Published in final edited form as:

J Mol Biol. 2016 June 5; 428(11): 2275–2288. doi:10.1016/j.jmb.2016.04.005.

Resolution of specific nucleotide mismatches by wild type and AZT-resistant reverse transcriptases during HIV-1 replication

Siarhei Kharytonchyk, Steven R. King, Clement B. Ndongmo¹, Krista L. Stilger, Wenfeng An², and Alice Telesnitsky*

Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109-5620

Abstract

A key contributor to HIV-1 genetic variation is reverse transcriptase errors. Some mutations result because reverse transcriptase (RT) lacks 3' to 5' proofreading exonuclease and can extend mismatches. However, RT also excises terminal nucleotides to a limited extent, and this activity contributes to AZT resistance. Because HIV-1 mismatch resolution has been studied *in vitro* but only indirectly during replication, we developed a novel system to study mismatched basepair resolution during HIV-1 replication in cultured cells, using vectors that force template switching at defined locations. These vectors generated mismatched reverse transcription intermediates, with proviral products diagnostic of mismatch resolution mechanisms. Outcomes for wild-type (WT) RT and an AZT-resistant (AZT^R) RT containing a thymidine analog mutation set --D67N, K70R, D215F, K219Q—were compared. AZT^R RT did not excise terminal nucleotides more frequently than WT, and for the majority of tested mismatches, both WT and AZT^R RTs extended mismatches in more than 90% of proviruses. However, striking enzyme-specific differences were observed for one mispair, with WT RT preferentially resolving dC-rC pairs either by excising the mismatched base or switching templates prematurely, while AZT^R RT primarily misaligned the primer strand, causing deletions via dislocation mutagenesis. Overall, the results confirmed HIV-1 RT's high capacity for mismatch extension during virus replication, and revealed dramatic differences in aberrant intermediate resolution repertoires between WT and AZT^R RTs on one mismatched replication intermediate. Correlating mismatch extension frequencies observed here with reported viral mutation rates suggests a complex interplay of nucleotide discrimination and mismatch extension drives HIV-1 mutagenesis.

Graphical Abstract

*Corresponding author: 1150 West Medical Center Dr. Room 5641, Ann Arbor, MI 48109-5620, Electronic: ; Email: ateles@umich.edu, Telephone: 734 936-6466, Fax: 734 764-3562

¹CDC, Lusaka, Zambia

²Department of Harmaceutical Sciences, South Dakota State University. Brookings, SD 57007

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Retroviral error mechanisms; forced copy choice recombination

Introduction

HIV-1 is a genetically diverse virus that persists as a quasispecies within infected individuals [1]. Many HIV-1 mutations reside in contexts characteristic of targets of host cytidine deaminases and likely result from host antiviral activities [2] [3]. The possibility that host RNA polymerases also contribute to retroviral errors cannot be ruled out [4] [5]. It is nonetheless clear that a significant source of genetic variation in HIV-1 populations is mutations introduced by reverse transcriptase (RT) during processive reverse transcription or upon template switching. Current consensus suggests errors in HIV-1 genomes arise around 1 time per 50,000 bases synthesized [6]. Template switching is remarkably frequent during retroviral DNA synthesis, with RT switching from one RNA template to homologous sequences on the co-packaged viral RNA roughly ten times during the synthesis of every proviral DNA [7–9].

Error rates based on *in vitro* experiments suggest that HIV-1 RT introduces approximately $2\text{--}5 \times 10^{-4}$ mutations per base pair, with significantly (perhaps 10-fold) lower rates observed during replication in cultured cells [10–13]. Retroviral mutagenesis generally involves base misinsertion followed by mismatch extension, and RT's error rates are much higher than those of the cellular replication machinery. Although it lacks 3' to 5' proofreading exonuclease activity and frequently extends mispairs without correction, HIV-1 RT is capable of excising incorporated bases to some extent by reversing the chemistry of polymerization in the presence of pyrophosphate or ATP [14, 15]. This reaction contributes to AZT resistance, and *in vitro* reverse transcription results suggest it contributes to replication fidelity as well [16].

It is now well established that some variant RTs differ from the wild type enzyme in nucleoside analog discrimination [17]. However, when RT mutations associated with AZT resistance in patients were first described, studies with purified enzymes failed to identify differences between RTs from AZT-sensitive and –resistant viruses, and the mechanism of HIV-1 RT's AZT resistance remained unexplained for several years [18]. It was eventually discovered that unlike many other nucleoside analog resistance-associated mutations, which act by affecting rates of nucleotide analog discrimination prior to incorporation [17], the rate of 3' terminal nucleotide excision is significantly increased for certain AZT-resistant (AZT^R) forms of RT [19] and that this enhanced level of primer unblocking contributes to HIV-1

resistance to AZT and some other nucleoside analogs. This history of differences between experimental conditions and intracellular replication masking mechanistic properties of RT underscores the need to understand RT error mechanisms in cells as well as in purified reactions.

RT is prone to mismatch insertion in purified reactions, with some forms of RT more prone to misinsertion than others [20, 21]. HIV-1 RT also corrects mismatches with a degree of selectivity: for example, G-T mismatches are rectified more frequently than C-T mismatches in purified reactions [22]. Although mutations that arise during virus replication are less well characterized mechanistically than those generated in purified reactions, differences in mutation frequency have been described for some drug resistant RT mutants during viral replication (eg: [23, 24]) as well as in purified reactions. While mismatch extension occurs more frequently than nucleotide excision in purified reverse transcription reactions and some reports find no contributions of NC to fidelity [25], the addition of viral nucleocapsid protein increased the efficiency of both WT and AZT^R RT *in vitro* base excision 10-fold in another study [16]. It is possible that these and additional parameters within cells may lead to outcomes during viral replication that differ from those reported for purified reactions. And indeed, a study of error hot spots observed during viral replication demonstrated that their pattern was different from those generated on the same template in reconstituted reactions *in vitro* [12]: prompting further investigation of additional differences between cells and standard *in vitro* reaction conditions that might explain differences between *in vitro* and intracellular reverse transcription outcomes [10].

Recombination, which results from template switching during reverse transcription, may also be a source of HIV-1 mutations. When RT reaches the end of a template, it can add non-templated nucleotides [26, 27]. Upon template switching *in vitro*, extension of nontemplated bases leads to mutations at the point of strand transfer in up to 30–50% of all reverse transcription products [28–30]. Recombination in cells appears to be much less error-prone [27, 28, 30, 31] but template switch-associated mutagenesis has been reported, with some reports suggesting that up to 20% of RT mutations may be associated with template switching [32–34]. However, the fact that recombinogenic template switching can occur at many if not all template positions in viral replication products complicates addressing whether or not mutations observed within crossover intervals arose upon template switching or by a different mechanism.

Although mismatch extension and excision by HIV-1 RT have been studied in purified reactions *in vitro*, there is currently no data directly addressing these processes in a cell-based system. In the current study, we describe a novel system for studying HIV-1 RT mismatch resolution during single rounds of replication in human cells. These studies use retroviral inside-out (RIO) vectors, which were designed to promote template switching at defined template positions during reverse transcription. The work here demonstrates that RIO vectors were packaged efficiently and able to successfully complete single rounds of reverse transcription and integration when mobilized by helpers harboring either WT or an AZT^R RT variant [19]. Additionally, forced copy choice recombination occurred as intended at the donor template's 5' terminus. Using this system, specific mismatches were introduced and three independent approaches were used to determine the mechanisms by which HIV-1

WT and AZT^R RTs resolved these. The results indicated that HIV-1 RT displays nucleotide-specific differences in mismatch extension during virus replication and that the spectra of mechanisms used for the resolution of at least one mismatch differs between WT and AZT^R RTs.

