Reviews

Vif and the Role of Antiviral Cytidine Deaminases in HIV-1 Replication ........................................ 2
Qin Yu, Nathaniel R. Landau, and Renate König

Forms and Function of Intracellular HIV DNA ........................................................................ 14
Andreas Meyerhans, Tanja Breinig, Jean-Pierre Vartanian, and Simon Wain-Hobson

The Microbial Forensic Use of HIV Sequences ......................................................................... 22
Gerald H. Learn and James I. Mullins

Mutations in Retroviral Genes Associated with Drug Resistance .............................................. 38
Urvi Parikh, Charles Calef, Shauna A. Clark, and John W. Mellors
Vif and the Role of Antiviral Cytidine Deaminases in HIV-1 Replication

Qin Yu, Nathaniel R. Landau and Renate König
The Salk Institute, La Jolla, CA

Abstract

Vif is required for HIV-1 replication in primary cells and other nonpermissive cell lines. The function of Vif was unknown until recent papers revealed that it prevents the encapsidation of APOBEC3G and APOBEC3F, two potent antiretroviral cytidine deaminases. APOBEC3G and APOBEC3F inhibit the replication of Δvif HIV-1 by deaminating the minus-strand of the viral reverse transcripts, introducing numerous G→A mutations. These two APOBEC3 family members appear to be the major contributors to G→A hypermutation in HIV-1 in vivo. Here we review recent advances in the field, emphasizing the sequence specificity of the antiviral APOBEC3 deaminases and their roles in molding the viral genome.

The Vif phenotype

vif (virion infectivity factor) was originally identified as an open reading frame in the central portion of the viral genome that was present in all of the lentiviruses except for equine anemia infectious virus (52, 62). vif encodes an approximately 220 amino acid cytoplasmic protein that is expressed from a multiply spliced, Rev-dependent mRNA (23, 66). In contrast, simpler retroviruses like murine leukemia virus (MLV) or avian sarcosis and leukosis virus (ASLV) do not contain vif. The more complex foamy viruses and human T cell leukemia virus (HTLV) also lack an obvious vif gene. vif along with nef, vpr/vpx, and vpu, is classified as an accessory gene because it is not required for virus replication in some cell types (17, 20, 66, 71).

Cell types in which vif is required for robust HIV-1 replication are termed “nonpermissive”. Conversely, cell types in which vif is not required are termed “permissive” (20, 71). Nonpermissive cells include primary T cells and transformed T cell lines such as PM1, Hut78, and CEM. Permissive cells include T cell lines SupT1, C8166, and CEMss as well as nonlymphoid cell lines such as HeLa (20, 71). The Vif phenotype is controlled by the producer cells and not the target cells (71). That is, Δvif HIV-1 produced in nonpermissive cells is noninfectious regardless of target cell phenotype. Nonpermissive cells produce normal amounts of Δvif virions but its infectivity was reduced two to three orders of magnitude compared to wild-type (24, 60, 61). Δvif virus was able to enter target cells but the infection aborts during or after reverse transcription and prior to integration (24, 60, 71). The protein and RNA content of Δvif virions was indistinguishable from wild-type virions in early studies (19, 21, 29, 58).

The first hint of the molecular basis of this cell-specific requirement for Vif was obtained in heterokaryon experiments. In fusions of permissive cells with nonpermissive cells, the nonpermissive phenotype was dominant (39, 59). This result predicted that nonpermissive cells expressed a dominant factor that was counteracted by Vif.

APOBEC3G and APOBEC3F as inhibitory host factors

Sheehy et al. identified the hypothesized inhibitory protein using a subtractive cDNA cloning approach (55). The strategy used involved the human T cell lines, CEM and CEMss, which are derived from a common parental cell yet CEM is nonpermissive whereas CEMss is permissive. Subtraction of CEMss cDNA from CEM cDNA yielded a cDNA termed CEM15. Transfer of CEM15 to CEMss caused it to become nonpermissive. Thus, CEM15, which was identical to the mRNA encoding apolipoprotein
B (apoB) mRNA-editing enzyme catalytic polypeptide-like 3G (APOBEC3G) (31), fits the definition of the Vif cofactor. In accord with the nonpermissive phenotype, CEM15 did not affect the amount of virus released from transfected cells, but dramatically reduced the infectivity of ∆vif but not wild-type HIV-1 (55).

