large substituent that precludes incorporation by the host cell DNA polymerase, but allows incorporation by HIV-1 RT.

REFERENCES AND NOTES

- 1. J. M. Coffin, Cell 46, 1 (1986)
- B. H. Hahn et al., Science 232, 1548 (1986).
 S. M. Desai et al., Proc. Natl. Acad. Sci. U.S.A. 83,
- 8380 (1986). 4. D. A. Steinheuer and J. J. Holland, Annu. Rev.
- Microbiol. 41, 409 (1987); H. M. Temin, Cancer Res. 48, 1697 (1988) 5. N. Weiss, H. Teich, J. Varmus, J. Coffin, Eds., RNA Tumor Viruses, vol. 2, Supplements and Appendixes
- (Cold Spring Harbor Laboratory, Cold Spring Har-bor, NY, ed. 2, 1985), p. 75. 6. I. M. Verma, Enzymes 14, 87 (1981); A. D. H.
- Hoffman, B. Banapour, J. A. Levy, Virology 147, 326 (1985) 7. N. Battula and L. A. Loeb, J. Biol. Chem. 249, 4086
- (1974); K. P. Gopinathan, L. A. Weymouth, T. A. Kunkel, L. A. Loeb, Nature 278, 857 (1979); T. A. Kunkel, F. Eckstein, A. S. Mildvan, R. M. Koplitz, L. A. Loeb, Proc. Natl. Acad. Sci. U.S.A. 78, 6734 (1981); T. A. Kunkel, R. M. Schaaper, R. Beckman, L. A. Loeb, J. Biol. Chem. 256, 9883 (1981); T. A. Kunkel, R. M. Schaaper, L. A. Loeb, Biochemistry 22, 2378 (1983); J. D. Roberts et al., Mol. Cell.
- Biol., in press.
 8. D. J. Jolly, R. C. Willis, T. Friedmann, Mol. Cell. Biol. 6, 1141 (1986).
- 9. M. Fry and L. A. Loeb, Animal Cell DNA Polymerases (CRC Press, Boca Raton, FL, 1986); M. E. Reyland and L. A. Loeb, J. Biol. Chem. 262, 10824 (1987).
- 10. F. Sanger and A. R. Coulson, J. Mol. Biol. 94, 441
- (1975).
 11. G. G. Hillebrand, A. H. McClusky, K. A. Abbott, G. G. Revich, K. L. Beattie, *Nucleic Acids Res.* 12, 3155 (1984).
- 12. M. S. Boosalis, J. Petruska, M. F. Goodman, J. Biol. Chem. 262, 14689 (1987)
- 13. B. D. Preston, B. J. Poiesz, L. A. Loeb, unpublished observations.
- , unpublished observations.
- A. Fersht, Enzyme Structure and Mechanism (Freeman, 15. San Francisco, CA, 1977), p. 16.
- 16. HIV-1 virions were obtained from 20 liters of cell culture-conditioned medium from HUT 78 cells that are stably infected with the HIV $_{AAV}$ isolate [S. Kwok *et al.*, *J. Virol.* **61**, 1690 (1987); B. Poiesz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7415 (1980)]. The purified virions were disrupted in 0.5% Triton X-100 and 0.6N HCl, and nucleic acids were removed by passing the lysate through a DEAEcellulose column that had been previously equili-brated with buffer A [0.5M NaCl, 25 mM tris-HCl (pH 7.5), 10% glycerol, 1 mM dithiothreitol, 1 mM EDTA, and 10 μ g of leupeptin and pepstatin A per milliliter] and with 15 ml of buffer A containing 0.1 mg of bovine serum albumin (BSA) per milliliter. Active fractions were combined, diluted in 10 volumes of modified buffer A (10 mM tris-HCl, pH 7.5) and immediately applied to a phosphocellulose P-11 column $(1 \times 6 \text{ cm})$ that was equilibrated with buffer A and washed with 40 ml of buffer A containing 0.1 mg of BSA per milliliter. Elution was with a 64-ml linear gradient of 0.05 to 0.5M NaCl in buffer A, containing 0.1 mg of BSA per milliliter, and 0.5-ml fractions were collected. RT activity eluted in a single sharp peak at 0.2M NaCl; the most active fraction was used in these studies
- 17. C. Majumdar, J. Abbotts, S. Broder, S. H. Wilson, Biochemistry 27, 901 (1988).
- 18. L. A. Weymouth and L. A. Loeb, Proc. Natl. Acad. Sci. U.S.A. 75, 1924 (1978); T. A. Kunkel, E. A. James, L. A. Loeb, in DNA Repair: A Laboratory Manual of Research Procedures, E. C. Friedberg and P. C. Hanawalt, Eds. (Dekker, New York, 1983), pp. 223-237; J. Abbotts and L. A. Loeb, Biochim. Biophys. Acta 824, 58 (1985).
- 19. T. A. Kunkel and P. S. Alexander, J. Biol. Chem. 261, 160 (1986)
- 20. F. W. Perrino and L. A. Loeb, unpublished results.

