



Template-switching mechanism of a group II intron-encoded reverse transcriptase and its implications for biological function and RNA-Seq

Received for publication, October 3, 2019, and in revised form, November 1, 2019. Published, Papers in Press, November 11, 2019, DOI 10.1074/jbc.RA119.011337

Alfred M. Lentzsch, Jun Yao, Rick Russell, and Alan M. Lambowitz¹

From the Institute for Cellular and Molecular Biology, Departments of Molecular Biosciences and Oncology, University of Texas at Austin, Austin, Texas 78712

Edited by Karin Musier-Forsyth

The reverse transcriptases (RTs) encoded by mobile group II introns and other non-LTR retroelements differ from retroviral RTs in being able to template-switch efficiently from the 5' end of one template to the 3' end of another with little or no complementarity between the donor and acceptor templates. Here, to establish a complete kinetic framework for the reaction and to identify conditions that more efficiently capture acceptor RNAs or DNAs, we used a thermostable group II intron RT (TGIRT; GsI-IIC RT) that can template switch directly from synthetic RNA template/DNA primer duplexes having either a blunt end or a 3'-DNA overhang end. We found that the rate and amplitude of template switching are optimal from starter duplexes with a single nucleotide 3'-DNA overhang complementary to the 3' nucleotide of the acceptor RNA, suggesting a role for non-templated nucleotide addition of a complementary nucleotide to the 3' end of cDNAs synthesized from natural templates. Longer 3'-DNA overhangs progressively decreased the template-switching rate, even when complementary to the 3' end of the acceptor template. The reliance on only a single bp with the 3' nucleotide of the acceptor together with discrimination against mismatches and the high processivity of group II intron RTs enable synthesis of full-length DNA copies of nucleic acids beginning directly at their 3' end. We discuss the possible biological functions of the template-switching activity of group II intron- and other non-LTR retroelement-encoded RTs, as well as the optimization of this activity for adapter addition in RNA- and DNA-Seq protocols.

This work was supported by National Institutes of Health Grants R01 GM37949 (to A. M. Lambowitz) and R35 GM131777 (to R. R.). Thermostable group II intron reverse transcriptase enzymes and methods for their use are the subject of patents and patent applications that have been licensed by the University of Texas and East Tennessee State University to InGex, LLC. A. M. Lambowitz, some former and present members of the Lambowitz laboratory, and the University of Texas are minority equity holders in InGex, LLC, and receive royalty payments from the sale of TGIRT enzymes and kits employing TGIRT template-switching activity for RNA-seq adapter addition and from the sublicensing of intellectual property to other companies. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article was selected as one of our Editors' Picks.

This article contains Figs. S1–S10 and Tables S1–S3.

RNA-Seq data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE138200.

¹ To whom correspondence should be addressed. Tel.: 512-232-3418; E-mail: lambowitz@austin.utexas.edu.

Reverse transcriptases (RTs)² typically have the ability to template-switch during cDNA synthesis, thereby joining discontinuous nucleic acid sequences. Such template switching plays a critical role in the replication cycle of retroviruses and other retroelements and also has biotechnological applications, particularly in RNA-Seq, where it can serve as an alternative to RNA tailing or ligation for adding RNA-Seq adapters to target RNAs (1–9). Although studied in greatest detail for retroviral RTs, the mechanism of template switching differs for the group II intron and other non-long terminal repeat (non-LTR) retroelement RTs in ways that are only just beginning to be fully appreciated and exploited for biotechnological applications.

Template switching was discovered in early studies of retroviral replication, which revealed an unexpected but crucial role for two template switches, also referred to as strand transfers, for the synthesis of full-length viral genomes (1–3). Both template switches occur when the RT reaches the end of one of the LTRs of the viral RNA and involve base pairing of a single-stranded cDNA, exposed by RNase H digestion of the RNA template, to a complementary sequence in the other LTR. Template switches between internal regions of the viral RNA are also common (10). Because retroviruses package two, often nonidentical, genomic RNAs and their RTs dissociate readily from RNA templates, template switching to complementary regions of another genomic RNA occurs frequently (3–12 times per replication cycle), leading to high rates of recombination between viral genomes (11, 12). Such recombination of genomic RNA sequences provides a means of rapidly propagating beneficial mutations that help retroviruses evade host defenses (13–15). Template switching by a retroviral RT is also employed for RNA-Seq adapter addition in a method called SMART-Seq (Switching Mechanism at the 5' end of the RNA Transcript-sequencing), which utilizes the ability of Moloney murine leukemia virus RT to add non-templated C residues to the 3' end of a completed cDNA, which can then base pair and enable template switching to the 3' end of an adapter DNA

² The abbreviations used are: RT, reverse transcriptase; 2'-O-Me, 2'-O-methyl; CV, column volume; LINE, long-interspersed nuclear element; LTR, long-terminal repeat; NTA, non-templated nucleotide addition; NTE, N-terminal extension; MRP, mitochondrial retroplasmid; mt, mitochondrial; RdRP, RNA-dependent RNA polymerase; TGIRT, thermostable group II intron RT; TGIRT-seq, RNA-Seq using thermostable group II intron RT; TPRT, target DNA-primed reverse transcription; nt, nucleotide; RNA-Seq, RNA-sequencing; ssDNA, single-strand DNA; PE, primer extension.

oligonucleotide ending with G residues (4, 7, 16). Biochemical studies showed that a related activity of HIV-1 RT (referred to as “clamping”) requires at least 2 bps between the 3' end of the cDNA and the acceptor template, one of which must be a GC or CG bp (17). Other retroviral and LTR retroelement RTs also required 2 bps for maximal clamping activity, but in some cases they showed weak activity (24–36%) with a single bp (18).