Recently, APOBEC3F, the most closely related APOBEC3 family member, was identified as a second anti-HIV host factor that is sensitive to Vif (74, 80). Like APOBEC3G, APOBEC3F is expressed in CEM but not CEMss cells, and specifically inhibits ∆vif not wild-type HIV-1 infectivity (7, 38, 74, 80).

APOBEC3G and APOBEC3F belong to the APOBEC family of cytidine deaminases, which includes APOBEC1 and activation induced deaminase (AID) arranged in tandem on chromosome 12; APOBEC2 on chromosome 6; and APOBEC3A-G arranged in tandem on chromosome 22 (31). The mouse genome contains only a single APOBEC3 gene (42), suggesting that APOBEC3A-G is a fairly recent evolutionary expansion (79). APOBEC1 is one of the prototypical mammalian RNA editing enzymes. RNA editing, as first described in protozoans, refers to post-transcriptional alteration of cellular RNA in which specific nucleotides are catalytically changed (3, 73). The alterations usually result in nonsense or missense mutations that change the coding capacity of the mRNA. The alterations are most frequently either A→I (inosine) or C→U, both of which are generated by deamination of the target base. APOBEC1 is expressed in the intestine and also the liver in some mammals. APOBEC1 specifically changes a single C (C6666) in apoB mRNA to U, which creates a premature stop codon that leads to a truncated form of apoB100 protein, apoB48 (13, 26, 51, 65). APOBEC1 is also active as a DNA deaminase in vitro and when expressed in bacteria (28, 50). AID, which is expressed in subsets of B lymphocytes, deaminates immunoglobulin variable region DNA to mediate somatic hypermutation and class switch recombination (2, 45, 46, 49). APOBEC3G and APOBEC3F, the most closely-related APOBEC3 members, are expressed primarily in lymphoid and myeloid cell lineages (31, 38, 74), but whether they play roles in the immune system other than as anti-retroviral proteins is not known. Two other APOBEC3 members, APOBEC3B and APOBEC3C, were recently analyzed against HIV-1. APOBEC3B was reported to inhibit HIV-1 whereas APOBEC3C did not affect the virus (27, 74, 80).

APOBEC3 family members contain a conserved motif at the active site domain that is characteristic of cytidine deaminases. The most conserved features of this motif are three zinc coordinating residues (His and two Cys) and a Glu residue that are involved in proton shuttling during catalysis (31). The active site domain is joined by a short linker to a pseudoactive site that contains the conserved His-Glu-Cys-Cys residues but does not coordinate Zn2+ (Fig. 1). In APOBEC3B, APOBEC3F, and APOBEC3G, this unit is duplicated. For APOBEC3G, the second active site was indicated to be the major catalytic site for deamination (57). APOBEC proteins generally homodimerize (31). Interestingly, APOBEC3F is able to form heterodimers with APOBEC3G (74).

Death by deamination

In nonpermissive cells, APOBEC3G and APOBEC3F are encapsidated in abundance into ∆vif HIV-1 virions, resulting in near 1,000-fold reduction in infectivity compared to wild-type (38, 42, 43, 56, 63, 74). Sequence analysis revealed that APOBEC3G/APOBEC3F generate numerous G→A mutations in the reverse transcripts of ∆vif HIV-1 following infection (Fig. 2) (7, 27, 35, 38, 40, 42, 74, 78).

Figure 1. Schematic of APOBEC3G domain organization. Amino acid 128 controls the species-specificity of APOBEC3G interaction with Vif.
These result from the deamination of cytosines in the minus-strand of the reverse transcript (Fig. 3). The minus-strand serves as the template for the plus-strand. Therefore, A is synthesized on the plus-strand opposite the U, resulting in G→A transitions (27, 35, 40, 42, 78). APOBEC3G induced G→A mutations were found over the length of the genome and were present with a graded frequency in the 5′→3′ direction (Fig. 4) (76).