Results

Establishing vectors to monitor mismatch resolution during HIV-1 replication

In this study, the mechanisms of primer-terminal mismatch resolution during HIV-1 replication were examined using novel vectors designed to promote template switching at defined template positions in cultured cells. These vectors, which are called retroviral inside-out (RIO) vectors, contain virus-derived sequences in a circularly permuted order (Figure 1A). These vectors were designed to circumvent one challenge to studying mechanisms of HIV mutagenesis in infected cells. Specifically, it is not straightforward to assess the relative contributions of template switching, base misincorporation, and mismatch extension to retroviral mutagenesis *in vivo*, because mutations arise throughout viral genomes. In fact, during HIV-1 replication, recombinogenic template switching can occur at many if not all positions along the viral genome, and point mutations—which arise when errors are introduced but not corrected—are also widespread [7]. The reverse transcription of RIO vectors forces template switching—which can be engineered to require mismatch extension—at a unique, defined position.

The general structure of RIO vectors is a circular permutation of a one-LTR (long terminal repeat) circle form of the HIV-1 genome, like those that can result when replication is aborted prior to integration ([35]; Figure 1A). Like pWA18puro, the standard HIV-based vector plasmid from which it was derived, the RIO plasmid included all *cis*-acting elements required for RNA packaging, reverse transcription, and integration as well as a selectable marker gene conferring puromycin resistance. Although RIO vectors contain a single modified LTR embedded in other viral sequences, their reverse transcription leads to reconstitution of a standard two-LTR lentiviral vector (Figure 1B). In the prototype RIO vector used in this study, an HIV-1 LTR was mutated to remove promoter and polyadenylation signals from U3 and R, and an SV40 virus-derived polyadenylation signal was placed downstream of vector sequences. The native polypurine tract (ppt) for initiating plus-strand viral DNA synthesis was maintained upstream of the modified solo LTR.

Transcription of the prototype HIV-1 RIO vector plasmid was initiated at a G residue using the Rous Sarcoma Virus (RSV) promoter (Figure 1A). Viral promoters are unusual in starting at one or only a few sites, and we have confirmed that the RSV promoters in RIO vectors initiate at unique template positions ([22, 36, 37] and data not shown). Because RNA polymerase II transcription preferentially initiates with purine residues, all RIO vector RNAs contained either an A or G residue at their 5' ends. As in HIV replication, reverse transcription of RIO vector RNAs is initially primed by a tRNA bound to the primer binding site (pbs; Figure 1B). However, the initial RIO vector minus-strand reverse transcription product is “pre-jumped”, in that it resembles the product that arises after the first strong stop template switch during normal retroviral replication [38]. Thus, the initial RIO vector minus-strand DNA product continues through the modified LTR until reaching the vector's 5' end,

from which position continued DNA synthesis requires a forced recombinogenic template switch.

To promote forced copy choice recombination at a defined site, identical 30 nucleotide donor (D) and acceptor (A) sequences were included at the 5' end and repeated near the 3' end of each RIO vector RNA. RIO minus strand DNA synthesis resumes after a forced template switch of the initial minus-strand product from (D) sequences at the 5' end of the RIO to (A) sequences near its 3' end (Figure 1B). At some time point during minus-strand DNA synthesis, plus-strand DNA synthesis initiates and proceeds, resulting in a linear viral DNA sequence similar to that produced by the standard HIV-1 based WA18puro vector (Figure 1A & B). Integration of RIO DNA into the cellular genome forms a provirus and confers puromycin resistance onto infected cells.

A series of RIO variants with altered donors and/or acceptors in pairwise combinations was constructed (Figure 2). As described above, RNA donors were engineered so that the 5' terminal nucleotide of each was either an A or a G (Figure 2A). Acceptor regions were engineered so that their 5' terminal nucleotides were G, A, U, or C. Except for these 5' terminal residues, all nucleotides within donor and acceptor sequences were identical, whereas the sequences flanking donors and acceptors were unique to allow the tracking of products' origins. The intention of this design was that when RIO vectors underwent forced copy choice recombination, those for which donor and acceptor shared the same 5' residue would generate replication intermediates with donor primer strand termini complementary to the acceptor (eg: in the case of dC:rG and dT:rA RIO vectors), while those for which donor and acceptor 5' ends differed would generate mismatched replication intermediates (for dC-rC, dC-rA, dC-rU, dT-rC, dt-rG, and dT-rU variants; Figure 2B). Modest sequence variation was introduced into RIO vectors immediately upstream of the recombinogenic acceptor template regions, to allow assessment of the mechanisms of mismatch resolution by restriction digestion. As a result, one restriction site would appear in proviral products if mismatch extension occurred faithfully and accurately, while a second restriction site would be generated in proviruses if template switching occurred earlier in the region of donor-acceptor homology, or if RT excised the mismatched donor terminal base and replaced it with a residue complementary to the acceptor template region (Figure 2C). Because forced copy choice recombination *in vitro* is often preceded by non-templated addition to the growing DNA prior to template switch, and non-templated addition is often followed by mismatch extension, +1 mutations were also a predicted outcome (Figure 2C). An alternate but unexpected outcome that was observed in some reverse transcription products in this study—deletions arising from dislocation mutagenesis—also are shown in Figure 2C.

Packaging of HIV RIO vector RNAs

In initial experiments designed to test replication step efficiency, RIO vector RNA expression in cells, and vector RNA packaging into virions, were assessed. 293T cells were co-transfected with helper function plasmids plus either a RIO vector expression plasmid or one encoding the conventional HIV-1 vector, pWA18puro. Virion and cellular RNA were harvested and analyzed by RNase protection assay (Figure 3A). Vector RNAs were detected using chimeric riboprobes complementary both to a puromycin resistance gene fragment

present in all vectors and to a region of 7SL, a small host cell RNA that is packaged by retroviruses such as HIV-1 in proportion to Gag, and thus can serve to normalize virion quantities [39].

As shown in Figure 3A, packaging of RIO vector RNAs was indistinguishable from that of conventional vector RNAs under the conditions used here (Figure 3A lanes 1, 2, and 3), with conventional vector WA18puro RNA slightly more abundant (<2-fold) in transfected cells than RIO vector RNA (Figure 3A lanes 6, 7 and 8). These results indicated that RIO vector RNAs were encapsidated well, despite the unnatural ordering of their genetic elements, including more central placement of packaging sequences than is associated with native HIV genomic RNAs. Interestingly, although RIO vectors possess only limited portions of the HIV-1 genome, attempts to reduce RIO vector size by removal of Rev Response Element (RRE) sequences severely diminished vector RNA packaging (data not shown, and see Materials and Methods). Thus, RIO vector packaging requires the RRE, as has previously been reported for other lentiviral vectors [40, 41]. All subsequent experiments were performed with RRE-containing RIO vectors.

HIV RIO vector replication

The ability of HIV RIO vectors to complete single rounds of replication was examined by determining puromycin resistant provirus titers. As shown in Figure 3B, puromycin resistance titers for RIO vectors with identical donor and acceptor sequences were approximately 50- to 100-fold less than those for pWA18puro, despite similar levels of packaging. These RIO vectors (dT:rA or dC:rG, depending on whether the vector RNA's 5' end was A or G) contained identical acceptor and donor sequences, which yielded complementary, or "matched" replication intermediates (Figure 2B). When these same vectors were mobilized using AZT^R RT, titers were 2- to 3-fold lower than those for WT RT (Figure 3B). Although all RIO vectors' titers were significantly lower than WA18 puro's, titers in the 10³ to 10⁴ cfu/ml range were readily achievable. Therefore, these vectors were suitable for examining mismatch resolution during intracellular reverse transcription.

