Arginine 72 in human immunodeficiency virus type 1 reverse transcriptase (RT), a highly conserved residue among retroviral polymerases and telomerases, forms part of the binding pocket for the nascent base pair. We show here that replacement of Arg72 by alanine strongly alters fidelity in a highly unusual manner. R72A reverse transcriptase is a frameshift and base substitution mutant polymerase whose increased fidelity results both from increased nucleotide selectivity and from a decreased ability to extend mismatched primer termini. Thus, Arg72-substrate interactions in wild-type human immunodeficiency virus type 1 RT can stabilize incorrect nucleotides allowing misinsertion and promoting extension of mismatched and perhaps misaligned template-primers. In contrast to the higher fidelity at most sites, R72A RT is highly error-prone for misincorporations opposite template T in the sequence context: 5'-CTGG. Surprisingly, this results mostly from a 1200-fold increase in the apparent $K_m$ for correct dAMP incorporation. Thus, Arg72 interactions with substrate are critical for the stability of the correct T'dAMP base pair when the 5'-CTGG sequence is present in the binding pocket for the nascent base pair. Collectively, the data show that a mutant polymerase may yield higher than normal average replication fidelity, yet paradoxically place specific sequences at very high risk of mutation.

The fidelity of DNA polymerization depends on discrimination between correct and incorrect nucleotides during insertion. Where exonucleolytic proofreading is possible, fidelity also depends on the preferential extension of the newly formed correct base pair but not mismatched or misaligned template-primers. Energetic differences between correct and incorrect base pairs in duplex DNA in solution are insufficient to account for the degree of DNA polymerase discrimination at the insertion step (reviewed in Ref. 1), implying important roles for DNA polymerases in determining replication fidelity. Recent structural studies of DNA polymerases (Refs. 2-7; briefly reviewed in Ref. 8) support the idea that discrimination may largely depend on geometric constraints in the binding pocket for the nascent base pair (reviewed in Ref. 9). The shape of this pocket can precisely accommodate complementary Watson-Crick base pairs, but may exclude non-complementary base pairs that have altered geometry and/or that form hydrogen bonds with solvent (10, 11).

One approach to probe the role of polymerases in determining replication fidelity is to study mutant polymerases whose specific protein-substrate interactions are altered. Here we use this strategy to investigate the fidelity of HIV-1 RT. HIV-1 RT lacks an intrinsic 3'→5' exonucleolytic activity (12) and therefore does not proofread replication errors. It has misinsertion properties typical of other proofreading-deficient DNA polymerases, but is promiscuous in its ability to extend mismatched primer termini (13, 14). This latter property has been suggested to reflect the ability and need to efficiently strand switch during viral replication in vivo (15, 16). HIV-1 RT is inaccurate for frameshift and base substitution errors initiated by template-primer slippage in homonucleotide runs (17–19). Important to this type of infidelity are five amino acids in HIV-1 RT, which interact with the DNA minor groove two to six base pairs upstream of the active site. When changed to alanine, each of the five mutant RTs has lower DNA binding affinity, lower processivity and altered fidelity for errors initiated by template-primer slippage (20).

More recently, the structure of a ternary complex of HIV-1 RT bound to template-primer and dTTP has been described (6). This structure in which the RT is poised for catalysis provides the location of amino acids at the active site that interact with the catalytic metals, the primer terminus and the incoming dNTP. Comparison to the structure of a RT-DNA complex (21) reveals that binding of the dNTP results in a "closed" conformation (Fig. 1A) wherein amino acids in the β3-β4 loop of the fingers subdomain move to form one wall of the binding pocket for the nascent base pair (Fig. 1B). Among these amino acids is Arg72 (red in Fig. 1, B and C). The guanidinium group of Arg72 stacks against the base of the incoming dNTP (Fig. 1, B and C) and donates hydrogen bonds to the α-phosphate (Fig. 1C). Although several residues in the dNTP binding pocket of RT are altered during acquisition of drug resistance to nucleoside inhibitors (e.g., Lys60, Leu74, and Gln151), RT mutants with an amino acid substitution at position 72 have not been reported among drug-resistant HIV-1 isolates, nor were Arg72 substitutions found in a screen for HIV-1 RT mutants that complement the growth defect of a temperature-sensitive E. coli DNA polymerase I mutant (22). These results, as well as the high conservation of Arg72 among retroviral reverse transcriptases and telomerases (23, 24) and the fact that an R72A mutant RT has a reduced $k_{cat}$, is resistant to a pyrophosphate analog, is deficient in pyrophosphorylase, and is compromised in trans-
location (25), point to the functional importance of Arg72. This, and the structural data indicating that Arg72 forms part of the binding pocket for the nascent base pair, prompted the current examination of the effects of modifying the interactions between Arg72 and the incoming nucleotide on the fidelity of DNA synthesis. Here we show that a HIV-1 RT mutant with an R72A substitution has fidelity properties that are unique among known DNA polymerases.

