Variation of HIV-1 Mutation Spectra among Cell Types

Colleen M. Holtz, Louis M. Mansky
Institute for Molecular Virology, Department of Diagnostic and Biological Sciences, MinnCResT Program, School of Dentistry, and Department of Microbiology, Medical School, University of Minnesota, Minneapolis, Minnesota, USA

The high rates of mutation, recombination, and replication drive HIV-1 diversity. In this study, we investigated how cell type affects viral mutation rate and mutation spectra. In studying four different cell types, no differences in mutation rate were observed, but intriguingly cell type differences impacted HIV-1 mutation spectra. This is the first description of significant differences in HIV-1 mutation spectra observed in different cell types in the absence of changes in the viral mutation rate.

Human immunodeficiency virus type 1 (HIV-1) infection remains a serious threat to global public health, with over 34 million people infected worldwide (http://www.unaids.org). In the absence of an effective vaccine, antiretroviral drug therapy remains the primary means for preventing transmission and disease progression, as well as new infections (1). The error-prone nature of HIV-1 reverse transcriptase (RT) remains a key determinant in the generation of mutations during HIV-1 replication. HIV-1 RT plays a prominent role in the high genetic diversity and evolution of retroviruses, which is driven by high rates of viral replication, mutation, and recombination (2). High genetic diversity allows HIV-1 to evade the immune system and become resistant to drug therapy. This high mutation rate and the high rate of virus evolution impact virus transmission and disease progression, as well as promotes the emergence of antiviral drug resistance (2).

Cellular proteins can exploit the propensity of HIV-1 to rapidly mutate by enhancing viral mutagenesis to the point where the virus is unable to replicate with enough fidelity to remain infectious (3). The apolipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like 3 (APOBEC3) family of proteins, for instance, act as cytidine deaminases during reverse transcription, causing G-to-A hypermutation in Vif-deficient HIV-1 (4, 5). APOBEC3G (A3G) has been particularly well characterized for its ability to induce cytosine deamination during the minus-strand DNA synthesis step of reverse transcription, resulting in G-to-A mutations in the plus-strand viral DNA (6, 7). The editing activities of APOBEC3G result in G-to-A hypermutation (and viral mutagenesis), although sublethal mutagenesis has been demonstrated (8). APOBEC3D, APOBEC3F, and APOBEC3H have also been demonstrated to restrict and cause G-to-A hypermutation of Vif-deficient HIV-1 (9).

Previous studies have provided evidence that HIV-1 genetic variation is impacted by RNA polymerase II transcription errors as well as minus-strand and plus-strand mutations that arise during HIV-1 reverse transcription (10–12). Deoxynucleoside triphosphate (dNTP) levels have been shown to have direct effects on RT fidelity (13, 14). Macrophages have been shown to have low dNTP levels, and this decreases the efficiency of viral DNA synthesis and increases the likelihood of mutations occurring during HIV-1 reverse transcription (15). A previous study investigated the differences in the HIV-1 mutation rate and mutation spectra observed between HIV-1 replication in HeLa cells versus that of CEM-A cells (16). No significant differences were observed in the viral mutation rates between these two cells. While the number of mutants that were characterized was small and prohibited a statistical analysis of potential differences in the rates of mutation for specific mutation types, the general trends suggested that there was no significant difference in the rates of base pair substitution mutations, frameshift mutations, and deletion or deletion-with-insertion mutations (16).

In the present study, we sought to further investigate the role of how cell type affects the mutation rate of HIV-1 as well as the virus mutation spectra. For analysis of mutation rate and mutation spectra, CEM-GFP (green fluorescent protein-expressing T lymphoblast cell line; NIH AIDS Research and Reference Reagent Program via J. Corbeil), U373-MAGI-CXCR4, CEM cells (glioblastoma cell line; via M. Emerman through the NIH AIDS Research and Reference Reagent Program), 293T cells (human embryonic kidney cells; American Type Culture Collection [ATCC]), and SupT1 cells (T lymphoblast cell line; ATCC) were transduced with an HIV-1 vector, pHHG (8). Mutant detection was determined by detection of cells that had a mouse heat-stable antigen-negative (HSA−) GFP+ phenotype. The calculation of mutant frequency was determined by dividing the number of infected cells harboring a mutated provirus phenotype (i.e., HSA− GFP+) by the total number of cells infected (i.e., HSA− GFP+ and HSA+ GFP+).

