Determinants of Selective Fidelity in Diversity-Generating Retroelements

Sumit Handa[1], Andres Reyna[1], Timothy Wiryaman[1], Partho Ghosh[1,*]

[1]Department of Chemistry & Biochemistry, 9500 Gilman Drive, La Jolla, CA, USA 92093-0375

* To whom correspondence should be addressed. Tel: 1-858-822-1139; Fax: 1-858-822-2871; Email: pghosh@ucsd.edu

Present Address: Andres Reyna, Department of Chemistry, University of Washington, Seattle, WA, USA 98195

ABSTRACT

Diversity-generating retroelements (DGRs) diversify proteins to the greatest extent known in the natural world. These elements are found in a wide variety of microbes, including constituents of the microbial 'dark matter' and the human microbiome. Diversification occurs through selective fidelity, in which genetic information in RNA is reverse transcribed to cDNA faithfully for all bases but adenine. We investigated the determinants of selective fidelity in the prototypical *Bordetella* bacteriophage DGR using an *in vitro* system that recapitulates this process. We found that the DGR reverse transcriptase bRT and associated protein Avd, along with a specific DGR RNA that contains a template region and flanking functional elements, had a markedly low catalytic efficiency across a template adenine. This was the case even for the correct base pair with an incoming TTP, in agreement with results from other low fidelity polymerases. We identified the C6 substituent of a template purine as a major determinant of selective fidelity, and also identified two bRT amino acids with counterparts in HIV RT based on an *in silico* model, R74 and I181, whose substitution led to increased fidelity. Our results provide the first elucidation of nucleobase and protein determinants of selective fidelity in DGRs.

INTRODUCTION

Adaptation by organisms to novel selective pressures requires variation. While this usually occurs over multiple generations and lengthy time scales, there are two examples of instantaneous adaptation that takes place within a single generation. These examples are the variation of antigen receptors by the vertebrate adaptive immune system, and the variation of select proteins belonging to diversity-generating retroelements (DGRs) (1). DGRs are prevalent in the microbial 'dark matter', which appear to comprise a major fraction of microbial life, and are widespread in the human virome and microbiome (2-7). The level of DGR variation exceeds by many orders of magnitude the 10^{14-16} of the vertebrate immune system (8). A DGR variable protein with 10^{20} possible sequences has been structurally characterized (9), and one with 10^{30} possible sequences has been identified (10). In the adaptive immune system, variation enables the recognition of novel targets and consequent adaptation to dynamic environments. A similar benefit appears to be provided by DGRs, as documented for the prototypical DGR of *Bordetella* bacteriophage (1). This DGR encodes the

bacteriophage's receptor-binding protein Mtd. Variation in Mtd enables the bacteriophage to adapt to the loss of potential surface receptors in its host *Bordetella*, which happens because of environmental changes or immune pressure (11,12).

DGRs diversify proteins through a fundamentally different mechanism from that of the adaptive immune system and indeed any biological system. Diversification by DGRs arises from selective fidelity, in which genetic information is transmitted faithfully for all bases but adenine. This occurs during reverse transcription of protein-coding RNA to cDNA. The RNA is part of the invariant template region (*TR*), and the resulting adenine-mutagenized cDNA homes to and replaces a part of the variable region (*VR*) in a DGR variable protein gene, such as *mtd*. As shown for the *Bordetella* bacteriophage DGR, adenines in RNA are reverse transcribed into cDNA incorrectly at an astonishing frequency of 50% (1,13,14).

We recently reconstituted selective fidelity *in vitro* (14). This reconstitution showed that the *Bordetella* bacteriophage DGR reverse transcriptase bRT and the associated pentameric protein Avd (15), along with a specific DGR RNA, are necessary and sufficient for selective fidelity. The DGR RNA, which is identical to the 'core' DGR RNA described previously (14), contains the 134-nucleotide (nt) *TR* flanked by two functional elements, a short 20-nt segment from *avd* at the 5' end and a longer 140-nt spacer (Sp) region separating *TR* and *brt* at the 3' end (Fig. 1a) (14). The mechanistic role of the 5' *avd* element is unknown, but the Sp region has been shown to provide an essential binding site for Avd and to supply the site from which reverse transcription is primed, Sp A56 (13,14). A number of pieces of evidence indicate that the 2'-OH of Sp 56 provides the priming nucleophile, resulting in a branched, covalently-linked RNA-cDNA molecule (14). An alternative model involving RNA cleavage at Sp 56 to expose its 3'-OH for priming has been proposed as well (13). The first nucleotide reverse transcribed is *TR* G117, and cDNAs typically extend from there to *TR* 22-24 (~90 nt), which includes all adenines in *TR* whose substitution leads to a coding change, or just further into *avd* (~120 nt) (Fig. 1a).

bRT-Avd is also capable of synthesizing cDNAs from non-DGR RNA templates (14). For this, an exogenous primer is required and only short cDNAs (~5-35 nt) are synthesized. These results indicate that template-priming and processive polymerization are both specific properties of the DGR RNA. Evidence suggests that this is because bRT-Avd and the DGR RNA combine to form a structured ribonucleoprotein (RNP) particle that aligns the priming site at Sp 56 with the reverse transcription start site at *TR* 117, and also maintains interaction between bRT-Avd and the RNA that is conducive to processive polymerization (14). Avd is still required for cDNA synthesis in these cases, indicating that Avd has a role in catalysis that is independent of its role in binding Sp. Notably, cDNAs produced from non-DGR RNA templates are adenine-mutagenized, indicating that selective fidelity is an intrinsic property of the bRT-Avd complex and independent of the RNA template and the mechanism of priming.

To understand the determinants of selective fidelity, we characterized the *in vitro* reverse transcription properties of bRT-Avd and the DGR RNA in detail. We found that the catalytic efficiency (k_{cat}/K_m) of bRT-Avd was markedly low across a template adenine, even for the correct base pair with TTP, consistent with results from other low fidelity polymerases (16). Using nucleobase analogs, we

identified the C6 position of the purine ring as a key determinant of fidelity. The amine at C6 of adenine had no role in modulating fidelity and in effect was invisible to mismatch detection mechanisms, while the carbonyl at C6 of guanine was a major factor in discriminating between correct and incorrect base pairs. The ribose 2'-OH also promoted fidelity. bRT-Avd was able to incorporate dNTPs across an abasic template site, albeit with a significant incidence of deletions and with only a partial preference for adenine as compared to A-rule polymerases (16). We also found that two bRT amino acids, Arg 74 and Ile 181, contributed to decreased fidelity. These amino acids are predicted by an *in silico* model to have counterparts in HIV reverse transcriptase (RT) and stabilize both correct and incorrect base pairs. Substitution of these amino acids with Ala and Asn, respectively, increased fidelity across template adenines, as has been observed for their putative equivalents in HIV RT. Substitutions at a number of other bRT amino acids predicted to be proximal to the catalytic site either did not change or instead decreased fidelity at template adenines. These results provide the first detailed characterization for the basis of selective fidelity in DGRs.

MATERIAL AND METHODS

Protein and RNA. The bRT-Avd complex was expressed in *Escherichia coli* and purified, and the core DGR RNA (*avd* 368 – *TR* 140) was produced through *in vitro* transcription with T7 polymerase and gel purified, both as previously described (14). Mutants of bRT were generated using QuickChange mutagenesis (Agilent), and expressed and purified as bRT-Avd complexes. Mutants of the DGR RNA were also generated through QuickChange mutagenesis.

RNA with nucleobase analogs. RNA oligonucleotides spanning *avd* 368 to *TR* 26 were chemically synthesized (Dharmacon), with adenine or base analogs at *TR* 23 and 24 (or just 24), and gel purified. RNA spanning *TR* 27 to Sp 140 was *in vitro* transcribed, gel purified, and then treated with alkaline phosphatase (NEB) according to the manufacturer's directions at 37 °C for 2 h to remove the triphosphate from the 5' end. The dephosphorylation reaction was quenched at 80 °C for 5 min to inactivate the phosphatase. A phosphate group was added to the 5' end of the RNA using T4 polynucleotide kinase and ATP, according to the manufacturer's directions (NEB). The RNA was then purified by phenol:chloroform extraction followed by a G-25 desalting column. The chemically synthesized RNA oligonucleotide (1.1 μM) and the *in vitro* transcribed RNA (2.2 μM) were annealed in the presence of 1.7 μM splint oligodeoxynucleotide P1 (Table S1), 8% DMSO, and 0.2x T4 RNA Ligase 1 (T4Rnl1) buffer in 250 μL. The splint was annealed to the RNA by heating at 95 °C for 3 min and cooling at 0.2 °C/min to 20 °C. A 250 μL solution consisting of 0.8x T4Rnl1 buffer, 2 mM ATP, 4% DMSO, and 540 units T4Rnl1 (NEB) were mixed with the annealing reaction. The resulting mixture was incubated for 8 h at 37 °C. The sample was then extracted with phenol:chloroform and ethanol-glycogen precipitated. The pellet was resuspended in water and gel purified.

