Published online 3 August 2017 *Nucleic Acids Research, 2017, Vol. 45, No. 17* **10143–10155**
doi: 10.1093/nar/gkx689

Stimulation of reverse transcriptase generated cDNAs with specific indels by template RNA structure: retrotransposon, dNTP balance, RT-reagent usage

Christophe Penno[1,2,3], Romika Kumari[1], Pavel V. Baranov[1], Douwe van Sinderen[2,3] and John F. Atkins[1,2,4,*]

[1]School of Biochemistry, University College Cork, Cork, Ireland, [2]School of Microbiology, University College Cork, Cork, Ireland, [3]Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland and [4]Department of Human Genetics, University of Utah, Salt Lake City, UT 84112-5330, USA

Received March 14, 2017; Revised July 19, 2017; Editorial Decision July 21, 2017; Accepted July 24, 2017

ABSTRACT

RNA dependent DNA-polymerases, reverse transcriptases, are key enzymes for retroviruses and retroelements. Their fidelity, including indel generation, is significant for their use as reagents including for deep sequencing. Here, we report that certain RNA template structures and G-rich sequences, ahead of diverse reverse transcriptases can be strong stimulators for slippage at slippage-prone template motif sequence 3′ of such 'slippage-stimulatory' structures. Where slippage is stimulated, the resulting products have one or more additional base(s) compared to the corresponding template motif. Such structures also inhibit slippage-mediated base omission which can be more frequent in the absence of a relevant stem–loop. Slippage directionality, base insertion and omission, is sensitive to the relative concentration ratio of dNTPs specified by the RNA template slippage-prone sequence and its 5′ adjacent base. The retrotransposon-derived enzyme TGIRT exhibits more slippage *in vitro* than the retroviral enzymes tested including that from HIV. Structure-mediated slippage may be exhibited by other polymerases and enrich gene expression. A cassette from *Drosophila* retrotransposon Dme1_chrX_2630566, a candidate for utilizing slippage for its GagPol synthesis, exhibits strong slippage *in vitro*. Given the widespread occurrence and importance of retrotransposons, systematic studies to reveal the extent of their functional utilization of RT slippage are merited.

INTRODUCTION

Non-standard events during the elongation phase of transcription can either enrich gene expression or contribute to erroneous and wasteful expression. An example of the former is selection for reverse transcriptase-mediated multiple alternative base substitutions to lead to pathogen surface variability to evade host defenses (1). A different type of productive non-standard polymerase action involves realignment of the template:product hybrid at a slippage-prone sequence to yield product with extra or fewer base(s) than present in the corresponding template sequence (2). This has been studied with DNA-dependent DNA polymerases, DNA-dependent RNA polymerases, RNA-dependent RNA polymerases and RNA-dependent DNA polymerases (reverse transcriptases, RTs).

Evolutionary selected transcription slippage is utilized in the expression of viruses such as the Paramyxoviruses, *Sendai virus* and *Parainfluenza virus* (3,4), the Filovirus, *Ebola virus* (5–7), the large *Potyviridae* family (8–10), chromosomal genes such as *Thermus thermophilus dnaX* (11), numerous genes in an endosymbiont (12), a variety of bacterial Insertion Sequences (13–16), several medically important plasmid genes of *Shigella flexneri* (17–19), and counterpart chromosomal toxin secretion genes in *Citrobacter rodentium* and *Yersinia pseudotuberculosis* (20). Further, an extensive bioinformatic analysis of bacterial genomes has revealed many candidates that have yet to be experimentally explored (13,15,16). Transcriptional indel errors are relevant to certain disease states (21–25) and maybe significant for aging (26,27).

One common bacterial type of transcriptional slippage-prone sequence involves 9 or more A's or T's (28); other repeats have also been analyzed (29). Dissociation of the nascent RNA from its template hybrid complement allows realigned pairing in either direction. A well known Paramyxovirus heteropolymeric slippage motif is composed of A's followed by G's with the identity of the mis-paired

*To whom correspondence should be addressed. Tel: +353 214 205 420; Fax: +353 238 855 147; Email: atkins@genetics.utah.edu or j.atkins@ucc.ie

© The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

base in the new re-aligned hybrid being important in determining slippage directionality (30,31). Nearly all work has focused on slippage involving a linear (unstructured) template. However, there is evidence that a protein roadblock or template structure ahead of a DNA-dependent RNA polymerase transcribing a slippage motif can stimulate realignment (2,32,33). Also there is one report of roadblock-mediated RT slippage where a polymerase bypasses an RNA-structure forming sequence prior to resumption of synthesis (34). The present work does not explore RT generation of product lacking sequence complementary to template sequence present in RNA structure.

Despite these studies of transcription slippage, and several studies of reverse transcriptase fidelity including (34–37), significant issues concerning RT mediated indel formation merit investigation. The use of RT as a lab reagent is one of the reasons why their slippage propensity is of interest (38). However, RTs do not contain 3′ exonuclease proofreading activity and their templates are prone to form structures at the ambient temperature at which these enzymes act. Better reagent polymerases have been developed from thermophilic DNA-dependent DNA polymerases by adapting their catalytic activity to function with RNA templates (39–41), or derived from existing RT polymerase by genetic engineering (42). Though the derived enzymes have the beneficial quality of lower base mis-incorporation due to higher accuracy for substrate selection (39), their indel fidelity remains to be explored.

The natural functional utilization of reverse transcriptase activities also enhances interest in deeper understanding of their propensity for indel formation. Reverse transcriptase activities are naturally essential for retroviruses and, retrotransposons, CRISPR spacer acquisition from RNA as a defense mechanism (43), and maintenance of chromosome ends (44). Further, the retron reverse transcriptase that yields msDNA (45) is significant for bacteria pathogenicity and colonization (46).

Here, we analyze the slippage propensity of different retroviral RTs as well as a retrotransposon counterpart. This study involves utilization of identical test sequences encompassing relevant stimulatory features and specific slippage-prone motifs. In addition, specific slippage candidate cassettes for natural RT slippage were also tested with their relevant RT enzymes.

The starting point for the present work was an unexpected result from a control for experiments in which reverse transcriptase slippage would confound the issue being addressed. A 6bp-stem 4nt-loop nascent transcript structure (here named 'model' stem–loop) stimulates *E. coli* DNA-dependent RNA polymerase transcriptional realignment at a 3′-A5G5–5′ motif which on its own is an inefficient slippage site (47). Analysis of the product RNA generated in that study involved reverse transcription by SuperScript™ III (derived from Moloney Murine Leukemia Virus RT). For the experiments included in that publication, the controls to distinguish whether indels in its product DNA derived from the initial DNA-dependent RNA polymerase step, or subsequently from product cDNA reverse transcriptase slippage, revealed no reverse transcriptase slippage. Follow-up work tested potential nascent RNA stem–loop structure stimulation of slippage at runs of A shorter than 9, the minimal needed for efficient slippage at such motifs. In this unpublished work, a significant proportion of the reverse transcriptase product of one 75 nt chemically synthesized RNA template with an inverted repeat with potential to form the 'model' stem–loop 5′ adjacent to 8 A's, had an extra T. This control experiment prompted the present investigation of RNA template stem–loop structure-mediated reverse transcriptase realignment.

MATERIALS AND METHODS

RNA template constructs

Preparation of RNA templates (quadruplex cassettes) with T7 RNA polymerase is described in Supplementary Methods. Chemically synthesized RNA templates and DNA oligonucleotides were from IDT-DNA (Supplementary Table S1).

Reverse transcription

Retroviral reverse transcriptase enzymes were purchased as follows: SuperScript™ III (Invitrogen), AMV (Biolabs), M-MulV (Biolabs), HIV-1 RT and HIV-2 RT (Abcam) and the retrotransposon TGIRT enzyme (InGex). In general when not indicated in the main text, RT reactions for SuperScript™ III, HIV-1, HIV-2 RT enzymes were with SuperScript™ III buffer 1X (50 mM Tris–HCl, pH 8.3 at 25°C, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT). For AMV (50 mM Tris–HCl pH 8.3 at 25°C, 75 mM KOAc, 8 mM $Mg(OAc)_2$, 10 mM DTT) and M-MulV (50 mM Tris–HCl pH 8.3 at 25 °C, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT). For the TGIRT reactions two alternative buffers were used. The buffer for the template switching reaction contained 450 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris–HCl pH 7.5 (48). The buffer for testing retrotransposon slippage candidate cassettes contained 75 mM KCl, 10 mM $MgCl_2$, 20 mM Tris–HCl pH 7.5 (49).

RT reactions with retroviral reverse transcriptases involved a pre-annealing step of the RNA template (100 ng): DNA Primer (2 pmol) (Supplementary Table S1), in the presence of the dNTP substrate (with the specific concentrations of each indicated in the main text), with the presence or absence of antisense where indicated (2, 20 or 200 pmol), in 10 μl reaction volumes. With wtSL or MUTsl RNA templates, incubation was at 65°C for 5 min before chilling on ice. For G-rich RNA templates with potential to form structure formation larger than that of the model stem–loop wtSL, the annealing mix had in addition 10 mM KCl and the annealing step was at 95°C for 30 s with a 1°C temperature decrease (from 95°C to 16°C) every 30 s. On completion one of several different 10 μl reaction mixes was added and incubated for 50 min at the temperature indicated. One reaction mix contained 100 units SuperScript™ III, 1X SuperScript™ III buffer and 20 mM DTT—this one was incubated at 52 °C. The AMV reaction mix contained 10 units of enzyme and 1× AMV buffer-incubation was at 37°C. The MuLV reaction contained 10 units of enzyme and 1× MuLV buffer-incubation 37 °C. The HIV-1 reaction mix contained 4 units enzyme (1.7 pmol), 1× SuperScript™ III buffer and 20 mM DTT-incubation 37°C. The HIV-2 reaction mix contained 0.2 units enzyme (1.7 pmol), 1× Super-

Script™ III buffer and 20 mM DTT-incubation 37°C. On completion a further incubation, which was at 85°C, followed for 5 min.

