

# SCIENTIFIC REPORTS



OPEN

## Transcriptional inaccuracy threshold attenuates differences in RNA-dependent DNA synthesis fidelity between retroviral reverse transcriptases

Alba Sebastián-Martín<sup>1</sup>, Verónica Barrioluengo<sup>1,2</sup> & Luis Menéndez-Arias<sup>1</sup>

In M13mp2 *lacZ* $\alpha$  forward mutation assays measuring intrinsic fidelity of DNA-dependent DNA synthesis, wild-type human immunodeficiency virus type 1 (HIV-1) RTs of group M/subtype B previously showed >10-fold higher error rates than murine leukaemia virus (MLV) and avian myeloblastosis virus (AMV) RTs. An adapted version of the assay was used to obtain error rates of RNA-dependent DNA synthesis for several RTs, including wild-type HIV-1<sub>BH10r</sub>, HIV-1<sub>ESP49r</sub>, AMV and MLV RTs, and the high-fidelity mutants of HIV-1<sub>ESP49</sub> RT K65R and K65R/V75I. Our results showed that there were less than two-fold differences in fidelity between the studied RTs with error rates ranging within  $2.5 \times 10^{-5}$  and  $3.5 \times 10^{-5}$ . These results were consistent with the existence of a transcriptional inaccuracy threshold, generated by the RNA polymerase while synthesizing the RNA template used in the assay. A modest but consistent reduction of the inaccuracy threshold was achieved by lowering the pH and Mg<sup>2+</sup> concentration of the transcription reaction. Despite assay limitations, we conclude that HIV-1<sub>BH10</sub> and HIV-1<sub>ESP49</sub> RTs are less accurate when copying DNA templates than RNA templates. Analysis of the RNA-dependent mutational spectra revealed a higher tendency to introduce large deletions at the initiation of reverse transcription by all HIV-1 RTs except the double-mutant K65R/V75I.

In retrovirus, the reverse transcriptase (RT) is the DNA polymerase responsible for the replication of the viral genome. Retroviral RTs use the (+) single-stranded RNA genome to synthesize a complementary minus-strand DNA, while the RNA template is being degraded by the RNase H activity of the RT. The newly synthesized complementary DNA (cDNA) is then used as template for the synthesis of plus-strand DNA to obtain a double-stranded proviral DNA<sup>1,2</sup>.

RTs are widely used in biotechnology for their ability to synthesize DNA using RNA templates. Advances introduced in the late 1990s such as cDNA microarrays and next-generation sequencing technologies have opened new possibilities for the identification of all RNA molecules in one cell or a population of cells, and the analysis of their expression levels<sup>3-5</sup>. RTs play a fundamental role behind these developments, and wild-type (WT) and engineered RT variants of avian myeloblastosis virus (AMV), murine leukaemia virus (MLV), human immunodeficiency virus type 1 (HIV-1) and *Geobacillus stearothermophilus* group II introns have been developed into more efficient tools to study gene expression by increasing catalytic efficiency, processivity, thermostability or fidelity of DNA synthesis<sup>6-9</sup>.

Unlike eukaryotic replicative polymerases, RTs lack exonuclease activity and are error-prone. Improvements in their intrinsic fidelity of DNA synthesis may have a positive impact on the reliability of whole transcriptome shotgun sequencing (i.e. RNA-seq) data<sup>10</sup>. Despite the large amount of research on the fidelity of retroviral RTs, most of the available studies have been devoted to the analysis of DNA-dependent DNA synthesis accuracy<sup>11,12</sup>. These studies have shown that oncoretroviral RTs (e.g. AMV and MLV RTs) are more faithful than lentiviral RTs

<sup>1</sup>Centro de Biología Molecular "Severo Ochoa" (Consejo Superior de Investigaciones Científicas & Universidad Autónoma de Madrid), c/Nicolás Cabrera, 1, Campus de Cantoblanco, 28049, Madrid, Spain. <sup>2</sup>Present address: DiaSorin Iberia S.A., Avenida de la Vega 1, 28108, Alcobendas (Madrid), Spain. Correspondence and requests for materials should be addressed to L.M.-A. (email: [lmendez@cbm.csic.es](mailto:lmendez@cbm.csic.es))

such as the HIV-1 RT. Thus, the intrinsic error rates of HIV-1 RT determined with the M13mp2 *lacZ* $\alpha$  forward mutation assay<sup>13</sup> were more than 10-fold higher than those obtained with AMV and MLV RTs<sup>14–16</sup>. However, in HIV-1 RTs, antiretroviral drug resistance-associated mutations such as K65R or the combination of K65R and V75I were shown to increase fidelity of DNA-dependent DNA synthesis to levels similar to those obtained with oncoretroviral RTs<sup>17,18</sup>.

The fidelity of RNA-dependent DNA synthesis of retroviral RTs has been evaluated in enzymatic assays by comparing kinetic parameters for the incorporation of correct and incorrect nucleotides, and the extension of matched and mismatched template-primers. Assays carried out under steady-state conditions with AMV, MLV and HIV-1 RTs did not reveal large differences when DNA templates were substituted by RNA templates<sup>19,20</sup>. However, using pre-steady-state kinetics, Kerr and Anderson<sup>21</sup> showed that misinsertion fidelity was 9–64 times (with duplex 45/25mer) and 14–23 times (with duplex 45/22mer) higher in RNA-templated than in DNA-templated reactions catalyzed by HIV-1 RT. These experiments were performed with synthetic duplexes bearing the same nucleotide sequences (except for having U instead of T in the RNA templates). Despite providing important mechanistic information, nucleotide incorporation assays are restricted to a few template-primers and therefore provide limited information on the propensity of RTs to introduce base substitutions.

In contrast, forward mutation assays based on the expression of target genes using phages such as M13 or  $\Phi$ X174 give error rate estimates over a wide range of mutational sites. Attempts to compare the fidelity of retroviral RTs in DNA polymerization reactions carried out with RNA or DNA templates using these methods did not provide consistent results. Studies performed with HIV-1 RT and the amber16 reversion assay using phage  $\Phi$ X174 showed that two out of the seven specific mismatches analysed had 20- and 7-fold lower mutation frequencies with DNA templates than with RNA templates<sup>22</sup>. On the other hand, M13-based assays using *lacZ* $\alpha$  or the HIV-1 *env* hypervariable region 1 (V1) as target genes showed little differences in fidelity between reactions carried out with HIV-1 RT using RNA or DNA templates. Thus, base substitution error rates determined with *lacZ* $\alpha$  were  $1.7 \times 10^{-4}$  and  $1.4 \times 10^{-4}$  for DNA-dependent and RNA-dependent reactions, respectively<sup>23</sup>. These small differences were also found for overall error rates in assays using a fragment of *env* as a target sequence (i.e.  $1.9 \times 10^{-4}$  with DNA templates and  $2.0 \times 10^{-4}$  with RNA templates)<sup>24</sup>. In contrast, and using a modified version of the M13mp2 *lacZ* $\alpha$  forward mutation assay, Boyer *et al.*<sup>25</sup> showed that the overall fidelity of HIV-1 RT was about two to six-fold higher while copying RNA templates than DNA templates. In these experiments, authors suggested the contribution of errors made by the T7 RNA polymerase (RNAP) while preparing the RNA used as template in the fidelity assays, acknowledging an undetermined impact on the calculated error rates.

The discrepancies in the fidelity assessments could be explained in part by methodological differences in the assays, partly due to the different WT HIV-1 group M/subtype B RTs used in these experiments [i.e., homodimers (p66/p66) *versus* heterodimers (p66/p51), or RT variants derived from different viral strains, such as NL4-3, HXB2, BH10 or NY5]. In the present study, we have determined the intrinsic fidelity of RNA-dependent DNA synthesis of several RTs that showed diverse error rates in DNA-dependent DNA polymerization assays, ranging from around  $1.4 \times 10^{-4}$  in the case of HIV-1<sub>BH10</sub> RT<sup>26–29</sup> to  $1.2 \times 10^{-5}$  for MLV RT or  $\sim 6.3 \times 10^{-6}$  for HIV-1 group O (ESP49 strain) RT mutants K65R (O\_K65R) and K65R/V75I (O\_K65R/V75I)<sup>18</sup>. We demonstrate how errors made by the T7 RNAP attenuate those differences while hampering an accurate determination of fidelity using RNA templates, despite improving the quality of the RNA template by using transcription conditions that could increase the accuracy of the RNA synthesis reaction. Despite those limitations, our results show that WT HIV-1<sub>BH10</sub> and HIV-1<sub>ESP49</sub> are less accurate when copying DNA than RNA templates.

## Results

**Fidelity of RNA-dependent DNA synthesis of retroviral RTs.** M13mp2 *lacZ* $\alpha$  forward mutation assays provide a broad estimate of the fidelity of RTs, based on a relatively large number of mutational sites and sequence contexts, although silent mutations cannot be detected by using this method. We determined mutant frequencies and error rates of RNA-dependent DNA synthesis for WT MLV, AMV, HIV-1<sub>BH10</sub> and HIV-1<sub>ESP49</sub> RTs, as well as mutant RTs O\_K65R and O\_K65R/V75I. For this purpose, a commercial T7 RNA polymerase (RNAP) was used to synthesize a *lacZ* $\alpha$  RNA template that was then reverse transcribed by retroviral RTs (the method is outlined in the Supplementary Fig. S1). The cDNA product was hybridised with a gapped M13mp2 DNA lacking the *lacZ* $\alpha$  gene in one of the two strands of the molecule. Errors made by RTs during reverse transcription result in a decrease in  $\alpha$ -complementation and could be detected by the altered colour phenotype of the mutant plaques (pale blue or colourless) when phages are grown on an appropriated indicator strain. Mutant frequencies obtained with the six studied RTs, and calculated as the ratio of mutant to total plaques, are given in Table 1. These assays showed less than two-fold differences in fidelity between the most accurate and the least faithful RTs.

The mutational specificity of the studied RTs was determined after sequencing the *lacZ* $\alpha$  mutants generated with the M13mp2-based mutation assays (Supplementary Figs S2–S7). All mutational spectra had insertions of one thymidine at the homopolymeric sequence located at positions +137/+139. In addition, U-to-C transitions at positions –36/–35 and U-to-A transversions at position +73 were observed in the spectra of five RTs. On the other hand, we also found large deletions in all mutational spectra, except in the one generated with the AMV RT. In the case of HIV-1<sub>BH10</sub>, HIV-1<sub>ESP49</sub> and the O\_K65R RTs, the large deletions clustered at positions +172/+173 (Supplementary Figs S2–S4), corresponding to the first two nucleotides incorporated during reverse transcription. These deletions could derive from misalignment errors occurring while extending the DNA primer.

