LA-UR- 08-7874

Approved for public release;
distribution is unlimited.

Title:	Human APOBEC3G Drives HIV-1 Evolution and the Development of Drug Resistance
Author(s):	Tanmoy Bhattacharya Eun-Young Kim Fransje A. Koning Michael H. Malim Steven M. Wolinsky
Intended for:	Science

Los Alamos
NATIONAL LABORATORY
EST.1943

Form 836 (7/06)

Human APOBEC3G drives HIV-1 evolution and the development of drug resistance

Eun-Young Kim[1], Tanmoy Bhattacharya[2], Fransje A. Koning[3], Michael H. Malim[3*], Steven M. Wolinsky[1*]

[1]Division of Infectious Diseases, The Feinberg School of Medicine, Northwestern University, Chicago, IL 60611 USA; [2]Los Alamos National Laboratory, Los Alamos, NM 87545, USA; [3]Department of Infectious Diseases, King's College London School of Medicine, London, UK.

[*]To whom correspondence should be addressed:

Steven M. Wolinsky, M.D., Division of Infectious Diseases, Northwestern University Feinberg School of Medicine, 645 North Michigan Avenue Suite 900, Chicago, Illinois 60611, e-mail: s-wolinsky@northwestern.edu, phone: (312) 695-5067

Michael H. Malim, Ph.D., Department of Infectious Diseases, 2nd Floor, Borough Wing, King's College London School of Medicine, Guy's Hospital, London Bridge, London SE1 9RT, E-mail: michael.malim@kcl.ac.uk, phone: +44 (0) 207 188 0149

Abstract

Human APOBEC3G (hA3G) is an innate virus restriction factor that induces deamination of specific cytidine residues in single-stranded human immunodeficiency virus type 1 (HIV-1) DNA. Whereas destructive hA3G editing leads to a profound loss of HIV-1 infectivity, more limited editing could be a source of adaptation and diversification. Here we show that the presence of hA3G in T-cells can drive the development of diversity in HIV-1 populations and that under selection pressure imposed by the nucleotide analog reverse transcriptase inhibitor 3TC ((-)2',3'-dideoxy-3'-thiacytidine), a single point mutation that confers 3TC resistance, methionine 184 to isoleucine (M184I), emerges rapidly and reaches fixation. These results provide strong evidence that mutation by hA3G is an important source of genetic variation on which natural selection acts to shape the structure of the viral population and drive the tempo of HIV-1 evolution.

The use of combinations of antiretroviral drugs has proven effective in delaying progression of human immunodeficiency virus type-1 (HIV-1) disease and prolonging survival; however, the benefits of therapy can be compromised by the appearance of mutations in the viral proteins targeted by the drugs that confer resistance (1, 2). The combination of rapid rates of viral replication, poor fidelity of the virus replicating enzymes (reverse transcriptase [RT] and RNA polymerase II) and recombinatory exchanges rapidly produces a swarm of viral genomes carrying a broad spectrum of mutations (3-10). Administration of a potent antiretroviral drug would modify the distribution of distinguishable variants by destabilizing the equilibrium and allowing the selection of genomes that encode a mutation that confers resistance (11). The speed of this process is contingent upon the magnitude of the selective advantage given by the mutation, the abundance of such a mutation within the viral population, and the concentration of drug at sites of viral replication (2).

The hA3G protein is a naturally expressed host enzyme that deaminates deoxycytidine to deoxyuridine in the (mostly) minus strand of newly synthesized HIV-1 DNA; these transition mutations correspond to deoxyguanosine-to-deoxyadenosine editing of viral plus strands (12-15). The potent antiviral phenotype of this enzyme is efficiently counteracted by the HIV-1 encoded Vif protein, which serves to induce the proteasomal degradation of hA3G (16-19).

While excessive deamination of viral DNA is profoundly deleterious to infectivity owing to the loss of genetic integrity, an important unanswered question is whether hA3G-mediated mutagenesis could benefit the virus by generating genomic diversity that would enable viral populations to adapt to changing environments (e.g., as imposed by drug treatment or an adaptive immune response).