TGIRT RT reactions involving template switching are described in Supplementary Methods. Analysis of the candidate retrotransposon slippage cassettes was performed using a specific DNA primer complementary to the 3′ end segment of the test RNA. A mix of 100 ng RNA with 4 μl 10 μM specific primer and 10 μl 2× TGIRT 'low salt' buffer in a total volume of 18 μl, was incubated at 65°C for 5 min and chilled on ice. Then 1 μl 10 μM TGIRT enzyme was added. The premix was pre-incubated at room temperature for 30 min. Reaction was initiated by adding substrate dNTPs as indicated in the text and incubated for 10 min at room temperature. In the final 20 μl reaction mix, the final concentration of primer was 2 μM and of TGIRT enzyme was 500 nM. Then, 1 μl 5 M NaOH was added and incubated at 95°C for 3 min. It was neutralized with 1 μl 5 M HCl. cDNA was then purified with a silica-based column following the procedure described in Supplementary Methods. Elution was with 20 μl RNase free water (Supplementary Table S1).

Polymerase chain reaction

Each specific cDNA was amplified using the corresponding set of forward and reverse primers (Supplementary Table S1). Standard PCR reactions were 50 μl volume and contained: 1× Thermo buffer (Biolabs), 2 μl cDNA or 4 nM DNA oligo, 200 μM each dNTP (Biolabs), 500 nM each specific primer, and 0.8 unit Taq DNA polymerase (Biolabs). The PCR cycle was: denaturation at 94°C for 5 min, then 25 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 30 s and elongation at 72°C for 30 s. This was followed by a final elongation at 72°C for 1 min.

Limited primer extension

IRD700 fluorescent 5′-labeled oligonucleotides were from IDT DNA. The standard limited primer extension reaction was in 12.5 μl volume with 1× Thermo buffer (Biolabs), 12 nM of a specific IRD700-labeled fluorescent primer (IDT-DNA, Supplementary Table S1), a mix of 1 μM of three dNTPs with the missing dNTP replaced by the corresponding terminator chain reaction acydNTP (Biolabs) at 50 μM, and 0.6 unit of Vent exo-polymerase (Biolabs). The quantity of (RT)-PCR template was about three times lower than that of the fluorescent primer. On average each primer molecule is utilized on 20 occasions for chain extension during the 60-cycle PCR reactions. The PCR cycle was: denaturation at 94°C for 2 min, then 60 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 30 s. The final elongation was at 72°C for 2 min. In all cases each RT reaction and its subsequent analysis was repeated at least twice. Reaction products were analyzed on 15% sequencing gels. Image capture was performed with a LiCor Sequencer.

RESULTS

RNA template stem–loop is a key factor for reverse transcriptase slippage directionality on 7A's and 6A's

Initial experiments investigating a possible role for stem–loops in stimulating indel formation utilized SuperScript™ III, the widely used genetically engineered RT and two chemically synthetized 75 nt RNA constructs containing 7A's. These specified the WT, or a variant, of the 'model' RNA stem–loop structure 5′ adjacent to 7A's. The first construct 'wtSL-A7' has the WT sequence specifying the 'model' stem–loop 5′-GCGGGCgcaaGCCCGC-3′, with the potential of base pairing indicated in upper case. The second 'MUTsl-A7' has the 5′ side sequence of the stem substituted by complementary nt bases, i.e. from 5′-GCGGGC-3′ to 5′-CGCCCG-3′ to prevent potential formation of the model stem–loop structure (Figure 1A and B). RT reactions were performed with all dNTP equimolar at 500 μM. The cDNAs were then amplified by PCR with Taq polymerase to yield the 'RT-PCR products'. The controls for Taq polymerase slippage used two chemically synthetized 75 nt DNAs, whose sequence corresponds to that of the test RNA sequence, used as template for PCR amplification. This yields the 'PCR products' referred to below. Next, the two RT-PCR and the two PCR products were used as templates for Limited Primer Extension (LPE) analysis for detecting the addition or omission of a base(s) in the T/A-tract derived sequence. LPE reactions were performed with one primer whose sequence is complementary to the template sequence adjacent to the T-tract present in one of the two strands of the RT-PCR and PCR products [the other DNA strand has the corresponding A-tract]. The conditions of the LPE reaction enable the primer to be extended to the first template base position at which termination was arranged to occur by incorporation of an acyclic dGTP (acyGTP) base. This leads to efficient termination at the first base C of the template encountered by the polymerase during extension of the primer as the corresponding dGTP standard substrate is absent from the reaction (see Materials and Methods). The C at which LPE termination occurs is 5′ adjacent to the T-tract (other sites and acyclic dNTPs are used as controls in Supplementary Data). The length of the LPE product also depends on the occurrence of any indel in the T-tract motif. In absence of slippage of the DNA polymerases used for amplification of the chemically synthesized DNA (Taq polymerase control) and subsequently for generation of the LPE product (Vent exo$^-$ polymerase), a homogeneous length LPE product is expected. This is used as a length marker. Comparison of the pattern of the LPE product(s) generated from RT-PCR with the marker reveal specific RT polymerase slippage-mediated base indels (Figure 1C).

With the wtSL-A7 construct, reverse transcription using SuperScript™ III enzyme and all dNTPs present at 500 μM, showed strong realignment-mediated addition of an extra A (Figure 1D, lane 9) but no addition with the MUTsl-A7 construct, where the potential for base-pair formation is greatly diminished (Figure 1D, lane 18). The corresponding control LPE marker showed no slippage addition for both the WT (Figure 1C and D, lane 19) and mutant con-

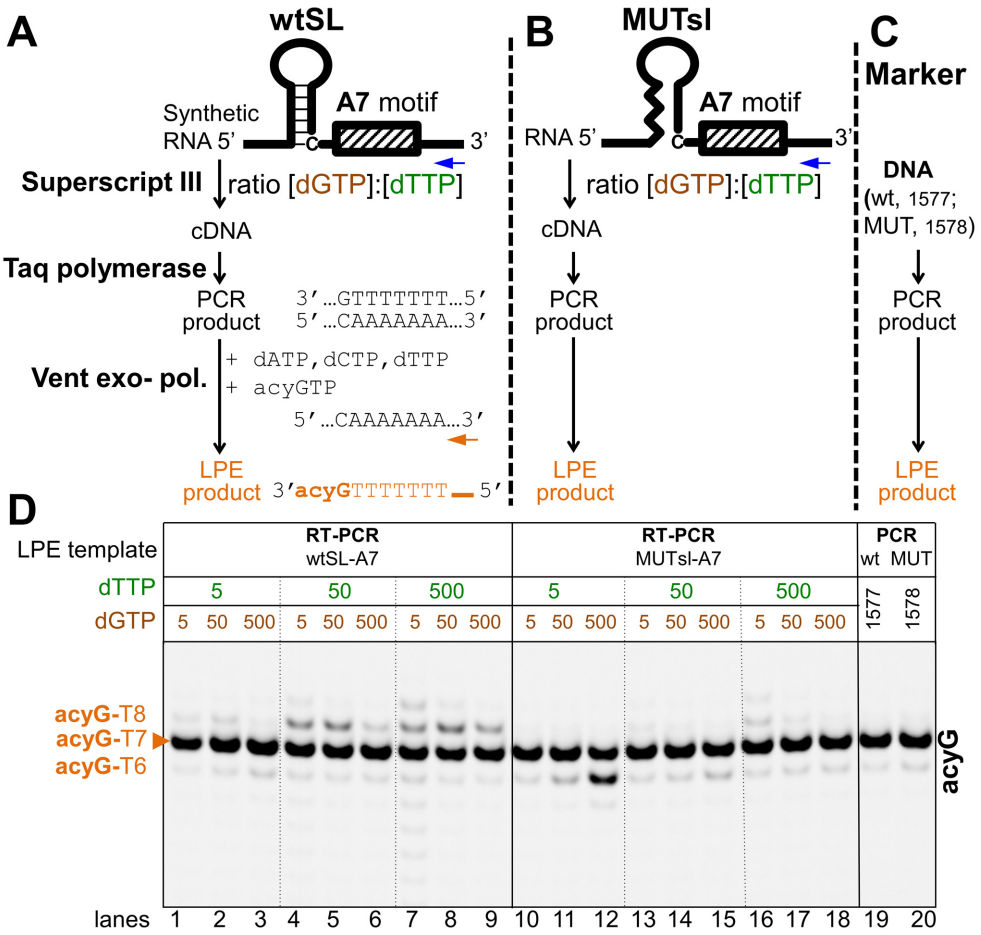

Figure 1. RNA stem–loop structure stimulator of reverse transcriptase slippage. Cartoon of chemically synthesized RNA template with presence (**A**) or absence (**B**) of stem–loop structure 5′ adjacent to an A-tract motif. The RT reactions involve varying the dGTP (in brown & specified by the base 5′ adjacent to the motif), dTTP (in green & the substrate specified by the motif), but keeping the dCTP and dATP concentrations constant (each at 500 μM). The subsequent analysis schemes are below each cartoon. LPE analysis (orange) involved a labeled primer that anneals 3′ adjacent to the slippage motif and terminates at the base 5′ adjacent to the motif. (**C**) LPE marker controls for DNA polymerase slippage. Chemically synthesized DNA counterparts of the corresponding RNA were used to generate a PCR product that served as template for LPE analysis. (**D**) LPE analysis with acyG terminators. Standard LPE products (whose synthesis did not involve slippage and reflect the original length of the motif in the chemically synthesized template) are indicated by an orange arrowhead.

structs (Figure 1C and D, lane 20). These LPE markers indicate that the DNA polymerase reagents (i.e. Taq and Vent exo-polymerases) are not responsible for the base addition. However, the wtSL-A7 and MUTsl-A7 constructs do show some omission of an A base (Figure 1D, lanes 9 and 18) with a similar signal detection level as the corresponding LPE markers (Figure 1D, lanes 19 and 20).

