10156–10167 *Nucleic Acids Research, 2017, Vol. 45, No. 17* *Published online 3 August 2017*
doi: 10.1093/nar/gkx690

Specific reverse transcriptase slippage at the HIV ribosomal frameshift sequence: potential implications for modulation of GagPol synthesis

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 13, 2017; Revised July 19, 2017; Editorial Decision July 21, 2017; Accepted July 24, 2017

ABSTRACT

Synthesis of HIV GagPol involves a proportion of ribosomes translating a U6A shift site at the distal end of the *gag* gene performing a programmed -1 ribosomal frameshift event to enter the overlapping *pol* gene. *In vitro* studies here show that at the same shift motif HIV reverse transcriptase generates -1 and +1 indels with their ratio being sensitive to the relative concentration ratio of dNTPs specified by the RNA template slippage-prone sequence and its 5′ adjacent base. The GGG sequence 3′ adjacent to the U6A shift/slippage site, which is important for ribosomal frameshifting, is shown here to limit reverse transcriptase base substitution and indel 'errors' in the run of A's in the product. The indels characterized here have either 1 more or less A, than the corresponding number of template U's. cDNA with 5 A's may yield novel Gag product(s), while cDNA with an extra base, 7 A's, may only be a minor contributor to GagPol polyprotein. Synthesis of a proportion of non-ribosomal frameshift derived GagPol would be relevant in efforts to identify therapeutically useful compounds that perturb the ratio of GagPol to Gag, and pertinent to the extent in which specific polymerase slippage is utilized in gene expression.

INTRODUCTION

Moloney Murine Leukemia retrovirus *gag* gene is 5′ adjacent to, and in the same frame as its *pol* gene, whereas the Rous Sarcoma and HIV counterpart *pol* genes are in the –1 frame with respect to their *gag* genes. Following the evidence that synthesis of Moloney Murine Leukemia Virus GagPol involves in-frame readthrough of the *gag* termination codon (1,2), attention focused on Rous Sarcoma and HIV GagPol synthesis. A ribosomal –1 frameshift event was shown to occur in the decoding of a sequence near the 3′ end of their *gag* genes that overlaps their *pol* genes (3,4). The demonstration for such frameshifting was a major discovery, and much subsequent work has confirmed the involvement of a –1 ribosomal frameshift event in GagPol synthesis (5,6). The HIV-1 *gag-pol* shift site, U6A, is AAU-UUU-UUA-GGG with the codons shown in the *gag* reading frame and the underlining highlighting the first two codons following realignment to the -1 frame, the *pol* reading frame.

The initial evidence for the frameshifting occurring at the translation level involved cell-free synthesis experiments. The argument against a small proportion of the mRNA used in the mammalian cell-free protein synthesis experiments having *gag* and *pol* in the same reading frame was the reported finding that *Escherichia coli* cell-free translation did not yield the fusion product (3). However, later work showed that *E. coli* ribosomes perform counterpart –1 ribosomal frameshifting at the same sequence with 40% efficiency of mammalian ribosomes (5,7). This ruled out the argument against some GagPol potentially being derived from reverse transcriptase slippage followed by standard in-frame decoding. During the subsequent 28 years, the possibility of reverse transcriptase slippage at the retroviral *gag-pol* frameshift sites has not been investigated, though deep sequencing has revealed heterogeneity at several sites in retroviral sequences (8). Numerous studies, some ongoing, seek compounds that specifically modulate HIV ribosomal frameshifting efficiency (9–15). The therapeutic goal of this work is expected to be dependent not only on their specificity and effectiveness in modulating ribosomal frameshifting, but also on whether a proportion of GagPol is derived from the alternative process of specific polymerase slippage. The present study is confined to *in vitro* experiments, but it points to the need for further work to assess the origin(s) of GagPol for retroviruses whose *pol* is in the

*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

–1 frame with respect to their *gag* coding sequence. In addition to practical considerations, historical aspects of the issues faced by one of the two groups that initially studied retroviral 'frameshifting' (16), the wider issue is the extent of functional utilization of specific polymerase slippage in gene expression. Early studies on mechanistically distinct utilization in the expression of paramyxoviruses (17,18), the filovirus Ebola (19), and a bacterial DNA polymerase gene (20) have been broadened by several studies including those on bacterial IS elements (21–23), Ebola virus (23–26), several secretion apparatus genes of *Shigella flexneri* (27–29), *Citrobacter rodentium* and *Yersinia pseudotuberculosis* (30), certain high genomic AT content endosymbiotic bacteria of insects (31,32) and a cell movement gene of the major plant virus family of Potyviruses (33–36). Despite these discoveries, the present work points to the need for further work on the relatively understudied phenomenon of specific transcriptional slippage in cellular and viral gene expression. One aspect relevant to searches for unknown cases of slippage utilization is context effects that may inhibit the propensity for polymerase slippage. This is investigated here by analysis of the GGG sequence 3′ adjacent to the ribosomal frameshift site and previously shown to be important for the ribosomal frameshifting (37). Another relevant aspect is potential stimulatory effects of 'partial roadblocking' RNA structures ahead of polymerase that could stimulate specific slippage at sites including those not recognizable as likely shift-prone. Such structures can stimulate reverse transcriptase slippage at a nearby 5′ sequence at which it would not otherwise detectably occur (38). Retroviral *gag-pol* ribosomal frameshifting is stimulated by RNA structural recoding elements 3′ of the shift site (16). This raises the possibility that the same structures, or components thereof, may also act to stimulate reverse transcriptase slippage.

Studies addressing potential, or demonstrably, functionally utilized specific polymerase slippage build on earlier fidelity investigations of different polymerases, including of reverse transcriptases. There have been a number of such studies with HIV-1 RT. One of the notable findings was several fold higher fidelity with RNA than DNA templates (39), and another discerned the effects of dNTP pool imbalances on frameshift fidelity with DNA templates (40). This was also studied in a comparative analysis of slippage-mediated product indel formation by several different reverse transcriptases (38). Furthermore, substrate imbalance is also well known to influence *E. coli* and *S. cerevisiae* DNA-dependent RNA polymerase slippage (23). The degree of influence of the ratio of the cognate substrate for the slippage site to that of the substrate base for the 5′ adjacent base, also features in the present study of specific HIV RT slippage at a site region that is also responsible for the ribosomal frameshifting that mediates GagPol synthesis. In distinction to these earlier studies, the present study involves the natural context of the relevant sequence, since it permits potential limiting or stimulatory features to be discerned. Also the deep sequencing employed permits a more refined analysis.

MATERIALS AND METHODS

RNA template constructs

Double-stranded DNA templates, for T7 RNA polymerase, were generated by an overlapping DNA oligonucleotide strategy (Supplementary Table S1) and 3′ extended with Vent DNA polymerase. Preparation of RNA template (200 nt HIV) was generated by T7 RNA polymerase *in vitro* in 20 μl reaction volumes with a Promega kit (T7 RiboMAX™ large scale RNA production system) according to the manufacturer's instructions. Reaction products were treated with 2 units of DNase I turbo (Ambion) in 50 μL reaction volumes at 37°C for 30 min. 500 μl Trizol (Invitrogen) was added to the reaction and incubated for 2 min. Then, 100 μl chloroform was added and incubated for 5 min at room temperature, before centrifugation, 12 000 × *g* for 15 min at 4°C. The upper aqueous phase was extracted, and an equal volume of 100% ethanol was mixed with it before it was loaded on a silica column (Anachem PCR purification kit). After centrifugation, 12 000 × *g* for 30 s, the column was washed with 600 μl of a low salt wash buffer (Anachem's 'PCR clean up' kit wash buffer to which ethanol was added to 80% concentration). The column was then treated with DNase I turbo, 10 units in a 100 μl volume for 1 h incubation at 37°C. This was followed by 600 μl of a high salt wash buffer (Anachem's 'PCR clean up' kit wash buffer to which ethanol was added to 40% concentration) and centrifugation. It was then washed twice, via centrifugation, with the low salt buffer, 600 and 300 μL. The final centrifugation was at 12 000 × *g* for 2 min to dry the column before elution with 100 μl of RNase free water.

