Error-prone Polymerization by HIV-1 Reverse Transcriptase

CONTRIBUTION OF TEMPLATE-PRIMER MISALIGNMENT, MISCODING, AND TERMINATION PROBABILITY TO MUTATIONAL HOT SPOTS

(Received for publication, October 15, 1992)

Katarzyna Bebenek, John Abbotts, Samuel H. Wilson, and Thomas A. Kunkel

From the Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709 and the Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

We have observed previously that DNA template-directed polymerization by the type 1 human immunodeficiency virus reverse transcriptase is error-prone for single-nucleotide substitution, addition and deletion errors at homopolymeric sequences. We have also noted strong termination of processive synthesis at these positions (Bebenek, K., Abbotts, J., Roberts, J. D., Wilson, S. H., and Kunkel, T. A. (1989) J. Biol. Chem. 264, 16948-16956). Here we have tested three models to explain errors at these hot spots: template-primer misalignment for deletion errors, and dislocation and direct miscoding for substitution errors. The approach involves introducing single-nucleotide changes within or flanking the homopolymeric hot spots and examining the effects that these changes have on human immunodeficiency virus type 1 (HIV-1) reverse transcriptase error rate, error specificity, and termination probability. The results obtained suggest that single-nucleotide deletion errors in homopolymeric runs result from template-primer misalignment and that both direct miscoding and template-primer dislocation contribute to the base substitution hot spots. The data also suggest that base substitution errors at one position can be templated by the preceding nucleotide or either of the next two nucleotides. Frameshift error rates at homopolymeric sites were affected by changes in the sequences flanking the runs, including single-nucleotide differences in the single-stranded template strand and in the double-stranded primer region as many as six nucleotides distant from the hot spot. Both increases and decreases in frameshift fidelity were observed, and most of these correlated with concomitant increases or decreases in the probability that HIV-1 reverse transcriptase terminated processive synthesis within the run. These data provide further support for a relationship between the frameshift fidelity and the processivity of DNA-dependent DNA synthesis by HIV-1 reverse transcriptase.

A reasonable explanation for high retroviral mutation rates in general and the extensive genomic heterogeneity exhibited by the type 1 human immunodeficiency virus (HIV-1) in particular is inaccurate DNA replication by the proofreading exonuclease-deficient retroviral reverse transcriptase. Studies in vitro using a synthetic template-primer (Takeuchi et al., 1988), nonsense codon reversion assays (Preston et al., 1988; Roberts et al., 1988; Weber and Grosse, 1988), kinetic assays for base misinsertion (Preston et al., 1988; Ricchetti and Buc, 1989; Yu and Goodman, 1992), or mispair extension (Ferrino et al., 1988; Ricchetti and Buc, 1990; Yu and Goodman, 1992) and a forward mutation assay (Roberts et al., 1988; Bebenek et al., 1989) all demonstrate that HIV-1 reverse transcriptase is not highly accurate. The last study revealed that many HIV-1 reverse transcriptase errors occur at specific hot spots for base substitution and one-nucleotide frameshifts. These sites share a common feature: they are all homopolymeric nucleotide sequences.

We suggested three models to account for the error specificity of HIV-1 reverse transcriptase (Bebenek et al., 1989). First, the one-nucleotide addition or deletion hot spots may have resulted from template-primer slippage during synthesis, a model originally proposed by Streisinger et al. (1966). Continued synthesis with the misaligned substrate yields the frameshift error. Second, the base substitution hot spots, which are at the 5' or 3'-end of homopolymeric sequences, could have resulted from a dislocation mechanism (Fowler et al., 1974; Kunkel and Alexander, 1986; Kunkel and Soni, 1988b). In this model, the initiating event is also template-primer slippage, followed by correct incorporation of the next correct nucleotide. However, unlike simple slippage, the template-primer then realigns prior to continued synthesis, creating a mispaired intermediate that leads to the observed base substitution error at the end of the homopolymeric run. The third model, and an alternative hypothesis for the substitution hotspots, is direct miscoding, i.e. the insertion of an incorrect nucleotide followed by correct incorporation from the mispair.

