VIROLOGY 199, 323–330 (1994)

Fidelity of HIV-1 Reverse Transcriptase Copying a Hypervariable Region of the HIV-1 *env* Gene

JIUPING JI AND LAWRENCE A. LOEB[1]

The Joseph Gottstein Memorial Cancer Research Laboratory, Department of Pathology SM-30, University of Washington, Seattle, Washington 98195

Received July 9, 1993; accepted November 3, 1993

The unusually high mutation frequency exhibited by the human immunodeficiency virus (HIV) is a major impediment to developing effective vaccines against the virus and to designing nucleotide analogs that inhibit viral replication. To investigate the molecular basis of HIV hypermutability, we established cell-free assays to measure the fidelity of HIV-1 reverse transcriptase (RT) in copying either DNA or both RNA and DNA templates that contain the hypervariable region 1 of the HIV-1 *env* gene (V-1). The fidelity of DNA synthesis was measured by repetitively copying the envelope gene (V-1) DNA by HIV-1 RT, followed by cloning and sequencing these newly synthesized DNA products. We found that the error rate of HIV RT copying either RNA or DNA of the *env* V-1 region is about one misincorporation per 5 kb polymerized. This rate is similar to that found with the M13mp2 forward mutation assay using the *lacZ*α gene as a template. This similarity suggests that the HIV *env* hypervariable sequence is not inherently hypermutable. The high error rate of HIV RT suggests that misincorporation by this enzyme is a major source of mutations throughout the viral genome and a determinant for rapid viral evolution. The spectrum of mutations produced by HIV RT *in vitro* partially correlates with the spectrum of HIV mutations observed in AIDS patients. The differences between these spectra highlight the contribution of phenotypic selection during HIV-1 infection. The overall uniformity of misincorporation of HIV-1 RT further suggests an alternative anti-HIV strategy based on increasing viral mutagenesis by nucleotide analogs.

Understanding the basis for the hypermutability of the human immunodeficiency virus (HIV) may be crucial for managing the acquired immunodeficiency syndrome (AIDS) (Coffin, 1986). HIV is a retrovirus with a 9.7-kb single-stranded RNA genome that codes for the major viral proteins including the reverse transcriptase (RT), pol, and the envelope protein, env (Gallo and Jay, 1991). Viral isolates from individual patients are not clonal; each isolate contains a population of closely related yet genetically distinct genomes, referred to as a quasispecies (Eigen *et al.*, 1989). This extensive intraclonal variation is likely to be the result of the high mutation rate of HIV. It is estimated that the error rate of HIV is approximately one mutation per replication per genome, the frequency of mutations is approximately a millionfold higher than that exhibited by human cells (Hahn *et al.*, 1986; Goodenow *et al.*, 1989; Nowak *et al.*, 1991).

Sequencing of the HIV-1 genome from AIDS patients reveals that the most extensive variation is in the *env* gene, in which four hypervariable regions have been delineated (Coffin, 1986; Wiley *et al.*, 1986; Myers *et al.*, 1989). The mutations are predominantly single base substitutions (Goodenow *et al.*, 1989; Vartanian *et al.*, 1991), suggesting that they arose as a result of misincorporations during RNA or DNA synthesis (Preston *et al.*, 1988). The nucleotide polymerization steps in the HIV life cycle include: (1) reverse transcription of the retroviral RNA genome by viral RT; (2) plus-strand DNA synthesis by RT copying the newly synthesized viral DNA; (3) DNA replication by mammalian cell DNA polymerases after viral DNA integration into the host genome, and (4) transcription of proviral DNA by host-cell RNA II polymerase. The error rate of HIV RT measured with the M13 or 0X174 fidelity assay is $1–5 \times 10^{-4}$, while the error rates of DNA polymerase α, β, or ϵ, the major host-cell polymerases, are at least one order of magnitude more accurate (Roberts *et al.*, 1988, 1989; Bebenek *et al.*, 1989; Perrino *et al.*, 1989; Weber and Grosse, 1989; Boyer *et al.*, 1992; Ji and Loeb, 1992). The fidelity of host RNA polymerase II is unknown. However, it has been demonstrated that the RNA polymerases from *Escherichia coli* as well as T3 or T7 bacteriophages are highly accurate (Blank *et al.*, 1986; Boyer *et al.*, 1992; Hubner *et al.*, 1992; Ji and Loeb, 1992). Therefore, the most error-prone polymerization steps in the HIV life cycle are likely to be viral reverse transcription and/or plus-strand DNA synthesis, both catalyzed by HIV RT.

The types of mutations found in the *env* gene of the HIV are unusual. Vartanian *et al.* (1991) reported that

[1] To whom reprint requests should be addressed.

0042-6822/94 $6.00

HIV mutations consisted almost exclusively of GC → AT transitions after multiple passages of HIV on lymphocytes and monocytes. Pathak and Temin (1990) analyzed the mutations produced during a single cycle of replication of spleen necrosis retrovirus and observed clusters of 15 GC → AT substitutions in a stretch of the 990-base genome in about 1 of 20 viral progeny. Both of these mutations were classified as retroviral hypermutations and proposed to be the direct results of misincorporation during viral reverse transcription (Pathak and Temin, 1990; Vartanian *et al.*, 1991). However, GC → AT transitions are not the most frequent mutations observed when HIV-1 RT is used to copy the *lacZα* gene (Boyer *et al.*, 1992; Ji and Loeb, 1992). Because the frequency and spectrum of mutations can be highly dependent on sequence context (Benzer, 1961; Ricchetti and Bue, 1990; Doi, 1991; Ji and Mathews, 1991; Yu and Goodman, 1992), we asked whether the nucleotide sequence of the *env* gene was inherently mutagenic and whether copying of these sequence by RT generated predominantly GC → AT transitions.

