Development in vivo of genetic variability of simian immunodeficiency virus

Michael Baier*, Matthias T. Dittmar, Klaus Cichutek, and Reinhard Kurth†

Paul-Ehrlich-Institute, Paul-Ehrlich-Strasse 51-59, D-66070 Langen, Federal Republic of Germany

Communicated by Maurice R. Hillenman, June 10, 1991 (received for review March 12, 1991)

ABSTRACT Rapid development of genetic variability may contribute to the pathogenicity of lentiviruses as it may allow escape from immune surveillance and/or from suppression of virus replication. Although apathogenic in African green monkeys, simian immunodeficiency virus isolated from African green monkeys is shown to display extensive genetic variability and defectiveness in the V1- and V2-like variable domains of the envelope protein comparable to that known for human immunodeficiency virus. However, in contrast to the situation in human immunodeficiency virus-infected individuals, a predominant major virus variant was detected neither in a monkey naturally infected for more than 10 years nor in two monkeys infected with a molecular virus clone for 15–20 months. Extensive variability evolves from a single genotype with a maximal rate of 7.7 mutations per 1000 nucleotides per year. A remarkable selection for nonsynonymous mutations that accounts for 92% of all changes indicates continuous selection of variants.

Simian immunodeficiency virus (SIV) from African green monkeys (AGMs) termed SIVagm is evolutionarily related to other SIVs and to the human immunodeficiency virus (HIV) 1 and HIV-2. In contrast to other lentiviruses, SIVagm does not cause any immune dysfunction in AGMs (1–5). The reasons for this apathogenicity are unclear.

Extensive genetic variability and defectiveness of HIV in a given patient has been implicated as being essential for the induction of AIDS (6–16). The observation that dominant immunologic epitopes change during the course of infection (17) has been taken as an indication that mutation and selection as a means to evade the immune system may occur. Variability may also provide a mechanism to change regulatory proteins or the long terminal repeat sequences of the virus allowing the selection of variants with improved replicative capacity—namely, from the slow/low to the rapid/high phenotype (14, 18, 19).

With the molecular cloning of biologically active immunodeficiency proviruses, the question of variability during infection, the modes of selection pressures, and implications for disease induction can now be investigated on a more defined experimental basis (20–23). We have shown (20) that a molecular virus clone of SIVagm termed SIVagm3mc is biologically active in vitro and in vivo. The complete nucleotide sequence was determined (21) and sequence comparison to other isolates (SIVagmTYO-1mc and SIVagm155mc) revealed a specific type of intragroup genomic variability (5, 21, 23). Divergence between various SIVagm isolates is higher for both pol and env relative to gag and is especially extensive for the regulatory genes tat and rev. The variable regions V1 and V2 in env have been localized (21).

We first asked the question whether those variable domains of the env gene are indeed subject to variation in naturally infected individual animals. The SIVagm loop corresponding to the variable and immunodominant V3 loop of HIV-1 (24–26) was found to be highly conserved and was, therefore, not investigated in the context of this study. The variable domains V1 and V2, which also correspond to V1 and V2 in HIV-1 (24–26), were indeed found to undergo extensive mutation.

In the two AGMs infected on average for 18 months with SIVagm3mc, a dichotomous emergence of variant viral genomes was observed. In one animal, amino acid sequences diverged up to 6% in the various subregions sequenced; whereas, in the other animal, only single mutations were observed. This investigation allows the estimation of a mutation rate for variable domains of an immunodeficiency virus after replication in its natural host starting from a single parental sequence.

MATERIALS AND METHODS

Animals and Viruses. Infection of seronegative and PCR-negative AGMs house in our colony with molecularly cloned SIVagm3mc and preparation of monkey peripheral blood mononuclear cells (PBMCs) has been described (1, 2, 21). SIVagm3mc stock virus was grown on MOLT 4/8 T-lymphoma cells for 3 weeks and stored in liquid nitrogen, and stock virus was used for infection. Renewed PCR-sequencing excluded development of variation due to the short-term in vitro cultivation.

PCR Amplification and Molecular Cloning. DNA preparation from AGM lymphocytes was performed according to standard methods (2, 27). Samples of PBMC DNA (1 µg) were subjected to PCR amplification by using the Taq polymerase under standard buffer conditions as suggested by the supplier (Perkin-Elmer/Cetus). To avoid contamination, all preparations of PBMCs, of PBMC DNA, and the PCRs were done in separate facilities. Negative control DNA from PBMCs of uninfected animals was simultaneously prepared and checked for contamination by nested PCR amplification before use of the PBMC samples.

