Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts

Bastien Mangeat*†, Priscilla Turelli*†, Gersende Caron‡, Marc Friedli*§, Luc Perrin‡ & Didier Trono*§

** Department of Genetics and Microbiology, ‡ Department of Medicine, and § 'Frontiers in Genetics' Research Program, University of Geneva, 1211 Geneva 4, Switzerland*
† These authors contributed equally to this work

Viral replication usually requires that innate intracellular lines of defence be overcome, a task usually accomplished by specialized viral gene products. The virion infectivity factor (Vif) protein of human immunodeficiency virus (HIV) is required during the late stages of viral production to counter the antiviral activity of APOBEC3G (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G; also known as CEM15), a protein expressed notably in human T lymphocytes[1–4]. When produced in the presence of APOBEC3G, *vif*-defective virus is non-infectious. APOBEC3G is closely related to APOBEC1, the central component of an RNA-editing complex that deaminates a cytosine residue in *apoB* messenger RNA[5–7]. APOBEC family members also have potent DNA mutator activity through dC deamination[8]; however, whether the editing potential of APOBEC3G has any relevance to HIV inhibition is unknown. Here, we demonstrate that it does, as APOBEC3G exerts its antiviral effect during reverse transcription to trigger G-to-A hypermutation in the nascent retroviral DNA. We also find that APOBEC3G can act on a broad range of retroviruses in addition to HIV, suggesting that hypermutation by editing is a general innate defence mechanism against this important group of pathogens.

As an initial step towards investigating the mechanism of APOBEC3G antiviral action, we asked whether the putative catalytic site of this protein is important for inhibiting *vif*-defective (*Δvif*) HIV-1. For this, we used single-round assays to measure the infectivity of wild-type and *Δvif* particles produced by transient transfection of 293T cells, in the presence or absence of this cellular factor (Fig. 1). APOBEC3G, which localized to the cytoplasm of the transfected cells (Fig. 1a), induced, as previously observed[4], a marked decrease in the infectivity of *Δvif* HIV-1 (Fig. 1b). We then tested a series of APOBEC3G point mutants, concentrating on residues found in the zinc-coordinating catalytic site of APOBEC family members[5,9]. Mutating critical glutamate and cysteine residues (E67Q, C100S, E259Q and C291S)—previously shown to be essential for APOBEC1 function—abrogated the inhibitory potential of APOBEC3G, whereas changing these amino acids at several other positions (E85Q, C221S, E323Q) was without consequence (Fig. 1c).

These results strongly suggested that APOBEC3G inhibits retroviral infectivity by means of editing. However, the target of this process remained undefined. We thus examined the breadth of APOBEC3G antiviral activity (Fig. 2a). Using HIV-derived lentiviral vector particles produced from a multiply-deleted packaging construct lacking all the HIV-1 accessory genes (*vif*, *vpr*, *vpu* and *nef*) and pseudotyped with the vesicular stomatitis virus G (VSV-G) envelope protein[10], we found that APOBEC3G induced a Vif-responsive inhibition of transduction. APOBEC3G action and its blockade by Vif are thus independent of the route of viral entry and from these other viral gene products. Because editing enzymes often act in a sequence-specific manner, we asked whether APOBEC3G could also inhibit HIV-derived vectors produced from two 'codon-optimized' packaging constructs. The *gag* and *pol* genes of these constructs contain close to 1,000 silent nucleotide substitutions compared with the original HIV-derived lentivector system, but they produce virions with the same protein composition[11]. In spite of these sequence differences, APOBEC3G was fully active on these viruses. Furthermore, it also inhibited the infectivity of retroviral particles only distantly related to HIV-1, such as vectors derived from simian immunodeficiency virus (SIV)[12], equine infectious anaemia virus (EIAV)[13] and murine leukaemia virus (MLV). In all