Titers for RIO vectors designed to test mismatch extension were also determined. Most RIO vectors with single base mismatch replication intermediates displayed titers 2- to 5-fold lower than matched intermediate vectors in the presence of either form of RT, while titers for those that generated dT-rU, dT-rG, or dC-rC mismatches were approximately 10-fold lower than those for matched intermediate RIO vectors (Figure 3B).

Outcomes of primer-template mismatch during RIO vector replication

The titer results above suggested that when RT encountered a mismatch upon template switching, DNA synthesis was aborted in most cases. RIO vector product sequences were analyzed to study the contributions of mismatch extension to mutagenesis in those rarer cases when DNA synthesis was completed despite primer-template mismatch. Because RIO vectors had been engineered to yield one restriction site when mismatches were extended and an alternate site when mismatched nucleotides were excised or template switching occurred before the donor template's end, restriction analysis could reveal which mismatch resolution mechanism had functioned during provirus synthesis. Donor and acceptor

sequences and the recombination junction sequences predicted to form in proviral products are shown in Figure 2.

To perform restriction analysis of reverse transcription outcomes, 293T target cells were transduced with RIO vectors, pools of cells containing proviruses were selected in puromycin, and proviral sequences were PCR amplified from pooled cells' DNA under conditions where one primer was radiolabeled. These PCR products were digested with diagnostic restriction enzymes, separated on agarose, and quantified by phosphorimager (Figure 4A). Pools consisting of 400–1000 proviral products of all 8 vectors shown in Figure 2B, mobilized in pairwise combinations with WT or AZT^R mutant helpers, were analyzed.

As quantified in Figure 4B, restriction analysis of the matched intermediate RIO vectors, dC:rG and dT:rA, showed that more than 90% of their viral DNA products were digestible by NotI and AhdI, respectively: verifying that for the most part, RIO vectors faithfully completed forced copy choice recombination as expected when mobilized by either WT or AZT^R RT. Furthermore, for both WT and AZT^R RT, restriction analysis indicated that most analyzed mismatches-- dT:rU, dT:rG, dC:rU and dC:rA -- were precisely extended 80–95% of the time (Figure 4A–B). In stark contrast, the dC:rC mismatch was extended only about ~10% or ~5% of the time by WT and AZT^R RTs, respectively (Figure 4B). Among the remaining products, restriction analysis suggested ~80% of dC:rC mismatch products generated by both RTs resulted from primer terminus excision or premature jump.

Products of a subset of RIO vectors, including those from the inefficiently extended mismatch dC:rC, were further analyzed using high throughput sequencing (Figure 5 and see Materials and Methods). The results confirmed that ~95% of the products of matched intermediate RIO vectors dC:rG and dT:rA contained the junctional sequences predicted to arise upon precise and faithful forced copy choice recombination: both for WT and AZT^R RT (Figure 5). Precise mismatch extension was observed at similar high levels among all products of the dT:rC mismatch.

Unexpectedly, high throughput sequencing revealed that a major class of products of the dC:rC mismatch had been misdiagnosed by restriction analysis. Whereas restriction enzyme data had classified ~80% of the dC:rC mismatch products of both WT and AZT^R RT as being products of primer terminal excision or premature jump (Figure 4B), sequencing revealed that the predominant products for WT and AZT^R differed. As predicted by the restriction digestion, 90% of the WT RT dC:rC products were indeed products of primer terminal excision or premature jump. In contrast, fully 80% of the AZT^R RT dC:rC products contained a one nucleotide deletion at the point of strand transfer: apparently due to misalignment of the primer terminus on the template, as has been reported to occur during both HIV-1 and murine leukemia virus reverse transcription ([34, 42] and see “dislocation” outcome in Figure 2C). This unanticipated major product class fortuitously generated the same junctional restriction site (Asc I) that was diagnostic of products of base excision/premature jump: albeit in a different sequence context. Therefore, although high throughput sequencing revealed that the reverse transcription products susceptible to Asc I digestion differed from original expectations, initial restriction analysis was consistent with the high throughput sequencing data.

For all high throughput sequencing samples, a moderate percentage (2.5–9% of the sequencing reads) contained single nucleotide substitutions one base “before” the point of forced copy choice template switch (–1 mutations; Figure 5). Unlike the above unanticipated deletion, whose sequence was present in a large number of high throughput sequencing reads, the –1 mutations summed in Figure 5 were a mixture of all possible substitutions at that position. +1 mutations, which have been reported to occur during recombinogenic template switching *in vitro*, were not observed at levels higher than the sporadic errors observed at non-junctional positions in the sequencing reads. We do not know the cause of these substitutions but note that they were detected only within the high throughput sequencing data. The discussion describes the similarity between RIO junction sequences and known hot spots for Illumina sequencing-associated mutations, thus suggesting these may be a technical artifact. The percentages of undigested products observed in the restriction enzyme analyses (Figure 4B) were in large part consistent with the –1 mutation frequencies detected by high throughput sequencing.

To further validate product distributions, proviral DNA sequences from a limited number of individual clonal integrants were also determined (Table I). Cellular DNA was isolated from a total of 120 individual puromycin-resistant 293T cell clones, each harboring an independent product generated by dC:rC, dC:rA, or dT:rG HIV-1 RIO vectors, mobilized by either WT or AZT^R RT. Recombination junction sequences were PCR amplified and sequenced. Table 1 shows the number of colonies with sequences corresponding to mismatch extension, mismatch excision/premature jump, and a one-base deletion for each sample. Although the limited sample sizes in this experiment resulted in somewhat different magnitudes of effects as those reported in the high throughput approaches described in Figures 4 and 5, in general, trends in single clone sequencing assessment of mismatch resolution outcomes were consistent with restriction analysis and high throughput sequencing. Note, however, that although restriction digestion and high throughput sequencing suggested junctional sequences suffered single –1 position substitutions in ~2–9% of reverse transcription products, none of the 120 individual products analyzed contained such mutations.

Discussion

Here we developed a unique system to examine forced copy-choice recombination during HIV-1 replication in human cells. Using this system, mismatched base pairs were induced at a defined location in replication intermediates, and the mechanisms by which RT resolved these mismatches were addressed. The vectors used here were based in part on forced copy choice approaches previously reported for murine leukemia virus [31, 38]. In the present study, circularly permuted HIV-1 derivatives allowed forced copy choice template switching between engineered donor and acceptor template regions on a single vector, thereby creating recombinogenic template switch complexes like those that arise about once every kilobase during HIV-1 DNA synthesis [7]. During viral replication, two replicative template switches—minus and plus strand strong stop transfer—occur in addition to the recombinogenic template switches [43]. Like the forced template switches induced on the RIO (Retroviral Inside-Out) templates used here, minus strand transfer involves homologous donor and acceptor templates on single RNAs. However, accurate minus strand transfer is guided by

factors in addition to primer: template complementarity [44–46]. Notably, acceptor template sequences downstream of the growing point for DNA synthesis—which are essential to the “acceptor template invasion” step of retroviral recombination—are completely dispensable for strong stop transfer [37, 47]. Thus, rather than study mismatch extension during strong stop transfer, where factors other than primer-template complementarity might mask the properties of interest here, RIO vectors were developed to force template switching at non-strong stop sites, to more accurately recapitulate the recombinogenic template switches that occur about once per kilobase during retroviral DNA synthesis.