Reverse transcription reactions and analysis. Reverse transcription reactions were carried out with wild-type or mutant bRT-Avd at 37 °C for 12 h, and resulting cDNA was purified, both as previously described (14). Purified cDNA was PCR amplified using primers P2 and P3 (Table S1) and *Pfu*

polymerase. The amplified PCR product was sequenced using Amplicon EZ sequencing (Genewiz). The quality scores for sequencing reactions were Q30, which is equivalent to an error probability of 1 in 1000. Fastq files generated from sequencing were aligned using bowtie2, and the output files were sorted and indexed using samtools (17-19). Incorporation and misincorporation frequencies in the template region were calculated using an integrative genomics viewer (20).

Reverse transcription reactions with Moloney Murine Leukemia Virus (MMLV) RT (BioBharati Life Sciences Pvt. Ltd) and HIV RT (Worthington Biochemical) were carried out according to the manufacturer's directions using primer P4 (Table S1). PCR amplification of cDNAs was carried out with *Pfu* polymerase and primers P2 and P3 (Table S1).

Single deoxynucleotide primer extension assay. Oligodeoxynucleotide P^{117} (Table S1) was 5'-[^{32}P]-labeled as previously described (14). The core DGR RNA (0.5 μM) containing *TR* G117A (and U116G, in the case of experiments with dATP) was mixed with primer P^{117} (0.5 μM), 5'-[^{32}P]-labeled P^{117} (0.05 μM), varying concentration of dNTPs, and 20 units RNase inhibitor (NEB) in 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 50 mM HEPES, pH 7.5 and 10% glycerol in a 20 μL final volume. The mixture was incubated at 37 °C, and then wild-type or mutant 1 μM bRT-Avd was added. A 2.5 μL aliquot of the reaction was removed at various time points, and quenched by addition to 7.5 μL of an ice-cold solution of proteinase K (1.3 mg/ml) followed by incubation at 50 °C for 20 min. The quenched reactions were then incubated with 0.5 μL RNase A/T1 mix (Ambion) at room temperature for 20 min in a final volume of 20 μL. The samples were then ethanol-glycogen precipitated overnight at –20 °C. Samples were centrifuged the next day, and pellets were air dried and resuspended in 20 μL of RNA loading dye. Five μL of the reaction sample was loaded on an 8% denaturing sequencing gel to resolve unreacted and extended primers. The radiolabeled products were visualized by autoradiography using a Typhoon Trio (GE Healthcare Life Sciences), and band densities were quantified using ImageQuant TL 8.1 (GE Healthcare Life Sciences). Background values determined from band densities prior to any reaction were subtracted. The steady-state initial velocity with respect to substrate concentration was fit to the Michaelis-Menten equation using nonlinear regression analysis in GraphPad Prism.

RESULTS

Adenine-mutagenesis of *TR*

To characterize adenine-mutagenesis of *TR* by bRT-Avd at fine-scale, we pursued next-generation sequencing (NGS) of cDNAs. The NGS read count of ~100,000 enabled conclusions to be drawn about the distribution of adenine-mutagenesis that were not possible due to the small number of sequences previously available from single clones (~30) (1). The template was the ~300-nt core DGR RNA (14), consisting of the *TR* (134 nt) flanked by upstream *avd* and downstream Sp regions that are functionally essential (Fig. 1a) (14). As described above, reverse transcription was template-primed by the core DGR RNA from Sp A56 and initiated from *TR* G117 (14). We found that the average misincorporation frequency across the 22 adenines in *TR* that lead to coding changes was ~50% (Fig. 1b), similar to the level previously observed *in vitro* and *in vivo* (1,13,14). The most frequently

misincorporated base was adenine itself (~23%), followed by cytosine (~16%) and then guanine (~11%). As a control, we examined the misincorporation frequency across *TR* using the high fidelity reverse transcriptase MMLV RT, which has an error rate of ~10^{-5} (21). We observed an average 0.5% misincorporation frequency for MMLV RT (Fig. S1a), which sets the baseline error level for our methodology. In the case of bRT-Avd, the misincorporation frequencies across template cytosines and uracils were at baseline level (~0.3% and ~0.5%, respectively), but significantly higher for guanine (1.6%) (Fig. S1b). Across template guanines, adenines were misincorporated most (0.85%), followed by guanine (0.58%) and thymine (0.15%). Thus, bRT-Avd generally has lower fidelity at template purines, but a markedly 33-fold lower fidelity at a template adenine as compared to a template guanine.

The misincorporation frequency of bRT-Avd varied widely across individual template adenines. *TR* A23 and A62 were especially prone to misincorporation, with frequencies of 75% (Fig. 1b). Notably these adenines are the first bases in AAC codons (for Mtd 344 and 357, respectively), indicating an enhanced potential for amino acid substitution at these positions (22). In contrast, *TR* A35 and A98 were especially resistant to misincorporation, with frequencies of only 23%. Equally notably, these adenines are the only ones in their codons (ACG for Mtd 348 and ATC for Mtd 369, respectively), indicating a curtailed potential for amino acid substitution at these codons (22). The differences in misincorporation frequencies were not attributable to any obvious RNA primary sequence patterns.

To determine whether misincorporation also occurred at artificially introduced adenines, we substituted adenines into the initial four positions that are reverse transcribed (*TR* 117-114). These positions are normally occupied by G, C, or U. Misincorporation was evident at all four positions when substituted by adenine. Positions 117, 116, and 115 were even more resistant to misincorporation than naturally occurring positions in *TR*, while 114 was within the range observed for naturally occurring positions (Fig. 1c). Thus adenine-mutagenesis can occur outside naturally occurring positions and is variable.

Enzymatic Parameters of bRT-Avd

We next examined the rate of single deoxynucleotide addition by bRT-Avd. As misincorporation occurred to a detectable level across *TR* G117A, we used oligodeoxynucleotide P^{117} (Table S1) to prime synthesis from *TR* G117A (of the core DGR RNA). We have previously shown that P^{117} primes cDNA synthesis from the natural start of reverse transcription, *TR* G117, and concurrently inhibits template-primed cDNA synthesis (14). The addition of a single deoxynucleotide to radiolabeled P^{117} was easily detectable (Fig. 2a). In the case of extension with dATP, the template contained both *TR* U116G and G117A substitutions, so as to avoid incorporation of a second dATP across *TR* U116. Using this single deoxynucleotide primer extension assay, we determined steady-state enzymatic parameters for bRT-Avd.

This analysis showed that k_{cat} varied little (only up to three-fold) between correct incorporation of TTP and misincorporation of the other dNTPs across the adenine at *TR* 117 (Figs. 2b-f and Table 1).

Similarly, the K_m for the incoming dNTP varied over a very small range, with the greatest difference being ~17-fold between dGTP and TTP (Table 1). The most favorable K_m was for dUTP, reflecting the previously noted preference in bRT-Avd for dUTP over TTP (14). The catalytic efficiency (k_{cat}/K_m) for the misincorporation of dUTP was 4-fold higher than that for the correct incorporation of TTP. The misincorporation ratios (catalytic efficiency for misincorporation to correct incorporation) for dCTP and dATP were 0.05, and 0.02 for dGTP (Table 1). These values correspond closely to the misincorporation frequencies at *TR* G117A determined by NGS from template-primed reverse transcription (Fig. 1c, 7%, 3%, and 2%, respectively). Thus, template- and P^{117}-primed synthesis resulted in similar misincorporation frequencies, which indicates that the enzymatic parameters determined here for oligodeoxynucleotide-primed synthesis are likely to be applicable to template-primed synthesis.

Nucleobase Determinants of Selective Fidelity

We next asked which features of nucleobases promote selective fidelity. To address this, we compared the nearly isosteric features of infidelity-promoting adenine and fidelity-promoting guanine. The differences between A and G primarily occur at the N1, C2, and C6 positions (Fig. 3a). We sought to probe these ring positions using base analogs. For this, short RNA oligonucleotides corresponding to the 5' portion of the core DGR RNA were chemically synthesized with nucleobase analogs at *TR* 23 and 24. The oligonucleotide was ligated to a longer *in vitro* transcribed RNA corresponding to the rest of the core DGR RNA. To validate this method, we first constructed the ligated core DGR RNA using an oligonucleotide that had adenines rather than analogs at *TR* 23 and 24. The cDNAs produced by bRT-Avd from the ligated core DGR RNA template had a near identical misincorporation frequency as the fully *in vitro* transcribed core DGR RNA template (Figs. 1b and 3b).

