AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 19, Number 12, 2003, pp. 1115–1123
© Mary Ann Liebert, Inc.

Hypermutation of HIV Type 1 Genomes Isolated from Infants Soon after Vertical Infection

IRENE N. KOULINSKA,[1] BETH CHAPLIN,[1] DAVIS MWAKAGILE,[2] MAX ESSEX,[1] and BORIS RENJIFO[1]

ABSTRACT

Hypermutation involving excessive G-to-A substitutions in the dinucleotide context GA or GG is common among the lentiviruses and results in multiple stop codons across the genome. Hypermutated viruses have been associated with slower disease progression and might reflect an antiviral cell-defense mechanism. However, it is unclear how soon G-to-A substitutions are generated after infection and whether they occur randomly along the genome. In this report we describe for the first time hypermutated sequences detected at delivery and in the first weeks of life, which suggests that they could be either generated *in utero* and soon after birth and/or vertically transmitted. Hypermutated C2–C5 *env* clones were harbored in 13.2% of 243 infants and 18.6% of 199 mothers. A lower extent of hypermutation was found in infants than in mothers (Fisher's exact $p = 0.034$), but there was no relationship between the percent hypermutated Gs and viral subtype or transmission status of the mother. Analyses of six hypermutated full-length HIV-1 clones showed that although all genes could be affected by G-to-A substitutions, there was a significant drop in the extent of hypermutation between the central polypurine tract and the beginning of *env*, indicating that hypermutation across the HIV-1 genome might occur in a specific pattern. The genomic regions most affected by hypermutation were *pol* and *env* while both polypurine tracts remained unaffected. A better understanding of the mechanism of hypermutation may reveal novel virus–host interactions that could be targeted in drug development.

INTRODUCTION

PURINE–PURINE TRANSITIONS are the most frequent substitutions in retroviruses, and overall the expected guanine-to-adenine (G-to-A) transition frequency is 5% of the number of Gs.[1] Excessive G-to-A substitutions, which exceed all other mutations in a viral sequence, have been reported among the lentiviruses, including HIV. The phenomenon is known as G-to-A hypermutation.[2–8] Hypermutated genomes are frequently nonviable since the G-to-A transitions occur preferentially in a GpA and GpG dinucleotide context resulting in multiple in-frame stop codons.[3,4] Information about the prevalence of hypermutated HIV-1 genomes is limited. Hypermutated sequences were found in 23 of 53 patients (43%) at different clinical stages and were predominant over time in long-term nonprogressors.[9–11] To better correlate hypermutation with disease progression and viral pathogenesis it would be valuable to know how soon after transmission hypermutated variants can be generated and whether they can be transmitted.

The mechanism of hypermutation is not well known. It has been suggested that it is induced by deoxynucleoside triphosphate (dNTP) imbalances during and after cell activation.[9] In this case, hypermutation would take place during negative DNA strand synthesis through reverse transcriptase (RT) slippage and realignment (GpA context) and by direct base misincorporation (GpG context), which are both facilitated by the capacity of lentiviral RTs to elongate beyond nucleotide mismatches.[12–16] However, a new mechanism of hypermutation involving deamination of newly synthesized viral DNA has been recently suggested.[17–19] Editing of minus-strand viral DNA by the cytidine deaminase APOBEC3G resulted in G-to-A substitutions in the HIV genome *in vitro* and this effect was counteracted by the Vif protein.[17–19] It is not known if these mechanisms take place randomly across the genome or follow specific restrictions. HIV-1 hypermutation has been reported in various regions of the genome,[3,4,9–11,20,21] but the only full-length hypermutated sequence that has been published belonged to the outlier group O, which itself is quite rare.[2] Although this sequence seemed

[1]Harvard AIDS Institute and Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115.
[2] Department of Microbiology and Immunology, Muhimbili Medical Center, Dar es Salaam, Tanzania.

to have random G-to-A substitutions, recent analyses of different regions of HIV-1 genomes from more common subtypes suggest that hypermutation might follow a specific pattern in which its extent increases progressively from the 5′ end of the genome to the central polypurine tract (PPT), then drops and rises again until the 3′ PPT.[22]

Here, we describe hypermutated HIV-1 C2–C5 *env* sequences found at birth or soon after vertical infection among 243 infants. This analysis indicates that hypermutation can occur early after transmission and/or that hypermutated viruses can be vertically transmitted. To investigate whether hypermutations were randomly distributed along the HIV-1 genome of vertically transmitted viruses, six near full-length hypermutated clones from three infants harboring HIV-1 intersubtype recombinant genomes were analyzed. One hundred ninety-nine mothers from the same cohort were also screened for hypermutated HIV-1 C2–C5 *env* sequences to compare the frequency and extent of hypermutation as well as its nucleotide preference in infants and in adults.

MATERIALS AND METHODS

Study subjects

Samples were collected from infants and mothers enrolled in a randomized double-blind placebo-controlled trial that investigated the role of vitamin supplements in the prevention of mother-to-child transmission (MTCT) of HIV-1 in Dar es Salaam, Tanzania.[23,24] A total of 199 mothers and 243 infants had been tested and there were 64 mother–infant pairs among them. Their HIV-1 status had been previously determined by polymerase chain reaction (PCR) amplification, and nucleotide sequences from *gag* and *env* regions were used for subtype classification.[25] Samples TZBFL0170, TZBFL0169, and TZBFL0210 had been found to contain hypermutated genomes in a previous full-length sequence analysis of recombinant HIV-1 viruses.[26] Samples TZBFL0169, TZBFL0170, and TZBFL0210 were drawn at 14, 13, and 10 weeks of age, respectively. All three specimens were collected between 1996 and 1997. No mothers or infants received antiretroviral medications before or during the study period. The project was approved by the Ethical Review Committee of the Harvard School of Public Health, the Research and Publications Committee of Muhimbili University College of Health Sciences, and the Ethical Committee of the National AIDS Control Program of the Tanzanian Ministry of Health.