Chemically synthesized RNA templates and DNA oligonucleotides were from IDT-DNA (Supplementary Table S1).

Reverse transcription

HIV-1 reverse transcriptase enzyme was purchased from Abcam. RT reactions involved a pre-annealing step of the RNA template (100 ng): DNA Primer (2 pmol) (Supplementary Table S1), in 1× SuperScript™ III buffer (50 mM Tris–HCl, pH 8.3 at 25°C, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT), in the presence of the dNTP substrate (with the specific concentrations of each indicated in the main text) and, where present, antisense (100 pmol), in a 10 μl reaction volume. With the HIV-1 200 nt cassette, the annealing step was at 95°C for 30 s with a 10% temperature decrease (from 95 to 16°C) every 30 s. With RT reactions involving the HIV-1 chemically synthesized RNA cassette combinations (Figure 3), the annealing step was at 65°C for 5 min before chilling on ice. Then, a 10 μl mix (1× SuperScript™ III buffer, 4 units HIV-1 RT enzyme, 20 mM DTT) was added and incubated at 37°C for 50 min, followed by 85°C for 5 min.

Polymerization 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 template 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.

Sample preparation for deep sequencing

A first round of amplification of the cDNA product was with a set of forward and reverse primers with an NGS adapter sequence at their 5′ ends and sequence complementary to the cDNA product at their 3′ ends. This set of primers included the common forward primer, HIV_NGS_F1697, whose 3′ sequence is complementary to the 3′ end of the cDNA. The reverse primer used for set I ('long RNA: long cDNA') was HIV_NGS_R1695, and for sets 2 and 3 ('long RNA: short cDNA' and 'short RNA: short cDNA', respectively) it was HIV_NGS_R1696. PCR products were then purified using Zymo-SpinTM I column (Zymo Research) with buffers from the PCR clean-up kit (Roche). Subsequently, a limited cycle PCR using Nextera XT indices (Illumina) was performed to incorporate sequencing adaptors and dual-index barcodes to each PCR product. Products were then purified with Ampure XP beads (Beckman Coulter), quantified using the Qubit fluorimeter (Thermo Fisher) and pooled at an equimolar concentration. The mix of 80 samples was sequenced on a MiSEq (Illumina) using a MiSeq Reagent Nano Kit (300 cycles) with V2 chemistry (Illumina).

Deep sequencing analysis

The paired-end cDNA libraries were obtained for 80 samples with approximate yield of ∼10 000/reads per sample. The reads were trimmed using Cutadapt (41). Adapter sequences were clipped from forward (5′-CTGTCTCTTATACACATCTCCGAGCCCACGAGAC-3′) and reverse (5′-CTGTCTCTTATACACATCTGACGCTGCCGACGA-3′) reads. Trimmed reads were aligned to the corresponding reference gene sequence using bowtie2 short read alignment program (42). Local alignments of paired end reads were carried out with default parameters. Sorted BAM alignment files were obtained using SAMtools version 1.3.1 (43) which were used for prediction of SNPs and INDELs. Variant calling was performed using SAMtools mpileup. Parameters used for mplileup predictions were (-g -uf –output-tags DP, AD -L 1000000 -d 1000000) which generated output in BCF format. BCFtools (44) view option was used to convert variant calls to the Variant Call Format (VCF). For graphical analysis of substitution rates in the slippery sites custom script based on seqLogo (45) R package was used to plot sequence logos. Heatmaps were generated using MeV (Multiple Experiment Viewer) (46).

RESULTS

Strategy for testing the effect of a 200 nt HIV1 *gag-pol* cassette

The sequence 3′ of the 5′-U6A-3′ HIV *gag-pol* frameshift site can form alternative structures. Though most work on ribosomal frameshifting has been done with cassettes with 65 nt 3′ of the shift site (47), additional flanking sequences may form other relevant structures (48–51). This was tested using a 140 nt cassette with 71 nt 3′, and 63 nt 5′, of the shift site. Structures similar to those in template RNA may also form in the nascent cDNA and are potentially relevant to reverse transcriptase slippage. Conceptual precedent comes from the finding that formation of a nascent RNA stem-loop structure can be a strong stimulator for DNA-dependent RNA polymerase slippage (23). To assess potential HIV RT enzyme slippage on the *gag-pol* U6A frameshift motif with putative stimulation by the flanking sequence context, the 140 nt RNA used in the ribosomal frameshifting studies plus 60 nt for primer annealing (total 200 nt), sequence was generated *in vitro* by T7 RNAP (Materials and Methods). This RNA had 70 nt 5′ of the shift site, the U6A shift site and 124 nt 3′ of it (Figure 1).

The cDNA products were amplified by polymerase chain reaction (PCR) using specific primers. As a control for DNA polymerase fidelity, a counterpart PCR experiment was performed with a chemically synthesized DNA whose sequence was the reverse complement of the 200 nt RNA template used above. The resulting RT-PCR and PCR products were quantitatively analyzed by next-generation sequencing (NGS) (Materials and Methods). For each sample, single nucleotide polymorphism (SNP) and nucleotide insertion/deletion (indel) frequency were determined at each nt location by comparison to their respective cDNA reference sequence. For indel analysis at the 5′-U6A-3′ *gag-pol* derived sequence, NGS reads were selected only when the sequence derived from the RNA template run of U's was free of substitutions.

Effect of the HIV1 *gag-pol* 140 nt frameshifting element RNA on RT slippage

With an A6 slippage motif HIV RT slippage directionality is sensitive to an imbalance of the concentrations of the cognate substrate dNTPs specified by a template slippage motif and the base 5′ adjacent to the motif (38). To ascertain if counterpart effects occur with 5′-AU6A, an imbalanced ratio of dATP, cognate for the U6 tract, and of dTTP, cognate for the template base 5′ adjacent to the U6, was tested. RT reactions were performed with 3 dNTP ratio conditions for the substrates: (i) all 4 dNTPs at 500 μM, (ii) dTTP 5 μM, other 3 at 500 μM, (iii) dATP at 5 μM other 3 at 500 μM.

At the 5′-U6A-3′ *gag-pol* derived sequence, the NGS data shows a correlation of indel frequency for the cDNA base A with the [dATP]:[dTTP] ratio in the RT reactions. With equimolar dNTP, base omission is 2.1% and base addition 1.4%; with the $[dATP]_{500\mu M}$:$[dTTP]_{5\mu M}$ ratio, base omission is 1.7% and addition 2.0%; with the $[dATP]_{5\mu M}$:$[dTTP]_{500\mu M}$ ratio, base omission is 2.6% and addition is 1.1%. The control for DNA polymerase slippage shows a lower frequency (1.2%) of base omission, and a background level (0.1%) for base addition (Figure 2). These NGS results show that HIV RT undergoes specific slippage-mediated base addition (especially with [dATP] > [dTTP]) and omission (especially with [dATP] < [dTTP]). Importantly, the natural *gag-pol* nucleotide context 5′ and 3′ to the U6 motif does not abolish the slippage propensity of the HIV-1 RT enzyme.