We first examined hypoxanthine, for which the N1 and C6 groups of adenine (N with a lone electron pair and amine, respectively) are substituted with those of guanine (NH and carbonyl, respectively). Hypoxanthine preferentially forms a Watson-Crick base pair with cytosine, and a less stable wobble base pair with adenine (23,24). This preference was verified through reverse transcription with MMLV RT (Fig. 3c). Cytosine was almost exclusively incorporated across template hypoxanthines by MMLV RT. Reverse transcription was then carried out with bRT-Avd. Significantly, bRT-Avd correctly incorporated cytosine with nearly the same frequency, 96-97% (Fig. 3c), as MMLV RT. Adenines constituted the rest. This result showed that an NH at N1 or a carbonyl at C6, or both, were positive determinants of fidelity.

We next asked if the amine at C6 in adenine impacted fidelity. To do so, we used purine (i.e., nebularine), which lacks a substituent at C6, as the base analog. Purine preferentially base pairs with thymine (25), which we verified with MMLV RT (Fig. 3d). With bRT-Avd, template purines led to misincorporation with very similar frequencies as those observed for template adenines (Fig. 3d). This result indicates that an amine at C6 has the same effect on misincorporation as no substituent, and thus an amine at C6 is a neutral determinant of fidelity.

To probe the C6 position further, we used N6-methyladenine (m^6A). As previously reported, reverse transcription of m^6A by MMLV results in the incorporation of thymine (Fig. 3e) (26). However, for bRT-Avd, m^6A resulted in a greater misincorporation frequency, most notably at *TR* 24, where misincorporation increased from 35% to 66% (Fig. 3e). Misincorporation also increased at *TR* 23, although less strikingly, from 71% to 82%. Thus, while an amine at the C6 position is a neutral determinant of fidelity, the bulkier methylamino group at the C6 position is a negative determinant of fidelity. Furthermore, these results suggested that the N1 position by itself had little if any role in modulating fidelity, as adenine and m^6A are identical at the N1 position but differed in misincorporation frequency.

The C2 position was probed using 2,6-diaminopurine (DAP), which is identical to adenine except that it contains an amine at C2 as well (as does guanine). DAP preferentially base pairs with thymine but can also form a wobble base pair with cytosine (27,28). In the case of MMLV RT, exclusive incorporation of thymine was observed (Fig. 3f). However for bRT-Avd, the major species incorporated was cytosine, followed to a lesser extent by thymine and adenine (Fig. 3f). This was the case even though the Watson-Crick base pair between DAP and thymine involves three hydrogen bonds, while the wobble base pair between DAP and cytosine involves only one.

A similar trend was seen for 2-amino purine (2AP), which is identical to DAP but lacks an amine at C6. 2AP forms a base pair with thymine but can also form a wobble or protonated base pair with cytosine (28-30). Both forms involve two hydrogen bonds. MMLV RT preferentially incorporated thymine but also a substantial level of cytosine. In the case of bRT-Avd, incorporation of cytosine was greatly preferred, followed by thymine and adenine (Fig. 3g), similar to the results with DAP (Fig. 3f). The similarity of misincorporation frequencies between DAP and 2-AP (which differ only in an amine at C6) confirms that an amine at C6 is a neutral determinant of fidelity for bRT-Avd.

Taken together, these results indicate that a carbonyl at the C6 is a key determinant of fidelity, even in the absence of any substituent at the C2 position (e.g., hypoxanthine). In the absence of a carbonyl at C6, an amine at C2 appears to position the template base and incoming dNTP to favor a wobble rather than a Watson-Crick base pair.

Abasic Site and the A-rule

We noticed that adenine was the most frequently misincorporated base across the 22 sites in *TR* (Fig. 1b). A number of nucleotide polymerases have the tendency to insert an adenine across an abasic template site, which is called the A-rule (31). We investigated whether bRT-Avd follows the A-rule by constructing a core DGR RNA template with abasic sites at *TR* 23 and 24. HIV RT has been documented to tolerate abasic sites (32), and thus we used HIV RT as a positive control. However, tandem abasic sites led to a high incidence of deletions at these and surrounding sites with HIV RT, as well as with MMLV RT and bRT-Avd (Fig. S2a). Thus, we limited the abasic site to *TR* 23. HIV RT incorporated adenine almost exclusively (98%) across the abasic site (Fig. 4a), and while deletions still occurred, they occurred much less frequently than with tandem abasic sites (Fig. S2). MMLV RT

has been reported to be intolerant to abasic sites (32), and indeed the incidence of deletion was approximately three-fold higher at abasic *TR* 23 for MMLV RT than for HIV RT (Fig. S2b). Nevertheless, for those cDNAs lacking deletions, MMLV RT incorporated adenine almost exclusively (99%) across the abasic site. A significant level of deletion at and around the abasic site was also seen with bRT-Avd, with an incidence that resembled that of MMLV RT (Fig. S2b). However, the near exclusive preference for incorporating adenine seen for HIV and MMLV RTs was not observed for bRT-Avd, and instead at best a strong preference for adenine was evident (65%) (Fig. 4a). Guanine was also incorporated at an appreciable frequency by bRT-Avd across the abasic site (Fig. 4a). These results indicate that bRT-Avd does not tolerate an abasic site as well as HIV RT, and does not follow the A-rule as strictly as HIV RT.

2'-OH

We also asked whether selective fidelity is maintained when a deoxynucleotide is reverse transcribed. We therefore instituted deoxyadenosines at *TR* 23 and 24 in the core DGR RNA template. As expected, thymine was incorporated exclusively by MMLV RT across *TR* dA23 and dA24 (Fig. 4b). In contrast, the misincorporation frequency increased for bRT-Avd. It went from 32% to 65% at *TR* 24 and from 71% to 83% at *TR* 23 (Fig. 4b). To ascertain whether selectivity to adenine was maintained, *TR* 23 and 24 were instituted with deoxyguanosines. MMLV RT incorporated cytosine (99%) almost exclusively across these sites, as did bRT-Avd (97%) (Fig. S3). These results indicate that the 2'-OH is a positive determinant of selective fidelity.

bRT amino acid that modulate selective fidelity

We next sought to identify bRT amino acids that modulate selective fidelity. We pursued this through structure-guided mutagenesis using a high-confidence model of bRT (100% confidence level for 95% of the amino acid sequence) that was generated using Phyre2 (33). This bRT model is based in part on the structures of group II intron maturases (34,35), including the high-fidelity GsI-IIc RT (36). This latter structure also contains a bound RNA template-DNA primer heteroduplex and an incoming dATP. The *in silico* bRT model consists of all the important functional elements of RTs — the canonical fingers, palm, and thumb domain (Fig. 5a). In addition to this, we relied on the extensive literature on the fidelity of HIV RT, and superposed the structure of HIV RT (37) with the *in silico* model of bRT to guide the choice of substitution sites.

Based on this superposition, bRT Arg 74 is predicted to form a part of the binding pocket for the incoming dNTP (Fig. 5b). Its putative homolog in HIV RT is Arg 72, which when substituted by Ala results in increased fidelity (38,39). We found a similar effect for bRT(R74A)-Avd. Substitution of bRT Arg 74 with Ala led to a marked increase in the correct incorporation of TTP across the 22 template adenines in *TR* from 50% to 92%, with some sites reaching 99% (Fig. 5c and Fig. S4a). The fidelity at other template bases resembled that of wild-type bRT (Fig. S4b). Single deoxynucleotide primer extension analysis was carried out with bRT R74A to determine the basis for this increased fidelity.

While there was almost no change in the K_m for TTP, a 40% decrease in k_{cat} was observed (Table 2). The catalytic efficiency of bRT(R74A)-Avd was 68% of that of wild-type bRT.

Ile 181 of bRT was also predicted to be proximal to the incoming dNTP, and corresponds to HIV RT Gln 151, whose substitution by Asn leads to increased fidelity (40-42). Ile 181 constitutes the initial part of the signature DGR RT motif [I/V/L]GxxxSQ (1). In most retroviral RTs, non-LTR retrotransposon RTs, and group II intron maturases, this motif is instead QGxxxSP. We found that bRT I181N yielded an increase in fidelity at template adenines, albeit to a lesser extent than did bRT R74A (Fig. 5c and S4c). The misincorporation frequency at template cytosines for bRT I181N did not change from wild-type bRT, but did increase for template guanines from 1.6% to 3%, and for template uracils from 0.5% to 1.2% (Fig. S4d). Single deoxynucleotide primer extension analysis of bRT I181N showed a 27-fold increase in the K_m for TTP but no appreciable change in k_{cat} (Table 2). The catalytic efficiency of bRT(I181N)-Avd was only 3% of that of wild-type.

We also substituted the last amino acid in the signature DGR RT motif [I/V/L]GxxxSQ, bRT Gln 187, to the proline in the QGxxxSP motif. Proline at this position in HIV RT, P157, has been seen to contact a base in the template strand (37). bRT(Q187P)-Avd was unaltered in fidelity across template adenines compared to wild-type bRT, although this complex showed an increased bias towards misincorporating adenines over other bases (Fig. 5c).