PCR amplification and cloning

Amplification and cloning of full-length genomes and C2–C5 *env* fragments were done directly from the patients' peripheral blood mononuclear cells (PBMCs) and have been described elsewhere.[25–28] Primers G49 (5′-aac-caa-ggg-gaa-ggg-aca-tag-3′) and p16 (5′-gtg-gta-ttc-cta-att-gaa-ctt-c-3′) were used to amplify a 1.4-kb *pol* fragment from samples TZBFL0169, TZBFL0170, and TZBFL210 to determine the presence of nonhypermutated variants among the patients' quasispecies. Dye-terminator sequencing was performed with an automated sequencer. The identification of crossover points in sample TZBFL0210 was based on the methodology previously used to describe the recombination patterns of samples TZBFL0170 and TZBFL0169.[26] Briefly, full-length alignments were conducted using the Clustal X software package[29] and including references recommended by the Los Alamos HIV Sequence Database.[30] The Simplot program was used to screen for recombination breakpoints followed by analysis with the Recombination Identification Program (RIP).[31,32] To confirm the location and significance of crossovers, the Maynard Smith maximum Chi-square method was used.[33] Regions adjacent to the apparent recombination points were individually subjected to phylogenetic analysis.

Identification and analysis of hypermutated clones

C2–C5 *env* sequences from all infected mothers and infants enrolled in the Tanzania vitamin trial up to August 2001 were screened for hypermutation.[25,28] Twelve hundred and forty-eight C2–C5 *env* clones from 199 mothers and 1379 C2–C5 *env* clones from 243 infants were available for this analysis. The percent adenine plus thymine (A + T) content was calculated for each clone. Sequences with A + T content above the median value of this population (65.0%) were aligned to control clones that were isolated from the same patient and had $\leq$65% A+T content and an open reading frame. The percent guanines (Gs) [number of Gs mutated to adenines (As) divided by the total number of Gs in the sequence], the ratio of G-to-A vs. A-to-G substitutions, and the dinucleotide context of G-to-A substitutions were calculated using in-house Perl scripts. A clone was categorized as hypermutated if the following conditions were present: a statistically significant increase in the percent Gs substituted relative to control clones (two-sample test of proportions), a preference for a GpG or GpA dinucleotide context, and a G-to-A vs. A-to-G ratio higher than 1 as evidence for bias in the transition substitutions.

All hypermutated C2–C5 *env* clones as well as the full-length genomes of samples TZBFL0169, TZBFL0170, and TZBFL0210, which had been previously found to contain multiple in-frame stop codons due to G-to-A substitutions, were studied with the Hypermut Program Package.[34] This program identifies mutations and their dinucleotide context with respect to a reference sequence included in each alignment. As a reference for the full-length clones, we used consensus sequences based on the subtype consensus sequences recommended by the Los Alamos HIV Sequence Database and constructed to individually match the respective recombinant patterns, thus allowing for comparison between genetic regions of the same subtype.[35]

A nonhypermutated clone from the same patient was used as a reference to study the extent of hypermutation among C2–C5 *env* sequences from the Tanzania clinical trial database. All 32 infants with hypermutated C2–C5 *env* clones had at least one nonhypermutated clone detected. Seven of 37 mothers carrying hypermutated sequences had no nonhypermutated clones available and were excluded from the analyses of the extent of hypermutation but not from the analyses of the dinucleotide preference of hypermutation. The context of G-to-A substitutions in those clones was determined using a consensus sequence of a matching length and subtype to allow for comparisons involving the dinucleotide preference of hypermutation. Two or more hypermutated clones per patient were included in the statistical analyses only if those clones had different extent

and distribution patterns of G-to-A substitutions indicating independent origin. Pearson Chi-square, Fisher's exact test for small binomial samples, Student's *t* test, and the two-sample test of proportions were used for statistical analysis.

RESULTS

Hypermutation of HIV-1 C2–C5 env sequences from mothers and infants

Previously amplified HIV-1 subtype A, C, D and recombinant C2–C5 *env* sequences from 199 mothers and 243 infants were screened for hypermutation. The average number of clones per patient was approximately the same among subtypes and between mothers and infants, but a higher number of clones had been sequenced from samples with recombinant C2–C5 *env* sequences (Table 1).

The proportion of sequences with stop codons among clones with different A + T content is shown in Figure 1. Among sequences with A+T content lower than the median for this population (65%), there was a very similar proportion of clones that contained stop codons (6.4–7.0%), which increased sharply for higher A+T values (Fig. 1). Based on this distribution and similar findings described by others, only sequences that had ≥65.1% A+T content were examined further for hypermutation.[9] Given that generally less than 3% of the total number of Gs (103–136) in the C2–C5 region analyzed were replaced by As when comparing control clones, at least a 5% increase in the percent Gs substituted was necessary for the designation of a hypermutant clone (two-sample test of proportions). The frequency in which hypermutated clones were detected was 13.2% among infants and 18.6% among mothers (two-sample test of proportions, $p = 0.118$) (Table 1).

The proportion of mothers and infants harboring hypermutated C2–C5 *env* sequences was very similar among HIV-1 subtypes A, C, and D, and although it seemed to be higher for recombinant samples as compared to pure subtypes, this difference was not significant (two-sample test of proportions, $p = 0.265$).