DNA-dependent RNA polymerase realignment is sensitive to the relative concentration of the substrate specified by the slippage site and by the DNA template base 5′ adjacent to it (47). To assess whether this also pertains with Reverse Transcriptase realignment, different dNTP concentration ratios were assayed. Nine dNTP ratio combinations with 5, 50 or 500 μM for the dTTP (specified by the A-tract slippage motif) and 5, 50 or 500 μM for the dGTP (specified 5′ adjacent to the template motif) were tested; the dATP and dCTP substrates were each present at 500 μM. Presence of the 'model' WT stem–loop stimulated addition of T (Figure 1D, lanes 4–9); this stimulation was increased with higher dTTP concentrations and higher ratios of [dTTP]:[dGTP] (Figure 1D, compare lane sets 1–3, 4–6 and 7–9). In the absence of the 'model' stem–loop, base omission of T was predominant. The most stimulatory dNTP condition was the lower ratio, 1:100, of [dTTP]:[dGTP], (Figure 1D, lane 12). At the highest dTTP concentration tested, a modest level of base addition is also observed but only at the highest ratio of [dTTP]:[dGTP] (lane 16). These results indicate that the RNA stem–loop is a strong stimulator for RT SuperScript™ III realignment directionality, promoting addition of an extra T in the cDNA, but not the omission of a T. Interestingly, in absence of the RNA stem–loop structure, realignment directionality is the inverse. This directionality difference indicates that the RNA stem–loop is also a strong inhibitor for omission of a base complementary to a template base. In summary, realignment efficiency and directionality is influenced by the relative dNTP concentrations and by RNA template structure.

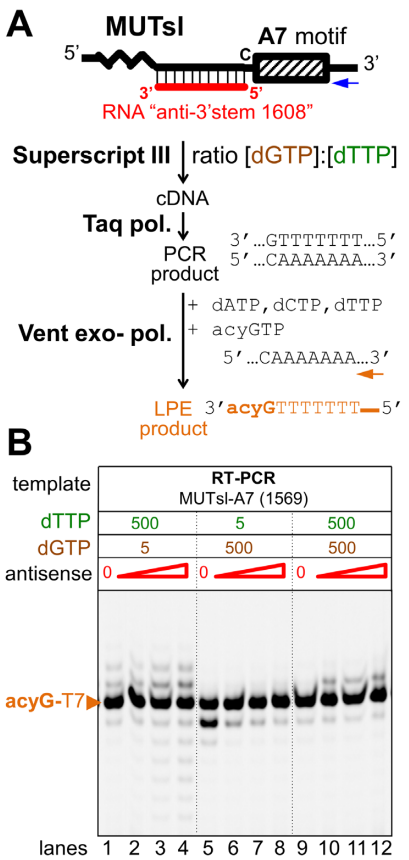

Figure 2. Inter-molecular structural counterparts of the stimulatory stem–loop for reverse transcriptase slippage. (**A**) Stimulation of slippage by an antisense (red) annealed 5′ adjacent to the A-tract motif. RT reactions were performed using 3 relative dNTP concentration conditions in the absence or presence, of titrated antisense RNA. (**B**) LPE analysis of the RT-PCR products derived from the cDNA template, and of the PCR products derived from the synthetic DNA template. Standard LPE products (whose synthesis did not involve slippage and reflect the original length of the motif in the chemically synthesized template) are indicated by an orange arrowhead.

As an alternative to the MUTsl-A7 construct whose potential for 'model' stem–loop structure formation is abolished by base substitution, we employed an RNA antisense strategy to decrease the potential for model stem–loop structure formation in the wtSL-A7 template. The result showed that presence of a 10 nt antisense RNA (anti-5′stem), complementary to the RNA sequence 9 nt 5′ to the 7A's, modestly decreases one base addition and enhances base omission. These experiments also showed that the efficiency and/or the directionality of the realignment are affected depending on the dNTP ratios (Supplementary Data and Figure S1).

To assess potential intermolecular stem–loop structure stimulatory action, we used an RNA antisense (anti-3′stem) complementary to the 10 nt sequence 5′ adjacent to the A7 motif in the MUTsl-A7' RNA construct (Figure 2A). The results, Figure 2B, show a strong effect of the antisense RNA on slippage directionality and efficiency. This antisense result with 'MUTsl-A7' is similar to the RT realignment without antisense with the 'wtSL-A7' construct (Figure 1). Increasing relative concentration of the antisense 'anti-3′stem' correlates: (i) at equimolar dNTP, with increasing base addition (Figure 2, lanes 9–12); (ii) at the lowest dNTP ratio (i.e. $[dTTP]_{5\mu M}$:$[dGTP]_{500\mu M}$), with a dramatically decreasing base omission of a base (lanes 5–8); (iii) at the highest dNTP ratio (i.e. $[dTTP]_{500\mu M}$:$[dGTP]_{5\mu M}$) with a slightly increasing base addition (lanes 1–4).

In conclusion, formation of an antisense RNA: template RNA hybrid 5′ adjacent to the motif, mimics the presence of an intramolecular RNA stem–loop structure with a similar effect on realignment directionality and efficiency.

RT catalytic center positioning

Formation of the RNA model stem–loop 5′ to the slippage motif should act as a physical roadblock for the transcribing RT polymerase on the A-tract. We first identified the minimal number of nucleotides 5′ of an A7 motif at which formation of the model stem–loop could stimulate slippage. Base C 5′ adjacent to the motif was maintained in all sequences. Derivatives of the 'wtSL-A7' construct were made with 1, 2 or 3 nt insertions between the stem–loop and the A-tract motif (Supplementary Figure S2, panel A). The LPE results showed that by increasing the distance between the model stem–loop and the A7 tract by just one nt, the stimulatory effect of the RNA stem–loop structure on base addition is abolished. The results also show that the inhibitory effect of the stem–loop on base omission is abolished as well. Base omission is now more sensitive to dNTP concentration ratio variation. This is most evident with a higher concentration of the dGTP substrate (specified by the template base 5′ adjacent to the A-tract motif), than that of the dTTP substrate (specified by the slippage motif) (Supplementary Figure S2A and C). The results with *E. coli* RNA polymerase generated RNA, showed that the model stem–loop 5′ adjacent to an A5 motif does not, at equimolar dNTP, stimulate SuperScript™ III-mediated base addition (data not shown). Interestingly, similar experiments using derivative constructs specifying the model stem–loop 0, 1, 2, or 3 nt 5′ to A5 motif, showed that though the RNA stem–loop does not stimulate base addition on A5, its inhibitory effect on base omission is present when the model stem–loop is 5′ adjacent to the A5 (Supplementary Figure S2B and C). The distance between the 'road-blocking' structure formation and the A7 slippage motif was also explored by antisense RNA experiments (Supplementary Results and Supplementary Figure S3).

To summarize, intra- or intermolecular 'stem' structures need to be 5′ adjacent to the re-alignment motif for optimal stimulation of base addition. They also need to be 5′ adjacent for maximal inhibition of base omission. Taken together, the results show that at the time of productive realignment, the catalytic center of RT is mostly located at the template position 3′ adjacent to the 'stem' structure.

G-rich sequences are also strong stimulators for slippage

To explore potentially relevant properties of G-rich sequences, four dsDNA constructs were made (Supplementary Methods). RNA generated from these with T7 RNA polymerase had the sequence GGCGGCGGCGG 5′ adjacent to the A7 motif or separated from it by 1, 2 or 3 nt (C,

UC or UUC) (Figure 3A). In the 5′ leader (UTR) of eukaryotic initiation factor-4A (eIF4A) mRNA this sequence forms an RNA quadruplex (50). However, the structure potentially formed in the transcripts utilized here could be different due the potential for pairing involving the U and C, where present, in the spacer, and was not explored.

LPE analysis was performed with primer, R_821, complementary to the sequence immediately adjacent to the A-tract in one DNA strand of the (RT)-PCR product. An acyC terminator mediates LPE termination at the site specified by the template base position underlined in the sequence 5′-G-spacer-A's-3′ (Figure 3B, left). The varied spacer lengths (0,1,2,3) determine the staggered LPE product sizes, markers, seen on the gel. The shift is related to the 1 nt length difference of the spacers involved (Figure 3B, PCR). With SuperScript™ III RT, at equimolar dNTP the RT-PCR derived LPE products from all four constructs contain detectable base addition (Figure 3B, lanes 1–4). With dNTP ratio conditions that favor base addition, both the efficiency of addition and number of bases added, increase with spacer length extensions (Figure 3B, lanes 5–8). In contrast, with dNTP ratio conditions that favor base omission, both the efficiency of base absence and number of bases missing, decreases with spacer length extensions (Figure 3B, lanes 9–12).

RT experiments have also been performed with RNA template variants of the eIF4A-derived G-rich sequence, two other G-rich sequences and their derivatives. With a subset, stimulatory effects are evident at specific relative dNTP concentration conditions (Supplementary Results and Supplementary Figure S4).

In conclusion, G- rich sequences can have a major impact on slippage and its directionality.