Interestingly, G→A hypermutation has been sporadically reported in patient isolates of HIV (9, 16, 18, 25, 37, 69) and in other lentiviruses (22, 48, 72). This was initially attributed to reverse transcription errors caused by nucleotide pool imbalances. But in retrospect, it was likely to have been caused, at least in part, by APOBEC3G and APOBEC3F. It will be of interest to determine whether G→A hypermutation in vivo is caused by viruses that encode a defective Vif, or by patients that over-express APOBEC3G/APOBEC3F or encode polymorphic enzymes that is relatively Vif-resistant.

### Strand preference of APOBEC3G

Nearly all of the APOBEC3G-induced mutations in HIV-1 are G→A, not C→T. This could result either from specific deamination of the minus-strand, or from deamination of both strands followed by repair of the plus-strand triggered by the mismatch between the two strands. Recent findings suggested that the near absence of C→T change results from the single-strand specificity of APOBEC3G (76).

In vitro, APOBEC3G was found to bind and deaminate single-stranded DNA, but not double-stranded DNA, or DNA/RNA hybrids. This is similar to APOBEC1 and AID, which also target single-stranded nucleic acid (12, 47, 50, 65). The requirement of single-stranded DNA target for APOBEC3G accounts for minus-strand specificity because only the minus-strand (for the most part) proceeds through a transient single-stranded phase during reverse transcription (Fig. 4). Indeed, APOBEC3G-induced C→T mutations were found at regions of the plus-strand that become transiently single-stranded during reverse transcription, including the 5′ U3 and the primer binding site (PBS) (76). The PBS mutation (642C→T) has been suggested to result from reverse transcription priming by an alternative tRNA, tRNAlysS (15, 70). While tRNAlysS priming may cause this PBS mutation in some viruses, the same mutation was found more frequently in reverse transcripts derived from APOBEC3G-containing viruses than those lacking the enzyme, suggesting that APOBEC3G can also cause this mutation. The graded frequency of G→A mutations over HIV-1 genome may be caused by the varied length of time that each region of the viral genome remains single-stranded during reverse transcription (Fig. 4) (76).

For the retrovirologists, APOBEC3G may be viewed as a biological probe of the molecular events in reverse transcription. The polypurine tract (PPT) near the 3′ LTR is relatively RNase H resistant and serves to prime plus-strand DNA synthesis. APOBEC3G did not deaminate the string of C nucleotides present at the PPT minus-strand, suggesting that the PPT is only briefly single-stranded, if at all (76). It may not be removed until it is displaced by the plus-strand DNA (Fig. 4). A central polypurine tract (cPPT) is thought to serve as an additional site of plus-strand initiation (11). Indeed, on a clade O HIV-1 isolate that contains an extremely high frequency of G→A mutations, both the PPT and the cPPT are protected from G→A mutations (68), supporting the use of both sites for plus-strand priming.

### Sequence preference of APOBEC3G and APOBEC3F

The target sequence preference of human APOBEC3G has been analyzed in virus culture and in vitro assays (4, 27, 76, 78). The HIV sequence database website provides a well-designed program, called Hypermut (http://www.hiv.lanl.gov/content/hiv-db/HYPERMUT/hypermut.html), for analyzing the G→A mutation rate, nucleotide context, and stop codon occurrence (Fig. 2) (53). In the dinucleotide context, human APOBEC3G preferentially targeted CC (where the underline designates the base that becomes deaminated), occasionally TC, but rarely AC or GC. Thus, APOBEC3G induced G→A mutations mostly in GG, less frequently in GA, but not in GT and GC (Fig. 2). In tetranucleotide context, the consensus target sequence for human APOBEC3G was CCCC/G (76). A G at the –2 position (GC) was disfavored. The target sequence preference of APOBEC3G found in the reverse transcripts of virus generated from cell lines was the same as on oligonucleotide model substrates in vitro, indicating that deamination hotspots were caused by the intrinsic preference of the enzyme for target sequences (76).
Figure 2. Examples of APOBEC3G induced G→A mutations in gag of the wild-type or Δvif HIV-1 produced in the presence or absence of indicated APOBEC. A representative clone of each virus group is depicted in the Hypermut program generated graph (76). Each mutation is illustrated as a vertical line color coded with respect to dinucleotide context: GG→AG (red), GA→AA (blue), GT→AT (green), non-G→A (black).