The mean age of all infants screened for hypermutation was 21 weeks and the mean age of infants in which hypermutated C2–C5 *env* sequences were found was 17 weeks, ranging from zero (at birth) to 119 weeks.

Five (15.6%) newborns had hypermutated C2–C5 *env* sequences at delivery, and 18 infants (56%) had hypermutated C2–C5 *env* sequences within the first 6 weeks of life. These proportions were not significantly different from the ones found in the study population—15.8% and 40.7% of the 1379 clones screened for hypermutation were derived from samples collected at birth and before 6 weeks, respectively. Although hypermutated viruses could be detected at birth, there was no direct evidence that they could be vertically transmitted. Of the 64 mother–child pairs analyzed, hypermutated virus could be isolated from nine mothers but not from their infants. Similarly, no hypermutation was detected in the six mothers whose infants harbored hypermutated viral variants.

Hypermutated sequences showed on average 14.6% G-to-A substitutions: 16% among mothers and 13% among infants (not shown) with the most affected clone having 37% of Gs hypermutated. To compare the extent of hypermutation between mothers and infants, all sequences for which a nonhypermutated control clone was available were divided into two groups based on the percent Gs substituted by As (Table 2). Using the average extent of hypermutation in the infants (13%) as a cutoff, the proportion of hypermutated C2–C5 *env* sequences in which at least 13% of Gs were substituted was significantly higher among mothers than among infants (Fisher's exact $p = 0.034$) (Table 2). In addition, the mean hypermutation level among mothers was significantly higher than the one observed among infants (*t* test $p < 0.0224$). This indicates that although hypermutation can be detected shortly after birth, its extent is lower than that found in sequences isolated from adults. For nine patients (six mothers and three infants), more than one clone was included in this analysis since such clones showed different extents and distributions of hypermutations. A high extent of hypermutation seemed to be associated with a preference for a GpG dinucleotide context (Fisher's exact $p = 0.047$) (Table 3), however, nonparametric rank-sum analysis did not confirm such association ($p = 0.186$).

No relationship was found between the percent hypermutated Gs and viral subtype or transmission status of the mother (not shown). There were 22 patients (18 mothers, 4 infants) from whom more than one hypermutated clone was amplified, and at least nine of them had hypermutated sequences with a different extent and/or dinucleotide context of G-to-A substitutions. In one newborn, hypermutated clones were amplified at birth but not at 14 weeks. Another infant harbored hypermutated sequences both at 14 and at 50 weeks, but the percent of Gs substituted had dropped from 15% to 8% and the dinucleotide context had switched from GpG to GpA, respectively. As observed in the full-length hypermutated genomes, most C2–C5 *env* sequences had G-to-A substitutions in the GpG context. The pro-

Table 1. Frequency of Hypermutated HIV-1 *env* Sequences in Mothers and Infants According to HIV-1 Subtype

	Infants			*Mothers*		
Subtype	*Number of clones*	*Number of infants (clones per infant)*	*Number of infants with hypermutated clones (%)*	*Number of clones*	*Number of mothers (clones per mother)*	*Number of mothers with hypermutated clones (%)*
A	549	99 (5.5)	14 (14.0)	569	76 (7.5)	12 (15.8)
C	424	71 (6.0)	8 (11.3)	290	56 (5.2)	10 (17.9)
D	279	51 (5.5)	7 (13.8)	212	47 (4.5)	9 (19.1)
Rec	127	22 (17.3)	3 (13.6)	177	20 (11.3)	6 (30.0)
Total	1379	243	32 (13.2)	1248	199	37 (18.6)

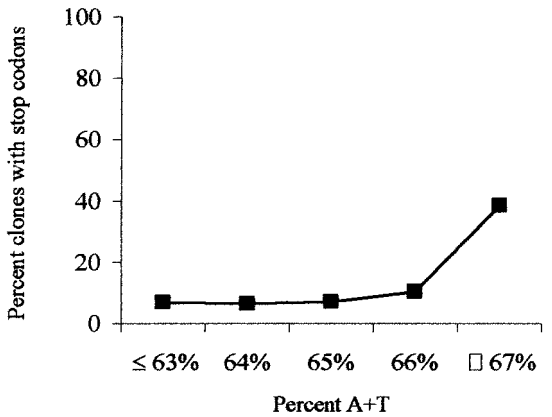


FIG. 1. Proportion of clones with stop codons among C2–C5 *env* sequences with different percent A+T content.

portion of clones with a GpG versus GpA preference for hypermutation did not differ significantly among mothers and infants or among subtypes (data not shown). Two infants and five mothers harbored both types of hypermutated HIV-1 clones, with a GpG and a GpA preference for G-to-A substitutions.

Hypermutation of full-length HIV-1 genomes

Two full-length clones from samples TZBFL0169, TZBFL0170, and TZBFL0210 were available for analysis. TZBFL0169 and TZBFL0170 are HIV-1 subtype A and D recombinant viruses, and the genomic position of their crossover points has been described elsewhere.[26] TZBFL0210 is a triple recombinant between HIV-1 subtypes A, D, and C with at least 16 crossover points (I. Koulinska, M. Essex, and B. Renjifo, unpublished data). Clones obtained from the same infant had the same recombinant subtype pattern.