HIV-1 RIO vector RNA packaging levels were as high as conventional HIV-1 vectors’ under the non-competitive packaging conditions used here. However, RIO proviral titers were ~50-fold lower than conventional vectors’. Since no defects in packaging were observed, this titer reduction likely resulted from inefficiency in forced copy choice recombination between the vectors’ 30 base acceptor and donor sequences. A longer region of donor: acceptor homology may have increased titers. Thus, whereas a short region of donor-acceptor homology was engineered to maximize the likelihood of forcing template switch at a unique position, proviral titers were reduced by this design. Not surprising in light of reports that mismatches often abort DNA synthesis in reconstituted reverse transcription reactions [48], mismatches further reduced RIO vector titers. Nonetheless, RIO vectors’ puromycin resistance colony forming titers were 10^3 /ml, which readily allowed the study of mismatch resolution.

Using RIO vectors, the mechanisms by which RT resolved mismatches were archived in proviral sequences. Mismatch resolution was studied independent of misinsertion—which ordinarily precedes mismatch resolution during RT-mediated mutagenesis—and in the context of template switching, which may differ from processive synthesis on a single template [48]. The results showed that dT:rU, dT:rC, dT:rG, dC:rU, and dC:rA mispairs nearly always either aborted DNA synthesis or were precisely extended by both WT and AZT^R RTs, thereby incorporating the mismatched residue as a mutation into proviral DNA. In contrast, the dC:rC mismatch was almost never extended and was resolved differentially by WT and AZT^R RTs. In most proviruses generated by WT RT, the dC:rC mismatch was resolved by excising the mismatched base or by switching templates prematurely, while for AZT^R RT the primer strand DNA was misaligned on the acceptor template in an interaction that has been referred to as dislocation [42], resulting in a one-base deletion in nearly all proviral products. The mechanisms contributing to these differences remain to be determined.

The focus of the current report was on establishing a system to study mismatch extension and on assessing its predominant outcomes, with three diagnostic methods explored. Restriction analysis using engineered restriction sites did, for the most part, prove prognostic of outcomes determined by other means, and high throughput sequencing confirmed results of the diagnostic restriction enzyme digestions. However, restriction analysis was limited by requiring template alterations that may affect replication outcomes at mismatches [49] and—as was revealed here by the frequency of dislocation mutations—was limited in its ability to identify unanticipated outcomes. A limitation of both pooled product restriction analysis and of high throughput sequencing as performed here was the potentially mutagenic

manipulation of reverse transcription products prior to analysis. As a result, switch-associated mutations detected by both approaches likely included errors introduced during PCR and/or by Illumina sequencing, as well as products of reverse transcription. Notable in this regard, a major trigger of Illumina misreads are GGC sequences or short inverted repeats, both of which were present in the recombination junctions of the RIO vectors used here [50]. In future work, employing Primer ID approaches will help resolve artifacts from rare reverse transcription outcomes, and increasing product pool sizes should help validate rare products [51]. However, despite the limitations of the initial approaches used here, the RIO vector system demonstrated its power to dissect molecular outcomes during virus replication and to detect functional differences among RT variants, such as the dramatically different mechanisms observed for the resolution of the dC:rC mismatch by WT and AZT^R RTs.

An unanswered question is why AZT^R and WT RT behaved so differently upon encountering the dC:rC mismatch. One possible explanation may be the enhanced pretranslocation dwell time associated with AZT^R RT's chemical cycle, which may allow additional sampling of alternate replication intermediate structures prior to the resumption of RT elongation. The high frequency of dislocation mutations observed here likely arose in part due to a 3-base tandem duplication of donor 5' end complementarity in the dC:rC vector's acceptor template region. All dC:rN mismatches had the potential to form at least two primer-terminal basepairs upon misalignment (Figure 2). However, high throughput sequencing and conventional Sanger sequencing revealed that dislocation was rare among the products of dC:rN vectors other than AZT^R RT-generated products of dC:rC.

Several previous studies have examined the ability of HIV RT to extend mismatched bases *in vitro*. In general, they found that pyrimidine:purine mismatches are more likely to be extended by RT than purine:purine or pyrimidine:pyrimidine mismatches [52–54]. In agreement with these observations, two pyrimidine:purine mismatches (dT:rG and dC:rA) tested here were preferentially extended by RT. However, mismatch extension was also the most frequent outcome of most pyrimidine: pyrimidine mismatches. In proviral products of the dT:rC, dT:rU and dC:rU vectors, both WT and AZT^R RT had extended most mismatches. A study examining HIV-1 mutations during replication [12], as well as phylogenetic analysis of HIV-1 isolates from infected individuals [55] revealed that T to A and C to A transversions are common events in HIV mutagenesis, and these can be products of dT:rU and dT:rC mismatch extensions, respectively. Although T to G transversions, some of which are likely products of dC:rU mismatches, are less common, they are still abundant among mutations observed in HIV-1 genomes [12, 55]. C to G transversions, on the other hand, are the rarest substitutions observed in phylogenetic analyses [55] and were not detected in a study of mutations that accrued during viral replication [12]. Therefore, our results that RT almost never extends dC:rC mismatches, as would be required to generate C to G substitutions, correlate well with previous findings and support a significant contribution of mismatch extension to HIV-1 replicative error rates.

With the notable exception of the dC:rC mismatch, HIV-1 RT displayed little nucleotide-specific differences in mismatch extension during virus replication for the mismatches studied here (dT:rC, dT:rG, dT:rU, dC:rA, and dC:rU). Additionally, these mismatches were

extended similarly by WT and AZT^R RTs. Because the work here dissected mutagenesis to allow examining the process of mismatch extension independent of base misinsertion frequencies, these new observations provide mechanistic insight into previous studies which have shown that AZT^R RT displays a 2.2- to 5-fold higher mutation rate compared to WT RT [56–58]. Specifically, the revelation of similar mismatch extension rates for WT and AZT^R RTs suggest that the elevated rate of AZT^R RT mutagenesis may be dictated more by differences in discrimination between correct and incorrect nucleotides during polymerization than by differences in resolution of the resulting mismatched intermediates. Note, however, that the current work was performed in actively dividing cells and conditions of high nucleotide pools: quite unlike conditions the virus may encounter *in vivo* that could affect nucleotide discrimination [59, 60].

One very common outcome of recombinogenic template switching in purified reactions is the addition a non-templated residue at the primer strand end prior to template switching, followed by mismatch extension upon the resumption of DNA synthesis on the acceptor template. This outcome is reported to occur up to 20% of all recombination products [61] [43]. However, no evidence of mutations generated by that mechanism was observed here.

Unlike what might simplistically have been expected based on its enhanced propensity to excise chain terminating thymidine analogs, AZT^R RT did not appear to increase primer strand unblocking, possibly reflecting the apparently low impediment to RT elongation that results from a mismatched basepair and the abundance of substrates known to inhibit the excision reaction in actively dividing cells (reviewed in [17]). Studying additional altered primer-templates, other RTs with resistance-associated alterations to active site geometries, or mutants that display functional defects *in vitro*—such as HIV RT mutants with differences in slippage-mediated error rates—in future RIO-based work may further advance understanding of RT: primer-template interactions and RT fidelity [62].

Materials and Methods

Plasmid construction: HIV retroviral inside-out vectors

HIV retroviral inside-out (RIO) vectors were generated by modifying the conventional HIV vector pWA18puro. pWA18puro is a derivative of pHIV LacPuro from which the CMV promoter-driven LacZ gene was removed [63]. Serial overlap extension PCR was used to create a fragment containing the following HIV elements in the order listed: polypurine tract (ppt), long terminal repeat (LTR) with a deletion of much of U3 and R, primer binding site (PBS), and 5' leader regions containing the packaging signal (Ψ). This fragment was ligated into pWA18puro, upstream of its RRE and SV40 promoter-driven puromycin resistance gene.