EXPERIMENTAL PROCEDURES

Mutagenesis and Protein Purification—Arg72 in HXB2 HIV-1 RT was changed to alanine by oligonucleotide-directed mutagenesis as described (26, 27). The entire coding sequence was confirmed to be correct by DNA sequence analysis. R72A and wild-type RT p66/p66 were expressed as described (27, 28) and purified free of contaminating exonuclease activities using a modification of the method described by Beard et al. (29).

Measurements of Fidelity during Gap-filling—The fidelity of RNA template-dependent gap-filling synthesis was measured as described (19), using M13mp2 DNA substrates. Synthesis errors in the forward assay are scored as light blue or colorless M13 plaques. Synthesis errors in the reversion assays are scored as blue-plaque revertants of M13 substrates containing mutations encoding colorless plaques. Base substitutions were monitored at a TGA codon at position 87–89 in the reporter gene, with any of eight different mispairs yielding blue plaques. Single-nucleotide deletion errors were monitored in a template TTTTT run (30). Synthesis reactions (25 μl) contained 35 fmol of gapped M13mp2 DNA, 1–5 pmol of RT, 1 mM each of the four dNTPs, 20 mM Tris-HCl, pH 7.5, 50 mM KCl, 10 mM MgCl2, and 2 mM dithiothreitol. Reactions were incubated for 1 h at 37 °C, and all reactions filled the gaps to apparent completion (19).

Steady-state Kinetics—Reactions (50 μl) contained 50 pmol Tris-HCl, pH 7.4, 10 mM MgCl2, 100 mM KCl, and 1 μM 3′-primer termini. The T-P was poly(rA)-oligo(dT)20, poly(rC)-oligo(dG)15, or poly(dC)-oligo(dG)15, and the variable nucleotide was [α-32P]TTP or [α-32P]dGTP. Reactions were initiated by the addition of enzyme, incubated at 22 °C and quenched by the addition of 20 μl of 0.5 M EDTA, pH 8.0. Quenched reaction mixtures were processed as described previously (29), and data were fitted to the Michaelis-Menten equation using nonlinear least-squares methods.

Missense Kinetics—Oligonucleotide template and primers were from Oligos Etc. Oligonucleotide primers were 5′-radiolabeled with T4 polynucleotide kinase and unincorporated nucleotides removed by gel filtration on a Super Sect-D, G-25 column (5 Prime—3 Prime, Inc.). T-Ps were prepared by incubating template and primer oligonucleotides in a 2:1 ratio at 85 °C for 1 min and then allowing the mixture to cool slowly to room temperature. The following T-Ps were used.

| Template/Primer and dNTP substrate | Enzyme | $K_{m, dNTP}$ | $k_{cat}$ | $k_{cat}/K_{m, dNTP}$ ($\times 10^3$) | $k_{cat}/K_{m, dNTP, R72A}$ |
|-----------------------------------|--------|---------------|-----------|-------------------------------------|--------------------------|
| Poly(rA)-oligo(dT)20, TTP         | Wild-type | 2.4 ± 0.2     | 0.22 ± 0.01 | 92                                  | 1150                     |
|                                  | R72A    | 92 ± 6        | 0.007 ± 0.001 | 0.08                                | 240                      |
| Poly(rC)-oligo(dG)15, dGTP        | Wild-type | 0.7 ± 0.1     | 0.17 ± 0.01  | 2.8                                 | 92                       |
|                                  | R72A    | 11.1 ± 0.9    | 0.029 ± 0.001 | 110                                 | 42                       |
| Poly(dC)-oligo(dG)15, dGTP        | Wild-type | 7.6 ± 1.5     | 0.83 ± 0.08  | 5.4                                 | 20                       |
|                                  | R72A    | 2.4 ± 0.03    | 0.013 ± 0.001 | 110                                  | 20                       |