Figure 3. A model for induction of G→A mutation by APOBEC3G-catalyzed deamination of HIV-1 minus strand DNA. Viral RNA is reverse transcribed to generate minus-strand DNA, which becomes single-stranded upon removal of the RNA template and subject to APOBEC3G C→U deamination. The U-containing minus-strand serves as template for plus-strand DNA synthesis, resulting in plus strand A opposite to the minus strand U.
Figure 4. A model for deamination of Δvif HIV-1 DNA by APOBEC3G during reverse transcription (adapted from Yu et al. (76)). Reverse transcription is initiated from the tRNA\textsuperscript{pre} primer annealed to the PBS (step 1). RNase H degrades the genomic RNA template as it is reverse transcribed. The minus-strand strong-stop DNA is transferred to the 3' end of the viral RNA and continues minus-strand synthesis (step 2). APOBEC3G deaminates the minus-strand DNA as it becomes single-stranded (red oval). Plus-strand DNA synthesis initiates from the RNase H-resistant PPT and cPPT, and is extended to copy the 3’ 18 nt of tRNA\textsuperscript{pre} to regenerate the PBS (step 3). After the removal of tRNA, the plus-strand PBS becomes transiently single-stranded and a target for deamination (step 4). Transfer of the plus-strand strong stop DNA to the 3’ end of minus-strand DNA leads to circularization of the DNA strands (step 5). The 5’ end of the plus-strand DNA becomes transiently single-stranded during the displacement synthesis (step 6) and is deaminated. The linear double-stranded DNA with LTRs is completed (step 7).
Unlike APOBEC3G, APOBEC3F preferentially targets TC, which would result in G→A mutation in GA context (7, 38, 74). The G→A hypermutation found in HIV-1 from patient samples and culture PBMCs occurred exclusively at GA and GG sequences, the preferred target sites of APOBEC3F and APOBEC3G (18, 30, 69). Thus, APOBEC3F and APOBEC3G are likely to be the major contributors to HIV-1 hypermutation in vivo.

African green monkey (AGM) APOBEC3G, which shares ~77% homology with the human APOBEC3G, also targets the same sequence in wild-type and Δvif HIV-1 genome (42, 76). Mouse APOBEC3 is ~30% identical to its human homolog and exhibits a less stringent target sequence preference. Mouse APOBEC3 deaminates TC and TT, with no preference for the nucleotide 3′ to the target C (76). Other APOBEC family members show different target sequence preferences compared to APOBEC3G. In a bacterial mutation assay and in vitro deaminase assay, recombinant APOBEC1 preferred TC while AID preferred A/G (4).

Cytosine deamination has two consequences for HIV-1 replication. First, it causes the degradation of most reverse transcripts prior to integration (42). The degradation probably is triggered by DNA repair enzymes such as uracil DNA glycosylase (UDG) that removes uracil from DNA and AP endonuclease that cleaves at abasic sites. Second, the G→A mutations introduce numerous stop codons into the open reading frames of the virus. Deamination at the consensus target sequence of human APOBEC3G generates TAG, TAA, and TGA triplets, corresponding to the three translational stop codons (76). Creation of abundant stop codons by G→A hypermutation is also common in HIV-1 field isolates (30). The ATG initiation codon of gag and nef are also frequently altered by human and AGM APOBEC3G (76).

APOBEC3G and APOBEC3F are encapsidated in abundance in Δvif HIV-1 whereas barely present in the wild-type virions (32, 42-44, 56, 63, 74, 80). Functionally inactive Vif mutants failed to prevent APOBEC3G encapsidation (32, 43, 44). These findings suggested that Vif functions to prevent APOBEC3G and APOBEC3F encapsidation, which was later found to occur through the formation of Vif:APOBEC3G/3F complexes that could be detected by coimmunoprecipitation (42-44, 56, 63, 74, 77).