In attempting to explain the large site-to-site variations in HIV-1 reverse transcriptase error rates, we also noted that the frequency of termination of processive DNA synthesis was higher at the frameshift hot spots than at other sequences in the target (Bebenek et al., 1989). This qualitative correlation between site-specific frameshift fidelity and processivity suggested that the formation and/or utilization of misaligned template-primers might be increased during the dissociation-reinitiation phase of the reaction.

With these observations in mind, we undertook the present study to examine three issues concerning error-prone DNA synthesis by the HIV-1 reverse transcriptase. First, we wished

*This work was supported by the National Institutes of Health Intramural AIDS Targeted Antiviral Program. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡To whom correspondence and reprint requests should be addressed: Laboratory of Molecular Genetics E2-01, National Institute of Environmental Health Sciences, P. O. Box 12233, Research Triangle Park, NC 27709. Tel.: 919-541-2644; Fax: 919-541-7613.

The abbreviations used are: HIV-1, type 1 human immunodeficiency virus; wt, wild type; pwt, pseudowild type.
to examine the idea that frameshift errors at the hot spots involved template-primer misalignment. Second, we wished to determine if the base substitution hot spots were consistent with the dislocation model, with direct miscoding, or both. Third, we wished to examine more closely the relationship between termination probability and frameshift fidelity. The approach we have taken for all three issues is to place typically silent single-nucleotide changes into the lacZa target sequence used for the fidelity assay. We then examined the effects that these changes have on HIV-1 reverse transcriptase frameshift and base substitution error rates.

EXPERIMENTAL PROCEDURES

Materials—The Escherichia coli strains, M13 bacteriophage derivatives, and materials for the fidelity experiments have been described previously (Kunkel, 1985; Kunkel and Soni, 1988a). The recombinant DNA polymerase reactions were performed as described previously (Bebenek et al., 1989). Oligonucleotides were purchased from Research Genetics (Huntsville, AL) and Synthecell Corp. (Rockville MD).

DNA Polymerase Reactions—Polymerase reactions (50 μl) contained 20 mM Hepes (pH 7.9), 5 mM dithiothreitol, 10 mM MgCl₂, 57 fmol of gapped DNA, 2.2 pmol of HIV-1 reverse transcriptase, and all four dNTPs at 1 mM. The mixtures were incubated for 1 h in 37 °C and terminated by adding EDTA to 15 mM. Twenty-μl aliquots were analyzed by agarose gel electrophoresis to ensure complete gap filling. All polymerase reactions reported here generated products filling all four primers. Mutants were scored as lighter blue or colorless plaque morphology of the resulting M13mp2 plaques.

RESULTS AND DISCUSSION

During synthesis templated by the wild type lacZa sequence, we previously observed substantial site-specific differences in HIV-1 reverse transcriptase error rates (see Fig. 1 in Bebenek et al., 1989). We began this study by introducing seven single-base changes into the wild type DNA sequence, creating a new template molecule referred to as pseudowild type 1 (pwt1). A second pseudowild type template (pwt2) containing seven single-base changes, two of which were in common with pwt1, was also constructed. Each of the changes (shown in Table 1 with the flanking nucleotide sequences) was chosen based on the models described below and the requirement that the changes did not significantly change the dark blue phenotype of the resulting M13mp2 plaques.

DNA sequences from these M13mp2 derivatives were used to construct gapped molecules, and synthesis reactions were performed in parallel with all three substrates. The reaction products were used to score the frequency of light blue and colorless lacZa mutants by transfection and plating. The mutant frequencies with the pwt1 and pwt2 substrates were 490 × 10⁻⁴ (the average of four determinations) and 450 × 10⁻⁴ (average of two determinations), respectively. These frequencies are similar to that obtained for reactions with the wild type substrate, 500 × 10⁻⁴ (average of five determinations). With all three templates, mutant frequencies were almost 100-fold higher than the mutant frequencies of the respective uncopied DNAs (5.1-6.7 × 10⁻⁴).

DNA sequence analysis was performed for 325 and 308 independent mutants, respectively, derived from the pwt1 and pwt2 templates. The results of these analyses are shown in Tables II-IV and Figs. 1 and 2. They are compared below with the error specificity of HIV-1 reverse transcriptase for single-base substitutions at the hot spots.