WT retroviral reverse transcriptases exhibit similar realignment

The reverse transcriptase from WT Moloney Murine Leukemia virus (MuLV), the parent of SuperScript™ III, plus the RTs from Avian Myeloblastosis virus (AMV) and from HIV-1 and HIV-2 were similarly tested with WTsl-A7, and mutSL-A7 under the nine dNTPs concentration conditions. In addition these RTs were tested with the WT, or mutated, 'model' stem–loop 5′ adjacent to an A6 motif (wtSL-A6 and MUTsl-A6) under equimolar dNTP concentration condition. The results show a similar LPE product pattern indicating that at identical RNA template and dNTP concentration conditions, the different RT polymerases tested share a clear similar response to slippage directionality. However, for specific reaction conditions, they can show marked differences in their slippage propensity (Supplementary Results and Supplementary Figure S5).

A retrotransposon RT mediates efficient slippage

Thermostable RTs encoded by group II introns from thermophilic bacteria are proving very useful for next generation RNA sequencing (49) and one of them, TGIRT, is commercially available and becoming widely used because of its thermostability (60°C) and advantageous template switching. TGIRT was first tested using the constructs specifying the WT or mutated 'model' stem–loop 5′ to the A7 slippage motif. As described more fully in Methods, the experimental conditions involved attachment of a preformed 41 bp DNA:RNA hybrid that is utilized as primer for reverse transcription of the test template by the TGIRT enzyme. The hybrid contained a one base overhang at the 3′ end of the DNA. It is complementary to the base at the 3′ end of the RNA test construct. The overhang base is utilized by the RT enzyme to switch from the RNA of the hybrid to the RNA test template. Such template switching (48) is utilized in preparation of samples for deep sequencing. The buffer conditions used for the preparation of the cDNA for deep sequencing were the same as used here for the study of TGIRT reagent slippage.

The first set of experiments was with WTsl-A7 and mutSL-A7constructs. Reactions were performed using three dNTP concentration conditions. RT reactions were performed with 3 dNTP ratio conditions for the substrates: (i) all 4 dNTPs at 1.25 mM, (ii) dTTP 12.5 μM, other 3 at 1.25 mM, (iii) dATP at 12.5 μM other 3 at 1.25 mM. The results show that TGIRT also responds to RNA template structure and specific dNTP concentration ratio. However, for its slippage-mediated base addition the range of the number of extra nucleotides was much greater and was from 1 to 50 nt (Supplementary Results and Supplementary Figure S6, panels A–C)

To determine the potential importance of the identity of the RNA template base, C, 5′ adjacent to the A7 motif in wtSL-A7 (= wtSL/C-A7) and MUTsl-A7 (= MUTsl/C-A7), the C was substituted by G to give the constructs 'wtSL/G-A7' and 'MUTsl/G-A7'. Also in the WT construct a compensatory base substitution was made in the sequence specifying the 5′ base of the 5′ side of the stem to maintain base pairing (Figure 4A). In the MUT construct a corresponding substitution to preclude base pairing was not necessary as its potential partner is already G (Figure 4B). The second set of constructs 'wtSL/U-A7' and 'MUTsl/U-A7' is as the first set except for the base adjacent to the motif being C with corresponding compensatory base substitutions (A and U respectively) to maintain (wt), or to abolish (MUT), stem–loop structure formation (Figure 4C and D). RT reactions were performed using three specific dNTP concentration ratios (1:1, 100:1 and 1:100) for the dNTP substrate specified by the slippage motif and the RNA base adjacent to the motif. LPE analysis showed a similar result as obtained with the WT and mut 'stem–loop' model structure where the last base of the sequence specifying the 3′ side of its stem has the base C 5′ adjacent to the A7 motif. The RT slippage followed the dNTP ratio 'rules' where higher substrate concentration specified by the slippage motif stimulates base addition (Figure 4E, lanes 2, 5, 8 and 11), and where higher substrate concentration specified by the template base 5′ adjacent to the motif, stimulates base omission (Figure 4E, lanes 3, 6, 9 and 12). The RT slippage also followed the rules of slippage directionality involving potential formation of the RNA structure 5′ of the motif. With the wtSL constructs, base addition is stimulated, and with the MUTsl constructs it is inhibited (Figure 4E, compare lane sets 1–2 with 7–8, and 4–5 with 10–11). In contrast, slippage omission of at least one base is favored in the ab-

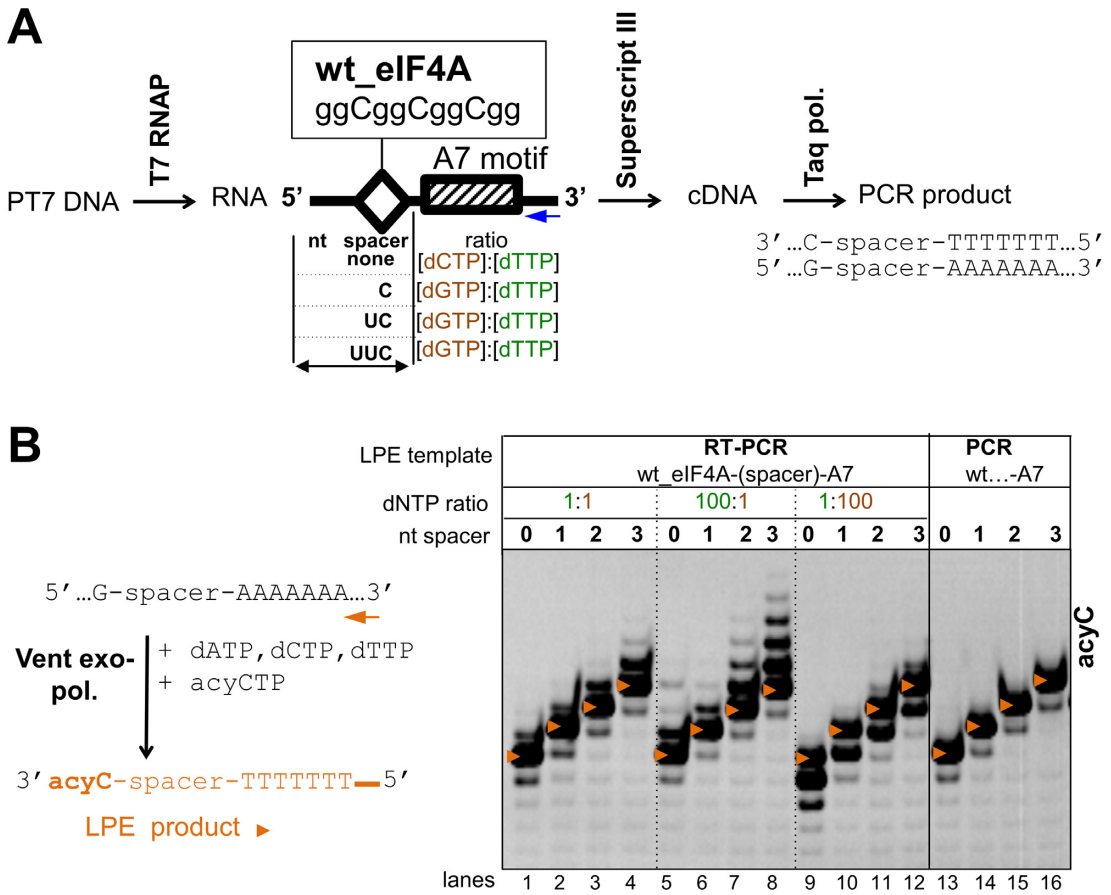

Figure 3. G-rich sequences are slippage stimulators. (**A**) Cartoon of the RNA templates generated *in vitro* using T7 RNA polymerase ('T7 RNAP') and PCR product template ('PT7 DNA'). Constructs contain a wild type (wt) G-rich sequence from eIF4A mRNA 5′ to the A7 motif. The nt spacing distance (from 0 to 3 nt) is between the last (3′) G of the G-rich sequence and the A7 motif (bottom). Combination of the G-rich sequence with different spacer lengths was tested as indicated in (**B**). LPE analysis of the RT-PCR product derived from the cDNA template, and of the PCR product derived from the 'PT7 DNA' template. Standard LPE products (whose synthesis did not involve slippage and reflect the original length of the motif in the chemically synthesized template) are indicated by an orange arrowhead.

sence of potential for RNA template stem–loop formation (Figure 4E, compare lane 9 with 3, and lane 12 with 6). In conclusion, the above result shows that the realignment for the TGIRT enzyme is independent of identity of the base located 5′ adjacent to the motif.

Next, we analyzed the stimulatory effect of the RNA road-blocking 'model' structure 5′ to A6 and to the U6 motifs (Supplementary Figure S6A and B). RT reactions were also performed using three specific dNTP ratios (1:1, 100:1 and 1:100) for the dNTP substrate specified by the slippage motif and the RNA base adjacent to the motif. LPE analysis showed a similar slippage pattern for A6 as shown for the A7 motif and it also followed the slippage rules involving dNTP ratio and potential RNA template structure formation (Supplementary Figure S6, panels D and E with A6 motif). Interestingly, slippage occurs with the U6 motif and follows the 'slippage rules' (Supplementary Figure S6, panels D and E with U6 motif).

In conclusion, these results show that the non-retroviral RT enzyme behaves similarly to the retroviral RT enzyme in terms of template structure and dNTP influences, although the number of bases inserted by TGIRT enzyme slippage is dramatically higher, ranging up to more than 50 bases instead of just 1.