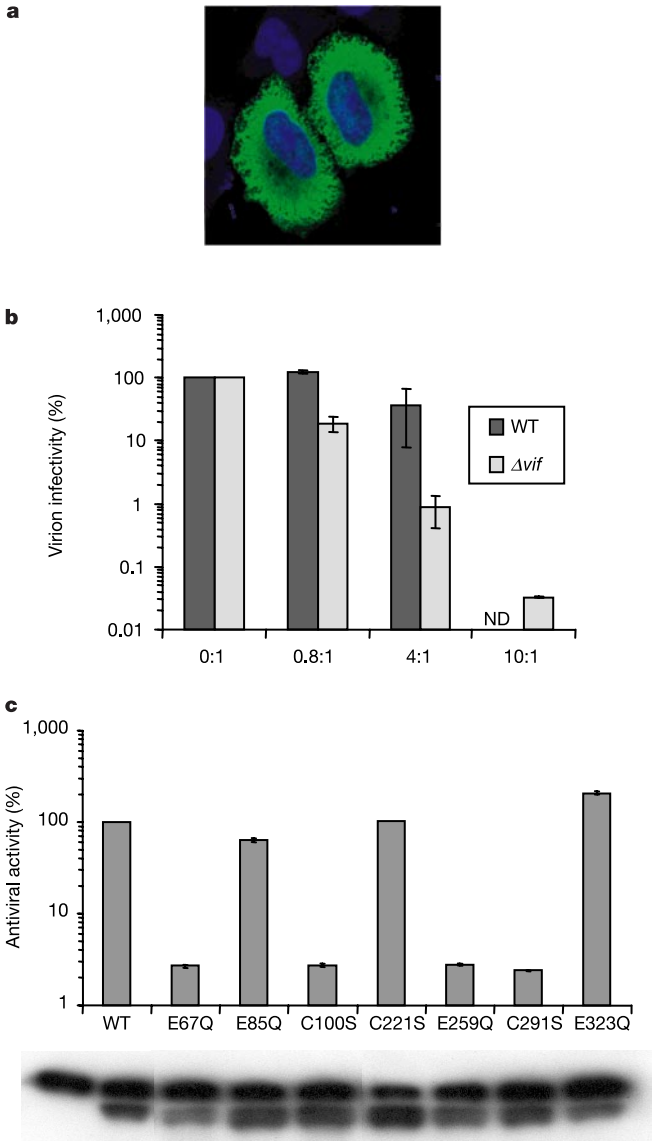


Figure 1 The putative catalytic site of APOBEC3G is important for HIV inhibition. **a**, Subcellular localization of APOBEC3G in transfected HeLa cells. The HA tag of the protein served for its detection by indirect immunofluorescence. 293T cells used for viral production gave the same image. **b**, Dose-dependent inhibition of *Δvif* HIV-1 infectivity by APOBEC3G, determined by single-round titration in $CD4^+$ P4 HeLa cells. The molar ratios of APOBEC3G compared with proviral DNA plasmids used for transfection are indicated on the *x* axis. Titres obtained with each virus produced in the absence of APOBEC3G were given the arbitrary value of 100%. ND, not determined; WT, wild type. **c**, Testing of indicated APOBEC3G point mutants, performed as in **b** using a 5:1 molar ratio of APOBEC3G/*Δvif* proviral DNA plasmids. Below is a western blot analysis of extracts from virus-producing 293T cells, using antibodies against tubulin (top band) and the HA tag of APOBEC3G (bottom band). The lane on the far left corresponds to untransfected 293T cells.

of these cases, inhibition was prevented by HIV-1 Vif, even for vectors derived from EIAV and MLV, two viruses devoid of the *vif* gene. Vif can thus act independently of other HIV-specific factors. Notably, APOBEC3G was detected in MLV-derived particles at levels roughly comparable to those found in HIV-derived virions (Fig. 2b).

The broad spectrum of APOBEC3G action suggested at least two possibilities. First, it might modify the sequence coding for a cellular cofactor essential for the generation of infectious retroviruses. Second, it might act on some viral genetic material, but in that case not in a sequence-specific manner considering the low level of identity between the genomes of HIV-1, SIV, EIAV and MLV. Although either hypothesis seemed plausible, the second one received support from a re-examination of the phenotypic consequences of APOBEC3G action.

We first noted that the infectivity defect of *Δvif* HIV-1 produced by proviral DNA transfection of H9 T lymphoid cells (which naturally express APOBEC3G) was only 50-fold at most, whereas this virus was completely unable to spread in this cell type, as if the magnitude of its replication defect increased with additional rounds of infection (data not shown). To probe this issue further, we infected T lymphoid SupT1 cells with VSV-G-pseudotyped wild-type and *Δvif* HIV-1 released from control or APOBEC3G-expressing 293T cells. SupT1 cells do not express APOBEC3G and are thus