- 21. M. M. Abboud, W. J. Sim, L. A. Loeb, A. S. Mildvan, J. Biol. Chem. 253, 3415 (1978)
- 22. G. N. Wilkinson, Biochem. J. 80, 324 (1961) B. D. Preston, B. Singer, L. A. Loeb, J. Biol. Chem. 262, 13821 (1987). 23
- F. Sanger et al., J. Mol. Biol. 125, 225 (1978).
 S.-Z. Hu, T. S. F. Wang, D. Korn, J. Biol. Chem. 259, 2602 (1984).
- 26. We thank F. W. Perrino for discussions and provid-

ing the calf thymus DNA polymerase α , S. H. Wilson for the generous gift of cloned HIV-1 RT, C. Stevens and M. Rubert for the HIV-1 preparation, and D. Lim, M. Whiting, and J. Hiltner for word processing. Supported in part by NIH grants CA-07263-03 (B.D.P.), N01AI72654 (B.J.P.), and R35-CA-39903 (L.A.L.).

15 July 1988; accepted 3 October 1988

The Accuracy of Reverse Transcriptase from HIV-1

John D. Roberts, Katarzyna Bebenek, Thomas A. Kunkel*

A study was conducted to determine the fidelity of DNA synthesis catalyzed in vitro by the reverse transcriptase from a human immunodeficiency virus type 1 (HIV-1). Like other retroviral reverse transcriptases, the HIV-1 enzyme does not correct errors by exonucleolytic proofreading. Measurements with M13mp2-based fidelity assays indicated that the HIV-1 enzyme, isolated either from virus particles or from Escherichia coli cells infected with a plasmid expressing the cloned gene, was exceptionally inaccurate, having an average error rate per detectable nucleotide incorporated of 1/1700. It was, in fact, the least accurate reverse transcriptase described to date, onetenth as accurate as the polymerases isolated from avian myeloblastosis or murine leukemia viruses, which have average error rates of $\sim 1/17,000$ and $\sim 1/30,000$, respectively. DNA sequence analyses of mutations generated by HIV-1 polymerase showed that base substitution, addition, and deletion errors were all produced. Certain template positions were mutational hotspots where the error rate could be as high as 1 per 70 polymerized nucleotides. The data are consistent with the notion that the exceptional diversity of the HIV-1 genome results from error-prone reverse transcription.

The ability of HIV-1 to escape the host's immune response (1), in part the result of extensive genetic variation, is one of the key features of this virus that complicates development of an effective vaccine for the prevention of AIDS. However, no information is available on the rate of mutation per HIV-1 growth cycle. In fact, the only calculation of a mutation rate for any retrovirus comes from a recent study (2) of an avian retroviral expression vector. Nevertheless, comparisons of HIV-1 isolates from patients in different geographic locations allow estimates of the frequency of change in terms of nucleotide substitutions per site per year, and values of ≥ 0.002 have been obtained (3, 4). Mutation frequencies have also been estimated from the nucleotide sequences of HIV-1 isolates taken from a single patient at different times after infection (5). Minimum estimates for the variation rate in the *env* gene range from 0.001 to 0.01 nucleotide substitutions per site per year, whereas mutation frequencies for the gag gene are about one-tenth as high. The most remarkable feature of HIV-1 variability is the presence of hypervariable regions

within specific portions of the env and nef genes (4). The diversity in these regions is characterized by a wide variety of changes, including transitions, transversions, insertions, and deletions (5).

