Virology 376 (2008) 101–111

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Variable contexts and levels of hypermutation in HIV-1 proviral genomes recovered from primary peripheral blood mononuclear cells

Gustavo H. Kijak [a,*], Luiz Mario Janini [a,b], Sodsai Tovanabutra [a], Eric Sanders-Buell [a], Miguel Angel Arroyo [a,c,1], Merlin L. Robb [a], Nelson L. Michael [c], Debora L. Birx [c,2], Francine E. McCutchan [a]

[a] Henry M. Jackson Foundation for the Advancement of Military Medicine/US Military HIV Research Program, 1600 East Gude Drive, Rockville, Maryland 20850, USA
[b] Infectious Diseases Division, Paulista School of Medicine, Federal University of São Paulo, R. Pedro de Toledo 781, 16 andar Vila Clementino, São Paulo, SP-04039, Brazil
[c] Division of Retrovirology, Walter Reed Army Institute of Research, 1600 East Gude Drive, Rockville, Maryland 20850, USA

ARTICLE INFO

Article history:
Received 5 January 2008
Returned to author for revision
20 February 2008
Accepted 8 March 2008
Available online 24 April 2008

Keywords:
HIV-1
Hypermutation
Innate immunity
Host restriction
APOBEC3G
APOBEC3F

ABSTRACT

APOBEC-mediated cytidine deamination of HIV-1 genomes during reverse transcription has been shown to be a potent mechanism of host restriction for HIV-1 infection *ex vivo* and *in vitro*. However, this defense system can be overcome by the viral protein Vif. Unlike other mechanisms of host restriction, the APOCEC–Vif interaction leaves an imprint on integrated proviruses in the form of G→A hypermutation. In the current work we systematically studied levels, contexts, and patterns of HIV-1 hypermutation *in vivo*. The analysis of 24 full-genome HIV-1 sequences retrieved from primary PBMCs, representing infections with several HIV-1 clades, and the inclusion of 7 cognate pairs of hypermutated/non-hypermutated sequences derived from the same patient sample, provided a comprehensive view of the characteristics of APOBEC-mediated restriction *in vivo*. Levels of hypermutation varied nearly 5-fold among the studied proviruses. GpG motifs were most frequently affected (22/24 proviruses). Levels of hypermutation varied across the genome. The reported "twin peak" pattern of hypermutation was observed in 18/24 hypermutants, but the remainder exhibited singular non-conforming patterns. These data suggest considerable complexity in the interplay of host restriction and viral defense during HIV-1 infection.

Introduction

Infection with the human immunodeficiency virus type 1 (HIV-1) is characterized by an early and extensive damage of key histological elements (the CD4+ cells), with a consequential loss of adaptive immune functions than is usually irreparable (Graziosi et al., 1998). This destruction occurs before the effectors of the adaptive immunity can even be detected (Brenchley et al., 2004; Mehandru et al., 2004). Especially during these early weeks, but still persisting throughout the whole course of infection, elements of innate immunity have a key role in controlling viral replication (Goff, 2004).

Only a few innate mechanisms of host restriction to HIV infection have been described so far(Gonzalez et al., 2005; Liu et al., 1996; Samson et al., 1996; Stremlau et al., 2004), and for each of them there is a recognized viral mechanism of evasion (Braaten and Luban, 2001; Connor et al., 1997). One of these mechanisms of host restriction to HIV infection is mediated by the members of the apolipoprotein B mRNA-editing catalytic polypeptide (APOBEC) enzyme family, with deaminase activity. APOBEC3G (Sheehy et al., 2002) and APOBEC3F (Zheng et al., 2004), and possibly other APOBECs (Rose et al., 2005), have been shown to possess strong activity in preventing HIV infection, especially in *ex vivo* models. These enzymes are incorporated into nascent viral particles and act during the next infection cycle, when a transient single-stranded DNA (ssDNA) replication intermediate form of the HIV genome becomes exposed during reverse transcription (Harris and Liddament, 2004). Deamination of cytidine to uracil predominantly occurs on the minus strand intermediate (Lecossier et al., 2003), which can be evidenced as guanine-to-adenine (G→A) changes in the plus strand. Thus, the effect of the APOBECs is to destroy the viral genetic information through extensive G→A substitution. HIV can counteract APOBEC-mediated restriction through the expression of the product of the *vif* gene. The Vif protein prevents the incorporation of APOBECs into the nascent viral particle, targeting the host deaminase for ubiquitylation and subsequent proteasomal degradation (Mehle et al., 2004). The Vif-mediated exclusion of APOBECs spares the virus from this host restriction system.

HIV genomes with APOBEC-mediated G→A changes can either be degraded or can be integrated into the host chromosome. More than a

* Corresponding author. Fax: +1 301 762 7460.
E-mail addresses: gkijak@hivresearch.org (G.H. Kijak), mjanini@pesquisa.epm.br (L.M. Janini), stovanabutra@hivresearch.org (S. Tovanabutra), esanders-buell@hivresearch.org (E. Sanders-Buell), Miguel.Arroyo@afrims.org (M.A. Arroyo), mrobb@hivresearch.org (M.L. Robb), nmichael@hivresearch.org (N.L. Michael), emg5@cdc.gov (D.L. Birx), fmccutchan@hivresearch.org (F.E. McCutchan).
[1] Current address: Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand.
[2] Current address: Centers for Disease Control and Prevention, Atlanta, Georgia, USA.

0042-6822/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2008.03.017

decade before the elucidation of the APOBEC restriction mechanism, HIV sequences showing "extensive and monotonous base substitutions of G→A" were first recognized in primary or cultured viral genomes, and they were termed HIV hypermutants (Vartanian et al., 1991), based on the term previously coined by Pathak and Temin for the spleen necrosis virus model (Pathak and Temin, 1990). Until the discovery of the APOBEC restriction system, hypermutation was often considered as an advantageous mechanism of HIV genetic variation that could accelerate its escape from anti-retroviral drugs and immune pressure (Diamond et al., 2001). HIV hypermutation has also been attributed to different factors, including Taq polymerase- (Overbaugh et al., 1996) or HIV reverse transcriptase (RT)-induced errors, or imbalances in intracellular nucleotide pools during reverse transcription (Martinez et al., 1994; Vartanian et al., 1994). However, in each of these studies, experiments failed to generate *in vitro* hypermutants exhibiting the dinucleotide contexts of G→A change consistently observed *in vivo* or in co-cultures, namely GpG→ApG (Fitzgibbon et al., 1993) and GpA→ApA (Vartanian et al., 1991). In 2002, before the discovery of the APOBEC restriction system, Janini et al. reported a series of experiments in which they could systematically retrieve, after *ex vivo* infection of primary human peripheral blood mononuclear cells (PBMCs), HIV hypermutants bearing the same signature changes observed in primary clinical materials (Janini et al., 2001b). Moreover, they showed that the levels and contexts of hypermutation could be modulated by altering the activation status of substrate cells.

Recent studies of hypermutants retrieved from primary clinical specimens, and those obtained *ex vivo* and *in vitro*, showed that the levels of hypermutation were not evenly distributed throughout the viral genome. The levels exhibited two increasing 5′-to-3′ clines, peaking just upstream of the central and the 3′ poly-purine tracts (PPT), the so-called "twin peaks" model (Janini et al., 2001a; Suspene et al., 2006). These findings led to the description of a model for the process of reverse transcription in which varying levels of hypermutation are ascribed to the time of exposure of ssDNA (Yu et al., 2004).

The recent years have seen a significant increase in the research of APOBEC–Vif interaction, which led to a broader insight on the subject. New, though sometimes conflicting, models of regulation of APOBEC expression (Peng et al., 2006; Sarkis et al., 2006; Ying et al., 2007) and post-translational processing (Chiu et al., 2005; Kreisberg et al., 2006; Soros et al., 2007; Stopak et al., 2007; Wang et al., 2007b), packaging into nascent virions (Wang et al., 2007a; Xu et al., 2007), and mechanism of cytidine deamination during reverse transcription have been proposed (Chelico et al., 2006; Coker and Petersen-Mahrt, 2007; Iwatani et al., 2007; Li et al., 2007; Miyagi et al., 2007), backed mostly by *ex vivo* and *in vitro* experiments. Nevertheless, the correspondence of each of these models to *in vivo* hypermutation remains to be evaluated. The fact that hypermutated HIV genomes can be integrated into the host chromosome and later retrieved for examination opens a unique window into this important host restriction mechanism, setting it apart from other mechanisms where the outcome of the host–virus interaction has to be deduced by indirect markers such as infection levels. So far, most of the information regarding host restriction of HIV by APOBECs derives from *ex vivo* and *in vitro* studies (Vartanian et al., 1991) or from partial genome sequences obtained from primary clinical material (Janini et al., 2001b; Simon et al., 2005), approaches subject to the rendering of an incomplete picture.

The objective of the current work is to systematically study the levels, contexts, and patterns of HIV hypermutation *in vivo*. The analysis of 24 full-genome HIV sequences retrieved from primary PBMCs and representing infections with several HIV-1 clades, and the inclusion of 7 cognate pairs of hypermutated/non-hypermutated sequences sampled from the same patient specimens, provide a comprehensive view of the marks left by APOBEC-mediated restriction *in vivo*.