The first round of PCR amplification comprised up to 40 cycles with oligonucleotide primers EX3(+) (5′-CTGACAGAAGACCTTATTGTA-3′) and EX4(−) (5′-CAATGCACACTGAAAACA-3′). They spanned nucleotides 6010–6557 of the SIVagm3 env gene (21). The second round of amplification starting from 5 µl of the reaction products of the first round comprised up to 40 cycles with a second internal primer pair IN1(+) (5′-CTGTGCACACTCATTGAGCAGA-3′) and IN2(−) (5′-CCCGTCGACACTCT-9)

Abbreviations: SIV, simian immunodeficiency virus; AGM, African green monkey; PBMC, peripheral blood mononuclear cell; HIV, human immunodeficiency virus.

*Present address: Interdisciplinary Research Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, England.
†To whom reprint requests should be addressed.
RESULTS

Cloning and Sequencing of V1- and V2-Like Variable Domains of SIVagm env. Sequence comparison between the molecular virus clones SIVagm3 (21), SIVagmTYO-1 (4), and SIVagm155 (5) revealed two variable domains located between amino acids 114 and 142 and between amino acids 151 and 212 of the gp130 external envelope glycoprotein. They resemble the V1 and V2 variable regions of HIV-1 in terms of size and location (Fig. 1). The presence of B-cell epitopes within these regions was revealed by antibody binding from sera of naturally infected AGMs to overlapping peptides derived from the envelope protein of SIVagm3 (unpublished data). Correspondingly, the V1-domain of HIV-1 has been shown to be part of at least one B-cell epitope (25) and the genetic diversity of the V1 and V2 regions of HIV-1 has been analyzed in some detail (6, 11). Thus, the comparable regions of the SIVagm env gene were chosen for a variability study of SIVagm in vivo.

A nested set of oligonucleotide primer pairs was used to amplify and clone the V1- and V2-like regions of the SIVagm env gene. Initially performed sensitivity tests of the nested amplification procedure demonstrated that one lymphocyte infected in vitro was detectable among $1 \times 10^5$ uninfected cells by amplification of a specific viral DNA fragment of 300 base pairs. To determine the mutation frequency introduced by Taq polymerase errors after 80 cycles of amplification, 50 copies of SIVagm3 plasmid mixed with 1 $\mu$g of genomic DNA

![Fig. 1. Amino acid sequence alignments of SIV variant subgenomes present in the naturally infected monkey AGM-3. Depicted are sequences from amino acids 114 to 213 of the external envelope glycoprotein of SIVagm3 of single variant subgenomes molecularly cloned from PBMC DNA of AGM-3. The sequence of variant 3/15 was used as a reference because it allowed the best alignments. Frame-shift mutations leading to premature stop codons are indicated by open squares. Dots indicate complete codon deletions.](image-url)
from uninfected lymphocytes were amplified. DNA was molecularly cloned and 12 clones were sequenced. No insertions or deletions were found. Only two base substitutions, one silent and one nonsilent, were found in a total of 3.6 kilobases sequenced (data not shown). This corresponded to an error rate of \(5.6 \times 10^{-4}\) for base substitutions. Thus, the contribution of errors caused by Taq polymerase was estimated to be less than 1 in 1000 nucleotides and would, therefore, not significantly add to the extent of variability measured.

**Extensive Variability and Defectiveness of SIVagm in a Naturally Infected Animal.** To study development of SIVagm genetic variability, we first analyzed variant strains in a naturally and long-term infected animal. Monkey AGM-3 was selected for the study, because isolate SIVagm3 had been obtained 3 years earlier by cocultivation of PBMCs from this animal with MOLT 4/8 lymphocytes (1, 2, 21). Molecular virus clone SIVagm3mc was derived from this isolate.

Approximately \(1 \times 10^7\) PBMCs from monkey AGM-3 were used as starting material for DNA preparation. Cell stimulation and culturing were avoided to prevent virus replication *in vitro*. Specific DNA fragments obtained after PCR amplification were molecularly cloned, and 22 independent molecular clones were sequenced and compared to each other (2, 21).

Each clone was found to represent a unique sequence and a major virus substrain or consensus sequence could not be identified (Fig. 1). One molecular clone (3/11) was found to differ from SIVagm3 only at amino acid 164 which was asparagine instead of lysine (Fig. 1). Although highly polymorphic, the analyzed virus population was still relatively homogeneous in comparison to the heterogeneity observed between SIVagm isolates from geographically different areas, as the degree of amino acid divergence found between different molecular clones was merely 1–15%. In contrast, the amino acid divergence of SIVagm3 versus SIVagmTYO-1 and SIVagm155 is about 35% in this region (21, 29).