The donor sequences used to produce RIO vector RNAs with either a guanine (G) or adenosine (A) as the 5' terminal nucleotide were generated by annealing the synthetic oligonucleotides 5'

AATTCCGCATTGCAGAGATATTGTATTAAAGTGCCTAGCTCGATACAATAAACGCCG
CATGT CAAGAGAGACAATTGC and 5'

AATTCCGCATTGCAGAGATATTGTATTAAAGTGCCTAGCTCGATACAATAAACACCG

CATGT CAAGAGAGACAATTGC, respectively, to complementary oligonucleotides: yielding short duplexes with ends designed to facilitate the following construction steps. An RSV promoter fragment amplified from pREP8 (Invitrogen) was ligated to the annealed oligonucleotides using an EcoRI restriction site. The resulting RSV promoter/initial transcribed region fragment was cloned into the modified pWA18puro construct, upstream of the ppt, using PstI and XhoI restriction sites.

The acceptor sequences for forced copy choice recombination, containing 30 nts of identity to the initial transcribed region donor sequences, were also generated by annealing synthetic oligonucleotides. The top strands of the four acceptor fragments were:

5' GTACCGACGCGACCGCATGTCAAGAGAGACAATTGCTCGAG, 5'

GTACCGACGCGTCCGCATGTCAAGAGAGACAATTGCTCGAG, 5'

GTACCGGCGCGCCCGCATGTCAAGAGAGACAATTGCTCGAG, and 5'

GTACCGACGCGGCCGCATGTCAAGAGAGACAATTGCTCGAG.

Acceptor sequences were first inserted into a pNGVL3 derivative [64] using BamHI and Asp718 to create an intermediate plasmid containing restriction sites useful in subsequent construction steps. Specifically, PstI-Asp718 fragments containing the RSV promoter-donor template elements were ligated into plasmids containing the acceptor sequences (Asp718-PstI) to produce eight HIV RIO vectors.

Unexpectedly, the production of some RIO vector plasmids in *E. coli* generated colonies that grew slowly. Plasmid sequencing revealed insertions of *E. coli* transposable elements in or around the RRE sequences of some RIO vector plasmids: all insertions were members of IS element classes previously demonstrated to transpose into plasmids [65–67]. Of the initial eight RIO plasmids tested, four (dC:rA, dC:rC, dT:rG, and dT:rU) contained insertions while four remained intact (dC:rG, dT:rA, dC:rU, and dT:rC). Subsequent deletion of IS sequences, followed by a comparison of the HIV-1 replication properties of vector RNAs generated from RIO plasmids with or without bacterial IS sequences, indicated that the insertions had no discernable effects on forced copy choice outcomes and little or no effect on titers in mammalian cells, but sporadic re-insertion of IS elements was observed during plasmid propagation in *E. coli*.

In attempts to remove the apparent source of RIO instability, RRE sequences were deleted from RIO vectors by re-circularization of SnaBI plus XmnI digested plasmids, yielding eight RRE RIO vectors. These plasmids were stable in *E. coli* and yielded proviral products with mismatch extension patterns indistinguishable from those generated by the cognate RRE-containing RIO vectors. However, removal of the RRE from RIO vectors resulted in a 25- to 100-fold decrease in vector RNA packaging and correlated titer reductions. Based on these findings, the original HIV RIO vectors with RRE were used in all experiments here.

Plasmid construction: AZT^R RT helper function plasmid

Wild-type HIV Gag, Pol, and accessory proteins were expressed from the helper plasmid pCMV R8.2 [68]. To construct an AZT^R RT-containing version of pCMV R8.2, the

following amino acid substitutions that confer resistance to AZT: D67N, K70R, T215F, and K219Q [69] were introduced into an HIV reverse transcriptase (RT) gene fragment using overlap extension PCR, sequenced, and used to replace wild-type sequences between AgeI and BclI in pCMV R8.2 to produce pCBN103-18, which contains an AZT-resistant (AZT^R) RT gene.

Virion production and titer

Retroviral vectors were produced by PEI mediated co-transfection of 293T cells [70] with RIO vectors or pWA18puro along with helper plasmids: either pCMV R8.2 [68] that expressed wild type viral proteins, or pCBN103-18, an AZT^R derivative. To pseudotype virions, co-transfections included a vesicular stomatitis virus (VSV) G protein expression plasmid (pHEF-VSVG) [71], which was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH, from Lung-Ji Chang. Culture media from transfected cells was collected 48 hours post transfection, filtered (0.22 µm) and used for infection or RNA extraction.

Titers were determined by puromycin resistant colony forming units/ml, with virus concentrations normalized by reverse transcriptase (RT) activity [72]. Serial dilutions of culture supernatants from transfected cells were used to infect 293T cells. After 48 hours, infected cells were treated with 2 µg/ml puromycin. Twelve days later, plates containing 500–1000 puromycin-resistant colonies were xed and stained using Coomassie stain (0.25% Coomassie Brilliant Blue R250, 10% v/v glacial acetic acid, and 45% methanol). Provirus-containing cell DNA was obtained from infected cells selected in puromycin and cultured in parallel, using the DNeasy Tissue and Blood kit (Qiagen, Valencia, CA). These DNA samples—either from individual or pooled puromycin resistant colonies were used as described below.

Viral RNA and DNA analyses

To examine viral RNA packaging, virus particles were concentrated by ultracentrifugation through a 20% sucrose cushion and RNA was extracted using TRIzol according to the manufacturer's protocol (Invitrogen). For analysis of cellular RNA, cells were lysed with TRIzol 48 hours after transfection. All viral samples were normalized by reverse transcriptase (RT) activity [72]. The amount of viral genomic RNA (gRNA) and host 7SL RNA present in samples was quantified by RNase protection assay as previously described [73] using two riboprobes. One was a chimeric 335 nt riboprobe generated by *in vitro* transcription of EcoRI-linearized pSKh64b-7SL/puro using T7 RNA polymerase and radiolabeled rCTP, and was used to detect a 145 nt puro gene fragment of gRNA and a 101 nt fragment of 7SL. The same 101 nt portion of 7SL was protected by hybridization of a second riboprobe that was generated by *in vitro* transcription of SalI digested pBru7SL. RNA fragments were separated by denaturing 8% polyacrylamide gel electrophoresis. Dried gels were visualized and bands quantified using a Typhoon Scanner phosphorimager (Amersham Biosciences) and ImageQuant LT software (GE Healthcare).

Products of forced copy choice recombination were examined by analyzing proviral DNA from pools of puromycin resistant cells. A 243 bp portion of HIV proviral DNA was ampli

ed using Taq DNA polymerase with ThermoPol buffer (New England Biolabs) using the following primer pairs: dirHIVpuro_2 (5' CCGCAACCTCCCCTTCTA) and rcSD1 (5' AAGTCATTGGTCTTAAAGGATCC). For restriction analysis, the dirHIVpuro_2 primer was γ -³²P-ATP prior to amplification. PCR products were analyzed by restriction enzyme digestion and the generated fragments were separated on non-denaturing 8% polyacrylamide gels and bands quantified by phosphorimager analysis.

For individual provirus analysis, single puromycin resistant colonies were harvested with cloning cylinders. Cells were grown in six-well plates to full confluency and total genomic DNA was isolated using the DNeasy Tissue and Blood kit (Qiagen, Valencia, CA). HIV proviral DNA was amplified with primers dirHIVpuro_2 and rcSD1 and PCR products were Sanger sequenced in the U-M Sequencing Core.