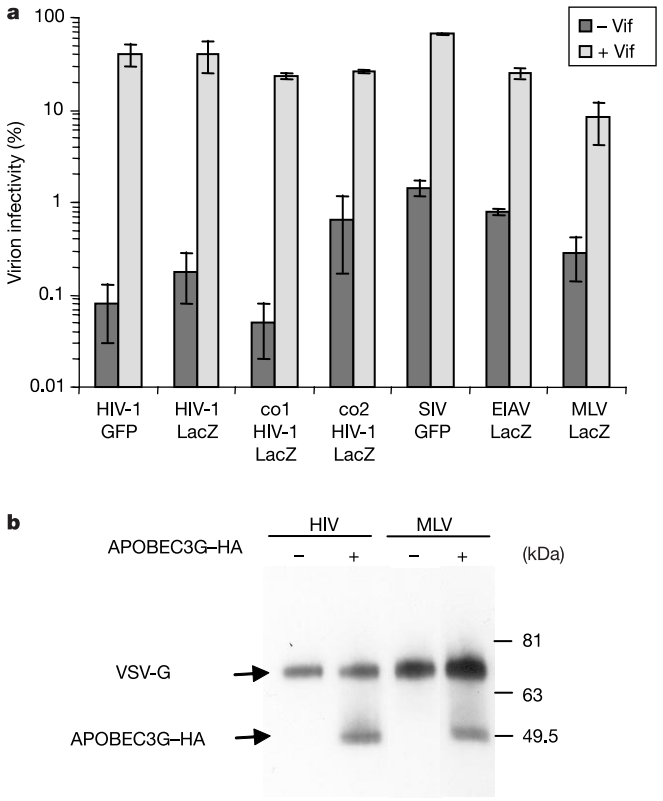
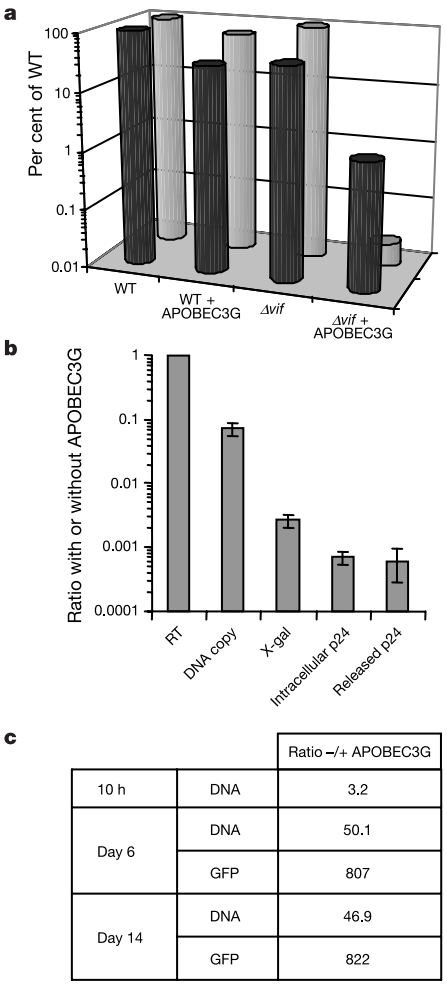


Figure 2 APOBEC3G is a broadly active antiretroviral. **a**, Single-round measurement of APOBEC3G-mediated inhibition of transduction of HeLa cells with particles derived from the following retroviral vector systems: multiply-deleted HIV-1-derived lentiviral vectors (HIV-1–GFP and HIV-1–LacZ); two 'codon-optimized' derivatives (co1 HIV-1–LacZ and co2 HIV-1–LacZ); non-human lentiviruses SIV (SIV–GFP) and EIAV (EIAV–LacZ); oncoretrovirus MLV (MLV–LacZ). Transduction efficiencies were scored by FACS (for GFP-expressing vectors) or 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) staining (for LacZ-expressing vectors). Transduction efficiency of vector produced in the absence of APOBEC3G was given the arbitrary value of 100% for each vector. Where indicated, HIV-1 Vif was expressed in vector producer cells. **b**, APOBEC3G is incorporated into MLV particles. VSV-G-pseudotyped HIV-1 and MLV particles were produced from transiently transfected 293T cells in the presence or absence of HA-tagged APOBEC3G. Normalized amounts (as assessed by reverse transcription activity) of Optiprep gradient-purified virions were analysed by western blotting, using antibodies against HA and VSV-G. Of note, the normalization is approximate, as it relies on measuring the activity of two distinct reverse transcriptases.



c

		Ratio –/+ APOBEC3G
10 h	DNA	3.2
Day 6	DNA	50.1
	GFP	807
Day 14	DNA	46.9
	GFP	822

Figure 3 Decreased viral DNA yields do not fully account for APOBEC3G-mediated inhibition. **a**, SupT1 human T lymphoid cells were infected with VSV-G-pseudotyped wild-type or *Δvif* HIV-1, produced in the presence or absence of APOBEC3G from transiently transfected 293T cells. After overnight incubation, the input virus was removed by extensive washing, and 60 h later the viral DNA content of the lymphoid cells was determined by real-time PCR, and their infectious virus output measured by titration on $CD4^+$ HeLa-derived P4 cells. The absence of significant amounts of the original 293T-produced VSV-G-pseudotyped virus in these supernatants was verified by parallel titration on $CD4^-$ HeLa cells (not shown). Black bars, viral DNA content of SupT1 cells/input virus (reverse transcription activity); grey bars, infectious output/viral DNA content of SupT1 cells. Numbers obtained for wild-type virus produced in the absence of APOBEC3G were given the arbitrary score of 100%. Results are representative of a duplicate experiment. **b**. $CD4^+$, HIV-1 LTR-β-gal-containing HeLa-derived P4 cells were infected as indicated above with *Δvif* HIV-1 exposed or not to APOBEC3G during production. After 24 h, target cells were extensively washed and 50 μM AZT was added. Sixty hours later, these cells were analysed for viral DNA load (by real-time PCR), Tat-induced β-gal expression (by X-gal staining), intracellular p24 content (by FACS) and virion release (by p24 ELISA on supernatants). Values were normalized for input virus (as determined by reverse transcription activity). **c**, HeLa cells were infected with a GFP-expressing, HIV-1-derived lentiviral vector produced without Vif in the presence or absence of APOBEC3G. Viral DNA content and rates of GFP expression of the target cells were evaluated at indicated times by real-time PCR and FACS analysis, respectively. Results are expressed as the ratio of values obtained from virus produced in the absence versus the presence of APOBEC3G.