Unlike retroviral RTs, which rely on relatively stable base-pairing interactions for template switching, the RTs encoded by non-LTR retroelements have the ability to template-switch from the 5' end of a donor template to the 3' end of an acceptor template with little or no complementarity between the cDNA and the acceptor (5). Non-LTR retroelements are a broad class that includes non-LTR retrotransposons, such as human long interspersed nuclear element 1 (LINE-1) and insect R2 retroelements, as well as mitochondrial retroplasmids (MRPs) and bacterial and organellar mobile group II introns, which are evolutionary ancestors of non-LTR retrotransposons and retroviruses in eukaryotes (19, 20). The RTs encoded by group II introns and other non-LTR retroelements differ structurally from retroviral RTs, with their distinctive features including an N-terminal extension (NTE) typically containing a conserved sequence block (RT0) and two structurally-conserved insertions (RT2a and 3a) between the universally-conserved RT sequence blocks (RT1–7; Fig. 1A) (21).

Biochemical studies of template switching by three non-LTR retroelement RTs have revealed common features as well as differences that may reflect adaptations of this activity for the replication cycle of different retroelements. The Mauriceville and Varkud MRPs found in some *Neurospora* spp. strains are closely related, small closed-circular DNAs, which encode an RT that functions in plasmid replication (22, 23). The MRPs are transcribed to yield full-length transcripts with a 3'-tRNA-like structure and 3'-CCA, which are recognized by the RT for the initiation of cDNA synthesis at the penultimate C of the 3'-CCA either *de novo* (i.e. without a primer) or by using the 3' end of either a DNA or RNA primer with little or no complementarity to the RNA template (24, 25). Upon reaching the 5' end of the plasmid transcript, template switching to the 3' end of the same or another plasmid transcript yields multimeric cDNAs, which can undergo intramolecular recombination to regenerate closed-circular plasmid DNA (25–29). End-to-end template switching also occurs between the MRP transcript and heterologous mtDNA transcripts, leading to recombinant plasmids containing an integrated mtDNA sequence, frequently a mitochondrial tRNA (27, 30, 31). Such recombinant plasmids have a replicative advantage over the WT plasmid, as well as the ability to stably integrate into mtDNA by homologous recombination at the corresponding mtDNA locus without relying on a specialized DNA integrase (27, 30, 31).

The insect R2 element is a non-LTR retrotransposon that inserts site-specifically into the large subunit rRNA gene by a target-primed reverse transcription (TPRT) mechanism in which the RT nicks one strand of the rDNA and then uses the 3' end of the cleaved DNA strand as a primer to synthesize a full-length cDNA beginning at the 3' end of the R2 RNA (32). Biochemical studies showed that upon reaching the 5' end of an RNA template, the RT can template switch to the 3' end of

another RNA or DNA template and that this template switching is facilitated by non-templated nucleotide addition (NTA) of short stretches of complementary nucleotides (“microhomologies”) to the 3' end of the completed cDNA (33, 34). Such template switching, utilizing microhomologies generated by NTA, has been proposed to play a role in linking the 5' end of the R2 element to host genomic sequences upstream of the cleavage site and in the formation of chimeric integration products containing additional heterologous sequences (31–33, 35–37).

Finally, mobile group II introns, the type of non-LTR retroelement whose RT is studied here, are bacterial and organellar retrotransposons that insert site-specifically into DNA target sites by a process called retrohoming. In this process, the excised intron lariat RNA integrates directly into a DNA strand by reverse splicing and is reverse-transcribed by the RT using either an opposite strand nick or a nascent DNA strand as a primer for reverse transcription (20, 38). Group II intron RTs have an end-to-end template-switching activity similar to that of MRP and R2 element RTs (6, 39), and this activity has been utilized for adapter addition in RNA-Seq using a thermostable group II intron RT (TGIRT-seq) (6, 8, 9). Unlike the R2 element RT (33, 34), group II intron RTs do not need to be actively elongating to template-switch and can do so directly from synthetic RNA template/DNA primer substrates with a single-nucleotide (1-nt) 3'-DNA overhang that base pairs with the 3' nucleotide of an RNA acceptor (6). This reaction, which is related to clamping by retroviral RTs, enables the precise attachment of 3' RNA-Seq adapters to target RNAs or DNAs in a single step without tailing or ligation (8, 9, 40).

Recently, we determined a crystal structure of a full-length thermostable group II intron RT (Gsl-IIC RT; sold commercially as TGIRT-III) in complex with an RNA template/DNA primer substrate and incoming dNTP (41). Pertinent to template-switching activity, this structure revealed that the NTE and RT0 loop present in non-LTR retroelement RTs likely form a binding pocket for the 3' end of an acceptor nucleic acid, a structure that does not exist in retroviral RTs (Fig. 1A). Additionally, we found that mutations affecting the RT0 loop of Gsl-IIC RT, which forms the lid of the binding pocket, strongly inhibit template switching while minimally affecting primer-extension activity (41). A similar finding was reported for alanine substitution in the conserved RT0 motif of the R2 element RT (42), suggesting that the RT0-lidded binding pocket plays a similar role in template switching in other non-LTR retroelement RTs. The crystal structure of Gsl-IIC RT opens the possibility of detailed structure–function analysis of the template-switching activity of group II intron and homologous non-LTR retroelement RTs, as well as its optimization for RNA-Seq.