The concentration of the wild-type enzyme in all reactions with the A and B templates was 30 nM. In reactions with the C template, the concentration of the enzyme was 30 nM for dATP insertion and 150 nM for dGTP insertion. The concentration of the R72A RT in reactions with all three templates was 30 nM and 150 nM for dATP and dGTP insertion, respectively. Reactions were initiated by the addition of a solution of RT and DNA, incubated at 37 °C, and quenched between 1 and 10 min by adding an equal volume of 95% formamide, 20 mM EDTA, 0.05% bromphenol blue, and 0.05% xylene cyanol FF. Products were analyzed on 16% polyacrylamide gels with 8 M urea. Substrate and product bands were quantified by phosphorimaging and analyzed as described previously (32). The data were fitted to the Michaelis-Menten equation by nonlinear least-squares methods, to obtain apparent $k_{cat}/K_m$ values.

Mismatch Extension Kinetics—Reactions were as outlined for insertions. Templates A and C were used with primers one nucleotide longer than the corresponding primers in the diagram to create either correctly matched (T-dATP) or mismatched (T-dGTP) primer termini. The insertion of the next correct nucleotide, dGTP, was measured. Varying the concentration of the next correct nucleotide, dGTP, enabled us to determine the apparent $k_{cat}/K_m$ values for extension of matched and mismatched base pairs (33, 34). In reactions with the A template, the RT concentrations were 10 and 15 nM for the wild-type and R72A RT, respectively. In reactions with the C template, the concentration of the wild-type RT was 15 nM and the concentration of the R72A enzyme was 30 nM.

RESULTS

Catalytic Efficiency on Homopolymeric Template-Primers—Kinetic parameters for polymerization by wild-type and R72A RT were determined with poly(rA)oligo(dT)20, poly(rC)oligo(dG)15, and poly(dC)oligo(dG)15 (Table I). The $K_{m, dNTP}$ of the mutant RT was higher than that of wild-type RT with the RNA templates, but not with the homopolymeric DNA template (i.e. poly(dC)). With all three substrates, the mutant RT had a lower $k_{cat}$ than wild-type. These differences result in catalytic efficiencies ($k_{cat}/K_m$) for the R72A mutant that are 1150- to 92-fold lower than for wild-type RT on poly(rA)oligo(dT)20, poly(rC)oligo(dG)15, and poly(dC)oligo(dG)15, respectively.

Reverse Transcriptase Fidelity during Gap-filling Syntheses—The fidelity of DNA-dependent DNA synthesis catalyzed by the R72A and wild-type RT was determined using assays that score RT errors as blue M13mp2 revertants of a colorless plaque phenotype. One assay detects eight different substitutions at a TGA codon (lacZ nucleotides 87–89), and the other detects one

---

| Template Primers |
|------------------|
| A (30-mer/20-mer): 5′-GCC GGA AGC ATA AAG TGT AT 3′-CCG CCT TCG TAT TTC ACA TAT CGG ACC CCA |
| B (45-mer/25-mer): 5′-GCC TCG CAG CCG TCC AAC CAA CTC A 3′-CCG AGC GTG GCC AGG TTT GTT GAG TTG GAG CTA GGT TAC GGC AGG |
| C (45-mer/25-mer): 5′-GTG CGG GCC TCT TCG CTA TTA GCC C 3′-CAC GCC CAG AGA AGC GAT AAG GCC GTC GAC GGT TCC CTC CCA ACA |
substitution errors at the opal codon in this particular sequence context. In marked contrast, the mutant is 20-fold more accurate for frameshift errors than wild-type. This indicates a significant difference in error specificity for the R72A and wild-type RT. To explore specificity in more detail, we employed a third fidelity assay that scores errors as light blue and colorless M13 plaques resulting from a variety of substitutions and frameshifts at over 100 template positions within the lacZ gene (19). With this assay, the R72A generated lacZ mutants at a frequency that was 1.6-fold higher than that of the wild-type RT (Table II). Despite this small change in mutant frequency, sequence analysis of lacZ mutants revealed a remarkable difference in the error specificity of R72A RT relative to wild-type (Fig. 2 and Table III).