Figure 2 shows the hypermutation frequency across genetic segments of each of the six clones compared to a consensus sequence constructed to match the respective subtype recombination pattern. A previously described full-length nonhypermutated subtype A and D recombinant genome from the same population, sample TZBFL0011, is shown for comparison[26] (Fig. 2). Sequences obtained from the same patient had different extents and distributions of hypermutations. This was pronounced in sample TZBFL0210, while the TZBFL0169 and TZBFL0170 clones had a similar percent of Gs substituted in most of the genome, except in *gag*. Overall, hypermutation occurred two to three times more frequently in the GpG context than in the GpA context (not shown). Even though many runs of Gs were substituted across the genome, none of the long G runs in the central and 3′ PPTs was hypermutated. Clone TZBFL0210 (1–2) had a three-base deletion (AAG) in the 5′ PPT, while the 3′ and 5′ PPTs in the other samples were intact. As previously reported for hypermutated HIV-1 sequences, short deletions, insertions, and in-frame stop codons were common in all clones.[3,36,37] Based on the consensus reference sequences, the average percent G-to-A substitutions in the complete genome ranged from 10% in TZBFL0210 (1–2) to 17% in TZBFL0210 (1–8) and did not equally affect all genomic segments.

The genes with the highest level of hypermutation were *pol* and *env*, but both TZBFL0169 clones and TZBFL0210 (1–2) had similarly high levels in the nef-LTR region (Fig. 2). Noticeably, all samples showed a drop in the extent of hypermutation between the central PPT and the beginning of *env* when compared to the neighboring *pol* and *env*. In fact, TZBFL0210 (1–2) had no G-to-A substitutions in this segment according to the criteria for hypermutation. Although the distribution of G-to-A substitutions followed the same pattern in TZBFL0169 (1–2), this was the only clone with no statistically significant difference between subgenomic regions (Fig. 2). The genomes of TZBFL0170 (3–3) and TZBFL0210 (1–8) had a similar distribution of hypermutations with the lowest levels after the central and 3′ PPT and the highest levels in *pol* and *env*. In contrast, hypermutation after the 3′ PPT was almost as high as in *pol* for both TZBFL0169 clones and was the highest for TZBFL0210 (1–2). Interestingly, the G-to-A substitutions in the TZBFL0170 (3–2) genome were almost exclusively in *pol* and *env*, where they had a pattern very close to the one observed in TZBFL0170 (3–3). Both TZBFL0169 clones and clone TZBFL0210 (1–8) had stop codons in the *vif* reading frame as a result of G-to-A substitutions relative to the reference sequence, however, the remaining three clones seemed to have a nondefective *vif* gene.

To investigate whether hypermutated viruses were predominant in the infants from which only full-length clones were isolated (TZBF0169, TZBFL0170, and TZBFL0210), multiple subgenomic *pol* clones obtained in a separate PCR were analyzed. All 16 *pol* sequences from TZBFL0169 were hypermutated and displayed two patterns of frequency and distribution of G-to-A substitutions (Fig. 3). In contrast, three of eight *pol* clones from TZBFL0170 were not hypermutated while the remaining five sequences showed variable degrees of G-to-A substitutions. Interestingly, hypermutation affected the same G residues in the 5′ half of four TZBFL0170 clones, and for three of them the overall difference in its extent was due to the stepwise progression of G-to-A substitutions into the 3′ half (Fig. 3). In sample TZBFL0210, only 2 of 18 *pol* clones were hypermutated in addition to the previously amplified TZBFL0210

TABLE 2. EXTENT OF G-TO-A HYPERMUTATIONS AMONG HIV-1 *env* SEQUENCES ISOLATED FROM MOTHERS AND INFANTS IN TANZANIA[a]

Percent G-to-A substitutions	*Infants*	*Mothers*	*Total*
High (>13%)	11	22	33
Low (≤13%)	24	16	40
Total[b]	35	38	73

[a]Fisher's exact p-value = 0.034.

[b]Includes all clones with different extents and distributions of hypermutation even if they belonged to the same patient. There were three infants and four mothers with two clones and two mothers with three clones that fell into this category. Seven mothers with no nonhypermutated clones available as a reference were excluded from this analysis.

TABLE 3. DINUCLEOTIDE CONTEXT OF G TO A HYPERMUTATION IN HIV-1 *env* SEQUENCES FROM MOTHERS AND INFANTS IN TANZANIA[a]

Percent G-to-A substitutions	*GpA*	*GpG*	*Total*
High (>13%)	7	26	33
Low (≤13%)	18	22	40
Total	25	48	73

[a]Fisher's exact p-value = 0.047.

(1–8) and (1–2). Similar to the four *pol* clones from TZBFL0170, TZBP0210 (2–18) had the same G residues substituted in the 5′ half of the amplified sequence, but the 3′ end was not affected.

DISCUSSION

This is the first report of HIV-1 hypermutation in newborns and infants and the first description of hypermutated full-length recombinant HIV-1 viruses. The analyses show that G-to-A hypermutation can be detected at birth and very soon after vertical transmission and that its distribution across the HIV genome is not random but may follow a specific pattern and restrictions.

Hypermutated C2–C5 *env* clones were detected in 13.2% of infants and 18.3% of mothers. However, such frequency is most probably an underestimate since it was based on random sequencing of clones that might have missed mutant variants, which represent only a minority of the circulating quasispecies. The design of new methods to detect G-to-A substitutions in proviral DNA has been recently described and their application led to the detection of hypermutants in at least 43% (23 of 53) of the patients studied.[9]

The existence of more than one type of hypermutant in some infants suggests that they might have originated after transmission. Generation and passive amplification of hypermutants already integrated in the host lymphocytes may be common in infants due to the rapid expansion of the $CD4^+$ pool in the first year of life and the increased chance of cell activation and infection to occur simultaneously.[38,39] The detection of hypermutated sequences at delivery suggests that they were either generated during *in utero* infection and/or that they were transmitted from the mother. MTCT of a free HIV-1 hypermutated virus might be possible after cell superinfection with a nondefective viral variant and packaging of the hypermutated viral RNA. The lower extent of hypermutation observed in infants as compared to mothers may reflect differences in the cells being infected and/or differences in the packaging efficiency of Vif/APOBEC3G in the virions.[40,41] In addition, the possibility of accumulation of hypermutations over time in a defective variant packaged after coinfection with a helper virus cannot be totally excluded.