Here, we took advantage of the ability of Gsl-IIC RT to template-switch directly from synthetic RNA template/DNA primer substrates to carry out a detailed kinetic analysis of the template switching and related NTA activities of a group II intron RT. We found that the rate and amplitude of template switching are optimal from starter duplexes with a single-nucleotide 3'-DNA overhang complementary to the 3' nucleotide of the acceptor RNA. We also show that this single bp to the 3' nucleotide of a target RNA dictates template switching with

high efficiency and yields seamless junctions, whereas template switching from a blunt-end duplex is inefficient and yields heterogeneous junctions. Our results inform possible biological functions of template switching for non-LTR retroelement RTs and suggest how to further optimize this activity for adapter addition in RNA-Seq.

Results

Overview of the template-switching reaction and determination of saturating enzyme concentrations

Fig. 1B depicts the assay that we used to analyze the template-switching activity of GsI–IIC RT, which is based on the method used for 3'-adapter addition in TGIRT-seq (6, 8, 9). In this assay, GsI–IIC RT bound to an artificial RNA template/DNA primer duplex (termed “starter duplex” or “donor”) initiates cDNA synthesis by template switching to the 3' end of an acceptor nucleic acid. For most of our experiments, the starter duplex was the same as that used for TGIRT-seq and consists of a 34-nt RNA oligonucleotide that contains an Illumina R2 adapter sequence and has a 3' blocking group (3SpC3; IDT) annealed to a complementary 35-nt DNA primer termed R2R (Table S1). The latter leaves a 1-nt 3' overhang (N) that can direct template switching by base pairing to the 3' nucleotide (N') of the acceptor nucleic acid (6). Upon the addition of dNTPs (an equimolar mix of dATP, dCTP, dGTP, and dTTP), GsI–IIC RT initiates reverse transcription by template switching to the 3' end of the acceptor nucleic acid and synthesizes a full-length cDNA with the R2R DNA primer attached to its 5' end. In biochemical experiments, the DNA primer was 5' end-labeled, allowing the products to be analyzed and quantified by denaturing PAGE (Fig. 1B, bottom left). Alternatively, in some experiments, we analyzed products by RNA-Seq, which involves the ligation of a 5'-adenylated R1R adapter to the 3' end of the cDNA followed by PCR to add capture sites and indexes for Illumina sequencing (Fig. 1B, bottom right).

To determine an optimal enzyme concentration for the reaction, we measured the rate and amplitude of the cDNA product formation at different concentrations of GsI–IIC RT using 100 nM of a 50-nt acceptor RNA with a 3'-C residue and 50 nM of a starter duplex with a complementary 1-nt 3'-G overhang in reaction medium containing 200 mM NaCl (Fig. 1C). At all enzyme concentrations, we observed a prominent template-switching product containing the labeled DNA primer attached to a full-length cDNA of the RNA acceptor (denoted 1x), as well as a ladder of larger products resulting from sequential end-to-end template switches from the 5' end of one acceptor RNA to the 3' end of another acceptor RNA (denoted 2x, 3x, etc.; Fig. 1C). In addition to template-switching products, a fraction of the DNA primer was extended by only 1–3 nucleotides due to NTA to the 3' end of the DNA primer (bands indicated by *star* in the gel of Fig. 1C).

Quantitation of the products resulting from template switching and extension by the RT (defined as those >2 nt larger than the primer) revealed a dominant fast phase of product formation and a minor slow phase, the latter possibly reflecting heterogeneity in the substrate and/or enzyme. For simplicity, we fit the data in this and subsequent figures using a single-exponen-

tial function to define the properties of the dominant fast phase. We found that 350–500 nM enzyme gives maximal values of the observed rate constant ($\sim 0.25\text{ s}^{-1}$) and reaction amplitude (0.85) (Fig. 1C). At these enzyme concentrations, nearly all of the starter duplex (50 nM) was utilized, and excess acceptor RNA was used for multiple template switches at the longer time points. Therefore, we used 500 nM GsI–IIC RT in all subsequent reactions.

Template switching to RNA or DNA is more efficient at lower salt concentrations

We next analyzed the effects of salt concentration on the kinetics of template switching to RNA and DNA acceptor oligonucleotides (Fig. 2). The observed rate constants for template switching to a 50-nt acceptor RNA or DNA were the same within error and relatively unaffected by NaCl concentrations between 100 and 300 mM, but decreased 2- to 3-fold at 400 mM NaCl (Fig. 2). The maximum amplitude of product formation was somewhat higher for the RNA than the DNA acceptor (0.85 and 0.77, respectively, at 200 mM NaCl), but in both cases decreased by $\sim 40\%$ at 400 mM NaCl, as did the proportion of products resulting from multiple template switches (Fig. 2). Additional experiments showed that neither the acceptor RNA nor the enzyme became limiting at 400 mM NaCl (Fig. S1), suggesting that the high-salt concentration may adversely affect the conformation of the enzyme or substrates for template switching. By contrast, in a parallel primer extension reaction in which GsI–IIC RT initiated synthesis from a DNA primer annealed to the 3' end of a 1.1-kb RNA template, the rate of extension decreased from $\sim 9\text{ nt/s}$ at 100 mM NaCl to $\sim 4\text{ nt/s}$ at 400 mM NaCl but reached the same amplitude irrespective of salt concentration (Fig. S2). These findings indicate that the decreased amplitude of reverse transcription reactions that require template switching results from decreased efficiency of the initial template switch and not the subsequent reverse transcription. A reciprocal experiment using nondenaturing gel electrophoresis to examine utilization of a 5'- ^{32}P -labeled acceptor RNA showed that a higher proportion of the acceptor was utilized for template switching at 200 mM than at 450 mM NaCl, a salt concentration used in TGIRT-seq to suppress NTA and multiple template switches (63 and 19% of acceptor utilized at 200 and 450 mM NaCl, respectively; Fig. S3). Based on these results, we used 200 mM NaCl in most subsequent experiments.