Retrotransposon *gag-pol* slippage candidates

A bioinformatic analysis of LTR retrotransposons revealed several that may utilize recoding in synthesis of their Gag-Pol, with some being candidates for utilization of transcription slippage (51). We selected three of these candidates for *in vitro* testing of TGIRT enzyme slippage during reverse transcription of cassettes. In the two *Drosophila melanogaster* candidates tested *pol* was in the –1 frame with respect to *gag*, whereas in the third candidate, which was from maize (*Zea mays*), its *pol* was in the +1 frame with respect to *gag*. *Drosophila* candidate Dme1_ChrX_2630566 has the motif 5′-AU6-3′ and was tested with a chemically synthetized RNA containing 22 nt 5′ and 26 nt 3′ to the motif (Figure 5, A). Candidate Dme1_Chr3_26087113 has the motif 5′-GA4U4-3′ and the chemically synthetized RNA to test it contained 18 nt 5′ and 32 nt 3′ of the motif (Supplementary Figure S7A and C). To more closely resemble physiological conditions, in these reactions the TGIRT-mediated reverse transcription was performed at room temperature

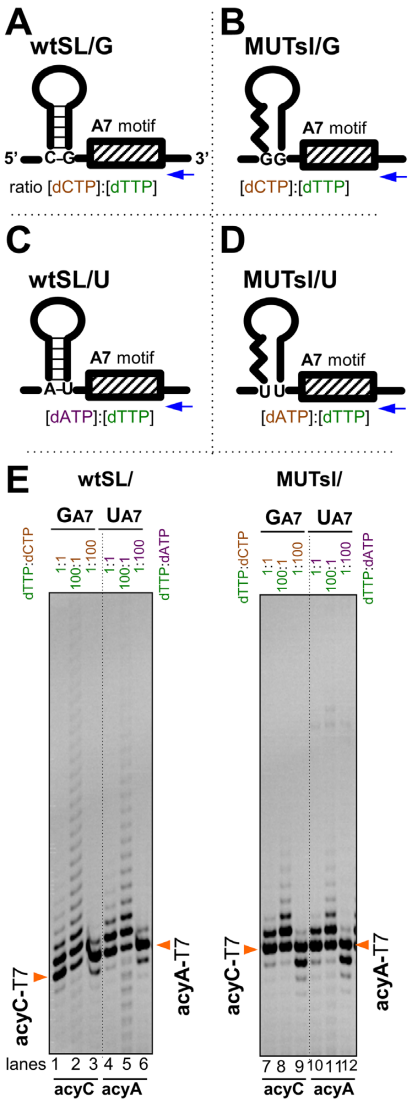

Figure 4. Retrotransposon TGIRT enzyme slippage. Chemically synthesized RNA template with variants of the model stem–loop (**A**, **C**) or no stem–loop structure (**B**, **D**) 5′ adjacent to an A7 slippage motif. The variants differ from the model stem–loop by the identity of the RNA base 5′ adjacent to the A7 with compensatory substitution to maintain base pairing at the bottom of the stem (A, C), or prevent base pairing at that position (B, D). *In vitro* TGIRT RT reactions were performed with three dNTP ratio concentration conditions between the substrate specified by the motif and that specified by the base 5′ adjacent to the motif (12.5 μM, 125 μM or 1.25 mM). The corresponding substrates are noted under their respective constructs and were at either 12.5 μM, 125 μM or 1.25 mM each (with each of the two other dNTPs at 1.25 mM). (**E**) LPE analysis of the RT-PCR products was performed with similar strategy as shown in Figure 1. Standard LPE products (whose synthesis did not involve slippage and reflect the original length of the motif in the chemically synthesized template) are indicated by an orange arrowhead.

and in low salt buffer (this differed from the 60°C and higher salt conditions utilized in the switching template experiment above). The RT reaction was performed with a sequence specific primer for each test candidate cassette. Each candidate was tested with 3 specific dNTP ratios (1:1, 100:1 and 1:100) for the dNTP substrate specified by the slippage motif and the RNA base 5′ adjacent to the motif. LPE analysis showed slippage for the candidate Dme1_ChrX_2630566 having the U6 tract in the RNA template: efficiency and distribution follow the dNTP rule for slippage (Figure 5B). Candidate Dme1_Chr3_26087113 showed no slippage (Supplementary Figure S7C).

The Maize candidate (gi_7262818_71383_R) has the motif 5′-UA4C3-3′. This candidate contains a conserved RNA template forming structure specified by 25 nt 5′ to the A4C3 motif (51), that is a candidate *cis*-acting RNA road-blocking element for stimulation of RT slippage. LPE analysis showed no relevant slippage for the (sub)-motif AC3 (Supplementary Figure S7B and D, with acyT LPE reaction) but showed marginal slippage-mediated addition of one base under all dNTP condition indicating that the A4 motif in the sequence UA4C3 is a poor but 'active' slippage-prone motif (Supplementary Figure S7B and D with acyA LPE reaction).

DISCUSSION

Two common features are evident for RT slippage directionality by all RT polymerases tested. The 'dNTP rule' is that a higher concentration of the cognate substrate specified by the template base 5′ adjacent to the slippage motif, than of the substrate specified by the motif, favors slippage-mediated base omission. When the ratio is reversed, slippage-mediated base addition is favored. The concentration of each dNTP used in our RT reaction with retroviral RT polymerase is in the μM range whereas with the Retrotransposon-derived TGIRT polymerase, the highest ratio is in the mM range. The equimolar dNTP conditions used in the present *in vitro* work are similar to those used in cDNA preparation for NGS analysis. The dNTP imbalance conditions used in the present work were relatively high, 10-fold. Even with ca. 3-fold dNTP imbalance substantial effects on DNA polymerase fidelity have been detected *in vivo*. Notably, the phenotype associated with an analogue of a colorectal cancer causing DNA polymerase mutator mutant is due to its causing an S-phase checkpoint-dependent elevation of dNTP pools (52). Other results also point to important correlations of dNTP pool levels with DNA polymerase mutator activities (53), and mutants of a deoxyribonucleoside triphosphate triphosphohydrolase that influence the level and balance of dNTP pools, are frequent in colon cancer cells (54).

With reverse transcriptases, road-blocking involves the template structure being crucial for slippage directionality because it limits RT polymerase access to the template base 5′ adjacent to the motif. This effect is irrespective of base identity. Road-blocking stimulates slippage-mediated base addition and inhibits slippage-mediated base omission. How does road-blocking and the 'dNTP' rule influence in which direction the RT polymerase will slip?

Standard synthesis of the cDNA transcript is achieved by successive nt addition at the 3′ end of the cDNA transcript. The RT is in its pre-translocated state when incorporation of the substrate occurs to yield the 3′ end of the cDNA in the polymerase's catalytic center (Figure 6A, C and E). After RT translocates one nt forward to the next template base, and its catalytic center is free of the cDNA 3′ end, the RT is in its post-translocated state (Figure 6B and

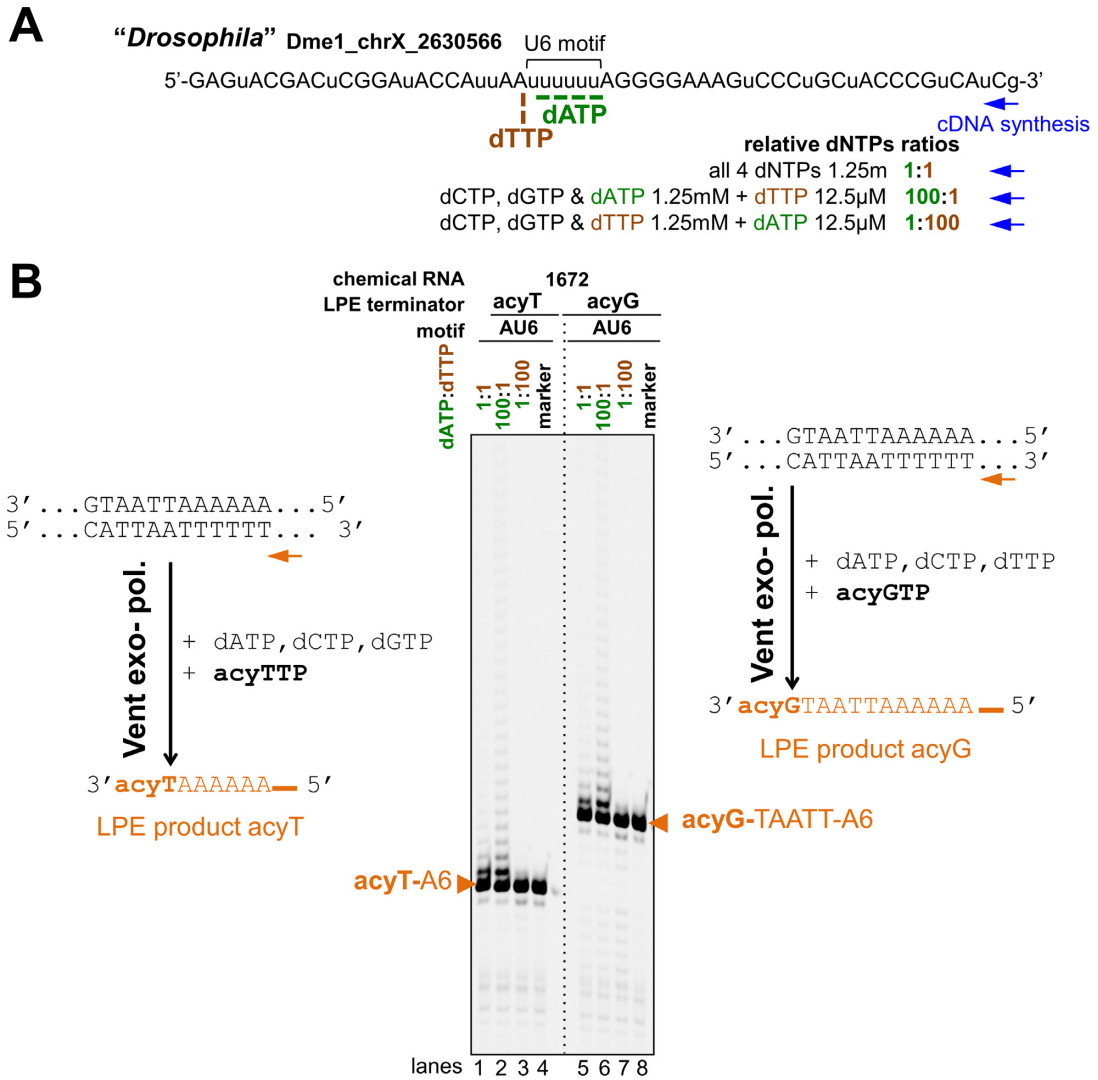

Figure 5. Retrotransposon slippage candidate genes. (**A**) Sequence of the *Drosophila* frameshifting candidate Dme1_ChrX_2630566 with motif 5′-AU6A-3′ (51). This RNA was chemically synthesized. (**B**) LPE analysis of the RT-PCR (using cDNA as template) and PCR (using synthetic DNA as template) products, were performed with similar strategy as shown in Figure 1 using specific primers. Standard LPE products (whose synthesis did not involve slippage and reflect the original length of the motif in the chemically synthesized template) are indicated by an orange arrowhead.