Substitutions in three bRT amino acids with putative equivalents in HIV RT had no effect on altering fidelity. These were bRT I176V, L184A, and M214V (Figs. 5b, c). The predicted structural equivalents in HIV RT are respectively Leu 74, whose substitution by Val increases fidelity (43); Lys 154, whose substitution by Ala increases fidelity (40); and Met 184, whose substitution by Val increases fidelity (44,45). These HIV RT amino acids interact with the template or primer strand proximal to the catalytic site (37). While none of these substitutions in bRT increased fidelity, they all curiously changed the misincorporation bias towards cytosine and away from adenine (Fig. 5c).

Lastly, we probed two bRT amino acids predicted to be proximal to the template base: Phe 66 and Ala 78 (Fig. 5b). The size and hydrophobicity of these positions were altered through F66V, F66S, A78V, and A78R substitutions, as well as, F66A/A78V and F66S/A78R double substitutions. None of these resulted in a significant increase in fidelity, and indeed, fidelity substantially decreased for bRT(A78I)-Avd and somewhat for bRT(A78R)-Avd (Fig. 5d).

DISCUSSION

Fidelity in the transmission of genetic information is crucial for the fixation of traits that confer fitness, while infidelity is required for adaptation through variation. Extremes in either are insupportable, resulting in either no variation or loss of function, respectively. DGRs appear to have evolved a balance between fixed and variable traits through selective fidelity. Variability occurs only in adenine-encoded amino acids, while non-adenine-encoded amino acids remain conserved. As seen in a number of DGR variable proteins (9,22,46-48), adenine-encoded amino acids are organized by the C-type lectin-fold of the variable protein into a solvent-exposed binding site. Non-adenine-encoded

amino acids form the invariable structural scaffolding for the variable binding site. AAY (Y = pyrimidine) codons are especially prevalent in DGR variable proteins, and as previously noted, adenine-mutagenesis of AAY codons captures the gamut of amino acid chemistry but precludes a stop codon (22). Adding a layer of complexity to adenine-mutagenesis is the distribution of misincorporation frequencies documented here. For example, this makes some amino acid positions more variable (e.g., Mtd 357) and some less (e.g., Mtd 369), and thereby shapes the repertoire of ligands functionally bound by DGR variable proteins. These positional effects on misincorporation frequencies are also seen in cDNAs synthesized *in vivo* (13) and *mtd* sequences that have undergone variation (in the absence of selection) (1). This is most noticeable for *TR* A62 (encoding Mtd 357), which consistently has a high misincorporation frequency (~70%). No obvious relationship was evident between positional variation in misincorporation frequency and primary sequence. It is possible that the secondary or even tertiary structure of the template plays a role in this.

Extensive works has shown that fidelity in nucleotide polymerases depends both on Watson-Crick hydrogen bonding and shape complementarity between base pairs (49-55). Indeed, shape complementarity appears to be the dominant discriminator (56), as supported by several lines of evidence, including the observation that hydrogen bonding is dispensable for fidelity in DNA polymerase I (57). The general consensus is that pairing between the template and incoming base is sterically evaluated by polymerases (31,51). If the pairing is correct, polymerases undergo an open to closed transition, which places catalytic groups in the right positions for chemistry to proceed. If the pairing is incorrect, the open to closed transition fails to occur, providing time enough for the incorrect dNTP to dissociate before chemistry can occur. A strong correlation has been noted between fidelity and catalytic efficiency (k_{cat}/K_m) for polymerases (16). High and low fidelity polymerases differ substantially in their catalytic efficiency for the correct pairing (up to 10^5-fold) but little in their catalytic efficiency for the incorrect pairing (only ~10^2-fold). That is, both sorts of polymerases are inefficient enzymes when faced with an incorrect base pair, but high fidelity polymerases become efficient given the correct base pair while low fidelity enzymes remain inefficient.

We found that bRT-Avd had a markedly low catalytic efficiency for the correct base pair across a template adenine, just as has been documented for low fidelity polymerases (16). The catalytic efficiency for an incorrect base pair was not much lower (only 18- to 49-fold). bRT-Avd most resembles members of the Y family of DNA polymerases, which are low fidelity polymerases responsible for replicating through DNA lesions (50). The Y family DNA polymerase ι misincorporates dGTP at a frequency of 0.72 across a template thymine (58-60), reminiscent of the overall 0.75 misincorporation frequency of bRT-Avd across certain template adenines, and has a catalytic efficiency for correct base pairs ranging between 10^{-1} to 10^{-4} $\mu M^{-1}min^{-1}$ (59), encompassing the 10^{-3} $\mu M^{-1}min^{-1}$ catalytic efficiency observed for bRT-Avd (Table 1). DNA Polι and other Y family DNA polymerases synthesize only short stretches of DNA to repair lesions (61). Indeed, DNA Polι tends to terminate synthesis after incorporating a dGTP across a template thymine (59). Likewise, bRT-Avd synthesizes only short cDNAs (5-35 nt) with non-DGR RNA templates. However, when the template is the DGR RNA with flanking functional elements, bRT-Avd becomes processive and synthesizes extended cDNAs (90- and 120-nt). This is likely due to the formation of a structured RNP by bRT-Avd

and the DGR RNA. The k_{cat} of bRT-Avd for single deoxynucleotide addition observed here was especially slow, consistent with the very slow rate (on the order of hours) of *in vitro* synthesis of extended cDNAs determined previously (14). It is possible that a low catalytic efficiency can be tolerated by DGRs, as the target for variation (e.g., the gene encoding the variable protein) exists in single copy number usually and thus requires only a single cDNA molecule to effect sequence variation. While the efficiency of cDNA synthesis by the *Bordetella* bacteriophage DGR *in vivo* requires further study, it is worth noting that the overall efficiency of sequence variation by this DGR is quite low (10^{-6}) (15).

To understand the nucleobase determinants that modulate selective fidelity, we took advantage of the near isosteric features of adenine and guanine. While a template guanine has a non-negligible misincorporation frequency of 1.6%, this is still 33-fold lower than that of a template adenine. Using nucleobase analogs that have adenine- or guanine-like groups, we found that the substituent at the C6 position was a major determinant of fidelity for purines. An amine at the C6 position, as in adenine, was a neutral determinant of fidelity and appeared to be effectively invisible to mismatch detection mechanisms, while a bulkier methylamine at the same position was a negative determinant and appeared to disengage these mechanisms. In contrast, a carbonyl at C6, as in guanine, was a positive determinant of fidelity and appeared to significantly engage mismatch detection mechanisms. In the case of hypoxanthine, which has a carbonyl at C6 but no substituent at C2, the misincorporation frequency was reduced from the ~50% of adenine to ~4%. The further 2% decrease to bring misincorporation to the ~2% level of guanine appears to be due to the C2 amine of guanine. In addition, we found that the 2'-OH aids in mismatch detection. This indicates second strand cDNA synthesis would lead to further mutagenesis at positions corresponding to thymines in the RNA template, and agrees with the finding that only first strand cDNA synthesis could be detected *in vivo* (13).

We also explored the possibility that adenine flips out of the catalytic site, leaving it empty. This has been suggested for DNA Polι, in which incorporation is more efficient across an abasic compared to a pyrimidine template site (59,60). However, an abasic site resulted in a substantial level of deletions for bRT-Avd, and the misincorporation pattern at the abasic site was not the same as with adenine. While both abasic and adenine sites led predominantly to adenine misincorporation, an abasic site led to preferential misincorporation of guanine over cytosine and an adenine site had this preference flipped. Thus, the adenine-mutagenesis pattern is not explained by adenine flipping out of the catalytic site. In addition, these results indicate that the catalytic site bRT-Avd is not as predisposed towards an incoming adenine as has been suggested for A-rule polymerases (31).

We sought to identify amino acids in bRT that have a role in modulating adenine-mutagenesis, and relied on a three-dimensional *in silico* model based primarily on group II intron maturases (34-36). We probed eight amino acids predicted to be located at or near the catalytic site, and identified two that modulated fidelity: R74 and I181. Based on the *in silico* model of bRT, Arg 74 corresponds to Arg 72 of HIV RT. This HIV RT amino acid contacts the base and phosphate of the incoming dNTP (37). Substitution of HIV RT Arg 72 with Ala leads to a significant increase in fidelity (about three-fold on average but up to 25-fold at specific sites) with a significant decrease in k_{cat} (30- to 100-fold) (38,39).

Similarly, substitution of bRT R74 with Ala led to an increase in fidelity and a decrease in k_{cat}. However, these effects were much more modest in bRT-Avd (1.8- and 1.7-fold, respectively). The second amino acid, Ile 181 of bRT, corresponds to HIV RT Q151. This amino acid in HIV RT contacts the ribose of the incoming dNTP (37). Substitution of HIV RT Gln 151 with Asn increases fidelity by 8- to 27-fold (41), and decreases affinity for the correct incoming dNTP by 120-fold and for incorrect ones to levels that are not measurable (42). A similar but smaller 1.5-fold increase in fidelity and a 27-fold increase in K_m for TTP was seen for bRT I181N. In HIV RT, Arg 74 and Gln 151 provide contacts that stabilize both correct and incorrect incoming dNTPs. In the absence of these nonspecific contacts, the contributions of correct hydrogen bonding and shape complementarity become more consequential and thereby favor fidelity. As indicated by the similar effects on K_m and k_{cat} in bRT and HIV RT, bRT Arg 74 and Ile 181 are likely to provide nonspecific stabilization of base pairs as well.