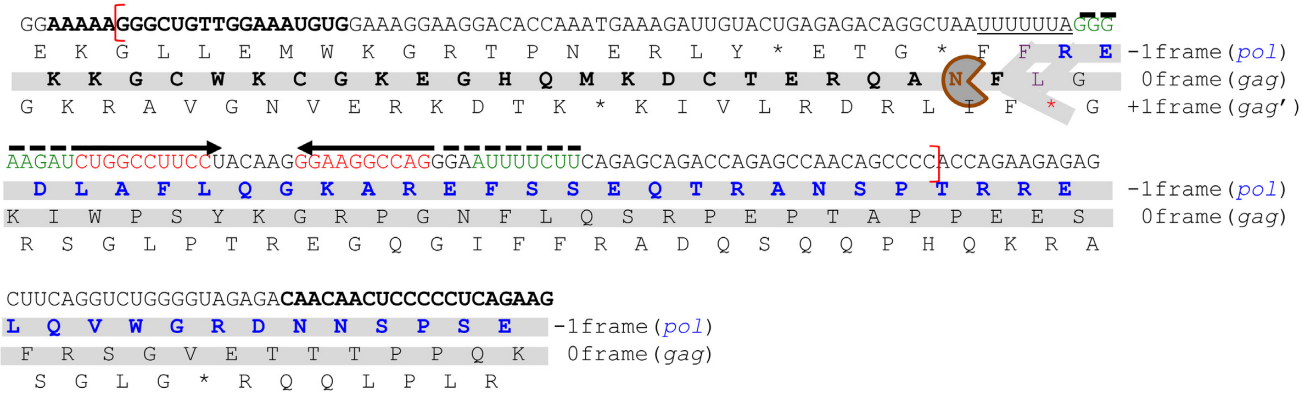

Figure 1. The HIV *gag-pol* ribosomal frameshift site also mediates HIV reverse transcriptase slippage. The *gag-pol* frameshift site, U_UUU_UUA (underlined) is shown with 30 nt 5′ of the shift site and 48 nt 3′ (for which one of the potential pairing possibilities is indicated in red and green). The 5′ and 3′ borders of an RNA sequence involved in a potential and relevant large alternative structure revealed in a prior study (48), are indicated by red brackets. RNA three frame translation is shown with the 0 and –1 reading frames being those of *gag* and *pol* respectively. The highlighted grey box indicates the amino-acid sequence of the ribosomal frameshift products and the corresponding sequence specified only by the original zero frame. The 5′ and 3′ nt sequences in bold indicate the nt annealing region of the primers used in the (RT)-PCR reactions.

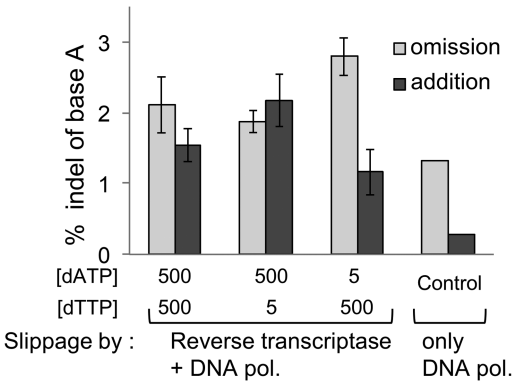

Figure 2. NGS analysis of indels in the run of A's derived from the 200 nt HIV sequence. Frequency of insertion, or omission of base A, derived from cDNA generated at three different relative dNTP concentration conditions. The standard deviation is for 4 independent reverse transcription reactions. The control reflects only DNA polymerase slippage.

The other type of fidelity error, base mis-incorporation was next analyzed to enable its comparison with indel formation. The SNP frequency is less than 0.5%. Its frequency at locations corresponding to each base in the cDNA run of A's (5′-$A_1A_2A_3A_4A_5A_6$–3′, with the number in subscript indicating the position of the base in the homopolymeric run) is similar whether all dNTPs are equimolar, or with the $[dATP]_{500\mu M}$:$[dTTP]_{5\mu M}$ ratio. In contrast, with the $[dATP]_{5\mu M}$:$[dTTP]_{500\mu M}$ ratio, the SNP frequency strongly correlates with the cDNA A-tract base position. It increases from 1.3% corresponding to A_1 to 7.1% at A_6 with a preference for mis-incorporation of the base G (from $\approx$ 1.1% at A_1 to 4.6% at A_6) (Supplementary Results). These results show that depending on the relative dNTP concentrations, SNP errors can be less significant than indel errors for the 5′-U6A-3′ derived sequence. [A graphical comparison of the frequency of indels and SNPs in the cDNA reference sequence 5′-TCCCTAAAAAATTAGCCTC-3′ for 4 repeated experiments is shown in Supplementary Dataset S1. A heatmap representation of the SNP frequency is also shown for 25 nt 5′ to the A-tract of the reference cDNA (Supplementary Figure S3). Supplementary Dataset S1 and Supplementary Figure S3 also show the repeat experiments for the constructs described below]. Responsiveness to alterations of the ratios of the different dNTPs is relevant to discrimination against effects of potential infidelity of the T7 RNA polymerase used to generate the template for this work.

The same derived RT-PCR and PCR products used in the NGS analysis were also analyzed by Limited Primer Extension (LPE) (Supplementary Information; Supplementary Figure S1, panel A). This has the advantage of allowing more conditions to be quickly and cheaply tested. Though it permits analysis of only a specific part of the template sequence, the segment analyzed here is the most relevant. LPE analysis yielded qualitatively similar results to NGS for both base indel and SNP, for products derived from the 5′-U6A-3′ containing sequence (Supplementary Results and Supplementary Figure S1, panel B). To explore a possible template road-blocking effect on HIV RT reverse transcriptase, an experiment was performed with a 10 nt antisense RNA complementary to the HIV RNA sequence 5′ adjacent to the U6-tract. LPE analysis revealed that presence of the antisense favors slippage-mediated base addition (Supplementary Results, Supplementary Figure S1, panel B) (LPE analysis was used for this characterization since it was also used for multiple reaction conditions with the templates described below).

Strategy for testing nucleic acid scaffolds for RT reactions with chemically synthetized RNA templates

Chemically synthesized RNA avoids potential complications due to T7 RNA polymerase infidelity. We explored slippage with synthetic RNAs close to the maximum practical length of single synthesis RNA (84 nt). The two RNAs contain 30 HIV nt 5′ of the shift site, the U6 tract, and, in RNA_1654, 48 HIV nt, and in RNA_1655, 25 HIV nt. 1654 is referred to below as 'WT long RNA' and 1655 as 'WT

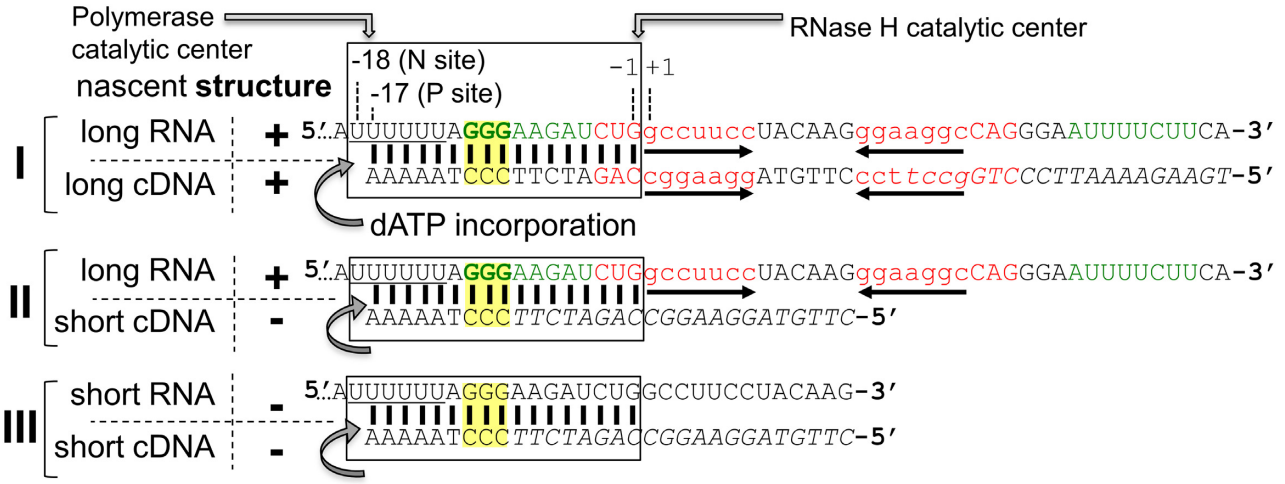

Figure 3. HIV *gag-pol* RT slippage stimulators. Strategy to assess possible stimulatory or inhibitory effects on slippage by potential structure formation in the RNA template and/or in its derived nascent cDNA, during reverse transcription of the U6A motif. Sequence set I shows the chemically synthesized RNA ('long RNA') and a potential for stem-loop formation (red and green). When the RT transcribes as far as the 5′ U of the U6 tract (to generate 'long cDNA'), the parts of the RNA or cDNA free behind the 18 bp hybrid, and potentially free to participate in stem-loop formation, are indicated by inverted arrows. Sequence set II shows the same 'long RNA' template reverse transcribed from a primer whose sequence is in italics at the 5′ end of the 'short cDNA'. Sequence set III has 'short RNA' without the potential for structure formation involving the sequence underlined by inverted arrows in sets I and II. The RNA triplet GGG present in both 'long and short RNA' templates, which was substituted by the sequence CCC (designated 'mut'), is highlighted in yellow.

short RNA'. In 'WT long RNA' at least 35 nt 3′ of the U6 tract have potential to be involved in stem–loop structure formation. In 'WT short RNA' the potential for stem–loop structure is limited (Figure 3).