D). In absence of substrate in the catalytic centre, RT can oscillate between post-translocation and pre-translocation conformations (Figure 6, sets A and B, C and D).

When the leading edge of the RT polymerase encounters relevant template RNA structure (Figure 6, H), its progression is restrained at the pre-translocation state (Figure 6C) increasing the propensity for backward realignment of the cDNA 3′ end. When the cDNA: RNA hybrid contains an appropriately positioned slippage-prone sequence there is pairing potential in the backward realigned cDNA (at least its 3′ end). Such a 1 nt backward realignment would mimic the situation where the RT polymerase is in the post-translocation state (Figure 6G). The template base in the RT catalytic centre is available for pairing with the substrate, and so productive substrate incorporation that yields slippage-mediated base addition (Figure 6, from G to C after incorporation of the substrate).

In the absence of template RNA structure, when RT polymerase transcribes a 'slippery' sequence the cDNA:RNA hybrid is prone to realign in either direction. cDNA backward realignment is at a lower level than when a template structure is at the leading edge of the polymerase (Figure 6, from C to G). cDNA forward realignment involves RT polymerase in the post-translocation state being transformed to the pre-translocation state (Figure 6, from B to F). Evidence for this assertion comes from several aspects of the results. Firstly, high relative substrate concentration specified by the motif base inhibits slippage that generates product lacking a base compared to the template. This is perhaps due to pairing of substrate base with the template base at the catalytic center as it would prevent pairing of the forward realigned cDNA 3′ end base (Figure 6B). Secondly, the finding that relatively higher concentration of the substrate, specified by the template base 5′ adjacent to the motif, strongly stimulates base omission, is explicable by re-

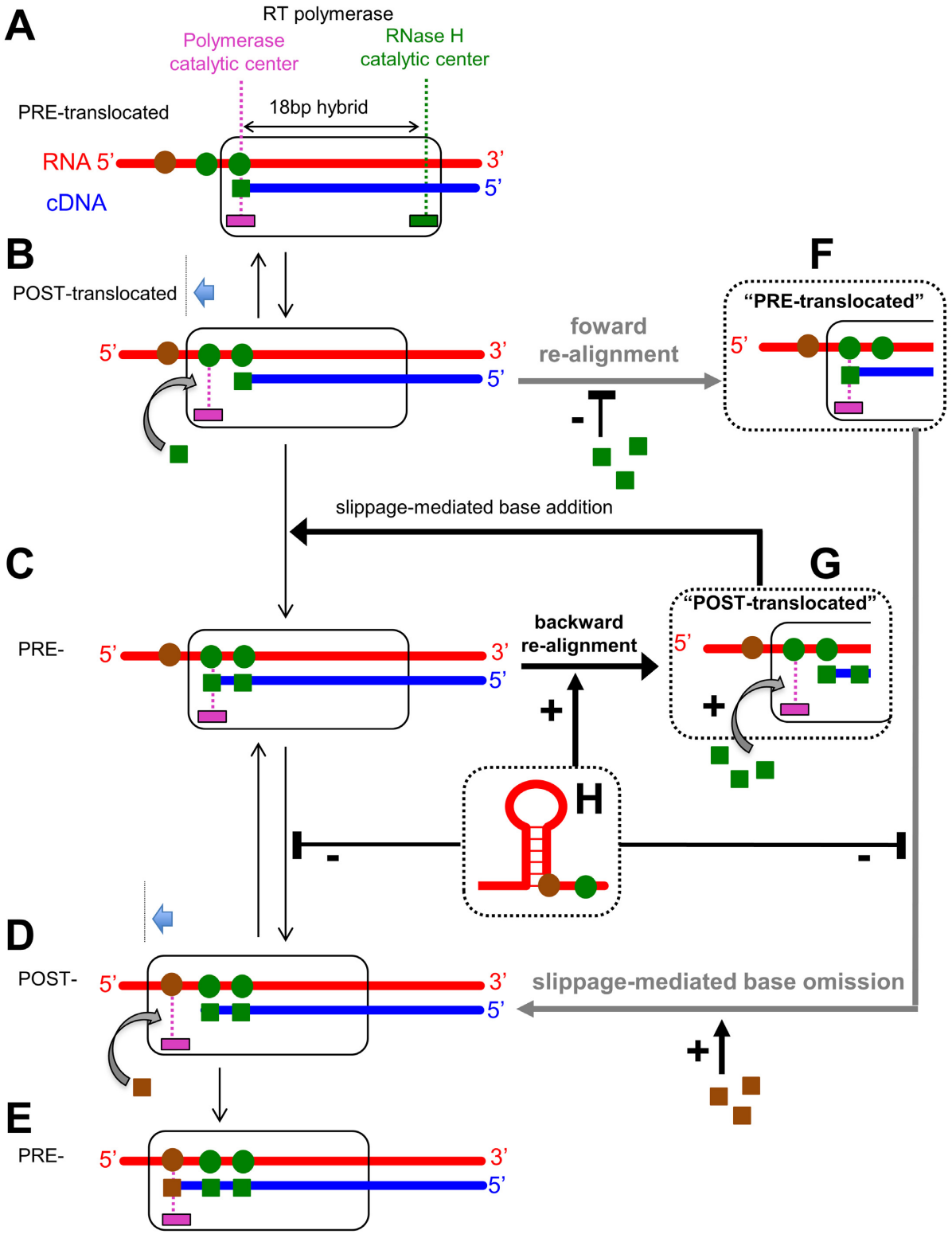

Figure 6. Model of RNA template structure influence on RT slippage. RT enzyme (open black rectangle) with the RNA template (red) and the nascent cDNA (blue). The polymerase and RNase H catalytic centers are pink and green rectangles respectively. The two 5′ bases of the RNA slippage motif are indicated by green closed circles and the base 5′ adjacent to the motif is indicated by a brown closed circle. Their corresponding cognate substrate is indicated with green and brown closed squares respectively. Inhibition and stimulation effect are indicated by – and + symbols. Standard RT transcription (**A**–**E**). Forward realignment-mediated base omission (B–F) occurs from a polymerase Post-translocation state in the absence of the cognate substrate in the catalytic center, and following polymerase forward translocation (F–D) is productively locked by incorporation of the substrate specified by the next template base (D–E). Backward realignment-mediated base addition (C–G) occurs from a polymerase Pre-translocation state stimulated by the formation of the template structure (H), and is productively locked by incorporation of the cognate substrate (G–C).

alignment locking following this template base locating in the catalytic center (Figure 6 from F to D). Consistent with this, availability of the base 5′ adjacent to the RNA motif is crucial for base omission since its sequestration in template stem pairing (intra-template or antisense pairing) (Figure 6H) inhibits base omission (Figure 6, from F to D). Accordingly, a requisite for productive forward realignment is that prior to its occurrence, the RT polymerase is in a post-translocation state with the 3′ end of the cDNA being base paired to the template base second from the 5′ end of the RNA motif (Figure 6B). Even though substrate base pairing serves to lock realigned hybrid pairing, the potential for reversal of realignments is indicated by the strong effect of relative dNTP concentration on modulating slippage directionality. This implies that substrate pairing is slower than realigned hybrid formation.

Our 'roadblock stimulatory RT slippage' model highlights the dependence of productive slippage on RT polymerase translocation state and cognate substrate incorporation. Though this model involves realignment of the 3′ end of the cDNA in either direction with respect to the template, it does not give any explanation about the trigger for the realignment process. A previous model for HIV-1 RT slippage-mediated base omission involves 'template-strand slippage' with a separate model for HIV-1 RT base addition involving 'primer-strand slippage' (35). These models feature either potential formation of an extrahelical base (a bulge) or base sharing. WT HIV-1 RT enzyme has a five-fold greater efficiency for base addition than base omission (35). Substitution of HIV-1RT Glu89 by other residues leads not only to a decrease of slippage-mediated base omission in favor of an increase of base addition, but it also leads to a decrease of base substitution errors. Glu89 is in close proximity to the sugar phosphate backbone of the template strand near the penultimate base pair. Its importance for fidelity and slippage directionality has been suggested to be due to increased dNTP binding pocket stability (35). An alternative explanation linked to our model, is that WT and position 89 variants differentially favor the pre- and post-translocation states. Knowledge of a possible correlation between WT and position 89 variants and both HIV-1 RT translocation state and slippage directionality, would be informative.