In summary, the weight of evidence suggests that selective fidelity in bRT-Avd is due to its markedly low catalytic efficiency, as has been observed for other highly error-prone polymerases. While such polymerases are typically capable of synthesizing only short stretches of DNA, bRT-Avd with the aid of functional elements in the DGR RNA is capable of synthesizing extended cDNAs that effect protein variation. Our results provide the first detailed characterization of the determinants of selective fidelity in DGRs.

SUPPLEMENTARY DATA

Supplementary Figures S1-S4.
Supplementary Table S1.

FUNDING

This work was supported by the National Institutes of Health [R01 GM132720 and P.G.].

CONFLICT OF INTEREST

No conflicts to declare.

REFERENCES

1. Liu, M., Deora, R., Doulatov, S.R., Gingery, M., Eiserling, F.A., Preston, A., Maskell, D.J., Simons, R.W., Cotter, P.A., Parkhill, J. *et al.* (2002) Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. *Science*, **295**, 2091-2094.
2. Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D. and Bushman, F.D. (2013) Rapid evolution of the human gut virome. *Proc Natl Acad Sci U S A*, **110**, 12450-12455.
3. Ye, Y. (2014) Identification of diversity-generating retroelements in human microbiomes. *Int J Mol Sci*, **15**, 14234-14246.

4. Paul, B.G., Bagby, S.C., Czornyj, E., Arambula, D., Handa, S., Sczyrba, A., Ghosh, P., Miller, J.F. and Valentine, D.L. (2015) Targeted diversity generation by intraterrestrial archaea and archaeal viruses. *Nat Commun*, **6**, 6585.
5. Paul, B.G., Burstein, D., Castelle, C.J., Handa, S., Arambula, D., Czornyj, E., Thomas, B.C., Ghosh, P., Miller, J.F., Banfield, J.F. *et al.* (2017) Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. *Nat Microbiol*, **2**, 17045.
6. Benler, S., Cobian-Guemes, A.G., McNair, K., Hung, S.H., Levi, K., Edwards, R. and Rohwer, F. (2018) A diversity-generating retroelement encoded by a globally ubiquitous Bacteroides phage. *Microbiome*, **6**, 191.
7. Yan, F., Yu, X., Duan, Z., Lu, J., Jia, B., Qiao, Y., Sun, C. and Wei, C. (2019) Discovery and characterization of the evolution, variation and functions of diversity-generating retroelements using thousands of genomes and metagenomes. *BMC Genomics*, **20**, 595.
8. Boehm, T., McCurley, N., Sutoh, Y., Schorpp, M., Kasahara, M. and Cooper, M.D. (2012) VLR-based adaptive immunity. *Annu Rev Immunol*, **30**, 203-220.
9. Le Coq, J. and Ghosh, P. (2011) Conservation of the C-type lectin fold for massive sequence variation in a Treponema diversity-generating retroelement. *Proc Natl Acad Sci U S A*, **108**, 14649-14653.
10. Wu, L., Gingery, M., Abebe, M., Arambula, D., Czornyj, E., Handa, S., Khan, H., Liu, M., Pohlschroder, M., Shaw, K.L. *et al.* (2018) Diversity-generating retroelements: natural variation, classification and evolution inferred from a large-scale genomic survey. *Nucleic Acids Res*, **46**, 11-24.
11. Melvin, J.A., Scheller, E.V., Miller, J.F. and Cotter, P.A. (2014) Bordetella pertussis pathogenesis: current and future challenges. *Nat Rev Microbiol*, **12**, 274-288.
12. Guo, H., Arambula, D., Ghosh, P. and Miller, J.F. (2014) Diversity-generating Retroelements in Phage and Bacterial Genomes. *Microbiol Spectr*, **2**.
13. Naorem, S.S., Han, J., Wang, S., Lee, W.R., Heng, X., Miller, J.F. and Guo, H. (2017) DGR mutagenic transposition occurs via hypermutagenic reverse transcription primed by nicked template RNA. *Proc Natl Acad Sci U S A*, **114**, E10187-E10195.
14. Handa, S., Jiang, Y., Tao, S., Foreman, R., Schinazi, R.F., Miller, J.F. and Ghosh, P. (2018) Template-assisted synthesis of adenine-mutagenized cDNA by a retroelement protein complex. *Nucleic Acids Res*, **46**, 9711-9725.
15. Alayyoubi, M., Guo, H., Dey, S., Golnazarian, T., Brooks, G.A., Rong, A., Miller, J.F. and Ghosh, P. (2013) Structure of the essential diversity-generating retroelement protein bAvd and its functionally important interaction with reverse transcriptase. *Structure*, **21**, 266-276.
16. Beard, W.A., Shock, D.D., Vande Berg, B.J. and Wilson, S.H. (2002) Efficiency of correct nucleotide insertion governs DNA polymerase fidelity. *J Biol Chem*, **277**, 47393-47398.
17. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-U354.

18. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Proc, G.P.D. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
19. Miller, J.R., Koren, S. and Sutton, G. (2010) Assembly algorithms for next-generation sequencing data. *Genomics*, **95**, 315-327.
20. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat Biotechnol*, **29**, 24-26.
21. Sebastian-Martin, A., Barrioluengo, V. and Menendez-Arias, L. (2018) Transcriptional inaccuracy threshold attenuates differences in RNA-dependent DNA synthesis fidelity between retroviral reverse transcriptases. *Sci Rep*, **8**, 627.
22. McMahon, S.A., Miller, J.L., Lawton, J.A., Kerkow, D.E., Hodes, A., Marti-Renom, M.A., Doulatov, S., Narayanan, E., Sali, A., Miller, J.F. *et al.* (2005) The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nat Struct Mol Biol*, **12**, 886-892.
23. Case-Green, S.C. and Southern, E.M. (1994) Studies on the base pairing properties of deoxyinosine by solid phase hybridisation to oligonucleotides. *Nucleic Acids Res*, **22**, 131-136.
24. Martin, F.H., Castro, M.M., Aboulela, F. and Tinoco, I. (1985) Base-Pairing Involving Deoxyinosine - Implications for Probe Design. *Nucleic Acids Research*, **13**, 8927-8938.
25. Rahman, M.S. and Humayun, M.Z. (1997) Nebularine (9-2'-deoxy-beta-D-ribofuranosylpurine) has the template characteristics of adenine in vivo and in vitro. *Mutat Res*, **377**, 263-268.
26. Harcourt, E.M., Ehrenschwender, T., Batista, P.J., Chang, H.Y. and Kool, E.T. (2013) Identification of a Selective Polymerase Enables Detection of N-6-Methyladenosine in RNA. *Journal of the American Chemical Society*, **135**, 19079-19082.
27. Cheong, C., Tinoco, I. and Chollet, A. (1988) Thermodynamic Studies of Base-Pairing Involving 2,6-Diaminopurine. *Nucleic Acids Research*, **16**, 5115-5122.
28. Patro, J.N., Urban, M. and Kuchta, R.D. (2009) Role of the 2-Amino Group of Purines during dNTP Polymerization by Human DNA Polymerase α. *Biochemistry*, **48**, 180-189.
29. Watanabe, S.M. and Goodman, M.F. (1982) Kinetic Measurement of 2-Aminopurine.Cytosine and 2-Aminopurine.Thymine Base-Pairs as a Test of DNA-Polymerase Fidelity Mechanisms. *P Natl Acad Sci-Biol*, **79**, 6429-6433.
30. Reha-Krantz, L.J., Hariharan, C., Subuddhi, U., Xia, S.L., Zhao, C., Beckman, J., Christian, T. and Konigsberg, W. (2011) Structure of the 2-Aminopurine-Cytosine Base Pair Formed in the Polymerase Active Site of the RB69 Y567A-DNA Polymerase. *Biochemistry*, **50**, 10136-10149.
31. Strauss, B.S. (2002) The "A" rule revisited: polymerases as determinants of mutational specificity. *DNA Repair (Amst)*, **1**, 125-135.
32. Kupfer, P.A., Crey-Desbiolles, C. and Leumann, C.J. (2007) Trans-lesion synthesis and RNaseH activity by reverse transcriptases on a true abasic RNA template. *Nucleic Acids Res*, **35**, 6846-6853.

33. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, **10**, 845-858.
34. Zhao, C. and Pyle, A.M. (2016) Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nat Struct Mol Biol*, **23**, 558-565.
35. Qu, G., Kaushal, P.S., Wang, J., Shigematsu, H., Piazza, C.L., Agrawal, R.K., Belfort, M. and Wang, H.W. (2016) Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol*, **23**, 549-557.
36. Stamos, J.L., Lentzsch, A.M. and Lambowitz, A.M. (2017) Structure of a Thermostable Group II Intron Reverse Transcriptase with Template-Primer and Its Functional and Evolutionary Implications. *Mol Cell*, **68**, 926-939 e924.
37. Huang, H., Chopra, R., Verdine, G.L. and Harrison, S.C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669-1675.
38. Sarafianos, S.G., Pandey, V.N., Kaushik, N. and Modak, M.J. (1995) Site-directed mutagenesis of arginine 72 of HIV-1 reverse transcriptase. Catalytic role and inhibitor sensitivity. *J Biol Chem*, **270**, 19729-19735.
39. Lewis, D.A., Bebenek, K., Beard, W.A., Wilson, S.H. and Kunkel, T.A. (1999) Uniquely altered DNA replication fidelity conferred by an amino acid change in the nucleotide binding pocket of human immunodeficiency virus type 1 reverse transcriptase. *J Biol Chem*, **274**, 32924-32930.
40. Weiss, K.K., Isaacs, S.J., Tran, N.H., Adman, E.T. and Kim, B. (2000) Molecular architecture of the mutagenic active site of human immunodeficiency virus type 1 reverse transcriptase: roles of the beta 8-alpha E loop in fidelity, processivity, and substrate interactions. *Biochemistry*, **39**, 10684-10694.
41. Kaushik, N., Talele, T.T., Pandey, P.K., Harris, D., Yadav, P.N. and Pandey, V.N. (2000) Role of glutamine 151 of human immunodeficiency virus type-1 reverse transcriptase in substrate selection as assessed by site-directed mutagenesis. *Biochemistry*, **39**, 2912-2920.
42. Weiss, K.K., Bambara, R.A. and Kim, B. (2002) Mechanistic role of residue Gln151 in error prone DNA synthesis by human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT). Pre-steady state kinetic study of the Q151N HIV-1 RT mutant with increased fidelity. *J Biol Chem*, **277**, 22662-22669.
43. Jonckheere, H., De Clercq, E. and Anne, J. (2000) Fidelity analysis of HIV-1 reverse transcriptase mutants with an altered amino-acid sequence at residues Leu74, Glu89, Tyr115, Tyr183 and Met184. *Eur J Biochem*, **267**, 2658-2665.
44. Pandey, V.N., Kaushik, N., Rege, N., Sarafianos, S.G., Yadav, P.N.S. and Modak, M.J. (1996) Role of methionine 184 of human immunodeficiency virus type-1 reverse transcriptase in the polymerase function and fidelity of DNA synthesis. *Biochemistry*, **35**, 2168-2179.
45. Wainberg, M.A., Drosopoulos, W.C., Salomon, H., Hsu, M., Borkow, G., Parniak, M.A., Gu, Z.X., Song, Q.B., Manne, J., Islam, S. *et al.* (1996) Enhanced fidelity of 3TC-selected mutant HIV-1 reverse transcriptase. *Science*, **271**, 1282-1285.

46. Miller, J.L., Coq, J.L., Hodes, A., Barbalat, R., Miller, J.F. and Ghosh, P. (2008) Selective Ligand Recognition by a Diversity-Generating Retroelement Variable Protein. *PLoS Biol*, **6**, e131.
47. Handa, S., Paul, B.G., Miller, J.F., Valentine, D.L. and Ghosh, P. (2016) Conservation of the C-type lectin fold for accommodating massive sequence variation in archaeal diversity-generating retroelements. *BMC Struct Biol*, **16**, 13.
48. Handa, S., Shaw, K.L. and Ghosh, P. (2019) Crystal structure of a Thermus aquaticus diversity-generating retroelement variable protein. *PLoS One*, **14**, e0205618.
49. Beard, W.A. and Wilson, S.H. (2003) Structural insights into the origins of DNA polymerase fidelity. *Structure*, **11**, 489-496.
50. Kunkel, T.A. (2009) Evolving views of DNA replication (in)fidelity. *Cold Spring Harb Symp Quant Biol*, **74**, 91-101.
51. Freudenthal, B.D., Beard, W.A. and Wilson, S.H. (2015) New structural snapshots provide molecular insights into the mechanism of high fidelity DNA synthesis. *DNA Repair (Amst)*, **32**, 3-9.
52. Echols, H. and Goodman, M.F. (1991) Fidelity mechanisms in DNA replication. *Annu Rev Biochem*, **60**, 477-511.
53. Kim, T.W., Delaney, J.C., Essigmann, J.M. and Kool, E.T. (2005) Probing the active site tightness of DNA polymerase in subangstrom increments. *Proc Natl Acad Sci U S A*, **102**, 15803-15808.
54. Sydow, J.F. and Cramer, P. (2009) RNA polymerase fidelity and transcriptional proofreading. *Curr Opin Struct Biol*, **19**, 732-739.
55. Menendez-Arias, L. (2002) Molecular basis of fidelity of DNA synthesis and nucleotide specificity of retroviral reverse transcriptases. *Prog Nucleic Acid Res Mol Biol*, **71**, 91-147.
56. Kool, E.T. (2001) Hydrogen bonding, base stacking, and steric effects in dna replication. *Annu Rev Biophys Biomol Struct*, **30**, 1-22.
57. Moran, S., Ren, R.X. and Kool, E.T. (1997) A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. *Proc Natl Acad Sci U S A*, **94**, 10506-10511.
58. Bebenek, K., Tissier, A., Frank, E.G., McDonald, J.P., Prasad, R., Wilson, S.H., Woodgate, R. and Kunkel, T.A. (2001) 5'-Deoxyribose phosphate lyase activity of human DNA polymerase iota in vitro. *Science*, **291**, 2156-2159.
59. Zhang, Y., Yuan, F., Wu, X. and Wang, Z. (2000) Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase iota. *Mol Cell Biol*, **20**, 7099-7108.
60. Johnson, R.E., Washington, M.T., Haracska, L., Prakash, S. and Prakash, L. (2000) Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. *Nature*, **406**, 1015-1019.
61. McCulloch, S.D. and Kunkel, T.A. (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res*, **18**, 148-161.

Table 1. Steady-state enzymatic parameters for (mis)incorporation by bRT-Avd

WT	K_m (μM)	k_{cat} (min^{-1}) x10^{-3}	k_{cat}/K_m (μM^{-1} min^{-1}) x 10^{-3}	Misincorporation ratio[b]
A:T[a]	7.97 ± 2.34	11.64 ± 0.71	1.46	-
A:dG	138.03 ± 48.33	3.62 ± 0.32	0.03	0.02
A:dC	84.35 ± 26.10	6.85 ± 0.56	0.08	0.05
A:dA	115.25 ± 22.31	9.33 ± 0.46	0.08	0.05
A:dU	1.73 ± 0.64	11.20 ± 0.96	6.47	4.43

[a]Template base:(mis)incorporated base
[b]Misincorporation ratio = [k_{cat} (incorrect)/K_m (incorrect)] / [k_{cat} (correct)/K_m (correct)]

Table 2. Steady-state enzymatic parameters for TTP incorporation by mutant bRT-Avd

bRT mutant	K_m (μM)	k_{cat} (min^{-1}) x10^{-3}	k_{cat}/K_m (μM^{-1} min^{-1}) x 10^{-3}	Efficiency[c]
R74A	7.09 ± 4.68	7.03 ± 0.78	0.99	0.68
I181N	218.85 ± 55.94	11.13 ± 1.04	0.05	0.03

[c]Efficiency = [k_{cat} (mutant)/K_m (mutant)] / [k_{cat} (wt)/K_m (wt)]