Lower salt concentration increases template switching to RNAs ending with 3'-phosphate or 2'-O-Me groups

Previous work showed that in a reaction medium containing 450 mM NaCl, template switching of group II intron RTs is most efficient to RNAs with a 3'-OH group and strongly inhibited for RNAs with either a 3'-phosphate or 2'-O-Me group (6). In light of our finding that template switching is more efficient at lower salt concentrations, we assayed the template-switching activity of GsI–IIC RT to otherwise identical 21-nt RNA acceptors with these modifications in reaction media containing 200 mM NaCl and made parallel measurements at 450 mM NaCl for comparison (Fig. 3). We found that template switching to RNAs with a 3'-phosphate or a 2'-O-Me group was substantially more efficient at 200 mM NaCl than at 450 mM NaCl, albeit with rates

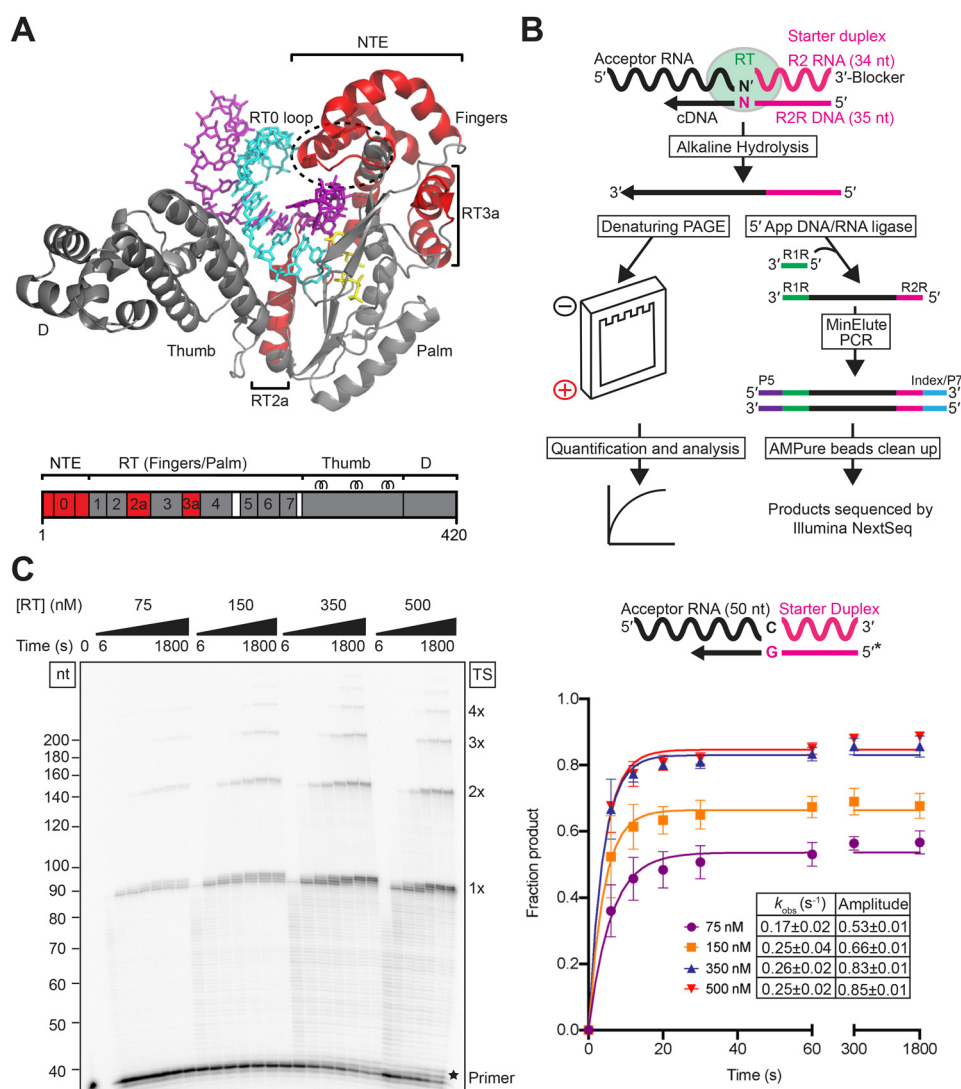


Figure 1. Overview of template-switching experiments and determination of saturating enzyme concentrations. *A*, structure of Gsl-IIC RT bound to RNA template/DNA primer and incoming dNTP (Protein Data Bank code 6AR1) (41). The NTE, RT2a, and RT3a insertions, which are not present in retroviral RTs, are colored *red* and delineated by *brackets*, with the RT0 loop *encircled by a dashed line*. Other protein regions are labeled and colored *gray*. The DNA primer and RNA template are shown in *stick* representation and are colored *cyan* and *purple*, respectively, and dATP bound at the RT active site is also shown in *stick* representation and colored *yellow*. The Gsl-IIC RT used to obtain the crystal structure has a C-terminal His₈ tag, whereas that used for biochemical analysis has an N-terminal maltose-binding protein tag to keep the protein soluble in the absence of bound nucleic acids (6). The schematic at the *bottom* shows the Gsl-IIC RT protein with different regions color-coded to the crystal structure. RT-1 to RT-7 are conserved RT sequence blocks found in all RTs. D denotes the C-terminal DNA-binding domain that functions in recognition of DNA target sites during retrohoming (38). *B*, outline of template-switching experiments. Gsl-IIC RT was pre-bound to a starter duplex (*magenta*) consisting of a 34-nt RNA oligonucleotide containing an Illumina Read 2 (R2) sequence annealed to a complementary 35-nt DNA primer (R2R) leaving a 1-nt, 3'-DNA overhang (N) (Table S1). The 3' overhang nucleotide (N) base pairs with the 3' nucleotide (N') of an acceptor RNA (*black*) for template switching, leading to the synthesis of a full-length cDNA