Potential utilization of structure stimulated slippage

The results here show that a cassette from *Drosophila* retrotransposon Dme1_chrX_2630566 containing an AU6 motif, exhibits strong slippage with sensitivity to relative dNTP concentration conditions. In addition, a cassette with a Maize retrotransposon sequence that has conserved potential for template stem–loop structure formation 5′ to a motif A4C3, showed marginal slippage-mediated addition of T. Given the widespread occurrence and importance of retrotransposons, these results highlight the need for systematic studies to reveal the extent of their functional utilization of RT slippage.

Replication of the single-stranded, positive sense, RNA genome of SARS Coronavirus involves a viral-encoded RNA-dependent RNA polymerase. Polymerase expression involves -1 ribosomal frameshifting at a U-UUA-AAC sequence (55–57). Together with 5′ bases, it is part of a GU5A3C sequence. Interestingly, a potential 10 bp-stem 4 nt-loop structure forms 2 nt 5′ to the GU5A3C sequence and causes reduced frameshift-derived product (58). During replication of the (+) strand such a stem–loop would be ahead (5′) of the U5A3 motif. This raises the possibility of it leading to road-block-induced slippage at the U5A3 motif, and so being a counterpart of the situation shown for the HIV frameshift site. The potential for HIV functional utilization of RT and its implications are considered in the accompanying ms (59).

The finding of RNA G-rich sequence stimulated RT slippage is of interest and its possible extension to DNA-dependent RNA polymerase slippage merits investigation. The widespread distribution of G-rich sequence in RNA has implications for the common use of reverse transcriptase in generating cDNA for deep-sequencing.

Interest in the potential of synthetic compensatory frameshifting near the sites of frameshift mutations to ameliorate a subset of genetic disease, prompted the testing of complementary oligonucleotides for frameshift stimulatory effects (60–63). Whether sequences that can bind to DNA, such as CRISPR-cas nickase mutants (64,65), would create a counterpart partial 'roadblock' structure for slippage stimulation, merits future work.

Perspectives

The results highlight the need for caution before assuming that RT products faithfully reflect template sequence. This caution extends to TGIRT. Though it is known to cause a very low rate of base substitution errors, nevertheless in the present work exhibits the highest level of slippage errors.

Extrapolating from the polymerase properties identified here to other polymerases, the recent increase in the modest number of known occurrences of productive utilization of transcription slippage for enriching gene expression, seems set to further increase. More generally, it extends awareness of the potential for template structure to stimulate slippage by diverse types of polymerase, and permits further parallels between context features that promote ribosomal frameshifting and transcription slippage.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Drs G. Loughran for facilitating the work, M. O'Connell Motherway, T. Cross and E. Dillan for help with LiCor sequencing. It is a pleasure to acknowledge constructive comments from Alan J. Herr and several anonymous reviewers.

FUNDING

Irish Research Council fellowship to C.P.; Science Foundation Ireland grants [12/IP1492, 13/1A/1853 to J.F.A., 12/1A/1335 to P.V.B., 12/RC/2273 to D.v.S.]. Funding for open access charge: Science Foundation Ireland.

Conflict of interest statement. None declared.

REFERENCES

1. Guo,H., Arambula,D., Ghosh,P. and Miller,J.F. (2014) Diversity-generating retroelements in phage and bacterial genomes. *Microbiol. Spectr.*, **2**, doi:10.1128/microbiolspec.MDNA3-0029-2014.
2. Anikin,M., Molodtsov,V., Temiakov,D. and McAllister,W.T. (2010) Transcript slippage and recoding. In: Atkins,JF and Gesteland,RF (eds). *Recoding: Expansion of Decoding Rules Enriches Gene Expression*. Springer, NY, pp. 409–432.
3. Kolakofsky,D., Roux,L., Garcin,D. and Ruigrok,R.W.H. (2005) Paramyxovirus mRNA editing, the 'rule of six' and error catastrophe: a hypothesis. *J. Gen. Virol.*, **86**, 1869–1877.
4. Kolakofsky,D. (2016) Paramyxovirus RNA synthesis, mRNA editing, and genome hexamer phase: a review. *Virology*, **498**, 94–98.
5. Sanchez,A., Trappier,S.G., Mahy,B.W.J., Peters,C.J. and Nichol,S.T. (1996) The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 3602–3607.
6. Shabman,R.S., Jabado,O.J., Mire,C.E., Stockwell,T.B., Edwards,M., Mahajan,M., Geisbert,T.W. and Basler,C.F. (2014) Deep sequencing identifies noncanonical editing of Ebola and Marburg virus RNAs in infected cells. *MBio*, **5**, e02011.
7. Volchkova,V.A., Dolnik,O., Martinez,M.J., Reynard,O. and Volchkov,V.E. (2015) RNA editing of the *GP* gene of Ebola virus is an important pathogenicity factor. *J. Infect. Dis.*, **212**, S226–S233.
8. Olspert,A., Chung,B.Y., Atkins,J.F., Carr,J.P. and Firth,A.E. (2015) Transcriptional slippage in the positive-sense RNA virus family Potyviridae. *EMBO Rep.*, **16**, 995–1004.
9. Rodamilans,B., Valli,A., Mingot,A., San Leon,D., Baulcombe,D., Lopez-Moya,J.J. and Garcia,J.A. (2015) RNA Polymerase slippage as a mechanism for the production of frameshift gene products in plant viruses of the Potyviridae family. *J. Virol.*, **89**, 6965–6967.
10. Hagiwara-Komoda,Y., Choi,S.H., Sato,M., Atsumi,G., Abe,J., Fukuda,J., Honjo,M.N., Nagano,A.J., Komoda,K., Nakahara,K.S. *et al.* (2016) Truncated yet functional viral protein produced via RNA polymerase slippage implies underestimated coding capacity of RNA viruses. *Sci. Rep.*, doi:10.1038/srep21411.
11. Larsen,B., Wills,N.M., Nelson,C., Atkins,J.F. and Gesteland,R.F. (2000) Nonlinearity in genetic decoding: homologous DNA replicase genes use alternatives of transcriptional slippage or translational frameshifting. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 1683–1688.
12. Tamas,I., Wernegreen,J.J., Nystedt,B., Kauppinen,S.N., Darby,A.C., Gomez-Valero,L., Lundin,D., Poole,A.M. and Andersson,S.G.E. (2008) Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 14934–14939.
13. Baranov,P.V., Hammer,A.W., Zhou,J.D., Gesteland,R.F. and Atkins,J.F. (2005) Transcriptional slippage in bacteria: distribution in sequenced genomes and utilization in IS element gene expression. *Genome Biol.*, **6**, R25.
14. Baranov,P.V., Fayet,O., Hendrix,R.W. and Atkins,J.F. (2006) Recoding in bacteriophages and bacterial IS elements. *Trends Genet.*, **22**, 174–181.
15. Sharma,V., Firth,A.E., Antonov,I., Fayet,O., Atkins,J.F., Borodovsky,M. and Baranov,P.V. (2011) A pilot study of bacterial genes with disrupted ORFs reveals a surprising profusion of protein sequence recoding mediated by ribosomal frameshifting and transcriptional realignment. *Mol. Biol. Evol.*, **28**, 3195–3211.
16. Antonov,I., Coakley,A., Atkins,J.F., Baranov,P.V. and Borodovsky,M. (2013) Identification of the nature of reading frame transitions observed in prokaryotic genomes. *Nucleic Acids Res.*, **41**, 6514–6530.
17. Penno,C., Sansonetti,P. and Parsot,C. (2005) Frameshifting by transcriptional slippage is involved in production of MxiE, the transcription activator regulated by the activity of the type III secretion apparatus in *Shigella flexneri*. *Mol. Microbiol.*, **56**, 204–214.
18. Penno,C. and Parsot,C. (2006) Transcriptional slippage in *mxiE* controls transcription and translation of the downstream *mxiD* gene, which encodes a component of the *Shigella flexneri* type III secretion apparatus. *J. Bacteriol.*, **188**, 1196–1198.
19. Penno,C., Hachani,A., Biskri,L., Sansonetti,P., Allaoui,A. and Parsot,C. (2006) Transcriptional slippage controls production of type III secretion apparatus components in *Shigella flexneri*. *Mol. Microbiol.*, **62**, 1460–1468.
20. Gueguen,E., Wills,N.M., Atkins,J.F. and Cascales,E. (2014) Transcriptional frameshifting rescues *Citrobacter rodentium* type VI secretion by the production of two length variants from the prematurely interrupted *tssM* gene. *PLoS Genet.*, **10**, e1004869.
21. Linton,M.F., Pierotti,V. and Young,S.G. (1992) Reading-frame restoration with an *apolipoprotein B* gene frameshift mutation. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 11431–11435.
22. Linton,M.F., Raabe,M., Pierotti,V. and Young,S.G. (1997) Reading frame restoration by transcriptional slippage at long stretches of adenine residues in mammalian cells. *J. Biol. Chem.*, **272**, 14127–14132.
23. Young,M., Inaba,H., Hoyer,L.W., Higuchi,M., Kazazian,H.H. and Antonarakis,S.E. (1997) Partial correction of a severe molecular defect in hemophilia A, because of errors during expression of the *factor VIII* gene. *Am. J. Hum. Genet.*, **60**, 565–573.
24. Laken,S.J., Petersen,G.M., Gruber,S.B., Oddoux,C., Ostrer,H., Giardiello,F.M., Hamilton,S.R., Hampel,H., Markowitz,A., Klimstra,D. *et al.* (1997) Familial colorectal cancer in Ashkenazim due to a hypermutable tract in *APC*. *Nat. Genet.*, **17**, 79–83.
25. Raabe,M., Linton,M.F. and Young,S.G. (1998) Long runs of adenines and human mutations. *Am. J. Med. Genet.*, **76**, 101–102.
26. Paoloni-Giacobino,A., Rossier,C., Papasavvas,M.P. and Antonarakis,S.E. (2001) Frequency of replication/transcription errors in (A)/(T) runs of human genes. *Hum. Genet.*, **109**, 40–47.
27. Atkins,J.F., Loughran,G., Bhatt,P.R., Firth,A.E. and Baranov,P.V. (2016) Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.*, **44**, 7007–7078.
28. Wagner,L.A., Weiss,R.B., Driscoll,R., Dunn,D.S. and Gesteland,R.F. (1990) Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*. *Nucleic Acids Res.*, **18**, 3529–3535.
29. Rockah-Shmuel,L., Toth-Petroczy,A., Sela,A., Wurtzel,O., Sorek,R. and Tawfik,D.S. (2013) Correlated occurrence and bypass of frame-shifting insertion-deletions (InDels) to give functional proteins. *PLoS Genet.*, **9**, e1003882.
30. Jacques,J.P., Hausmann,S. and Kolakofsky,D. (1994) Paramyxovirus messenger RNA editing leads to G deletions as well as insertions. *EMBO J.*, **13**, 5496–5503.
31. Hausmann,S., Garcin,D., Delenda,C. and Kolakofsky,D. (1999) The versatility of paramyxovirus RNA polymerase stuttering. *J. Virol.*, **73**, 5568–5576.
32. Broxson,C., Beckett,J. and Tornaletti,S. (2011) Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human *c-myb* gene. *Biochemistry*, **50**, 4162–4172.
33. Pavco,P.A. and Steege,D.A. (1991) Characterization of elongating T7 and SP6 RNA polymerases and their response to a roadblock generated by a site-specific DNA binding protein. *Nucleic Acids Res.*, **19**, 4639–4646.
34. Zhang,Y.J., Pan,H.Y. and Gao,S.J. (2001) Reverse transcription slippage over the mRNA secondary structure of the *LIP1* gene. *Biotechniques*, **31**, 1286–1294.
35. Hamburgh,M.E., Curr,K.A., Monaghan,M., Rao,V.R., Tripathi,S., Preston,B.D., Sarafianos,S., Arnold,E., Darden,T. and Prasad,V.R. (2006) Structural determinants of slippage-mediated mutations by human immunodeficiency virus type 1 reverse transcriptase. *J. Biol. Chem.*, **281**, 7421–7428.
36. Bebenek,K., Roberts,J.D. and Kunkel,T.A. (1992) The effects of dNTP pool imbalances on frameshift fidelity during DNA replication. *J. Biol. Chem.*, **267**, 3589–3596.
37. Boyer,J.C., Bebenek,K. and Kunkel,T.A. (1992) Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 6919–6923.
38. Fungtammasan,A., Tomaszkiewicz,M., Campos-Sanchez,R., Eckert,K.A., DeGiorgio,M. and Makova,K.D. (2016) Reverse transcription errors and RNA-DNA differences at short tandem repeats. *Mol. Biol. Evol.*, **33**, 2744–2758.
39. Ellefson,J.W., Gollihar,J., Shroff,R., Shivram,H., Iyer,V.R. and Ellington,A.D. (2016) Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science*, **352**, 1590–1593.