High-levels of resistance to the nucleotide analog RT inhibitor 3TC is determined by a single point mutation, methione 184 to isoleucine (M184I), in the conserved Tyr-Met-Asp-Asp motif of the catalytic site of RT as the result of a G-to-A change (AUG-GAT to AUA-GAT) (20-24). The substitution of this single amino acid interferes with the proper positioning of 3TC within the catalytic site. When 3TC is used as a single antiretroviral agent, drug-resistant virus outcompetes wild-type virus in a few weeks (25-27), suggesting that the M184I mutation preexists in the viral population. Because hA3G preferentially mutates GpG dinucleotides to ApG (12, 15, 28, 29), and therefore has the potential to generate the M184I change, we tested the hypothesis that limited hA3G-mediated editing may enable viral diversification and accelerate adaptive change in response to strong positive selection, in this case the influence of 3TC.

To do this, CEM-SS cells, which do not express significant levels of the APOBEC3 family proteins, were stably transduced with a hA3G-expressing

retroviral vector or the corresponding parental vector. Both cultures were challenged with wild-type HIV-1NL4-3 and maintained continuously without 3TC or with escalating concentrations of 3TC. Cultures were maintained by replenishment with fresh cells transduced with the appropriate transgene (Table 1). The virus was able to replicate in presence of hA3G in these experiments owing to the presence of Vif (30). Supernatant virions were recovered at frequent intervals and viral sequences in the RT region of *pol* were amplified from individual virion RNAs by RT polymerase chain reaction (PCR) following limiting dilution, and then sequenced directly (31-35).

The hA3G protein increased the rate at which mutations in the HIV-1 genome appeared (Table 1). Indeed, the viral population diversity was more pronounced in CEM-SS cells in the presence of hA3G than the control culture in which HIV-1 remained genetically homogeneous. The maximum genetic diversity within a time point ranged from 0.038% at day 20 to 0.225% at day 44 with high diversity in the CEM-SS-hA3G population (range, 0% - 2.0%) and low diversity in the CEM-SS-control population (range, 0% - 0.5%). Even though most mutations seriously compromise some facet of viral function and in so doing reduce fitness (36), the overall rate of virus replication was not appreciably different as measured by culture supernatant levels of virus (Table 1). As expected, many of the accumulated mutations were G-to-A changes occurring at GpG sites (14-fold more relative to the control); and, importantly, the M184I mutation was clearly

identified even in the absence of 3TC (Fig. 1, hA3G column). Occasional hypermutated viral sequences were found, however, suggesting that full-length copies of a replication incompetent and competent genomic RNA could be packaged together (7). Thus, despite the high error rate of the HIV-1 replicating enzymes and the number of rounds of replication that occurred, the genetic complexity of the viral population was determined to great extent by hA3G.

Inspection of the sequences of viruses grown in the presence of 3TC revealed striking positive selection for the M184I change in the presence of hA3G; specifically, 20 of 24 sequences at days 37 and 44 carried this mutation in hA3G expressing cells, whereas no such sequences were found in the hA3G-null cells (Fig. 1). Other amino acid changes that facilitate 3TC resistance (e.g., K64R and M184V) did not emerge in either culture. Viral diversity in hA3G expressing cells became limited in the presence of 3TC relative to the drug-free cultures, reflecting purifying selection (0.225% verses 0.232%, respectively). Interestingly, there was selection of independent genomes as the M184I mutation was present in different variants in the viral population. Despite strong positive selection, amino acid substitutions that facilitate 3TC resistance were not detected in the viral population propagated in the absence of hA3G.

HIV-1 is able to overcome the burden of harmful mutations by its high rate of replication so that there is a reservoir of individual genetic variants within the

population with potentially beneficial (i.e., fitness-enhancing) mutations that facilitate adaptation to change. Because even small increases in the mutation rate can have profound fitness effects, it is postulated that viruses have mutation rates close to the tolerable error threshold (8). Deviation from the tolerable error threshold would affect the rate of evolutionary change: for instance, higher mutation rates lead to the viral genome accumulating deleterious mutations such that the viral population can be driven to extinction (8). Accordingly, there is much interest in the development of pharmacologic inhibitors of Vif function that would de-repress the natural antiviral activity of hA3G and increase the mutation frequency of HIV-1 above the tolerable error threshold (38). Based on our findings, however, the use of a Vif protein inhibitor should be approached with caution; the consequence of establishing an error rate below the tolerable error threshold in response to these new agents could be a HIV-1 population able to adapt to new challenges encountered during infection.