Using a cell-free assay for HIV RT fidelity that mimics the replicative steps in the viral life cycle, we have identified the errors introduced by HIV RT during DNA-dependent as well as RNA-dependent DNA polymerization in the *env* hypervariable region 1 (V-1). The frequency and types of errors induced by HIV RT could account for the high mutability and rapid evolution of the HIV genome. We find that the nucleotide sequence of the *env* gene is not inherently hypermutable and that both GC → AT and AT → GC transitions are the most frequent base substitutions. We conclude that the enhanced localization of mutations to the envelope gene and the specificity of these mutants *in vivo* is determined by selection and factors other then misincorporation by HIV RT.

MATERIALS AND METHODS

Enzymes and chemicals

HIV-1 RT was purified as previously described (Preston *et al.*, 1988; Ji and Loeb, 1992). *Bam*HI was from United States Biochemical Corp. T4 polynucleotide kinase and *Eco*RI were from New England BioLabs. T7 RNA polymerase and *Hin*dIII were from Promega. *Taq* polymerase was from Boehringer Mannheim Corp. T4 ligase was from BRL. The deoxyoligonucleotide primers used for extension reactions, DNA amplification, and sequencing were synthesized by Operon Technologies, Inc. Deoxy- and dideoxyribonucleoside triphosphates (dNTPs and ddNTPs), dATP/ddATP, dGTP/ddGTP, dCTP/ddCTP, and dGTP/ddGTP were from Sigma. Ribonucleoside triophosphates (rNTPs), rATP, rGTP, rCTP, and rGTP were from Promega. [α-^{32}P]ATP (3000 Ci/mmol) was from Amersham Corp.

Bacteria, bacteriophage, and plasmids

E. coli MC1061 [*hsdR, mcrB, araD, 139Δ-(araABC-leu), 7679ΔlacX74, galU, galK, rpsL, thi*] was the transformation host. An F′ derivative of *E. coli* CSH50 [*Δ(pro-lac) thi ara strA/F′ (proAB lacIa-zΔM15)*] was used as an indicator for the transfected *E. coli*. Both MC1061 and CSH50 were provided by T. A. Kunkel (NIEHS). Bacteriophages M13mp18 and M13mp19 were from this laboratory. The TA Cloning kit was from Invitrogen. pME235, a pBluescript KS$^+$-derivative plasmid containing the entire coding sequences of gp120 from an HIV-1 BRU isolate, was kindly provided by M. Emerman (FHCRC, Seattle).

Construction of the nucleotide template

The *env* V-1 (317–533 from the *env* translational start site in HIV-1 BRU isolate) in pBluescript KS$^+$-derivative plasmid pME235 was PCR-amplified. Two synthetic oligonucleotide primers, Olig 1 (AAT TTA ATA CGA CTC ACT ATA GGG ATA TAA TCA GTT TAT GGG ATC) and Olig 2 (AAT TTA ATA CGA CTC ACT ATA GGG ATA TTC TTT CTG CAC CTT ACC) were used. Each primer contains a T7 RNA polymerase binding sequence and the preferred transcriptional initiation site at a 5′ terminus, as well as the HIV-1 *env* sequence complementary to the regions 317–339 and 513–533 from the *env* translational start site. The reaction was conducted in 100 μl containing 10 m*M* Tris–HCl (pH 8.3), 50 m*M* KCl, 1.5 m*M* MgCl, 10 μg gelatin, 0.5 m*M* of each of the four dNTPs, 100 ng pME254 DNA, 20 pmol of each primer, and 2 μg *Taq* polymerase. The DNA was amplified for 20 PCR cycles consisting of 1 min at 95°, 1 min at 54°, and 1.5 min at 72°. An aliquot of the PCR product was cloned into the plasmid pCR1000 using the TA cloning kit (Invitrogen). A DNA clone contained the *env* V-1 fragment was identified by sequencing using an M13 universal primer (GTA AAA CGA CGG CCA GT). Ten micrograms of plasmid DNA were incubated at 37° for 60 min in a 5-μl reaction containing 50 m*M* Tris–HCl (pH 8.0), 10 m*M* $MgCl_2$, 100 m*M* NaCl, and 10 μg *Eco*RI. The digested DNA was precipitated with ethanol and then incubated in 50 μl of reaction mixture containing 6 m*M* Tris–HCl (pH 7.5), 6 m*M* $MgCl_2$, 50 m*M* NaCl, 1 m*M* DTT, 10 μg bovine serum albumin, and 10 μg *Hin*dIII. The *Eco*RI/*Hin*dIII-digested 350-bp *env* fragments were resolved by 1.5% agarose gel electrophoresis, extracted from the gel with phenol/chloroform, and precipitated with ethanol. The DNA was dissolved in 10 m*M* Tris–HCl/1 m*M* EDTA (pH 8.0) and used as a template for RT-mediated amplification.

Amplification from DNA template

An HIV RT-mediated PCR was used to synthesize double-stranded DNA. Two internal *env* primers were used: Olig 3 with a *Bam*HI recognition sequence at the 5′ terminus (ATG GGA TCC AAG CCT AAA GCC ATG TG) and Olig 4 with an *Eco*RI recognition sequence at the 5′ terminus (GGA ATT CTT TCT GCA CCT TAC CTC). The reaction was conducted in 50 μl containing 60 m*M* Tris–HCl (pH 8.2), 1.0 m*M* dithiothreitol, 7 m*M* $MgCl_2$, 0.5 m*M* EDTA, 10 m*M* KCl, 500 μ*M* each of the four dNTPs, 20 ng template, and 2 pmol of each primer. The DNA was amplified by 25 PCR cycles of 2 min at 95°, 2 min at 25°, and 3 min at 37° with the addition of 30 μg HIV-1 RT prior to incubation at 37° during each cycle.

Amplification from RNA/DNA templates

HIV RT-catalyzed DNA amplification using both RNA and DNA template was modified from the procedure described by Guatelli *et al.* (1990). The reaction was conducted at 37° for 60 min in 50 μl containing 60 m*M* Tris–HCl (pH 8.2), 20 ng *env* *Eco*RI/*Hin*dIII V-1 fragment, 10 m*M* $MgCl_2$, 10 m*M* KCl, 2 m*M* spermidine–HCl, 2.5 m*M* dithiothreitol, 0.5 m*M* each of dATP, dCTP, dGTP, and dTTP, 2 m*M* each of ATP, CTP, GTP, and UTP, 20 pmol each of primers Olig 1 and Olig 2, 90 μg of HIV-1 RT, 100 μg of T7 RNA polymerase, and 2 units of *E. coli* RNase H. The exponential DNA accumulation at constant temperature during amplification was monitored by agarose gel electrophoresis and scintillation counting.