Using this analysis, it was determined that there was no significant difference (P > 0.05, Student’s t test) in the frequencies of mutants recovered from these cell types in parallel analyses (Table 1). A similar lack of observed difference in mutation rates had previously been made in an analysis conducted comparing HeLa cells and CEM-A cells (16).

To determine if the HIV-1 mutation spectra were influenced by cell type, proviral HSA mutation target sequences were analyzed from the four cell lines after parallel virus infections. Interestingly, a significant difference was noted when comparing T-cell line to CXCR4 cell line (Table 1). A similar lack of observed difference in mutation rates had previously been made in an analysis conducted comparing HeLa cells and CEM-A cells (16).

Using this analysis, it was determined that there was no significant difference (P > 0.05, Student’s t test) in the frequencies of mutants recovered from these cell types in parallel analyses (Table 1). A similar lack of observed difference in mutation rates had previously been made in an analysis conducted comparing HeLa cells and CEM-A cells (16).

To determine if the HIV-1 mutation spectra were influenced by cell type, proviral HSA mutation target sequences were analyzed from the four cell lines after parallel virus infections. Interestingly, a significant difference was noted when comparing T-cell line to CXCR4 cell line (Table 1). A similar lack of observed difference in mutation rates had previously been made in an analysis conducted comparing HeLa cells and CEM-A cells (16).
higher degree of G-to-A mutations per sequence in CEM-GFP and GFP, U373-MAGI, and 293T cells, there was a trend toward a
mutations per sequence were not statistically different between CEM-
analyzed (Fig. 1A). Although the average numbers of G-to-A mu-
mutations from each of the cell lines was
recovered that contained multiple G-to-A mutations in the target
gene sequence.
recovered among the 4 cell types analyzed were observed (data not
shown). This indicates that the frequency of G-to-A transition
mutations was significantly influenced by the proviral sequences
recovered that contained multiple G-to-A mutations in the target
gene sequence.
Proviral sequences with multiple G-to-A mutations in the tar-
gen sequence were recovered from CEM-GFP, U373-MAGI, and
293T cells (Table 3). There were significant differences (Fish-
ner’s exact test) in the frequencies of sequences with multiple G-
to-A mutations recovered between CEM-GFP and 293T cells (P =
0.03), CEM-GFP and SupT1 cells (P = 0.0005), and U373-MAGI
and SupT1 cells (P = 0.006). The mutational load for G-to-A
mutations per proviral sequence from each of the cell lines was
analyzed (Fig. 1A). Although the average numbers of G-to-A
mutations per sequence were not statistically different between CEM-
GFP, U373-MAGI, and 293T cells, there was a trend toward a
higher degree of G-to-A mutations per sequence in CEM-GFP and
U373-MAGI cells relative to that of 293T cells. The locations of
G-to-A mutations analyzed indicated that certain G residues were
hot spots for G-to-A mutations (Fig. 1B). Of these mutations, 73%
of the G-to-A mutations occurred at GA dinucleotides, while 17%
ocurred at GG dinucleotides, 5% occurred at GT dinucleotides,
and 5% occurred at GC dinucleotides.
The observation that G-to-A mutations preferentially oc-
curred at GA dinucleotides suggested that these mutations may be
due to the expression of APOBEC3 proteins. In order to test
whether there was a correlation between the frequency of recov-
ering proviral sequences with multiple G-to-A mutations and that
of APOBEC3 gene expression, mRNA expression levels of the
APOBEC3 proteins were analyzed in each of the cell lines under
study. The study of mRNA expression levels was done as a surro-
gate for assessment of APOBEC3 protein levels, given that anti-
obodies that can readily differentiate the family of proteins are not
currently available. A3A, A3B, A3C, A3D/E, A3F, A3G, A3H, and
TATA-binding protein expression vectors were used as control
standards (21). Quantitative reverse transcriptase PCR (qRT-
PCR) analysis was done on each cell line for A3A, A3B, A3C,
A3D/E, A3F, A3G, and A3H (Fig. 1C). A3C mRNA expression
levels suggested a possible correlation with the prevalence of pro-
viral sequences harboring multiple G-to-A mutations from each
line (i.e., CEM-GFP > U373-MAGI > 293T > SupT1). However,
analysis using a Pearson correlation coefficient test did not
support this correlation (P = 0.052). The A3G mRNA expression
also appeared to have an expression pattern that appeared to cor-
relate to the proviral sequences with multiple G-to-A mutations
from each cell line. In particular, significant differences were ob-
served in A3C expression between CEM-GFP and SupT1 (P =
0.0015, Student’s t-test), CEM-GFP and 293T (P = 0.0213), U373-
MAGI and 293T (P = 0.0286), U373-MAGI and SupT1 (P <
0.001), and 293T and SupT1 (P < 0.001) cells. However, the ob-
served dinucleotide specificity (i.e., GA) was not consistent with
the dinucleotide specificity of A3G (i.e., GG). All of the other
APOBEC3s either lacked mRNA expression in one of the cell lines
where multiple G-to-A mutations were observed, or there was no
significant difference in mRNA levels for a particular A3 between
the 4 cell lines analyzed. It is interesting to note that while both
CEM-GFP and Sup T1 cells are T lymphoblast cell lines, these
findings indicate that distinct virus mutation spectra were ob-
served. Taken together, the origins of the multiple G-to-A muta-
tions in recovered proviral sequences could not be directly attribu-
ted to A3 gene expression in the 4 target cell types.
To date, the nature of how cell type influences the mutation