of the acceptor RNA with the R2R adapter linked to its 5' end. The cDNAs were incubated with NaOH to degrade RNA and neutralized with equimolar HCl prior to further analysis (see "Experimental procedures" for details). For the biochemical experiments (*left branch*), the R2R DNA primer in the starter duplex was 5'-³²P-labeled, and the cDNA products were analyzed by electrophoresis in a denaturing polyacrylamide gel, which was dried and quantified with a phosphorimager. For RNA-Seq experiments (*right branch*), the cDNAs were cleaned up by using a Qiagen MinElute column (not shown) prior to ligating a 5'-adenylated R1R adapter to the 3' end of the cDNA using the Thermostable 5' App DNA/RNA Ligase (New England Biolabs). After another MinElute cleanup, Illumina RNA-Seq capture sites (P5 and P7) and indexes were added by PCR, and the resulting libraries were cleaned up by using AMPure XP beads (Beckman Coulter) prior to sequencing on an Illumina NextSeq 500. *C*, determination of saturating enzyme concentrations. Template-switching reactions included various concentrations of Gsl-IIC RT as indicated, 50 nM RNA template/DNA primer starter duplex (5'-³²P-labeled on DNA primer indicated by *), 100 nM of a 50-nt RNA acceptor template, and 4 mM dNTPs (an equimolar mix of 1 mM dATP, dCTP, dGTP, and dTTP) in reaction medium containing 200 mM NaCl at 60 °C. Aliquots were quenched at times ranging from 6 to 1,800 s, and the products were analyzed by denaturing PAGE, as described under "Experimental procedures." The *numbers to the left* of the gel indicate size markers (a 5'-³²P-labeled ssDNA ladder; ss20 DNA Ladder, Simplex Sciences) run in a parallel lane, and the *labels to the right* of the gel indicate the products resulting from the initial template switch (1×) and subsequent end-to-end template switches from the 5' end of one acceptor to the 3' end of another (2×, 3×, etc.). The *star at the bottom right* of the gel indicates the position of bands resulting from NTA to the 3' end of the DNA primer. The plot at *right* shows time courses for the production of template-switching products (*i.e.* products >2 nt larger than the primer), with each data set fit by a single-exponential function, and the *error bars* indicate the standard deviations for three experiments. The *inset table* indicates the k_{obs} and amplitude parameters obtained from the fit of an exponential function to the average values from three independent determinations, along with standard errors obtained from the fit (see "Experimental procedures").