Table 1
Virtually complete hypermutated HIV-1 proviral sequences obtained from primary clinical PBMCs

Strain	Subtype	Country	Accession number[a]
99TH.C1416Hyper	B	Thailand	EF165363 (AY945711)
01TH.OUR033iHyper	B/CRF01_AE	Thailand	EF165361 (AY358071)
00TH.C1705HYP	CRF01_AE	Thailand	AY945715 (DQ789392)
99TH.C2266Hyp	C/CRF01_AE	Thailand	AY945735 (AY262830)
98US.MSC5016Hyper	C	United States	EF165359 (AY444801)
98UG57146Hyper	D	Uganda	EF165364 (AF484513)
98UG57132 Hyper	D	Uganda	EF165362 (AF484506)
00TH.R1496HYP	CRF01_AE	Thailand	AY945729
00KE_KER2018	A	Kenya	AF457057
01TH.OUR700I	CRF01_AE	Thailand	AY358058
00TH.C1468HYP	CRF01_AE	Thailand	AY945714
00TH.C4141HYP	CRF01_AE	Thailand	AY945723
02TH.OUR658I	CRF01_AE	Thailand	AY358055
01TZBD28-11	A	Tanzania	EF165365
00TH.OUR736I	CRF01_AE	Thailand	AY358054
99IL.99ET8	C	Ethiopia	AY255828
03TZ.CO783iHyper	A	Tanzania	EF165366
00KE_KNH1214	A	Kenya	AF457071
99UGC27305	A	Uganda	AF484484
03TZ.CO883P	C	Tanzania	AY734557
01TZ.BD24-8Hyper	C	Tanzania	EF165360
99KE_KSM4023-29	A	Kenya	AF457076
01TH.OUR598I	CRF01_AE	Thailand	AY358053
00KE_NKU3007	A	Kenya	AF457091

[a] Accession numbers of non-hypermutated sequences, retrieved from the same sample, are provided in parenthesis.

Results

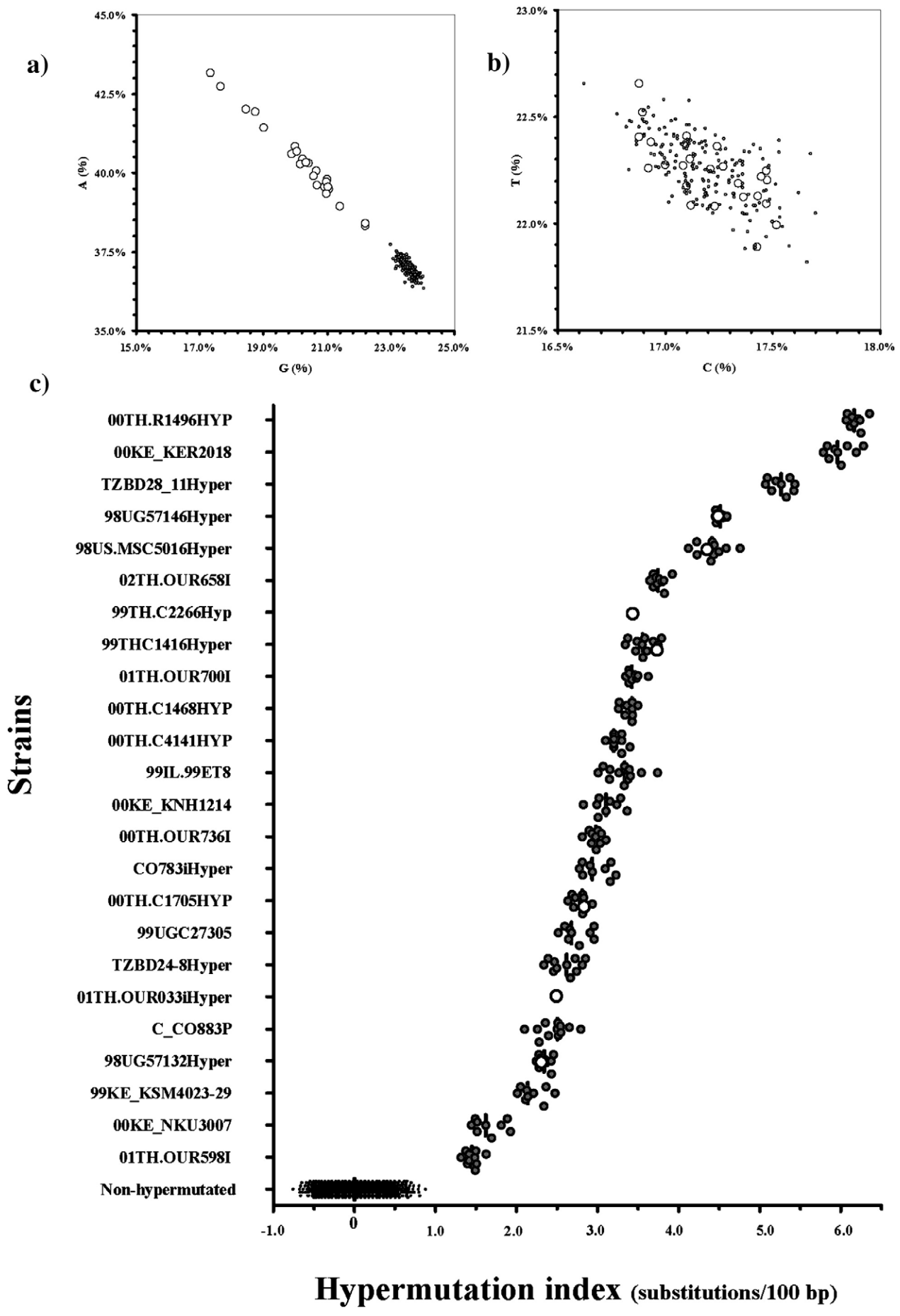
HIV-1 sequences

Twenty-four hypermutated HIV-1 full-genome sequences recovered from primary PBMCs from sero-positive patients were studied (Table 1). The hypermutants represented infections with HIV-1subtypes A, B, C, D, CRF01_AE, and unique recombinant forms, and were sampled in East Africa, Southeast Asia, and North America. Specimens 01TH.OUR598I, 01TH.OUR700I, and 00TH.OUR736I were obtained within 6 months of sero-conversion and 01TH.OUR033iHyper and 02TH.OUR658I were obtained within 1–2 years of sero-conversion (Tovanabutra et al., 2004). Dates of infection were not available for the other samples.

In seven cases, a non-hypermutated cognate sequence (Table 1, top) was retrieved from the same sample that had yielded the hypermutant. Each of these cognate pairs clustered phylogenetically; the non-hypermutants were the closest available matches to the hypermutants, differing from the hypermutants mostly in G→A changes (data not shown).

Hypermutated genomes exhibited premature stop codons in the majority of their protein coding sequences, and showed characteristically long branches in phylogenetic trees, as previously reported (Arroyo et al., 2004, 2005; Dowling et al., 2002; Harris et al., 2003, 2002; Tovanabutra et al., 2004, 2005; Watanaveeradej et al., 2006).

Fig. 1. Levels of hypermutation in HIV-1 proviral genomes. a) Adenine and guanine and b) Cytidine and thymine content in sequences from 24 hypermutated (large open circles) and 203 non-hypermutated strains (small filled circles). c) Hypermutation index (HI). Using multiple non-hypermutated references, the HI (APOBEC-mediated G→A substitutions/100 bp) was calculated for 24 hypermutants (large gray circles) and non-hypermutants (small filled circles). Open circles depict the HI for pair-wise comparisons using normal reference sequences from the same DNA sample. Bars represent the median. See text for details.


a)
45.0%
42.5%
40.0%
37.5%
35.0%
A (%)
15.0%
17.0%
19.0%
21.0%
23.0%
25.0%
G (%)
b)
23.0%
22.5%
22.0%
21.5%
T (%)
16.5%
17.0%
17.5%
18.0%
C (%)
c)
Strains
00TH.R1496HYP
00KE_KER2018
TZBD28_11Hyper
98UG57146Hyper
98US.MSC5016Hyper
02TH.OUR658I
99TH.C2266Hyp
99THC1416Hyper
01TH.OUR700I
00TH.C1468HYP
00TH.C4141HYP
99IL.99ET8
00KE_KNH1214
00TH.OUR736I
CO783iHyper
00TH.C1705HYP
99UGC27305
TZBD24-8Hyper
01TH.OUR033iHyper
C_CO883P
98UG57132Hyper
99KE_KSM4023-29
00KE_NKU3007
01TH.OUR598I
Non-hypermutated
-1.0
0
1.0
2.0
3.0
4.0
5.0
6.0
Hypermutation index (substitutions/100 bp)

Levels of hypermutation

The nucleotide composition of hypermutants was compared to that of non-hypermutated reference strains, as described in Materials and methods. The range of adenine content was significantly higher among hypermutants compared to normal reference strains (38.3–43.2% versus 36.3–37.7%, respectively; $p<0.001$) (Fig. 1a). The adenine increase was directly proportional to the decrease in guanine (slope=−0.9887; $r^2=0.9848$). Levels of adenine varied widely among the hypermutant strains, with increases of 3% to 16%, compared to reference sequences. Cytidine and thymine levels varied within a narrow range and were comparable in hypermutants and reference sequences (Fig. 1b).

Hypermutation can also be assessed by comparing aligned query and reference sequences, and determining the rates of G→A change. We devised a parameter termed the Hypermutation Index (HI), in which G→A substitutions resulting from cytidine deamination by the APOBECs are corrected for the "background" G→A mutations. In pairs of normal, non-hypermutated genomes of HIV-1 matched for genetic subtype, the rates of G→A versus A→G substitutions are equilibrated. The balance of these rates in pair-wise comparisons among normal sequences was narrowly distributed around zero (Fig. 1c, bottom left). Thus, for pairs of hypermutated and normal genomes, the observed G→A substitution rate can be corrected by subtraction of the A→G mutation rate to represent G→A substitutions attributed to APOBEC-mediated hypermutation.

Fig. 1c shows the range of HI observed. Each hypermutated sequence was evaluated using several normal reference sequences, including those from the same DNA sample when available. HI varied nearly 5-fold, ranging from 1.3 to 6.4 substitutions/100 bp. The choice of a particular normal reference sequence did not substantially impact these results; coefficients of variation for HI determined using different reference sequences of a given subtype, or sequences from the same versus a different individual, did not exceed 15% (data not shown). There were no obvious associations between HI and the genetic subtype or the geographic origin of hypermutated strains.