Identification of RT errors

After amplification of the *env* gene by HIV RT, the target was reamplified with *Taq* DNA polymerase for 5 cycles for higher cloning efficiency using Olig 3 and Olig 4 to introduce *Eco*RI and *Bam*HI cleavage sites. Both the amplified *env* V-1 fragment from DNA template or RNA/DNA template and the double-stranded M13mp18 and mp19 vector DNAs were digested with *Bam*HI and *Eco*RI by incubation at 37° for 60 min in a 50-μl reaction mixture containing 50 m*M* Tris–HCl (pH 8.0), 10 m*M* $MgCl_2$, 100 m*M* NaCl, and 10 μg each of *Eco*RI and *Bam*HI. To ensure complete digestion, an additional 10 μg each of *Eco*RI and *Bam*HI were added and the incubation was repeated. The DNA was extracted with phenol/chloroform and precipitated with ethanol. Nonelongated oligonucleotide primers were removed by Centricon-30 filtration (Amicon). The *env* fragment was ligated into an M13 vector by incubation at 16° for 12 hr in a 10-μl reaction containing 50 m*M* Tris–HCl (pH 7.6), 10 m*M* $MgCl_2$, 1 m*M* ATP, 1 m*M* dithiothreitol, 5% polyethylene glycol-8000, and 2 μg T4 DNA ligase. *E. coli* MC1061 cells were transformed with the ligated DNA using a Gene-pulser electroporator (Bio-Rad). The individual viral plaques were isolated and DNA was sequenced as described (Ji and Loeb, 1992).

RESULTS

Fidelity of HIV-1 RT copying *env* V-1 DNA

In order to study the fidelity of HIV-1 RT in copying the *env* V-1 DNA sequence, we developed a cell-free assay based on an HIV RT-mediated PCR (Fig. 1A). A DNA fragment containing the *env* V-1 sequence was copied repetitively by HIV RT. Unlike the thermostable *Taq* polymerase, HIV RT is thermolabile and must be added after each primer annealing step. The product of the reaction was ligated into an M13 DNA vector and then transfected into *E. coli*. Clones containing the nucleotide insert copied by HIV RT were isolated and the sequence of nucleotides within the *env* V-1 fragment was established.

The types of mutations obtained after 25 cycles of the HIV RT-mediated PCR are shown in Fig. 2. From 160 individual clones that were isolated and sequenced, we identified 64 mutations at 36 different positions along the 153-base *env* V-1 insert. The predominant errors introduced by HIV RT in copying the DNA template were single-base substitutions; transitions represented 76% of these. The most frequent substitutions were the GC → AT transitions at position 13 and the AT → GC transitions at position 141. Some of single-base substitutions that were observed are compatible with the dislocation mechanism proposed by Kunkel (1985), wherein a substitution is initiated by a template–primer transient misalignment with the next downstream complementary base. Substitutions fulfilling this criterion include mutations occurring at positions 48, 93, and 133. The other substitutions presumably result from direct misinsertions opposite the noncomplementary template nucleotides. Seventeen of 64 errors identified were frameshift mutations opposite the polyadenine tract between positions 119 and 124; 12 of these were +1 additions, 3 were +2 additions, and 2 were −1 deletions. The fidelity of HIV RT in copying the *env* V-1 DNA is both sequence- and position-dependent. The error rates for different types of substitutions range from 1/8100 to 1/120,000, in the order of GC → AT > AT → GC > frameshift > transversion (Table 1). The distribution of different types of errors does not appear to be random from this limited data set; for example, the addition frameshift error rate around positions 119 to 124 is about 1/890, while the average frameshift error rate is about 1/19,000.

Fidelity of HIV-1 RT copying RNA and DNA

The fidelity assay for measuring the ability of HIV RT to copy both RNA and DNA templates (RNA/DNA) (Fig. 1B) is modified from the method of Guatelli *et al.* (1990)