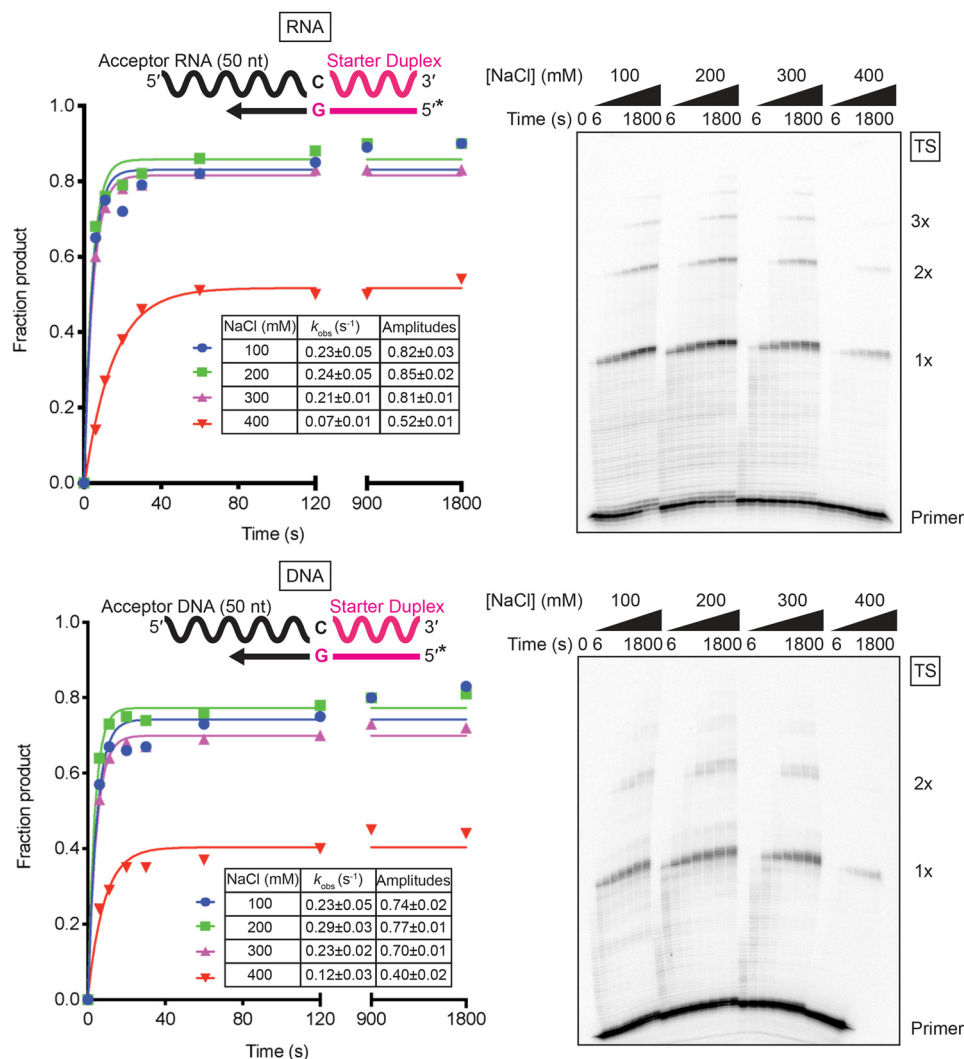


Figure 2. Template switching to RNA and DNA is more efficient at lower salt concentrations. Template-switching reaction time courses with 100 nM of 50-nt RNA (top) or DNA (bottom) acceptors of identical sequence (Table S1) were done with 500 nM GsI-IIC RT and 50 nM starter duplex with a 5'-³²P-labeled (*) DNA primer in reaction medium containing 100, 200, 300, or 400 mM NaCl. Time points were taken at intervals ranging from 6 to 1,800 s, and the products were analyzed by denaturing PAGE, as described in Fig. 1B. The plots to the left of the gel show the data fit by a single-exponential function to calculate the k_{obs} and amplitude for each time course, and the values are summarized in the inset tables together with the standard error of the fit. The gels are labeled as in Fig. 1.

and amplitudes remaining lower than those for the same RNA with a 3'-OH group (Fig. 3). The inhibitory effects of these modifications presumably reflect that they affect the binding or alignment of the 3' end of the acceptor nucleic acid in the template-switching pocket. We also examined template switching to DNA acceptors of identical sequence without or with a didoxynucleotide at their 3' end. Template switching was not impeded by the absence of a 3'-OH group, indicating that recognition of a 2'- or 3'-OH group of the acceptor is not required for template switching.

Non-templated nucleotide addition activity of GsI-IIC RT using mixed dNTPs

The ability of GsI-IIC RT to add non-templated nucleotides to the 3' end of the DNA primer in the starter duplex could affect the efficiency of template switching positively or negatively. To probe this NTA activity quantitatively, we incubated GsI-IIC RT with a blunt-end version of the starter duplex and varying concentrations of an equimolar mixture of all four

dNTPs. Gel analysis of NTA reactions showed progressive addition of up to three nucleotides (Fig. 4A). The product band reflecting addition of a single nucleotide accumulated and persisted throughout the reaction, indicating that the first nucleotide addition was faster than the second and/or that a fraction of the complex entered a paused or stopped state, a scenario discussed further below. Accordingly, the intensities of all of the product bands were summed to estimate the rate of step 1, and product bands 2 and 3 were summed to estimate the rate of step 2, and product band 3 was used to estimate the rate of step 3 (Fig. 4B and Fig. S4A).

Our analysis of the rate dependence on dNTP concentration showed that the first dNTP addition was the fastest and most efficient, with a maximal rate constant (k_{add}) of $0.064 \pm 0.002 s^{-1}$ and a second-order rate constant ($k_{add}/K_{1/2}$) of $66 M^{-1} s^{-1}$ (Fig. 4B and Fig. S4A, left). The observed rate constants for the second and third NTA were lower than that for the first NTA at each dNTP concentration and did not approach a maximum at 4 mM dNTP (Fig. 4B and Fig. S4A, right), resulting in lower

