene-

Table 4. Genetic diversification caused by A3F and A3G

| A3 Deaminated sequence | Mutated clones | Total codons deaminated | NS NS NS in EGFP NS total NS |
|------------------------|----------------|-------------------------|-----------------------------|-----------------------------|
| A3F EGFP               | 20             | 118                     | 38                          | 38                          | 32                          |
| A3G EGFP               | 20             | 184                     | 83                          | 83                          | 45                          |
| A3F Env–EGFP           | 20             | 283                     | 24                          | 201                         | 71                          |
| A3G Env–EGFP           | 20             | 870                     | 60                          | 567                         | 65                          |

NS, Non-synonymous substitutions.

Fig. 3. Fluorescence of mutated EGFP and Env–EGFP reporter proteins. (a) Relative mean fluorescence intensity (MFI) of the EGFP reporter protein mutated by A3F (left) and A3G (right). Relative MFI (green bars) and the total number of G-to-A mutations (black bars) are indicated for each clone. Light blue bars indicate the number of NS amino acid substitutions in the EGFP sequence of the clone. Red hexagons indicate the presence of one or more stop codons in a clone (the exact number is specified inside the hexagon); Asterisks (*) indicate clones with determined EGFP-inactivating mutations other than stop codons. (b) Relative MFI of the Env–EGFP reporter protein mutated by A3F (left) and A3G (right). Only mutations in N-terminal Env and EGFP are displayed (N-Ter.). Fluorescence results represent MFI ± SD normalized to a non-mutated plasmid control and were compiled from at least three independent transfections.
inactivating potential of A3F and A3G, especially in light of the fact that these two enzymes have different DNA target specificities and biochemical processes for deaminating cytidines in single-stranded DNA (Ara et al., 2014). We focused our interest on using the EGFP reporter protein as a model because it offers an equivalent distribution of 5’-TC and 5’-CC putative deamination targets. This thereby provided us with the ability to directly compare how DNA target specificity impacts gene inactivation and retroviral sequence evolution. This is an issue of particular importance as A3F, more than A3G, has been shown to promote HIV-1 diversification and evolution in vivo (Ara et al., 2014; Armitage et al., 2012, 2014; Kim et al., 2014; Sato et al., 2014; Wood et al., 2009). However, the extent to which this diversification correlates with alterations in gene function had not been assessed until now.

Biochemical analyses have previously demonstrated that A3G is a more potent cytosine deaminase than A3F, in that purified A3G is able to deaminate more cytidines in a single-stranded DNA substrate within a defined unit of time (Ara et al., 2014). These assays, however, were conducted using DNA oligos containing the preferred consensus target site for each enzyme. Here we revisited this issue using a retroviral restriction assay and analysing a sequence offering nearly identical 5’-TC and 5’-CC target sites. Our results also support the idea that A3G is a more potent deaminase by introducing approximately 34% more G-to-A mutations in the EGFP gene than A3F (178 mutations for A3G compared with 118 for A3F) (Table 2). Despite the fact that the ability of both deaminases to reduce EGFP fluorescence was remarkably similar, A3G generated 20% more sequences that did not encode a fluorescent protein (Fig. 3 and Table 3). This result is in line with our observation that A3G produced larger numbers and a greater proportion of NS substitutions in the gene coding for EGFP than A3F (Table 4).

Our analysis of mutations in the full-length Env–EGFP gene yielded quite contrasting results. We found that A3G was predominantly more effective in mutating and inactivating reporter fluorescence than A3F (Fig. 3 and Table 2). This was primarily due to a combination of factors, notably a larger number of overall 5’-CC dinucleotides than 5’-TC, and also, more importantly, a large number of Trp codons in the sequence. In fact, every A3G-mutated Env–EGFP clone analysed contained at least two stop codons upstream of the EGFP coding sequence (Fig. 3 and Table 3). Only four A3F clones contained a termination codon in the N terminus of the Env gene, all of which at position W48 (Fig. 2c and Table S2). The termination codon generated in this case is TGA, thereby indicating that it is the result of A3F deaminating in the less favourable 5’-CC context. This shows that although A3G is much more efficient at generating stop codons, A3F can also target the 5’-CC dinucleotide opposite to the TGG of the Trp codon, but at much lower frequencies. Our assays also revealed that A3F targets 5’-CC in 13% of the compiled mutations (Fig. 2c). Although A3F was less efficient than A3G at inactivating EGFP fluorescence, it did however introduce similar NS substitutions in the Env–EGFP gene (Table 4). With a lower gene inactivation potential and a high rate of NS substitution generation, our results support that A3F is much more effective than A3G at promoting viral sequence evolution. This is also consistent with previous findings that employed other analysis methods and systems (Ara et al., 2014; Armitage et al., 2012, 2014; Kim et al., 2014; Sato et al., 2014; Wood et al., 2009).