HIV RT slippage may be stimulated by a structure in the nascent cDNA that is a counterpart of an RNA template structure. To investigate this possibility two different sites were used for priming, only one of which is in 'WT short RNA'. The primer for this site, R_1648, yields a 25 HIV nt nascent cDNA sequence 5′ of the A-tract, 'short cDNA', with limiting stem–loop structure formation (Figure 3). The additional primer for 'WT long RNA', (R_1645) yields a 48 HIV nt nascent cDNA sequence 5′ of the A-tract, 'long cDNA', with the potential for structure formation (Figure 3). Three different 'RNA template: Primer' combinations were tested (Figure 3).

The first combination involved 'WT long RNA: long cDNA' with both having structure potential, (Figure 3, combination I). The second combination involved 'WT long RNA: short cDNA' with only the template having relevant structure potential (Figure 3, combination II). The third combination involved 'short RNA: short cDNA' with neither having structure potential (Figure 3, combination III).

Due to the effect of dNTP imbalance on RT slippage directionality (resulting in base addition or omission), RT reactions were performed with 9 dNTP concentration conditions. In all conditions dCTP and dGTP were constant at 500 μM. The concentrations of dATP (the substrate specified by the U6 motif) and dTTP (the substrate specified by the RNA template base 5′ adjacent to the U6) were either at 5, 50 or 500 μM.

Run of U's derived indels studied with several template:primer combinations

For each sample, NGS reads were selected for slippage analysis only when sequence derived from the RNA template run of U's was free of substitutions. For the 3 'RNA template: Primer' combinations (Figure 3), sequence derived from the run of U's exhibits both types of indel due to slippage in both directions, and with a higher frequency of base omission than base addition (Figure 4). This means that potential to form a small cDNA structure adjacent, or almost adjacent, to the hybrid does not influence product indel formation. Also, potential by the 'long RNA' template containing the GGG to form structure does not, at least, significantly influence the efficiency of indel generation. However, specific dNTP ratios are seen to influence the slippage. With cognate dATP at any of 5, 50 or 500 μM concentrations and the dTTP concentration increasing from 5 to 50 to 500 μM, the frequency of specific base omission increased from 0.5% up to 2.5%. With the same substrate concentrations, the efficiency of base addition is less than 0.5% except with the highest dATP:dTTP ratios (Figure 4). Qualitatively similar results were also obtained by LPE analysis (Supplementary Results and Supplementary Figure S2A and C).

Does the highly conserved GGG 3′ adjacent to the U6A motif influence RT slippage?

The highly conserved GGG triplet 3′ adjacent to the U6A motif (37) is in the RNA:DNA hybrid when HIV RT reverse transcribes the U-tract. This GGG is potentially relevant to hybrid stability and realignment.

We used a similar strategy as in the previous section to test for possible effects of substituting GGG with CCC (Figure 3, highlighted in yellow). The three combinations, 'mut

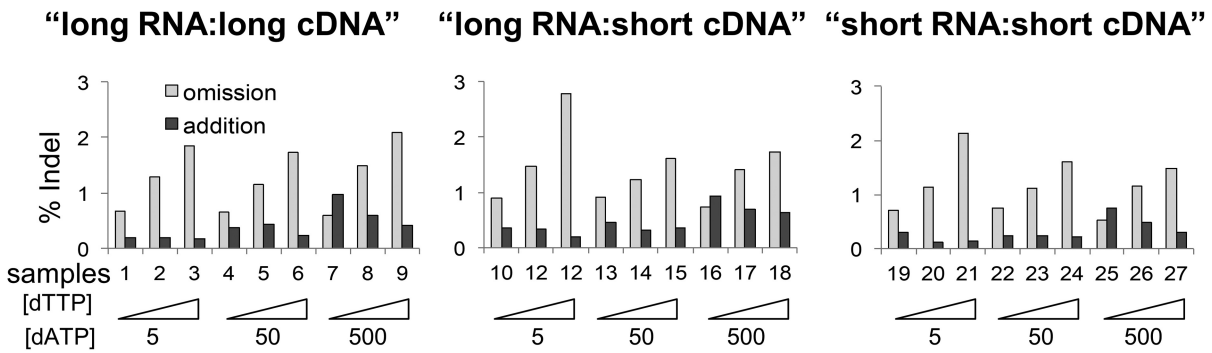

Figure 4. NGS analysis of 'WT' HIV cDNA derived sequence: indel distribution. Frequency of base addition, or omission, in the cDNA run of A's. This is shown for the 3 'RNA template: Primer' combinations indicated above each graph. Sample identity and dNTP concentration ratio conditions are indicated at the bottom. Indel results are from representative samples 1–27 (SI Results & SI Dataset 1).

long RNA: long cDNA', 'mut long RNA: short cDNA' and 'mut short RNA: short cDNA' were tested with nine dNTP concentration conditions. In all conditions dCTP and dGTP were at 500 μM. The concentrations of dATP (specified by U6) and dTTP (specified by the template base 5′ adjacent to U6) were either at 5, 50 or 500 μM. The 27 different RT-PCR products derived from cDNA generated by the nine dNTP concentration conditions for each of the 3 'RNA template: Primer' combinations, were analyzed by LPE (Supplementary Results and Supplementary Figure S2B and D). Nine of these different RT-PCR products were also analyzed by NGS; they correspond to the 3 'RNA template: Primer' combinations tested with the three following dNTP ratio conditions (i) all four dNTPs at 500 μM, (ii) dTTP 5 μM, other three at 500 μM, (iii) dATP at 5 μM other three at 500 μM.

With the two combinations, 'mut long RNA: short cDNA' and 'mut short RNA: short cDNA', slippage-mediated base addition occurs at a higher frequency than with their WT counterpart (Figure 5A). With equimolar dNTP concentration, base addition is ∼3 times more efficient with the mut combination compared with the WT equivalent. Interestingly, the efficiency of base addition of the mut combinations at equimolar dNTPs is even higher than the efficiency of base addition for the corresponding WT combination at the dNTP concentration ratio which favors base addition, i.e. $[dATP]_{500\mu M}$:$[dTTP]_{5\mu M}$. In contrast, mut combinations exhibit a similar frequency of product base omission as their WT counterparts (Figure 5A).

With the mut sequence, base substitutions are more prominent at each position of the cDNA run of A's (5′-$A_1A_2A_3A_4A_5A_6$–3′), especially with 5 μM dATP and the other three dNTPs at 500 μM. This SNP difference between mut and WT can be seen with sequence logo illustrations (Figure 5, panel B and Supplementary Dataset 1). Compared to the 'WT RNA: cDNA' combinations, the corresponding mut combinations, there is a marked SNP increase at cDNA base position A_2 and this high proportion is maintained at a similar level for positions A_3–A_6. In contrast, with the corresponding WT combination, there is a gradual increase of substitutions from positions A_1 to A_6 in the cDNA. This result suggests that mutating the conserved RNA template GGG triplet to CCC, influences RT polymerase fidelity at least at the template base position U_2. As judged only from mutating the conserved RNA template GGG to CCC, the identity of the bases at this position in WT serves to constrain the frequency of base A errors in the run of A's in the product cDNA.