Figure 1

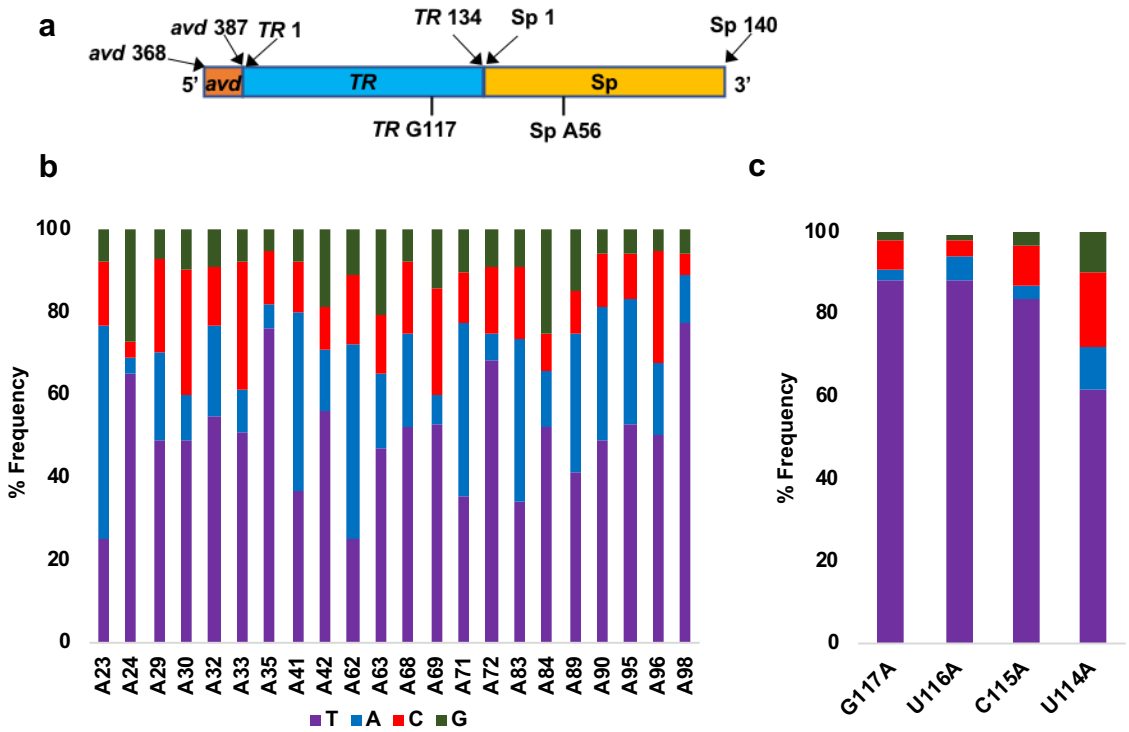

Figure 1. Adenine-specific mutagenesis.
a. *Bordetella* bacteriophage DGR core RNA: segment of *avd* (brown), *TR* (blue), and Sp (yellow). Sp A56 acts as the priming site for reverse transcription, and the first nucleotide reverse transcribed is *TR* G117.
b. Frequency of deoxynucleotides incorporated (thymine, purple) or misincorporated (adenine, blue; cytosine, red; guanine, green) across template adenines in *TR*. This color coding is used throughout.
c. Frequency of deoxynucleotides (mis)incorporated across template adenines individually substituted at *TR* 114-117.

Figure 2

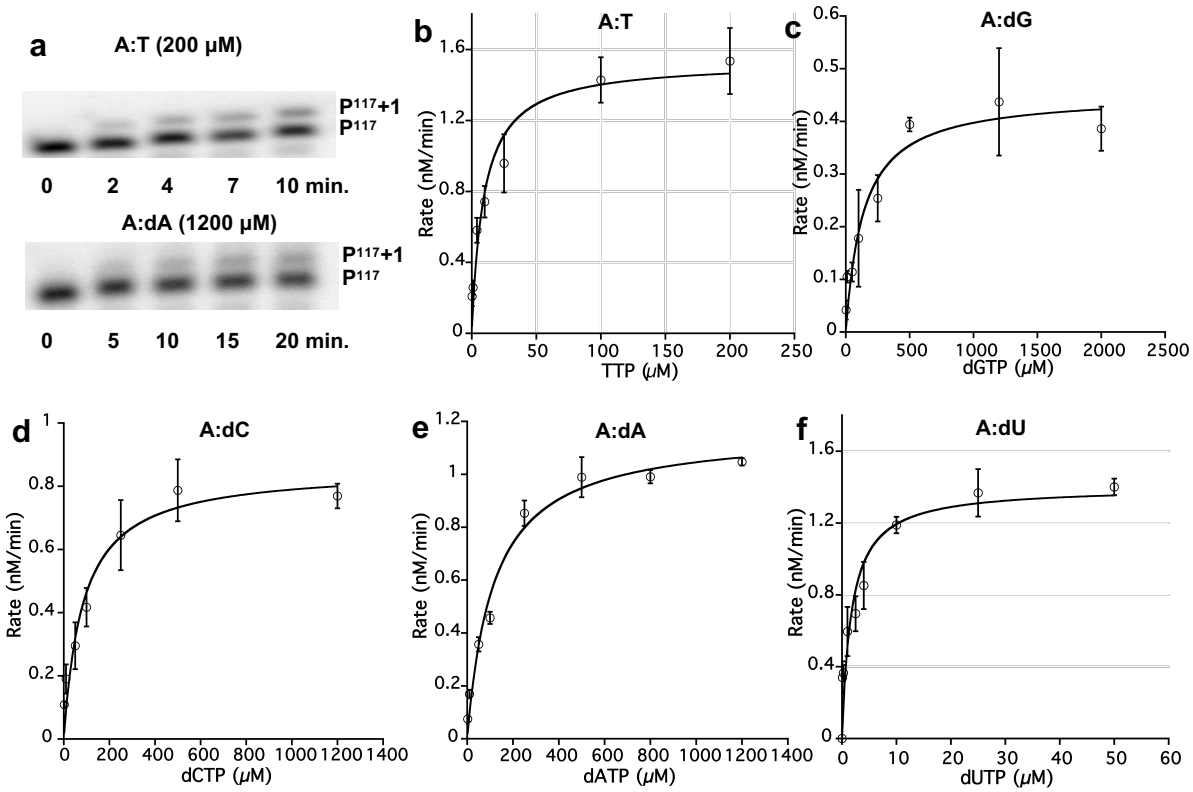

Figure 2. Kinetics of single deoxynucleotide (mis)incorporation.
a. Single deoxynucleotide primer extension by bRT-Avd of [^{32}P]-labeled P^{117} by TTP (top) or dATP (bottom), as templated by the core DGR RNA containing *TR* G117A. The extended product (P^{117} +1) was resolved from the reactant (P^{117}) with an 8% sequencing gel.
b-f. Steady-state kinetic characterization of single deoxynucleotide (mis)incorporation of TTP, dGTP, dCTP, dATP and dUTP, across the core DGR RNA containing *TR* G117A (and *TR* U116G, in the case of dATP) by bRT-Avd. Error bars represent standard deviations from three independent measurements.