As expected from earlier studies (17, 18), wild-type RT generates both substitution and frameshift errors at numerous template positions. The spectrum (Fig. 2A) is dominated by deletions in mononucleotide runs at template positions 70–73, 88–90, 91–94, 132–136, and 137–139 and by base substitutions at −36. Previous studies (18) had suggested that the deletions, and most of these base substitutions, result from template-primer slippage. In contrast, only two frameshifts are present in the R72A mutant spectrum (Fig. 2B), and there is only one substitution at site −36. When the results are used to calculate error rates per detectable nucleotide incorporated (Table III), the mutant RT is 25-fold more accurate than wild-type for one-nucleotide deletions and for T → C substitutions at position −36. Thus, R72A RT is a strong antimutator polymerase for misalignment-mediated errors in a variety of sequence contexts. Also absent from the R72A RT error spectrum are the majority of base substitutions at other template locations (Fig. 2). The base substitution fidelity of the R72A mutant, representing an average value for all 12 mismatches at a variety of locations, is 3.3-fold higher than that of wild-type RT (Table III).

| Enzyme | No. of determinations | Mutant frequency (× 10⁻⁴) | Ratio R72A/wild-type |
|--------|-----------------------|---------------------------|---------------------|
| Base substitution reversion assay | | | |
| Wild-type | 4 | 2.2 ± 0.91 | 37 |
| R72A | 4 | 82.0 ± 8.0 |
| One-nucleotide deletion reversion assay | | | |
| Wild-type | 88 | 32.0 ± 10.0 | 0.05 |
| R72A | 2 | 1.6 ± 0.08 |
| Forward mutation assay | | | |
| Wild-type | 10 | 210 ± 45 | 1.6 |
| R72A | 5 | 340 ± 38 |

| Type of mutation | Wild-type | R72A | R72A/wild-type |
|------------------|-----------|------|----------------|
| Frameshifts | 88 | 0.70 | 2 | 0.03 | 0.04 |
| Base substitutions | 74 | 0.93 | 0.15 | 0.30 | 0.30 |
| T → C at: | | | | | |
| −36 | 46 | 71 | 1 | 3.10 | 0.04 |
| 147 | 1 | 1.60 | 101 | 320 | 200 |
| 87 | 1 | 1.60 | 43 | 140 | 87 |

A second remarkable feature of the error specificity of R72A RT are the hot spots for misincorporation at positions 87 and 147 of the lacZ gene (Fig. 2B). Both positions are template thymines within the sequence 5′-CTGG. The R72A RT error rates for T → C transitions at these sites are 87- and 200-fold higher, respectively, than for wild-type RT (Table III). T → G transversions are also generated at position 147, at a rate that is at least 24-fold higher than that of wild-type RT.

### Analysis of Insertion Kinetics
In gap-filling fidelity assays, the recovery of lacZ mutants containing base substitutions requires both misinsertion and mismatch extension. To determine the contribution of selectivity during the insertion step to the observed changes in error specificity, we compared the kinetic parameters for insertion of correct dATP and incorrect dGMP (Table IV). This primarily reflects much higher $K_m$ values for the incoming nucleotides, with $k_{cat}$ values much less affected. Like wild-type RT, R72A inserts incorrect dGMP with substantially lower efficiencies than it inserts correct dAMP. For both enzymes, this is primarily a result of a ≥100-fold increase in $K_m$ for the incorrect 5′-GCTAT— or 5′-GCTAT—

| Enzyme | dNTP | $K_m$ (μM) | $k_{cat}/K_m$ (μM⁻¹ min⁻¹) |
|--------|-------|------------|----------------------------|
| Wild-type T-A | dATP | 0.07 ± 0.01 | 0.24 ± 0.01 |
| | dGTP | 8.50 ± 1.6 | 0.20 ± 0.01 |
| R72A T-A | dATP | 6.80 ± 1.1 | 0.38 ± 0.02 |
| | dGTP | 910 ± 95 | 0.10 ± 0.04 |