Ten of 22 clones (45%) were obviously derived from defective proviruses because they contained one or more frame-shift mutations leading to premature stop codons in the external envelope protein (Fig. 1). All but three of the amino acid mutations were located within the V1 and V2 regions. Previously unrecognized clusters of pronounced heterogeneity were found between amino acids 134 and 138 within V1 and between amino acids 202 and 212 within V2 (Fig. 1). In addition to base substitutions, several in-frame insertions and deletions had occurred within the V2 region, which were also found by comparison of SIVagm3mc to SIVagmTYO-1mc (data not shown; ref. 21).

**Variability and Mutation Rate of SIVagm After Experimental Infection with a Molecular Virus Clone.** The molecular virus clone SIVagm3mc is replication competent *in vitro* and *in vivo* (2). Twenty months ago two monkeys termed AGM-15 and AGM-16 were inoculated intravenously with a cell-free virus stock of SIVagm3mc (2). Until now, no signs of an immunodeficiency-like disease have been observed. Nevertheless, both monkeys displayed the characteristics of a persistent lentivirus infection, as virus could be repeatedly isolated and anti-SIVagm antibody response remained con-

---

**Fig. 2.** Amino acid sequence alignments of SIV variant subgenomes present in monkey AGM-16 infected with molecular virus clone SIVagm3mc. As in Fig. 1, amino acid exchanges versus the parental sequence of SIVagm3 from amino acids 114 to 213 of the external envelope protein are shown. #: Silent mutations.
sistantly positive over time without a decline in titer (data not shown).

Genetic analysis of viral variants was performed by PCR amplification, cloning, and sequencing of the variable regions V1 and V2 as described above. Only 2 of 20 clones representing the variable env regions of viral DNA obtained from PBMCs of AGM-16 were found to be identical (16/14 and 16/15; Fig. 2). No frame-shift mutations or in-frame stop codons were detected, indicating that accumulation of defective proviruses had not yet occurred at a detectable level. Amino acid divergence of the clones was in the range of 1–6% relative to the parental sequence of SIVagm3mc. One change of Ser-151 to Asn was found to be common to all the clones obtained and other preferred amino acid changes were also observed (Fig. 2). One of the conservative amino acid changes (namely, Arg-208→Lys) was found in 6 clones. This codon change was also predominant among the clones obtained from naturally infected monkey AGM-3 (amino acid 164, Fig. 1). All other amino acid alterations found in the viral genomes from AGM-16 were either absent or little represented in viral DNA from AGM-3.

Surprisingly, sequence analysis of viral DNA from PBMCs of AGM-15 revealed a different development of variation. Of the 23 clones sequenced 15 months after infection, 15 contained only up to two base changes and were summarized to genotype cluster 15A1 in Fig. 3. Eight clones had a 7-amino acid deletion between codons 128 and 134 in common (variants 15/2A in Fig. 3). Interestingly, this deletion was found to predominate 20 months after infection (15/B in Fig. 3). However, specific DNA amplification of the env region from PBMC DNA of this animal was difficult in comparison to AGM-16 and only occasionally successful. PCR quantitation of viral DNA in PBMCs indicated at least 5 times less viral DNA copies than in PBMCs of AGM-16 (data not shown). Therefore, due to the very low peripheral virus load, it was not possible to obtain a statistically significant representation of viral DNA clones from this animal. Hence, we did not include the data obtained from AGM-15 clones in the calculation of mutation rates given below.

Mutation Rate of Cloned SIVagm in AGM-16. The frequencies of individual mutations calculated from the sequence analysis of virus clones obtained from AGM-16 are depicted in Table 1. Only two transversions were detected among a total of 69 nucleotide alterations. Normalization of substitution frequencies for base composition demonstrated that, as has been shown for HIV-1, the most frequent base substitution was a G→A transition that accounted for all G→N mutations observed, where N is any base (Table 1; ref. 11).

The 69 nucleotide changes in a total of 20 clones averages to 3.45 nucleotide changes per clone after 18 months or 2.3 changes per year. This rate obtained with clones 300 base pairs long corresponds to 7.7 base changes per 1000 nucleotides per year. A strong preference for nonsilent mutations was observed, as 63 of the 69 base changes occurred in the first two codon positions. A rate of 0.7 base substitution per 1000 nucleotides per year for silent mutations was calculated

### Table 1. Nucleotide substitution frequencies of SIVagm3 in monkey AGM-16

<table>
<thead>
<tr>
<th>Base change</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>G→N</td>
<td>18.5</td>
<td>111.5</td>
<td>0.0</td>
<td>31.5</td>
</tr>
<tr>
<td>A→N</td>
<td>6.8</td>
<td>2.9</td>
<td>0.0</td>
<td>3.6</td>
</tr>
<tr>
<td>T→N</td>
<td>2.2</td>
<td>2.6</td>
<td>9.7</td>
<td>5.5</td>
</tr>
<tr>
<td>C→N</td>
<td>8.3</td>
<td>19.7</td>
<td>0.0</td>
<td>12.7</td>
</tr>
<tr>
<td>X→N</td>
<td>9.0</td>
<td>22.5</td>
<td>3.0</td>
<td>11.5</td>
</tr>
</tbody>
</table>