TABLE 2 Mutation spectra in the HSA mutation target gene of HIV-1
proviral sequencesa

<table>
<thead>
<tr>
<th>Mutation</th>
<th>CEM-GFPb</th>
<th>U373-MAGIb</th>
<th>293Tc</th>
<th>SupT1c</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-to-A</td>
<td>29</td>
<td>26</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>A-to-G</td>
<td>27</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>T-to-C</td>
<td>27</td>
<td>29</td>
<td>30</td>
<td>36</td>
</tr>
<tr>
<td>C-to-T</td>
<td>10</td>
<td>10</td>
<td>14</td>
<td>11</td>
</tr>
</tbody>
</table>

a Cells were transduced with an HIV-1 vector (HIG) pseudotyped with vesicular
stomatitis virus protein G (VSVG), and the HSA reporter gene sequence from the
proivirus in infected cells was PCR amplified and sequenced. The percentage of
mutations for each mutation type compared to the total mutations identified is
indicated.
b Total no. of sequences, 238; total no. of mutations, 365.
c Total no. of sequences, 225; total no. of mutations, 329.
d Total no. of sequences, 154; total no. of mutations, 199.
e Total no. of sequences, 173; total no. of mutations, 219.

TABLE 3 Proportion of HIV-1 proviral sequences recovered per
infected cell line analyzed possessing multiple G-to-A mutations in the
HSA mutation target genea

<table>
<thead>
<tr>
<th>Target cells</th>
<th>No. (%) of mutant sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEM-GFP</td>
<td>14/238 (5.9)</td>
</tr>
<tr>
<td>U373-MAGI</td>
<td>10/225 (4.4)</td>
</tr>
<tr>
<td>293T</td>
<td>2/154 (1.3)</td>
</tr>
<tr>
<td>SupT1</td>
<td>0/173 (0)</td>
</tr>
</tbody>
</table>

a Cells were transduced with an HIV-1 vector (HIG) pseudotyped with vesicular
stomatitis virus protein G (VSVG), and the HSA reporter gene sequence from the
proivirus in infected cells was PCR amplified and sequenced. The number of proviral
sequences having HSA reporter gene sequences with multiple G-to-A mutations from
each cell type was divided by the total number of proviral sequences identified with
mutations.
rate and mutation spectra of HIV-1 has not been extensively studied. The discovery of the APOBEC3 proteins provided clear evidence of the potential for the host cell to extensively edit and mutate the HIV-1 genome, which may lead to mutations that can shape HIV-1 evolution (8). In this study, we have investigated several cell lines that are commonly used to study HIV-1 replication in cell culture. Using an HIV-1 vector to study the mutation rate and mutation spectra, we observed that cell type did not influence the viral mutation rate, as had been observed in an analysis of HIV-1 mutation rates between HeLa and CEM-A cells (16). Importantly, we observed for the first time distinct differences in HIV-1 mutation spectra in parallel analyses of 4 different target cell types. In particular, there was a significant difference in the frequency of G-to-A transition mutations observed between CEM-GFP and 293T cells, U373-MAGI and 293T cells, and between CEM-GFP and SupT1 cells. A significant difference in T-to-C transition mutations was also observed between CEM-GFP and SupT1. Interestingly, analysis of the proviral sequences from CEM-GFP and U373-MAGI cells led to the characterization of sequences harboring multiple G-to-A mutations in the reporter gene sequence, ranging from 2 to 13 G-to-A mutations. Furthermore, there were significant differences in frequencies between CEM-GFP and 293T cells, CEM-GFP and SupT1 cells, and U373-MAGI and SupT1 cells. The proviral sequences with multiple G-to-A mutations occurred mainly at GA dinucleotides (i.e., 73%) and GG dinucleotides (i.e., 17%).