Genetic variability in HIV-1 may involve one or all of several steps in the viral life cycle (6), including reverse transcription, integration into the host chromosome to establish the proviral state, provirus replication by the host replication machinery, and transcription by RNA polymerase II into RNAs to be packaged into virions. In this study we focused on the HIV-1 reverse transcriptase (RT) as a potential source of diversity. We have used the RT as purified from virus particles as well as the purified enzyme that had been overproduced in E. coli carrying the cloned gene for HIV-1 RT. This recombinant enzyme appears to have the same subunit structure and kinetic properties as the natural enzyme (7).

We first examined HIV-1 RT for associated $3' \rightarrow 5'$ proofreading exonuclease activity using two separate assays for excision of a mismatched base from a primer terminus. In the first assay (8), a mismatched substrate was constructed by hybridizing a 15-base oligonucleotide labeled with ³²P at the 5' end to M13mp2 single-strand DNA; this created an A (template): G mispair at the 3' primer terminus. The products of excision

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709

^{*}To whom correspondence should be addressed.

reactions were analyzed by electrophoresis in a denaturing 20% polyacrylamide gel (Fig. 1). Efficient excision of the terminal nucleotide occurs within 5 min when the Klenow fragment of E. coli DNA polymerase I is used as a positive control (Fig. 1, lane 4). Under identical conditions, only a trace of $3' \rightarrow 5'$ exonuclease activity was detected in the viral HIV-1 RT preparation after 2 hours of incubation (lane 3). On the basis of the amounts of polymerase present in these reactions (7), we calculated that the natural HIV-1 RT preparation contained about 1/10,000 as much $3' \rightarrow 5'$ exonuclease activity as the Klenow polymerase. The exonuclease activity of the bacterially expressed recombinant HIV-1 polymerase was one-fifth that of the natural enzyme (lane 2). The contribution to fidelity of this trace amount of exonuclease in the HIV-1 RT preparations is likely to be negligible, particularly since the contribution to fidelity of



Fig. 1. Electrophoretic analysis of terminal mismatch excision. Reactions (25 µl) contained 50 mM tris-HCl (pH 7.8), 2 mM dithiothreitol, 10 mM MgCl₂, 300 ng of mismatched DNA [see (8)], and 0.176 pmol of Klenow fragment of DNA polymerase I or ~7.4 pmol of HIV-1 polymerase. After incubating at 37°C for 5 min with the Klenow polymerase or 120 min with the HIV-1 polymerases, reactions were terminated as described (8). Aliquots (4 µl) were analyzed by electrophoresis, and the radioactivity in each band was quantitated as described (8). Lane 1, no enzyme; lane 2, recombinant HIV-1 RT; lane 3, natural HIV-1 RT; and lane 4, Klenow polymerase. The calculation of exonuclease activity is based on the percentage of radioactivity present in bands shorter than the original 15-base oligonucleotide: no enzyme, 2.6%; recombinant HIV-1 RT, 4.2%; natural HIV-1 RT, 11%; and Klenow, 82%. The relative amount of exonuclease activity was estimated by subtracting the background (no enzyme), dividing by the picomoles of enzyme present in the assay, then dividing by the time of incubation. This calculation yields a value of 90% removed per picomole per minute for Klenow polymerase [(82% - 2.6%)/0.176 pmol per 5 min], 0.0095% per picomole per minute for natural HIV-1 RT [(11% - 2.6%)/2.6%)/7.4 pmol per 120 min], and 0.0018% per picomole per minute for recombinant HIV-1 RT (4.2% - 2.6%)/7.4 pmol per 120 min]. Thus, the difference in activity between Klenow polymerase and HIV-1 RT is ~10,000-fold for the natural enzyme and ~50,000-fold for the recombinant enzyme.

exonucleolytic proofreading by *E. coli* polymerase I is modest (9).