| Enzyme | dNTP | $K_m$ (μM) | $k_{cat}/K_m$ (μM⁻¹ min⁻¹) |
|--------|-------|------------|----------------------------|
| R72A T-G | dATP | 240 ± 53 | 0.21 ± 0.01 |
| | dGTP | 1400 ± 445 | 0.03 ± 0.005 |

| Enzyme | T-P terminus | $K_m$ (μM) | $k_{cat}/K_m$ (μM⁻¹ min⁻¹) |
|--------|-------------|------------|----------------------------|
| Wild-type | dATP | 0.20 ± 0.05 | 0.16 ± 0.01 |
| | dGTP | 140 ± 24 | 0.01 ± 0.0005 |
| R72A | dATP | 240 ± 53 | 0.21 ± 0.01 |
| | dGTP | 1400 ± 445 | 0.03 ± 0.005 |

| Enzyme | T-P terminus | $K_m$ (μM) | $k_{cat}/K_m$ (μM⁻¹ min⁻¹) |
|--------|-------------|------------|----------------------------|
| Wild-type | dATP | 0.05 ± 0.03 | 0.07 ± 0.006 |
| | dGTP | 0.27 ± 0.99 | 0.90 ± 0.045 |
| R72A | dATP | 0.12 ± 0.04 | 0.35 ± 0.028 |
| | dGTP | 170 ± 69 | 1.50 ± 0.27 |
Fidelity of R72A HIV-1 Reverse Transcriptase

The DNA template strand is gray, and the primer strand is yellow; the incoming dTTP is purple. B, the binding pocket for the nascent base pair in HIV-1 RT. The backbone of the DNA template strand is red, and the templating base is yellow; the primer strand is light purple, and the incoming dTTP is orange. The backbone of the β3-β4 loop is shown in blue, and the Arg72 side chain is in red. The surface of the DNA and the protein that form the binding pocket for the nascent base pair is gray, and the parts of the surface contributed by Lys65, Arg72, Leu74, and Gln151 are colored cyan, red, green, and magenta, respectively. C, interactions of Arg72 with the incoming deoxynucleotide triphosphate. A fragment of the backbone of the β3-β4 loop is in blue with the side chains of Leu74 (green) and Arg72 (red). The template nucleotides are white. The incoming nucleotide (orange) is sandwiched between the primer terminal nucleotide (yellow) and the Arg72 side chain (red). The distances from the NH2 of Arg72 and the O2 and O3 of the α-phosphate are indicated. Images were generated based on the coordinates of the crystal structure of HIV-1 RT in ternary complex (1RTD, Protein Data Bank) determined by Ref. 6. The image in panel A was made with MOLSCRIPT (52) and Raster3D (53), in panel B with GRASP (54), and in panel C with Insight® II version 97.0.

The above kinetic analysis with T-Ps A and B suggest that the higher fidelity of the R72A RT is partly due to increased discrimination against misinsertion of dGMP opposite T at these two template positions.

Next, we examine insertion kinetics at the hottest site of misincorporation by R72A RT (position 147, T-P C). Kinetic parameters for correct dAMP insertion by wild-type RT were similar to those observed with the other sequences (Table IV). However, the efficiency of dGMP misinsertion by wild-type RT is significantly lower at the hot spot than at the two other sites. At position 147, the wild-type RT has a 700-fold higher $K_m$ and a 16-fold lower $k_{cat}$ for misinsertion of dGMP than for correct dAMP insertion, resulting in a 10,000-fold lower efficiency for dGMP misinsertion than for correct insertion (Table IV). Consequently, the dGMP misinsertion frequency at the 147 site by wild-type RT is 70- and 170-fold lower than at thymidines in templates A and B, respectively (Fig. 3).