DISCUSSION

Reverse transcriptase slippage is shown to mediate generation of DNA which when transcribed would yield RNA with the sequence AAU-UUU-UUU-AGG. The underlined U is absent in the product, AAU-UUU-UUA-GG, generated without slippage. Standard translation of the slippage derived RNA would yield Asn-Phe-Phe-Arg, identical to the minor product from -1 ribosomal frameshifting (Figure 6) [two GagPol products are generated by –1 ribosomal frameshifting, 80% with Asn-Phe-Leu-Arg and 20% with Asn-Phe-Phe-Arg specified by the frame junction sequence (4,5,52)]. Despite all the work on retroviral ribosomal frameshifting, it is only relatively recently that a product from ribosome shifting into the alternative +1 frame has been detected. The frameshifting involved was detected with reporter cassettes in reticulocyte lysate translation experiments, and the shift involved is –2 rather than +1 (Figure 6). It was considered likely insignificant as its occurrence during HIV translation would lead to a Gag product with the C-terminal sequence Asn-Phe-Leu-stop (termination being specified by the +1 frame UAG above) (53). Reverse transcriptase slippage is here shown to yield DNA sequence whose derived RNA with the sequence AAU-UUU-UAG would be translated Asn-Phe-stop, i.e. the same as the –2 ribosomal frameshift derived product except for the absence of one amino acid, Leu (Figure 6). WT HIV Gag is proteolytically cleaved after its amino acid Asn encoded by the AAU whose U is the 5′ U of the homopolymeric component of the ribosomal frameshift site. In WT Gag multiple amino acids follow Asn and presumably cleavage does not occur when only two AA are C-terminal to Asn. If so, and if the novel termination product is synthesized in virus-infected cells, its generation could be significant. There are five sequential cleavage sites (54–59) and alteration of the 3rd cleavage site may have an effect on the fourth and fifth proteolytic cleavages of the Gag precursor protein. Irrespective, its potential functional role merits investigation, in

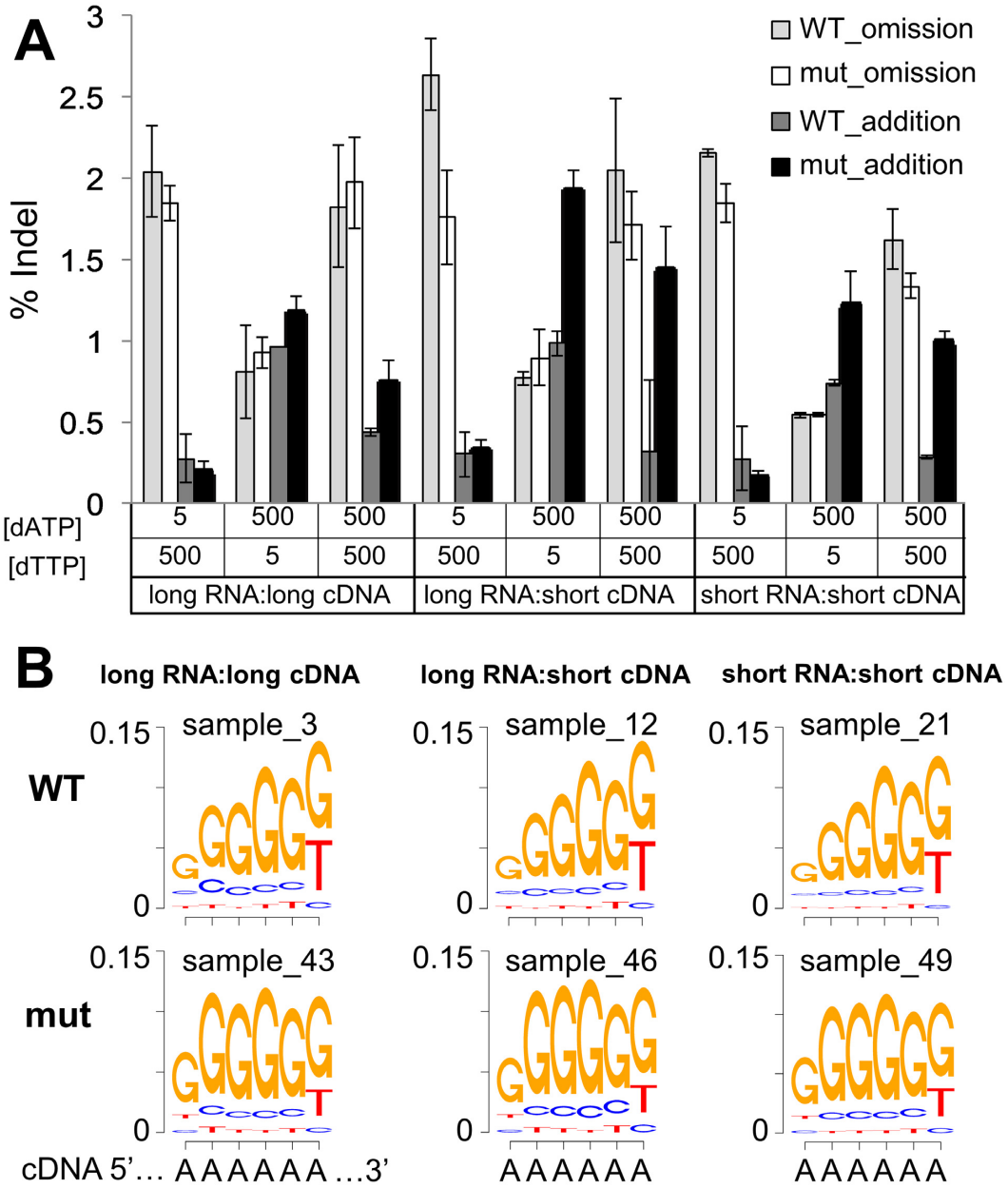

Figure 5. Effect of the nt triplet one nt 3′ to the run of U's on HIV RT fidelity. (**A**) Frequency of slippage-mediated base addition, or omission, for the WT and mut combinations tested with RT reactions involving three different dNTP concentrations conditions. SEM error bars are indicated for two independent experiments. (**B**) Sequence logo of the substituted base in the run of A's. The cognate substrate was omitted to enhance visualization of the substituted bases in the run of A's in the cDNA. The identity of the 'RNA template: Primer' combination and the dNTP concentration conditions, are indicated above the WT (top row), and below the corresponding mut combination (bottom row). The name of the representative sample is indicated; its corresponding dataset is in SI Dataset S1.

part because of possible relevance to resistance mechanisms against drugs targeting the protease involved.

Protease mutants

In some protease resistance mutants there is a compensatory C-terminal amino acid substitution of the NC-SP2-P6 Gag-derivative product (60). The mutations involved may enhance reverse transcriptase slippage potential. For instance, at the proteolytic cleavage site of NC/SP2, in the mutation that causes an A431V substitution, the codon GCU, 5′ adjacent to the *gag* sequence AAU_UUU_UUA_GGG that contains the ribosomal frameshifting site, is changed to GUU (60). This mutation may increase potential formation of a weak GNRA capped stem-loop 5′ adjacent to the U-tract (Supplementary Figure S6). At the proteolytic cleavage site for SP2/P6, in the mutation that causes the L449F substitution, the codon CUU (bold) in the *gag* sequence AAU_UUU_**CUU**_CAG is replaced by one of the Phe codons, which are UUU and UUC (61,62). Either would