Our results provide direct experimental evidence that mutation by hA3G is a source of genetic variation on which natural selection acts to shape the structure of the viral population. These data support a model wherein hA3G contributes to the extent of diversity and the rate of evolution of HIV-1 by influencing the mutation rate that enable the virus to evolve and adapt to changing conditions encountered during infection (8-10). Thus, the high rates of evolutionary change that characterize HIV-1 infection can be influenced not only by the poor fidelity of

the virus replicating enzymes, but also by the host antiviral factor hA3G. Our study suggests that hA3G is an important determinant in explaining the mutational dynamic that underlies the rate of evolutionary change in HIV-1 and contributes to the pathogenesis of disease.

References and Notes

1. F. J. Palella *et al.*, *N. Engl. J. Med.* **338**, 853 (1998).

2. F. Clavel & A. J. Hance, *N. Engl. J. Med.* **350**, 1023 (2004).

3. M. Eigen, P. Schuster, *Naturwissenschaften* **64**, 541 (1977).

4. J. W. Drake, *Proc. Natl. Acad. Sci.* **90**, 4171 (1993).

5. M. C. Boerlijst, S. Bonoeffer, & M. A. Nowak, *Proc. R. Soc. Lon. B Biol. Sci.* **263**, 1577 (1996).

6. T. Rhodes, H. Wargo, & W. S. Hu, *J. Virol.* **77**, 11193 (2003).

7. A. Jung *et al.*, *Nature* **418**, 144 (2002).

8. E. Domingo & J. J. Holland, *Ann. Rev. Microbiol.* **51**, 151 (1997).

9. I. M. Rouzine, A. Rodrigo, & J. M. Coffin, *Microbiol. Mol. Biol. Rev.* **65**, 151 (2001).

10. A. Moya, E. C. Holmes, & F. Gonzalez-Candelas, *Nature Rev. Micro.* **2**, 279 (2004).

11. J. M. Coffin, *Science* **267**, 483 (1995).

12. R. S. Harris *et al.*, *Cell* **133**, 803 (2003).

13. B. Mangeat *et al.*, *Nature* **424**, 99 (2003).

14. H. Zhang *et al.*, *Nature* **424**, 94 (2003).

15. Q. Yu *et al.*, *Nature Struct. Mol. Biol.* **11**, 435 (2004).

16. K. Stopak *et al.*, *Mol. Cell* **12**, 591 (2003).

17. A. M. Sheehy, N. C. Gaddis, & M. H. Malim, *Nature Med.* **9**, 1404 (2003).

18. M. Marin *et al.*, *Nature Med.* **9**, 1389 (2003).

19. X. Yu *et al.*, *Science* **302**, 1056 (2003).

20. C. A. Boucher *et al.*, *Antimicrob. Agents Chemother.* **37**, 2231 (1993).

21. L. C. F. Mulder, A. Harari, & V. Simon, *Proc. Natl. Acad. Sci.* **105**, 5501 (2008).