High-throughput sequencing

Proviral DNA from uncloned populations of 293T cells (comprised of 600–1000 puromycin resistant colonies per population) that had been infected with one of the following RIO vectors: dC:rG, dT:rA, dT:rC, or dC:rC, mobilized by either WT or AZT^R RT, was amplified with primers dirHIVpuro_2 and rcSD1, using Taq DNA polymerase and ThermoPol buffer (NEB). A second round of PCR was performed with the primer dirHIVpuro_2 and a reverse primer that was uniquely barcoded for each DNA sample to allow multiplexing using Pfu with Pfu Ultra buffer (Agilent). Illumina high-throughput sequencing and analysis was performed by the University of Michigan Medical School Sequencing and Bioinformatics Cores. Approximately 90% of a total of 6,859,630 reads passed the imposed filters, including low degeneracy and possession of sequences from +4 upstream of the engineered transfer point to position –4 downstream. At least 470,000 sequence reads for each multiplexed sample (~500X coverage), including all sequences that passed the filters with greater than 30 reads were considered, with those with minor sequence differences outside the transfer interval clustered into related major sequence classes. These sequences were analyzed to infer how each terminally matched/mismatched RIO vector was processed during reverse transcription.

Acknowledgments

This work was supported by NIH grant P50 GM103297 to AT, with early stage support from R01 GM64479. We thank Cleo Burnett for preparing the figures, Philip Smaldino for critical reading of the manuscript, Rupak Neupane for experimental help at early stages of this project in this work, and an anonymous reviewer for helpful input on the manuscript.

References

1. Wain-Hobson S. Human immunodeficiency virus type 1 quasispecies in vivo and ex vivo. *Curr Top Microbiol Immunol*. 1992; 176:181–93. [PubMed: 1600752]
2. Malim MH, Emerman M. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell Host Microbe*. 2008; 3(6):388–98. [PubMed: 18541215]
3. Cuevas JM, et al. Extremely High Mutation Rate of HIV-1 In Vivo. *PLoS Biol*. 2015; 13(9):e1002251. [PubMed: 26375597]
4. O'Neil PK, et al. Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. *J Biol Chem*. 2002; 277(41):38053–61. [PubMed: 12151398]

5. Preston BD, Dougherty JP. Mechanisms of retroviral mutation. *Trends Microbiol.* 1996; 4(1):16–21. [PubMed: 8824790]
6. Sanjuan R, et al. Viral mutation rates. *J Virol.* 2010; 84(19):9733–48. [PubMed: 20660197]
7. Onafuwa-Nuga A, Telesnitsky A. The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol Mol Biol Rev.* 2009; 73(3):451–80. Table of Contents. [PubMed: 19721086]
8. Rhodes T, Wargo H, Hu WS. High rates of human immunodeficiency virus type 1 recombination: near-random segregation of markers one kilobase apart in one round of viral replication. *J Virol.* 2003; 77(20):11193–200. [PubMed: 14512567]
9. Jetzt AE, et al. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J Virol.* 2000; 74(3):1234–40. [PubMed: 10627533]
10. Achuthan V, et al. Human immunodeficiency virus reverse transcriptase displays dramatically higher fidelity under physiological magnesium conditions in vitro. *J Virol.* 2014; 88(15):8514–27. [PubMed: 24850729]
11. Mansky LM, Temin HM. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol.* 1995; 69(8):5087–94. [PubMed: 7541846]
12. Abram ME, et al. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol.* 2010; 84(19):9864–78. [PubMed: 20660205]
13. Menendez-Arias L. Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. *Viruses.* 2009; 1(3):1137–65. [PubMed: 21994586]
14. Arion D, et al. Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase. *Biochemistry.* 1998; 37(45):15908–17. [PubMed: 9843396]
15. Sarafianos SG, et al. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol.* 2009; 385(3):693–713. [PubMed: 1902262]
16. Bampi C, et al. Nucleotide excision repair and template-independent addition by HIV-1 reverse transcriptase in the presence of nucleocapsid protein. *J Biol Chem.* 2006; 281(17):11736–43. [PubMed: 16500895]
17. Menendez-Arias L. Molecular basis of human immunodeficiency virus drug resistance: an update. *Antiviral Res.* 2010; 85(1):210–31. [PubMed: 19616029]
18. Goldschmidt V, Marquet R. Primer unblocking by HIV-1 reverse transcriptase and resistance to nucleoside RT inhibitors (NRTIs). *Int J Biochem Cell Biol.* 2004; 36(9):1687–705. [PubMed: 15183338]
19. Meyer PR, et al. A mechanism of AZT resistance: an increase in nucleotide-dependent primer unblocking by mutant HIV-1 reverse transcriptase. *Mol Cell.* 1999; 4(1):35–43. [PubMed: 10445025]
20. Lwatula C, Garforth SJ, Prasad VR. Lys66 residue as a determinant of high mismatch extension and misinsertion rates of HIV-1 reverse transcriptase. *FEBS J.* 2012; 279(21):4010–24. [PubMed: 22925131]
21. Rubinek T, et al. The fidelity of 3' misinsertion and mispair extension during DNA synthesis exhibited by two drug-resistant mutants of the reverse transcriptase of human immunodeficiency virus type 1 with Leu74-->Val and Glu89-->Gly. *Eur J Biochem.* 1997; 247(1):238–47. [PubMed: 9249032]
22. Marchand B, Gotte M. Impact of the translocational equilibrium of HIV-1 reverse transcriptase on the efficiency of mismatch extensions and the excision of mispaired nucleotides. *Int J Biochem Cell Biol.* 2004; 36(9):1823–35. [PubMed: 15183347]
23. Abram ME, et al. Mutations in HIV-1 reverse transcriptase affect the errors made in a single cycle of viral replication. *J Virol.* 2014; 88(13):7589–601. [PubMed: 24760888]
24. Wainberg MA, et al. Enhanced fidelity of 3TC-selected mutant HIV-1 reverse transcriptase. *Science.* 1996; 271(5253):1282–5. [PubMed: 8638110]
25. DeStefano J, et al. High fidelity of internal strand transfer catalyzed by human immunodeficiency virus reverse transcriptase. *J Biol Chem.* 1998; 273(3):1483–9. [PubMed: 9430686]