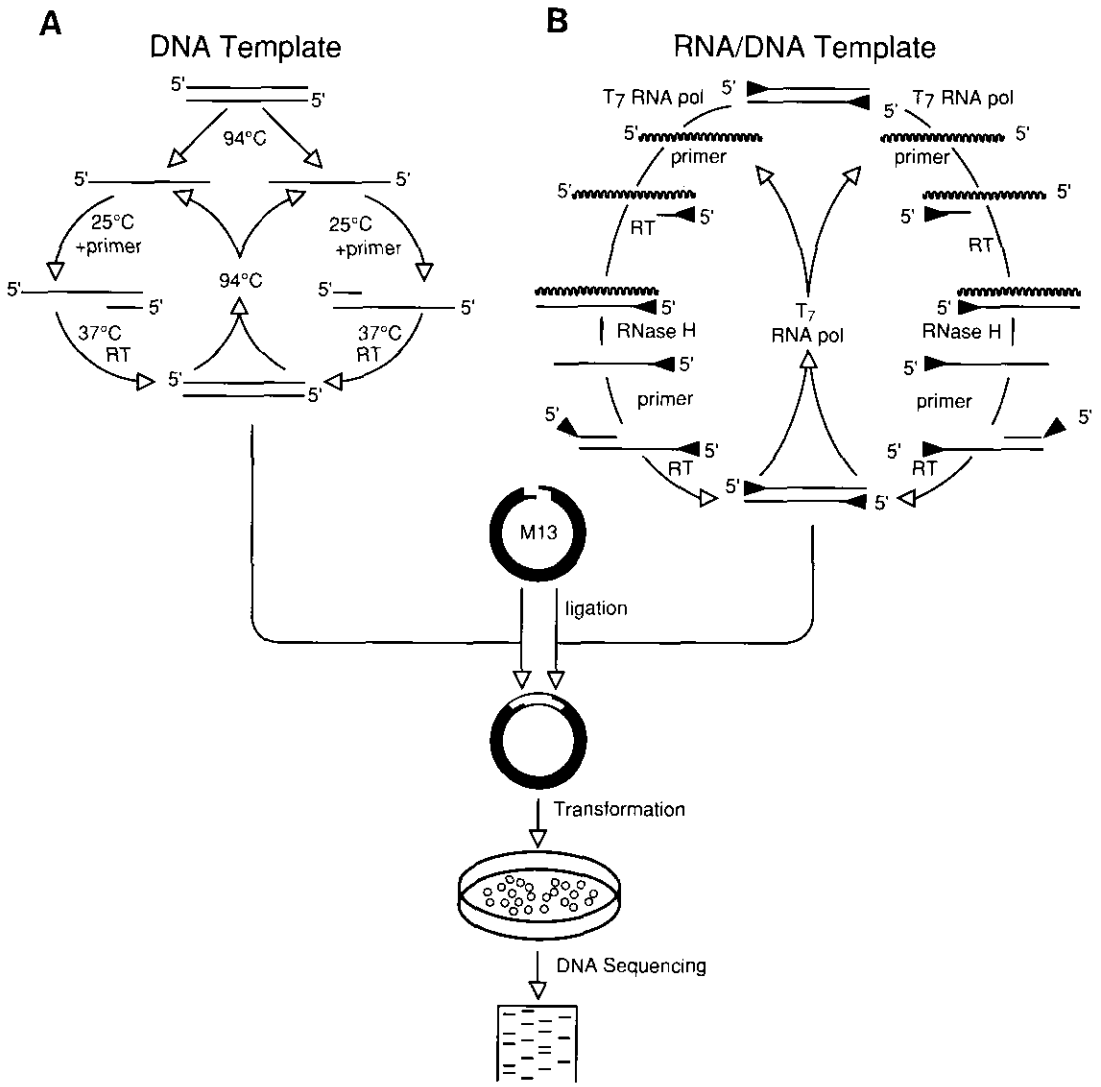

FIG. 1. Assays for fidelity of nucleotide polymerization on DNA and RNA templates containing the *env* V-1 gene. T7 promoter sequences were introduced both upstream and downstream from the HIV-1 *env* V-1. HIV-1 RT was used to copy the V-1 fragment using (A) a DNA template, in which the HIV-1 *env* hypervariable fragment is amplified by the HIV-1 RT-mediated PCR, and using (B) sequential RNA and DNA templates, in which RNA was synthesized by T7 RNA polymerase, and T7 RNA transcripts were used as templates for the production of cDNA copies by HIV-1 RT. The amplified V-1 DNA products were cloned into an M13 vector, transfected into *E. coli,* and sequenced.

and in principle similar to that used with the DNA template (Fig. 1A). Amplification proceeds by reiterated RNA synthesis by T7 RNA polymerase using the *env* V-1 fragment as a template followed by RNA-dependent and then DNA-dependent DNA synthesis catalyzed by HIV RT to produce a double-stranded DNA product. The initial template that we constructed contains T7 promoters on both ends of the *env* V-1 insert, resulting in the synthesis of both strands of RNA upon the addition of T7 RNA polymerase. The RNA serves as template for the synthesis of cDNA copies by HIV RT. The complementary DNA is then synthesized by HIV RT after degradation of RNA in the RNA–DNA hybrid by RNase H. Since the oligonucleotide primers specify the target and contain 5′ extensions encoding the T7 RNA polymerase binding site, the resultant double-stranded DNA is a competent transcription template for subsequent amplification at constant temperature. The RT errors introduced during RNA/DNA amplification were identified by the same procedures as those utilized for DNA amplification. Since T7 RNA polymerase is at least eightfold more accurate then HIV-1 RT (Boyer *et al.,* 1992; Hubner *et al.,* 1992), the infidelity of T7 RNA polymerase after multiple-cycle amplifications would be included in the background mutation frequency in this system.

The mutation spectrum and error rates of HIV RT during RNA- and DNA-dependent DNA polymerization are shown in Fig. 2 and tabulated in Table 1. Fifty-seven mutations were identified among the 136 clones obtained after 13 cycles of RNA/DNA amplification. The combined error rate for both reverse transcription