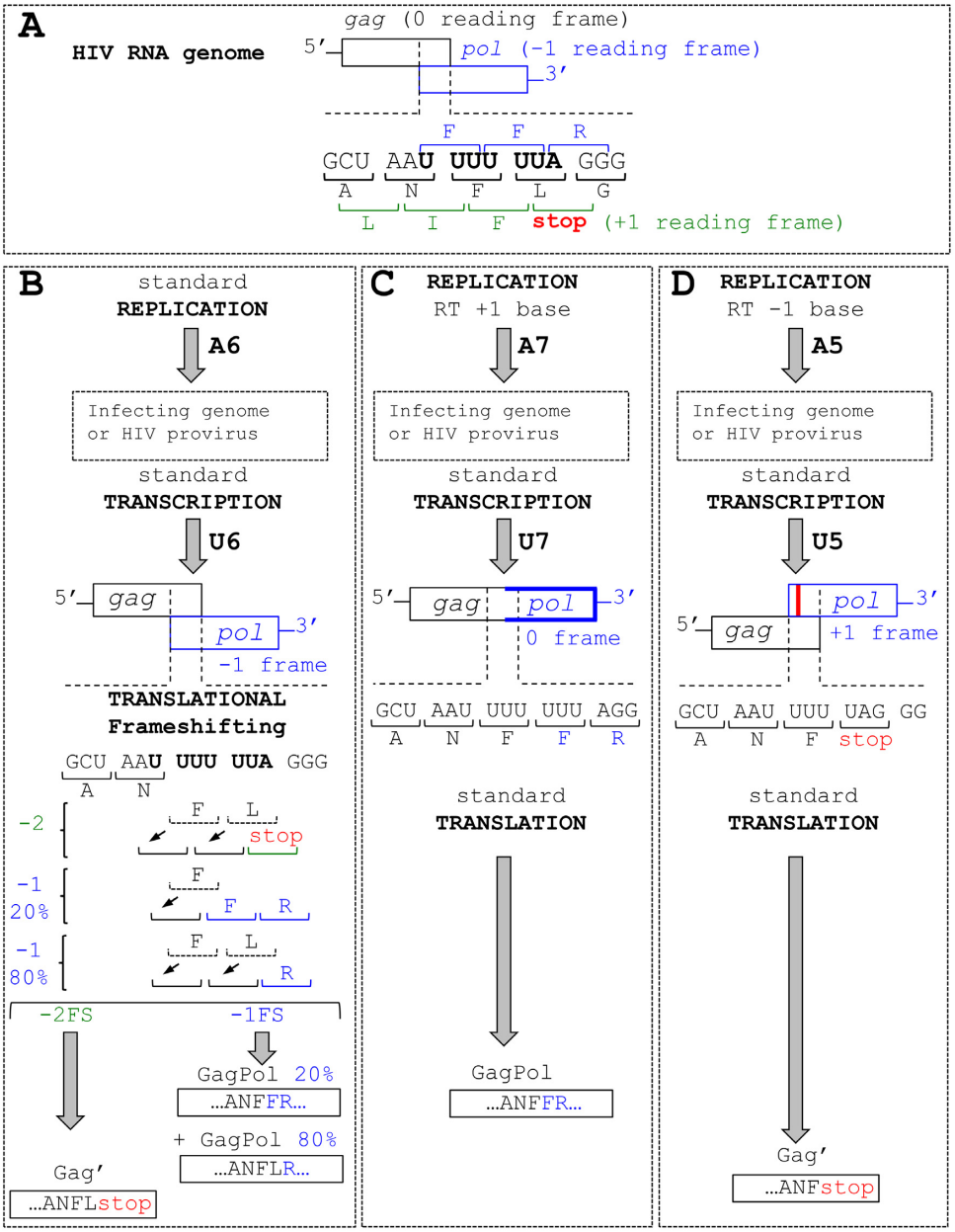

Figure 6. Expression of HIV *gag-pol*. (**A**) Genetic organization of the *gag* and *pol* open reading frames. With the nomenclature in common use, standard translation of RNA with an extra base is the counterpart of –1 ribosomal frameshifting on WT mRNA (and from the protein product perspective, base omission is the counterpart of +1 ribosomal frameshifting). The *gag-pol* ribosomal frameshift sequence is indicated in bold with the corresponding amino-acid sequences indicated for the *gag* reading 0 frame (black), the *pol* reading –1 frame (blue) and the +1 reading frame (green). (**B**) Ribosomal frameshifting yield to Gag' and GagPol products. The alternative amino-acid sequences in prior analyses of –1 ribosomal frameshift products with the major transframe encoded Phe-Leu containing GagPol product being 80% of the total and its Phe–Phe counterpart constituting the remaining 20% (16). Brackets with dashed lines under codons indicate initial tRNA anticodon pairing prior to realignment. Potential transcription slippage counterparts due to RT slippage-mediated one base addition (**C**) or one base omission (**D**) are considered in the Discussion.

create the potential for an additional slippage motif with a 5′ sequence with potential to form an RNA structure that could act as a 'partial roadblock' RT slippage stimulator.

Substrate ratios

dUTP incorporation by RT HIV has a protective affect against host chromosome integration (63). The effect of dUTP on reverse transcriptional slippage was not explored in the present manuscript. However, this could also be relevant with the natural target of HIV, as CD4 T cells and macrophages contain a 1:4 ratio of dUTP:dTTP in contrast to other cells (1:100 ratio) (63 and refs therein)

dNTP depletion is part of the host's response to viral infection. The deoxynucleotide triphosphohydrolase SAMHD1 depletes dNTP pools which interferes with RT function in myeloid cell (64). Lentiviruses, such as HIV-2, have developed a *vpx* gene to counteract this effect (65–67).

[Contrastingly, in many colon cancer cells SAMHD1 is directly debilitated by mutants in its coding sequence with consequences for dNTP levels and ratios (68); mutator phenotypes of colon cancer cells due to DNA polymerase mutants also affects dNTP levels and ratios (69).] Future work is needed to explore the potential effect on RT HIV slippage of the relative intracellular dNTP concentration in virus-infected cells. The potential selective advantage of base insertion or omission in the *gag-pol* region of chromosomally integrated HIV virus sequence may also be relevant. Such an investigation could involve *gag-pol* sequencing of specific tissues of latent phase infected patients.

Slippage versus ribosomal frameshifting: hybrid lengths

A big distinction between ribosomal frameshifting and polymerase slippage is that with frameshifting at each position a much shorter template sequence is involved in base pairing. The RNA:cDNA hybrid in reverse transcriptase is ∼18 bp (70,71). It extends from the RNase H catalytic center to the polymerase catalytic center. The length of the RT slippage U6 motif investigated here is much shorter than the length of the hybrid, suggesting that only a portion of the 18 bp are involved in the re-alignment. RT modelling to identify locations that could accommodate potential hybrid bulge formation (due to an extra helical base), pointed to such potential only corresponding to a hybrid region whose cDNA component is 5–6 bp from its 3′ end (72). However, it is possible that the portion of the hybrid between the slippage motif and the RNase H catalytic center may also affect slippage, perhaps due to its helical topology. A constraint on the latitude of sequence variation relevant to both frameshifting and slippage is retention of an Asn codon (overlapping the 5′ end of the shift motif) since Asn is important for proteolytic cleavage (55,73). While this restricts the number of possible heptanucleotide frameshift sequences that could have been utilized at this site, there are still some that are unlikely to mediate reverse transcriptase slippage.

Other features of comparative importance of the shift/slippage cassette are also noteworthy. The highly conserved triplet, GGG, separated by 1 base from the 3′ end of the U-tract, is significant for frameshifting. This sequence has generally been considered not to have reached the ribosomal A-site when frameshifting occurs. However, a recent study has proposed that it is in a (perhaps distorted) A-site when the ribosomal frameshifting occurs (37). In the present work when GGG was substituted with CCC, efficient RT slippage still occurred but it was qualitatively different. There was also an effect on product base substitutions and further work may reveal relevance of the GGG for shift/slippage site preservation. Formally structural RNA recoding signals can potentially stimulate frameshifting or slippage due to effects on their initial unwinding/unpairing or when the sequence involved (re)forms structure on or after, exit from the ribosome/polymerase. With the WT GGG just discussed, structure involving at least most of this ribosomal frameshift relevant sequence only minimally affects slippage. However, when the GGG is substituted with CCC, the further 3′ ribosomal frameshift relevant sequence has also a marked effect on slippage. However, to avail of the advantages of using homogeneous chemically synthesized—though necessarily shorter - template RNA, slippage studies with the full extent of the 3′ recoding signal important for ribosomal frameshifting, were limited. Accordingly, deductions from the part of the work involving the 3′ RNA structure need to be treated with caution. Future experiments could employ RNA construct variants with increased pairing potential 3′ of the U-tract and at different distances from it, or a similar strategy involving complementary antisense oligonucleotides.