Despite the similarities found among the mutational spectra, there were also remarkable differences. Thus, for example, AMV RT generated one-nucleotide insertions of A in the homopolymeric region located at positions +91/+94 (Supplementary Fig. S7), whereas MLV RT showed one hotspot at position +144, including various deletions and transversions (A-to-C and A-to-T) (Supplementary Fig. S6). In our analyses, hotspots were defined as those positions where at least four mutations were found. The statistical differences in hotspot distribution between different RTs were determined by using a two-tailed Fisher's exact test, and are shown in

| RTs                    | Mutant plaques | Total plaques | Mutant frequency <sup>a</sup> |
|------------------------|----------------|---------------|-------------------------------|
| HIV-1 <sub>BH10</sub>  | 52             | 12,836        | 0.00405                       |
| HIV-1 <sub>ESP49</sub> | 46             | 13,347        | 0.00345 (1.18)                |
| O_K65R                 | 54             | 18,284        | 0.00295 (1.37)                |
| O_K65R/V75I            | 69             | 23,512        | 0.00293 (1.38)                |
| MLV                    | 36             | 13,569        | 0.00265 (1.53)                |
| AMV                    | 28             | 11,250        | 0.00249 (1.63)                |

**Table 1.** RNA-dependent DNA synthesis fidelity of WT and mutant RTs in M13mp2 *lacZ* $\alpha$  forward mutation assays. For each enzyme, mutant plaques were obtained after transfection of gapped DNA hybridised with the cDNA product of ten synthesis reactions. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNAP (Promega) in a transcription buffer containing 40 mM Tris-HCl pH 7.9 and 6 mM MgCl<sub>2</sub> (full composition given in Materials and Methods). <sup>a</sup>Background frequencies in these assays were estimated to be less than one in 20,000 plaques<sup>18</sup>. Numbers between parentheses indicate the fold-increase in fidelity relative to the WT HIV-1<sub>BH10</sub> RT.

the Supplementary Table S1. Interestingly, the spectrum of the O\_K65R/V75I RT showed an important hotspot at position +147 (U-to-C substitutions) (Supplementary Fig. S5). This hotspot is absent in the mutational spectra of the other analysed RTs ( $P < 0.005$ , for all five comparisons) (Supplementary Table S1). On the other hand, mutations induced by HIV-1<sub>ESP49</sub> RT seemed to be scattered throughout the whole *lacZ* $\alpha$  sequence (Supplementary Fig. S3). The spectrum induced by the single-mutant O\_K65R had one hotspot at position +109 (Supplementary Fig. S4), while in the one obtained with HIV-1<sub>BH10</sub> RT, G-to-T changes accumulated at position +149 (Supplementary Fig. S2).

Error rates for all RTs are summarized in Table 2. The highest error rates were obtained with WT HIV-1<sub>BH10</sub> RT, while one of the most faithful enzymes was the double-mutant O\_K65R/V75I that showed 1.4-fold increased accuracy relative to the HIV-1<sub>BH10</sub> RT. Mutant O\_K65R/V75I RT had a low tendency to introduce frameshifts, although it had a relatively high base substitution error rate. This enzyme was prone to generate transitions (81%), which accumulated at two major hotspots (at positions -36 and +147) and three minor hotspots (located at nucleotides -35, -7 and +87). Unlike in the case of O\_K65R/V75I, this strong bias towards the generation of transitions was not observed with the other RTs studied. On the other hand, compared with the other RTs, the HIV-1<sub>BH10</sub> polymerase was prone to introduce frameshift errors, which were predominantly deletions.

By using the forward mutation assay we cannot determine whether the errors were made by the T7 RNAP during the synthesis of the RNA template (transcription) or by RTs during cDNA synthesis (reverse transcription). The overall error rates obtained with the RTs used in this study were in the range of  $2.5 \times 10^{-5}$  to  $3.5 \times 10^{-5}$ . These values are relatively close to reported estimates of transcription error rates obtained in different organisms<sup>30-33</sup>, and suggest the existence of an inaccuracy threshold imposed by the T7 RNAP while synthesizing the RNA template.

**Fidelity of promoter-dependent transcription by T7 RNA polymerase.** In order to reduce the possible inaccuracy threshold imposed by the T7 RNAP in M13mp2-based assays, we analysed its fidelity under different reaction conditions. A linearized pTRI- $\beta$ -actin-Mouse plasmid containing a T7 promoter sequence was used as template for *in vitro* transcription using T7 RNAP (Ambion) in the presence of [ $\alpha$ -<sup>32</sup>P]CTP. Reactions were carried out at different pH and Mg<sup>2+</sup> concentrations, either in the presence of all four NTPs or with biased NTP pools (i.e. lacking or with a very low concentration of one ribonucleotide) (Supplementary Fig. S8). In this assay, when all NTPs were present (lanes marked with an asterisk), the polymerase synthesized an RNA of 53 nucleotides. When using biased pools, the amount of full-length RNA was reduced, since the insertion of incorrect nucleotides would be needed to complete the synthesis reaction.

Relative amounts of full-length products in reactions carried out with biased NTP mixtures and in the presence of all NTPs provided a rough estimate of the accuracy of T7 RNAP in different assay conditions. It should be noted that G is required to initiate transcription, and therefore T7 RNAP cannot synthesize RNA in the absence of this nucleotide (lanes 3 in Supplementary Fig. S8). Inefficient RNA synthesis was also observed in the absence of A or U (lanes 1 and 6). When ATP was supplied at 1  $\mu$ M while maintaining the three other NTPs at 100  $\mu$ M (lanes 2), a small but significant amount of full-length RNA was observed at higher magnesium concentrations, as well as above pH 7.0. Similar findings were obtained in reactions with NTP mixtures having low concentrations of GTP (10–20  $\mu$ M) (lanes 4 and 5). The amount of full-length products in biased reactions was greater at 6 mM MgCl<sub>2</sub> and at pH 7.5, suggesting a lower accuracy in those conditions. At pH 5.5–6.0 we did not detect any transcription products. In addition, the efficiency of the reaction was very low in the presence of 0.5 mM MgCl<sub>2</sub>. Taken together, our data indicate that T7 RNAP showed good activity and improved accuracy at pH 6.5–7.0, and in the presence of Mg<sup>2+</sup> at 1–3 mM.

**Nucleotide incorporation kinetics and template-primer binding affinity of T7 RNA polymerase.** RNA templates used in forward mutation assays were synthesized under optimal conditions for transcription by T7 RNAP<sup>34</sup>. These reactions were carried out in 40 mM Tris-HCl buffer pH 7.9, containing 6 mM MgCl<sub>2</sub>. However, promoter-dependent transcription assays revealed qualitative differences in fidelity when RNA was synthesized at lower pH and in the presence of reduced amounts of magnesium. These observations indicated

| Mutation type        | HIV-1 <sub>BH10</sub> RT |            | AMV RT        |            | MLV RT        |            | HIV-1 <sub>ESP49</sub> RT |            | O_K65R RT     |            | O_K65R/V75I RT |            |
|----------------------|--------------------------|------------|---------------|------------|---------------|------------|---------------------------|------------|---------------|------------|----------------|------------|
|                      | No. of errors            | Error rate | No. of errors | Error rate | No. of errors | Error rate | No. of errors             | Error rate | No. of errors | Error rate | No. of errors  | Error rate |
| All classes          | 54                       | 1/28524    | 43            | 1/31395    | 46            | 1/35397    | 44 <sup>a</sup>           | 1/36401    | 58            | 1/37829    | 71             | 1/39739    |
| Base substitutions   | 26                       | 1/36434    | 25            | 1/33210    | 29            | 1/34531    | 19                        | 1/51843    | 29            | 1/46530    | 53             | 1/32739    |
| Transitions          | 11 (42%)                 |            | 10 (40%)      |            | 12 (41%)      |            | 10 (53%)                  |            | 19 (66%)      |            | 43 (81%)       |            |
| Transversions        | 15 (58%)                 |            | 15 (60%)      |            | 17 (59%)      |            | 9 (47%)                   |            | 10 (34%)      |            | 10 (19%)       |            |
| Frameshifts          | 28                       | 1/48960    | 18            | 1/66750    | 17            | 1/85245    | 25                        | 1/57018    | 29            | 1/67336    | 18             | 1/139505   |
| Insertions           | 7 (25%)                  |            | 14 (78%)      |            | 9 (53%)       |            | 13 (52%)                  |            | 10 (34%)      |            | 9 (50%)        |            |
| Deletions            | 21 (75%)                 |            | 4 (22%)       |            | 8 (47%)       |            | 12 (48%)                  |            | 19 (66%)      |            | 9 (50%)        |            |
| At runs <sup>b</sup> | 10 (36%)                 | 1/29266    | 10 (56%)      | 1/25650    | 7 (41%)       | 1/44196    | 12 (48%)                  | 1/25359    | 13 (45%)      | 1/32067    | 12 (67%)       | 1/44673    |
| At non-runs          | 18 (64%)                 | 1/59901    | 8 (44%)       | 1/118125   | 10 (59%)      | 1/113980   | 13 (52%)                  | 1/86242    | 16 (55%)      | 1/95991    | 6 (33%)        | 1/329168   |

**Table 2.** Summary of error rates for WT and mutant RTs, for various classes of mutations in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNAP (Promega) in a transcription buffer containing 40 mM Tris-HCl pH 7.9 and 6 mM MgCl<sub>2</sub>. <sup>a</sup>The total number of errors with this RT was smaller than the number of mutant plaques (Table 1), because five of those mutant plaques were either recombinants or had complex arrays of mutations, and were not included in the analysis of error rates. <sup>b</sup>A run is considered when there is a row of three or more identical nucleotides.

that transcriptions could be more accurate when carried out in buffers containing 40 mM Bis-Tris pH 6.75 and 1.5 mM MgCl<sub>2</sub>.