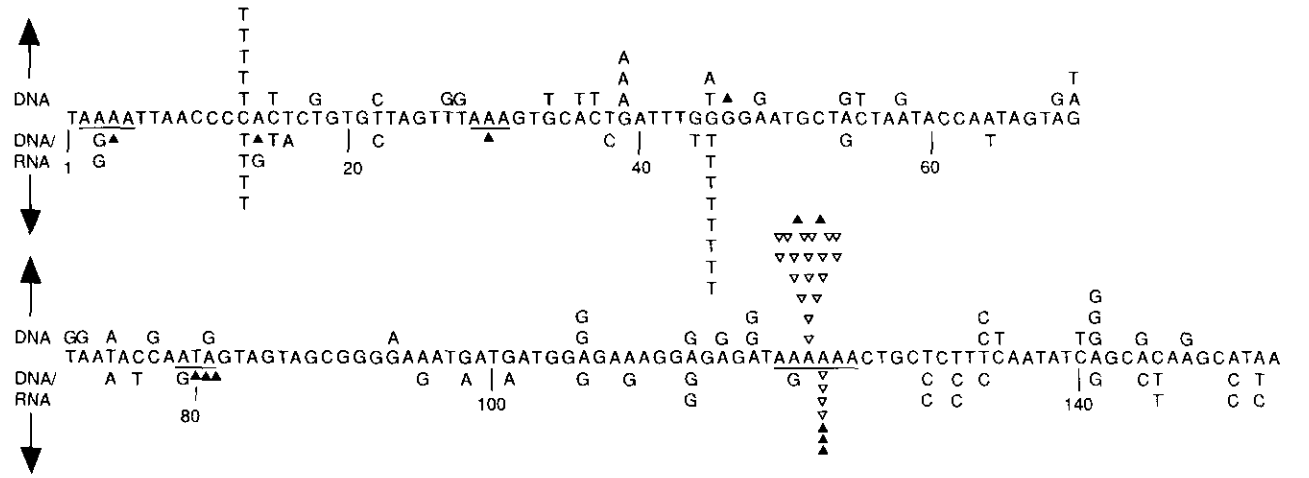

FIG. 2. Spectra of mutations induced by HIV RT. The 153-base target of the HIV-1 *env* viral (+) strand is shown. Number one is 358 bases downstream of the *env* translational start site of the HIV BRU isolate. The spectrum shown above the substituted nucleotide in the wild-type sequence is obtained from DNA amplification and that below the wild-type sequence is from RNA/DNA amplification. Single-base substitutions are displayed above or below the wild-type sequence. Deletions are shown as ▲, and additions as ∇ for +1 and ∇∇ for +2, with a line indicating the presumed origin. The ▲▲▲ at position 80 refers to a 3 nucleotide deletion. The following multiple mutations were observed in individual clones obtained by DNA amplification: (1) position 36 A → T, 81 A → G, and 113 A → G; (2) 37 C → T and 121 +2A; (3) 106 A → G and 133 T → C; (4) 117 A → G and 121 +1A; and (5) 150 A → C and 28 T → G. The following multiple mutations were observed in individual clones obtained by the RNA and DNA amplification: (1) 3 A → G and 22 T → C; (2) 13 C → T and 45 G → T; (3) 106 A → G and 121 +1A; and (4) 150 A → C and 152 A → T.

and plus-stranded DNA synthesis is estimated as one error per 2600 nucleotides replicated (Table 1, see legend). The most frequent mutations were observed at positions 13, 45, and 119 to 124. A wide range in the numbers of mutations at different positions was observed; for examples, the frequency of GC → TA transversion was 1/110 at position 45 compared to an average of 1/17,000 and the frequency of GC → AT transitions was 1/220 at position 13 compared to an average of 1/6100. In each cycle of RNA/DNA amplification, RT synthesizes DNA twice, first with an RNA template and then with a DNA template. Position of coincidence in the mutation spectra with a DNA and an RNA/DNA templated reaction presumably represents misincorporation during the DNA-dependent DNA polymerization step, examples of which are the GC → AT transitions at 13 and the additions opposite the poly(dA/T) tract. The errors unique to the RNA/DNA templated reactions presumably represent errors introduced by RT on the RNA template. The error rate for HIV-1 RT copying RNA was estimated at about 1/5000, nearly the same as that copying DNA. However, different types of errors were produced on the RNA template versus DNA template (Table 1). The transversion and frameshift error rate on *env* V-1 RNA is 1/12,500 and 1/4100, about twofold higher or lower, respectively, than those obtained on *env* V-1 DNA.

RT errors and AIDS mutations

To estimate the contribution of HIV RT in generating the hypermutability that is characteristic of the *env* V-1 sequence, we compared the types of mutations introduced by HIV-1 RT *in vitro* with those reported in AIDS patients (Fig. 3). Since frameshift mutations due to RT infidelity are unlikely to be viable *in vivo*, we limited this comparison to only base-substitution mutations. We calculated nucleotide variability *in vivo* by numbers of different nucleotides at a given position divided by frequency of the most frequent nucleotide found at that position after aligning 44 individual AIDS isolates (Myers *et al.*, 1989). In general, there is no strong correlation between the nucleotide positions where HIV-1 RT misincorporates at high frequency and the mutations observed in AIDS patients. The region from 48 to 68 is not the location for high misincorporation by HIV RT but is among the most variable *in vivo*. Likewise, the hot spots for misincorporation by HIV RT at positions 13 and 45 are not observed at high frequency in AIDS patients.

DISCUSSION

To determine the contribution of the nucleotide sequence of the envelope gene to the high frequency of the mutations observed in this region, we developed two new cell-free assays for HIV RT fidelity. Vectors containing the *env* V-1 insert provide a homogeneous template for measuring the fidelity of HIV RT *in vitro*, avoiding the complexity of an initial heterogeneous viral population. Misincorporations by HIV RT in the *env* V-1 region were detected after multiple rounds of replication without selection. We found the error rate of HIV RT is about one error per 5300 nucleotides copying DNA and about one in 5000 copying RNA. This exceptionally high error rate could account for three to

TABLE 1

EFFECTS OF TEMPLATE AND SEQUENCE-CONTEXT ON HIV-1 RT FIDELITY

Sequence changes	Position	DNA template		RNA/DNA template		RNA template	Error ratio
		Number of mutations	e^a ($\times 10^{-5}$)	Number of mutations	e ($\times 10^{-5}$)	e ($\times 10^{-5}$)	RNA/DNA
Base substitutions		46	13	46	30	17	1.3
Transitions		35	10	29	19	9	0.9
GC to AT		16	12	10	16	4	0.3
	13	5	250	4	450	200	0.8
	39	4	150	0	<16	<16	<0.1
AT to GC		19	8	19	17	9	1.1
	113	1	48	2	220	180	3.8
	141	3	150	1	110	<9	<0.1
Transversions		11	3	17	11	8	2.7
GC to TA		3	1	9	6	5	5.0
	45	1	48	8	910	830	17.3
AT to CG		5	1	4	3	1	1.0
Others		3	1	4	3	2	2.5
Frameshifts		18	5	11	8	2	0.4
Deletions	119–125	2	15	3	53	37	2.5
Additions	119–125	15	110	4	71	<8	<0.1
Overall		64	19	57	38	20	1.1