Through this study we also identified a number of gene-inactivating mutations in the EGFP gene caused by DNA deamination. The most potent mutations that completely abolished fluorescence are E133K and E223K; the latter has been found to define the charged state on the β-barrel chromophore structure of EGFP and is critical for fluorescence (Arpino et al., 2012). Both these substitutions were caused by the deamination of a glutamate codon (GAA) into a lysine (AAA) in a 5’-TC context by both A3F and A3G. Other residues such as E214K (5’-TC context) and M219I (5’-CC context) also significantly reduced fluorescence, but were influenced by the presence of other mutations. For example, the E214K mutation reduced fluorescence to 1.5% when present alone (clone A3F-5), but fluorescence was restored to 74% when accompanied with a S209N mutation (clone A3F-15). This emphasizes the epistatic-like effect of mutations when A3 deaminates proviral DNA. But it also underlines a possible shortcoming when using EGFP fluorescence intensity to measure infection by a reporter virus. Another interesting observation is that some mutations outside the EGFP coding sequence also influenced fluorescence. Clones Env-A3F-2, Env-A3F-4 and Env-A3F-7 do not contain NS substitutions in the EGFP sequence but display reduced fluorescence (Fig. 3b and Tables 5, S2 and S7). These examples suggest that A3 deamination events may have effects not only on protein function but also on various other parameters such as mRNA translation efficiency, and protein and mRNA stability. Determining the exact cause of reduced fluorescence for each affected clone is beyond the scope of this study.

TZM-bl and CEM-GFP indicator cell lines are laboratory workhorses frequently used to monitor HIV-1 infection and the restriction potential of various A3 proteins. Such indicator cells require Tat and or Rev protein expression from the infecting provirus for cell-encoded HIV-1 LTR-mediated reporter gene transcription and expression. Interestingly, all HIV-1 genes contain several Trp residues, however, early genes that code for Tat and Rev only contain one (Table S9). As such, these HIV reporter cells rely on the expression of these early genes, which may not be inactivated through stop codon generation if the intensity of hypermutation is low and especially if the investigated A3 proteins deaminate with a preferred 5’-TC consensus. This could result in viral restriction being largely underestimated. Because most human A3 proteins other than A3G target preferentially 5’-TC dinucleotides, it would be more advisable to measure inhibition of proviral DNA integration or p24 particle release from infected cells to compare the restriction efficiency of various A3 proteins.
Here we have functionally demonstrated that A3G is more efficient at mutating retroviral DNA than A3F when given equal numbers of preferred putative deamination targets. But more importantly, we have also shown that A3G will potently inactivate genes containing Trp codons through stop codon generation even when the mutation burden is very low, which was not the case for A3F. Therefore, A3F, like all other A3 proteins that prefer 5’-TC deamination targets, is more amenable to contribute to retroviral sequence evolution than restriction, especially when the mutation burden is low. The ability to diversify the genetic code of retroviruses without inactivating viral genes further strengthens the possible implications of some A3 proteins in immune evasion, but also in retroviral drug resistance. This may become a special concern for the new HIV-1 therapies comprised of highly neutralizing antibodies that bind to viral epitopes that could be putative A3 deamination targets.

### METHODS

**Cells.** HEK 293T human embryonic kidney epithelium (293T) and NIH mouse embryonic fibroblasts (NIH 3T3) were cultured in HyClone Dulbecco’s Modified Eagle Medium (DMEM)/high glucose medium supplemented with 10% fetal bovine serum, 100 U penicillin and 100 μg streptomycin ml⁻¹.

**Plasmids.** Flag-tagged A2, A3F and A3G expression plasmids were constructed previously (Langlois et al., 2005). The single-cycle HIV[p8.9] pseudovirus expressing EGFP and the replicative Moloney murine leukaemia virus (M-MLV) expressing EGFP as an Env–EGFP fusion have been previously described (Bélanger et al., 2013; Langlois et al., 2009). HIV[p8.9] (referred to as HIV-1 ΔVif in the text) expresses EGFP from an internal spleen focus-forming virus promoter (Langlois et al., 2005). M-MLV expresses EGFP from the viral LTR promoter as a nested fusion protein with Env (Becker et al., 2005). The single-cycle Env–EGFP reporter for three independent experiments

**Viral mutator assays, reporter gene cloning and sequencing.** Infections using HIV-1 ΔVif pseudoviruses and M-MLV produced in presence of A3F or A3G were carried out as previously described (Bélanger et al., 2013). A3F and A3G expression in virus producing cells were monitored by Western blot and were consistently equal in each transfection assay (Fig. 1). Viral titres were normalized by p24 or p30 ELISA. Target cells were infected with the same p24 or p30 amounts that yield an m.o.i. of 0.5 as when the viruses are produced in presence of the A2 negative control (Bélanger et al., 2013). The restriction profile of each A3 was monitored 24h after infection by measuring EGFP fluorescence in target cells by flow cytometry analysis.