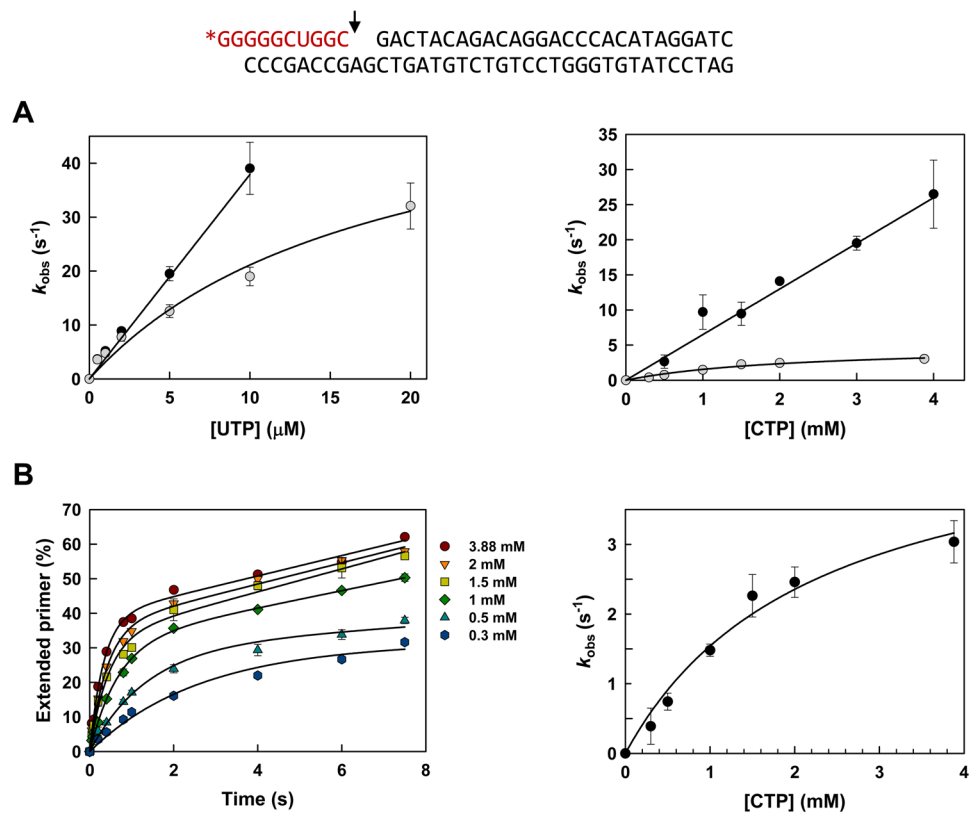
The efficiency and nucleotide specificity of the T7 RNAP was determined in single-nucleotide incorporation assays using a VSR10 hybrid containing a dsDNA and a 10-nucleotide RNA primer (Fig. 1). Due to the requirement of relatively large amounts of enzyme, these assays were carried out with purified recombinant T7 RNAP. Nucleotide incorporation rates were determined at different concentrations of correct (UTP) or incorrect ribonucleotides (ATP, CTP and GTP) in standard buffer conditions (40 mM Tris-HCl pH 7.9, containing 6 mM MgCl<sub>2</sub>), as well as in conditions enhancing T7 RNAP fidelity (40 mM Bis-Tris buffer pH 6.75 and 1.5 mM MgCl<sub>2</sub>). Kinetic constants for UTP incorporation could not be obtained due to the high efficiency of the polymerization reaction. Thus, at nucleotide concentrations above 300 μM, the T7 RNAP showed UTP incorporation rates ( $k_{obs}$ ) above 150 s<sup>-1</sup> in both assay conditions. At low concentrations of UTP, the polymerase showed different kinetic behaviours at pH 7.9/6 mM Mg<sup>2+</sup> and pH 6.75/1.5 mM Mg<sup>2+</sup> (Fig. 1A). The relationship between  $k_{obs}$  and [UTP] was linear at pH 7.9/6 mM Mg<sup>2+</sup> but hyperbolic at pH 6.75/1.5 mM Mg<sup>2+</sup>. Similar kinetics were observed in CTP incorporation reactions, although nucleotide incorporation efficiencies were much lower at pH 6.75/1.5 mM Mg<sup>2+</sup> than in reactions carried out in 40 mM Tris-HCl buffer pH 7.9, containing 6 mM MgCl<sub>2</sub>. These data revealed that relative incorporation rates of CTP (incorrect) versus UTP (correct) were lower at pH 6.75/1.5 mM Mg<sup>2+</sup> than at pH 7.9/6 mM Mg<sup>2+</sup>. However, reliable estimates of kinetic constants were obtained only for the misincorporation of C at pH 6.75/1.5 mM Mg<sup>2+</sup> (Fig. 1B).

Unlike in the case of pyrimidines, the efficiency of ATP and GTP misincorporation by T7 RNAP was much lower. Kinetic parameters, given in the Supplementary Table S2, showed that T7 RNAP incorporated non-complementary purines with 2.6- to 3.0-fold higher efficiencies at pH 7.9/6 mM Mg<sup>2+</sup> than at pH 6.75/1.5 mM Mg<sup>2+</sup>. ATP was incorporated more efficiently than GTP in both reaction conditions. The catalytic efficiency of ATP incorporation was 29.15 ± 15.17 M<sup>-1</sup>s<sup>-1</sup> at pH 7.9/6 mM Mg<sup>2+</sup> and 11.05 ± 1.53 M<sup>-1</sup>s<sup>-1</sup> at pH 6.75/1.5 mM Mg<sup>2+</sup>. Similar differences were obtained for GTP, although at pH 6.75/1.5 mM the  $k_{po}/K_d$  of the T7 RNAP was only 6.60 ± 1.16 M<sup>-1</sup>s<sup>-1</sup>. Taken together, these data provide further support to the notion that at pH 6.75 and 1.5 mM Mg<sup>2+</sup> the T7 RNAP showed a modest but consistent improvement in its fidelity of RNA synthesis.

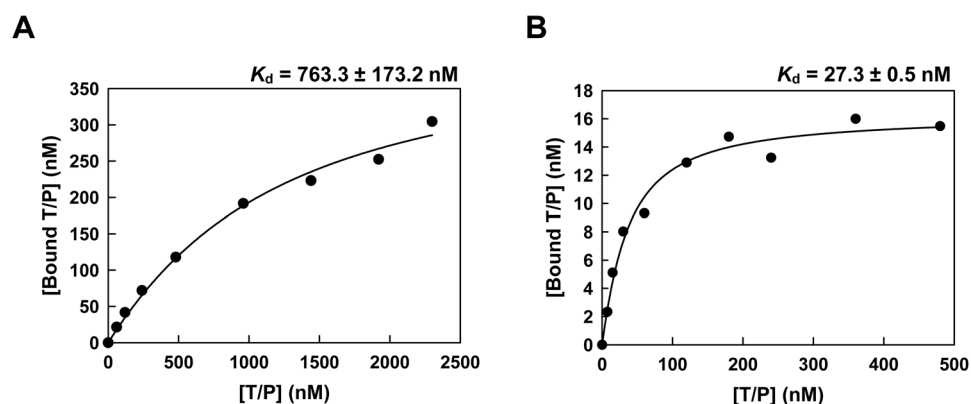
Interestingly, in addition to its effects on fidelity, the assay conditions had a strong influence on template-primer binding. At pH 7.9 and 6 mM Mg<sup>2+</sup>, the dissociation equilibrium constant ( $K_d$ ) for T7 RNAP and the VSR10 heteroduplex was about 30 times higher than at pH 6.75 and 1.5 mM MgCl<sub>2</sub> (Fig. 2). The higher affinity for VSR10 of the T7 RNAP at lower pH and Mg<sup>2+</sup> concentration could represent an additional advantage towards increasing the efficiency of RNA synthesis in the presence of limiting amounts of template.

### Estimates of RNA-dependent DNA synthesis fidelity of HIV-1 RTs using more faithful RNA templates.

Although the biochemical studies reveal that it is possible to increase transcription fidelity by reducing the pH and magnesium concentration of the reaction, it is not clear whether this would affect estimates of fidelity of RNA-dependent DNA synthesis of retroviral RTs. As shown in Tables 1 and 2, the highest mutant frequency and error rate obtained with RNA templates synthesized at pH 7.9 and in the presence of 6 mM MgCl<sub>2</sub> were found in assays carried out with the HIV-1<sub>BH10</sub> RT. We used this enzyme to test the impact of the RNA template on mutant frequencies obtained with M13mp2 *lacZα* forward mutation assays. Theoretically, more faithful RNAs (i.e. those carrying less transcription errors) would reduce the mutational threshold of the assay. RNAs containing the *lacZα* sequence were synthesized using recombinant T7 RNAP in the two different conditions (i.e. pH 7.9/6 mM Mg<sup>2+</sup> and pH 6.75/1.5 mM Mg<sup>2+</sup>). Lower mutant frequencies were obtained with HIV-1<sub>BH10</sub> RT using RNA templates synthesized at lower pH and Mg<sup>2+</sup> concentrations, although the differences were not large (Table 3). The analysis of the mutational spectra showed similar distributions of transitions and transversions using both RNAs, but at the lower pH and Mg<sup>2+</sup> concentration, we observed an increased tendency



**Figure 1.** Nucleotide incorporation kinetics for UTP and CTP into VSR10 template-primer by recombinant T7 RNA polymerase. Nucleotide sequences of the VSR10 complex<sup>56</sup> are shown above. VSR10 is made of a double-stranded DNA (black) annealed to a 10-nt RNA primer (red). The RNA is labelled at its 5' end with [ $\gamma$ -<sup>32</sup>P] ATP. The arrow indicates the nucleotide incorporation site. (A) UTP and CTP incorporation rates at different nucleotide concentrations, obtained at pH 7.9 and 6 mM MgCl<sub>2</sub> (black circles), and at pH 6.75 and 1.5 mM MgCl<sub>2</sub> (grey circles). (B) Pre-steady-state kinetics of CTP incorporation on VSR10 by T7 RNAP, at pH 6.75 (Bis-Tris buffer) and 1.5 mM MgCl<sub>2</sub>. In the left panel, continuous lines represent the best fit of the data to the single-exponential equation, obtained at different nucleotide concentrations. The right panel shows the nucleotide concentration dependence of CTP incorporation. The continuous line represents the best fit of the  $k_{\text{obs}}$  data to the Michaelis-Menten equation. The obtained polymerization rates ( $k_{\text{pol}}$ ) and apparent equilibrium dissociation constant ( $K_{\text{d}}$ ) of CTP were  $4.98 \pm 0.71 \text{ s}^{-1}$  and  $2.23 \pm 0.61 \text{ mM}$ , respectively. Results were obtained from three independent experiments.