Insertion of correct dAMP at position 147 by the R72A RT is characterized by a remarkably high $K_m$ of 240 μM. This value is 1200-fold higher than for wild-type RT at this same site, and it is approximately 30-fold higher than those observed for R72A RT with the non-hot spot T-Ps A and B. In contrast to the large enzyme- and DNA-dependent differences in $K_m$ for correct nucleotide, R72A RT has $k_{cat}$ values for correct dAMP insertion and $K_m$ and $k_{cat}$ values for incorrect insertion of dGMP at position 147 that are similar to those observed with the other T-Ps (Table IV). As a consequence, the R72A RT has a misinsertion frequency at position 147 that is 220-fold higher than that of wild-type RT (Fig. 3). This parallels the 200-fold increased error rate for stable misincorporation of dGMP opposite T at these two sites.

The efficiency of dGMP misinsertion by wild-type RT is significantly lower at the hot spot than at the two other sites. At position 147, the wild-type RT has a 700-fold higher $K_m$ and a 16-fold lower $k_{cat}$ for misinsertion of dGMP than for correct dAMP insertion, resulting in a 10,000-fold lower efficiency for dGMP misinsertion than for correct insertion (Table IV). Consequently, the dGMP misinsertion frequency at the 147 site by wild-type RT is 70- and 170-fold lower than at thymidines in templates A and B, respectively (Fig. 3).

Analysis of Extension Kinetics by Wild-type and R72A RT—The above kinetic analysis with T-Ps A and B suggest that the higher fidelity of the R72A RT is partly due to increased discrimination during insertion. To test the hypothesis that higher fidelity might also result from increased discrimination against extension of mismatched primer termini, we compared the ability of wild-type and R72A RT to extend a template-primer with a terminal T-G mispair. With T-P A, the wild-type enzyme extends the correctly paired and mispaired T-P termini with similar efficiency (Table V). This is consistent with the
well known promiscuity of wild-type RT for mismatch extension (13, 14) and the fact that the T-G mismatch is among the most readily extended of the 12 possible mispairs (e.g., see Refs. 9 and 33–35). The R72A RT extends the correctly paired template with an efficiency similar to that of the wild-type RT. However, extension of the T-G mismatch by R72A RT is 330-

![Fig. 2](image)

**Spectra of single-base mutations produced in the forward assay.** A, spectrum of mutations generated by the wild-type HIV-1 RT. B, spectrum of mutations generated by R72A RT. Three lines of the primary DNA sequence for the lacZ a-complementation gene in M13mp2 are shown. The sequence is that of the viral (†)-template strand. Position +1 is the first transcribed base. The base substitutions shown above the lines of the wild-type DNA sequence indicate the new base found in the viral DNA. Frameshifts are shown below the lines of the sequence; a loss of a base is indicated by Δ, and an addition of a base is indicated by s. For frameshifts at iterated nucleotide positions, it is not possible to distinguish which base was lost or added; therefore, the symbol is centered under the run.
fold less efficient than extension of the correctly paired terminus, due to a 1400-fold increase in the $K_m$ for the incoming correct dGTP (170 $\mu M$, Table V). With template-primer A, R72A RT is 380-fold less efficient in extending the mismatch than is wild-type RT (Table V). A similar difference (330-fold) in efficiency is also observed between R72A and wild-type RT for extension of the T-G mispair at the position 147 hot spot, again due to a high $K_m$ for the incoming correct dGTP (320 $\mu M$, Table V). The strongly reduced ability to extend a T-dGMP mismatch in two different sequence contexts is consistent with the generally higher fidelity of R72A RT at most locations.