RNA structure effects

The potential for RNA structure formation ahead a slippage motif can strongly influence the directionality of HIV reverse transcriptase slippage (38). Interestingly, an RNA antisense specific to the sequence 5′ adjacent to the HIV *gag-pol* U6 tract also influences, though moderately, HIV-1 RT polymerase slippage. Testing whether native *gag* RNA template structure formation 5′ to the run of U's influence HIV GagPol synthesis via reverse transcriptase slippage is outside the scope of the present work.

During synthesis of HIV double-stranded DNA from single-stranded cDNA, before the polymerase reaches 3′-TA6TCCC-5′, it encounters the complement of the sequence that forms the ribosomal frameshift stimulatory structure. Studies with a model stem-loop 5′ adjacent to an A6 RNA motif have shown the potential for 'road-blocking' RNA structure to enhance slippage-mediated base addition (38) raising the possibility that HIV RT undergoes structure stimulated slippage during DNA second strand synthesis. In contrast to RT slippage at 5′-AU6AGGG-3′ in the generation of single-stranded cDNA, a higher ratio of dATP compared to dTTP is expected to positively influence slippage-mediated base omission on second strand DNA synthesis.

While any conclusion about whether the presence of a cassette that can mediate both ribosomal frameshifting and reverse transcriptase slippage is fortuitous or has been selected to facilitate both processes, is premature, dual function may not be coincidental.

Perspective

One type of recoding and splicing are alternatively used in different occurrences of one set of othologs (74–76), in several instances one type of recoding has been replaced by a different type (16). One example of the latter is *dnaX* where in *E. coli* expression of a second protein product involves a ribosomal frameshift event and in other bacteria including *Thermus thermophilus*, synthesis of the counterpart product involves transcription slippage (20,77–79). The present work raises the possibility that several retroviruses use a single cassette for productive ribosomal frameshifting and polymerase slippage. Such dual function could be relevant to studies aiming to ameliorate retroviral infection with compounds that alter only the efficiency of one of the two mechanisms involved.

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; F. Crispie and T. Cotter in the APC Microbiome Institute for Illumina sequencing and advice; and Alan J. Herr and several anonymous reviewers for constructive comments on the ms.

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. Philipson,L., Andersson,P., Olshevsky,U., Weinberg,R., Baltimore,D. and Gesteland,R. (1978) Translation of MuLV and MSV RNAs in nuclease-treated reticulocyte extracts: enhancement of the Gag-Pol polypeptide with yeast suppressor tRNA. *Cell*, **13**, 189–199.
2. Yoshinaka,Y., Katoh,I., Copeland,T.D. and Oroszlan,S. (1985) Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 1618–1622.
3. Jacks,T. and Varmus,H.E. (1985) Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. *Science*, **230**, 1237–1242.
4. Jacks,T., Power,M.D., Masiarz,F.R., Luciw,P.A., Barr,P.J. and Varmus,H.E. (1988) Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature*, **331**, 280–283.
5. Weiss,R.B., Dunn,D.M., Shuh,M., Atkins,J.F. and Gesteland,R.F. (1989) *E. coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol.*, **1**, 159–169.
6. Li,G.D. and De Clercq,E. (2016) HIV genome-wide protein associations: a review of 30 years of research. *Microbiol. Mol. Biol. Rev.*, **80**, 679–731.
7. Yelverton,E., Lindsley,D., Yamauchi,P. and Gallant,J.A. (1994) The function of a ribosomal frameshifting signal from human immunodeficiency virus-1 in *Escherichia coli*. *Mol. Microbiol.*, **11**, 303–313.
8. Flynn,W.F., Chang,M.W., Tan,Z.Q., Oliveira,G., Yuan,J.Y., Okulicz,J.F., Torbett,B.E. and Levy,R.M. (2015) Deep sequencing of protease inhibitor resistant HIV patient isolates reveals patterns of correlated mutations in Gag and protease. *PLoS Comput. Biol.*, **11**, e1004249.
9. Brakier-Gingras,L., Charbonneau,J. and Butcher,S.E. (2012) Targeting frameshifting in the human immunodeficiency virus. *Expert Opin. Ther. Targets*, **16**, 249–258.
10. Ofori,L.O., Hilimire,T.A., Bennett,R.P., Brown,N.W. Jr, Smith,H.C. and Miller,B.L. (2014) High-affinity recognition of HIV-1 frameshift-stimulating RNA alters frameshifting *in vitro* and interferes with HIV-1 infectivity. *J. Med. Chem.*, **57**, 723–732.
11. Cardno,T.S., Shimaki,Y., Sleebs,B.E., Lackovic,K., Parisot,J.P., Moss,R.M., Crowe-McAuliffe,C., Mathew,S.F., Edgar,C.D., Kleffmann,T. *et al.* (2015) HIV-1 and Human PEG10 frameshift elements are functionally distinct and distinguished by novel small molecule modulators. *PLoS One*, **10**, e0139036.
12. Hilimire,T.A., Bennett,R.P., Stewart,R.A., Garcia-Miranda,P., Blume,A., Becker,J., Sherer,N., Helms,E.D., Butcher,S.E., Smith,H.C. *et al.* (2016) N-methylation as a strategy for enhancing the affinity and selectivity of RNA-binding peptides: application to the HIV-1 frameshift-stimulating RNA. *ACS Chem. Biol.*, **11**, 88–94.
13. Dinman,J.D., Ruiz-Echevarria,M.J. and Peltz,S.W. (1998) Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends Biotechnol.*, **16**, 190–196.
14. Hung,M., Patel,P., Davis,S. and Green,S.R. (1998) Importance of ribosomal frameshifting for human immunodeficiency virus type 1 particle assembly and replication. *J. Virol.*, **72**, 4819–4824.
15. Marcheschi,R.J., Tonelli,M., Kumar,A. and Butcher,S.E. (2011) Structure of the HIV-1 frameshift site RNA bound to a small molecule inhibitor of viral replication. *ACS Chem. Biol.*, **6**, 857–864.
16. 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.
17. 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.
18. Kolakofsky,D. (2016) Paramyxovirus RNA synthesis, mRNA editing, and genome hexamer phase: a review. *Virology*, **498**, 94–98.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. Mehedi,M., Falzarano,D., Seebach,J., Hu,X.J., Carpenter,M.S., Schnittler,H.J. and Feldmann,H. (2011) A new ebola virus non structural glycoprotein expressed through RNA Editing. *J. Virol.*, **85**, 5406–5414.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. Wernegreen,J.J., Kauppinen,S.N. and Degnan,P.H. (2010) Slip into something more functional: selection maintains ancient frameshifts in homopolymeric sequences. *Mol. Biol. Evol.*, **27**, 833–839.
33. Olspert,A., Chung,B.Y.W., 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.
34. Untiveros,M., Olspert,A., Artola,K., Firth,A.E., Kreuze,J.F. and Valkonen,J.P. (2016) A novel sweet potato potyvirus ORF is expressed via polymerase slippage and suppresses RNA silencing. *Mol. Plant. Pathol.*, **17**, 1111–1123