**Figure 2.** Dissociation equilibrium constants ( $K_{\text{d}}$ ) for T7 RNA polymerase and the VSR10 dsDNA/RNA heteroduplex. Data shown were obtained from representative assays carried out at pH 7.9 and 6 mM MgCl<sub>2</sub> (panel A) and at pH 6.75 and 1.5 mM MgCl<sub>2</sub> (panel B). The solid line represents the best fit of the data to the quadratic equation relating the template-primer bound to the T7 RNAP and the total concentration of the VSR10 heteroduplex. Reported  $K_{\text{d}}$  values represent averages  $\pm$  standard deviations obtained from at least three independent experiments.

| RTs                   | T7 RNAP source | Transcription conditions |                          | Mutant plaques | Total plaques | Mutant frequency |
|-----------------------|----------------|--------------------------|--------------------------|----------------|---------------|------------------|
|                       |                | pH                       | [Mg <sup>2+</sup> ] (mM) |                |               |                  |
| HIV-1 <sub>BH10</sub> | Recombinant    | 7.9                      | 6                        | 45             | 4,650         | 0.00968          |
| HIV-1 <sub>BH10</sub> | Recombinant    | 6.75                     | 1.5                      | 70             | 11,226        | 0.00624          |
| O_K65R/V75I           | Recombinant    | 6.75                     | 1.5                      | 56             | 14,735        | 0.00380          |
| O_K65R/V75I           | Commercial     | 6.75 <sup>a</sup>        | 1.5                      | 51             | 19,466        | 0.00262          |

**Table 3.** RNA-dependent DNA synthesis fidelity of HIV-1 RTs in M13mp2 *lacZ*<sub>α</sub> forward mutation assays. Template RNA used in these assays was obtained in T7 RNAP-catalyzed reactions carried out at different pH and [Mg<sup>2+</sup>], as indicated. <sup>a</sup>This assay was carried out with the T7 RNAP from Promega in PIPES buffer (pH 6.75), and results were obtained after pooling the products of six reverse transcription reactions.

| Mutation type        | HIV-1 <sub>BH10</sub> RT                              |            |  |            | O_K65R/V75I RT   |            |  |            |
|----------------------|---|------------|--|------------|--|------------|--|------------|
|                      | Recombinant T7 RNAP (pH 7.9; 6 mM MgCl <sub>2</sub> ) |            | Recombinant T7 RNAP (pH 6.75; 1.5 mM MgCl <sub>2</sub> ) |            | Recombinant T7 RNAP (pH 6.75; 1.5 mM MgCl <sub>2</sub> ) |            | Commercial T7 RNAP <sup>a</sup> (pH 6.75; 1.5 mM MgCl <sub>2</sub> ) |            |
|                      | No. of errors   | Error rate | No. of errors  | Error rate | No. of errors  | Error rate | No. of errors  | Error rate |
| All classes          | 47  | 1/11872    | 74 <sup>b</sup>  | 1/18204    | 59   | 1/29969    | 53   | 1/44074    |
| Base substitutions   | 26  | 1/13199    | 42   | 1/19726    | 40   | 1/27186    | 32   | 1/44893    |
| Transitions          | 14 (54%)  |            | 25 (60%)   |            | 30 (75%)   |            | 19 (59%)   |            |
| Transversions        | 12 (46%)  |            | 17 (40%)   |            | 10 (25%)   |            | 13 (41%)   |            |
| Frameshifts          | 21  | 1/23649    | 32   | 1/37467    | 19   | 1/82826    | 21   | 1/98999    |
| Insertions           | 10 (48%)  |            | 19 (59%)   |            | 16 (84%)   |            | 9 (43%)  |            |
| Deletions            | 11 (52%)  |            | 13 (41%)   |            | 3 (16%)  |            | 12 (57%)   |            |
| At runs <sup>c</sup> | 10 (48%)  | 1/10602    | 21 (66%)   | 1/12188    | 15 (79%)   | 1/22397    | 11 (52%)   | 1/40348    |
| At non-runs          | 11 (52%)  | 1/35509    | 11 (34%)   | 1/85726    | 4 (21%)  | 1/309435   | 10 (48%)   | 1/163514   |

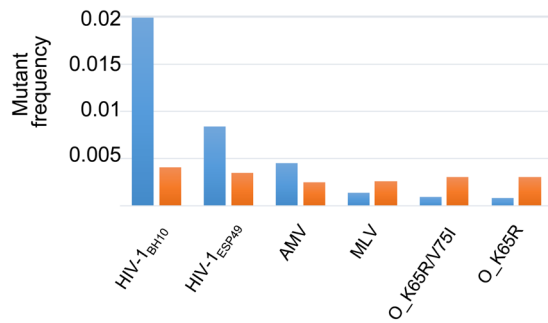
**Table 4.** Summary of error rates for RNA-dependent DNA synthesis catalyzed by HIV-1 RTs, and obtained with RNAs synthesized by the T7 RNAP under different assay conditions. <sup>a</sup>The T7 RNAP used in these assays was obtained from Promega, and buffers for the corresponding RNA synthesis reactions were prepared with PIPES. <sup>b</sup>Recombinant mutants are excluded from the analysis. <sup>c</sup>A run is considered when there is a row of three or more identical nucleotides.

to introduce frameshift errors in homopolymeric *versus* heteropolymeric sequences (Supplementary Figs S9 and S10). In any case, base substitution and frameshift error rates were about 49–58% lower in assays carried out with RNA obtained using the pH 6.75/1.5 mM Mg<sup>2+</sup> conditions (Table 4).

In previous studies, we found that the mutant O\_K65R/V75I RT was >15-fold more accurate than the HIV-1<sub>BH10</sub> RT in forward mutation assays measuring fidelity of DNA-dependent DNA synthesis<sup>18,29</sup>. Although mutant frequencies obtained with the double-mutant RT using RNAs synthesized at pH 6.75 and in the presence of 1.5 mM MgCl<sub>2</sub> were lower than those obtained with HIV-1<sub>BH10</sub> RT (Table 3), the effects were not large. Error rates of  $3.3 \times 10^{-5}$  were estimated for the O\_K65R/V75I RT under those conditions (Table 4). Interestingly, the analysis of mutational spectra revealed that major hotspots obtained with O\_K65R/V75I RT while reverse transcribing an RNA template synthesized at pH 7.9/6 mM Mg<sup>2+</sup> (Supplementary Fig. S5) were absent from the spectra obtained with the RNA made at pH 6.75/1.5 mM Mg<sup>2+</sup> (Supplementary Fig. S11). Examples of hotspots that seemed to be lost with RNAs obtained at lower pH and Mg<sup>2+</sup> concentration included clusters of U-to-C transitions found at positions -36, -35 and +147.

The analysis of data shown in Tables 1–4 also revealed that RNA synthesized by recombinant T7 RNAP had lower quality than the RNA made with commercial preparations of the enzyme that rendered lower mutant frequencies in M13mp2-based assays, even at pH 7.9 and high Mg<sup>2+</sup> concentration. Thus, error rates obtained with the HIV-1<sub>BH10</sub> RT using RNA synthesized at pH 7.9/6 mM Mg<sup>2+</sup> were estimated to be  $8.4 \times 10^{-5}$  with the recombinant T7 RNAP and  $3.5 \times 10^{-5}$  with the commercial enzyme. Similar experiments with O\_K65R/V75I RT and RNAs transcribed at pH 6.75/1.5 mM Mg<sup>2+</sup> showed error rates of  $3.3 \times 10^{-5}$  and  $2.3 \times 10^{-5}$  for recombinant and commercial T7 RNAPs, respectively.

The smaller differences observed with the O\_K65R/V75I RT relative to the HIV-1<sub>BH10</sub> RT can be attributed to the existence of a transcriptional threshold that has a bigger influence on more faithful RTs. This limit seems to be achieved with RNAs obtained at pH 6.75/1.5 mM Mg<sup>2+</sup> with the commercial T7 RNAP. However, when the commercial enzyme is used, the pH and Mg<sup>2+</sup> concentration had a relatively small influence in the error rate ( $2.5 \times 10^{-5}$ , with RNAs obtained at pH 7.9/6 mM Mg<sup>2+</sup>) while differences may not be significant. Interestingly, and in agreement with our proposal, the error rate of HIV-1<sub>BH10</sub> RT was reduced by 1.6-fold when the RNAs were synthesized by recombinant T7 RNAP at pH 6.75/1.5 mM Mg<sup>2+</sup>, as compared with those obtained at pH 7.9/6 mM Mg<sup>2+</sup>. Meanwhile, the smaller reduction in the error rate (1.1-fold) observed with O\_K65R/V75I RT when changing transcription conditions might be attributed to the use of commercial T7 RNAP to synthesize the RNA templates.



**Figure 3.** Comparison of RNA- and DNA-dependent DNA synthesis fidelities of retroviral RTs. Estimates of fidelity of RNA-dependent DNA synthesis are based on mutant frequencies obtained with the adapted M13mp2 *lacZα* forward mutation assay, reported in this paper (orange bars). Blue bars represent previously reported values, obtained using M13mp2-based assays measuring the accuracy of DNA-dependent DNA synthesis of HIV-1<sub>BH10</sub><sup>29</sup>, HIV-1<sub>ESP49</sub><sup>36</sup>, AMV<sup>14</sup>, and MLV, O\_K65R/V75I and O\_K65R RTs<sup>18</sup>.

The origin of the T7 RNAP had a minor effect on hotspot distribution in the O\_K65R/V75I spectra when the template RNA was synthesized at lower pH and magnesium concentration (Supplementary Figs S11 and S12). Thus, major hotspots at positions +87 and +139 were found with RNAs synthesized at pH 6.75/1.5 mM Mg<sup>2+</sup>, with recombinant or commercial T7 RNAP. Differences in hotspot distribution along the *lacZα* sequence between both assays were not significant statistically.