26. Peliska JA, Benkovic SJ. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science*. 1992; 258(5085):1112–8. [PubMed: 1279806]
27. Patel PH, Preston BD. Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. *Proc Natl Acad Sci U S A*. 1994; 91(2):549–53. [PubMed: 7507249]
28. Wu W, et al. Strand transfer mediated by human immunodeficiency virus reverse transcriptase in vitro is promoted by pausing and results in misincorporation. *J Biol Chem*. 1995; 270(1):325–32. [PubMed: 7529228]
29. DeStefano JJ, Raja A, Cristofaro JV. In vitro strand transfer from broken RNAs results in mismatch but not frameshift mutations. *Virology*. 2000; 276(1):7–15. [PubMed: 11021989]
30. Zhuang J, et al. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol*. 2002; 76(22):11273–82. [PubMed: 12388687]
31. Pfeiffer JK, Telesnitsky A. Effects of limiting homology at the site of intermolecular recombinogenic template switching during Moloney murine leukemia virus replication. *J Virol*. 2001; 75(23):11263–74. [PubMed: 11689606]
32. Schlub TE, et al. Fifteen to twenty percent of HIV substitution mutations are associated with recombination. *J Virol*. 2014; 88(7):3837–49. [PubMed: 24453357]
33. Chin MP, et al. Long-range recombination gradient between HIV-1 subtypes B and C variants caused by sequence differences in the dimerization initiation signal region. *J Mol Biol*. 2008; 377(5):1324–33. [PubMed: 18314135]
34. Kulpa D, Topping R, Telesnitsky A. Determination of the site of first strand transfer during Moloney murine leukemia virus reverse transcription and identification of strand transfer-associated reverse transcriptase errors. *EMBO J*. 1997; 16(4):856–65. [PubMed: 9049314]
35. Shoemaker C, et al. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. *Proc Natl Acad Sci U S A*. 1980; 77(7):3932–6. [PubMed: 6449003]
36. Kadonaga JT. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol*. 2012; 1(1):40–51. [PubMed: 23801666]
37. Marr SF, Telesnitsky A. Mismatch extension during strong stop strand transfer and minimal homology requirements for replicative template switching during Moloney murine leukemia virus replication. *J Mol Biol*. 2003; 330(4):657–74. [PubMed: 12850138]
38. Duggal NK, et al. Effects of identity minimization on Moloney murine leukemia virus template recognition and frequent tertiary template-directed insertions during nonhomologous recombination. *J Virol*. 2007; 81(22):12156–68. [PubMed: 17804514]
39. Onafuwa-Nuga AA, Telesnitsky A, King SR. 7SL RNA, but not the 54-kd signal recognition particle protein, is an abundant component of both infectious HIV-1 and minimal virus-like particles. *RNA*. 2006; 12(4):542–6. [PubMed: 16489186]
40. Blissenbach M, et al. Nuclear RNA export and packaging functions of HIV-1 Rev revisited. *J Virol*. 2010; 84(13):6598–604. [PubMed: 20427541]
41. Grewe B, et al. The HIV-1 Rev protein enhances encapsidation of unspliced and spliced, RRE-containing lentiviral vector RNA. *PLoS One*. 2012; 7(11):e48688. [PubMed: 23133650]
42. Bebenek K, et al. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J Biol Chem*. 1989; 264(28):16948–56. [PubMed: 2476448]
43. Telesnitsky, A.; Goff, SP. Reverse Transcriptase and the Generation of Retroviral DNA. In: Coffin, JM.; Hughes, SH.; Varmus, HE., editors. *Retroviruses*. Cold Spring Harbor (NY): 1997.
44. Piekna-Przybylska D, Bambara RA. Requirements for efficient minus strand strong-stop DNA transfer in human immunodeficiency virus 1. *RNA Biol*. 2011; 8(2):230–6. [PubMed: 21444998]
45. Galvis AE, et al. Impairment of HIV-1 cDNA synthesis by DBR1 knockdown. *J Virol*. 2014; 88(12):7054–69. [PubMed: 24672043]
46. Topping R, et al. Cis-acting elements required for strong stop acceptor template selection during Moloney murine leukemia virus reverse transcription. *J Mol Biol*. 1998; 281(1):1–15. [PubMed: 9680471]

47. Brincat JL, Pfeiffer JK, Telesnitsky A. RNase H activity is required for high-frequency repeat deletion during Moloney murine leukemia virus replication. *J Virol.* 2002; 76(1):88–95. [PubMed: 11739674]
48. Liesch GR, DeStefano JJ. Analysis of mutations made during active synthesis or extension of mismatched substrates further define the mechanism of HIV-RT mutagenesis. *Biochemistry.* 2003; 42(19):5925–36. [PubMed: 12741851]
49. Meyer PR, et al. Effects of primer-template sequence on ATP-dependent removal of chain-terminating nucleotide analogues by HIV-1 reverse transcriptase. *J Biol Chem.* 2004; 279(44):45389–98. [PubMed: 15308646]
50. Nakamura K, et al. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* 2011; 39(13):e90. [PubMed: 21576222]
51. Zhou S, et al. Primer ID Validates Template Sampling Depth and Greatly Reduces the Error Rate of Next-Generation Sequencing of HIV-1 Genomic RNA Populations. *J Virol.* 2015; 89(16):8540–55. [PubMed: 26041299]
52. Perrino FW, et al. Extension of mismatched 3' termini of DNA is a major determinant of the infidelity of human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci U S A.* 1989; 86(21):8343–7. [PubMed: 2479023]
53. Bakhianashvili M, Hizi A. The interaction of the reverse transcriptase of human immunodeficiency virus type 1 with 3'-terminally mispaired DNA. *Arch Biochem Biophys.* 1996; 334(1):89–96. [PubMed: 8837743]
54. Menendez-Arias L. Studies on the effects of truncating alpha-helix E' of p66 human immunodeficiency virus type 1 reverse transcriptase on template-primer binding and fidelity of DNA synthesis. *Biochemistry.* 1998; 37(47):16636–44. [PubMed: 9843431]
55. Hillis DM, Huelsenbeck JP, Cunningham CW. Application and accuracy of molecular phylogenies. *Science.* 1994; 264(5159):671–7. [PubMed: 8171318]
56. Chen R, et al. Human immunodeficiency virus mutagenesis during antiviral therapy: impact of drug-resistant reverse transcriptase and nucleoside and nonnucleoside reverse transcriptase inhibitors on human immunodeficiency virus type 1 mutation frequencies. *J Virol.* 2005; 79(18):12045–57. [PubMed: 16140780]
57. Mansky LM, Bernard LC. 3'-Azido-3'-deoxythymidine (AZT) and AZT-resistant reverse transcriptase can increase the in vivo mutation rate of human immunodeficiency virus type 1. *J Virol.* 2000; 74(20):9532–9. [PubMed: 11000223]
58. Mansky LM, Pearl DK, Gajary LC. Combination of drugs and drug-resistant reverse transcriptase results in a multiplicative increase of human immunodeficiency virus type 1 mutant frequencies. *J Virol.* 2002; 76(18):9253–9. [PubMed: 12186909]
59. Amie SM, Noble E, Kim B. Intracellular nucleotide levels and the control of retroviral infections. *Virology.* 2013; 436(2):247–54. [PubMed: 23260109]
60. Julias JG V, Pathak K. Deoxyribonucleoside triphosphate pool imbalances in vivo are associated with an increased retroviral mutation rate. *J Virol.* 1998; 72(10):7941–9. [PubMed: 9733832]
61. Peliska JA, Benkovic SJ. Fidelity of in vitro DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Biochemistry.* 1994; 33(13):3890–5. [PubMed: 7511410]
62. Hamburgh ME, et al. Structural determinants of slippage-mediated mutations by human immunodeficiency virus type 1 reverse transcriptase. *J Biol Chem.* 2006; 281(11):7421–8. [PubMed: 16423828]
63. An W, Telesnitsky A. Frequency of direct repeat deletion in a human immunodeficiency virus type 1 vector during reverse transcription in human cells. *Virology.* 2001; 286(2):475–82. [PubMed: 11485415]
64. Yang S, et al. Generation of retroviral vector for clinical studies using transient transfection. *Hum Gene Ther.* 1999; 10(1):123–32. [PubMed: 10022537]
65. Timmerman KP, Tu CP. Complete sequence of IS3. *Nucleic Acids Res.* 1985; 13(6):2127–39. [PubMed: 2987844]
66. Reinitz DM, Inverso JA, Mansfield JM. Complete nucleotide sequence of an E.coli IS5 insertion element containing an internal 88 base pair direct repeat (IS5-D). *Nucleic Acids Res.* 1989; 17(10):3990. [PubMed: 2543960]