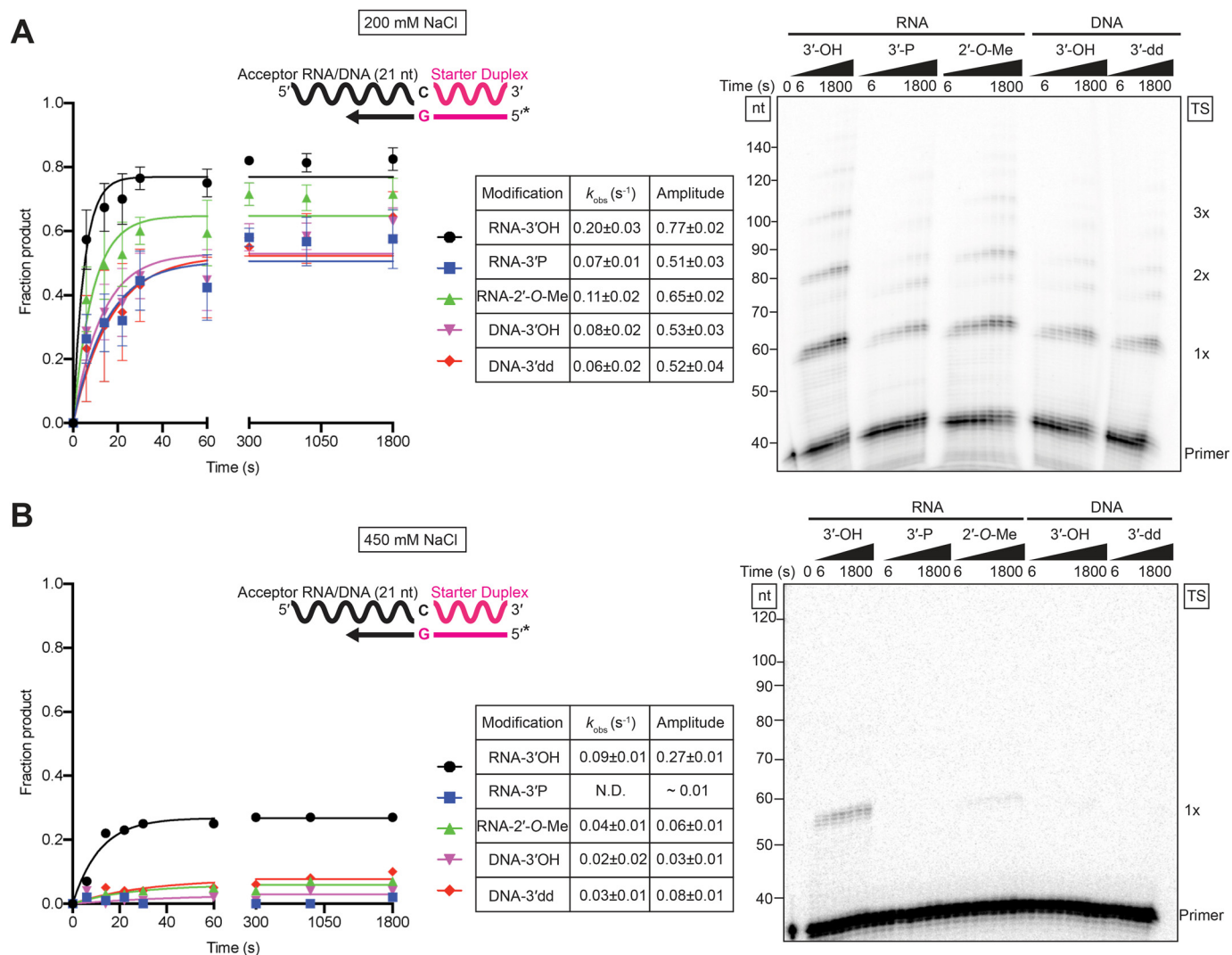


Figure 3. Lower salt concentration increases template switching to acceptor RNAs ending with 3'-phosphate or 2'-O-Me groups. *A* and *B*, time courses of template switching to 21-nt RNA and DNA acceptors of identical sequence but different 3' end modifications (Table S1) at 200 and 450 mM NaCl, respectively. The RNA acceptors had a 3-hydroxyl (OH), 3'-phosphate (3'-P), or 2'-O-methyl (2'-O-Me) group, and the DNA acceptors had a 3' hydroxyl (OH) or a dideoxy (3'-dd) terminus. Reactions were done using 500 nM Gsl-IIIc RT, 100 nM acceptor RNA or DNA, and 50 nM starter duplex with a 5'-³²P-labeled (*) DNA primer. Time points were taken at intervals ranging from 6 to 1,800 s, and the products were analyzed by denaturing PAGE, as described in Fig. 1. The plots to the left of the gel show the data fit by a single-exponential function to calculate the k_{obs} and amplitude for each reaction, and the values and standard errors of the fit are shown in the tables to the right of the plots. The gels are labeled as in Fig. 1. N.D., not determined.

$k_{\text{add}}/K_{1/2}$ values of 8 and $3 \text{ M}^{-1} \text{ s}^{-1}$, respectively. Thus, each step of NTA is progressively slower, with the second step ~ 8 -fold slower than the first step, and the third step even slower than the second step (Fig. 4B). We also measured NTA using a starter duplex with a 1-nt overhang (Fig. S4B). As expected, the kinetic parameters for addition of the first nucleotide from a 1-nt overhang duplex were similar to those determined for the second step starting from a blunt-end starter duplex (compare Fig. S4B with Fig. 4A and Fig. S4A). Thus, the data indicate that the first NTA occurs most efficiently, whereas the second and third additions are substantially and increasingly less efficient.

Nucleotide preferences for non-templated nucleotide addition activity

The above reactions examined the kinetics of NTA with mixed dNTPs. To probe the nucleotide preferences of NTA activity, we compared NTA reactions performed with each of

the four dNTPs individually. This comparison showed that the order of efficiencies of the first NTA from a blunt-end duplex was $A > G \gg C \approx T$ (Fig. 5A, Fig. S5, and Table S2), indicating a preference for purines and suggesting that incorporation is largely governed by the "A-rule," which is followed by a variety of DNA polymerases and RTs (34, 43–46). Interestingly, for the first NTA from a starter duplex with a 3'-G overhang (the equivalent of a second NTA for a blunt-end duplex), dATP addition was even more strongly preferred (7–17-fold higher $k_{\text{add}}/K_{1/2}$ value than for dGTP, dTTP, or dCTP; Fig. 5A).

Closer analysis of the NTA progress curves revealed that even at the reaction endpoints, significant fractions of the primer do not undergo the maximum of 3 to 4 NTAs (e.g. Fig. 4A), suggesting complex kinetics with potential side reactions and/or pausing and stopping mechanisms. To explore these possibilities and to generate a model that recapitulates all features of the data, we analyzed the single dNTP reactions by