[a] The error rate (e), defined as errors per base per replication cycle was estimated as $2\,(p/n - p'/n')/d$ (Hayes, 1965; Saiki *et al.*, 1988), in which p is percentage of clones with mutations after multiple cycles (d) of replication with n as length of nucleotide copied and p'/n' as background mutations during DNA amplification. The number of nucleotide targets (n) that yield transitions or transversions is 55 for G or C and 98 for A to T/U among 153 bases of the *env* V-1 fragment. To estimate error rate (e') of *Taq* polymerase during PCR, we sequenced 50 clones derived from 5 cycles of PCR by *Taq* polymerase and observed one GC → AT transition at 134. The estimated error rate of *Taq* polymerase, 1.0×10^{-4}, is consistent with earlier reports (Tindall and Kunkel, 1988; Saiki *et al.*, 1990). With 5 cycles (d') of PCR, the mutation (p'/n') introduced by *Taq* polymerase is $1/2\ e'd'$, about 1/3600 nucleotides polymerized. The HIV RT-mediated amplification cycles were monitored by agarose gel electrophoresis and scintillation counting. Sixty-four errors were observed among 160 clones ($64/160 \times 153$) after 25 cycles (d) using DNA amplification, while there were 57 errors among 136 clones ($57/136 \times 153$) after 13 cycles (d) using RNA/DNA amplification. Therefore, the error rates during DNA and RNA/DNA amplification are about 1/5300 $\{2 \times [(64/160 \times 153) - (1/3600)/25] = 1.9 \times 10^{-4}\}$ and 1/2600 $\{2 \times [(57/153 - 1/3600)]/13 = 3.8 \times 10^{-4}\}$, respectively. In each cycle of RNA/DNA amplification, RT synthesizes DNA twice, first from the RNA template and second from the c-DNA template for second-strand DNA synthesis. The error rate from the RNA template alone was calculated by subtracting the error rate with DNA alone. The error rate of HIV RT is estimated to be about 1/5000 (2.0×10^{-4}) for the RNA template, and this includes the small proportion of PCR-dependent errors due to the infidelity of T7 RNA polymerase (Boyer *et al.*, 1992; Hubner *et al.*, 1992; Ji and Loeb, 1992). Mutations increased as a function of the number of amplification cycles. With DNA amplification, there were 3 mutations in 20 clones after 10 cycles and 64 mutations in 160 clones after 25 cycles. With the RNA/DNA template, there were 3 mutations in 21 clones after 4 cycles and 57 mutations in 153 clones after 13 cycles.

four mutations per HIV genome per replication cycle and could be responsible for the high frequency of point mutations, which are the primary variations observed among viral isolates. Furthermore, errors made by RT could also contribute to the high frequency of recombination and genome rearrangements that are found in viral isolates (Hahn *et al.*, 1986; Goodenow *et al.*, 1989; Hu and Temin, 1990). Single base substitutions made by HIV RT could be the sites for repair, leaving single-stranded gaps that enhance recombination. The similarity in error rates of HIV RT in copying the *env* gene segment and M13 *lacZα* (Ji and Loeb, 1992) suggests that the HIV hypervariable region is not a sequence that is copied at a higher error frequency than that of the rest of the viral genome.

The mutation spectra produced by HIV RT within the *env* V-1 sequence is nonrandom. The similarities that exist between the RT errors *in vitro* with mutations *in vivo* (Myers *et al.*, 1989; Vartanian *et al.*, 1991) suggest that certain mutations are due to misincorporation by the RT directly. The GC → AT transition is one of the major types of error produced by HIV RT in copying both the RNA and the DNA templates (Table 1). Some of the mutations observed by Vartanian *et al.* (1991), such as the GC → AT transition at position 39, may be the sites for increased misincorporation. However, our studies of RT fidelity do not reveal an exceptionally high frequency of G → A transitions and thus suggest that the preponderance of G → A substitutions observed *in vivo* (Pathak and Temin, 1990; Vartanian *et al.*, 1991)

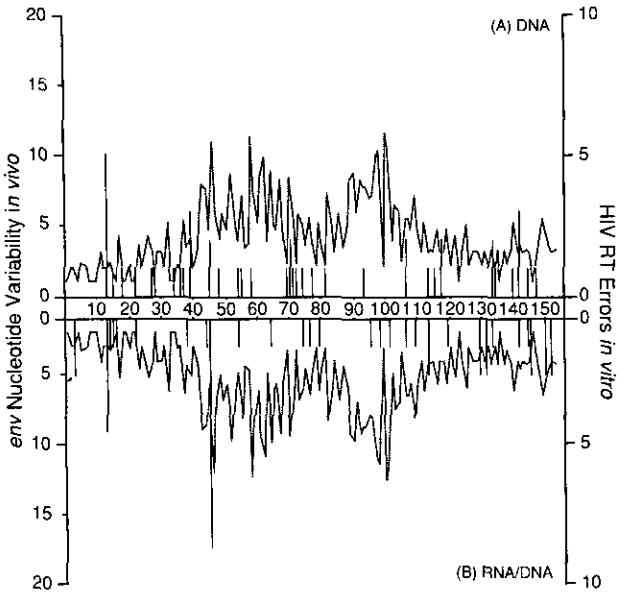

FIG. 3. Comparison of *env* nucleotide variability in AIDS patients with HIV-1 RT errors *in vitro*. The vertical bars represent the number of base substitutions introduced by HIV-1 RT at each given positions during DNA amplification (A) and RNA/DNA amplification (B). The data are taken from Fig 2. The curve represents the nucleotide variability *in vivo*, calculated by numbers of different nucleotides at a given position divided by frequency of the most frequent nucleotide found at that position (Devereuv *et al.*, 1984; Wiley *et al.*, 1986). The data *in vivo* are from 44 individual HIV-1 *env* sequences from AIDS patients (Myers *et al.*, 1989).

are not due to the specificity of errors by HIV RT. Other potential sources for G → A substitutions include selective advantage for viral pathogenesis and transmissibility (Overbaugh *et al.*, 1991), an increase in the pool size of dATP or dTTP in HIV-infected cells (Mathews and Ji, 1992), and mutations in the polymerase during viral infection that enhance GC → AT transitions (Hu and Temin, 1990; Pathak and Temin, 1990).