## Discussion

Previous studies showed large variations in estimates of fidelity of retroviral RTs while synthesizing DNA using DNA templates (reviews<sup>12,35</sup>). The results obtained in the present study show that the intrinsic fidelity of RNA-dependent DNA synthesis varies within the range of  $2.3 \times 10^{-5}$  to  $5.5 \times 10^{-5}$ , in assays carried out with RNA transcribed by T7 RNAP at pH 6.75 and in the presence of 1.5 mM Mg<sup>2+</sup>. This narrow window of variation was also observed with RNA templates obtained in conditions where the T7 RNAP was found to be less faithful (i.e. pH 7.9/6 mM Mg<sup>2+</sup>), rendering overall error rates of  $2.5 \times 10^{-5}$  to  $8.4 \times 10^{-5}$ . The comparison of mutant frequencies obtained with M13mp2 *lacZα* forward mutation assays using DNA or RNA templates (Fig. 3) illustrates these differences. While HIV-1<sub>BH10</sub> RT is >15-fold less accurate than HIV-1<sub>ESP49</sub> mutant RTs K65R and K65R/V75I in DNA-dependent DNA synthesis reactions, differences between the most and least faithful RTs fall within less than two-fold in assays carried out with RNA templates (Table 1). Nonetheless, HIV-1<sub>BH10</sub> RT remains as the least faithful enzyme in both types of assays. All the results reported in Fig. 3 were obtained in our lab following the same experimental procedures, except in the case of mutation frequencies determined for AMV RT with DNA-dependent DNA synthesis fidelity assays. Our previous studies measuring the fidelity of DNA-dependent DNA synthesis have shown that the mutant frequency variability obtained in forward mutation assays was under 30% for HIV-1<sub>BH10</sub> and HIV-1<sub>ESP49</sub> RTs<sup>28,29,36</sup>, in agreement with values reported by other laboratories using mutant and WT RTs<sup>37,38</sup>.

Intrinsic error rates of RTs in DNA-dependent DNA synthesis can be as low as  $6.3 \times 10^{-6}$  (e.g. for O\_K65R and O\_K65R/V75I RTs)<sup>18</sup>, with RNA templates we obtained error rates that were more than three-fold higher with all tested enzymes. Structural studies with HIV-1 RT bound to RNA/DNA template-primers<sup>39–41</sup> have shown that the RT adopts a similar conformation than in complexes containing DNA/DNA hybrids. However, in the structures containing RNA/DNA complexes, there are additional contacts involving residues of the p51 subunit as well as 2'-OH groups of the RNA template. These interactions contribute to the higher affinity of RT for RNA/DNA hybrids reported in biochemical studies<sup>42</sup>. Although differences in the interaction with RNA/DNA complexes *versus* DNA/DNA template-primers could be potentially responsible for the different error rates obtained in assays measuring fidelity of RNA- and DNA-dependent DNA synthesis, a more likely possibility is that those differences are due to transcription errors affecting the quality of the RNA template used in the assays.

Transcription errors in prokaryotic and eukaryotic cells are very difficult to quantify and large differences in error rates have been reported by several groups. Nonetheless, rough estimates of around  $10^{-5}$  per nucleotide have been reported<sup>30–32,43</sup>. Moreover, in a recent study, Traverse and Ochman<sup>33</sup> found conserved transcription error rates (ranging from  $2.3 \times 10^{-5}$  to  $5.2 \times 10^{-5}$  per nucleotide) across different species of bacteria, determined under different growth conditions and by considering either messenger or ribosomal RNA sequences. Those values were consistent with error rates obtained in our study, suggesting that T7 RNAP might be responsible for the inaccuracy threshold observed in our assays. Unfortunately, we are not aware of any reliable estimates of fidelity for T7 RNAP, although misincorporation rates have been estimated around  $5 \times 10^{-5}$ <sup>44</sup>, in good agreement with average base substitution error rates of  $3 \times 10^{-5}$  determined by Remington *et al.*<sup>45</sup> with a codon reversion assay.

Our results also show that the accuracy of the T7 RNAP can be manipulated to some extent by changing transcription conditions, although reductions of the inaccuracy threshold were modest. T7 RNAP synthesizes more faithful RNA at lower pH (6.75 vs. 7.9) and Mg<sup>2+</sup> concentration (1.5 mM vs. 6 mM). This increase in fidelity at lower pH conditions has also been previously reported for several DNA polymerases (e.g. *Taq* DNA polymerase, exonuclease-deficient Klenow fragment of *E. coli* DNA polymerase I, human DNA polymerase  $\alpha$  and HIV-1 RT)<sup>46–48</sup>. Thus, increases of fidelity of up to 50-fold have been reported at pH 6.2 compared to pH 9.8 for base substitutions using the Klenow fragment of *E. coli* DNA polymerase I<sup>47</sup>. Authors attributed these effects to altered

template binding properties of the enzyme at a lower pH. In the case of T7 RNAP, nucleotide incorporation studies with nonpolar thymidine analogues have shown that hydrogen bonds are important for polymerization efficiency due to their role in stabilizing the closed ternary complex of the polymerase, but electrostatic interactions with the minor groove of the substrate were critical for fidelity<sup>49</sup>. On the other hand, it has been shown that HIV-1 RT and *Taq* DNA polymerase display higher fidelity at low concentrations of  $Mg^{2+}$  (e.g. 0.25 mM), although this property was not observed in experiments carried out with the MLV RT<sup>46,50</sup>. Despite improvements in fidelity obtained by reducing pH and  $Mg^{2+}$  concentrations, there are obvious limitations to this approach due to the loss of catalytic activity of the enzymes when departing from the optimal conditions for their polymerization activity.

By analysing the mutational spectra induced by the studied RTs we also obtained relevant information on preferred hotspots and types of errors on different templates (i.e. RNA *versus* DNA), and the contribution of specific errors likely made by T7 RNAP and found in RNA templates used in the assays. The mutational spectra of the six RTs studied (gathered in Fig. 4) show that RTs are prone to introduce errors when copying RNA or DNA templates at positions  $-36/-34$  (mostly, U-to-C or T-to-C transitions),  $+139/+149$ , and within nucleotides  $+165$  and  $+171$ . In the latter case, transversions are largely predominant with the DNA template, while transitions, frameshift errors and large deletions dominated the RNA-dependent mutational spectra. In addition, we observed a different distribution of hotspots when comparing spectra obtained with RNA or DNA templates. The DNA-dependent mutational spectra showed scattered hotspots at positions  $-66$ ,  $+108$ ,  $+118$ ,  $+136$  and clustered at nucleotides  $+87$  to  $+92$ . In contrast, the RNA-dependent spectra contained hotspots at positions  $-7$ ,  $+73$ ,  $+87$ ,  $+109$  and  $+121$ . Among them, the hotspot at position  $+73$  was represented mainly by U-to-A transversions and one-nucleotide frameshifts.

The mutational spectra induced by RTs during RNA-dependent DNA synthesis contain a prominent hotspot at the homopolymeric region located at  $+137/+139$ . Errors at this position consist of insertions of one T in a run of three thymidines. These errors were previously detected by Boyer *et al.*<sup>25</sup> in spectra obtained with HIV-1 and AMV RTs, and attributed to errors made by the T7 RNAP and therefore, found in the RNA template used by retroviral RTs in the cDNA synthesis reaction. Our results are consistent with that proposal, since we also found one-nucleotide insertions at this site in all of the analysed mutational spectra, without statistically significant differences between them (Supplementary Table S1). Interestingly, the insertions were not detected in the mutational spectrum induced by HIV-1 RT using *lacZ $\alpha$*  RNA templates synthesized with the T3 RNAP<sup>23</sup>, supporting the notion that this type of errors are specifically made by the T7 RNAP. Another important hotspot represented by many large deletions occurs at positions  $+167/+173$ , at the 3' end of the primer used for cDNA synthesis by the retroviral RTs. Deletions were more frequent at positions that correspond to the incorporation of the first and second nucleotides (i.e.  $+172$  and  $+173$ ), and probably result from either inefficient extension of the primer during DNA synthesis or aberrant RT/template-primer interactions. Interestingly, frameshift errors near the 3' end of the primer used in cDNA synthesis had been previously reported by Boyer *et al.*<sup>25</sup>, although in that study the hotspot induced by HIV-1 RT appeared at position  $+158$ . Interestingly, the authors also noted that this hotspot was absent from the mutational spectra generated with AMV RT. In our experiments, the large deletions at the 3' end of the primer were found only in the mutational spectra induced by HIV-1<sub>BH10</sub>, HIV-1<sub>ESP49</sub> and the O\_K65R RT, but not with oncoretroviral RTs such as MLV or AMV RTs and the HIV-1<sub>ESP49</sub> mutant K65R/V75I. The presence of this particular hotspot could be partly explained by the relatively high template-primer dissociation rate of HIV-1 RT during the first nucleotide incorporation events<sup>51</sup>.