Figure 3

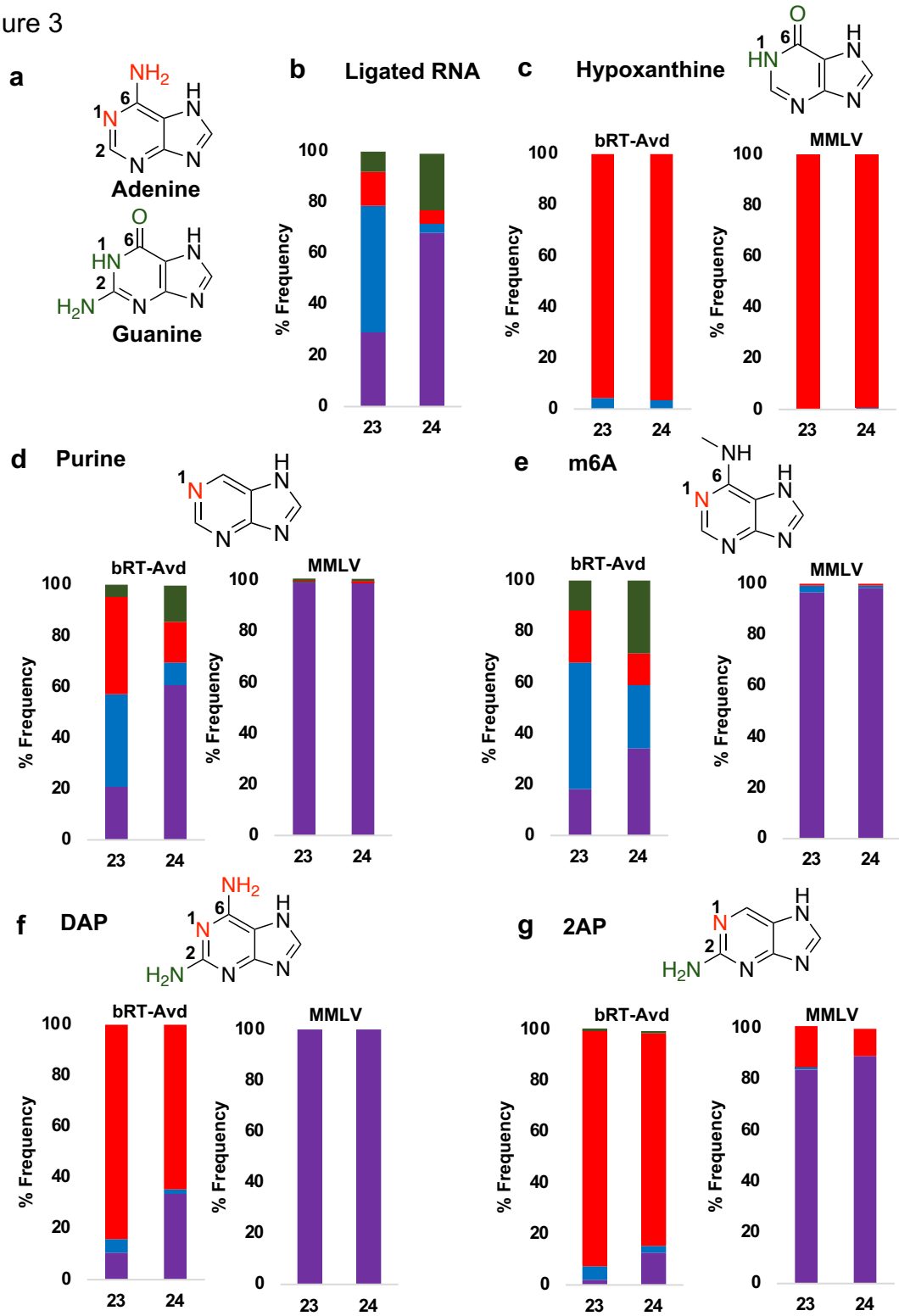

Figure 3. Nucleobase determinants of fidelity.
a. Adenine and guanine differ at N1, C2, and C6 positions (red for A, green for G).
b. (Mis)incorporation frequencies at *TR* A23 and *TR* A24 using the core DGR RNA template that had been ligated from chemically synthesized and *in vitro* transcribed sections.

Figure 3 continued. Nucleobase determinants of fidelity.
c. Top, structure of hypoxanthine. Bottom, (mis)incorporation frequencies of bRT-Avd (left) and MMLV RT (right) with hypoxanthine at *TR* 23 and 24.
d. Top, structure of purine. Bottom, (mis)incorporation frequencies of bRT-Avd (left) and MMLV RT (right) with purine at *TR* 23 and 24.
e. Top, structure of N6-methyladenine (m6A). Bottom, (mis)incorporation frequencies of bRT-Avd (left) and MMLV RT (right) with m6A at *TR* 23 and 24.
f. Top, structure of DAP. Bottom, (mis)incorporation frequencies of bRT-Avd (left) and MMLV RT (right) with DAP at *TR* 23 and 24.
g. Top, structure of 2,6-diaminopurine (DAP). Bottom, (mis)incorporation frequencies of bRT-Avd (left) and MMLV RT (right) with DAP at *TR* 23 and 24.

Figure 4

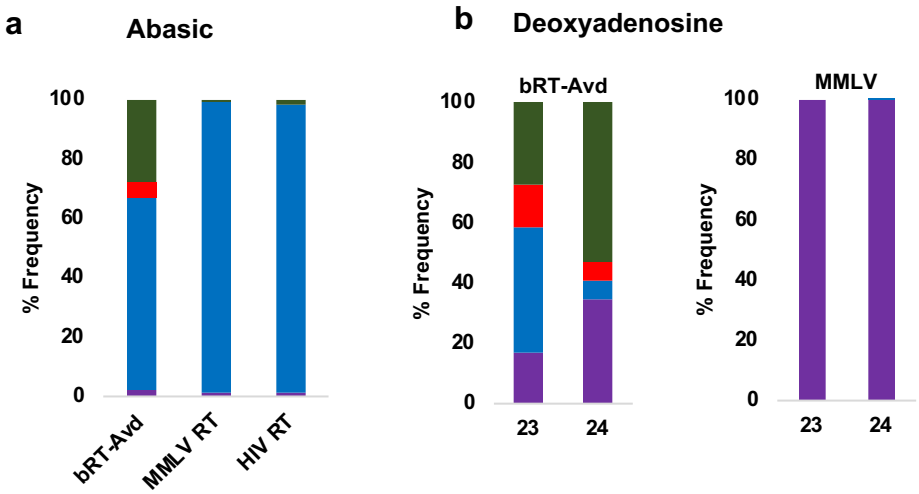

Figure 4. Abasic and Deoxy.
a. (Mis)incorporation frequencies with an abasic site at *TR* 23 for bRT-Avd, MMLV RT, and HIV RT.
b. (Mis)incorporation frequencies of bRT-Avd (left) and MMLV RT (right) with deoxyadenosines at *TR* 23 and 24.

Figure 5

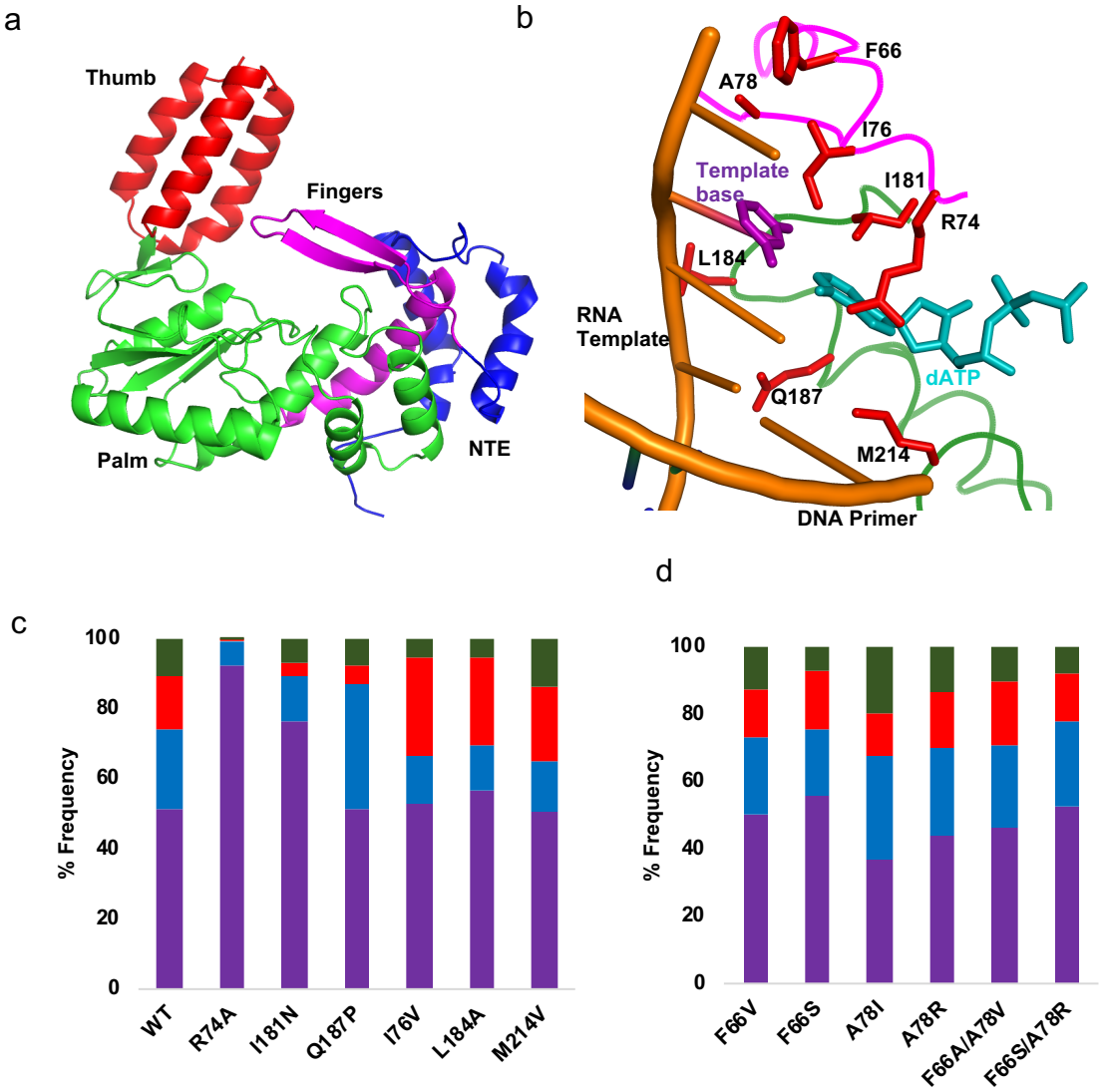

Figure 5. bRT amino acids that modulate selective fidelity.
a. *In silico* model of bRT in cartoon representation, showing the N-terminal extension (NTE, blue), Fingers (magenta), Palm (green), and Thumb (red) subdomains.
b. *In silico* model of bRT with amino acids subjected to mutagenesis shown as red bonds. The main chain is shown as a coil, colored by subdomains as in panel a. For reference, the RNA template-DNA primer heteroduplex (gold) and incoming dATP (cyan) from the structure of the GsI-IIc group II intron maturase is shown. The template base (thymine) is purple.
c. Average (mis)incorporation frequencies across the 22 *TR* adenines for wild-type bRT and bRT containing substitutions at amino acids implicated in modulating selective fidelity.
d. Average (mis)incorporation frequencies across the 22 *TR* adenines for wild-type bRT and bRT containing substitutions at amino acids predicted to be proximal to the templating base.