Analysis of APOBEC3 mRNA expression levels did not allow for a correlation between the expression level of any one particular APOBEC3 and the observed sequences with multiple G-to-A mutations. There have been previous reports implicating an APOBEC3 protein as having a target cell effect in generating G-to-A mutations (22,23). Koning et al. hypothesized that A3A was responsible for G-to-A editing of HIV-1 cDNA in macrophages (23), while Bourara et al. hypothesized that A3C could mutate HIV-1 viral DNA in the target cell (22). In both studies, the G-to-A mutations were likely sublethal and did not result in G-to-A hypermutation. It is possible that in these previous studies, G-to-A hypermutation did occur but was at a low frequency and was not detected.

Our studies here are distinct in that mutation spectrum differences involving both sequences containing multiple G-to-A mutations in the mutation target gene sequence (which is indicative of G-to-A hypermutation), and differences in the frequencies of other mutation types were observed. While the origins of these mutations were not determined, differences in cellular protein levels that could edit HIV-1 DNA or differences in the fidelity of HIV-1 reverse transcriptase and/or cellular RNA polymerase II are likely responsible for the changes in mutation spectra observed in this study. As indicated earlier, differences in nucleotide pool levels in various cell types can influence the likelihood of RT-mediated mutation (15). Further studies to determine the molecular basis for these differences in mutation spectra would be of particular interest, as well as studies that would extend these studies to human primary cells (e.g., primary T cells and macrophages). Additional studies to investigate the ability of APOBEC3 proteins are warranted and would be enhanced with antibody reagents that allow for specific detection of each APOBEC3 protein as well as specific small interfering RNAs (siRNAs) for mRNA depletion studies.

FIG 1 G-to-A mutational load and mutation location in HIV-1 proviruses with multiple G-to-A mutations. (A) G-to-A mutational load in HIV-1 proviruses from CEM-GFP, U373-MAGI, and 293T cells. Each provirus with a mutant HSA sequence containing with multiple G-to-A mutations is indicated by a black circle (CEM-GFP), black square (MAGI), or black triangle (293T). The average G-to-A mutational load and standard deviation are indicated. (B) Location of G-to-A mutations in recovered proviruses harboring multiple G-to-A mutations. The red, green, and blue circles above G residues indicate the locations of G-to-A mutations (one circle per mutation identified) in proviruses recovered from CEM-GFP, U373-MAGI, and 293T, respectively. The start and stop codons of the HSA gene are identified by black rectangular boxes. (C) Quantitative RT-PCR was performed to determine the relative levels of APOBEC3 mRNA expression among the cell lines under investigation (i.e., CEM-GFP, SupT1, U373-MAGI, and 293T). The asterisk indicates that the APOBEC3 level from 293T cells was set to 1, and all other values are relative to this measurement. The mRNA expression levels were normalized to TATA-binding protein (TBP) mRNA levels. For these analyses, data with a relative to this measurement. The mRNA expression levels were normalized to APOBEC3C level from 293T cells was set to 1, and all other values are (i.e., CEM-GFP, SupT1, U373-MAGI, and 293T). The asterisk indicates that the APOBEC3 level from 293T cells was set to 1, and all other values are relative to this measurement. The mRNA expression levels were normalized to APOBEC3C level from 293T cells was set to 1, and all other values are (i.e., CEM-GFP, SupT1, U373-MAGI, and 293T). The asterisk indicates that the APOBEC3 level from 293T cells was set to 1, and all other values are relative to this measurement. The mRNA expression levels were normalized to APOBEC3C level from 293T cells was set to 1, and all other values are
REFERENCES