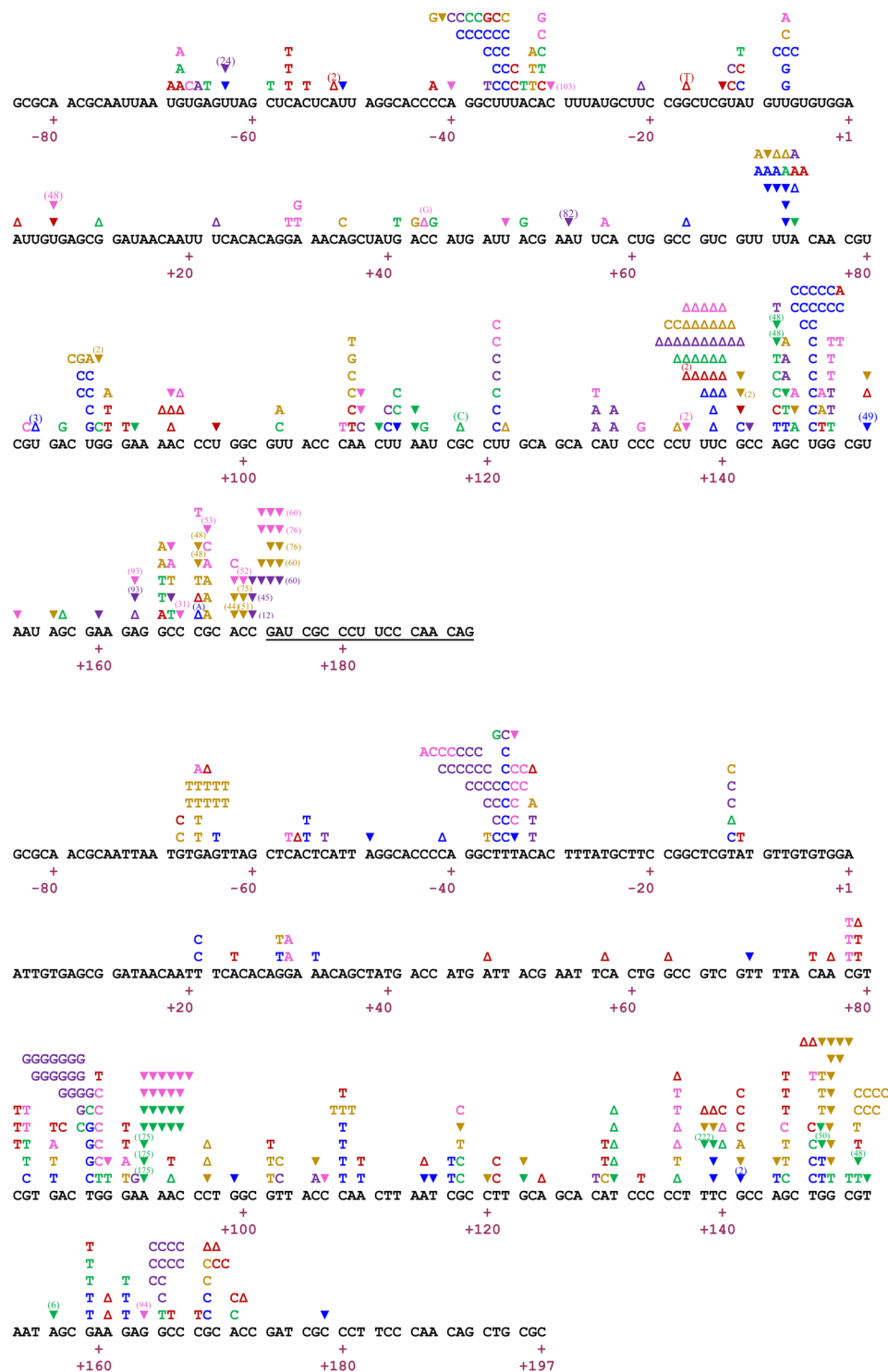
In summary, our study further demonstrates that retroviral RTs such as HIV-1<sub>BH10</sub> (group M subtype B) and HIV-1<sub>ESP49</sub> (group O) are more faithful when copying RNA than when copying DNA templates. For enzymes showing higher accuracy such as AMV RT, MLV RT and the mutant HIV-1<sub>ESP49</sub> RTs O\_K65R and O\_K65R/V75I, differences in accuracy cannot be determined in a reliable manner due to the sequence heterogeneity of the RNA used as template. Transcription errors, and specifically in our assays those made by the T7 RNAP, hamper a correct assessment of fidelity in RNA-dependent DNA synthesis reactions, even after changing the reaction conditions to improve the quality of the RNA template. Further methodological developments including a better assessment of RNA-dependent DNA synthesis fidelity of retroviral RTs are expected to be helpful for improving next-generation sequence platforms that require reverse transcription for RNA sequencing.

## Experimental Procedures

**Enzymes.** Heterodimeric (p66/p51) wild-type (WT) RTs of HIV-1 group M strain BH10 (HIV-1<sub>BH10</sub>), HIV-1 group O strain ESP49 (HIV-1<sub>ESP49</sub>) and HIV-1<sub>ESP49</sub> mutants K65R and K65R/V75I (O\_K65R and O\_K65R/V75I, respectively) were obtained as previously described<sup>18,36,52</sup>. RT p66 subunits with His<sub>6</sub> tags at their C-terminus were co-expressed with HIV-1 protease in *Escherichia coli* XL1 Blue. RT heterodimers were purified to homogeneity by ionic exchange on cellulose phosphate P11 (Whatman), followed by affinity chromatography on Ni<sup>2+</sup>-nitriloacetic-agarose (ProBond<sup>TM</sup> resin, Invitrogen)<sup>53</sup>. Purity of the enzymes was assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). RT concentrations were determined spectrophotometrically using a molar extinction coefficient of  $2.6 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$  at 280 nm, and RTs were titrated to determine their active site concentration<sup>54</sup>. MLV and AMV RTs were obtained from Promega (catalogue #M170A and #M5101, respectively).

T7 RNA polymerase (RNAP) was obtained from commercial sources (Promega catalogue #P2075 and Ambion catalogue #AM1312), or purified after expression in *E. coli* BL21 (pREP4), using the plasmid pQE9T7 encoding the full-length enzyme with a His<sub>6</sub> N-terminal extension<sup>55</sup>. The plasmid pQE9T7 was kindly provided by Ralf Ehrlich (Alere Technologies GmbH, Jena, Germany). Recombinant T7 RNAP was purified by affinity chromatography on Ni<sup>2+</sup>-nitriloacetic-agarose, as previously reported<sup>55</sup>. The purity of the enzyme was assessed by SDS-PAGE and its concentration was calculated from its absorbance at 280 nm and using a molar extinction coefficient of  $1.4 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$ .





**Figure 4.** Comparison of mutational spectra induced by retroviral RTs during RNA- and DNA-dependent DNA synthesis. (Up) Combined mutational spectra derived from RNA-dependent DNA synthesis reactions including all RTs analysed in this study (Supplementary Figs S2 to S7). RTs included are those of WT HIV-1<sub>BH10</sub> (pink), HIV-1<sub>ESP49</sub> (purple), AMV (red), and MLV (green), and mutants O\_K65R/V75I (blue) and O\_K65R (brown). (Bottom) Combined mutational spectra induced during DNA-dependent DNA synthesis reactions, and taken from previously published reports<sup>15,18,36</sup>. RT colour codes are the same as above. For all RTs and mutational spectra, single-nucleotide substitutions are indicated by the letter corresponding to the new base above the template sequence of the *lacZ* target. Open upright triangles represent insertions and inverted triangles indicate deletions. The triangles are positioned at the 3' end of the frameshift, with the number of inserted/deleted nucleotides indicated between parentheses. If not specified, one-nucleotide insertions correspond to the duplication of the base where the triangle is positioned.

**Promoter-dependent transcription assays.** The plasmid pTRI- $\beta$ -actin-Mouse (included in the Ambion MAXscript<sup>®</sup> T7 Transcription Kit) was used by T7 RNAP as the promoter-dependent template in transcription assays. These assays were used to assess nucleotide misincorporation efficiencies in the absence of one NTP or with unbalanced mixtures of NTPs under different transcription conditions. First, the plasmid was digested for 90 min at 37 °C with EcoRI (New England Biolabs) in 50  $\mu$ l of 100 mM Tris-HCl buffer pH 7.5, containing 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.025% Triton X-100 and 3 units of the enzyme per  $\mu$ g of plasmid. The resulting linearized plasmid was analysed in 1% agarose gels and purified with the NucleoSpin<sup>®</sup> Gel and PCR Clean-up Kit (Macherey-Nagel) according to the manufacturer's instructions.

Transcription assays were carried out at 37 °C for 0–30 min in different buffers depending on the desired pH (i.e. MES, PIPES, Bis-Tris and Tris-HCl). Reactions were done in 20  $\mu$ l of 40 mM buffer (pH 5.5–7.5), containing 10 mM NaCl, 0.5–6 mM MgCl<sub>2</sub>, 2 mM spermidine, 10 mM dithiothreitol (DTT), 5 units of RNasin<sup>®</sup> Plus RNase inhibitor (Promega), 15 units of T7 RNAP (Ambion) and 0.5  $\mu$ g of the linearized pTRI- $\beta$ -actin-Mouse plasmid. Reactions were initiated by the addition of different combinations of NTPs at 100  $\mu$ M each [e.g. \* (ATP, GTP, CTP, UTP); –A (GTP, CTP, UTP); –G (ATP, CTP, UTP); and –U (ATP, GTP, CTP)], as well as with unbalanced mixtures of the four NTPs, including one that is added in a low concentration [e.g. –A<sup>+</sup> (GTP, CTP and UTP at 100  $\mu$ M each, plus 1  $\mu$ M ATP); or –G<sup>++</sup> (ATP, CTP and UTP at 100  $\mu$ M each, plus 20  $\mu$ M GTP)], always in presence of 0.2  $\mu$ Ci of [ $\alpha$ -<sup>32</sup>P]CTP (Perkin Elmer). High-purity ribonucleotides were from GE Healthcare (cat. no. 27-2025-01). Reactions were quenched at appropriate times by adding the stop solution [10 mM ethylenediaminetetraacetic acid (EDTA) in 90% formamide, containing 3 mg/ml xylene cyanol FF and 3 mg/ml bromophenol blue]. The RNA products were resolved on denaturing polyacrylamide gel electrophoresis (20% polyacrylamide and 8 M urea). Then, gels were exposed to a phosphor screen and scanned with a BAS1500 phosphorimager instrument (Fuji).

**Single-nucleotide incorporation assays.** The transcription (DNA-dependent RNA synthesis) fidelity of T7 RNAP was studied by single-nucleotide incorporation assays with correct and incorrect nucleotides at pH 7.9 in the presence of 6 mM MgCl<sub>2</sub>, and at pH 6.75 in the presence of 1.5 mM MgCl<sub>2</sub>. Assays were carried out with the VSR10 template-primer<sup>36</sup>, a hybrid containing double-stranded DNA and a short RNA oligonucleotide, but lacking the T7 RNAP promoter (Fig. 1). Template DNA, non-template DNA and primer RNA were mixed at a 1:1.5:1 ratio, heated at 95 °C for 20 min, and then progressively cooled at 75, 55, 45, 20 and 4 °C for 20 min each. The 10-nt primer RNA was previously labelled at its 5' end with [ $\gamma$ -<sup>32</sup>P]ATP (Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs) in 70 mM Tris-HCl buffer pH 7.6, containing 10 mM MgCl<sub>2</sub> and 5 mM DTT. Before annealing, the labelled RNA primer was purified with a Micro bio-Spin<sup>™</sup> column (BioRad) loaded with Sephadex G-25 (GE Healthcare) to eliminate the excess of labelled ATP that could interfere with the transcription reaction.

Kinetic parameters for UTP and CTP incorporation were determined at 37 °C using a rapid quench-flow instrument (model QFM-400, Bio-Logic Science Instruments, Claix, France), upgraded with a mixer cross and a special mixer (Bio-Logic). Assays were carried out in 24  $\mu$ l of 40 mM Tris-HCl buffer pH 7.9, containing 6 mM MgCl<sub>2</sub> (or Bis-Tris buffer pH 6.75 with 1.5 mM MgCl<sub>2</sub>), 10 mM NaCl and 10 mM DTT, in the presence of increasing concentrations of NTP. The concentration of VSR10 was 240 nM and the recombinant T7 RNAP was added to a final concentration of 1  $\mu$ M. Reactions were stopped with EDTA (0.3 M final concentration). Pre-steady-state kinetic data were fit to a burst equation:  $[P] = A \times [1 - \exp(-k_{\text{obs}} \times t)] + k_{\text{ss}} \times t$ , where  $[P]$  is the product concentration,  $A$  is the amplitude of the burst,  $k_{\text{obs}}$  is the apparent kinetic constant of formation of the phosphodiester bond and  $k_{\text{ss}}$  is the kinetic constant of the steady-state linear phase. The dependence of  $k_{\text{obs}}$  on the NTP concentration is described by the equation:  $k_{\text{obs}} = k_{\text{pol}} \times [\text{dNTP}] / (K_{\text{d}} + [\text{dNTP}])$ , where  $K_{\text{d}}$  is the equilibrium constant and  $k_{\text{pol}}$  is the catalytic rate constant of the nucleotide incorporation reaction. Kinetic parameters were determined using curve-fitting tools provided by the SigmaPlot software (Systat Software Inc.).