It is interesting to note that HIV RT has a tendency to misincorporate adenine into its genome, which might present an evolutionary bias for maintaining HIV as an A-rich genome (twofold) (Myers *et al.*, 1989; Williams and Loeb, 1992). We not only observed a higher frequency of GC → AT transitions and GC → TA transversion mutations, but also a high frequency of adenine additions on the plus strand or thymidine on the minus strand during DNA-dependent DNA polymerization. Using a limited amount of RT to copy the *env* V-1 region, we found a strong pausing of HIV RT copying viral plus-strand DNA around 119–125 of *env* V-1, but not on minus-strand DNA or RNA (data not shown). The strong correlation between pausing sites and frameshift hot spots around polydeoxyadenine regions suggests that the additions occur through a primer–template slippage mechanism (Streisinger *et al.*, 1986).

From an analysis of viral isolates obtained during the course of AIDS in two patients, Nowak *et al.* (1991) found that diversity of *env* sequences increased with time. Their data suggest that the immunodeficiency syndrome becomes manifested when the diversity of the *env* gene exceeds the host immune capacity, supporting the model in which the generation of *env* diversity is a cause, and not a consequence, of AIDS (Nowak *et al.*, 1991). The implication is that effective antiviral drugs would be those that inhibit viral replication and thus diminish the extensive nucleotide variation in the HIV *env* region, thereby delaying the progression of the disease. Although drugs that inhibit viral DNA synthesis such as 3′-azido-3′-deoxythymidine (AZT) and 2′,3′-dideoxyinosine (ddI) are effective, their efficiency is limited by the emergence of resistant variants after a short course of chemotherapy (Merluzzi *et al.*, 1990; Pauwels *et al.*, 1990; Baba *et al.*, 1991; Clair *et al.*, 1991; Goldman *et al.*, 1991; Nunberg *et al.*, 1991). The present study indicates that the hypervariations observed in the *env* gene are not the result of an exceptionally high error rate that is inherent to copying the nucleotide sequence of that gene. Instead the error rate observed in copying the *env* V-1 fragment is similar to that previously reported in studies using the *lacZα* gene (Ji and Loeb, 1992). Thus, it seems likely that the disproportionately high frequency of mutations observed in the *env* gene of viral isolates is the result of amino acid substitutions that permit the virus to circumvent the host's immune defense mechanisms or that enhance the rate of viral replication by a yet to be defined mechanism. The extraordinary high RT error rates observed in this study and others (Preston *et al.*, 1988; Roberts *et al.*, 1988, 1989; Bebenek *et al.*, 1989; Perrino *et al.*, 1989; Weber and Grosse, 1989; Boyer *et al.*, 1992; Ji and Loeb, 1992) suggest that the rapid evolution of the virus (Myers *et al.*, 1989; Williams and Loeb, 1992), the high mutation rate (Goodenow *et al.*, 1989; Nowak *et al.*, 1991; Vartanian *et al.*, 1991), and the emergence of drug-resistant strains (Merluzzi *et al.*, 1990; Pauwels *et al.*, 1990; Baba *et al.*, 1991; Clair *et al.*, 1991; Goldman *et al.*, 1991; Nunberg *et al.*, 1991) are due to misincorporation by HIV RT. Since errors by RT occur throughout the genome including the *env*, *pol*, or *gag* genes, it seems likely that many mutants are nonviable. Agents that increase the frequency of errors by the RT might exceed the threshold for virus viability and thus restrict replication of the AIDS virus (Eigen *et al.*, 1989). The infidelity of HIV RT has been exploited for the site-specific insertion of both noncomplementary nucleotides and bulky nucleotide analogues (Cheng *et al.*, 1991) and might be advantageous to the selective insertion of nucleotide analogues that miscode at high frequencies.

ACKNOWLEDGMENTS

We express appreciation to Dr. Michael Emerman for providing pME235 plasmid, to Richard Duncan for assistance in DNA se-

quencing, and to David Adler, David Basiji, and Kris Carroll for custom software and graphic arts. This work was supported by Grant OIG R-35-CA-39903 from the U.S. National Institutes of Health.

REFERENCES

BABA, M., DECLERCQ, E., TANAKA, H., UBASAWA, M., TAKASHIMA, H., SEKIYA, K., NITTA, I., UMEZU, K., NAKASHIMA, H., MORI, S., SHIGETA, S., WALKER, R. T., and MIYASAKA, T. (1991). *Proc. Natl. Acad. Sci. USA* **88,** 2356–2360.

BEBENEK, K., ABBOTTS, J., ROBERTS, J. D., WILSON, S. H., and KUNKEL, T. A. (1989). *J. Biol. Chem.* **264,** 16948–16956.

BENZER, S. (1961). *Proc. Natl. Acad. Sci. USA* **47,** 403–415.

BLANK, A., GALLANT, J. A., BURGESS, R. R., and LOEB, L. A. (1986). *Biochemistry* **25,** 5920–5928.

BOYER, J. C., BEBENEK, K., and KUNKEL, T. A. (1992). *Proc. Natl. Acad. Sci. USA* **89,** 6919–6923.

CHENG, K., CAHILL, D. S., PRESTON, B. D., SINGLE, B., and LOEB, L. A. (1991). *Proc. Natl. Acad. Sci. USA* **88,** 9974–9978.

CLAIR, M. H. S., MARTIN, J. L., TUDOR-WILLIAMS, G., BACH, M. C., VAVRO, C. L., KING, D. M., KELLAM, P., KEMP, S. D., and LARDER, B. A. (1991). *Science* **253,** 1557–1559.

COFFIN, J. M. (1986). *Cell* **46,** 1–4.

DEVEREUX, J., HAEBERLI, P., and SMITHIES, O. (1984). *Nucleic Acids Res.* **12,** 387–395.

DOI, H. (1991). *Proc. Natl. Acad. Sci. USA* **88,** 9282–9286.

EIGEN, M., MCCASKILL, J., and SCHUSTER, P. (1989). *Adv. Chem. Phys.* **75,** 149–263.

GALLO, R. C., and JAY, G. (1991). "The Human Retroviruses." Academic Press, San Diego, CA.