fully permissive for *Δvif*[1,2]. Three days after infection, we measured the proviral DNA load and the infectious virus output of the lymphoid cells. This revealed that the *Δvif* virus initially exposed to APOBEC3G not only generated decreased levels of viral DNA in the SupT1 cells but that, for a given number of proviral copies, cells exposed to this virus yielded markedly decreased amounts of infectious virions (Fig. 3a). We similarly infected $CD4^+$ HeLa cells containing a LacZ reporter under control of an integrated HIV promoter[14]. HeLa cells are also APOBEC3G-negative cells. We then scored infection through a series of parameters reflecting sequential steps of the viral life cycle: viral DNA load (reverse transcription); β-galactosidase expression (integration-dependent production of a functional viral Tat transactivator); p24 capsid production in the infected cells (expression of the late viral genes, synthesis and proteolytic processing of the Gag/Pol precursor); and p24 release in their supernatant (virion assembly and budding) (Fig. 3b). The *Δvif* virus initially produced in the presence of APOBEC3G exhibited a gradient of increasing defectiveness as one moved towards parameters reflecting later steps of viral replication. This suggested that its replication potential had been irreversibly altered by its one-time exposure to the antiviral factor, and that its phenotype resulted from the cumulative effect of multiple defects.

To investigate this issue further, we examined the viral DNA load and the transgene expression profile of HeLa cells exposed to a replication-defective, green fluorescent protein (GFP)-expressing, HIV-1-derived lentiviral vector made from a *vif*-less packaging construct (Fig. 3c). At 10 h after infection, levels of reverse transcripts induced by this vector were decreased threefold if it had been produced in the presence of APOBEC3G. At six days, this difference had increased to 50-fold. Viral DNA levels in the target cell population then remained constant, as assessed by a third measurement performed at 14 days after infection. Because these cells were rapidly dividing, the viral DNA detected from the sixth day on was stably integrated. The increasing viral DNA defect noted at 6 and 14 days compared with 10 h after infection was consistent with a destabilizing effect of APOBEC3G on the HIV-1 reverse transcription complex, which not only compromized the synthesis of the viral DNA but triggered its degradation before it could integrate, as previously described for *Δvif* HIV-1 produced from restrictive cells[1,3]. An additional and crucial piece of information was obtained by comparing the integrated viral DNA content with the apparent rate of transduction of the target cells. The 50-fold lower proviral DNA load of cells exposed to the vector produced in the presence of APOBEC3G was indeed markedly outmatched by their 800-fold