---

**Table I**

Silent nucleotide sequences changes in templates used in this study

<table>
<thead>
<tr>
<th>Nucleotide change</th>
<th>Nucleotide no.</th>
<th>Template designations</th>
<th>Flanking nucleotide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>From</td>
<td>To</td>
<td>ss template</td>
<td>ds primer stem</td>
</tr>
<tr>
<td>C A</td>
<td>-41</td>
<td>pwt1</td>
<td>G G C A C C C</td>
</tr>
<tr>
<td>C G</td>
<td>-37</td>
<td>pwt1</td>
<td>C C C A A G G</td>
</tr>
<tr>
<td>C A</td>
<td>-30</td>
<td>pwt2</td>
<td>A T T T A C A A</td>
</tr>
<tr>
<td>G C</td>
<td>69</td>
<td>pwt1</td>
<td>G G C G T C T C</td>
</tr>
<tr>
<td>G A</td>
<td>69</td>
<td>pwt2</td>
<td>G G C G T C T C</td>
</tr>
<tr>
<td>A G</td>
<td>92</td>
<td>pwt2</td>
<td>A C T G G G G A</td>
</tr>
<tr>
<td>T A</td>
<td>68</td>
<td>pwt1</td>
<td>A A A A C C C C</td>
</tr>
<tr>
<td>G C</td>
<td>102</td>
<td>pwt1</td>
<td>C C C A G G G</td>
</tr>
<tr>
<td>G A</td>
<td>102</td>
<td>pwt2</td>
<td>C C C T G G C</td>
</tr>
<tr>
<td>C T</td>
<td>107</td>
<td>pwt1 and 2</td>
<td>G C C T T A C T C</td>
</tr>
<tr>
<td>T A</td>
<td>137</td>
<td>pwt1 and 2</td>
<td>A T C C C C C C</td>
</tr>
</tbody>
</table>

**Table II**

Frequency of various classes of mutants

The mutant frequency for each mutational class was calculated based on the number of mutants sequenced (325 and 308 from reactions with pwt1 and pwt2 templates, respectively), the number of mutants in each class, and the overall mutant frequencies (490 × 10⁻⁴ and 450 × 10⁻⁴ for reactions with pwt1 and pwt2 templates, respectively). In the "other" category are mutants with more than 1 base deleted or added and mutants with complex changes.

<table>
<thead>
<tr>
<th>Mutational class</th>
<th>No. of mutants</th>
<th>Mutation frequency (× 10⁻⁴)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pwt1 template</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-base frameshifts</td>
<td>143</td>
<td>220</td>
</tr>
<tr>
<td>Single-base substitutions</td>
<td>113</td>
<td>170</td>
</tr>
<tr>
<td>Multiple mutants</td>
<td>49</td>
<td>74</td>
</tr>
<tr>
<td>Other</td>
<td>21</td>
<td>32</td>
</tr>
<tr>
<td>pwt2 template</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-base frameshifts</td>
<td>90</td>
<td>130</td>
</tr>
<tr>
<td>Single-base substitutions</td>
<td>157</td>
<td>230</td>
</tr>
<tr>
<td>Multiple mutants</td>
<td>53</td>
<td>77</td>
</tr>
<tr>
<td>Other</td>
<td>8</td>
<td>12</td>
</tr>
</tbody>
</table>

---

**Table III**

Frameshifts-During synthesis templated by the wild type lacZa sequence, we previously observed substantial site-specific differences in HIV-1 reverse transcriptase error rates (see Fig. 1 in Bebenek et al., 1989). We began this study by introducing seven single-base changes into the wild type DNA sequence, creating a new template molecule referred to as pseudowild type 1 (pwt1). A second pseudowild type template (pwt2) containing seven single-base changes, two of which were in common with pwt1, was also constructed. Each of the changes (shown in Table 1 with the flanking nucleotide sequences) was chosen based on the models described below and the requirement that the changes did not significantly change the
Bebenek et al., 1989). All five hot spots were homopolymeric 
sequences, suggesting that the errors resulted from template-
primer slippage. According to this model, after some 
amount of correct incorporation within the run, slippage could allow 
a misaligned intermediate to form which is stabilized by 
correct base pairs (e.g. see Fig. 3, top). Continued synthesis 
from the misaligned intermediate would fix the extra nucleo-
tide within double-stranded DNA, leading to a deletion if the 
extra nucleotide is in the template strand (as in Fig. 3) or an 
addition if the extra nucleotide is in the primer strand.