For gene inactivation analysis, the gDNA of 293T cells infected with HIV-1 ΔVif or NIH 3T3 cells infected with M-MLV was extracted and purified. PCR was used to amplify EGFP (HIV-1 ΔVif) or Env–EGFP (M-MLV) from integrated proviral sequences. The following primers were used for the PCR; EGFP-FWD-NheI: 5’-GCTAGCTATCCACCGGTCCGCACCAT-3’, EGFP-REV-HindIII: 5’-AAGCTTTCGTCGACTCTAGAT-3’, ENV-EGFP-FWD-NheI: 5’-GCTAGCTATCCACCGGTCCGCACCAT-3’, ENV-EGFP-REV-HindIII: 5’-AAGCTTATGCGCGCGTGCTTGAAG-3’. PCR products were then digested and cloned into the pcDNA3.1 expression vector (Invitrogen) downstream of the CMV promoter using the NheI and HindIII restriction sites. In total, 42 EGFP clones derived from HIV-1 ΔVif (20 mutated by A3F and 20 by A3G; and one unmutated clone for each); and 42 Env–EGFP clones derived from M-MLV (20 mutated by A3F and 20 by A3G; and one unmutated clone for each) were randomly selected from a small library of clones that were sequenced. We ensured that each mutated clone selected displayed a unique mutational profile.

**EGFP reporter assay.** EGFP and Env–EGFP clones were co-transfected along with a transfection control plasmid into 293T cells in triplicate in three independent experiments. Cells were harvested by trypsinization (0.25%, w/v, Trypsin, 1 mM EDTA) 48h post-transfection, washed twice in PBS and analysed by flow cytometry on a Beckman Coulter CyAN ADP flow cytometer. Transfection efficiency for all clones was consistently between 75 and 90% as determined by the reporter control. Mean EGFP fluorescence was assessed on the gated transfected cell population delineated by the reporter control. Mutated clones with fewer than 5% of gated cells expressing mean EGFP fluorescence above background were registered as having a relative mean fluorescence of 0. Data are presented as the relative mean EGFP fluorescence of the mutated clones compared with the WT EGFP or Env–EGFP reporter for three independent experiments ± SD.

**ACKNOWLEDGEMENTS**

The authors thank Laura Goodwin for technical assistance, Maria Rosales Gerpe and Tyler Renner for helpful discussions and Kristin Kemmerich for comments on the manuscript. We are grateful to the NIH AIDS Research and Reference Reagent Program for reagents. K. B. holds an Ontario Graduate Scholarship and a scholarship from le Fonds de Recherche en Santé du Québec. M.-A. L. holds a Canada Research Chair in Molecular Virology and Intrinsic Immunity. This research was supported by the Canadian Institutes of Health Research (grant #89774) and an Early Researcher Award from the Ontario Ministry of Research and Innovation to M.-A. L.

### Table 5. Amino acid substitutions outside of the EGFP sequence that influence fluorescence

<table>
<thead>
<tr>
<th>Substitution type and position in Env–EGFP</th>
<th>N-terminal Env</th>
<th>C-terminal Env</th>
</tr>
</thead>
<tbody>
<tr>
<td>E119K</td>
<td>D211N</td>
<td>G722E</td>
</tr>
<tr>
<td>L755K</td>
<td>L814I</td>
<td>M858I</td>
</tr>
<tr>
<td>C865Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparison of the gene-inactivating potential of A3F and A3G

**EGFP reporter assay.** EGFP and Env–EGFP clones were co-transfected along with a transfection control plasmid into 293T cells in triplicate in three independent experiments. Cells were harvested by trypsinization (0.25%, w/v, Trypsin, 1 mM EDTA) 48h post-transfection, washed twice in PBS and analysed by flow cytometry on a Beckman Coulter CyAN ADP flow cytometer. Transfection efficiency for all clones was consistently between 75 and 90% as determined by the reporter control. Mean EGFP fluorescence was assessed on the gated transfected cell population delineated by the reporter control. Mutated clones with fewer than 5% of gated cells expressing mean EGFP fluorescence above background were registered as having a relative mean fluorescence of 0. Data are presented as the relative mean EGFP fluorescence of the mutated clones compared with the WT EGFP or Env–EGFP reporter for three independent experiments ± SD.
REFERENCES


hotspots in the human immunodeficiency virus genome favors reduced recognition by CD8+ T cells. *PloS One* 9, e93428.