40. Yasukawa,K., Konishi,A., Shinomura,M., Nagaoka,E. and Fujiwara,S. (2012) Kinetic analysis of reverse transcriptase activity of bacterial family A DNA polymerases. *Biochem. Biophys. Res. Commun.*, **427**, 654–658.
41. Jozwiakowski,S.K. and Connolly,B.A. (2011) A modified family-B archaeal DNA polymerase with reverse transcriptase activity. *Chembiochem.*, **12**, 35–37.
42. Baranauskas,A., Paliksa,S., Alzbutas,G., Vaitkevicius,M., Lubiene,J., Letukiene,V., Burinskas,S., Sasnauskas,G. and Skirgaila,R. (2012) Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Eng. Des. Sel.*, **25**, 657–668.
43. Silas,S., Mohr,G., Sidote,D.J., Markham,L.M., Sanchez-Amat,A., Bhaya,D., Lambowitz,A.M. and Fire,A.Z. (2016) Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science*, doi:10.1126/science.aad4234.
44. Servant,G. and Deininger,P.L. (2015) Insertion of retrotransposons at chromosome ends: adaptive response to chromosome maintenance. *Front. Genet.*, doi:10.3389/fgene.2015.00358.
45. Inouye,M. (2017) The first demonstration of the existence of reverse transcriptases in bacteria. *Gene*, **597**, 76–77.
46. Elfenbein,J.R., Knodler,L.A., Nakayasu,E.S., Ansong,C., Brewer,H.M., Bogomolnaya,L., Adams,L.G., McClelland,M., Adkins,J.N. and Andrews-Polymenis,H.L. (2015) Multicopy single-stranded DNA directs intestinal colonization of enteric pathogens. *PLoS Genet.*, **11**, e1005472.
47. Penno,C., Sharma,V., Coakley,A., Motherway,M.O., van Sinderen,D., Lubkowska,L., Kireeva,M.L., Kashlev,M., Baranov,P.V. and Atkins,J.F. (2015) Productive mRNA stem loop-mediated transcriptional slippage: crucial features in common with intrinsic terminators. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E1984–E1993.
48. Katibah,G.E., Qin,Y.D., Sidote,D.J., Yao,J., Lambowitz,A.M. and Collins,K. (2014) Broad and adaptable RNA structure recognition by the human interferon-induced tetratricopeptide repeat protein IFIT5. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 12025–12030.
49. Mohr,S., Ghanem,E., Smith,W., Sheeter,D., Qin,Y.D., King,O., Polioudakis,D., Iyer,V.R., Hunicke-Smith,S., Swamy,S. *et al.* (2013) Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA*, **19**, 958–970.
50. Wolfe,A.L., Singh,K., Zhong,Y., Drewe,P., Rajasekhar,V.K., Sanghvi,V.R., Mavrakis,K.J., Jiang,M., Roderick,J.E., Van der Meulen,J. *et al.* (2014) RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature*, **513**, 65–70.
51. Gao,X., Havecker,E.R., Baranov,P.V., Atkins,J.F. and Voytas,D.F. (2003) Translational recoding signals between *gag* and *pol* in diverse LTR retrotransposons. *RNA*, **9**, 1422–1430.
52. Mertz,T.M., Sharma,S., Chabes,A. and Shcherbakova,P.V. (2015) Colon cancer-associated mutator DNA polymerase δ variant causes expansion of dNTP pools increasing its own infidelity. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E2467–E2476.
53. Williams,L.N., Marjavaara,L., Knowels,G.M., Schultz,E.M., Fox,E.J., Chabes,A. and Herr,A.J. (2015) dNTP pool levels modulate mutator phenotypes of error-prone DNA polymerase ϵ variants. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E2457–E2466.
54. Rentoft,M., Lindell,K., Tran,P., Chabes,A.L., Buckland,R.J., Watt,D.L., Marjavaara,L., Nilsson,A.K., Melin,B., Trygg,J. *et al.* (2016) Heterozygous colon cancer-associated mutations of SAMHD1 have functional significance. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 4723–4728.
55. Baranov,P.V., Henderson,C.M., Anderson,C.B., Gesteland,R.F., Atkins,J.F. and Howard,M.T. (2005) Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology*, **332**, 498–510.
56. Plant,E.P., Perez-Alvarado,G.C., Jacobs,J.L., Mukhopadhyay,B., Hennig,M. and Dinman,J.D. (2005) A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol.*, **3**, 1012–1023.
57 Brierley,I. and Dos Ramos,F.J. (2006) Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res.*, **119**, 29–42.
58. Cho,C.P., Lin,S.C., Chou,M.Y., Hsu,H.T. and Chang,K.Y. (2013) Regulation of programmed ribosomal frameshifting by co-translational refolding RNA hairpins. *PLoS One*, **8**, e62283.
59. Penno,C., Kumari,R., Baranov,P.V., van Sinderen,D. and Atkins,J.F. (2017) Specific reverse transcriptase slippage at the HIV ribosomal frameshift sequence: Potential implications for modulation of GagPol synthesis. *Nucleic Acids Res.*, doi:10.1093/nar/gkx690.
60. Howard,M.T., Gesteland,R.F. and Atkins,J.F. (2004) Efficient stimulation of site-specific ribosome frameshifting by antisense oligonucleotides. *RNA*, **10**, 1653–1661.
61. Olsthoorn,R.C., Laurs,M., Sohet,F., Hilbers,C.W., Heus,H.A. and Pleij,C.W. (2004) Novel application of sRNA: stimulation of ribosomal frameshifting. *RNA*, **10**, 1702–1703.
62. Henderson,C.M., Anderson,C.B. and Howard,M.T. (2006) Antisense-induced ribosomal frameshifting. *Nucleic Acids Res.*, **34**, 4302–4310.
63. Yu,C.H., Noteborn,M.H. and Olsthoorn,R.C. (2010) Stimulation of ribosomal frameshifting by antisense LNA. *Nucleic Acids Res.*, **38**, 8277–8283.
64. Luo,M.L., Mullis,A.S., Leenay,R.T. and Beisel,C.L. (2015) Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.*, **43**, 674–681.
65. Luo,M.L., Leenay,R.T. and Beisel,C.L. (2016) Current and future prospects for CRISPR-based tools in bacteria. *Biotechnol. Bioeng.*, **113**, 930–943.