35. Mingot,A., Valli,A., Rodamilans,B., San Leon,D., Baulcombe,D.C., Garcia,J.A. and Lopez-Moya,J.J. (2016) The *P1N-PISPO* trans-frame gene of sweet potato feathery mottle potyvirus is produced during virus infection and functions as an RNA silencing suppressor. *J. Virol.*, **90**, 3543–3557.
36. 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.*, **6**, doi:10.1038/srep21411.
37. Mathew,S.F., Crowe-McAuliffe,C., Graves,R., Cardno,T.S., McKinney,C., Poole,E.S. and Tate,W.P. (2015) The highly conserved codon following the slippery sequence supports -1 frameshift efficiency at the HIV-1 frameshift site. *PLoS One*, **10**, e0122176.
38. Penno,C., Kumari,R., Baranov,P.V., van Sinderen,D. and Atkins,J.F. (2017) Stimulation of reverse transcriptase generated cDNAs with specific indels by template RNA structure: Retrotransposon, dNTP balance, RT-reagent usage. *Nucleic Acids Res.*, doi:10.1093/nar/gkx689.
39. 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.
40. 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.
41. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, doi:10.14806/ej.17.1.200.
42. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
43. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Proc,G.P.D. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
44. Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., DePristo,M.A., Handsaker,R.E., Lunter,G., Marth,G.T., Sherry,S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
45. Bembom,O. (2016) SeqLogo: Sequence logos for DNA sequence alignments. *SeqLogo package version*. 1.40.0 **(**1990):1-5.
46. Howe,E.A., Sinha,R., Schlauch,D. and Quackenbush,J. (2011) RNA-Seq analysis in MeV. *Bioinformatics*, **27**, 3209–3210.
47. Dulude,D., Baril,M. and Brakier-Gingras,L. (2002) Characterization of the frameshift stimulatory signal controlling a programmed-1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res.*, **30**, 5094–5102.
48. Low,J.T., Garcia-Miranda,P., Mouzakis,K.D., Gorelick,R.J., Butcher,S.E. and Weeks,K.M. (2014) Structure and dynamics of the HIV-1 Frameshift element RNA. *Biochemistry*, **53**, 4282–4291.
49. Huang,X., Yang,Y., Wang,G., Cheng,Q. and Du,Z. (2014) Highly conserved RNA pseudoknots at the *gag-pol* junction of HIV-1 suggest a novel mechanism of -1 ribosomal frameshifting. *RNA*, **20**, 587–593.
50. Mouzakis,K.D., Dethoff,E.A., Tonelli,M., Al-Hashimi,H. and Butcher,S.E. (2015) Dynamic motions of the HIV-1 frameshift site RNA. *Biophys. J.*, **108**, 644–654.
51. Qiao,Q., Yan,Y., Guo,J., Du,S., Zhang,J., Jia,R., Ren,H., Qiao,Y. and Li,Q. (2016) A review on architecture of the *gag-pol* ribosomal frameshifting RNA in human immunodeficiency virus: a variability survey of virus genotypes. *J. Biomol. Struct. Dyn.*, **35**, 1629–1653.
52. Liao,P.Y., Choi,Y.S., Dinman,J.D. and Lee,K.H. (2011) The many paths to frameshifting: kinetic modelling and analysis of the effects of different elongation steps on programmed -1 ribosomal frameshifting. *Nucleic. Acids Res.*, **39**, 300–312.
53. Lin,Z., Gilbert,R.J. and Brierley,I. (2012) Spacer-length dependence of programmed -1 or -2 ribosomal frameshifting on a U6A heptamer supports a role for messenger RNA (mRNA) tension in frameshifting. *Nucleic Acids Res.*, **40**, 8674–8689.
54. Erickson-Viitanen,S., Manfredi,J., Viitanen,P., Tribe,D.E., Tritch,R., Hutchison,C.A. 3rd, Loeb,D.D. and Swanstrom,R. (1989) Cleavage of HIV-1 gag polyprotein synthesized *in vitro*: sequential cleavage by the viral protease. *AIDS Res. Hum. Retroviruses*, **5**, 577–591.
55. Henderson,L.E., Bowers,M.A., Sowder,R.C., Serabyn,S.A., Johnson,D.G., Bess,J.W., Arthur,L.O., Bryant,D.K. and Fenselau,C. (1992) Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences. *J. Virol.*, **66**, 1856–1865.
56. Pettit,S.C., Moody,M.D., Wehbie,R.S., Kaplan,A.H., Nantermet,P.V., Klein,C.A. and Swanstrom,R. (1994) The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.*, **68**, 8017–8027.
57. Wiegers,K., Rutter,G., Kottler,H., Tessmer,U., Hohenberg,H. and Krausslich,H.G. (1998) Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J. Virol.*, **72**, 2846–2854.
58. Muller,B., Anders,M., Akiyama,H., Welsch,S., Glass,B., Nikovics,K., Clavel,F., Tervo,H.M., Keppler,O.T. and Krausslich,H.G. (2009) HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 Infectivity. *J. Biol. Chem.*, **284**, 29692–29703.
59. Ning,J.Y., Erdemci-Tandogan,G., Yufenyuy,E.L., Wagner,J., Himes,B.A., Zhao,G.P., Aiken,C., Zandi,R. and Zhang,P.J. (2016) *In vitro* protease cleavage and computer simulations reveal the HIV-1 capsid maturation pathway. *Nat. Commun.*, **7**, doi:10.1038/ncomms13689.
60. Dam,E., Quercia,R., Glass,B., Descamps,D., Launay,O., Duval,X., Krausslich,H.G., Hance,A.J., Clavel,F. and ANRS 109 study group (2009) Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS Pathog.*, **5**, e1000345.
61. Doyon,L., Croteau,G., Thibeault,D., Poulin,F., Pilote,L. and Lamarre,D. (1996) Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J. Virol.*, **70**, 3763–3769.
62. Knops,E., Brakier-Gingras,L., Schulter,E., Pfister,H., Kaiser,R. and Verheyen,J. (2012) Mutational patterns in the frameshift-regulating site of HIV-1 selected by protease inhibitors. *Med. Microbiol. Immunol.*, **201**, 213–218.
63. Yan,N., O'Day,E., Wheeler,L.A., Engelman,A. and Lieberman,J. (2011) HIV DNA is heavily uracilated, which protects it from autointegration. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 9244–9249.
64. Lahouassa,H., Daddacha,W., Hofmann,H., Ayinde,D., Logue,E.C., Dragin,L., Bloch,N., Maudet,C., Bertrand,M., Gramberg,T. *et al.* (2012) SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat. Immunol.*, **13**, 223–228.
65. Hrecka,K., Hao,C., Gierszewska,M., Swanson,S.K., Kesik-Brodacka,M., Srivastava,S., Florens,L., Washburn,M.P. and Skowronski,J. (2011) Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature*, **474**, 658–661.
66. Laguette,N., Sobhian,B., Casartelli,N., Ringeard,M., Chable-Bessia,C., Segeral,E., Yatim,A., Emiliani,S., Schwartz,O. and Benkirane,M. (2011) SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*, **474**, 654–657.
67. Lenzi,G.M., Domaoal,R.A., Kim,D.H., Schinazi,R.F. and Kim,B. (2015) Mechanistic and kinetic differences between reverse transcriptases of *vpx* coding and non-coding lentiviruses. *J. Biol. Chem.*, **290**, 30078–30086.
68. 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.
69. 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.
70. Lapkouski,M., Tian,L., Miller,J.T., Le Grice,S.F.J. and Yang,W. (2013) Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. *Nat. Struct. Mol. Biol.*, **20**, 230–236.
71. Nowak,E., Miller,J.T., Bona,M.K., Studnicka,J., Szczepanowski,R.H., Jurkowski,J., Le Grice,S.F.J. and Nowotny,M.

(2014) Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry. *Nat. Struct. Mol. Biol.*, **21**, 389–396.
72. 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.
73. Baril,M., Dulude,D., Gendron,K., Lemay,G. and Brakier-Gingras,L. (2003) Efficiency of a programmed -1 ribosomal frameshift in the different subtypes of the human immunodeficiency virus type 1 group M. *RNA*, **9**, 1246–1253.
74. Blaby-Haas,C.E., Padilla-Benavides,T., Stube,R., Arguello,J.M. and Merchant,S.S. (2014) Evolution of a plant-specific copper chaperone family for chloroplast copper homeostasis. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E5480–E5487.
75. Meydan,S., Klepacki,D., Karthikeyan,S., Margus,T., Thomas,P., Jones,J.E., Khan,Y., Briggs,J., Dinman,J.D., Vazquez-Laslop,N. *et al.* (2017) Programmed ribosomal frameshifting generates a copper transporter and a copper chaperone from the same gene. *Mol. Cell*, **65**, 207–219.
76. Atkins,J.F., Loughran,G. and Baranov,P.V. (2017) A [Cu]rious ribosomal profiling pattern leads to the discovery of ribosomal frameshifting in the synthesis of a copper chaperone. *Mol. Cell*, **65**, 203–204.
77. Flower,A.M. and Mchenry,C.S. (1990) The gamma subunit of DNA Polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 3713–3717.
78. Tsuchihashi,Z. and Kornberg,A. (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 2516–2520.
79. 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.