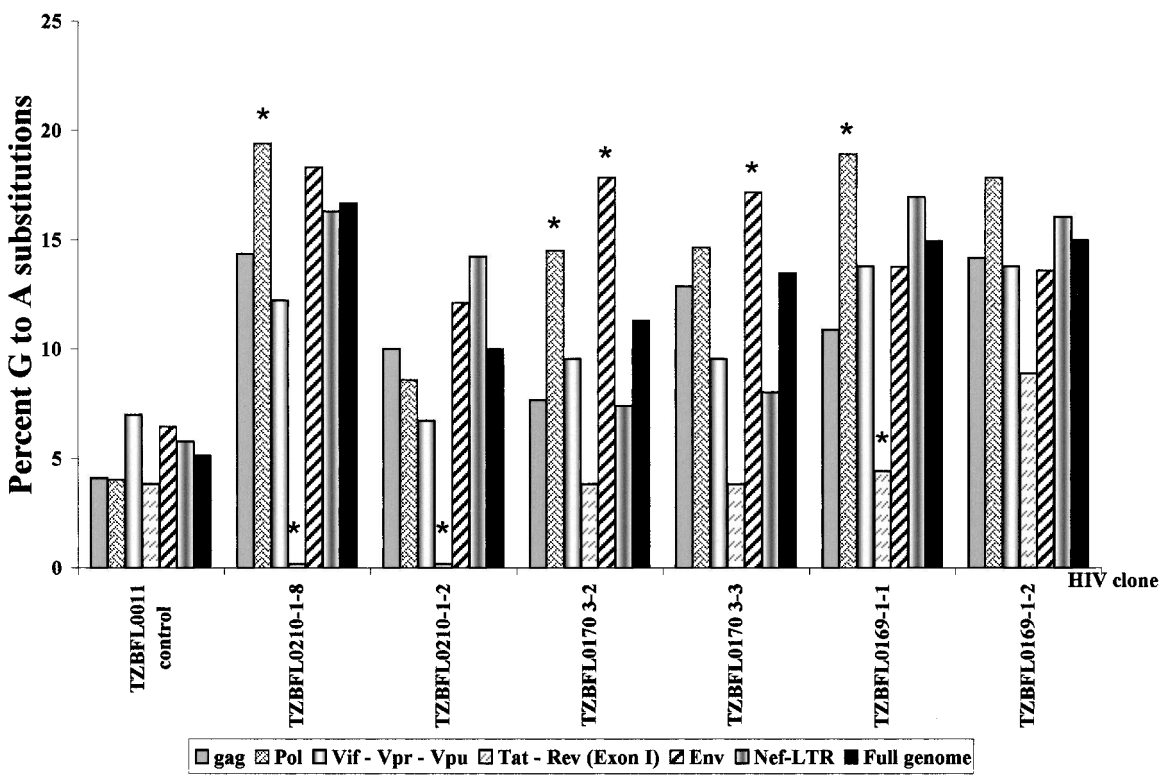


FIG. 2. Distribution of G-to-A substitutions across the genomes of six hypermutated full-length HIV-1 clones. A nonhypermutated virus, TZBFL0011, is shown for comparison. Statistically significant differences in the extent of hypermutation between neighboring subgenomic regions are shown by an asterisk (two-sample test of proportions).