To test the slippage model, we altered the sequences at two 
of the previously observed slip-jump hot spots by introducing 
phenotypically silent base changes (boldface type in Fig. 3B).

If HIV-1 reverse transcriptase deletion errors at these sites 
occur via misaligned intermediates, the changes should reduce 
the error rates with the pwt1 template because the predicted 

misaligned intermediates (Fig. 3B) would be stabilized by 
fewer correct base pairs. The fidelity analyses with the pseu-
dowild type templates support this idea. When the TTT run 
at positions 137-139 was altered to ATT, the error rate 
decreased ≥ 75-fold (compare parts A and B in Fig. 3). When 
the CCC run at positions 106-108 was altered to CTC, the 
error rate decreased ≥ 36-fold. Note that the error rate at an 
unaltered GGG run at positions 88-90 was in the same range 
for the two templates. This serves as an internal control 
validating the reproducibility of the hot spot assignment at 
this position. The observed decrease in error rates at the
HIV-1 Reverse Transcriptase Mutational Hot Spots

**FIG. 1. Spectrum of single-base errors on the pwtl template.** Three lines of viral (+) strand sequence are shown. The base substitutions shown above the template indicate the new base found in the viral DNA. For frame shifts (below the lines) the loss of a base is indicated by a △, and the addition of a base is indicated by a ▲. When a frameshift occurs in homopolymeric run, it is not known which base was lost or added. The seven changes introduced into this substrate are underlined and in bold type.

**FIG. 2. Spectrum of single-base errors on the pwt2 template.** For an explanation, see the legend to Fig. 1.

altered sites in the pwtl template indicate that one-nucleotide frameshift hot spots require a homopolymeric run and are consistent with a local template-primer misalignment during synthesis by HIV-1 reverse transcriptase.

During DNA synthesis with the wild type substrate, HIV-1 reverse transcriptase generated a high frequency of single-base substitutions at several different template sequences (see Figs. 1 and 2 in Bebenek et al., 1989). Two pathways that may explain how these errors arose are illustrated in Fig. 4, using as an example the T → C hot spot at position -36. Direct misinsertion of dGTP opposite the template T (Fig. 4A, left),
followed by extension from the mispair, could explain the observed hot spot. Alternatively, following misalignment to create an unpaired template T and then correct incorporation of dGTP opposite template C, realignment could generate the same base substitution error (Fig. 4A, center).

In an attempt to distinguish between these two pathways, we changed the nucleotide at position -37 from C to either A (in the pwt2 template) or to G (in the pwt1 template). With template A at position -37, miscoding still predicts T→C errors at position -36 (Fig. 4B, left), whereas dislocation with an unpaired template base predicts T→A errors (Fig. 4B, center). Similarly, with template G at position -37, miscoding predicts T→C errors at position -36, whereas dislocation with an unpaired template base predicts T→G errors (Fig. 4C, center). Hypothetically, the dislocation mechanism could operate via an unpaired nucleotide in the primer strand (Fig. 4, A–C, right). This predicts that whatever nucleotide is present at position -37 (underlined) will be changed to a template T, the 3′-neighboring template nucleotide in all three substrates.

Interpretations using this experimental strategy are valid only for those base substitutions known to yield a detectable mutant plaque color in the forward mutation assay performed with each template. We therefore generated each of the required changes relevant to testing these models and found that they indeed had detectable light blue plaque phenotypes (Table V).

The results of the fidelity analysis with all three templates are shown in Fig. 5A, and the error rates per detectable nucleotide polymerized are given in Table VI. The effect of changing the nucleotide at position -37 on error rates at position -36 were substantial. The rate for T→C substitutions at position -36 when -37 is a C (1/60) is reduced by 6.7-fold (to 1/400) when a template A was placed at position -37, and by 17-fold (to 1/1,000) when a template G was placed at position -37. Concomitantly, when a template A was present at -37, the error rate for T→A at position -36, at 1/130, was 52-fold and 31-fold greater than when -37 was a C (error rate ≤ 1/6,800) or a G (error rate ≤ 1/4,000), respectively. These data strongly suggest that many of the T→C and T→A errors at position -36 result from dislocation involving an unpaired template nucleotide, as depicted in Fig. 4, A and B (center). Consistent with this were the -T errors detected at the TTT run, also implying a misaligned intermediate involving an unpaired template nucleotide that did not realign prior to further polymerization.