22. T. L. Keefer *et al.*, *J. Virol.* **79**, 1975 (2005).

23. W. Keulen *et al.*, *J. Virol.* **82**, 6536 (2008).

24. S. A. Knoepfel *et al.*, *J. Virol.* **65**, 3973 (1991).

25. C. A. Boucher *et al.*, *Antimicrob. Agents Chemo.* **37**, 2231 (1993).

26. S. G. Sarafianos *et al.*, *Proc. Natl. Acad. Sci.* **96**, 10027 (1999).

27. R. Schuurman *et al.*, *J. Infec. Dis.* **171**, 1411 (1995).

28. R. Suspene *et al.*, *Nucleic Acids Res.* **34**, 4677 (2006).

29. K. N. Bishop *et al.*, *Current Biology* **14**, 1392 (2004).

30. A. M. Sheehy *et al.*, *Nature* **418**, 646 (2002).

31. S. Palmer *et al.*, *J. Clin. Microbiol.* **43**, 406 (2005).

32. D. Shriner *et al.*, Genetics **167**, 1573 (2004).

33. A. Meyerhans, J. P. Vartanian, & S. Wain-Hobson, *Nucleic Acids Res.* **18**, 1687 (1990).

34. B. Keele *et al.*, *Proc. Natl. Acad. Sci.* **105**, 7552 (2008).

35. The RT region of the pol gene (1.3-kb) was amplified by nested PCR using High Fidelity Platinum Taq Polymerase (Invitrogen). The probes were designed against HIV-1 N-L43 and avoided the hA3G target motif. RT primer position and sequence were: external sense primer: 5'-CAG AGC AGA CCA GAG CCA AC-3' (2140→2156) and external antisense primer: 5'-CTG CTA TTA AGT CTT TTG ATG GGT C-3' (3508←3528 reverse complement), internal sense primer: 5'-CAG CCC CAC CAG AAG AGA GC-3' (2159→2175), and internal antisense primer: 5'-TCT AGC TCT GCT TCT TCT GTT AGT GGT A -3' (3427←3451 reverse complement). Single genome sequencing was performed as described (31). Viral sequences are available by GenBank accession numbers xxx – xxx.

36. S. Duffy, L. A. Shackelton, & E. C. Holmes, *Nature Rev. Genetics* **11**, 435 (2004).

37. R. Nathans *et al.*, *Nature Biotechnol.* **26**, 1187 (2008).

38. We thank Bette Korber for her technical assistance. This work was supported by grants from the National Institutes of Health, National Institutes of Allergy and Infectious Diseases (S. M. W.), and the U.K. Medical Research Council (M. H. M.). F. A. K. is a Fellow of the European Molecular Biology Organization. The authors declare no competing financial interests.

Tables and Figure Legends

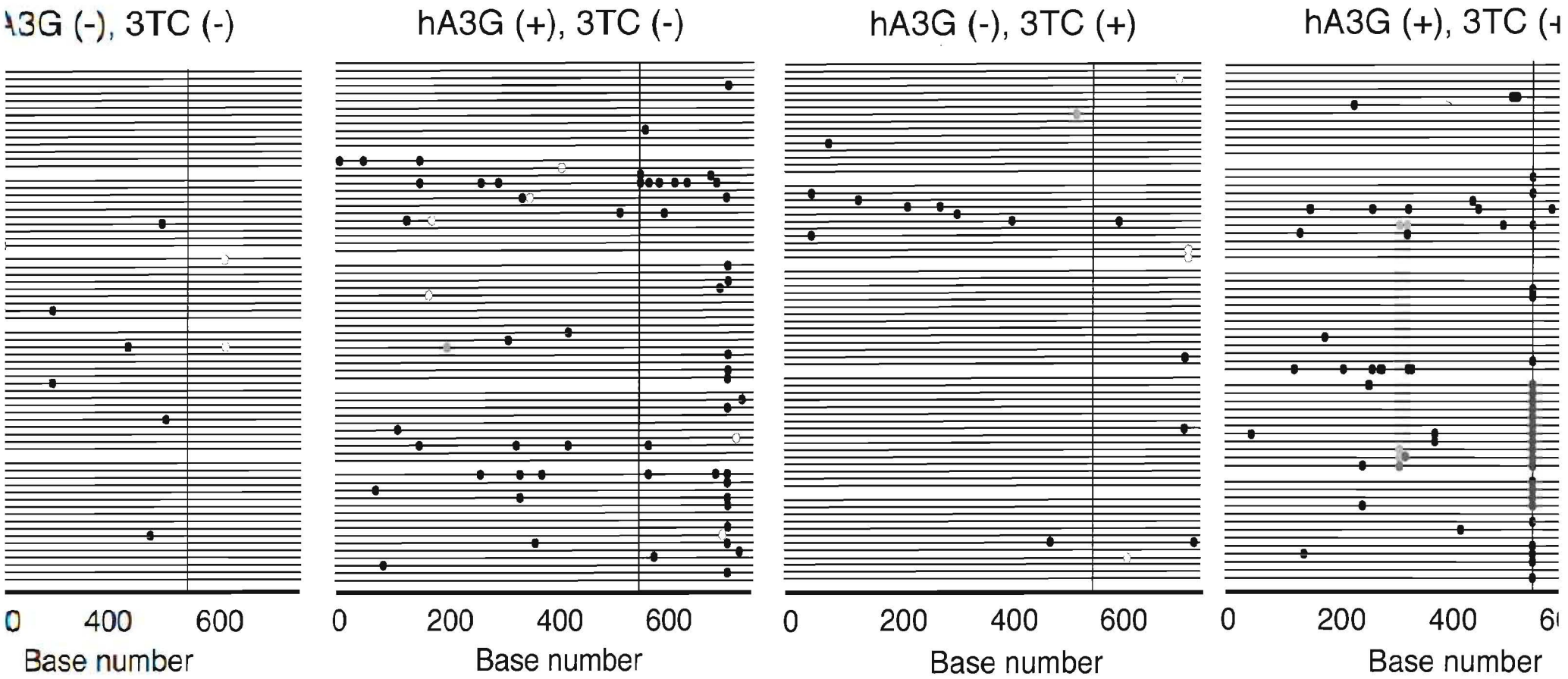
Table 1. RNA levels and genomic diversity for CEM-SS-hA3G and CEM-SS-control populations