67. Kovarik A, et al. Transposition of IS10 from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts. *Mol Genet Genomics*. 2001; 266(2):216–22. [PubMed: 11683262]
68. Naldini L, et al. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science*. 1996; 272(5259):263–7. [PubMed: 8602510]
69. Larder BA, Kemp SD. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science*. 1989; 246(4934):1155–8. [PubMed: 2479983]
70. Boussif O, et al. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc Natl Acad Sci U S A*. 1995; 92(16):7297–301. [PubMed: 7638184]
71. Chang LJ, et al. Efficacy and safety analyses of a recombinant human immunodeficiency virus type 1 derived vector system. *Gene Ther*. 1999; 6(5):715–28. [PubMed: 10505094]
72. Telesnitsky A, Blain S, Goff SP. Assays for retroviral reverse transcriptase. *Methods Enzymol*. 1995; 262:347–62. [PubMed: 8594360]
73. Onafuwa-Nuga AA, King SR, Telesnitsky A. Nonrandom packaging of host RNAs in moloney murine leukemia virus. *J Virol*. 2005; 79(21):13528–37. [PubMed: 16227273]

Highlights

Dissection of HIV-1 RT intracellular error mechanisms is challenging

A novel system to study mismatch resolution during HIV-1 replication was developed

Outcomes for WT and an AZT-resistant RT were quantified using three approaches

Most mismatches were extended to similar high extents by both enzymes

Striking mechanistic differences were revealed during dC-rC mispair resolution

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Schematic explanation of Retroviral Inside-Out (RIO) vector design, and steps of RIO vector reverse transcription that differ from those of HIV-1 and conventional HIV-1 based vectors. **(A)** *Schematic of the conventional pWA18puro vector* (Figure 1A top line; not to scale). The conventional HIV-based vector plasmid, pWA18puro contains native 5' and 3' LTRs (long terminal repeats consisting of U3, R, and U5), other essential cis-acting elements {primer binding site (pbs), packaging signal (Ψ), Rev response element (RRE), and polypurine tract (ppt)} and a selectable marker gene for puromycin resistance gene (Puro) driven by the SV40 virus promoter. *How RIO vectors mimic recombination intermediates* (Figure 1A; WA18 puro RNA vectors' recombination). Heavy lines indicate nascent DNA strand; single-

stranded wavy lines represent RNA; double-stranded wavy lines indicate host/non-viral DNA; portions shaded blue indicate the reverse-transcribed regions whose sequences contribute to recombination products; (D) indicates recombination donor and (A) indicates recombination acceptor sequences. (Figure 1A; RIO RNA) The curved dotted lines indicate how RIO vector design involved fusing segments that contribute to retroviral recombination products in reversed order (Figure 1A; RIO plasmid). The RIO RNA expression plasmid, which contains a Rous Sarcoma virus-derived promoter and SV40 polyadenylation signal, is indicated. **(B) RIO vector reverse transcription.** (Figure 1B: RIO vector initial minus strand DNA product) Minus strand DNA synthesis initiates from a tRNA annealed to the pbs on a RIO RNA transcript. However, whereas retroviral RNAs' 5' ends terminate at R, RIO vector RNAs contain additional upstream sequences. Thus, the 5' ends of RIO RNAs resemble templates that arise after minus strand transfer for conventional vectors, and contain forced copy choice donor sequences (D) at their 5' ends (Figure 1B: RIO vector forced template switch) After reaching RIO vectors' 5' end, the donor-terminated minus strand DNA switches to the acceptor site: allowing continued minus strand synthesis. (Figure 1B: RIO provirus) The double-stranded DNA that results after completion of subsequent RIO vector reverse transcription steps.

Figure 2.

RIO vector forced copy choice recombination, mismatch generation and resolution (A)

Formation of RIO RNA 5' ends. Start site junctions in plasmid DNA are indicated below the schematic of RIO plasmids' transcription start site. Two start site sequences were used in this study: one engineered to introduce an A at RIO RNA's 5' end, and the second with the native RSV 5' G. Note that because this study's focus is on template switch outcomes, the residue upstream of the transcription start site is designated +1, because it marks the position of the first residue incorporated after template switching. (B) *Matched and mismatched replication intermediates that result during RIO vector reverse transcription.* Each of the two 5' end sequences represented in Figure 2A is paired with one of four acceptor template sequences, to generate the two matched-end and six mismatched-end intermediates indicated in Figure 2B. (C) *Outcomes of RIO vector reverse transcription.* Anticipated possible proviral DNA outcomes, using the dC:rC mismatch as an example, are shown. For this mismatch, precise extension of the mispair would generate a Not I site that is not present in the donor template region. If template switching occurred before completion of the initial minus strand DNA product, or if the terminal base were excised before synthesis resumed upon template switching, then proviral products would contain a junctional Asc I site. Two additional possible mechanisms of mismatch resolution and their outcomes are presented.

Figure 3. Packaging and replication of HIV RIO vectors

(A) *Vector RNA packaging.* Conventional WA18puro and RIO vector RNAs transiently expressed with helper function plasmids in 293T cells (lanes 6–10) and packaged into virus particles (lanes 1–5) were detected by RNase protection assay (RPA). Probes protect portions of vector RNA (puro) and cellular 7SL RNA (7SL), respectively. Lane marked PS contains undigested probes; lane marked M contains molecular size standards of lengths indicated at the panel's left. (B) *Puromycin resistant colony forming titers.* Titers were determined for conventional HIV vector WA18puro and terminal matched and mismatched RIO vectors, mobilized by either WT or AZT^R RT. The Y axis indicates cfu titers per ml after samples' virion RNA content was normalized to values for WA18puro, with values based on infections using virus from three independent transfections.

Figure 4. Distribution of HIV RIO vector reverse transcription products revealed by diagnostic restriction enzyme digestion

Proviral DNA fragments were amplified from uncloned pools of RIO-transduced cells and digested with restriction enzymes diagnostic for mismatch extension or excision/premature jump. **(A)** Primary data example, showing a polyacrylamide gel of digested PCR fragments containing pooled dC-rA and dC-rC products generated by either WT or AZT^R RT, is shown. Enzymes used are indicated, with those used in odd-numbered lanes diagnostic of mismatch extension, and those in even-numbered lanes diagnostic of terminal base excision or premature template switching **(B)** Quantification of products, as determined by restriction

analysis. Each RIO vector was designed so that precise extension and premature jump/excision led to the generation of one of two restriction sites. Data are from 3–5 independent digestion experiment repetitions. The percent (on Y axis) of total PCR product digestible by each diagnostic enzyme, as well as the % not cut by either enzyme, are represented in the graphs.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5.

Mechanisms of mismatch resolution, as determined by high throughput sequencing. The graph depicts sequencing reads for products of two matched intermediate RIO vectors and two mismatched intermediate vectors generated by wild type and AZT^R RT. Reads were collated and the data indicate the percent of total high quality reads that were diagnostic of each mechanism of mismatch resolution coded at the bottom of the figure. See Materials and Methods for quality controls and other experimental details.

Table 1
Distribution of mismatch resolution as determined by sequencing individual proviruses

Terminal mismatch	RT	Total colonies	Mismatch extension	Excision/ premature jump	One base deletion
dC-rC	WT	22	3 ¹ (13.5%) ²	16 (73%)	3 (13.5%)
	AZT ^R	25	2 (8%)	11 (44%)	12 (48%)
dC-rA	WT	26	23 (88.5%)	3 (11.5%)	0
	AZT ^R	24	23 (96%)	1 (4%)	0
dT-rG	WT	12	12 (100%)	0	0
	AZT ^R	12	12 (100%)	0	0

¹ number of colonies with the indicated sequence type
² percent of total analyzed colonies for this mismatch/enzyme with this sequence type