The error rate for T→C substitutions at -36 continued to be high even when A or G was present at -37 (error rates of 1/400 and 1/1,000, respectively, Table VI and see Fig. 5A). These errors cannot be explained by dislocation to adjacent template nucleotides. They could reflect direct misincorporation of dGTP opposite the template T, or, less likely, dislocation to more distant nucleotides.

Further support for a dislocation model involving an extra
A. Template C at -37

Misincorporation of dGTP

\[ A\rightarrow T \rightarrow G \rightarrow \cdot \cdot \cdot \cdot \]
\[ 5\:'-G\rightarrow C\rightarrow T\rightarrow T\rightarrow A\rightarrow C\rightarrow \]

Error fixation

\[ G\rightarrow A\rightarrow T\rightarrow G \rightarrow \cdot \cdot \cdot \cdot \]
\[ 5\:'-G\rightarrow G\rightarrow T\rightarrow T\rightarrow A\rightarrow C\rightarrow \]

B. Template A at -37

\[ G\rightarrow A\rightarrow T\rightarrow G \rightarrow \cdot \cdot \cdot \cdot \]
\[ 5\:'-G\rightarrow A\rightarrow T\rightarrow T\rightarrow A\rightarrow C\rightarrow \]

Realignment to form mispair

\[ T\rightarrow A\rightarrow A\rightarrow T\rightarrow A\rightarrow C\rightarrow \]
\[ 5\:'-G\rightarrow T\rightarrow T\rightarrow T\rightarrow A\rightarrow C\rightarrow \]

C. Template G at -37

\[ G\rightarrow A\rightarrow A\rightarrow T\rightarrow G \rightarrow \cdot \cdot \cdot \cdot \]
\[ 5\:'-G\rightarrow G\rightarrow T\rightarrow T\rightarrow A\rightarrow C\rightarrow \]

Realignment to form mispair

\[ C\rightarrow A\rightarrow A\rightarrow T\rightarrow A\rightarrow C\rightarrow \]
\[ 5\:'-G\rightarrow T\rightarrow T\rightarrow T\rightarrow A\rightarrow C\rightarrow \]

HIV-1 Reverse Transcriptase Mutational Hot Spots

Fig. 4. Models for base substitution errors at template sites -37 and -36. Shown are the models leading to substitution errors at position -37 (underlined) and position -36, when the template base at position -37 is a C (part A), an A (part B), or a G (part C).
The mutant color with that of wild type on the same plate. Base substitutions are shown above the original mutant spectrum (see Fig. 2 in Bebenek et al., 1989)

**Table V**

<table>
<thead>
<tr>
<th>Substitution</th>
<th>Mutant phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>T36C</td>
<td>CTTT ATTT GTTT</td>
</tr>
<tr>
<td>T36A</td>
<td>2 2 2</td>
</tr>
<tr>
<td>T36G</td>
<td>2 2 2</td>
</tr>
<tr>
<td>X3TT</td>
<td>2 3 1</td>
</tr>
<tr>
<td>G90A</td>
<td>3 3 3</td>
</tr>
<tr>
<td>T70C</td>
<td>CTTTT ATTTTT GTTTT</td>
</tr>
<tr>
<td>T71C</td>
<td>2 3 4</td>
</tr>
<tr>
<td>T70A</td>
<td>3 3 3</td>
</tr>
<tr>
<td>T70G</td>
<td>1 2 2</td>
</tr>
<tr>
<td>X69T</td>
<td>3 3 3</td>
</tr>
</tbody>
</table>