Contexts of hypermutation

In hypermutated sequences, the dinucleotide motifs for G→A substitution are GpG and GpA, where the underline denotes the edited base. These correspond to the reverse complements of the preferred dinucleotide substrates of cytidine deaminases APOBEC3G and APOBEC3F, respectively (Beale et al., 2004; Liddament et al., 2004). In the studied sample set the vast majority of the G→A changes occurred within these two dinucleotide contexts (data not shown). In

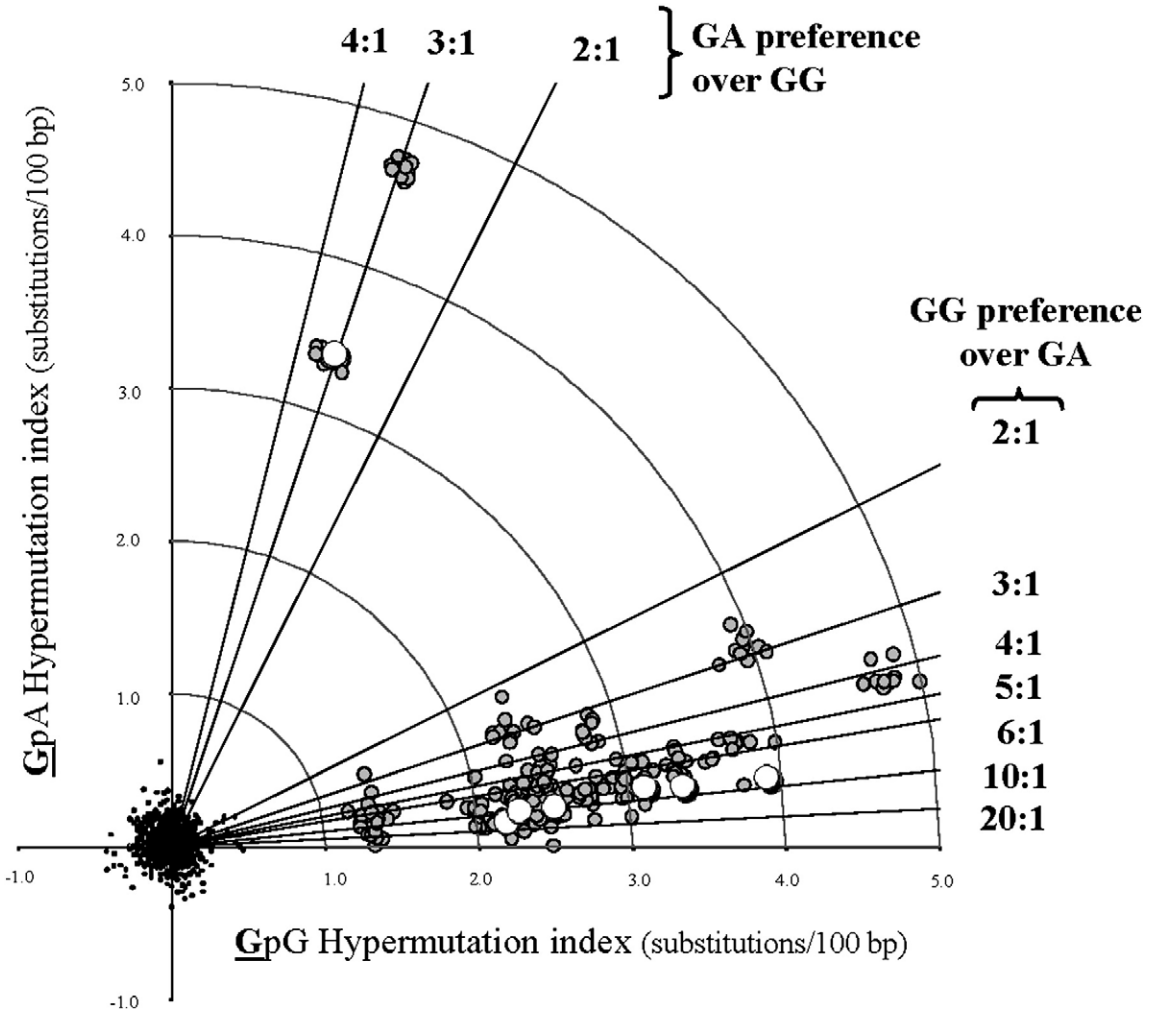


Fig. 2. Preferred context of hypermutation. Using multiple non-hypermutated references, the levels of GpG hypermutation (number of GpG→ApG substitutions−number of ApG→GpG) and GpA hypermutation (number of GpA→ApA substitutions−number of ApA→GpA substitutions) was calculated for 24 hypermutants (large gray circles) and non-hypermutants (small filled circles). Large open circles depict the GpG and GpA hypermutation values for pair-wise comparisons using cognate references. Diagonals radiating from the origin with different slopes are shown as a reference. See text for details.

order to infer the different contributions of APOBEC3G and APOBEC3F, the dinucleotide context of hypermutation was assessed by computing the number of GpG→ApG and GpA→ApA substitutions. In each case, the values were corrected for background mutation by subtracting the rates of ApG→GpG and ApA→GpA substitutions, respectively.

When performing pair-wise comparisons involving solely non-hypermutated sequences, calculated values for GpG and GpA hypermutation displayed a narrow distribution around 0.00 substitutions/bp for both contexts (95% CI = −0.02 to −0.002, and 0.008 to 0.02 substitutions/100 bp, respectively) (Fig. 2).

When analyzing hypermutants, two separate clusters were defined. Twenty-two out of 24 hypermutants had GpG as the preferred context. A wide variation in the ratio of GpG preference over GpA was observed, ranging from ratios of 2:1 to nearly 20:1 (interquartile range = 5:1 to 8:1).

The second cluster was represented by two strains (00TH. R1496HYP and 98UG57146Hyper, respectively). Both of them had a 3:1 preference of GpA over GpG, and a relatively high HI compared to the other group. In both clusters, the choice of reference did not affect the HI or the context preference, with a coefficient of variation lower than 7% for comparisons involving the same hypermutant but different references (data not shown). In each cluster, sequences with the highest hypermutation for the predominant context were also the ones exhibiting the highest HIs (data not shown). Thus, GpG was the preferred context in almost all of the hypermutant genomes in primary PBMCs, with a wide range of hypermutation levels. A preference for editing within the GpA motif was recorded, but rarely.

Previous reports, based *ex vivo* and *in vitro* experiments, have shown that the sequence context for APOBEC-mediated hypermutation extends beyond the dinucleotide level (Beale et al., 2004; Yu et al., 2004). Thus, we examined this large hypermutant dataset representing *in vivo* hypermutation for wider sequence contexts. The frequency of G→A changes in tetranucleotides of the form N_{-1}pGpGpN_{+2} (where the edited base is underlined, and N_{-1} and N_{+2} denote the bases located at positions −1 and +2 relative to the targeted G, respectively) was computed. To minimize the impact of background mutation, only the 6 samples that had a predominant GpG hypermutation context and for which a cognate non-hypermutated sequence was available (Table 1) were studied. Fig. 3 shows the abundance of each of the motifs in the non-hypermutated sequences, and their frequency of hypermutation. Some tetranucleotides were over- or under-represented, as expected from previous studies (Berkhout and van Hemert, 1994). Adenine-containing motifs (green) were the most abundant in non-hypermutated genomes, consistent with known

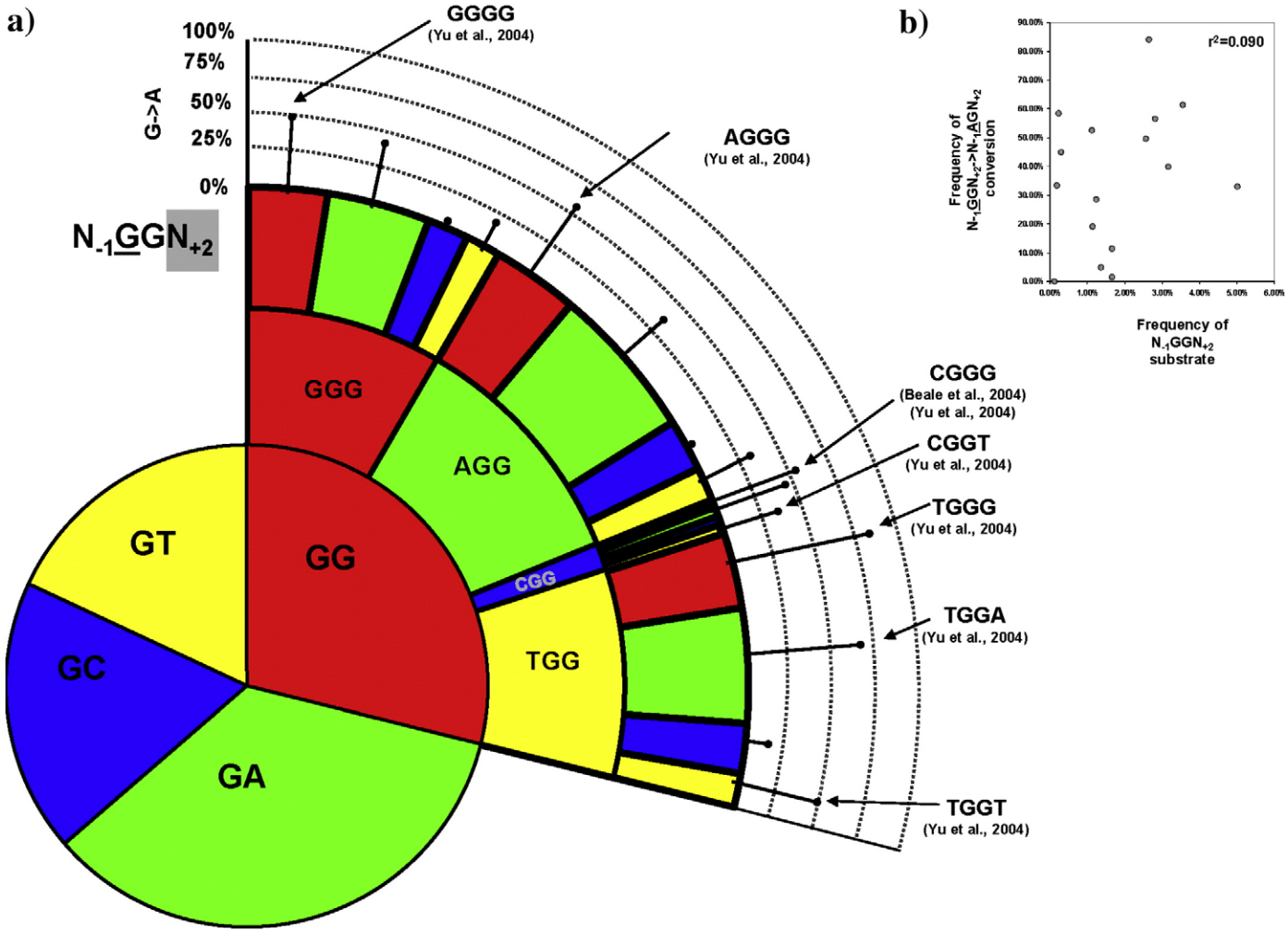


Fig. 3. Abundance and frequency of hypermutation of tetranucleotide motifs in HIV-1 proviruses. a) The colored circle represents the abundance of di-, tri-, and tetranucleotide motifs containing GpG in 195 non-hypermutated nearly-full proviral HIV-1 genomes from different subtypes and CRFs. The bars represent the median frequency of N_{-1}pGpGpN_{+2}→N_{-1}pApGpN_{+2} conversion as computed from 6 hypermutants compared to their non-hypermutated cognates. Motifs previously shown to be preferred substrates of APOBEC3G *ex vivo* and *in vitro* are shown, with their respective bibliographic reference. b) Relationship between abundance of N_{-1}pGpGpN_{+2} motifs in non-hypermutated HIV-1 proviral genomes, and N_{-1}pGpGpN_{+2}→N_{-1}pApGpN_{+2} substitution frequency in hypermutants. The resulting coefficient of correlation is shown. Data is derived from the same cognate pairs as in Fig. 3a. See text for details.

base-composition bias in HIV-1 (Berkhout and van Hemert, 1994). Tetranucleotides containing CpG (blue) were the least frequent because their cytidines tend to be methylated, leading to reduced transcriptional activity (Cooper, 2000), and are under-represented in eukaryotic genomes. We observed that the tetranucleotide contexts of *in vivo* hypermutation were consistent with, but broader than those identified *ex vivo* and *in vitro* (Beale et al., 2004; Yu et al., 2004). Among the seven tetranucleotides preferred *ex vivo* and *in vitro*, six were hypermutated at levels of 50% or greater *in vivo* (Fig. 3a, and Supplemental Table 2). However, *in vivo* hypermutation also occurred at rates between 30% and 50% in the context CpGpGpT, identified *ex vivo* and *in vitro*, and in four newly identified preferred contexts, namely GpGpGpA, CpGpGpA, ApGpGpA, and ApGpGpT. They were low but measurable rates in TpGpGpC and GpGpGpT contexts as well.

We tested the hypothesis that hypermutation had contributed to the current base-composition bias in HIV, by depleting the cytidine deaminases' preferred substrates. The frequency of each tetranucleotide motif in HIV-1 was compared to its frequency of hypermutation, but the analysis yielded no direct relationship between these two factors (see Fig. 3b).

Patterns of hypermutation

We (Janini et al., 2001a) and others (Suspene et al., 2006) have reported that the levels of hypermutation can be unevenly distributed across HIV-1 genomes. A pattern of rising and falling levels has been seen, ascribed to different durations of exposure of single-stranded DNA to APOBECs during HIV replication. Too few complete viral genomes have been analyzed to determine the consistency of this pattern, and its range of variation. We applied a newly developed software package, "HyperPack" (Kijak et al., 2007) to examine 203 normal, non-hypermutated (Supplemental Table 1) and 24 hypermutated (Table 1) HIV-1 genomes for the pattern of hypermutation across the genome.

First, the distribution of substrates for hypermutation across the genomes was analyzed (Supplemental Fig. 1). Normal genomes (n=203) were analyzed using SubstrateScan, which determined the abundance of mono- or oligonucleotide motifs in a series of overlapping, sliding windows. The G content averaged 24% and varied from 20–27% in different genome regions. The average frequency of the GpG dinucleotide across the genome was 7%, and the range of variation was 4–9%. The motif GpA averaged 8% of dinucleotides, and varied from 6–9% across the genome. Thus, the substrates of hypermutation are abundant throughout the HIV-1 genome, albeit with some small regional variations consistent among many different isolates (Supplemental Fig. 1).

The distribution of hypermutation events was estimated across the genome using HyperScan, a similar sliding-window approach that determines what fraction of a given substrate, evaluated in a normal reference strain or strains, is converted to product. We empirically determined that windows of 600 bp overlapping by 10 bp provided adequate signal-to-noise ratios (data not shown). For the HyperScan of G→A substitutions, 18 of 24 hypermutated genomes showed a consistent pattern (Fig. 4, top). Hyperscans of individual strains are shown in Supplemental Fig. 2. Substitution levels were low at the beginning of *gag*, and steadily increased until the RNAseH domain of *pol*. The rate dropped dramatically, reaching a nadir at the 3′ end of the integrase gene. Thereafter, levels again increased, mirroring the pattern of increase in the 3′ half of the genome, until the end of gp41, where they seemed to start diminishing. This pattern was reiterated in the GpG→ApG HyperScans (Fig. 4, bottom), except for the presence of two local minimums, in the middle of the RT coding region and in the C3 domain of gp120, respectively. The GpA→ApA HyperScans showed no consistent patterns among the strains analyzed (Supplemental Fig. 2).

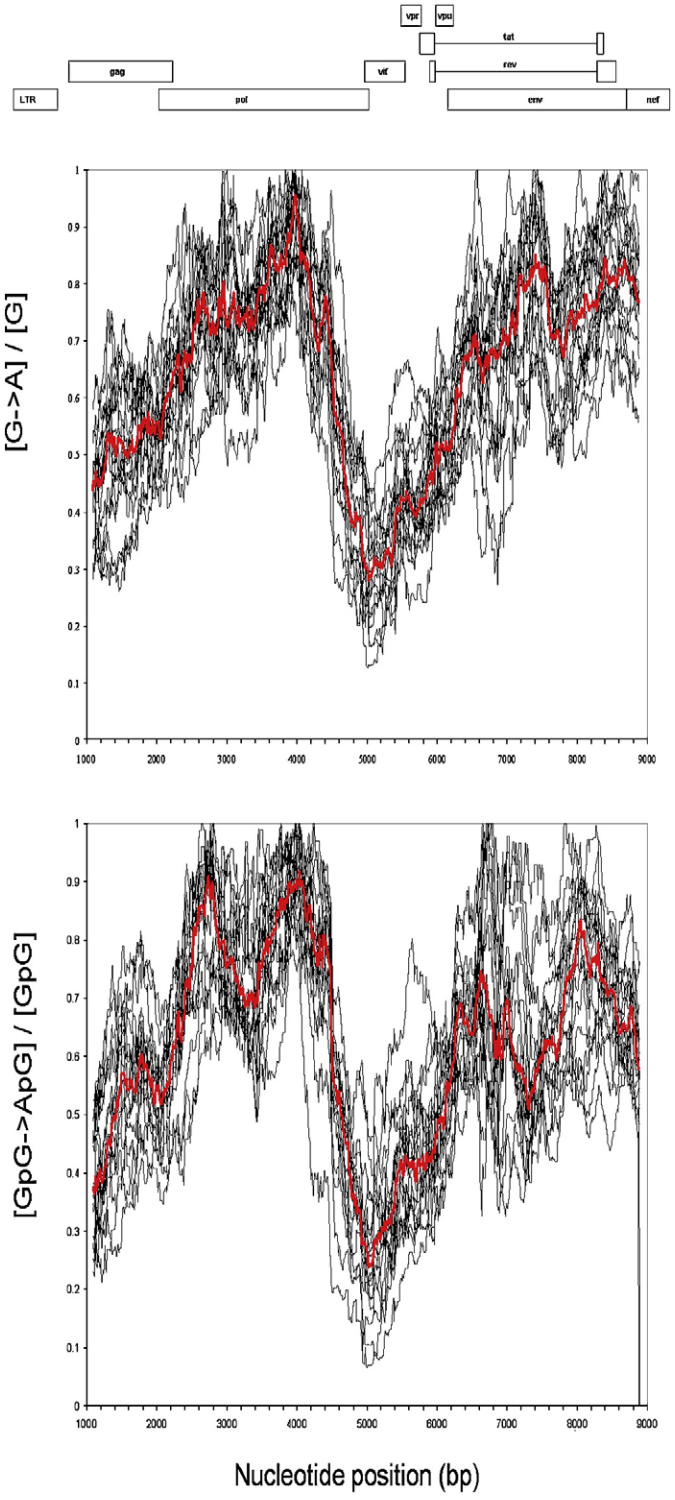


Fig. 4. Hypermutation patterns of 18 HIV-1 hypermutants. Hypermutated sequences were aligned with non-hypermutated references and a) the frequency of G→A substitution/G content and b) the frequency of GpG→ApG substitution/GpG content in sliding windows of 600 bp overlapping by 10 bp was calculated. Black lines depict the median of measurements of the same hypermutants versus at least 10 references, and the red line depicts the median between all hypermutants. In interest of clarity, the levels of hypermutation shown in the *y*-axis were normalized to the maximum level in each strain. Nucleotide positions are relative to HXB2 reference strain (GenBank accession number K03455).

Although consistent, the above-described hypermutation pattern was not universal (Fig. 5). Six strains showed atypical patterns, including the two GpA context hypermutants 98UG_57146 and

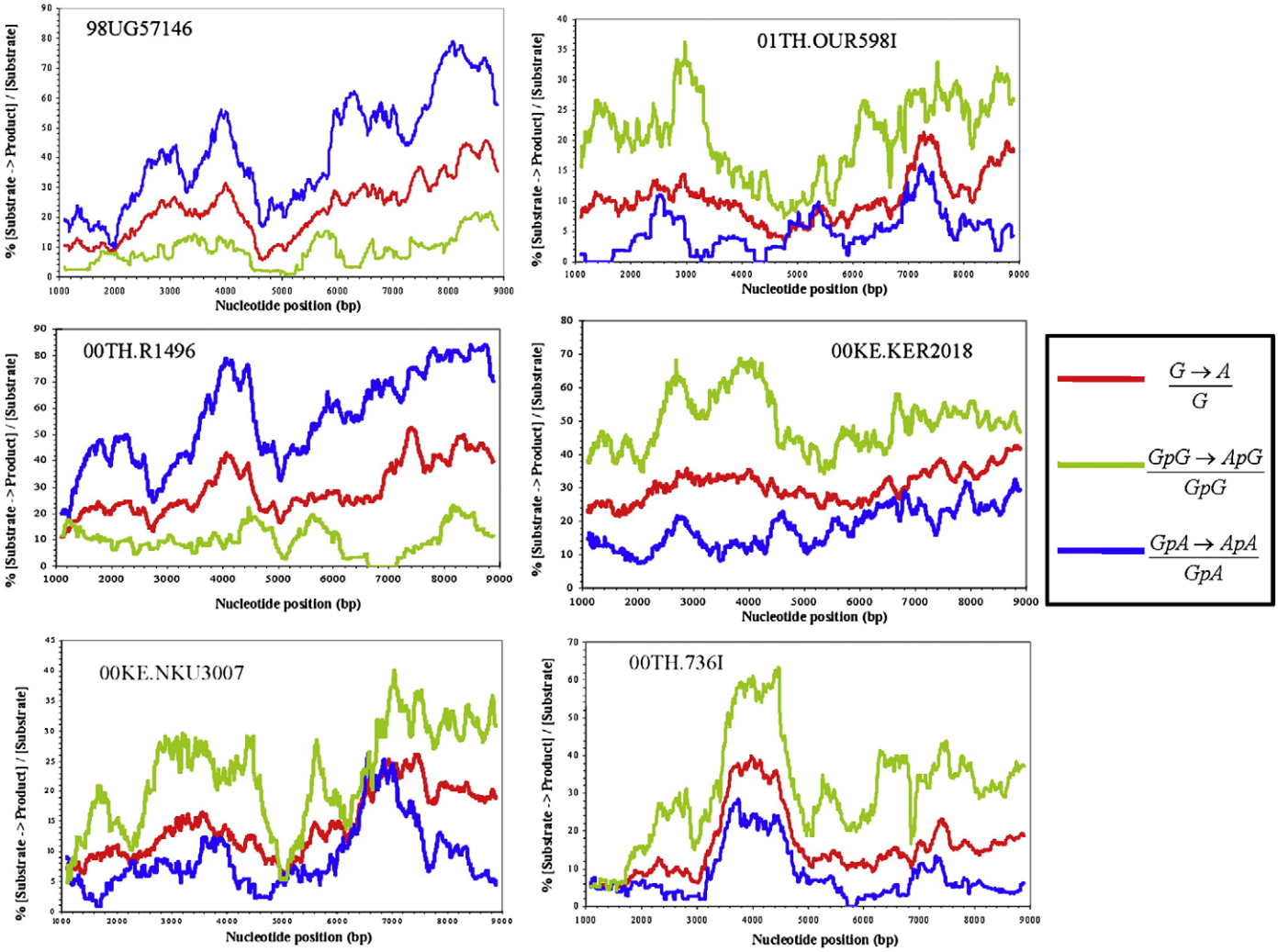


Fig. 5. Hypermutation patterns of 6 outlier HIV-1 hypermutants. Hypermutated sequences were aligned with non-hypermutated references and the frequency of G→A substitution/G content (red), the frequency of GpG→ApG substitution/GpG content (green) and the frequency of GpA→ApA substitution/GpG content (blue) in sliding windows of 600 bp overlapping by 10 bp were calculated. Each line represents the median of measurements of the same hypermutants versus at least 10 references. Nucleotide positions are relative to HXB2 reference strain (GenBank accession number K03455).

00TH_R1496, respectively. In these two strains, there was a steady increase in hypermutation levels from the 5′ to the 3′ end of the genome, with some inconsistent peaks and troughs. Three of the other outliers were the two strains with the lowest overall levels of hypermutation among those with GpG context preference, namely 00KE_NKU3007 and 01TH_OUR5981, and that with the second-to-highest level, 00KE_KER2018. In the latter strain, there was a relatively constant level of hypermutation across the genome. Finally, strain 00TH_OUR736I had low hypermutation across the genome except for a patch at the 3′ end of *pol*.

Overall, 18 of 22 strains with a preferred GpG hypermutation context conformed to the "twin peak" pattern shown in Fig. 4. Exceptionally high or low levels of GpG context hypermutation seemed associated with aberrant patterns, as did the two hypermutants with a preferred GpA context.

Evaluation of proposed models of APOBEC3G-mediated hypermutation

Recently, *in vitro* experiments performed on synthetic substrates have led to a model of APOBEC3G-mediated deamination with a series of sliding and jumping events (Chelico et al., 2006). Following random binding of the APOBEC3G to the ssDNA substrate, the enzyme would act processively 3′→5′ on the anti-sense strand, until structural constrains are found, which would lead to the jumping of the enzyme to another region along the same genome. Patches of highly hypermutated areas separated by stretches relatively devoid of hypermutation would be expected from this model.

We analyzed cognate pairs where GpG was the preferred context for the distribution of GpG sites and the local distribution of hypermutation events. The GpG sites were dense in the genome. In non-hypermutated sequences, 55% of the GpG sites were separated by 10 bp or fewer, with 40% of them being separated by 5 bp or fewer (data not shown). We selected a few regions of the genome representing different levels of hypermutation for a more detailed analysis. The hypermutated sites tended to be spread evenly throughout the area with no obvious alternation between highly-edited patches and cold-spots (Supplemental Fig. 3).

Discussion

APOBEC3G and APOBEC 3F are expressed in the lymphoid cells that are primary targets of HIV-1 infection *in vivo*, and their activity is counterbalanced by a range of viral Vif variants (Harris and Liddament, 2004). The activities of APOBEC3G and APOBEC 3F are distinguishable in part by their dinucleotide context for G→A substitution, GpG and GpA, respectively (Bishop et al., 2004). This study used an analysis of

hypermutated complete HIV-1 proviral genomes recovered from primary PBMCs to help dissect the relative contributions of different APOBECs and the range of suppression mediated by Vif *in vivo*. While previous studies reported results obtained in *ex vivo* or *in vitro* experiments, the use of primary clinical materials permitted us to preserve the imprints left *in vivo* by the APOBEC restriction system.

The main features of the APOBEC–Vif interaction identified here include: 1) levels of hypermutation varying by nearly 5-fold, ranging from 1.3 to 6.4 G→A substitutions/100 bp (Fig. 1c); 2) Preference for the GpG context (22/24 strains), implicating APOBEC3G as a major restricting element in the cells harboring HIV-1 provirus in the bloodstream of infected individuals (Fig. 2); 3) Variable bias for the GpG over the GpA context, ranging from 2:1 to nearly 20:1, which in turn suggests variation in the relative contributions of APOBEC3G and APOBEC3F in different hypermutated proviruses (Fig. 2); 4) Confirmation of the previously reported preferred tetranucleotide contexts for GpG hypermutation and identification of five additional substrates converted at 25%–50% levels; and 5) Conservation of the "twin peak" pattern of hypermutation across the genome at different levels of hypermutation and identification of non-conforming patterns, mainly at the extremes of the range of hypermutation levels in GpG context, and in both of the GpA context hypermutants identified. Overall, these results suggest a wider range of APOBEC–Vif interactions than previously appreciated, and show that examination of hypermutated genomes provides new opportunity to gain insight into parameters controlling levels, contexts, and patterns of hypermutation *in vivo*. Moreover, these results help validate some of the proposed models for APOBEC–Vif interaction.

The observed variation in levels of hypermutation may be attributed to a number of factors, probably acting in concert: different number of APOBEC molecules packaged in each viral particle, polymorphic variants of APOBEC with different activities, variation in the kinetics of reverse transcription that result in varying exposure times of ssDNA substrate. APOBEC3F and -G are known to be expressed at different levels depending on factors such as cell type (Sarkis et al., 2006), cell activation (Peng et al., 2006), or HIV disease stage (Jin et al., 2005, 2007). In turn, APOBECs are present within cells mostly in two states, an inactive high-mass molecular (HMM) form, and an active low-mass molecular form (LMM) (Chiu et al., 2005). Mitogen activation can lead to the conversion from the LMM to the HMM form, and is associated with rendering the activated cells permissive to HIV infection (Stopak et al., 2007). Altogether, this expression and post-translational regulation of APOBECs can result on variable availability of these molecules for incorporation into the nascent viral particle. Xu et al. have shown that delta-*vif* viruses produced in culture carry in average between 3 and 11 APOBEC3G molecules per virion (Xu et al., 2007). Thus, the nearly 5-fold variation of hypermutation levels observed *in vivo* is generally consistent with a 4-fold variation of APOBEC content in viral particles. Moreover, the small number of APOBEC particles incorporated per virion would tend to make the distribution of HI values behave as a discrete variable, which is also apparent from our results. For instance, in our dataset 17/24 hypermutants had HI that concentrated in the range of 2–3 substitutions/100 bp, probably reflecting the mode of the distribution of number of APOBEC molecules per virion. Also, aberrant pattern of genome-wide hypermutation observed at the extremes of HI distribution may reflect the inclusion of a high number of APOBEC molecules, leading to saturation of G→A substitution (strain 00KE.KER2018), or incorporation of a very low number of APOBEC molecules in the viral particle accompanied by low overall HI (strains 00KE.NKU3007 and 01TH.OUR598I). In the latter, and following the model proposed by Soros et al., by which virion-incorporated APOBEC3G is associated with HIV-1 genomic RNA but remains latent until its activity is restored during reverse transcription by action of HIV RNase H (Soros et al., 2007), incorporation of a very limited number of APOBEC molecules will be manifested as genomes bearing significant marks of hypermutation in few areas. Polymorphic variants of APOBEC may also be responsible of varying levels of hypermutation, but would not necessarily lead to the observed discrete distribution of HI values.

In the current analysis, we were able to retrieve from primary PBMCs strains predominantly carrying the GpG motif. Janini et al. have shown in *ex vivo* experiments that, by modifying cell activation around the time of infection, the context of hypermutation can be modulated (Janini et al., 2001b). According to their model, where GpG was the predominant context of hypermutation in resting PBMCs but the GpA context predominated in the genomes recovered after activation, our results would seem to indicate that most of the hypermutated proviruses that we retrieved might have derived from viral particles produced in resting PBMCs. If that was the case, APOBEC3G would be mediating the protection of resting PBMCs, and APOBEC3F would be the restriction system acting in activated cells. The underlying reasons for these phenomena are probably related to different stimuli and mediators of the recruitment of LMM into HMM form for each of the APOBECs, an important subject that remains to be elucidated.

In our results it was evident that in each studied sequence there was a clear predominance of one of the dinucleotide contexts over the other, which was concordant with previously reported *in vivo* data based on partial HIV genome sequences (Janini et al., 2001b; Simon et al., 2005). The preference of APOBEC3G was estimated *in vitro* to be 2.5:1 for GpG versus GpA (Suspene et al., 2004), but the ratio of context preference for APOBEC3F has not been yet systematically addressed. In the sequences analyzed here exhibiting a predominant GpG hypermutation context, the ratio of motif preference ranged from 2:1 to 20:1 in favor of GpG. Among the explanations that can be proposed for this distribution is the occurrence of APOBEC3G in human populations as allelic forms with different dinucleotide or larger-context substrate specificity, or the concurrent involvement of other members of APOBEC family, as homo- or heterodimers. Only a limited number of representatives of the GpA-preferred motif were detected in the current sample set, and their ratio of context preference was 3:1 in favor of GpA. Interestingly, these two sequences had also HI levels that were among the highest in the sample set, suggesting that, while not a prevalent mechanism of restriction, APOBEC3F hypermutation can be a very effective one.

In aggregate, the presence of two clearly separate clusters based on dinucleotide context preference would support the hypothesis that APOBEC3F and -G may not be naturally co-packaged into nascent virions, at least in equivalent proportions for their respective active forms. Our results contrast with the *ex vivo* experiments of Liddament et al. (2004), that showed that both enzymes are co-expressed and that they can act in concert in controlled experimental settings. Still, in the latter, the burden of high enzyme concentration resulting from *ex vivo* over-expression may have distorted the natural balance of host restriction and viral defense mechanisms, and could have surpassed the Vif-mediated APOBEC-exclusion system. In addition, the two enzymes seem to have different sensitivities to Vif-mediated exclusion (Liddament et al., 2004). It will be important to conduct experiments that can clarify the relationship between natural expression and regulation of the different members of the APOBEC family and anti-retroviral restriction.

Ex vivo and *in vitro* experiments have shown that the context preference of APOBECs extends beyond the dinucleotide level, and tetranucleotide motifs have been identified, facilitated by the knowledge of the sequence of the initial infecting strain in the controlled experimental setting. Deducing preferred tetranucleotide motifs from hypermutated sequences retrieved from primary clinical materials can be hindered by the absence of closely related reference sequences. Our results from the comparison of cognate sequence pairs from the same sample were consistent with *ex vivo* and *in vitro* data (Beale et al., 2004; Yu et al., 2004), reiterating the avoidance of APOBEC3G

from editing NpGpGpC motifs. Beale et al. showed that the optimal substrate for APOBEC3G *in vitro* was CGGG (Beale et al., 2004). We showed that this motif was very infrequent in HIV genomes but, when present, it was affected in more than 50% of its occurrences. We were also able to identify five additional substrates converted at significant levels, which indicates that the available sites for APOBEC-mediated hypermutation are extensively present throughout the viral genome. Due to the small number of sequences exhibiting the GpA motif, we could not clearly establish an extended context for APOBEC3F.

Even before the recognition of APOBECs as the elements responsible for HIV hypermutation, it was described that G→A editing did not follow an erratic configuration (Janini et al., 2001a; Suspene et al., 2006). A pattern that consisted in two ascending gradients in 5′→3′ direction, peaking at the central-PPT and 3′-PPT tracts, respectively, was defined in several HIV hypermutants. This pattern was later ascribed to a replication model with two initiation sites for the synthesis of the sense DNA strand during reverse transcription (Suspene et al., 2006; Yu et al., 2004). According to this model, uneven levels of hypermutation along the genome were due to the fact that different genomic areas in the anti-sense strand remain as ssDNA for different periods of time, as only ssDNA and neither ssRNA, RNA/DNA heteroduplexes, nor dsDNA are subject to APOBEC-mediated deamination (Suspene et al., 2004). While most of the genomes analyzed would support this model, we did observe several exceptions, mostly at the low and high extremes of the HI and in the GpA preference hypermutants. Additional complexities of the APOBEC–Vif interaction, stoichiometry of APOBEC incorporation into the viral particle (Xu et al., 2007), the dynamics of reverse transcription and concurrent degradation of the RNA template by the viral RNase H, which in turn is intimately related to the activation of APOBECs (Soros, Yonemoto, and Greene, 2007), may underlie these observations.

In the current study, we evidenced hypermutation as a phenomenon that occurs in multiple viral genetic subtypes and inter-subtype recombinant forms, and in patients from diverse geographic areas. No association was evident between any of the abovementioned factors with the levels, contexts, or patterns of hypermutation. However, given the extent of inter-clade variation in the *vif* gene and the limited number of representatives of each clade included in the current study, we cannot exclude the possibility that there are parameters of hypermutation that vary in different populations and geographic regions. Both the viral defense protein Vif (Pace et al., 2006) and human APOBECs (Do et al., 2005; Valcke et al., 2006) are genetically variable. Some forms of the Vif protein may be more efficient than others when preventing the incorporation of APOBEC molecules into the nascent virion. In the same vein, some allelic variants of APOBECs may be less susceptible to Vif. While polymorphism H186R in APOBEC3G was associated with progression to AIDS in North America and Western Europe (An et al., 2004), the molecular basis of this phenomenon remains to be revealed, and the crucial interaction of APOBEC and Vif warrants further studies on co-variation of these genes and the potential for variable strategies of host restriction reflected by different levels and patterns of hypermutation, in different populations.

The APOBEC–Vif interaction is the only known host–virus restriction system in HIV-1 biology where an imprint of outcome of the event can be retrieved from infected cells. This added dimension warrants continuing investigating HIV hypermutation for the information it may provide leading to the manipulation of this interaction in favor of the host.

Materials and methods

Studied hypermutated sequences

Twenty-four hypermutated HIV type 1 (HIV-1) virtually full-length genome sequences were analyzed, each of them obtained from a different patient. All of the sequences were retrieved from DNA extracted from primary PBMCs and without co-culture. Specimens were sampled in the setting of HIV-1 global molecular epidemiology studies. All of these sequences were retrieved through the same standard protocol (see below), and they were recognized as hypermutants immediately after obtainment, due to their extensive G→A substitution levels. In 7 cases, cognate pairs of hypermutants/non-hypermutated sequences retrieved from the same DNA sample could be obtained, and these were used for analyses that required closely matched hypermutated and normal reference sequences. Official sequence names, HIV-1 clade assignment, country of sampling, and accession numbers are indicated in Table 1. Sixteen of these hypermutated sequences were reported previously (Arroyo et al., 2004, 2005; Dowling et al., 2002; Harris et al., 2003, 2002; Tovanabutra et al., 2004, 2005; Watanaveeradej et al., 2006) but had not been analyzed in detail. No information regarding anti-retroviral treatment history, plasma viral load, or CD4 cell counts was available.

Normal reference sequences

A total of 203 non-hypermutated full-genome HIV-1 sequences retrieved from the Los Alamos Database (http://www.hiv.lanl.gov) were used as normal references for analysis (Supplemental Table 1). Sequences included subtypes A ($n=39$), B ($n=11$), C ($n=69$), D ($n=30$), and CRF01_AE ($n=54$), and their proportions were balanced to reflect the clade distribution in the hypermutant dataset.

Levels and contexts of hypermutation

Nucleotide base frequencies, substitution frequencies, and contexts of hypermutation were determined using the *Bases*, *Substitutions*, and *Context* modules, respectively, implemented in HyperPack (Kijak et al., 2007). For each hypermutant, levels and context of hypermutation were assessed through pair-wise comparisons in optimized alignments with at least 10 reference sequences matched for HIV subtype, to control for within-clade genetic variation. Unique recombinant forms (URFs) were compared only to their respective cognate reference.

Levels of hypermutation were determined through the hypermutation index (HI), defined as:

$$\mathrm{HI} = \frac{\text{G}\rightarrow\text{A substitutions(bp)} - \text{A}\rightarrow\text{G substitutions(bp)}}{\text{sequence length(bp)}}$$

In the HI, the level of A→G substitution is used to correct for background mutation.