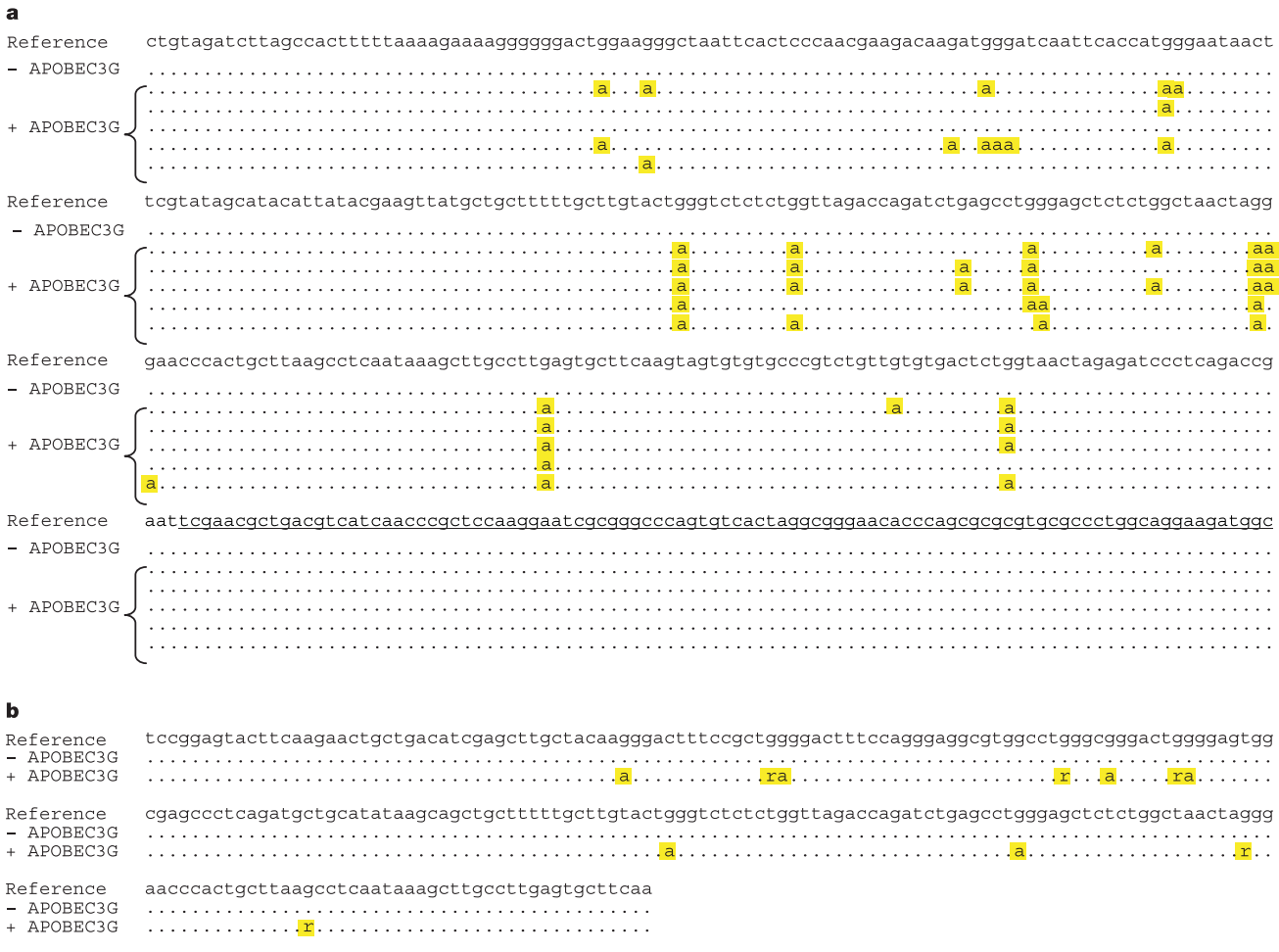


Figure 4 APOBEC3G triggers G-to-A hypermutation in the retroviral DNA. **a**, Sequence of representative individual clones obtained by PCR-mediated cloning of viral DNA fragments amplified at 6 days after infection from HeLa cells exposed to a replication-defective, HIV-1-derived lentiviral vector produced in the presence or absence of APOBEC3G. Underlined nucleotides are derived from the plasmid backbone used to clone these fragments. The starting pWPTS vector plasmid is taken as reference. **b**, Direct sequencing of representative PCR fragments amplified at 10 h after infection from $CD4^+$ HeLa cells exposed to similarly produced *Δvif* HIV-1 virions. R indicates a mixture of G and A. In both cases, the depicted sequences were recovered with primers mapping in HIV-1 Nef and U5. Similar types of results were obtained by sequencing other regions of the viral genomes, for instance the HIV-1 RT gene and the GFP transgene of integrated HIV and MLV vectors. Altogether, more than 10,000 bases of viral DNA were sequenced, and the results depicted are highly representative of the overall picture.

lower rate of GFP positivity, compared with cells infected with the control vector. Therefore, only one in every 16 transduced cells appeared capable of expressing the vector transgene. A similar type of result was obtained with MLV-derived vectors (data not shown).

This accumulation of non-functional proviruses prompted us to sequence viral DNAs in the target cells. This revealed that *vif*-less vector produced in the presence of APOBEC3G yielded DNA with a very high number of G-to-A substitutions (Fig. 4a). These were noted with equal frequency in DNA isolated at 10 h, 6 days and 14 days after infection (data not shown). Altogether, approximately 13% of G residues were mutated to A in the regions examined. The likelihood of this nucleotide change was higher if the original sequence contained stretches of two or more G residues (4% compared with 32% chance of G-to-A mutation for single versus multiple G residues). Importantly, no other type of nucleotide substitution was found in more than 10,000 sequenced bases. A similarly high rate of G-to-A mutation was detected in proviral DNA derived from *Δvif* HIV-1 virions (Fig. 4b) and from MLV retroviral vector particles (data not shown) produced from APOBEC3G-expressing cells. Correlating its ability to counter the inhibitory effect of APOBEC3G on HIV-1 replication, Vif expression in virus producer cells prevented the accumulation of G-to-A mutations (data not shown).

These results are consistent with a model in which the previously demonstrated cytidine deaminase activity of APOBEC3G leads to the replacement of cytosine by uracil in the nascent retroviral DNA. Indeed, if the viral RNA were the target of this modification, C-to-T rather than G-to-A changes would be observed in the viral DNA. Confirming this point, partial sequencing of the genomic RNA of APOBEC3G-exposed *Δvif* HIV-1 virions did not reveal any mutation (data not shown). The increased frequency of G-to-A changes in poly-G stretches indicates that cytosine residues are targeted by APOBEC3G in the context of the DNA, and not as free nucleotides. It has already been noted that APOBEC3G favours certain hotspots as substrates for cytidine deamination[8].

The replacement of cytosine by uracil in the minus strand DNA during reverse transcription leads to G-to-A transitions in the plus strand. In contrast, a C-to-U change in the plus strand would probably be corrected using the minus strand as a template. Therefore, we cannot exclude the possibility that APOBEC3G affects both DNA strands. However, it is interesting to note that the related enzyme activation-induced cytidine deaminase (AID) can deaminate cytidines on single-stranded but not double-stranded DNA[15].