A. Spectra of errors at positions -37 to -34 with a C, A or G at position -37.

B. Spectra of errors at position 90 with an A or G at position 92.

C. Spectra of errors at positions 69 to 73 with a C, A or G at position 69.

**Fig. 5.** Effects of template nucleotide changes on substitutions at three positions. The sequences of three mutational hot spots are shown. The underlined bases in parts A, B, and C are at positions -37, 69, and 92, respectively. Base substitutions are shown above the nucleotide in the original sequence that has changed and indicate the new base found in the viral DNA. Frameshifts are indicated by triangles below the lines of sequence; Δ, one-nucleotide deletion; ■, two-nucleotide deletion; Δ, one-nucleotide addition. The data are taken from Fig. 1 and Table II in Bebenek et al. (1989) and from Figs. 1 and 2 and Tables III and IV. The 5'-most T (position -29) was a hot spot for a phenotypically silent T → C change that was recovered, always in conjunction with another detectable change, seven times in the collection of 438 sequenced mutants. This observation was confirmed in the present collection of mutants containing two or more changes obtained with the pwt1 template (Table III). Here, eight of 326 mutants contained this same T → C substitution. No T → A substitutions at position -29 were recovered in either collection. However, when the template sequence motif was changed to ATTT (in pwt2, Fig. 2), among the collection of multiple mutants, two silent T → A substitutions were recovered at position -29 out of 308 mutants sequenced (Table IV), precisely as predicted by the dislocation model.

Interestingly, when a template G was at position -37, the error rate for G → T errors at this site was high (1/300, Table VI) relative to the rate for C → T or A → T errors at this site when either C or A was present (Fig. 5A and Table VI). This fact, and the high error rate for +T at the TTT run when position -37 was a G, suggest the involvement of a misaligned intermediate containing an unpaired primer-strand nucleotide, as depicted in Fig. 3C (right). This was also suggested by the observation with the wild type template (Bebenek et
Errors at position 70 are only observed when position 69 is an 
A, and at position 69, the error rate for T
T errors at the ATTTT site in the pwt2 template (Fig. 5C,
and Table VI) are remarkably consistent with the observa-
tions described above at position -36 and provide further
support for the dislocation model. Thus, when a template C
positions 69 to 73—Using the 
M13mp2 forward mutation assay, we had previously observed high error rates for -T framenesses and G → T base substitutions by a cellular DNA polymerase, pol β, at a GTTTT sequence at position 69-73 and at a GTT sequence at positions 102-104 (Kunkel, 1985). Later analyses (Kunkel and Soni, 1988b and Boosalis et al., 1989) suggested that the base substitution errors resulted from dislocation. Encouraged by the observation that -T and +T framenesses were also generated at these positions by HIV-1 reverse transcriptase (see Fig. 1 in Bebenek et al., 1989), we introduced G → C changes (pwt1 template) or G → A changes (pwt2 template) at positions 69 and 102 to determine if errors consistent with the dislocation model could be detected.

The results at positions 102-104 with the two altered tem-
plates (Figs. 1 and 2) were not highly informative, since only a few mutants were recovered. However, the three T → C
mutations observed at position 103 when position 102 is a template C (Fig. 2) are consistent with the dislocation model. Furthermore, the results for errors at position 69-73 (Fig. 5C and Table VI) are remarkably consistent with the observa-
tions described above at position -36 and provide further support for the dislocation model. Thus, when a template C
is at position 69, the error rate for T → C at position 70 (1/300) is more than 10-fold higher than when a template A or 
G is present at position 69 (Table VI). Similarly, T → A
errors at position 70 are only observed when position 69 is an 
A. The observation of both A → T errors at position 69 and
+T-errors at the ATTTT site in the pwt2 template (Fig. 5C,
middle) are both consistent with a misaligned intermediate containing an extra primer-strand nucleotide.

The error specificity analysis at these sequences also sug-
gests that base substitution errors via transient misalignment 
may involve more than a single unpaired nucleotide. Note that with the 5'-CTTTTT motif (Fig. 5C, left), the rate for T → C substitutions at the 5'-penultimate nucleotide in the run (position 71, underlined above) is 1/400, more than 10-fold higher than with the other two templates (Table VI). Notably, it is only with this template that -TT mutants were observed (Fig. 5C, left, closed squares). Both outcomes can be explained by misaligned template-primers containing two unpaired template T residues.

Effects of Local Sequence Environment on Frameshift Error Rate—We have observed previously that the HIV-1 reverse transcriptase error rate for addition and deletions in homopolymeric sequences depends not only on the length and nucleotide composition of the run but also on the flanking nucleotide sequence (Bebenek et al., 1989). The clearest example of a neighboring-nucleotide effect was the ≥ 49-fold difference in rate of loss of a C in a CCC run at position 106-108 versus a CCC run at position 95-97 (top two lines in Table VII). Comparison of the flanking nucleotide sequences at the two runs reveals that they have an identical template 5'-neighbor, an A.