Depicted are substitution frequencies for each X→N exchange (where X is any base mutated to a different base, N) normalized for codon position and base composition. The given values are calculated by dividing the substitution frequency of a specific X→N exchange at a specific codon position by the frequency of base X at that position within the sequenced region and by the number of independent clones sequenced multiplied by 1000 (11). The base composition of the analyzed part of the env gene (300 bases) is as follows: A, 36.7%; C, 18.5%; G, 20.3%; T, 24.7%.

from six synonymous codon mutations in the 20 clones analyzed. This low silent mutation rate may reflect the mutation rate that would be observed in the absence of any selection pressure.

### DISCUSSION

HIV and SIVmac from rhesus macaques (Macaca mulatta) are genetically related lentiviruses that persistently infect their hosts and induce immunosuppression after a latency period of months or years. Genetic variability localized in the variable domains V1 to V5 of the external envelope glycoproteins of these viruses may allow escape from suppression of virus replication by immunologic or cellular mechanisms and may thus contribute to disease induction (12, 17, 20). The experiments described above were set out to understand whether variability of the apathogenic SIVagm is comparable with the variability of the pathogenic lentiviruses.

Our study of variant subgenomes of apathogenic SIVagm after natural and clonal infections proves that variability and defectiveness of the viral DNA in PBMCs can also be a hallmark of apathogenic lentiviruses. These characteristics of a natural infection with lentiviruses are, therefore, not correlated with immunosuppression per se. We were able to demonstrate that every variant virus developing from a single input virus is unique and has a different genotype than the original virus. Thus, genetic variability inherently evolves in a lentivirus population and is not necessarily a result of an infection with multiple virus substrains. The overall mutation rate of 7.7 mutations per 1000 nucleotides per year accounts for an average divergence of 6.5% between variant subgenomes in AGM-16 after an in vivo replication period of 18 months. The surprisingly low overall divergence of only 15% between the variable env regions detected in AGM-3 more than 10 years after natural infection may accordingly also be the consequence of an infection either with a single virus strain or with related variants.

![Fig. 3. Amino acid sequence alignments of mutated regions of SIV variant subgenomes present in monkey AGM-15 infected with molecular virus clone SIVagm3mc. Variant subgenomes from AGM-15 are grouped into two clusters (15/A1 and 15/A2 versus 15/B). Clones A were obtained 15 months and clones B 20 months after experimental infection. Codon deletions are indicated by bars. #, Number of clones/total number of sequenced clones.](image-url)
As defective subgenomes were not found after clonal infections of up to 20 months in both AGM-16 and AGM-15, generation of defectiveness may either require longer replication periods in vivo or defectiveness is already introduced during the natural infection. The high percentage of defective subgenomes in AGM-3 may also be explained by assuming that defective viral DNA persists in circulating blood cells as a relic of former nonproductive infections, whereas replication competent proviruses persist for shorter time periods because of virus-induced lysis of infected cells and the immune response to cells expressing virus proteins.

The overall mutation rate of SIVagm in AGM-16 of 7.7 mutations per 1000 nucleotides per year is similar to the mutation rates estimated for HIV or SIVmac, which range from 2 to 12 mutations per 1000 nucleotides per year in various variant env regions (11, 12, 20, 28). However, a 92% bias of mutations in favor of amino acid changes indicates a continuous selection for virus variants that may be better adapted for in vivo growth. This selection seems to comprise the two variable regions V1 and V2, in contrast to the short epitope containing V3 region where a 100% bias of mutations in favor of amino acid changes was previously found for SIVmac (20). In contrast to the situation in HIV-1-infected individuals, major SIVagm variants could not be identified, even after long-term natural infection. In the absence of neutralizing antibodies (30), continuous selection during the in vivo replication of SIVagm may be due to cellular immune surveillance or yet unidentified effective intracellular suppression of SIVagm replication.

We thank Mrs. Y. Buckendahl for excellent technical assistance; Drs. R. Seinsche, J. Löwer, J. Denner, J. Ennen, H. König, and S. Norley for helpful discussions; Dr. S. Hartung for veterinary supervision; and Mrs. B. Brandi for editorial assistance. This study was supported by grants from the Federal Ministry for Research and Technology and the Federal Ministry of Health of the Federal Republic of Germany.