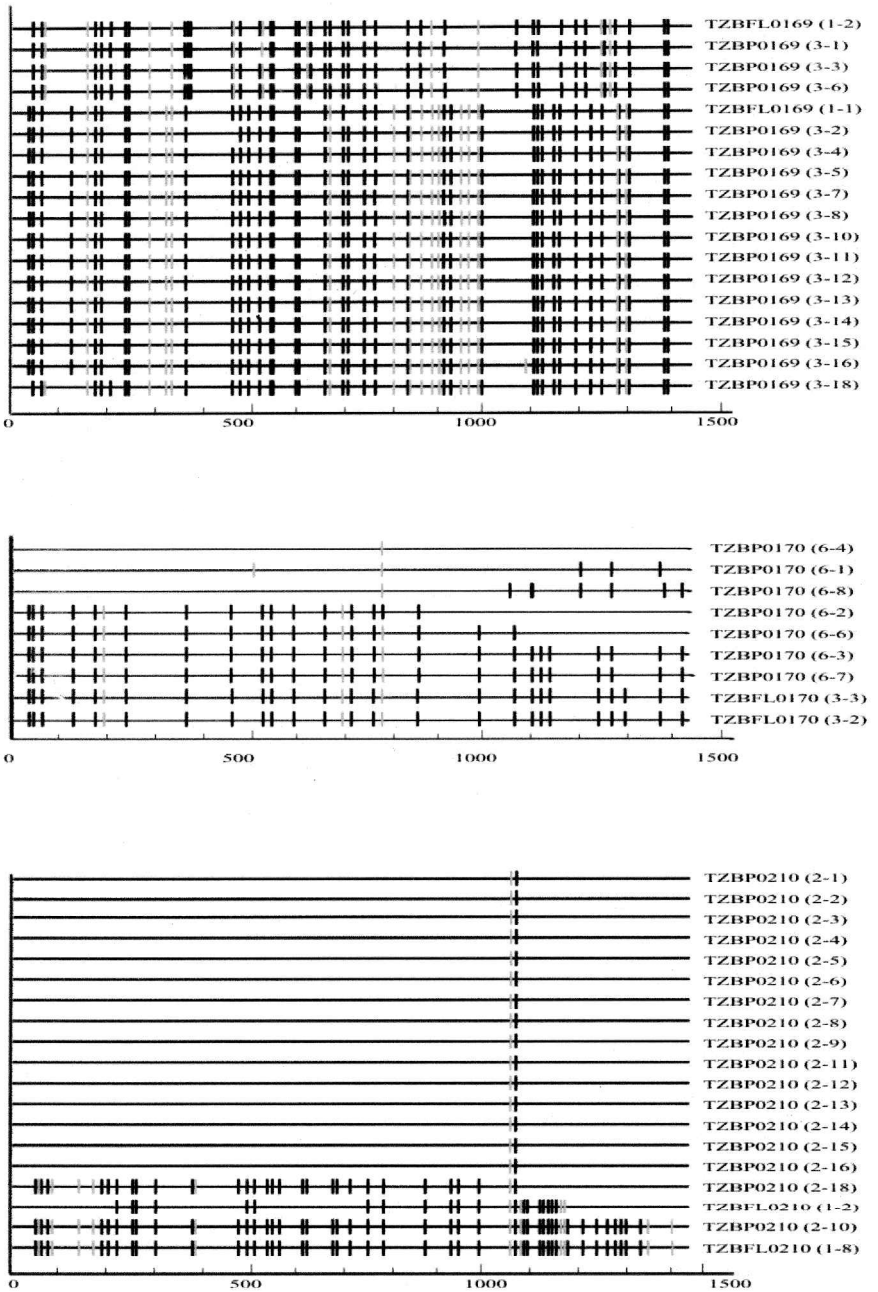


FIG. 3. Distribution of G-to-A hypermutations among subgenomic *pol* sequences from three HIV-1 viruses isolated from infants in Tanzania. A nonhypermutated clone was used as a reference for samples TZBFL0170 and TZBFL0210. A consensus sequence was used as a reference for sample TZBFL0169 since all TZBP0169 *pol* clones were hypermutated. Hypermutated *pol* sequences from each of the full-length clones of samples TZBFL0169, TZBFL0170, and TZBFL0210 are also shown for comparison. G-to-A substitutions occurring in a GpG context are in black, G-to-A substitutions occurring in a GpA context are in gray.

Even though it is possible that HIV subtypes differ in the efficiency of the Vif-APOBEC3G interaction and/or in the error rate of reverse transcription in the presence of a dNTP imbalance, the proportion of hypermutated viruses among HIV-1 subtypes A, C, and D was similar in our study. A tendency for increased frequency of hypermutation among HIV-1 subtype D viruses has been recently suggested, however, it was based on observations in only three HIV-1 subtype D-infected patients.[9]

In contrast to most published reports, hypermutation occurred predominantly in the GpG as opposed to the GpA context in the majority of the infant and maternal sequences.[3,4,9,21] Both GpG and GpA preference mutants were found to coexist in at least seven cases. Given that HIV-1 viruses isolated from

infants are M-tropic[42] and that both GpG and GpA preference mutants were found among them, the dinucleotide context of hypermutation is unlikely to be restricted by M or T cell tropism. Subtle differences in dNTP ratios at the time of viral cell entry and the relative proportion of resting versus activated cell targets in the host might influence the context of G-to-A substitutions. Alternatively, the GpG or GpA preference for hypermutation may depend on the specificity of cytidine deaminases or necessary cofactors.

All genes in the HIV-1 genome were affected by G-to-A substitutions but there was a significant drop in the extent of hypermutation between the central PPT and the beginning of *env*. It is not clear what determines this lower hypermutation level, since the G content in this region ranged from 22% to 23%, a level similar to the one of the neighboring *gag* and *env*. The gradual decrease in the extent of hypermutation after the 3′ PPT, interpreted by others as evidence for dual initiation origins during (−) DNA strand synthesis,[22] was observed in three of the six full-length clones analyzed: TZBFL0210 (1–8), TZBFL0170 (3–2), and TZBFL0170 (3–3). It is still unclear if hypermutation induced by DNA editing follows a specific pattern across the genome and how this might be related to levels of APOBEC3G or other cofactors. Variable degrees of hypermutation across the genome may also be related to fluctuations in the dNTP ratios in activated and resting cells. Productive infection of activated T cells leads to accumulation of viral DNA within the first 4–6 hr, and it is not known if the increased dNTP pools associated with cell proliferation can be maintained for the same duration.[43,44] In resting mononuclear phagocytes, substrate limitations slow but do not arrest HIV-1 reverse transcription[44] and it has been shown that mitogenic stimulation of infected resting cells induces reinitiation of DNA synthesis from partial reverse transcripts.[45] Pausing and/or reinitiation of synthesis after the central PPT might be responsible for the drop in the extent of hypermutation in the accessory gene region. This would also facilitate intrastrand RT jumps along this genomic segment, which seems to be a hot spot for recombination.[26,27]

In agreement with previous reports, the genomic regions most affected by hypermutations were *pol* and *env*, while both PPTs remained unaffected.[2,22,46] This finding might reflect a negative selection against DNA sequences with altered PPTs that may not be suitable for the initiation of positive DNA strand synthesis and for the generation of a double-stranded provirus, which is required for integration. Decreased levels of reverse transcripts and lower integration efficiency have been found with viruses hypermutated by APOBEC3G editing *in vitro*.[18] If this is the case, then hypermutation should occur much more frequently than estimated by studying proviral DNA since many hypermutated sequences would never become integrated into the cell genome.