In an attempt to understand over what distance flanking nucleotides might affect frameness fidelity, we introduced a silent T → A change at position 98 in the pwt1 template. This change now generates two CCC runs with identical immediately flanking 5'- and 3'-neighbors. Despite this, the frequency of loss of a C at positions 95-97 (1.5 × 10⁻⁴, Table VII) remained 36-fold lower than at positions 106-108. We then extended the analysis to consider a third site, a CCCC run at positions -41 through -44. Results with the wild type template (Table VII, line 4) indicated that the -C error rate was 15-fold lower than at positions 105-108, despite the fact that it was an even longer run of identical nucleotides and shared a common template 5'-A neighbor. Introducing a single silent C → A change at position -41 into the pwt1 template (Table VII, last line) created a six-nucleotide sequence that was identical to positions 105-110 in the wild type template (Table VII, top line). Despite this, the rate of
loss of a C in the CCC run was still 12-fold different for the two sites. Thus frameshift fidelity is affected by nucleotide sequence differences beyond the immediate 5' neighbor in the template strand and the two immediate 3' neighbors in the primer stem flanking the run. This illustrates that frameshift error rates depend not only on the length and nucleotide composition of the homopolymeric run itself but also on the longer range sequence context.

**Frameshift Fidelity and Termination Probabilities**—We examined previously whether the processivity of synthesis correlates with frameshift fidelity by comparing hot spots for frameshift errors with the positions at which HIV-1 reverse transcriptase terminated processive synthesis (Bebenek et al., 1989). Those results indicated a positive correlation between termination probability and frameshift error rates in homopolymeric sequences. One interpretation of this correlation is that the formation and/or utilization of misaligned template-primers may be increased during the dissociation-reinitiation phase of the reaction, leading to an increased frameshift error rate. Alternatively, mutations may result from some other low frequency event that is not detected by the gel electrophoretic analysis of product molecules. One of the major objectives of the present study was to attempt to distinguish between these possibilities by altering the template nucleotide sequence and reexamining frameshift fidelity and termination probability.

Detailed quantitative analysis of termination at each position in the wild type template (Abbotts et al., 1993) revealed high and low termination likelihood sequences for HIV-1 reverse transcriptase. When a parallel analysis was performed with the wt1 and wt2 templates, substantial differences in site-specific termination probabilities were observed, several of which were predicted by the deduced trend sequences and the nucleotide changes introduced into each template (Abbotts et al., 1993). Some of these sites were homopolymeric runs, providing instances in which a correlation with frameshift fidelity could be examined (Table VIII). For example, consider the difference in termination probability and the rate of addition of a T to the TTTT run at positions 70–73 which accompanies the G → C change at position 69 (the first comparison in Table VIII). Using the simplest (but not the only) model for T addition errors, termination after incorporating the final complementary A at position 70 would yield the maximally stable intermediate, having an unpaired primer-strand A stabilized by four correct A-T base pairs. In defining nucleotide 70 as position 0, nucleotide 69 is in the template strand at the +1 position. The change of a G to a C at position 69 generates a sequence motif that is one nucleotide further diverged from the high termination likelihood sequence, which has a G at the +1 position, and is one nucleotide closer to the low termination likelihood sequence, which has a C at the +1 position. This predicts a decrease in termination probability at position 70, as observed. Including all three sites in which misalignment following incorporation could yield a + one nucleotide intermediate stabilized by at least one terminal A-T base pair (i.e., positions 70, 71, and 72), termination probability decreased by 13.8%. If the relationship between termination probability and fidelity holds, this predicts a decrease in the rate of + T errors in the TTTT run. This is exactly what was observed. Using the error specificity data for the wt and wt1 templates (Bebenek et al., 1989, Fig. 1 and Table II), the calculated error rate for + T errors at positions 70–73 decreased 4.4-fold (Table VIII).