The incorporation of uracil into minus strand DNA was recently shown to interfere with initiation of plus strand synthesis during lentiviral reverse transcription[16]. This may contribute to the decreased amounts of viral DNA found in *Δvif*-infected cells. Although our results do not directly explain why APOBEC3G also leads to the degradation of synthesized yet unintegrated reverse transcripts, this latter phenomenon may be one consequence of the handling of uridine-rich DNA by repair pathways. In this respect, HIV-1 packages the uracil DNA glycosylase hUNG-2, a host-derived enzyme that removes uracil bases from DNA as part of the base excision repair pathway. Although it may correct some of the G-to-A changes induced by APOBEC3G, it remains that HIV-1 proviruses established under the influence of the antiviral are highly mutated and as a consequence are largely non-functional.

The relative lack of sequence specificity of APOBEC3G-mediated editing explains the broad spectrum of activity of this antiviral factor, which can both inhibit a wide range of retroviruses and compromise expression from very different proviruses. The APOBEC3G-sensitive MLV and EIAV do not encode a *vif* gene. Perhaps a *vif*-like activity buried in some other gene product allows them to counter the orthologue of APOBEC3G probably present in their cognate host. Alternatively, they may preferentially replicate in APOBEC3G-negative cells, as this gene exhibits a narrowly restricted expression pattern, at least in humans[4].

G-to-A hypermutation was previously observed in HIV and other lentiviruses[17–21], and to a lesser extent in other retroviruses such as human T-cell leukaemia virus[22], spleen necrosis virus[23], a close relative of MLV, and in the pararetrovirus hepatitis B virus[24]. Whether APOBEC3G-mediated editing has a role in these cases as well will need to be explored. □

Methods

Constructs, virus productions, infections and titrations

Wild-type and *vif*-defective HIV-1 proviral clones, as well as the HIV-derived lentiviral vector system, have been described previously[1,10]. For the latter, packaging and envelope functions were provided by plasmids pCMV-ΔR8.91 and pMD.G, respectively, whereas the transfer vector was the GFP-containing pWPTS plasmid (see http://www.tronolab.unige.ch for details). The codon-optimized HIV vector packaging systems pHDMHgpm2 and pSYNGP[11] were obtained from R. Mulligan and K. Mitrophanous, respectively. The SIV (obtained from F.-L. Cosset) and EIAV vector systems were as described[12,13]. The plasmid expressing a haemagglutinin (HA)-tagged form of APOBEC3G[4] was a gift from M. Malim. APOBEC3G mutants were constructed with the Quickchange Mutagenesis kit (Stratagene). HIV-1 and retroviral vector particles were produced by transient transfection of 293T cells with Fugene (Roche) (see http://www.tronolab.unige.ch for details). We performed infections and titrations as previously described[25].

Protein analyses

HA-tagged APOBEC3G indirect immunofluorescence was performed as previously described[25], using the monoclonal HA.11 antibody (BAbCO). For HA- or VSV-G-specific western blot analyses, the mouse monoclonal 3F10-peroxidase-conjugated antibody (Roche) and a rabbit polyclonal antibody (a gift from M. Matsuda) were used, respectively. Virions were isolated by ultracentrifugation through a sucrose cushion followed by purification through an Optiprep (Axis-Shield) gradient[26], to minimize contamination by nonspecific microvesicles. Intracellular p24 fluorescence-activated cell sorting (FACS) analysis was performed with the Cytofix/Cytoperm kit (BD Biosciences).

Real-time PCR measurement of viral DNA load

Cells were infected with DNase-treated viruses or vectors and DNA was extracted with DNeasy kit (Qiagen). Taqman quantitative polymerase chain reaction (PCR) (Applied Biosystem) of virus sequences was performed with late reverse transcripts primers and probe[27]. Vector sequences were detected with GFP primers (GFP sense, 5′-CTGCTGCCCGACAACCAC-3′; GFP antisense, 5′-ACCATGTGATCGCGCTTCTC-3′; GFP probe, FAM-CCAGTCCGCCCTGAGCAAAGACC-TAMRA). The specific amplification of neo-synthesized reverse transcripts was controlled by AZT (zidovudine) treatment of a subset of target cells.

Sequencing

DNA was obtained as for Taqman PCR. Limited dilutions were either submitted to amplification (high fidelity Pfu Turbo polymerase; Stratagene) with primers specific for 5′ long terminal repeat (LTR) (PB.nef-1 sense, 5′-AGGCAGCTGTAGATCTTAGCCACTT-3′; PB.U5-1 antisense, 5′-GGTCTGAGGGATCTCTAGTTAC-3′) and directly sequenced[28], or submitted to amplification with the same primers flanked by 5′ *Bam*H1 and 3′ *Eco*R1 sites, cloned and sequenced. RNA from DNase-treated virions used for the infections was sequenced via the PCR-mediated amplification of complementary DNA synthesized with AMV reverse transcriptase.

Received 14 April; accepted 13 May 2003; doi:10.1038/nature01709.
Published online 28 May 2003.