Cultures were performed in 24-well plates in triplicate and maintained by replenishment with fresh CEM-SS cells transfected with the appropriate plasmid. Upon observing HIV-1-induced syncytium formation, the concentration of 3TC in the cell culture medium was increased 2-fold. The concentrations of 3TC in the absence of (hA3G (-)) were 0.24 μM (days 20 and 27), 0.48 μM (days 30 and 37), and 0.96 μM (day 44). The concentrations of 3TC in the presence of hA3G (hA3G (+)) were 0.24 μM (day 20), 0.48 μM (days 27, 30, and 37), and 0.96 μM (day 44). HIV-1 cDNAs were derived by single genome amplification of cell culture virion-associated RNA followed by direct sequencing of the PCR-product DNA to obviate *Taq* polymerase-induced nucleotide misincorporation and recombination and template resampling (31-34).

Figure 1. Highlighter analysis of diversity in the reverse transcriptase region of the *pol* gene

Highlighter tracings show diversity of the population following infection of CEM-SS-control and CEM-SS-hA3G cells by a single virus in the presence and absence of 3TC (+ and -, respectively). The median number of virion RNA RT sequences per time point was 13 (range, 9-17). Virions were recovered from the cell supernatant at frequent intervals. Viral sequences in the RT region of pol

were derived by single genome amplification (31). Days 20, 27, 30, 37 and 44 are shown. In the presence of hA3G (+), G-to-A changes that conformed to the predicted motif were distributed in multiple sequences. Hypermutated sequences conformed to model prediction. The G-to-A change that confers resistance to 3TC is located at nucleotide position 552 (vertical line). All viral sequences were aligned to HIV-1 wild-type HIV-1NL4-3 (top line).


A3G (-), 3TC (-)
hA3G (+), 3TC (-)
hA3G (-), 3TC (+)
hA3G (+), 3TC (+
0
400
600
0
200
400
600
0
200
400
600
0
200
400
6
Base number
Base number
Base number
Base number

	Day 20	Day 27	Day 30	Day 37	Day 44
hA3G (-) 3TC (-)					
vRNA (copies/ml)	4.59×10^8	1.69×10^8	1.66×10^8	1.18×10^7	3.49×10^6
Total number of mutations	4 / 9,750	3 / 7,500	5 / 6,750	7 / 12,750	2 / 12,750
Mutations per genome	0.29	0.27	0.50	0.39	0.11
hA3G (+) 3TC (-)					
vRNA (copies/ml)	8.79×10^8	9.12×10^8	5.75×10^8	1.32×10^8	4.95×10^8
Total number of mutations	2 / 8,250	21 / 9750	9 / 12,000	8 / 7,500	20 / 11,250
Mutations per genome	0.17	1.50	0.53	0.73	1.25
hA3G (-) 3TC (+)					
vRNA (copies/ml)	5.72×10^8	5.50×10^8	1.17×10^8	2.08×10^7	6.58×10^6
Total number of mutations	3 / 11,250	10 / 8,250	1 / 10,500	1 /12,000	3 / 9,000
Mutations per genome	0.19	0.83	0.07	0.06	0.23
hA3G (+) 3TC (+)					
vRNA (copies/ml)	8.09×10^7	7.92×10^8	4.90×10^7	1.06×10^8	2.96×10^8
Total number of mutations	12 / 8,250	21 / 9,000	14 / 9,750	21 / 8,250	12 / 8,250
Mutations per genome	1.00	1.62	1.00	1.75	0.86