**DISCUSSION**

Structural studies (Ref. 6 and references therein) suggest that the binding pocket for the nascent base pair in HIV-1 RT assembles through movement of the fingers subdomain to bring into position amino acid side chains that interact with the primer-terminal base pair, the templating base and the incoming dNTP. One of these is Arg$^{\text{72}}$ in the $\beta3$-$\beta4$ loop of the fingers, a residue that is highly conserved in the RT family of DNA polymerases. The importance of Arg$^{\text{72}}$ to catalysis is clearly demonstrated by the lower catalytic efficiencies of the R72A mutant with three different homopolymeric template-primers (Table I; also see Ref. 25). This defect is consistent with the location of Arg$^{\text{72}}$, which stacks with the base of the incoming nucleotide, donates hydrogen bonds to the $\alpha$-phosphate, and is contacted by Leu$^{\text{74}}$, whose side chain interacts with the templating base (Fig. 1C). Removal of these interactions by replacing the arginine side chain with the methyl group of alanine may impair catalysis by affecting any of several steps in the reaction cycle. The higher apparent $K_m$ for incorporation with poly(rA) and poly(rC) (Table I) and for correct insertion of a single nucleotide (Table IV) suggests that dNTP binding affinity is reduced, consistent with loss of H-bonding or stacking with the base. In addition, Sarafianos et al. (25) reported that R72A RT has reduced processivity, is defective in pyrophosphorolysis, and is resistant to phosphonofomate. These latter two defects are consistent with loss of H-bonding to the $\alpha$-phosphate (Fig. 1C). Substitution of alanine for Arg$^{\text{72}}$ reduced catalytic efficiency with poly(rA) and poly(rC) templates much more strongly than with a poly(dC) template (Table I). This primarily reflects an elevation of the $K_m$ for incorporation with ribo-homopolymers that is not observed with poly(dC). This is consistent with the possibility that copying RNA templates may be generally more problematic than copying DNA templates. In support of this, R72A RT effectively copied several hundred nucleotides of heteropolymeric DNA template used for the DNA-templated M13mp2 fidelity assays, but an initial attempt to measure fidelity with an RNA template, using the method described by Boyer et al. (36), was unsuccessful because R72A RT was unable to completely copy RNA of similar length and identical sequence (data not shown).

R72A RT has strongly altered fidelity compared with wild-type HIV-1 RT. This is consistent with the hypothesis that the nucleotide selectivity of DNA polymerases results partly from selection for correct Watson-Crick base pairing geometry (reviewed in Ref. 9). In support of the geometric selection hypothesis, the nascent base pair binding pockets of HIV-1 RT (6) and other DNA polymerases are indeed shaped to tightly accommodate correct base pairs (2–5, 7). Moreover, substituting amino acids for residues that form part of the binding pocket has been shown to alter the fidelity of DNA polymerase $\beta$ (37–39), the Klenow fragment of E. coli DNA polymerase I (40–42), and HIV-1 RT (43–47). These reports and the present study showing that changing Arg$^{\text{72}}$ side-chain interactions with the incoming dNTP and/or with surrounding side chains (e.g., Leu$^{\text{74}}$, Lys$^{\text{65}}$, Gln$^{\text{151}}$, Fig. 1B) alters fidelity illustrate the importance to fidelity of interactions between the substrates and amino acid side chains that form the binding pocket.

The observation that the R72A mutant has higher fidelity than wild-type RT for both base substitutions and single-nucleotide additions and deletions (Tables II and III) indicates that normally error production depends on Arg$^{\text{72}}$ interactions. The lower error rate for base substitutions due to direct miscoding partly reflects higher discrimination by R72A RT against direct misinsertion, exemplified by reduced insertion of dGMP opposite T (Fig. 3, 4-fold with substrate A and 92-fold with substrate B). This suggests that Arg$^{\text{72}}$ interactions stabilize incorrect incoming dNTPs. This might occur by stacking of the guanidinium group of Arg$^{\text{72}}$ with the base of an incorrect dNTP and/or through interactions between Arg$^{\text{72}}$ and Leu$^{\text{74}}$ and/or Gln$^{\text{151}}$ that affect the stability of the templating base. The $\alpha$-bonds to the $\alpha$-phosphate observed in the ternary complex likely contribute to initial binding of incoming dNTPs (correct and incorrect) prior to closure of the fingers, but may also influence the stability of incorrect nucleotides in the binding pocket and/or influence chemistry.

The lower error rate for base substitutions due to direct miscoding by R72A RT also reflects a lower efficiency of correct dNTP incorporation onto mismatched primer termini. As illustrated by results in Table V, the lower efficiency of T-G mismatch extension results from a several hundred-fold higher $K_m$ for incorporation of the correct dGTP by R72A RT compared with wild-type. Strongly reduced mismatch extension efficiency would be anticipated based on the fact that the binding pocket, wherein the incoming dNTP is tightly sandwiched between Arg$^{\text{72}}$ and the primer terminal base (Fig. 1C), would be distorted when the terminus is mismatched and the arginine side chain is replaced with alanine.