**Species-specific Vif:APOBEC3G interaction**

Of the primate APOBEC3G, chimpanzee APOBEC3G is most similar to its human homolog (95% identity). AGM and macaque APOBEC3G are more divergent, with 77% and 75% identity to human APOBEC3G, respectively. These three primate proteins as well as mouse APOBEC3 are active against Δvif HIV-1. Interestingly, they also block wild-type HIV-1 replication, with the exception of the chimpanzee homolog (42). The failure of HIV-1 Vif to counteract primate APOBEC3G may provide a species-specific barrier to primate lentivirus transmission. Furthermore, the ability of Vif to adapt to chimpanzee APOBEC3G may have been a prerequisite for the zoonosis of the virus to humans. In contrast, SIVmac Vif neutralized all of the APOBEC3G proteins that were tested. Whereas SIVagm Vif did not interact with human APOBEC3G nor prevent its encapsidation (42). By using human-AGM chimeric APOBEC3G proteins, several groups mapped the determinant of this species specific interaction to a single amino acid at position 128 (Fig. 1) (8, 41, 54, 75). Interchange of Asp128 in human APOBEC3G with Lys128 of AGM APOBEC3G switched the physical association of the proteins and their sensitivity to HIV-1 and SIVagm Vif. The specificity of the Vif:APOBEC3G interaction suggests that the proteins associate directly without the requirement for a common intermediate such as viral RNA.

The encapsidation of APOBEC3G into retroviral virions is less virus-specific. APOBEC3G can be encapsidated by HIV, SIV, and murine leukemia virus (MLV) through a mechanism thought to involve binding to viral RNA, cellular RNA, or the viral nucleocapsid protein (1, 10, 36, 64). Interestingly, APOBEC3G can also be encapsidated in hepatitis B virus and is thought to inhibit its replication without deamination of the viral DNA (67).

**Vif-induced degradation of APOBEC3G**

Exclusion of APOBEC3G from virions by Vif could have been caused by masking of an encapsidation domain on APOBEC3G or by inducing APOBEC3G degradation. Although initial pulse-
chase analysis did not detect an effect of Vif on APOBEC3G half life (42), analyses with very short pulses showed that Vif causes extremely rapid degradation of APOBEC3G, reducing its half-life to about two minutes (43). Some groups have reported a half-life of 30 min to 4 h in the presence of Vif (14, 32, 44, 56, 63, 77). Vif-induced rapid degradation of APOBEC3G accounted for its effective exclusion from virions. Proteasome inhibitors which prevented Vif-mediated APOBEC3G degradation, restored APOBEC3G to steady-state levels and led to encapsidation of Vif:APOBEC3G complexes into virions (14, 32, 44, 56, 63, 77). Interestingly, these Vif:APOBEC3G complexes appeared to remain active, indicating that Vif binding does not block cytidine deaminase activity. An effect of Vif on the translation of APOBEC3G mRNA has been reported to further reduce APOBEC3G encapsidation (63).

The mechanism of Vif-induced APOBEC3G degradation was elucidated by Yu et al. who found that Vif associates with an elongin/cullin/SOCS box E3 ligase complex (77). This complex is similar to the Skp2/cullin/F-box E3 ligase complex and ubiquitinates a large number of cellular proteins, including cytokine receptors, signaling proteins, and cell cycle proteins (33). The SOCS box protein serves as an adaptor that links a substrate cellular protein to the E3 ligase (33). Once bound, the substrate is ubiquitinated by E3 ligase and E2-ubiquitin conjugating enzyme that associates with the complex through Rbx1. The SLQLXLA motif in Vif resembles the central region of some SOCS boxes. Mutation in this motif prevents the association of Vif with E3 and blocks the Vif-induced APOBEC3G degradation (77).

Vif does not fully protect the virus from APOBEC3G and APOBEC3F, leaving small amounts of these enzymes encapsidated in wild-type virions (42, 43, 63, 74, 80). This may cause a continual low level of G→A mutations in viruses that replicate in vivo. A thorough analysis of HIV-1 proviral sequences from patient PBMCs showed that at least 43% of patients harbored abundant G→A hypermutated viral DNA (30). Interestingly, in cultured primary CD4+ T cells, G→A hypermutation is most pronounced when HIV-1 infection occurs within a few hours of phytohemagglutinin (PHA) activation (30). After 12h of PHA activation, most HIV-1 sequences were normal. This suggested that expression of APOBEC3G and APOBEC3F increases significantly upon PHA activation.