1. von Schwedler, U., Song, J., Aiken, C. & Trono, D. Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. *J. Virol.* **67,** 4945–4955 (1993).
2. Gabuzda, D. H. *et al.* Role of vif in replication of human immunodeficiency virus type 1 in $CD4^+$ T lymphocytes. *J. Virol.* **66,** 6489–6495 (1992).
3. Simon, J. H. & Malim, M. H. The human immunodeficiency virus type 1 Vif protein modulates the postpenetration stability of viral nucleoprotein complexes. *J. Virol.* **70,** 5297–5305 (1996).
4. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418,** 646–650 (2002).
5. Jarmuz, A. *et al.* An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79,** 285–296 (2002).
6. Powell, L. M. *et al.* A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **50,** 831–840 (1987).
7. Teng, B., Burant, C. F. & Davidson, N. O. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260,** 1816–1819 (1993).
8. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* **10,** 1247–1253 (2002).
9. MacGinnitie, A. J., Anant, S. & Davidson, N. O. Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity. *J. Biol. Chem.* **270,** 14768–14775 (1995).
10. Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery *in vivo*. *Nature Biotechnol.* **15,** 871–875 (1997).
11. Kotsopoulou, E., Kim, V. N., Kingsman, A. J., Kingsman, S. M. & Mitrophanous, K. A. A Rev-independent human immunodeficiency virus type 1 (HIV-1)-based vector that exploits a codon-optimized HIV-1 gag-pol gene. *J. Virol.* **74,** 4839–4852 (2000).
12. Negre, D. *et al.* Characterization of novel safe lentiviral vectors derived from simian

immunodeficiency virus (SIVmac251) that efficiently transduce mature human dendritic cells. *Gene Ther.* **7,** 1613–1623 (2000).
13. Mitrophanous, K. *et al.* Stable gene transfer to the nervous system using a non-primate lentiviral vector. *Gene Ther.* **6,** 1808–1818 (1999).
14. Charneau, P. *et al.* HIV-1 reverse transcription. A termination step at the center of the genome. *J. Mol. Biol.* **241,** 651–662 (1994).
15. Jayanta, C., Ming, T., Chan, K., Eric, P. & Alt, F. W. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **422,** 726–730 (2003).
16. Klarmann, G. J., Chen, X., North, T. W. & Preston, B. D. Incorporation of uracil into minus strand DNA affects the specificity of plus strand synthesis initiation during lentiviral reverse transcription. *J. Biol. Chem.* **278,** 7902–7909 (2003).
17. Borman, A. M., Quillent, C., Charneau, P., Kean, K. M. & Clavel, F. A highly defective HIV-1 group O provirus: evidence for the role of local sequence determinants in G → A hypermutation during negative-strand viral DNA synthesis. *Virology* **208,** 601–609 (1995).
18. Vartanian, J. P., Meyerhans, A., Asjo, B. & Wain-Hobson, S. Selection, recombination, and G → A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol.* **65,** 1779–1788 (1991).
19. Wain-Hobson, S., Sonigo, P., Guyader, M., Gazit, A. & Henry, M. Erratic G → A hypermutation within a complete caprine arthritis-encephalitis virus (CAEV) provirus. *Virology* **209,** 297–303 (1995).
20. Johnson, P. R., Hamm, T. E., Goldstein, S., Kitov, S. & Hirsch, V. M. The genetic fate of molecularly cloned simian immunodeficiency virus in experimentally infected macaques. *Virology* **185,** 217–228 (1991).
21. Turelli, P., Guiguen, F., Mornex, J. F., Vigne, R. & Querat, G. dUTPase-minus caprine arthritis-encephalitis virus is attenuated for pathogenesis and accumulates G-to-A substitutions. *J. Virol.* **71,** 4522–4530 (1997).
22. Mansky, L. M. *In vivo* analysis of human T-cell leukemia virus type 1 reverse transcription accuracy. *J. Virol.* **74,** 9525–9531 (2000).
23. Pathak, V. K. & Temin, H. M. Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc. Natl Acad. Sci. USA* **87,** 6019–6023 (1990).
24. Gunther, S. *et al.* Naturally occurring hepatitis B virus genomes bearing the hallmarks of retroviral G → A hypermutation. *Virology* **235,** 104–108 (1997).
25. Turelli, P. *et al.* Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. *Mol. Cell* **7,** 1245–2154 (2001).
26. Dettenhofer, M. & Yu, X. F. Highly purified human immunodeficiency virus type 1 reveals a virtual absence of Vif in virions. *J. Virol.* **73,** 1460–1467 (1999).
27. Butler, S. L., Hansen, M. S. & Bushman, F. D. A quantitative assay for HIV DNA integration *in vivo*. *Nature Med.* **7,** 631–634 (2001).
28. Yerly, S. *et al.* Transmission of antiretroviral-drug-resistant HIV-1 variants. *Lancet* **354,** 729–733 (1999).

Acknowledgements We thank S. Vianin for technical assistance, and M. Malim, R. Mulligan, K. Mitrophanous, F.-L. Cosset and M. Matsuda for the gift of reagents. This work was supported by the Swiss National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.T. (didier.trono@medecine.unige.ch).

Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation

Phuong Pham, Ronda Bransteitter, John Petruska & Myron F. Goodman

Departments of Biological Sciences and Chemistry, Hedco Molecular Biology Laboratories, University of Southern California, University Park, Los Angeles, California 90089-1340, USA

Activation-induced cytidine deaminase (AID) is a protein required for B cells to undergo class switch recombination and somatic hypermutation (SHM)—two processes essential for producing high-affinity antibodies[1]. Purified AID catalyses the deamination of C to U on single-stranded (ss)DNA[2–4]. Here, we show *in vitro* that AID-catalysed C deaminations occur preferentially on 5′ WR*C* sequences in accord with SHM spectra observed *in vivo*. Although about 98% of DNA clones suffer no mutations, most of the remaining mutated clones have 10–70 C to T transitions per clone. Therefore, AID carries out multiple C deaminations on individual DNA strands, rather than jumping from one strand to another. The avid binding of AID to ssDNA could result from its large net positive charge (+11) at pH 7.0, owing to a basic amino-terminal domain enriched in arginine and lysine. Furthermore, AID exhibits a 15-fold preference for C deamination on the non-transcribed DNA strand exposed by RNA polymerase than the transcribed strand protected as a RNA–DNA hybrid. These deamination results on ssDNA bear relevance to three characteristic features of SHM: preferential mutation at C sites within WR*C* hotspot sequences, the broad clonal mutagenic heterogeneity of antibody variable regions targeted for mutation[5,6], and the requirement for active transcription to obtain mutagenesis[7,8].

AID was identified three years ago[9,10], and knowledge of its function in the generation of high-affinity antibodies has helped to reveal why children lacking AID exhibit an inefficient antibody response[10]. Despite the similarity of AID in sequence to APOBEC1, a cytidine deaminase involved in messenger RNA editing[11], initial AID purifications indicated no discernible activity on nucleic acid substrates *in vitro*[12]. Recently, however, we were able to show that human AID purified from *Baculovirus*-infected insect cells catalyses the deamination of C → U on ssDNA[2], but not on double-stranded (ds)DNA, RNA–DNA hybrids or ssRNA[2], with similar conclusions drawn by other groups using *Escherichia coli*[4] and activated B-cell extracts[3] to purify AID. SHM-like spectra were observed in *E. coli*, DT40 chicken B cells and in mice lacking Ung activity expressing endogenous levels of AID[6,13,14], demonstrating that AID-catalysed C deamination has a definitive role in the mutational targeting process. In an effort to reconstitute SHM *in vitro*, we have examined the action of AID on a long ssDNA target sequence including a *lacZ* alpha (*lacZa*) reporter gene. We find that the action of AID yields a mutation spectrum with hotspots and cold spots much like those observed for SHM *in vivo*. The action of AID also shows a highly diverse clonal pattern of mutations, leaving many ssDNA molecules unmutated while introducing multiple mutations into a few molecules (Fig. 1).

AID was incubated with phage M13mp2 circular DNA substrate containing a 230-nucleotide target of the *lacZa* reporter sequence in a 365-nucleotide, single-stranded gapped region (Fig. 1a). This system has been used extensively to measure the fidelity of DNA polymerases by scoring mutated (white and light-blue plaques) and non-mutated (dark-blue plaques) phage progeny after transfection into *E. coli*[15]. AID-catalysed deamination events occurring on individual ssDNA *lacZa* reporter sequences are detected as C → T

Table 1 **AID-catalysed C → U sequences**

Hotspots		Cold spots	
Site	No. of C → U	Site	No. of C → U
ATA**C**GC	34	TGC**C**CG	0
ATG**C**TT	34	TCC**C**GA	0
CAG**C**TA	33	CGC**C**TT	0
CAA**C**TT	32	CGC**C**AG	0
ATG**C**AG	31	GGC**C**GA	0
AAA**C**CC	31	TtT**C**AC	1
TTA**C**CC	31	CGC**C**TC	1
AAA**C**CA	29	TCa**C**AC	1
TAG**C**TG	28	ACa**C**AG	1
AcG**C**AA	28	GGC**C**GT	1
AAA**C**CG	28	CGT**C**GT	2
TTG**C**AG	28	TaaCAA	2
CAG**C**AC	28	TGg**C**CG	3
AAA**C**AG	26	CGC**C**CA	3
TTA**C**GA	26	TCT**C**AC	3
		CGa**C**AG	3
		CGT**C**GT	4
WR*C*		**SY*C***	

Hotspots and cold spots were derived from the AID-catalysed deamination spectrum shown in Fig. 2. Hotspots are defined as C deamination sites (underlined) occurring in ≥50% of the mutant phage. Cold spots are C sites occurring in ≤8% of the mutated phage. Mismatches within the consensus motifs are indicated by lower-case letters. All sequences are indicated 5′ to 3′. The consensus motifs are indicated in bold. W = A or T; S = G or C; R = purine; Y = pyrimidine.