The incorporation of non-complementary nucleotides GTP and ATP was performed manually with template-primer VSR10 (240 nM). Reactions were carried out in 20  $\mu$ l of 40 mM Tris-HCl buffer pH 7.9 (or Bis-Tris buffer pH 6.75), containing 6 mM MgCl<sub>2</sub> (or 1.5 mM MgCl<sub>2</sub>), 10 mM NaCl, 10 mM DTT, 2 mM spermidine and 24 units of RNasin<sup>®</sup> Plus RNase inhibitor. The T7 RNAP was supplied at 1  $\mu$ M. Samples were pre-incubated at 37 °C for 10 min, and the reaction was then initiated by adding increasing concentrations of NTP. Reactions were stopped at different times (0–30 s) by mixing an aliquot of 4  $\mu$ l with an equal volume of stop solution. Samples were incubated at 90 °C for 10 min before loading in polyacrylamide-urea gels. Results were analysed by phosphorimaging and nucleotide incorporation data were fitted to the Michaelis-Menten equation, as described above.

**Binding affinity of T7 RNAP for template-primer VSR10.** The equilibrium dissociation constant ( $K_{\text{d}}$ ) for T7 RNAP binding to the template-primer VSR10 was determined after pre-incubating the enzyme with increasing concentrations of the 5'-<sup>32</sup>P-labelled VSR10 (60 nM–2.3  $\mu$ M) at 37 °C for 10 min and then initiating the incorporation reaction by adding UTP to a final concentration of 5  $\mu$ M. Reactions were carried out in 20  $\mu$ l of 40 mM Tris-HCl buffer pH 7.9 containing 6 mM MgCl<sub>2</sub> (or Bis-Tris buffer pH 6.75 containing 1.5 mM MgCl<sub>2</sub>), and 10 mM NaCl, 10 mM DTT, 2 mM spermidine and 24 units of RNasin<sup>®</sup> Plus RNase inhibitor. The T7 RNAP was supplied at 0.75–1  $\mu$ M in reactions carried out at pH 7.9 with 6 mM MgCl<sub>2</sub>. At pH 6.75 and 1.5 mM MgCl<sub>2</sub>, the T7 RNAP concentration was 120–250 nM. Aliquots of 4  $\mu$ l were taken at 10, 20, 30 and 40 s, quenched with stop solution and analysed by denaturing polyacrylamide gel electrophoresis and phosphorimaging, as described above. The burst amplitudes (RT bound to template-primer at time zero) were plotted as a function of the template-primer concentration, and the data were fitted to the following quadratic equation:

$$[E \bullet T/P] = 0.5 \times \left[ (K_d + E_T + T/P) - \sqrt{(K_d + E_T + T/P)^2 - 4 \times E_T \times T/P} \right]$$

where  $E_T$  and  $T/P$  are the total active enzyme and template-primer concentration used in the assay, respectively, and  $K_d$  is the equilibrium dissociation constant for T7 RNAP binding to the template-primer.

**M13mp2-based forward mutation assays.** The fidelity of RNA-dependent DNA synthesis of retroviral RTs was determined using M13mp2-based forward mutation assays (Supplementary Fig. S1)<sup>57</sup>. Briefly, the dsDNA genome of the M13mp2 bacteriophage containing the promoter for T7 RNAP upstream the *lacZα* gene (T7-M13mp2) was digested with FspI (New England Biolabs). Five or six μg of linearized T7-M13mp2 were used as template for transcription by T7 RNAP (either the recombinant enzyme or the commercially available polymerase from Promega). Reactions were carried out in a total volume of 50 μl. Two different transcription conditions were used: (i) 40 mM Tris-HCl buffer (pH 7.9), containing 6 mM MgCl<sub>2</sub> and 0.5 mM of each NTP; and (ii) 40 mM Bis-Tris or PIPES buffer (pH 6.75), containing 1.5 mM MgCl<sub>2</sub> and 0.2 mM of each NTP. In both cases, reactions also contained 10 mM NaCl, 10 mM DTT, 2 mM spermidine and 50–60 units of RNasin® Plus RNase inhibitor. Samples were incubated at 37 °C for 2 hours to obtain a 313-nucleotide RNA product. The T7-M13mp2 dsDNA template was then digested with 5–6 units of RQ1 RNase-free DNase (Promega) at 37 °C during 15 min, and the RNA product was purified by extraction with phenol:chloroform:isoamyl alcohol (25:24:1, by volume), followed by ethanol precipitation in the presence of sodium acetate.

The purified RNA was used as template in reverse transcription reactions carried out with WT AMV, MLV, HIV-1<sub>BH10</sub> and HIV-1<sub>ESP49</sub> RTs, and mutant O\_K65R and O\_K65R/V751 RTs. The cDNA synthesis reactions were carried out at 37 °C for 2 hours in 50 μl of 25 mM Tris-HCl buffer pH 8.0, containing 100 mM KCl, 2 mM DTT, 4 mM MgCl<sub>2</sub>, 250 μM of each dNTP and 100 nM RT (active site concentration). Each reaction contained 2 pmol of RNA, previously heated to 65–70 °C for 5 min, and primed with a two-fold molar excess of the Rtr174/18 oligonucleotide (5'-CTGTTGGGAAGGGCGATC-3'). The reactions were stopped by adding EDTA at a final concentration of 15 mM and incubated at 80 °C for 5 min. Six to ten reverse transcription reactions were performed for each enzyme. The RNA template used in reverse transcription was digested at 37 °C for 1 hour, after adding 1 μl of RNase Cocktail™ containing RNase A and RNase T1 (Ambion). The obtained cDNA products were then purified by phenol extraction as described above. The cDNA was then phosphorylated with T4 polynucleotide kinase and 1 mM of ATP at 37 °C for 1 hour. The reaction was stopped after incubation at 65 °C for 5 min.

Finally, the cDNA was hybridised with gapped DNA (M13mp2 dsDNA lacking the *lacZα* gene in one of the two strands of DNA) in a solution containing 300 mM NaCl and 30 mM sodium citrate. Gapped DNA was obtained as previously described<sup>13</sup>. The cDNA was added in excess over gapped DNA and the mixture was heated to 70–75 °C for 5 min and slowly cooled at room temperature. The annealed products were analysed by electrophoresis in 0.8% agarose gels. These hybrid products were electroporated into *E. coli* MC1061 cells, and the transformed cells were plated onto M9 medium-containing plates with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and isopropyl-β-D-galactopyranoside (IPTG) with *E. coli* CSH50 lawn cells. In *E. coli*, the phosphorylation of the cDNA facilitates its ligation to the gapped DNA while repairing the nick occurring between positions +191 and +192 of the *lacZα* gene. Mutant frequencies were calculated as the ratio of mutant (light blue or colourless) plaques to the total number of plaques screened. Mutant phenotypes were confirmed by nucleotide sequencing of the phage replicative-form DNA using primer 5'-GCTTGCTGCAACTCTCTCAG-3' (Macrogen Inc., Seoul, South Korea).

Specific error rates were calculated as previously described<sup>57</sup>, but taking into account that the primer used for reverse transcription hybridised at positions +174 to +191 of the *lacZα* nucleotide sequence. Errors were detected at 123 template positions for base substitutions and 178 for frameshifts (38 at runs, and 140 at non-runs). Despite the different target size for base substitutions and frameshifts, we determined overall error rates to facilitate the comparison of different retroviral RTs by considering the 200 template positions where any phenotypic change could be detected. It should be noted that silent mutations and mutations that do not affect the β-galactosidase activity may not be detected in these assays. It has been estimated that this limitation results in a 2- to 3-fold underestimation of the actual mutation rates<sup>58</sup>. Statistical analysis about the differences in the proportion of mutations at specific sites in the spectra of different RTs was determined by using a two-tailed Fisher's exact test, using the GraphPad software.

## References

- Hu, W. S. & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med.* **2**, a006882 (2012).
- Menéndez-Arias, L., Sebastián-Martín, A. & Álvarez, M. Viral reverse transcriptases. *Virus Res.* **234**, 153–176 (2017).
- Kurimoto, K. & Saitou, M. Single-cell cDNA microarray profiling of complex biological processes of differentiation. *Curr. Opin. Genet. Dev.* **20**, 470–477 (2010).
- Mutz, K. O., Heilkenbrinker, A., Lönne, M., Walter, J. G. & Stahl, F. Transcriptome analysis using next-generation sequencing. *Curr. Opin. Biotechnol.* **24**, 22–30 (2013).
- Hrdlickova, R., Toloue, M. & Tian, B. RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA* **8**, e1364 (2017).
- Arezi, B. & Hogrefe, H. Novel mutations in Moloney murine leukemia virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Res.* **37**, 473–481 (2009).
- Baranauskas, A. *et al.* Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Eng. Des. Sel.* **25**, 657–668 (2012).
- Matamoros, T., Barrioluengo, V., Abia, D. & Menéndez-Arias, L. Major groove binding track residues of the connection subdomain of human immunodeficiency virus type 1 reverse transcriptase enhance cDNA synthesis at high temperatures. *Biochemistry* **52**, 9318–9328 (2013).
- Mohr, S. *et al.* Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**, 958–970 (2013).