**Evolutionary skewing of retroviral genomes by host deaminases**

APOBEC3G and APOBEC3F appear to have played a role in molding the HIV genome over evolution, resulting in skewing of the nucleotide content and codon usage in retroviral genomes (5, 6). The HIV-1 genome is A-rich, while other retroviruses, such as HTLV-I and MLV, are C-rich. HIV contains 36% A bases and the third base positions of codons have a 60% bias for A (5, 6). The preferred targets for APOBEC3G are the sequences complementary to CGGG and TGGG. Interestingly, CGGG sequences, but not TGGG sequences, are strongly suppressed in HIV-1 genome. CGG encodes Arg which can be encoded by six codons, CGN and AGG/A. In HIV-1, CGG is rarely used to encode Arg (3%). Instead, there is an overwhelming bias (65%) for AGA, which is not a target for APOBEC3G (76). For HTLV-I, in contrast, AGA is not favored (12%) and CGG is not suppressed (17%) as Arg codon (76). TGGG sequences, although they are favored targets of APOBEC3G, contain TGG, the single codon for Trp. Because only one codon encodes Trp, these sequences cannot be suppressed without losing Trp residues. The skewing of nucleotides and codons in the HIV-1 genome may have been caused by evolutionary pressure from APOBEC3G, APOBEC3F, and potentially from other APOBEC family members as well. Despite the controversy on the effect of APOBEC3G towards MLV replication (27, 34, 42), the differences in nucleotide content suggest that HIV-1 is susceptible to APOBEC3 but that HTLV-1 and MLV are not. This is not likely to be due to lack of the APOBEC enzymes in the target cells, as both HTLV-1 and HIV replicate in human CD4+ T cells. Thus, HTLV-1 and MLV, which do not encode Vif, must avoid APOBEC3 deamination through an alternative mechanism that remains to be determined.

**Future directions**

Several interesting questions remain to be addressed. How did Vif originate in the lentiviral genome? Do APOBEC3G and APOBEC3F have a function in the immune system in addition to viral deamination? Do they act on other retroviruses or other viruses that replicate through a single-stranded
intermediate? What are the functions of the various APOBEC3 family members and why is a single
APOBEC3 sufficient in the mouse? And most importantly, can the recent advances in Vif lead to develop-
ment of novel therapeutics? Clearly, a small molecule that disrupted the Vif:APOBEC3G/3F interac-
tion but did not interfere with deaminase activity would be desirable.

References

71:817–46.
Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes:
retroviral genomes is consistent with virus-specific nucleotide pressure. AIDS Res. Hum.
Retroviruses 18:133–41.
[6] Berkhout, B., and F. J. van Hemert. 1994. The unusual nucleotide content of the HIV RNA ge-
ence in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion
HIV-1 group O provirus: evidence for the role of local sequence determinants in G→A
between HIV-1 Gag and APOBEC3G. J. Biol. Chem. 279:33177–84.
unintegrated linear DNA defined by a central copy of the polypurine tract. J. Virol. 65:2415–21.
P. Deslypere, M. Rosseneu, and et al. 1987. Apolipoprotein B-48 is the product of a messenger
[14] Conticello, S. G., R. S. Harris, and M. S. Neuberger. 2003. The Vif protein of HIV triggers degra-
ciency virus primer-binding site suggests the use of an alternative tRNA(Lys) molecule in reverse
virus type 1 nef and long terminal repeat sequences over 4 years in vivo and in vitro. J. Virol.
Vif & antiviral cytidine deaminases


Vif & antiviral cytidine deaminases


Vif & antiviral cytidine deaminases


single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to

[76] Yu, Q., R. Konig, S. Pillai, K. Chiles, M. Kearney, S. Palmer, D. Richman, J. M. Coffin, and N. R.


APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication.