Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication

Rodolphe Suspène, Christophe Rusniok[1], Jean-Pierre Vartanian and Simon Wain-Hobson*

Unité de Rétrovirologie Moléculaire, CNRS URA1930 and [1]Unité de Génomique des Microorganismes Pathogènes, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris cedex 15, France

ABSTRACT

The human immunodeficiency virus (HIV) Vif protein blocks incorporation of two host cell cytidine deaminases, APOBEC3F and 3G, into the budding virion. Not surprisingly, on a *vif* background nascent minus strand DNA can be extensively edited leaving multiple uracil residues. Editing occurs preferentially in the context of TC (GA on the plus strand) and CC (GG) depending on the enzyme. To explore the distribution of APOBEC3F and –3G editing across the genome, a product/substrate ratio (AA + AG)/(GA + GG) was computed for a series of 30 edited genomes present in the data bases. Two highly polarized gradients were noted each with maxima just 5′ to the central polypurine tract (cPPT) and LTR proximal polypurine tract (3′PPT). The gradients are in remarkable agreement with the time the minus strand DNA remains single stranded. *In vitro* analyses of APOBEC3G deamination of nascent cDNA spanning the two PPTs showed no pronounced dependence on the PPT RNA:DNA heteroduplex ruling out the competing hypothesis of a PPT orientation effect. The degree of hypermutation varied smoothly among genomes indicating that the number of APOBEC3 molecules packaged varied considerably.

INTRODUCTION

Cytidine deaminases of the human APOBEC3 family, particularly APOBEC3B, 3C, 3F and 3G, edit cytidine residues only in the context of single-stranded DNA (1–12). The genes are relatively well expressed in lymphocytes and some may be induced by interferon-α and -γ *in vitro* and *in vivo* (13–15). As such these enzymes represent a potential barrier to the replication of retroviruses that replicate via a single-stranded minus strand DNA intermediate. To counter restriction by APOBEC3F and 3G human immunodeficiency virus-1 (HIV-1) encodes a *vif* gene whose product, the Vif protein, is capable of excluding them from budding virions (3–7, 9,11,16). As all but one of the other lentiviruses encode a *vif* gene it is presumed that they too are susceptible to genetic editing by APOBEC3 orthologs (17–21). On the background of a lesion in the HIV-1 *vif* gene, either APOBEC3F or 3G is incorporated into the budding virion (2–7,9,11,12,16). Following infection of a target cell, cDNA synthesis ensues. However, C residues in the nascent single DNA strand can now be edited to U resulting in the collapse of viral information. Monotonous substitution of C for U on the minus strand shows up as numerous G→A transitions on the reference viral (+) strand, hence their name, G→A hypermutants (2–7,9,11,12,16). The fine substrate specificities of human APOBEC3F and 3G are subtly different, the former preferring the TC dinucleotide (GA on the plus strand) to CC (GG) where the edited base is underlined. APOBEC3G shows an inverse preference, i.e. CC (GG) > TC (GA) (1–3,5,8,9,12,22).

HIV minus strand DNA synthesis is primed by tRNAlys3 bound to the primer binding site (PBS) and does not differ qualitatively to that of other retroviruses. However, plus strand DNA synthesis is initiated at two RNA polypurine tracts (PPT) that remain associated with the minus strand cDNA. The first of these RNA primers is located within the integrase gene while the second maps immediately upstream of the 3′-LTR, the two being referred to as cPPT and 3′PPT, respectively. Dual initiation sites of plus strand DNA is a particular trait of the lentiviruses and spumaviruses (23,24). Given the priming of DNA synthesis at fixed sites on the genome and a finite velocity of DNA synthesis, it follows that not all cytidine residues will remain single stranded for the same period of time. Accordingly, the extent of editing should be dependent on the lifetime a C residue remains single stranded, assuming no product inhibition, which in turn suggests that the distribution of edited sites across the genome may well be non-random. Ideally such a hypothesis could be best addressed by comparisons of full-length HIV hypermutated genomes along with a normal sequence from the same sample.

A collection of 30 nearly full-length hypermutated sequences was recovered from the HIV databases and the

distribution of APOBEC3 edited sites across the genome analyzed. Although nearly half were lightly edited and hence relatively uninformative, two relatively smooth gradients of APOBEC3 editing of HIV-1 minus strand DNA were apparent culminating in maxima just 5′ to the two PPTs. *In vitro* editing experiments show that there is no orientation effect of the PPT RNA:DNA heteroduplex on editing. The twin gradients are consistent with the replication dynamics of HIV-1.

MATERIALS AND METHODS

Hypermutated HIV-1 sequences were identified by screening the databases with a variety of key words such as 'HIV, hypermutation' and 'HIV, hypermutated'. A total of 29 nearly full-length and 1 complete genome were identified. As the 5′ and 3′ ends of the sequences differed somewhat—reflecting the use of different amplification primers—they were trimmed so as to generate two groups of sequences encompassing ∼95% (n = 22) and ∼92% (n = 7) of unique sequence HIV-1. All have a common 5′ extremity mapping to the beginning of the *gag* coding region while the 3′ extremities mapped to the U3 region of the distal LTR. The extensively G→A hypermutated and reference sequences derived from the HIV-1 group O strain Vau have been published previously (25). A total of 20 sequences were used as non-hypermutated controls. They were selected from the Los Alamos HIV database and spanned all HIV-1 clades. The non-hypermutated HIV-1 O sequence Vau was also analyzed. The accession numbers of all sequences are given in Table 1.

For the calculation of the 'product substrate ratio' (PS ratio) across the genome, a simple program was written in Java, a copy of which can be found at http://www.pasteur.fr/recherche/unites/gmp/sitegmp/APOBEC3-PSratio.html. It calculates the number of AG, AA, GG, GA dinucleotides and the AG + AA/GG + GA ratio within a sliding window of variable length and interval. We found a window of 600 bp displaced at 50 bp intervals to be visually clear.

In vitro deamination of PPT loci

Two loci spanning the central and 3′PPTs were analyzed *in vitro* for APOBEC3G deamination of nascent minus strand cDNA. DNA spanning the cPPT and 3′PPT were amplified from the LAI molecular clone using primers pairs T7/cPPT+3′/cPPT and T7/3′PPT+3′/3′PPT, respectively. T7/cPPT, GCGAATTTAATACGACTCACTATAGGGAGCA-GCAATTTCACCAGTACTACGGTT; 3′/cPPT, CCTTCA-CCTTTCCAGAGGAGCTTT; 5′/cPPT, GCAGCAATTTCA-CCAGTACTACRGTT; T7/3′PPT, GCGAATTTAATACG-ACTCACTATAGGGACTTGGAAAGGATTTTGCTATAA; 3′/3′PPT, ACAAGCTGGTGTTCTCTCCTTTA; 5′/3′PPT, GCGCTTRRAAARRATTTTGCTATAA, where R = G or A. The T7 RNA polymerase promoter is underlined. RNA was made from uncloned PCR products while cDNA synthesis, as well as the deamination assay, were performed as described previously (8,26). Reaction volumes, times and temperatures were 50 μl, 3 h and 37°C, respectively. Sufficient material for subsequent cloning was recovered from the cDNA reactions by PCR in 12 cycles using the primer pairs 5′/cPPT + 3′/cPPT and 5′/3′PPT + 3′/3′PPT, respectively.

Table 1. Collection of hypermutated genomes and selected reference sequences

Hypermutated sequences			Reference sequences		
Code	Accession number	Clade*	Code	Accession number	Clade*
H1	AY037273	B/F	R1	AF004885	A
H2	AF362994	CRF01_AE/B	R2	AF005496	H
H3	AF442566	A/D	R3	AF061641	G
H4	AF442568	A/D	R4	AF067155	C
H5	AF442570	A/D	R5	AF077336	F1
H6	AF457057	A	R6	AF082394	J
H7	AF457060	A2/D	R7	AF286224	C
H8	AF457071	A	R8	AF286237	A2
H9	AF457074	A/D	R9	AF361872	A
H10	AF457076	A	R10	AF361874	C
H11	AF457091	A	R11	AF377956	F2
H12	AF484484	A	R12	K02013	B
H13	AY037274	B	R13	K03454	D
H14	AY037276	B/F	R14	M26727	B
H15	AY237165	A/C/D	R15	M62320	A1
H16	AY237166	A/D	R16	U34604	B
H17	AY237167	A/C/D	R17	U46016	C
H18	AY255828	C	R18	U88824	D
H19	AY531116	B	R19	X04415	A/D
H20	AY781125	B	R20	AF286238	A
H21	AY829213	B			
H22	AY358053	CRF01_AE			
H23	AY358054	CRF01_AE			
H24	AY358055	CRF01_AE			
H25	AY358058	CRF01_AE			
H26	AY358061	CRF01_AE			
H27	AY037279	CRF12_BF			
H28	AY734557	C			
H29	AY734561	C			
HVau	AF407419	O	Vau	AF407418	O

A simplified designation for each sequence is given along with the accession number. Although some sequences belonged to 'pure' HIV-1 M group clades, 18 were recombinants. The asterisk serves as a reminder that HVau and Vau do not belong to group M, but to the HIV-1 O group.

PCR products were cloned into the TOPO TA cloning vector and DNA sequenced using Big Dye terminators. HIV-1 RT was purchased from Amersham. Human APOBEC3G was expressed as a GST fusion protein using a recombinant Baculovirus and purified as described previously (8).

RESULTS AND DISCUSSION

Twin gradients for APOBEC3 editing

To date the only full-length G→A hypermutated HIV-1 sequence was fortunately accompanied by an unedited reference sequence from the same DNA sample (25). Comparison of the number of edited residues within a non-overlapping sliding window of 200 bp across the genome shows a distinctly non-random distribution (Figure 1). At a macroscopic level there were two minima 3′ to the PBS and cPPT and two maxima both 5′ to the cPPT and 3′PPT (black line). The signal was strongest in the context of CC and TC dinucleotides (blue) to the detriment of GC and AC (red), which is in keeping with the preference of APOBEC3F and 3G for these targets (1–3,5,8,9,12,22).

To see if this was a general finding a total of 29 nearly full-length G→A hypermutated sequences (92–95%) were recovered from GenBank. Unfortunately none was accompanied

by an unedited sequence. To overcome this, a simple metric was computed to highlight the effects of editing. The preferred product of APOBEC3F editing is AA from a GA substrate, while that for APOBEC3G is AG from a GG dinucleotide. As both deaminases may be packaged by a Δvif virion and edit neo-synthesized cDNA, the ratio

Figure 1. High-resolution analysis of the hypermutated full-length genome (HVau) compared to its unedited reference sequence (Vau). To reduce noise due to small numbers, a 200 bp non-overlapping window was chosen for analysis. Whether substitutions are normalized to the G (C on the negative strand) content (black), GG + GA (CC + TC, blue, the preferred context for APOBEC3 deamination) twin gradients were always observed. No significant gradient was observed for the GT + GC (AC + GC, red which are avoided by APOBEC3). The proviral form of the sequence is given along with the cPPT and 3′PPTs.

(AA + AG)/(GA + GG)—termed PS ratio for 'product substrate'—within a sliding window across the sequence should reveal evidence of APOBEC3 editing. For an ostensibly hypermutated sequence, if editing was infrequent the profile of PS ratios across the genome should be close to those for unedited genomes. For a highly edited sequence there should be considerable deviation from a set of references sequences. In this latter situation it is possible that AG (CT on the minus strand) and GA (TC) may be affected. However, as this would affect both the numerator and denominator, the impact on the PS ratio may not be too great.

The PS ratio was first computed for a series of 20 unedited HIV-1 M reference sequences (Table 1), as well as the unedited HIV-1 group O Vau sequence, to see if there were local variations across the genome that could hinder interpretation. As can be seen from Figure 2A for a sliding window of 600 bp the PS ratios varied little through to 5500 with a local high ∼5800 bp followed by a decline towards a ratio close to 1 at ∼8400 bp. This local peak reflects a known elevated A content centered on the first hypervariable region of gp120 and naturally increases the PS ratio. In contrast, the LTR region is relatively rich in G and so it is not surprising that the PS ratio is at a minimum from 7500 bp to the end. As the 20 sequences varied somewhat in length (8751–8864 bp, $\Delta = 1.2\%$), they were aligned by ClustalW and individual PS profiles recomputed. As this changed little the PS profiles (data not shown), unaligned reference sequences were used
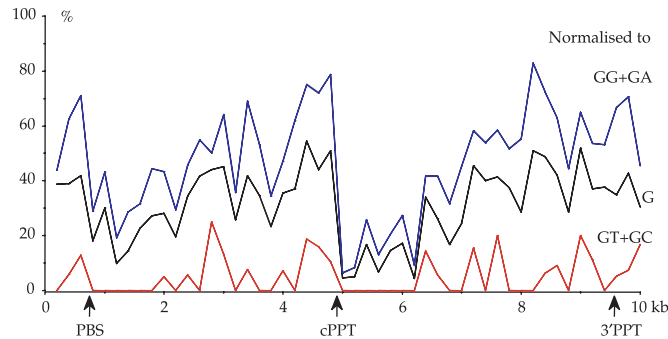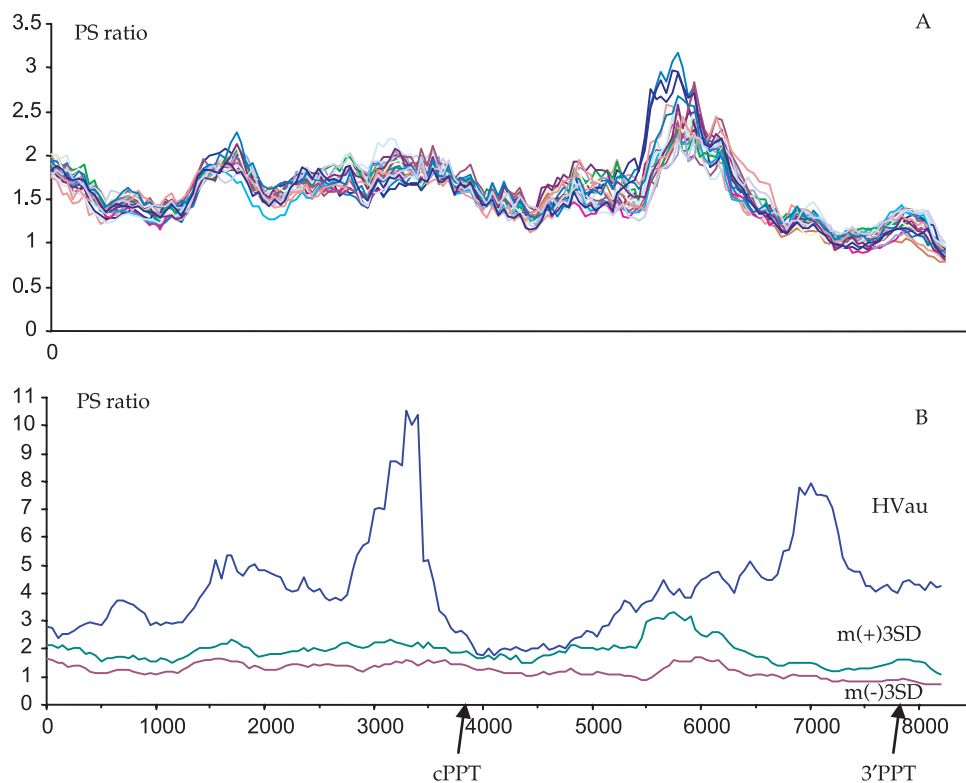

Figure 2. APOBEC3 product/substrate ratio (PS) scans across HIV genomes. (**A**) PS scans for a collection of 20 unedited HIV-1 M references sequences (Table 1) starting from the beginning of the Gag orf and running through to the U3 and R regions of the LTR. Hence the numbering system is not that of complete HIV sequences. The window length was 600 bp displaced at 50 bp intervals. The peak at ∼5800 corresponds to a known A-rich region centered on the first hypervariable region in the gp120 coding region. (**B**) PS scan for HVau along with the mean ± 3 SDs clearly shows that all regions of the HVau genome were significantly edited by APOBEC3 molecules albeit to different degrees. The positions of the cPPT and 3′PPT are indicated.

from here on. The PS profile of the group O Vau reference sequence is indistinguishable from those of HIV-1 group M sequences (data not shown).

The PS profile for the hypermutated Vau sequence (HVau) is shown in Figure 2B along with the mean PS ± 3 SDs computed from the reference sequence set. Throughout, the PS profile of HVau was significantly greater compared to controls. The profile captures the essential features of the distribution of edited sites across the genome with twin gradients 5′ to the two PPTs (Figure 1).

Wide range in degree of APOBEC3 editing

Given the variable PS ratios across the genome, in the following analyses of hypermutated sequences, the mean PS ratio computed from the 20 HIV-1 M reference sequences was subtracted. As can be seen the normalized PS ratios, referred to as PS∗, 13 sequences (H2, 3, 5, 7, 11, 13, 14, 17, 19, 22, 26, 27 and 29; Figure 3A) were only slightly edited compared to controls (PS∗ ⩽ 1) despite the fact that they were annotated as hypermutated. A further 13 sequences (H1, 4, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25 and 28; Figure 3B) were moderately edited with a PS∗ ratio rarely exceeding 2. In order to enhance the signal, the PS∗ values were averaged among the

collection of 13 sequences represented in Figure 3B. As can be seen twin gradients were apparent (Figure 3C). Only three sequences (H6, 21 and 23; Figure 3D) exhibited PS∗ ratios >2. For this last group, H23 apart (see below), twin gradients with maxima just 5′ to the cPPT and 3′PPT were in evidence. Even for moderately edited sequences gradients were evident particularly in the first half of the sequence (Figure 3B–D, 1–3500 bp). In short, extensively APOBEC3 edited genomes such as H6, 21, H23 and Vau (4/30 or ∼13%) are relative rarities.

To explore further the variable degree of APOBEC3 editing of these HIV-1 genomes, their overall base composition was examined (Figure 4). As expected the accumulation of A is negatively and monotonously correlated to the depletion of G. Four sequences in particular were particularly lightly edited, yet appear at the upper end of the distribution for the reference sequences. Assuming that this small difference (∼0.3%) is due to n APOBEC3 molecules per virion then a hypermutated genome with an overall A content of ∼42.2% would correspond to ∼19-fold more [(42.2 − 36.5)/0.3]. Although the minimal value for $n = 1$, in the close confines of a HIV replication complex this translates into a concentration of the order of 10 µM (27). Hence a 19-fold increase is synonymous with an enzyme concentration of ∼0.2 mM
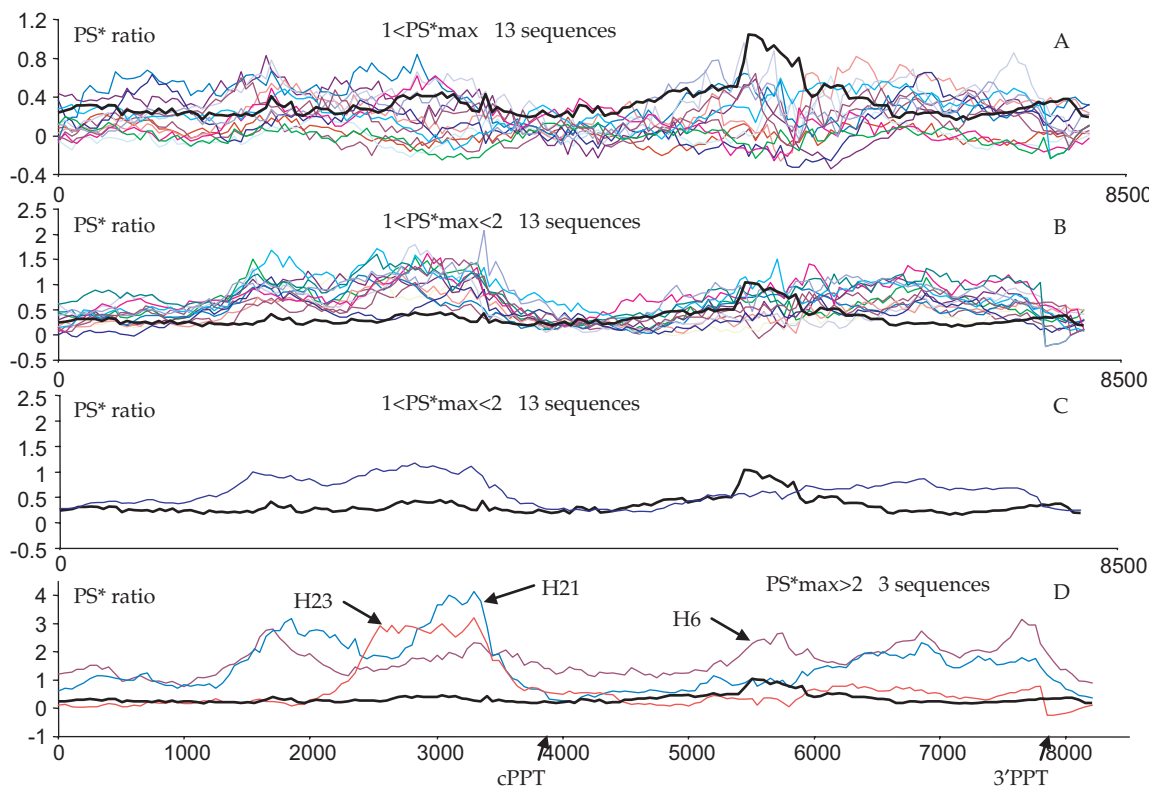

Figure 3. APOBEC3 product/substrate ratio scans for three groups of hypermutated genomes. Given the somewhat irregular PS ratio across the HIV genome (Figure 2A) the mean PS ratios for the reference sequences were subtracted from those of the individual hypermutated sequences, yielding the PS∗ ratio. (**A**) A collection of 13 sequences where PS∗max <1. The profiles in bold for all four graphs represents the mean + 3 SD derived from the reference sequences. (**B**) A collection of 13 sequences where 1<PS∗max<2. (**C**) In order to enhance the signal the PS∗ values of the 13 sequences represented in B) were averaged. (**D**) A collection of 3 sequences where PS∗max>2. Note that the ordinate scale varies for all three graphs. The positions of the cPPT and 3′PPT are given. As the PS∗ ratio for the interval $x \rightarrow x + 600$ bp is reported at position x, the PS∗ ratio starts to decay at cPPT-600 and 3′PPT-600. Hence the twin gradients do indeed reach maxima just 5′ to the two PPTs. The sharp break in the PS∗ ratio for six sequences ∼7800 results from their being slightly shorter than the reference sequence set (Figure 5C).
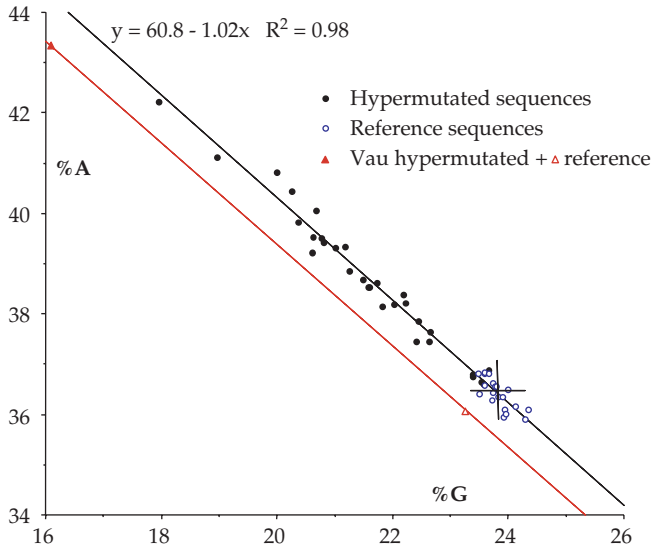
Figure 4. Smooth distribution in the A and G content of HIV-1 M group hypermutated genomes. The increase in A content is strictly related to the depletion of G. The large cross represents the mean (35.6%) for the 20 reference sequences. The two data points for the Vau sequences (HIV-1 O) are displaced towards a lower G content. However, as the gradient is parallel to that for HIV-1 M group sequences, it may be presumed that there is no qualitative difference in the way the hypermutated Vau sequences was edited.

which is considerable. Hence it seems plausible, but by no means proven, that the most lightly edited sequences reflect editing by a single APOBEC3 enzyme.

No PPT orienting of APOBEC3

Given the asymmetry of editing around the two PPT motifs, it might be posited that the RNA:DNA PPT heteroduplex remaining following reverse transcription had some orientation effect on APOBEC3G editing. To explore this we performed an *in vitro* deamination assay on two loci spanning the two PPTs in a manner comparable to a previous study (8). RNA corresponding to 300–400 bp segments spanning the central and 3′PPTs was made *in vitro* and cDNA was made using HIV-1 RT which would leave PPT RNA hybridized to the cDNA. Baculovirus APOBEC3G was present from the outset to mimic the events occurring within the HIV replication complex of a Δ*vif* genome derived from a non-permissive cell. As can be seen from Figure 5A and B, editing occurred on either side of both PPTs at similar frequencies. These observations show that the RNA:DNA heteroduplex had no orienting effect on APOBEC3G editing.

As can be seen in Figure 5C, 5/18 and 1/20 cDNAs derived from the *in vitro* reactions had edited cPPT and 3′PPT tracts, respectively, meaning that either there is a little breathing of the RNA:DNA heteroduplex or displacement of the RNA
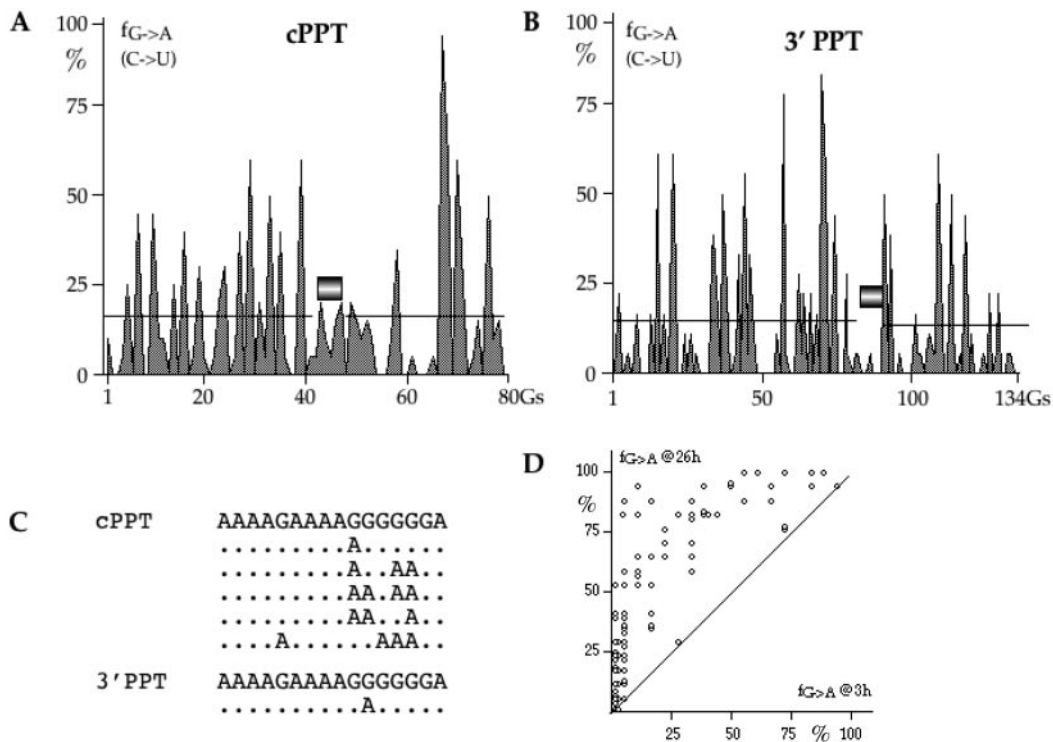

Figure 5. *In vitro* APOBEC3G editing of nascent HIV PPT cDNA. (**A**) Cytidine specific deamination frequencies within 338 bp fragment spanning the cPPT. It encodes 83 G(C) residues and is shown with respect to the reference plus strand. The ordinate gives the frequency of editing for every C residue among a collection of 20 clones. The box denotes the PPT sequence while the horizontal bars denote the mean cytidine editing frequency on either side of the PPT. (**B**) Cytidine specific deamination frequencies within 461 bp fragment spanning the 3′PPT, and comprises a total of 143 G(C) residues. (**C**) Edited PPT sequences from the two loci shown with respect to the reference sequence. (**D**) Graphic representation of % site-specific deamination frequencies across the template for 3 and 26 h reactions are plotted on the *x* and *y* axes, respectively. The mean site-specific editing frequencies over all sites were 21.9% at 3 h and 53.0% at 26 h. Hence editing was 2.4-fold more extensive after 26 h incubation. As most of the preferred CC and TC targets are >90%, the 26 h reaction showed signs of saturation.

strand by APOBEC3G. To explore the temporal component to APOBEC3G editing, a second cPPT cDNA reaction was incubated with APOBEC3G for 26 h as opposed to 3 h. Material was recovered by amplification and individual clones were sequenced. As shown in Figure 5D the degree of editing was increased ~2.4-fold compared to the 3 h reaction. Most GG and GA targets were >75% substituted at 26 h. As would be expected for a simple reaction without product inhibition, the longer the incubation time, the greater the degree of editing.

Editing gradients are compatible with HIV replication dynamics

As there was no orientation effect of the PPT RNA:DNA heteroduplex on APOBEC3G editing, how might the twin gradients of APOBEC3 editing be understood? That the deamination frequency was related to the reaction time (Figure 5D) suggests a solution related to the mechanics of HIV replication. HIV-1 DNA synthesis starts with the single-stranded minus strand primed by tRNAlys3. As soon as the short PPT RNA primers are generated by the RNaseH function of RT, double-stranded DNA synthesis proceeds. However, as only single-stranded DNA is a substrate for APOBEC3 editing, not all regions remain singled stranded for the same amount of time. For example, a base 3' proximal (vis-à-vis the plus strand) to either of the PPTs will remain single stranded for a very short time compared to those several kilobases downstream (Figure 6A). Assuming

a constant velocity for the synthesis of minus (V_R) and plus strand DNA (V_D), it is simple to show that $f_{c\to u} = b[\text{APOBEC3}]t_c = b[\text{APOBEC3}](d/V_R + d/V_D)$, where $f_{c\to u}$ is the fraction of cytidine residues edited within a small interval at a distance d between the target cytidine and the upstream PPT, t_c is the time a C residue remains single stranded and b is a constant. As V_R and V_D are constants, as is the concentration of APOBEC3 for a given virion, this reduces to $f_{c\to u} = b'd$, where b' is a constant. In short, the amplitude of editing is proportional to the distance from the PPT assuming a relatively smooth distribution of dinucleotide targets across the genome (Figure 6B).

The PS∗ profile of sequence H23 (28) does not fit the schema. As shown in Figure 3 the PS∗ ratios between 2200 and 3500 indicate extensive deamination. By contrast the surrounding PS∗ ratios are indistinguishable from those of the reference sequence set. It is as though H23 represented a recombinant between two genomes one edited, the other not. Retroviral recombination presupposes that a hypermutated provirus may be transcribed and its genomic RNA packaged. Although this is a fascinating question, a posteriori analysis cannot distinguish it from the result of PCR recombination (29). Yu *et al.* (10) noted a gradient in editing across the HIV genome and sparsely edited LTR sequences immediately downstream of 3'PPT. However, they studied small PCR fragments across the genome and not near full-length genomes. As they did not analyze a segment encompassing the cPPT they concluded that there was a gradient across the entire genome when in fact there are two gradients in
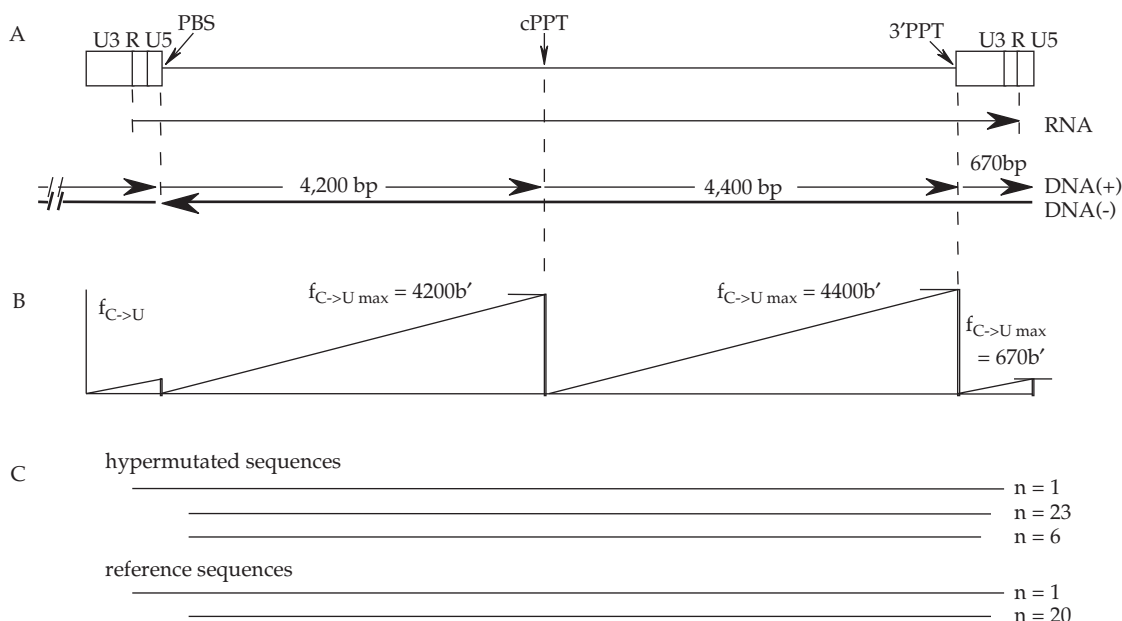

Figure 6. Kinetic model of APOBEC3 editing of nascent HIV-1 DNA. (**A**) Schematic representation of HIV proviral transcription and the first steps of reverse transcription drawn to scale. PBS, primer binding site; cPPT and 3'PPT, central and 3' polypurine tracts. (**B**) The amplitude of cytidine deamination $f_{c\to u}$ is proportional to the time (t) a base remains single stranded. Assuming that the velocities of RNA- and DNA-dependent reverse transcription (V_R and V_D) are constant across the genome, equal access to all sites and a smooth distribution of targets across the sequence, then $f_{c\to u} = b[\text{APOBEC3}]t = bd(1/V_R + 1/V_D)$ where d is the distance of any cytidine residue to the downstream PPT, [APOBEC3] is the concentration of APOBEC3 molecules for a single virion, and b is a constant. This reduces to $f_{c\to u} = b'd$, where b' is a constant. Given that the maximum values of d are essentially the same, 4.2 and 4.4 kb for the two regions primed by synthesis from the 3'PPT-PBS and cPPT, this results in twin gradients of similar amplitude. The model predicts that the LTR minus strand is lightly edited compared to other parts of the genome. Although very few complete hypermutated LTR sequences exist, two papers reported infrequent editing just 3' of the 3'PPT (10,30). (**C**) Linear schematic indicating the lengths and numbers of sequences used in this study.

the same orientation reflecting the mechanics of HIV DNA synthesis. The observation that the LTR is lightly edited was recently confirmed by Wurtzer *et al.* (30).

A rather elegant *in vitro* study of APOBEC3G editing of DNA showed that for an editing hotspot, the deaminase was able to jump from 3′ to 5′ over ∼100 bp (31). This would serve to amplify the positional bias in substrate editing across the genome but could not generate the double gradients with minima just downstream of the two PPTs. Given that the study concerned editing hotspots, the phenomenon may introduce local regions of intense editing so explaining additional peaks in the PS profile, e.g. those ∼1600–2000 bp for H21, H23 and Vau (Figure 3). In contrast the *in vitro* findings of aberrant DNA priming on a minus strand DNA template resulting from total substitution of TTP by dUTP (32) are not comforted by the present findings *in vivo*. Were this to be the case twin gradients would not be anticipated. This may be due to the fact that PPT DNA is relatively well protected from deamination (Figure 5A and B).

The variation in the degree of editing of near full-length genomes indicates a wide range in the number of APOBEC3F/G molecules packaged per virion (Figures 3 and 4). For a Δ*vif* virus this could reflect differences in the concentration of APOBEC3 molecules in the donor cell and/or the vagaries of APOBEC3 packaging. An additional hypothesis may be that some wild-type genomes encode functional, yet suboptimal *vif* alleles allowing packaging of small numbers of APOBEC3 molecules. As considerable variation in Vif function has been shown among alleles derived from clinical isolates this may be a plausible hypothesis (33). At a practical level, choosing a locus mapping just 5′ to either of the PPTs would provide a more sensitive readout of function in any analysis of G→A hypermutation, e.g. analysis of the impact of a mutation in the APOBEC3G gene.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Beale,R.C., Petersen-Mahrt,S.K., Watt,I.N., Harris,R.S., Rada,C. and 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**, 585–596.
2. Bishop,K.N., Holmes,R.K., Sheehy,A.M., Davidson,N.O., Cho,S.J. and Malim,M.H. (2004) Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr. Biol.*, **14**, 1392–1396.
3. Harris,R.S., Bishop,K.N., Sheehy,A.M., Craig,H.M., Petersen-Mahrt,S.K., Watt,I.N., Neuberger,M.S. and Malim,M.H. (2003) DNA deamination mediates innate immunity to retroviral infection. *Cell*, **113**, 803–809.
4. Lecossier,D., Bouchonnet,F., Clavel,F. and Hance,A.J. (2003) Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science*, **300**, 1112.
5. Liddament,M.T., Brown,W.L., Schumacher,A.J. and Harris,R.S. (2004) APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 *in vivo*. *Curr. Biol.*, **14**, 1385–1391.
6. Mangeat,B., Turelli,P., Caron,G., Friedli,M., Perrin,L. and Trono,D. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*, **424**, 99–103.
7. Mariani,R., Chen,D., Schrofelbauer,B., Navarro,F., Konig,R., Bollman,B., Munk,C., Nymark-McMahon,H. and Landau,N.R. (2003) Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell*, **114**, 21–31.
8. Suspene,R., Sommer,P., Henry,M., Ferris,S., Guetard,D., Pochet,S., Chester,A., Navaratnam,N., Wain-Hobson,S. and Vartanian,J.P. (2004) APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. *Nucleic Acids Res.*, **32**, 2421–2429.
9. Wiegand,H.L., Doehle,B.P., Bogerd,H.P. and Cullen,B.R. (2004) A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *EMBO J.*, **23**, 2451–2458.
10. Yu,Q., Konig,R., Pillai,S., Chiles,K., Kearney,M., Palmer,S., Richman,D., Coffin,J.M. and Landau,N.R. (2004) Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nature Struct. Mol. Biol.*, **11**, 435–442.
11. Zhang,H., Yang,B., Pomerantz,R.J., Zhang,C., Arunachalam,S.C. and Gao,L. (2003) The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature*, **424**, 94–98.
12. Zheng,Y.H., Irwin,D., Kurosu,T., Tokunaga,K., Sata,T. and Peterlin,B.M. (2004) Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. *J. Virol.*, **78**, 6073–6076.
13. Bonvin,M., Achermann,F., Greeve,I., Stroka,D., Keogh,A., Inderbitzin,D., Candinas,D., Sommer,P., Wain-Hobson,S., Vartanian,J.P. and Greeve,J. (2006) Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology*, **43**, 1364–1374.
14. Peng,G., Lei,K.J., Jin,W., Greenwell-Wild,T. and Wahl,S.M. (2006) Induction of APOBEC3 family proteins, a defensive maneuver underlying interferon-induced anti-HIV-1 activity. *J. Exp. Med.*, **203**, 41–46.
15. Tanaka,Y., Marusawa,H., Seno,H., Matsumoto,Y., Ueda,Y., Kodama,Y., Endo,Y., Yamauchi,J., Matsumoto,T., Takaori-Kondo,A. *et al.* (2006) Anti-viral protein APOBEC3G is induced by interferon-alpha stimulation in human hepatocytes. *Biochem. Biophys. Res. Commun.*, **341**, 314–319.
16. Sheehy,A.M., Gaddis,N.C., Choi,J.D. and Malim,M.H. (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, **418**, 646–650.
17. Fitzgibbon,J.E., Mazar,S. and Dubin,D.T. (1993) A new type of G→A hypermutation affecting human immunodeficiency virus. *AIDS Res. Hum. Retroviruses*, **9**, 833–838.
18. Gao,F., Yue,L., White,A.T., Pappas,P.G., Barchue,J., Hanson,A.P., Greene,B.M., Sharp,P.M., Shaw,G.M. and Hahn,B.H. (1992) Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature*, **358**, 495–499.
19. Perry,S.T., Flaherty,M.T., Kelley,M.J., Clabough,D.L., Tronick,S.R., Coggins,L., Whetter,L., Lengel,C.R. and Fuller,F. (1992) The surface envelope protein gene region of equine infectious anemia virus is not an important determinant of tropism *in vitro*. *J. Virol.*, **66**, 4085–4097.
20. Vartanian,J.P., Meyerhans,A., Asjo,B. and Wain-Hobson,S. (1991) Selection, recombination, and G→A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol.*, **65**, 1779–1788.
21. Wain-Hobson,S., Sonigo,P., Guyader,M., Gazit,A. and Henry,M. (1995) Erratic G→A hypermutation within a complete caprine arthritis-encephalitis virus (CAEV) provirus. *Virology*, **209**, 297–303.
22. Langlois,M.A., Beale,R.C., Conticello,S.G. and Neuberger,M.S. (2005) Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res.*, **33**, 1913–1923.
23. Charneau,P., Alizon,M. and Clavel,F. (1992) A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *J. Virol.*, **66**, 2814–2820.
24. Kupiec,J.J., Tobaly-Tapiero,J., Canivet,M., Santillana-Hayat,M., Flugel,R.M., Peries,J. and Emanoil-Ravier,R. (1988) Evidence for a gapped linear duplex DNA intermediate in the replicative cycle of human and simian spumaviruses. *Nucleic Acids Res.*, **16**, 9557–9565.

25. Vartanian,J.P., Henry,M. and Wain-Hobson,S. (2002) Sustained G→A hypermutation during reverse transcription of an entire human immunodeficiency virus type 1 strain Vau group O genome. *J. Gen. Virol.*, **83**, 801–805.

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

27. Benjamin,J., Ganser-Pornillos,B.K., Tivol,W.F., Sundquist,W.I. and Jensen,G.J. (2005) Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography. *J. Mol. Biol.*, **346**, 577–588.

28. Tovanabutra,S., Beyrer,C., Sakkhachornphop,S., Razak,M.H., Ramos,G.L., Vongchak,T., Rungruengthanakit,K., Saokhieo,P., Tejafong,K., Kim,B. *et al.* (2004) The changing molecular epidemiology of HIV type 1 among northern Thai drug users, 1999 to 2002. *AIDS Res. Hum. Retroviruses*, **20**, 465–475.

29. Meyerhans,A., Vartanian,J.P. and Wain-Hobson,S. (1990) DNA recombination during PCR. *Nucleic Acids Res.*, **18**, 1687–1691.

30. Wurtzer,S., Goubard,A., Mammano,F., Saragosti,S., Lecossier,D., Hance,A.J. and Clavel,F. (2006) Functional central polypurine tract provides downstream protection of the human immunodeficiency virus type 1 genome from editing by APOBEC3G and APOBEC3B. *J. Virol.*, **80**, 3679–3683.

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

32. Klarmann,G.J., Chen,X., North,T.W. and Preston,B.D. (2003) 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.

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