For deletion errors, consider the difference in termination probability and rate of loss of A from an AAAA run at positions 91–94 which accompanies the A change at position 98 (second comparison in Table VIII). For the loss of an A, termination after incorporating a complementary A at positions 92, 93, and 94 is considered in the analysis because in each case a deletion intermediate could form which has an unpaired template A potentially stabilized by at least one correct A-T base pair. Position 91 is not counted here, since insertion of the fourth complementary T precludes formation of a deletion intermediate. Position 98 is in the double-stranded primer region, at the −6 position relative to position 92 and at the −5 position relative to position 93. The change of a T to an A at position 98 thus generates a sequence motif that, in relation to positions 92 and 93, is one nucleotide close to the deduced high termination likelihood sequence, which has an A at position −6 and −5. Consistent with prediction, the termination probability in the run increased concomitantly with the T → A change at position 98 (Table VIII). Consistent with the correlation between termination probability and frameshift error rate, the error rate also increased.

In a similar manner, the third through sixth pairwise comparisons in Table VIII also reveal a concomitant increase or decrease in fidelity and termination probability. Effects were seen for changes in either the single-stranded template or the double-stranded primer region, again implying that both regions are important. The seventh comparison, predicting an increase in + T errors at positions −36 through −34, was noninformative. No such errors were scored on either template, although termination probability did increase. The final three comparisons in Table VIII represent exceptions to the simple prediction. In each instance only one of the two parameters (fidelity or termination probability) responded as predicted by the high and low termination likelihood sequences.

Two aspects of this comparative analysis deserve special comment. First, the reaction conditions for the two assays are not identical. The ratio of enzyme to template-primer is high (typically 30:1) for fidelity measurements, to assure complete gap-filling synthesis. However, it is much lower (1:5) for analysis of termination. We have demonstrated that under the latter conditions, reinitiation on previously extended primers did not occur to a detectable extent, such that enzymes terminating synthesis were more likely to reinitiate on
processivity is important for frameshift fidelity, for two reasons. First, for the comparisons shown in Table VIII, the changes in fidelity and termination probability are not constant. However, we do not necessarily provide a constant probability of generating a frameshift intermediate. Secondly, the relationship between frameshift fidelity and termination probability is not absolute. For any single site, frameshift errors are consistently less frequent than termination (Table VIII). Furthermore, any comparison of two related sequences (Table VIII) changes in fidelity and termination probability are not constant. However, we do not consider these facts to be inconsistent with the idea that processivity is important for frameshift fidelity, for two reasons. First, for the comparisons shown in Table VIII, the absolute increase in termination far exceeds the absolute increase in frameshift errors. For example, termination in the AAAA run increased by 0.012 whereas -A91 errors increased by 0.001063. Thus, the seemingly modest increase in termination is more than sufficient to account for the substantial increase in frameshifts. Second, other factors may preclude a direct quantitative relationship. Studies with DNA polymerase I large fragment suggest that termination of synthesis may occur at any of (at least) four steps in an ordered reaction mechanism (Abbott et al., 1988). Extrapolating this to synthesis by HIV-1 reverse transcriptase, it may be that only a subset of total termination events is related to frameshift fidelity. As noted previously (Bebenek et al., 1989; Kunkel, 1990), several parameters influence frameshift error rates, including the length and base composition of the run, the influence of neighboring bases, and the structure of the polymerase and its contacts with the template-primer. Given these variables, there is no reason to expect that all template-primer-template-primer discontinuity reassociation events should necessarily result from both base miscoding and template-primer misalignment; (ii) base substitutions result from both base miscoding and template-primer misalignment.

Summary—The results presented here for DNA-dependent DNA synthesis by HIV-1 reverse transcriptase suggest that (i) single-nucleotide deletion errors in homopolymeric runs result from template-primer misalignment; (ii) base substitutions result from both base miscoding and template-primer misalignment.
dislocation; (iii) base substitution errors at one position can be templated by the preceding nucleotide or either of the next two nucleotides; (iv) frameshift error rates at homopolymeric sites were affected by changes in the sequences flanking the runs, including single-nucleotide differences in the single-stranded template strand and in the double-stranded primer region; (v) increases and decreases in fidelity often correlate with concomitant increases or decreases in the probability that HIV-1 reverse transcriptase terminated processive synthesis within the run, supporting the idea that some frameshift errors are associated with enzyme dissociation-reassociation.

Acknowledgments—We thank Lisa A. Clarke for expert technical assistance in the sequence analysis of mutants and R. Stephen Lloyd and Roel M. Schaaper for critical evaluation of the manuscript.

REFERENCES
Rochetti, M., and Buc, H. (1990) EMBO J. 9, 1583-1593