There is ample opportunity *in vivo* for recombination, which would allow short stretches of hypermutated sequences to become incorporated into a viable background and provide an additional mechanism for viral diversification. Recombination is a possible explanation for the stepwise progression of hypermutations in the partial *pol* sequences from TZBFL0170 and for the almost exclusive presence of G-to-A substitutions in *gag* and *env* only as observed in clone TZBFL0170 (3–2). The magnitude of these effects may be an important factor in shaping the characteristic phenotypes of the different viral lineages. Excess A and G transitions are proposed to enhance the infectious potential of lentiviruses by increasing the heterogeneity of the viral population, while a preponderance of C and T transitions is correlated with the reduced infectious potential of HTLV-related oncoviruses.[47] Although it has been recently shown that hypermutation can be induced by defects in the Vif protein, three of the six full-length clones in this study had no obvious defects in the *vif* reading frame and stop codons found in the other three clones might have resulted only from hypermutation of nondefective viruses. A better understanding of the mechanism of hypermutation may also reveal novel virus–host interactions that could be targeted in drug development.

SEQUENCE DATA

All new sequences have been deposited in the GenBank Sequence Database under the following accession numbers: TZBFL0170 (3–2)—AY237166, TZBFL0210 (1–2)—AY237167, TZBFL0210 (1–8)—AY237165; TZBP0169 *pol* clones—AY231242 to AY231257, TZBP0210 *pol* clones—AY231258 to AY231275, and TZBP0170 *pol* clones—AY231276 to AY231283.

ACKNOWLEDGMENTS

This study was supported in part by Grants R35 CA39805, RO1 HD37783, and by Grant D43 TW00004 from the Fogarty International Center. We thank Wafaie Fawzi, Gernard Msamanga, and all members of the Tanzania Vitamin and HIV Infection Trial Team for their efforts in the collection and analysis of these samples and Chanc E. VanWinkle for editorial assistance.

REFERENCES

1. Wain-Hobson S: Retroviral G → A Hypermutation. In: *Human Retroviruses and AIDS 1996: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Myers G, Korber BT, Foley BT, Jeang K-T, Mellors JW, and Wain-Hobson S, eds.). Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, 1996, pp. III-57–III-63.
2. Borman AM, Quillent C, Charneau P, Kean KM, and Clavel F: A highly defective HIV-1 group O provirus: Evidence for the role of local sequence determinants in G → A hypermutation during negative-strand viral DNA synthesis. Virology 1995;208:601-609.
3. Vartanian JP, Meyerhans A, Asjo B, and Wain-Hobson S: Selection, recombination, and G → A hypermutation of human immunodeficiency virus type 1 genomes. J Virol 1991;65:1779–1788.
4. Fitzgibbon JE, Mazar S, and Dubin DT: A new type of G → A hypermutation affecting human immunodeficiency virus. AIDS Res Hum Retroviruses 1993;9:833–838.
5. Wain-Hobson S, Sonigo P, Guyader M, Gazit A, and Henry M: Erratic G → A hypermutation within a complete caprine arthritis-encephalitis virus (CAEV) provirus. Virology 1995;209:297–303.
6. Gao F, Yue L, White AT, *et al.*: Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. Nature 1992;358: 495–499.