The second assay, which measures both the ability of an exonuclease to excise a mismatched base and the ability of its associated polymerase to extend from a terminal mispair (8), requires the use of an M13mp3 DNA substrate containing a 363-base single-strand gap and an A (template):C mispair at the 3' primer terminus. The codons resulting from the two bases in the mispair code for two different lacZ phenotypes. Excision of the cytosine followed by gap-filling synthesis generates a homoduplex DNA product that, upon transfection, yields exclusively faint blue M13 plaques. Polymerization from the mispair without excision generates a heteroduplex molecule that yields a high proportion of easily distinguishable medium blue plaques. The ratio of faint to medium blue plaques thus describes the extent of mismatch excision prior to polymerization. Polymerase reactions were performed with the Klenow polymerase and the recombinant form of HIV-1 RT; each efficiently filled the single-strand gap. Transfection of the Klenow polymerase reaction products yielded few medium blue plaques, demonstrating that this enzyme excised 97% of the mismatched cytosines prior to polymerization (8, 10, 11). In contrast, HIV-1 RT efficiently extended from the mispair without excising the mismatched cytosine, since the proportion of medium blue plaques (341 out of 891 total plaques, 41%) observed upon transfection of the reaction products was similar to the 43% (355 out of 830 total plaques) obtained when the exonuclease-deficient avian myeloblastosis virus (AMV) RT (12) was used. These data confirm the absence of detectable proofreading exonuclease activity and also demonstrate the ability of HIV-1 RT to polymerize from mispaired termini.

We next examined the fidelity of DNA synthesis catalyzed by HIV-1 RT using a M13mp2-based nonsense codon reversion assay that detects single base substitution errors (10). An M13mp2 DNA molecule was constructed with a 361-nucleotide gap containing a single base change, $G \rightarrow A$ in the viral (plus) template strand at position 89 of the $lacZ\alpha$ coding sequence. This change creates a chain-termination TGA codon, resulting in a colorless plaque phenotype. This DNA was used as a substrate for a polymerization reaction in vitro. A portion of the products was analyzed by agarose gel electrophoresis, confirming gap-filling synthesis. The remaining products were used to

Table 1. Fidelity of HIV-1 and AMV reverse transcriptases and cellular DNA polymerases in a base subsitution reversion assay. The background revertant frequency for the assay was 2×10^{-6} . Error rates are calculated by subtracting the background, correcting for the expression [60%, see (8)] of errors in the complementary (minus) strand, and dividing by the number of detectable sites (three) [see (8, 18)]. The AMV RT data are from (17); polymerase α -primase (13); polymerase β and polymerase δ (10); and polymerase γ (8). Multiple experiments with the same enzymes indicate that the M13mp2-based mutation assays are highly reproducible with standard deviations that are usually between 10% and 20% of the mean value [for example, see (19)].

Enzyme	Plaques scored		Revertant	
	Total	Blue	(×10 ⁻⁶)	Error rate
Recombinant HIV-1 RT	730,000	75	100	1/18,000
AMV RT	370,000	29	78	1/24,000
Polymerase α-primase (calf thymus)	370,000	84	230	1/8.000
Polymerase B (rat hepatoma)	84,000	151	1800	1/1.000
Polymerase δ (calf thymus)	2,300,000	12	5.2	1/1.700.000
Polymerase γ (chick embryo)	600,000	5	8.3	1/290,000

Table 2. Forward mutant frequencies in DNA copied by HIV-1 RT. Wild-type $lacZ\alpha$ DNA was copied and used to transfect competent *E. coli*. The background mutant frequency in this assay (6×10^{-4}) was determined as described (14). The three mutant frequency values for HIV-1 RT yield an average of 393 (± 61) $\times 10^{-4}$. The polymerase β data were taken from (14), the AMV and M-MuLV data from (17).

Enzyme	Plaques scored		Mutant	Error
	Total	Mutant	$(\times 10^{-4})$	rate
Recombinant HIV-1 RT				
Experiment 1	6,109	284	460	1/1,500
Experiment 2	6.511	221	340	1/2.000
Natural HIV-1 RT	5,741	216	380	1/1.800
AMV RT	4.221	19	45	1/17.000
M-MuLV RT	19.112	53	28	1/30.000
Polymerase β (rat)	10,646	495	460	1/1,500

transfect competent α-complementation host cells to score the colors of the resulting M13mp2 plaques. Base substitution errors were detected as blue plaques and the reversion frequency (the proportion of blue to total plaques) reflected the base substitution error rate.