Considerable evidence suggests that the $T \rightarrow C$ substitutions at template position 36 and the single-nucleotide frameshifts in short homopolymeric sequences generated by wild-type HIV-1 RT (Table III, Fig. 2) involve misaligned template-primers (18, 20). Their virtual absence from the R72A error spectrum (Fig. 2B) represents a 25-fold increase in fidelity, the strongest antimaturation effect reported for any HIV-1 RT derivative. A possible explanation for higher fidelity of R72A RT for strand-slipage errors is a reduced ability to extend template-primers containing unpaired nucleotides in the duplex template-primer stem just upstream of the active site. This and alternative explanations related to DNA binding affinity and processivity are under investigation.

Unanticipated in the present study were the base substitution hot spots at template positions 87 and 147 (Fig. 2B). Among 33 possible template T residues in the lacZ target sequence at which base substitution errors can be detected (19),
these are the only two\(^2\) that share the same flanking nucleotides, a 5'-C and two 3'-G nucleotides. They are substitution hot spots despite the fact that R72A RT has strongly reduced mismatch extension efficiency at position 147 (Table V). However, we intentionally performed gap-filling DNA synthesis reactions using a high dNTP concentration in order to increase the probability that any misinsertions by RT would be extended. In fact, the 200-fold higher error rate for T → C substitution at position 147 by R72A RT (Table III) matches a similar increase in the rate of misinsertion of dGMP opposite substitution at position 147 by R72A RT relative to wild-type RT (Fig. 3, template C). This strongly indicates that the hot spot for stable misincorporation observed in the M13 fidelity assay results from reduced discrimination at the insertion step. The result is remarkable in that it is opposite to the higher misinsertion fidelity of R72A for the same error but in two different sequence contexts (Fig. 3). Loss of discrimination is a consequence of a 1200-fold increase in the \(K_m\) for the correct dATP for R72A as opposed to the higher misinsertion from reduced discrimination at the insertion step. The result is 5'-CTGG. Although more work will be required to understand this unique sequence specificity, we do note that among the possible dinucleotide sequences, 5'-TG/CA is one of the most flexible base steps both in naked B-DNA and in DNA-protein complexes; in some sequence contexts it may assume an unusually high twist (48–51). Experiments are currently under way to define the contributions of individual flanking nucleotides to the strong sequence dependence of misincorporation by R72A RT.

Acknowledgments— We thank Dr. Thomas A. Darden for helpful discussions during the course of this work and Drs. William C. Copeland and Thomas A. Darden for their critical comments on the manuscript.

REFERENCES

1. Loeb, L. A., and Kunkel, T. A. (1982) Annu. Rev. Biochem. 51, 429–457
2. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994) Science 264, 1891–1903
3. Pelletier, H., Sawaya, M. R., Welfle, W., Wilson, S. H., and Kraut, J. (1996) Biochemistry 35, 12742–12761
4. Doublé, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) Nature 391, 251–258
5. Kieder, J. R., Mao, C., Braman, J. C., and Beece, L. S. (1998) Nature 391, 304–307
6. Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998) Science 282, 1669–1675
7. Li, Y., Korovets, S., and Waksman, G. (1998) EMBO J. 17, 7514–7525
8. Kunkel, T. A., and Wilson, S. H. (1998) Nat. Struct. Biol. 5, 95–99
9. Echols, H., and Goodman, M. F. (1991) Annu. Rev. Biochem. 60, 477–511
10. Petrussa, J., Sowers, L. C., and Goodman, M. F. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 1559–1562
11. Kool, E. T. (1998) Biopolymers 48, 3–17
12. Roberts, J. D., Bebenek, K., and Kunkel, T. A. (1988) Science 242, 1171–1173
13. Herrick, F. W., Preston, B. D., Sandell, L. L., and Loeb, L. A. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 8343–8347
14. Yu, H., and Goodman, M. F. (1992) J. Biol. Chem. 267, 10888–10896
15. Temin, H. M. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 6900–6903

\(^2\) A T in the same 5'-CTGG sequence context is at position 61 in the lacZ α target sequence. However, a T → C transition at this site results in a plaque phenotype very similar to that of the wild-type M13mp2. Thus, mutants with a T → C substitution at site 61 are difficult to ascertain and might have escaped detection.