GOLDMAN, M. E., HUNBERG, J. H., O'BRIEN, J. A., QUINTERO, J. C., SCHLEIF, W. A., FREUND, K. F., GAUL, S. L., SAARI, W. S., WAI, J. S., HOFFMAN, J. M., ANDERSON, P. S., HUPE, D. J., EMINI, E. A., and STERN, A. M. (1991). *Proc. Natl. Acad. Sci. USA* **88,** 6863–6867.

GOODENOW, M., HUET, T., SAURIN, W., KWOK, S., SNINSKY, J., and WAIN-HOBSON, S. (1989). *J. AIDS* **2,** 344–352.

GUATELLI, J. C., WHITFIELD, K. M., KWOH, D. Y., BARRINGER, K. J., RICHMAN, D. D., and GINGERAS, T. R. (1990). *Proc. Natl. Acad. Sci. USA* **87,** 1874–1878.

HAHN, B. H., SHAW, G. M., TAYLOR, M. E., REDFIELD, R. R., MARKHAM, P. D., SALAHUDDIN, S. Z., WONG-STAAL, F., GALLO, R. C., PARKS, E. S., and PARKS, W. P. (1986). *Science* **232,** 1548–1553.

HAYES, W. (1965). "The Genetics of Bacteria and Their Viruses." Wiley, New York.

HU, W. S., and TEMIN, H. M. (1990). *Science* **250,** 1227–1233.

HUBNER, A., KRUHOFFER, M., GROSSE, F., and KRAUSS, G. (1992). *J. Mol. Biol.* **223,** 596–600.

JI, J., and LOEB, L. A. (1992). *Biochemistry* **31,** 954–958.

JI, J., and MATHEWS, C. K. (1991). *Mol. Gen. Genet.* **226,** 257–264.

KUNKEL, T. A. (1985). *J. Biol. Chem.* **260,** 5587–5596.

MATHEWS, C. K., and JI, J. (1992). *BioEssays* **14(5),** 1–7.

MERLUZZI, V. J., HARGRAVE, K. D., LABADIA, M., GROZINGER, K., SKOOG, M., WU, J., SHIH, C. K., ECKNER, K., HATTOX, S., ADAMS, J., ROSENTHAL, A. S., FAANES, R., ECKNER, R. J., KOUP, R. A., and SULLIVAN, J. L. (1990). *Science* **250,** 1411–1413.

MYERS, G., RABSON, A. B., JOSEPHS, S. F., SMITH, T. F., WONG-STAAL, F., and BERZOFSKY, J. A. (1989). "Human Retroviruses and AIDS." Los Alamos Natl. Lab., Los Alamos, NM.

NOWAK, M. A., ANDERSON, R. M., MCLEAN, A. R., WOLFS, T. F., GOUDSMIT, J., and MAY, R. M. (1991). *Science* **254,** 963–969.

NUNBERG, J. H., SCHLIEF, W. A., BOOTS, E. J., O'BRIEN, J. A., QUINTERO, J. C., HOFFMAN, J. M., EMINI, E. A., and GOLDMAN, M. E. (1991). *J. Virol.* **65,** 4887–4892.

OVERBAUGH, J., RUDENSEY, L. M., PAPENHAUSEN, M. D., BENVENISTE, R. E., and MORTON, W. R. (1991). *J. Virol.* **65,** 7025–7031.

PATHAK, V. K., and TEMIN, H. (1990). *Proc. Natl. Acad. Sci. USA* **87,** 6019–6023.

PAUWELS, R., ANDRIES, K., DESMYTER, J., SCHOLS, D., KUKLA, M. J., BRESLIN, H. J., RAEYMAECKERS, A., GELDER, J. V., WOESTENBORGHS, R., HEYKANTS, J., SCHELLEKENS, K., JANSSEN, M. A. C., CLERCQ, E. D., and JASSEN, P. A. J. (1990). *Nature* **343,** 470–474.

PERRINO, F. W., PRESTON, B. D., SANDELL, L. L., and LOEB, L. A. (1989). *Proc. Natl. Acad. Sci. USA* **86,** 8343–8347.

PRESTON, B. D., POIESZ, B. J., and LOEB, L. A. (1988). *Science* **242,** 1168–1171.

RICCHETTI, M., and BUE, H. (1990). *EMBO. J.* **9,** 1583–1593.

ROBERTS, J. D., BEBENEK, K., and KUNKEL, T. A. (1988). *Science* **242,** 1171–1173.

ROBERTS, J. D., PRESTON, B. D., JOHNSON, L. A., SONI, A., LOEB, L. A., and KUNKEL, T. A. (1989). *Mol. Cell. Biol.* **9,** 469–475.

SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARF, S. J., HIGUCHI, R., HORN, G. T., MULLIN, K. B., and ERLICH, H. A. (1988). *Science* **239,** 487–491.

STREISINGER, G., OKADA, Y., EMRICH, J., NEWTON, J., TSUGITA, A., TERZAGHI, E., and INOUYE, M. (1986). *Cold Spring Harbor Symp. Quant. Biol.* **31,** 77–84.

TINDALL, K. R., and KUNKEL, T. A. (1988). *Biochemistry* **27,** 6008–6013.

VARTANIAN, J. P., MEYERHANS, A., ASJO, B., and WEIN-HOBSON, S. (1991). *J. Virol.* **65,** 1779–1788.

WEBER, J., and GROSSE, F. (1989). *Nucleic Acids Res.* **17,** 1379–1393.

WILEY, R. L., RUTLEDGE, R. A., DIAS, S., FOLKS, T., THEODORE, T., BUCKLER, C. E., and MARTIN, M. A. (1986). *Proc. Natl. Acad. Sci. USA* **83,** 5038–5042.

WILLIAMS, K. J., and LOEB, L. A. (1992). Current topics in microbiology and immunology. *In* "Genetic Diversity of RNA Viruses" (J. Holland, Ed.), pp. 165–180. Springer-Verlag, Berlin.

YU, H., and GOODMAN, M. F. (1992). *J. Biol. Chem.* **267,** 10888–10896.