HIV-1 RT generated revertants at a frequency of 100×10^{-6} (Table 1). This is more than 50 times above the spontaneous background reversion frequency of uncopied DNA, showing that errors are indeed the result of DNA polymerization in vitro. The error rate at the three template nucleotides of the TGA codon is 1 for each 18,000 nucleotides polymerized (for calculation, see legend to Table 1). For comparison in the same assay, Table 1 also shows results for AMV RT and for the four known DNA polymerases (α , β , δ , and γ) in animal cells. HIV-1 RT has accuracy similar to that of AMV RT and an exonuclease-deficient DNA polymerase α -primase complex that had been freshly purified from calf thymus (13). The HIV-1 enzyme is 18 times more accurate than the DNA polymerase β for base substitution errors at this TGA codon. None of these DNA polymerases contain associated exonuclease activity. In contrast, HIV-1 RT is error-prone compared to the highly accurate DNA polymerases γ and δ , both of which contain associated proofreading exonuclease activity (8, 10)

Other exonuclease-deficient DNA polymerases commit a variety of errors in addition to single base substitutions, including simple frameshift errors and extensive and complex deletions (11, 14, 15). In addition, the known diversity of the HIV-1 genome described by DNA sequence analyses of individual isolates of HIV-1 includes diverse base substitutions, additions, and deletions (4, 5). Furthermore, for reasons yet unknown, these changes are not randomly distributed but are clustered within certain regions of the genome (4, 5).

Thus, in order to score a broad spectrum of polymerization errors, we next examined the fidelity of HIV-1 RT in the M13mp2 forward mutation assay (14). This assay uses the wild-type $lacZ\alpha$ sequence. Correct polymerization to fill a 390-nucleotide gap produces DNA that, when used to transfect the appropriate E. coli host strain, produces dark blue M13 plaques. Errors during synthesis are scored as lighter blue or colorless plaques. Since the assay measures loss of a gene function that is not essential for phage production (α -complementation of β -galactosidase activity), a wide variety of mutations at many different sites can be recovered and scored (14, 15).

Gap-filling polymerization reactions were performed with both the recombinant and

25 NOVEMBER 1988

the natural form of HIV-1 RT. Transfection of reaction products yielded the mutant frequencies shown in Table 2. Both sources of HIV-1 RT are highly inaccurate, producing errors about ten times as frequently as RTs from either AMV or Moloney murine leukemia virus (M-MuLV). HIV-1 RT is, in fact, as inaccurate in the forward mutation assay as DNA polymerase β , the least accurate DNA polymerase described to date. This result was unexpected, since HIV-1 RT appears much more accurate than polymerase β in the base substitution reversion assay (Table 1). This suggests that the error rate of these polymerases differs substantially from site to site.

We therefore determined the DNA sequences of 40 independent and randomly chosen mutants, 20 generated by the recombinant and 20 by the natural enzyme. Both enzyme preparations yielded similar results, producing three classes of errors: single base substitutions (7 out of 20 and 9 out of 20 sequenced mutations generated by the recombinant and natural enzymes, respectively), -1 errors (11 out of 20 and 10 out of 20) and +1 errors (2 out of 20 and 1 out of 20). The errors were dramatically nonrandom in their distribution within the target sequence. For example, although there are at least 110 different sites at which base substitution errors can be scored, 10 of the 16 single base substitution errors were $T \rightarrow C$ transitions at a single template nucleotide, position -36. This site is so mutable that even the small number of mutants sequenced allows us to calculate the error frequency for this T:dGTP mispair to be an extraordinarily high 1 out of 70. Similarly, although there are a variety of monotonous template nucleotide runs in the target sequence, 8 out of 21 detected -1 errors were within a single TTT sequence at positions 137 to 139. Here, the error rate per detectable nucleotide incorporated is 1/250, presumably because of a Streisinger slippage mechanism (16).

We also note that even after the removal of the base substitution and frameshift errors at the hot spots from the collection of HIV-1 mutations sequenced thus far, the base substitution error rate for HIV-1 RT is several times that for AMV RT. In addition, both the base substitution and frameshift hot spots observed with HIV-1 RT are unique when compared to the mutational spectra of eight other DNA polymerases whose error specificities have been defined (11, 14, 15, 17).