Dinucleotide contexts of hypermutation were determined through the comparison of context-dependent HIs:

$$\mathrm{HI}_{\mathrm{GpG}} = \frac{\text{GpG}\rightarrow\text{ApG substitutions(bp)} - \text{ApG}\rightarrow\text{GpG substitutions(bp)}}{\text{sequence length(bp)}}$$

and

$$\mathrm{HI}_{\mathrm{GpA}} = \frac{\text{GpA}\rightarrow\text{ApA substitutions(bp)} - \text{ApA}\rightarrow\text{GpA substitutions(bp)}}{\text{sequence length(bp)}}$$

where the levels of ApG→GpG and ApA→GpA substitution are used to correct for the HIV-1 RT contribution (background mutation) to sequence variation.

The occurrence of tetranucleotide motifs of the form N_{-1}pGpGN_{+2} in normal sequences, where N_{-1} and N_{+2} denote the bases located at positions −1 and +2 relative to the targeted G, respectively, and their corresponding frequencies of hypermutation to N_{-1}pApGN_{+2} were assessed. The investigation of wider contexts of hypermutation is more sensitive to the effects of background genetic variation, so only cognate pairs, which represent the closest matches of query and reference sequences, were used.

Patterns of hypermutation

The distribution of substrate motifs for hypermutation across the HIV-1 genome, and their rates of APOBEC-mediated G→A conversion, were determined using the SubstrateScan and HyperScan modules, implemented in the HyperPack as previously described (Kijak et al., 2007). Briefly, optimized pair-wise alignments of a query and a reference sequence are parsed into overlapping sliding windows, and the occurrence of substrates (in SubstrateScan) and their rates of hypermutation (in HyperScan) are computed. Each hypermutant was separately compared to at least 10 reference sequences, matched for HIV subtype, to control for within-clade genetic variation, except from URFs, which were only compared to their respective cognate reference.

HIV-1 full-genome amplification, sequencing and phylogenetic analysis

Sequences were retrieved from DNA extracted from primary PBMCs, using the same standard full-genome amplification method by nested PCR by which non-hypermutated isolates are retrieved, and as previously reported (Carr et al., 1999; Salminen et al., 1995b). Nucleotide sequencing was carried out with big dye terminators using an ABI 3100 capillary sequencer (Applied Biosystems Inc., Foster City, CA), as described. Sequences were analyzed, edited, and assembled with Sequencher 3.0 (Ann Arbor, MI). Phylogenetic analysis was performed using the SEQBOOT, DNADIST, NEIGHBOR, CONSENSE modules of the Phylip software package (distributed by Joseph Felsenstein, Department of Genetics, University of Washington, Seattle [ftp://evolution.genetics.washington.edu/pub/phylip/]), and with TREETOOL (provided by Maciukenas S: Treetool, 2.0.2. Ribosomal DNA Database Project, University of Illinois Board of Trustees, Urbana, IL). Bootscanning (Salminen et al., 1995a) and sub-genomic neighbor joining trees were used to characterize recombinant strains.

Statistical analysis

The analyses of the data distribution, Mann–Whitney *U* test, and linear regression were performed using Prism 4.0 (GraphPad Software).

Acknowledgments

This work was supported through a cooperative agreement between the Henry M. Jackson Foundation for the Advancement of Military Medicine and the U.S. Department of Defense, by the National Institute for Allergy and Infectious Diseases, National Institutes of Health. The views and opinions expressed herein do not necessarily reflect those of the U.S. Army or the Department of Defense.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2008.03.017.

References

An, P., Bleiber, G., Duggal, P., Nelson, G., May, M., Mangeat, B., Alobwede, I., Trono, D., Vlahov, D., Donfield, S., Goedert, J.J., Phair, J., Buchbinder, S., O'Brien, S.J., Telenti, A., Winkler, C.A., 2004. APOBEC3G genetic variants and their influence on the progression to AIDS. J. Virol. 78 (20), 11070–11076.

Arroyo, M.A., Hoelscher, M., Sanders-Buell, E., Herbinger, K.H., Samky, E., Maboko, L., Hoffmann, O., Robb, M.R., Birx, D.L., McCutchan, F.E., 2004. HIV type 1 subtypes among blood donors in the Mbeya region of southwest Tanzania. AIDS Res. Hum. Retrovir. 20 (8), 895–901.

Arroyo, M.A., Hoelscher, M., Sateren, W., Samky, E., Maboko, L., Hoffmann, O., Kijak, G., Robb, M., Birx, D.L., McCutchan, F.E., 2005. HIV-1 diversity and prevalence differ between urban and rural areas in the Mbeya region of Tanzania. Aids 19 (14), 1517–1524.

Beale, R.C., Petersen-Mahrt, S.K., Watt, I.N., Harris, R.S., Rada, C., Neuberger, M.S., 2004. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra *in vivo*. J. Mol. Biol. 337 (3), 585–596.

Berkhout, B., van Hemert, F.J., 1994. The unusual nucleotide content of the HIV RNA genome results in a biased amino acid composition of HIV proteins. Nucleic Acids Res. 22 (9), 1705–1711.

Bishop, K.N., Holmes, R.K., Sheehy, A.M., Davidson, N.O., Cho, S.J., Malim, M.H., 2004. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr. Biol. 14 (15), 1392–1396.

Braaten, D., Luban, J., 2001. Cyclophilin A regulates HIV-1 infectivity, as demonstrated by gene targeting in human T cells. EMBO J. 20 (6), 1300–1309.

Brenchley, J.M., Schacker, T.W., Ruff, L.E., Price, D.A., Taylor, J.H., Beilman, G.J., Nguyen, P.L., Khoruts, A., Larson, M., Haase, A.T., Douek, D.C., 2004. CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. J. Exp. Med. 200 (6), 749–759.

Carr, J.K., Laukkanen, T., Salminen, M.O., Albert, J., Alaeus, A., Kim, B., Sanders-Buell, E., Birx, D.L., McCutchan, F.E., 1999. Characterization of subtype A HIV-1 from Africa by full genome sequencing. Aids 13 (14), 1819–1826.

Chelico, L., Pham, P., Calabrese, P., Goodman, M.F., 2006. APOBEC3G DNA deaminase acts processively 3′ → 5′ on single-stranded DNA. Nat. Struct. Mol. Biol. 13 (5), 392–399.

Chiu, Y.L., Soros, V.B., Kreisberg, J.F., Stopak, K., Yonemoto, W., Greene, W.C., 2005. Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. Nature 435 (7038), 108–114.

Coker, H.A., Petersen-Mahrt, S.K., 2007. The nuclear DNA deaminase AID functions distributively whereas cytoplasmic APOBEC3G has a processive mode of action. DNA Repair (Amst.) 6 (2), 235–243.

Connor, R.I., Sheridan, K.E., Ceradini, D., Choe, S., Landau, N.R., 1997. Change in coreceptor use coreceptor use correlates with disease progression in HIV-1-infected individuals. J. Exp. Med. 185 (4), 621–628.

Cooper, G.M., 2000. The Cell, a Molecular Approach. ASM Press.

Diamond, T.L., Kimata, J., Kim, B., 2001. Identification of a simian immunodeficiency virus reverse transcriptase variant with enhanced replicational fidelity in the late stage of viral infection. J. Biol. Chem. 276 (26), 23624–23631.

Do, H., Vasilescu, A., Diop, G., Hirtzig, T., Heath, S.C., Coulonges, C., Rappaport, J., Therwath, A., Lathrop, M., Matsuda, F., Zagury, J.F., 2005. Exhaustive genotyping of the CEM15 (APOBEC3G) gene and absence of association with AIDS progression in a French cohort. J. Infect. Dis. 191 (2), 159–163.

Dowling, W.E., Kim, B., Mason, C.J., Wasunna, K.M., Alam, U., Elson, L., Birx, D.L., Robb, M.L., McCutchan, F.E., Carr, J.K., 2002. Forty-one near full-length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A-containing recombinants. Aids 16 (13), 1809–1820.

Fitzgibbon, J.E., Mazar, S., Dubin, D.T., 1993. A new type of G→A hypermutation affecting human immunodeficiency virus. AIDS Res. Hum. Retrovir. 9 (9), 833–838.

Goff, S.P., 2004. Retrovirus restriction factors. Mol. Cell 16 (6), 849–859.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J., Freedman, B.I., Quinones, M.P., Bamshad, M.J., Murthy, K.K., Rovin, B.H., Bradley, W., Clark, R.A., Anderson, S.A., O'Connell, R.J., Agan, B.K., Ahuja, S.S., Bologna, R., Sen, L., Dolan, M.J., Ahuja, S.K., 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307 (5714), 1434–1440.

Graziosi, C., Soudeyns, H., Rizzardi, G.P., Bart, P.A., Chapuis, A., Pantaleo, G., 1998. Immunopathogenesis of HIV infection. AIDS Res. Hum. Retrovir. 14 (Suppl 2), S135–S142.

Harris, M.E., Serwadda, D., Sewankambo, N., Kim, B., Kigozi, G., Kiwanuka, N., Phillips, J.B., Wabwire, F., Meehen, M., Lutalo, T., Lane, J.R., Merling, R., Gray, R., Wawer, M., Birx, D.L., Robb, M.L., McCutchan, F.E., 2002. Among 46 near full length HIV type 1 genome sequences from Rakai District, Uganda, subtype D and AD recombinants predominate. AIDS Res. Hum. Retrovir. 18 (17), 1281–1290.

Harris, M.E., Maayan, S., Kim, B., Zeira, M., Ferrari, G., Birx, D.L., McCutchan, F.E., 2003. A cluster of HIV type 1 subtype C sequences from Ethiopia, observed in full genome analysis, is not sustained in subgenomic regions. AIDS Res. Hum. Retrovir. 19 (12), 1125–1133.

Harris, R.S., Liddament, M.T., 2004. Retroviral restriction by APOBEC proteins. Nat. Rev., Immunol. 4 (11), 868–877.

Iwatani, Y., Chan, D.S., Wang, F., Maynard, K.S., Sugiura, W., Gronenborn, A.M., Rouzina, I., Williams, M.C., Musier-Forsyth, K., Levin, J.G., 2007. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic Acids Res. 35 (21), 7096–7108.

Janini, M., Carr, J., Birx, D., McCutchan, F., 2001a. AIDS Vaccine. Philadelphia, PA, USA.

Janini, M., Rogers, M., Birx, D.R., McCutchan, F.E., 2001b. Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells. J. Virol. 75 (17), 7973–7986.

Jin, X., Brooks, A., Chen, H., Bennett, R., Reichman, R., Smith, H., 2005. APOBEC3G/CEM15 (hA3G) mRNA levels associate inversely with human immunodeficiency virus viremia. J. Virol. 79 (17), 11513–11516.

Jin, X., Wu, H., Smith, H., 2007. APOBEC3G levels predict rates of progression to AIDS. Retrovirology 4, 20.

Kijak, G.H., Janini, M., Tovanabutra, S., Sanders-Buell, E.E., Birx, D.L., Robb, M.L., Michael, N.L., McCutchan, F.E., 2007. HyperPack: a software package for the study of levels, contexts, and patterns of APOBEC-mediated hypermutation in HIV. AIDS Res. Hum. Retrovir. 23 (4), 554–557.

Kreisberg, J.F., Yonemoto, W., Greene, W.C., 2006. Endogenous factors enhance HIV infection of tissue naive CD4 T cells by stimulating high molecular mass APOBEC3G complex formation. J. Exp. Med. 203 (4), 865–870.

Lecossier, D., Bouchonnet, F., Clavel, F., Hance, A.J., 2003. Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science 300 (5622), 1112.

Liddament, M.T., Brown, W.L., Schumacher, A.J., Harris, R.S., 2004. APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 *in vivo*. Curr. Biol. 14 (15), 1385–1391.

Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A., Landau, N.R., 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 86 (3), 367–377.

Li, X.Y., Guo, F., Zhang, L., Kleiman, L., Cen, S., 2007. APOBEC3G inhibits DNA strand transfer during HIV-1 reverse transcription. J. Biol. Chem. 282 (44), 32065–32074.

Martinez, M.A., Vartanian, J.P., Wain-Hobson, S., 1994. Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. Proc. Natl. Acad. Sci. U. S. A. 91 (25), 11787–11791.

Mehandru, S., Poles, M.A., Tenner-Racz, K., Horowitz, A., Hurley, A., Hogan, C., Boden, D., Racz, P., Markowitz, M., 2004. Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. J. Exp. Med. 200 (6), 761–770.

Mehle, A., Strack, B., Ancuta, P., Zhang, C., McPike, M., Gabuzda, D., 2004. Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. J. Biol. Chem. 279 (9), 7792–7798.

Miyagi, E., Opi, S., Takeuchi, H., Khan, M., Goila-Gaur, R., Kao, S., Strebel, K., 2007. Enzymatically active APOBEC3G is required for efficient inhibition of HIV-1. J. Virol. 81 (24), 13346–13353.

Overbaugh, J., Jackson, S.M., Papenhausen, M.D., Rudensey, L.M., 1996. Lentiviral genomes with G-to-A hypermutation may result from Taq polymerase errors during polymerase chain reaction. AIDS Res. Hum. Retrovir. 12 (17), 1605–1613.

Pace, C., Keller, J., Nolan, D., James, I., Gaudieri, S., Moore, C., Mallal, S., 2006. Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and *vif* genetic variation. J. Virol. 80 (18), 9259–9269.

Pathak, V.K., Temin, H.M., 1990. 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. U. S. A. 87 (16), 6019–6023.

Peng, G., Lei, K.J., Jin, W., Greenwell-Wild, T., Wahl, S.M., 2006. Induction of APOBEC3 family proteins, a defensive maneuver underlying interferon-induced anti-HIV-1 activity. J. Exp. Med. 203 (1), 41–46.

Rose, K.M., Marin, M., Kozak, S.L., Kabat, D., 2005. Regulated production and anti-HIV type 1 activities of cytidine deaminases APOBEC3B, 3F, and 3G. AIDS Res. Hum. Retrovir. 21 (7), 611–619.

Salminen, M.O., Carr, J.K., Burke, D.S., McCutchan, F.E., 1995a. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res. Hum. Retrovir. 11 (11), 1423–1425.

Salminen, M.O., Koch, C., Sanders-Buell, E., Ehrenberg, P.K., Michael, N.L., Carr, J.K., Burke, D.S., McCutchan, F.E., 1995b. Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. Virology 213 (1), 80–86.

Samson, M., Libert, F., Doranz, B.J., Rucker, J., Liesnard, C., Farber, C.M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C., Muyldermans, G., Verhofstede, C., Burtonboy, G., Georges, M., Imai, T., Rana, S., Yi, Y., Smyth, R.J., Collman, R.G., Doms, R.W., Vassart, G., Parmentier, M., 1996. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382 (6593), 722–725.

Sarkis, P.T., Ying, S., Xu, R., Yu, X.F., 2006. STAT1-independent cell type-specific regulation of antiviral APOBEC3G by IFN-alpha. J. Immunol. 177 (7), 4530–4540.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., Malim, M.H., 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418 (6898), 646–650.

Simon, V., Zennou, V., Murray, D., Huang, Y., Ho, D.D., Bieniasz, P.D., 2005. Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. PLoS Pathog. 1 (1), e6.

Soros, V.B., Yonemoto, W., Greene, W.C., 2007. Newly synthesized APOBEC3G is incorporated into HIV virions, inhibited by HIV RNA, and subsequently activated by RNase H. PLoS Pathog. 3 (2), e15.

Stopak, K.S., Chiu, Y.L., Kropp, J., Grant, R.M., Greene, W.C., 2007. Distinct patterns of cytokine regulation of APOBEC3G expression and activity in primary lymphocytes, macrophages, and dendritic cells. J. Biol. Chem. 282 (6), 3539–3546.

Stremlau, M., Owens, C.M., Perron, M.J., Kiessling, M., Autissier, P., Sodroski, J., 2004. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature 427 (6977), 848–853.

Suspene, R., Sommer, P., Henry, M., Ferris, S., Guetard, D., Pochet, S., Chester, A., Navaratnam, N., Wain-Hobson, S., Vartanian, J.P., 2004. APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. Nucleic Acids Res. 32 (8), 2421–2429.

Suspene, R., Rusniok, C., Vartanian, J.P., Wain-Hobson, S., 2006. Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. Nucleic Acids Res. 34 (17), 4677–4684.

Tovanabutra, S., Beyrer, C., Sakkhachornphop, S., Razak, M.H., Ramos, G.L., Vongchak, T., Rungruengthanakit, K., Saokhieo, P., Tejafong, K., Kim, B., De Souza, M., Robb, M.L., Birx, D.L., Jittiwutikarn, J., Suriyanon, V., Celentano, D.D., McCutchan, F.E., 2004. The changing molecular epidemiology of HIV type 1 among northern Thai drug users, 1999 to 2002. AIDS Res. Hum. Retrovir. 20 (5), 465–475.

Tovanabutra, S., Brodine, S.K., Mascola, J.R., Sankale, J.L., Sanders-Buell, E., Kim, B., Birx, D.L., McCutchan, F.E., 2005. Characterization of complete HIV type 1 genomes from non-B subtype infections in U.S. military personnel. AIDS Res. Hum. Retrovir. 21 (5), 424–429.

Valcke, H.S., Bernard, N.F., Bruneau, J., Alary, M., Tsoukas, C.M., Roger, M., 2006. APOBEC3G genetic variants and their association with risk of HIV infection in highly exposed Caucasians. Aids 20 (15), 1984–1986.

Vartanian, J.P., Meyerhans, A., Asjo, B., Wain-Hobson, S., 1991. Selection, recombination, and G——A hypermutation of human immunodeficiency virus type 1 genomes. J. Virol. 65 (4), 1779–1788.

Vartanian, J.P., Meyerhans, A., Sala, M., Wain-Hobson, S., 1994. G→A hypermutation of the human immunodeficiency virus type 1 genome: evidence for dCTP pool imbalance during reverse transcription. Proc. Natl. Acad. Sci. U. S. A. 91 (8), 3092–3096.

Wang, T., Tian, C., Zhang, W., Luo, K., Thi, P., Sarkis, N., Yu, L., Liu, B., Yu, Y., Yu, X.F., 2007a. 7SL RNA mediates virion packaging of the antiviral cytidine deaminase APOBEC3G. J. Virol. 81 (23), 13112–13124.

Wang, X., Dolan, P.T., Dang, Y., Zheng, Y.H., 2007b. Biochemical differentiation of APOBEC3F and APOBEC3G proteins associated with HIV-1 life cycle. J. Biol. Chem. 282 (3), 1585–1594.

Watanaveeradej, V., Benenson, M.W., Souza, M.D., Sirisopana, N., Nitayaphan, S., Tontichaivanich, C., Amphaipit, R., Renzullo, P.O., Brown, A.E., McNeil, J.G., Robb, M.L., Birx, D.L., Tovanabutra, S., Carr, J.K., McCutchan, F.E., 2006. Molecular epidemiology of HIV Type 1 in preparation for a Phase III prime-boost vaccine trial in Thailand and a new approach to HIV Type 1 genotyping. AIDS Res. Hum. Retrovir. 22 (8), 801–807.

Xu, H., Chertova, E., Chen, J., Ott, D.E., Roser, J.D., Hu, W.S., Pathak, V.K., 2007. Stoichiometry of the antiviral protein APOBEC3G in HIV-1 virions. Virology 360 (2), 247–256.

Ying, S., Zhang, X., Sarkis, P.T., Xu, R., Yu, X., 2007. Cell-specific regulation of APOBEC3F by interferons. Acta Biochim. Biophys. Sin. (Shanghai) 39 (4), 297–304.

Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J.M., Landau, N.R., 2004. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat. Struct. Mol. Biol. 11 (5), 435–442.

Zheng, Y.H., Irwin, D., Kurosu, T., Tokunaga, K., Sata, T., Peterlin, B.M., 2004. Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. J. Virol. 78 (11), 6073–6076.