7. Johnson PR, Hamm TE, Goldstein S, Kitov S, and Hirsch VM: The genetic fate of molecularly cloned simian immunodeficiency virus in experimentally infected macaques. Virology 1991;185:217–228.
8. Pathak VK and Temin HM: Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: Substitutions, frameshifts, and hypermutations. Proc Natl Acad Sci USA 1990;87:6019–6023.
9. Janini M, Rogers M, Birx DR, and McCutchan FE: 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 2001;75:7973–7986.
10. Huang Y, Zhang L, and Ho DD: Characterization of gag and pol sequences from long-term survivors of human immunodeficiency virus type 1 infection. Virology 1998;240:36–49.
11. Yamada T and Iwamoto A: Comparison of proviral accessory genes between long-term nonprogressors and progressors of human immunodeficiency virus type 1 infection. Arch Virol 2000;145:1021–1027.
12. Meuth M: The molecular basis of mutations induced by deoxyribonucleoside triphosphate pool imbalances in mammalian cells. Exp Cell Res 1989;181:305–316.
13. Bebenek K, Abbotts J, Roberts JD, Wilson SH, and Kunkel TA: Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. J Biol Chem 1989; 264:16948–16956.
14. Kunkel TA and Soni A: Mutagenesis by transient misalignment. J Biol Chem 1988;263:14784–14789.
15. Perrino FW, Preston BD, Sandell LL, and Loeb LA: Extension of mismatched 3′ termini of DNA is a major determinant of the infidelity of human immunodeficiency virus type 1 reverse transcriptase. Proc Natl Acad Sci USA 1989;86:8343–8347.
16. Preston BD, Poiesz BJ, and Loeb LA: Fidelity of HIV-1 reverse transcriptase. Science 1988;242:1168–1171.
17. Zhang H Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, and Gao L: The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature 2003;424:94–98
18. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, and Trono D: Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 2003;424:99–103.
19. Lecossier D, Bouchonnet F, Clavel F, and Hance AJ: Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science 2003; 300:1112.
20. Goodenow M, Huet T, Saurin W, *et al.*: HIV-1 isolates are rapidly evolving quasispecies: Evidence for viral mixtures and preferred nucleotide substitutions. J Acquir Immune Defic Syndr 1989;2: 344–352.
21. Li Y, Kappes JC, Conway JA, *et al.*: Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: identification of replication-competent and -defective viral genomes. J Virol 1991;65:3973–3985.
22. Janini M, Carr JK, Birx DR, and McCutchan FE: Hypermutation of HIV-1 genomes follows a defined pattern consistent with dual initiation points for first-strand viral DNA synthesis. AIDS Vaccine 2001 (2001, Abstr. 328).
23. Fawzi WW, Msamanga G, Hunter D, *et al.*: Randomized trial of vitamin supplements in relation to vertical transmission of HIV-1 in Tanzania. J Acquir Immune Defic Syndr 2000;23:246–254.
24. Fawzi WW, Msamanga GI, Spiegelman D, *et al.*: Randomised trial of effects of vitamin supplements on pregnancy outcomes and T cell counts in HIV-1–infected women in Tanzania. Lancet 1998; 351:1477–1482.
25. Renjifo B, Gilbert P, Chaplin B, *et al.*: Emerging recombinant human immunodeficiency viruses: Uneven representation of the envelope V3 region. AIDS 1999;13:1613–1621.
26. Koulinska IN, Msamanga G, Mwakagile D, Essex M, and Renjifo B: Common genetic arrangements among human immunodeficiency virus type 1 subtype A and D recombinant genomes vertically transmitted in Tanzania. AIDS Res Hum Retroviruses 2002; 18:947–956.
27. Koulinska IN, Ndung'u T, Mwakagile D, *et al.*: A new human immunodeficiency virus type 1 circulating recombinant form from Tanzania. AIDS Res Hum Retroviruses 2001;17:423–431.
28. Renjifo B, Fawzi W, Mwakagile D, *et al.*: Differences in perinatal transmission among human immunodeficiency virus type 1 genotypes. J Hum Virol 2001;4:16–25.
29. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, and Gibson TJ: Multiple sequence alignment with Clustal X. Trends Biochem Sci 1998;23:403–405.
30. Carr J, Foley B, Leitner T, *et al.*: Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic. In: *Human Retroviruses and AIDS 1998: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Korber B, Kuiken CL, Foley B, Hahn B, McCutchan F, Mellors JW, and Sodroski J, eds.). Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, 1998, pp. III10–III19.
31. Salminen MO, Carr JK, Burke DS, and McCutchan FE: Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res Hum Retroviruses 1995;11:1423–1425.
32. Siepel AC, Halpern AL, Macken C, and Korber BT: A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. AIDS Res Hum Retroviruses 1995;11:1413–1416.
33. Smith JM: Analyzing the mosaic structure of genes. J Mol Evol 1992;34:126–129.
34. Rose PP and Korber BT: Detecting hypermutations in viral sequences with an emphasis on G → A hypermutation. Bioinformatics 2000;16:400–401.
35. HIV-1 subtype reference alignments. HIV Sequence Database, 2002: http://hiv-web.lanl.gov/content/hiv-db/SUBTYPE_REF/align.html.
36. Pezo V, Martinez MA, and Wain-Hobson S: Fate of direct and inverted repeats in the RNA hypermutagenesis reaction. Nucleic Acids Res 1996;24:253–256.
37. Vartanian JP, Meyerhans A, Sala M, and Wain-Hobson S: G → A hypermutation of the human immunodeficiency virus type 1 genome: Evidence for dCTP pool imbalance during reverse transcription. Proc Natl Acad Sci USA 1994;91:3092–3096.
38. Age-related standards for T lymphocyte subsets based on uninfected children born to human immunodeficiency virus 1–infected women. The European Collaborative Study. Pediatr Infect Dis J 1992;11:1018–1026.
39. Denny T, Yogev R, Gelman R, *et al.*: Lymphocyte subsets in healthy children during the first 5 years of life. JAMA 1992;267: 1484–1488.
40. Kao S AH, Khan MA, Dettenhofer M, Yu XF, and Strebel K: Human immunodeficiency virus type 1 Vif is efficiently packaged into virions during productive but not chronic infection. J Virol 2003; 77:1131–1140.
41. Mariani R, Chen D, Schrofelbauer B, Navarro F, Konig R, Bollman B, Munk C, Nymark-McMahon H, and Landau NR: Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell 2003;114:21–31.
42. Matala E, Hahn T, Yedavalli VR, and Ahmad N: Biological characterization of HIV type 1 envelope V3 regions from mothers and infants associated with perinatal transmission. AIDS Res Hum Retroviruses 2001;17:1725–1735.
43. Kim SY, Byrn R, Groopman J, and Baltimore D: Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus

infection: Evidence for differential gene expression. J Virol 1989; 63:3708–3713.
44. O'Brien WA, Namazi A, Kalhor H, *et al.*: Kinetics of human immunodeficiency virus type 1 reverse transcription in blood mononuclear phagocytes are slowed by limitations of nucleotide precursors. J Virol 1994;68:1258–1263.
45. Zack JA, Haislip AM, Krogstad P, and Chen IS: Incompletely reverse-transcribed human immunodeficiency virus type 1 genomes in quiescent cells can function as intermediates in the retroviral life cycle. J Virol 1992;66:1717–1725.
46. Vartanian JP, Henry M, and Wain-Hobson S: Sustained G → A hypermutation during reverse transcription of an entire human immunodeficiency virus type 1 strain Vau group O genome. J Gen Virol 2002;83:801–805.
47. Bronson EC and Anderson JN: Nucleotide composition as a driving force in the evolution of retroviruses. J Mol Evol 1994;38: 506–532.

Address reprint requests to:
Max Essex
Harvard School of Public Health
Harvard AIDS Institute and Department of Immunology and Infectious Diseases
651 Huntington Avenue
EXR 401
Boston, Massachusetts 02115

E-mail: messex@hsph.harvard.edu