These results are consistent with the possibility that the infidelity of reverse transcription is responsible for the DNA sequence diversity among different HIV-1 isolates. A more detailed description of error specificity in vitro will allow comparisons to the in vivo database to further examine this possibility. Since it is possible that polymerase-mediated fidelity differences may reflect differences in the structure of active sites, it is also possible that error specificity differences between HIV-1 RT and normal cellular DNA polymerases may provide information that could be exploited for intervention therapy.

REFERENCES AND NOTES

- 1. A. S. Fauci, Science 239, 617 (1988). J. P. Dougherty and H. M. Temin, Mol. Cell. Biol. 168, 4387 (1987).
- J. M. Coffin, Cell 46, 1 (1986).
- 4. G. Myers et al., Human Retroviruses and AIDS, 1987 (Los Alamos National Laboratory, Los Alamos, NM. 1987)
- 5. B. H. Hahn et al., Science 232, 1548 (1986).
- R. A. Weiss et al., Eds., Molecular Biology of Tumor 6. Viruses: RNA Tumor Viruses (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, 1985), vols. 1 and 2, respectively
- C. Majumdar, J. Abbotts, S. Broder, S. H. Wilson, J. Biol. Chem., in press; S. H. Wilson, personal communication. Natural HIV-1 RT and bacterially expressed recombinant HIV-1 RT were the kind gift of S. H. Wilson. Due to the limited availability of the natural HIV-1 RT, we have examined it only for exonuclease activity and for forward mutation frequency. We have estimated the amount of recombinant HIV-1 RT in our assays based on an estimated molecular size of 95,000 [A. Chandra, T. Gerber, P. Chandra, FEBS Lett. 197, 84 (1986)]. We have used equivalent amounts of activity of the natural and recombinant enzymes in our assays based on their ability to fill the single-strand gap in the M13mp2 substrate.
- 8. T. A. Kunkel and A. Soni, J. Biol. Chem. 263, 4450 (1988).
- 9. T. A. Kunkel, Cell 53, 837 (1988).
-, R. D. Sabatino, R. A. Bambara, Proc. Natl. Acad. Sci. U.S.A. 84, 4865 (1987).
 K. R. Tindall and T. A. Kunkel, Biochemistry 27,
- 6008 (1988).
- 12. N. Battula and L. A. Loeb, J. Biol. Chem. 251, 982 (1976)
- 13. T. A. Kunkel et al., in Mechanisms and Consequences of DNA Damage Processing, ICN-UCLA Symposia on Molecular and Cellular Biology, New Series, E. Friedberg and P. Hanawalt, Eds. (Liss, New York, 1988), p. 521. The calf thymus DNA polymerase α was an intact form of the enzyme containing four subunits, including an active primase. This preparation was tested within 24 hours of its elution from the immunoaffinity column, was not frozen at any time before use, and contained no contaminating immunoglobulin G. The data obtained with this enzyme were consistent with previous results from experiments with human KB cell enzyme (15) and human HeLa cell polymerase α enzyme [J. D. Roberts and T. A. Kunkel, *Proc. Natl. Acad. Sci.* U.S.A. 85, 7064 (1988)] prepared by immunoaffinity chromatography.
- 14. T. A. Kunkel, J. Biol. Chem. 260, 5787 (1985). 15. _, ibid., p. 12866.
- ibid. 261, 13581 (1986); G. Streisinger et 16. al., Cold Spring Harbor Symp. Quant. Biol. 31, 77 (1986); G. Streisinger and J. E. Owen, Genetics 109, 633 (1985).
- 17. J. D. Roberts et al., Mol. Cell. Biol., in press.
- 18. T. A. Kunkel and P. S. Alexander, J. Biol. Chem. **261**, 160 (1986). T. A. Kunkel, *ibid.* **263**, 14784 (1988).
- We thank J. Drake, K. Tindall, and R. Swanstrom 20. for their helpful comments and critical evaluation of the manuscript. This work was supported in part by the NIH Intramural AIDS Targeted Antiviral Program.

26 July 1988; accepted 23 September 1988