10. Oszolák, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87–98 (2011).
11. Menéndez-Arias, L. Molecular basis of fidelity of DNA synthesis and nucleotide specificity of retroviral reverse transcriptases. *Prog. Nucleic Acid Res. Mol. Biol.* **71**, 91–147 (2002).
12. Menéndez-Arias, L. Molecular basis of human immunodeficiency virus type 1 drug resistance: Overview and recent developments. *Antiviral Res.* **98**, 93–120 (2013).
13. Bebenek, K. & Kunkel, T. A. Analyzing fidelity of DNA polymerases. *Methods Enzymol.* **262**, 217–232 (1995).
14. Roberts, J. D., Bebenek, K. & Kunkel, T. A. The accuracy of reverse transcriptase from HIV-1. *Science* **242**, 1171–1173 (1988).
15. Roberts, J. D. *et al.* Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis *in vitro*. *Mol. Cell Biol.* **9**, 469–476 (1989).
16. Barrioluengo, V., Wang, Y., Le Grice, S. F. & Menéndez-Arias, L. Intrinsic DNA synthesis fidelity of xenotropic murine leukemia virus-related virus reverse transcriptase. *FEBS J.* **279**, 1433–1444 (2012).
17. Shah, F. S. *et al.* Differential influence of nucleoside analog-resistance mutations K65R and L74V on the overall mutation rate and error specificity of human immunodeficiency virus type 1 reverse transcriptase. *J. Biol. Chem.* **275**, 27037–27044 (2000).
18. Barrioluengo, V., Álvarez, M., Barbieri, D. & Menéndez-Arias, L. Thermostable HIV-1 group O reverse transcriptase variants with the same fidelity as murine leukaemia virus reverse transcriptase. *Biochem. J.* **436**, 599–607 (2011).
19. Yu, H. & Goodman, M. F. Comparison of HIV-1 and avian myeloblastosis virus reverse transcriptase fidelity on RNA and DNA templates. *J. Biol. Chem.* **267**, 10888–10896 (1992).
20. Bakhanashvili, M. & Hizi, A. The fidelity of the reverse transcriptases of human immunodeficiency viruses and murine leukemia virus, exhibited by the mispair extension frequencies, is sequence dependent and enzyme related. *FEBS Lett* **319**, 201–205 (1993).
21. Kerr, S. G. & Anderson, K. S. RNA dependent DNA replication fidelity of HIV-1 reverse transcriptase: evidence of discrimination between DNA and RNA substrates. *Biochemistry* **36**, 14056–14063 (1997).
22. Hübner, A., Kruhoffer, M., Grosse, F. & Krauss, G. Fidelity of human immunodeficiency virus type I reverse transcriptase in copying natural RNA. *J. Mol. Biol.* **223**, 595–600 (1992).
23. Ji, J. & Loeb, L. A. Fidelity of HIV-1 reverse transcriptase copying RNA *in vitro*. *Biochemistry* **31**, 954–958 (1992).
24. Ji, J. & Loeb, L. A. Fidelity of HIV-1 reverse transcriptase copying a hypervariable region of the HIV-1 *env* gene. *Virology* **199**, 323–330 (1994).
25. Boyer, J. C., Bebenek, K. & Kunkel, T. A. Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc. Natl. Acad. Sci. USA* **89**, 6919–6923 (1992).
26. Bebenek, K., Abbotts, J., Roberts, J. D., Wilson, S. H. & Kunkel, T. A. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J. Biol. Chem.* **264**, 16948–16956 (1989).
27. Bebenek, K., Abbotts, J., Wilson, S. H. & Kunkel, T. A. Error-prone polymerization by HIV-1 reverse transcriptase – Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. *J. Biol. Chem.* **268**, 10324–10334 (1993).
28. Álvarez, M., Barrioluengo, V., Afonso-Lehmann, R. N. & Menéndez-Arias, L. Altered error specificity of RNase H-deficient HIV-1 reverse transcriptases during DNA-dependent DNA synthesis. *Nucleic Acids Res.* **41**, 4601–4612 (2013).
29. Álvarez, M., Sebastián-Martín, A., García-Marquina, G. & Menéndez-Arias, L. Fidelity of classwide-resistant HIV-2 reverse transcriptase and differential contribution of K65R to the accuracy of HIV-1 and HIV-2 reverse transcriptases. *Sci. Rep.* **7**, 44834 (2017).
30. Gout, J. F., Thomas, W. K., Smith, Z., Okamoto, K. & Lynch, M. Large-scale detection of *in vivo* transcription errors. *Proc. Natl. Acad. Sci. USA* **110**, 18584–18589 (2013).
31. Imashimizu, M., Oshima, T., Lubkowska, L. & Kashlev, M. Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Res.* **41**, 9090–9104 (2013).
32. Magnuson, B., Bedi, K. & Ljungman, M. Genome stability versus transcript diversity. *DNA Repair (Amst.)* **44**, 81–86 (2016).
33. Traverse, C. C. & Ochman, H. Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. *Proc. Natl. Acad. Sci. USA* **113**, 3311–3316 (2016); **113**, E4257–E4258 (erratum) (2016).
34. Tunitskaya, V. L. & Kochetkov, S. N. Structural-functional analysis of bacteriophage T7 RNA polymerase. *Biochemistry (Mosc.)* **67**, 1124–1135 (2002).
35. Menéndez-Arias, L. Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. *Viruses* **1**, 1137–1165 (2009).
36. Álvarez, M., Matamoros, T. & Menéndez-Arias, L. Increased thermostability and fidelity of DNA synthesis of wild-type and mutant HIV-1 group O reverse transcriptases. *J. Mol. Biol.* **392**, 872–884 (2009).
37. Bebenek, K. *et al.* Reduced frameshift fidelity and processivity of HIV-1 reverse transcriptase mutants containing alanine substitutions in helix H of the thumb subdomain. *J. Biol. Chem.* **270**, 19516–19523 (1995).
38. Rezende, L. F., Curr, K., Ueno, T., Mitsuya, H. & Prasad, V. R. The impact of multidideoxynucleoside resistance-conferring mutations in human immunodeficiency virus type 1 reverse transcriptase on polymerase fidelity and error specificity. *J. Virol.* **72**, 2890–2895 (1998).
39. Sarafianos, S. G. *et al.* Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J.* **20**, 1449–1461 (2001).
40. Lapkouski, M., Tian, L., Miller, J. T., Le Grice, S. F. J. & Yang, W. Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. *Nat. Struct. Mol. Biol.* **20**, 230–236 (2013).
41. Das, K., Martínez, S. E., Bandwar, R. P. & Arnold, E. Structures of HIV-1 RT-RNA/DNA ternary complexes with dATP and nevirapine reveal conformational flexibility of RNA/DNA: insights into requirements for RNase H cleavage. *Nucleic Acids Res.* **42**, 8125–8137 (2014).
42. Bohlayer, W. P. & DeStefano, J. J. Tighter binding of HIV reverse transcriptase to RNA-DNA versus DNA-DNA results mostly from interactions in the polymerase domain and requires just a small stretch of RNA-DNA. *Biochemistry* **45**, 7628–7638 (2006).
43. Ninio, J. Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. *Genetics* **129**, 957–962 (1991).
44. Huang, J., Brieba, L. G. & Sousa, R. Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* **39**, 11571–11580 (2000).
45. Remington, K. M., Bennett, S. E., Harris, C. M., Harris, T. M. & Bebenek, K. Highly mutagenic bypass synthesis by T7 RNA polymerase of site-specific benzo[a]pyrene diol epoxide-adducted template DNA. *J. Biol. Chem.* **273**, 13170–13176 (1998).
46. Eckert, K. A. & Kunkel, T. A. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res.* **18**, 3739–3744 (1990).
47. Eckert, K. A. & Kunkel, T. A. Effect of reaction pH on the fidelity and processivity of exonuclease-deficient Klenow polymerase. *J. Biol. Chem.* **268**, 13462–13471 (1993).
48. Eckert, K. A. & Kunkel, T. A. Fidelity of DNA synthesis catalyzed by human DNA polymerase alpha and HIV-1 reverse transcriptase: effect of reaction pH. *Nucleic Acids Res.* **21**, 5212–5220 (1993).
49. Ulrich, S. & Kool, E. T. Importance of steric effects on the efficiency and fidelity of transcription by T7 RNA polymerase. *Biochemistry* **50**, 10343–10349 (2011).
50. Achuthan, V., Keith, B. J., Connolly, B. A. & DeStefano, J. J. Human immunodeficiency virus reverse transcriptase displays dramatically higher fidelity under physiological magnesium conditions *in vitro*. *J. Virol.* **88**, 8514–8527 (2014).

51. Majumdar, C., Abbotts, J., Broder, S. & Wilson, S. H. Studies on the mechanism of human immunodeficiency virus reverse transcriptase. Steady-state kinetics, processivity, and polynucleotide inhibition. *J. Biol. Chem.* **263**, 15657–15665 (1988).
52. Matamoros, T. *et al.* Suppression of multidrug-resistant HIV-1 reverse transcriptase primer unblocking activity by  $\alpha$ -phosphate-modified thymidine analogues. *J. Mol. Biol.* **349**, 451–463 (2005).
53. Boretto, J. *et al.* An integrated system to study multiply substituted human immunodeficiency virus type 1 reverse transcriptase. *Anal. Biochem.* **292**, 139–147 (2001).
54. Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* **267**, 25988–25997 (1992).
55. Ellinger, T. & Ehrlich, R. Single-step purification of T7 RNA polymerase with a 6-histidine tag. *Biotechniques* **24**, 718–720 (1998); **25**, 640 (erratum) (1998).
56. Anand, V. S. & Patel, S. S. Transient state kinetics of transcription elongation by T7 RNA polymerase. *J. Biol. Chem.* **281**, 35677–35685 (2006).
57. Boyer, J. C., Bebenek, K. & Kunkel, T. A. Analyzing the fidelity of reverse transcription and transcription. *Methods Enzymol.* **275**, 523–537 (1996).
58. Abram, M. E., Ferris, A. L., Shao, W., Alvord, W. G. & Hughes, S. H. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J. Virol.* **84**, 9864–9878 (2010).

## Acknowledgements

We thank Thomas Kunkel and Katarzyna Bebenek for technical advice on the M13mp2 DNA synthesis fidelity assay, and critical reading of the manuscript, and Ralf Ehrlich for providing us with the plasmid used for the expression and purification of T7 RNAP. This work was supported by grants of the Spanish Ministry of Economy and Competitiveness (BIO2013–48788-C2-1-R and BIO2016-76716-R (AEI/FEDER, UE)), and an institutional grant of the Fundación Ramón Areces. A. S.-M. is a recipient of a predoctoral fellowship of the Spanish Ministry of Education, Culture and Sport (FPU2013-00693).

## Author Contributions

This study was conceived and designed by L.M.-A. A.S.-M. and V.B. performed experiments. All three authors analysed data. A.S.-M. and L.M.-A. wrote the manuscript, which was critically revised by V.B. All authors read and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-18974-8>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018