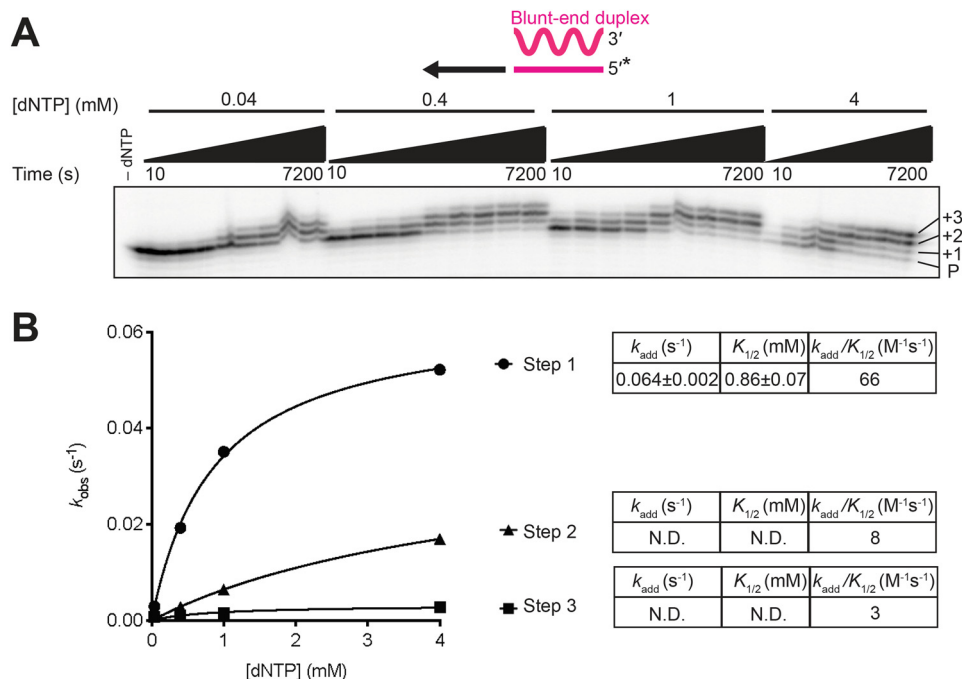


Figure 4. Non-templated nucleotide addition activity of GsI-III RT using a mixture of all four dNTPs. Reactions included 500 nM GsI-III RT and 50 nM of a blunt-end starter duplex with 5'-³²P-labeled (*) DNA primer in reaction medium containing 200 mM NaCl and varying dNTP concentrations (0.04, 0.4, 1, and 4 mM, where 4 mM is an equimolar mix of 1 mM dATP, dCTP, dGTP, and dTTP). Aliquots were stopped after times ranging from 10 to 7,200 s, and the products were analyzed by electrophoresis in a denaturing polyacrylamide gel, which was dried and scanned with a phosphorimager. Each product band was quantified individually and summed to estimate the rate of step 1. Product bands 2 and 3 were summed to estimate the rate of step 2, and product band 3 was used to estimate the rate of step 3. The data were plotted and fit by a single-exponential function to calculate the k_{obs} and amplitude parameters for each reaction. *A*, representative gel showing the labeled DNA primer (*P*) and bands resulting from NTA of 1, 2, and 3 nucleotides to the 3' end of the DNA. *B*, plot of k_{obs} values as a function of dNTP concentration fit by a hyperbolic function to calculate k_{add} , the catalytic rate at saturating substrate concentration; $K_{1/2}$, the substrate concentration at half-maximum k_{add} ; and $k_{add}/K_{1/2}$ for each NTA step (values summarized in tables to the right of the plots). The individual parameter values k_{add} and $K_{1/2}$ were not well-defined for steps two and three because saturation was not reached at 4 mM dNTP, and they are therefore indicated as *N.D.* (not determined). Although the progress curve of the second and third NTA products (Fig. S4) would be expected to include kinetic lags in principle, the rate constants for NTA are progressively lower with repeated additions, such that the data for these additions are adequately described by simple exponential functions without lag phases. All reactions were performed at least twice, and some time points were collected three times. Data were averaged for each time point, and these averages were fit by a single-exponential function to obtain the k_{obs} values.

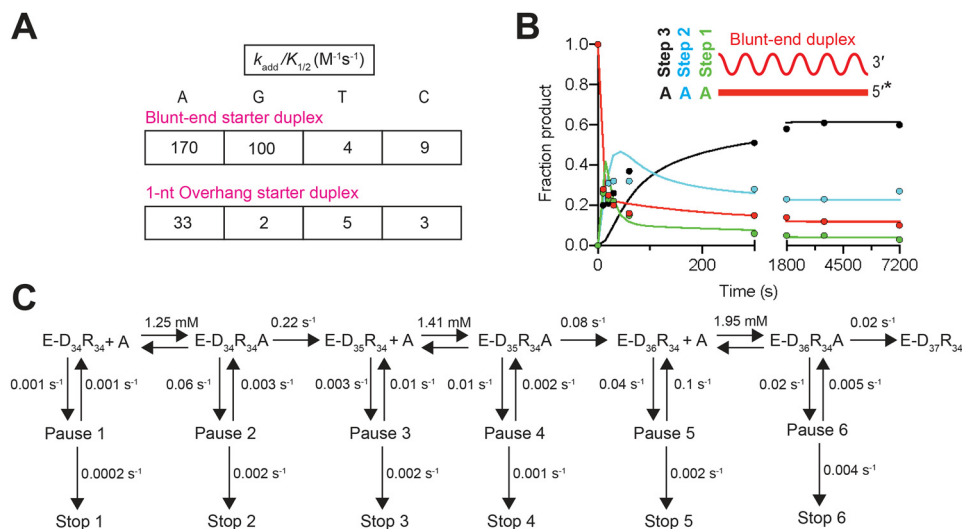


Figure 5. Nucleotide preferences for non-templated nucleotide addition activity. *A*, second-order rate constants of NTA activity using individual dNTPs for the first step (blunt-end starter duplex, top) and second step (1-nt G overhang starter duplex, bottom) are shown. The kinetic parameters for k_{add} and $K_{1/2}$ were obtained as described in Fig. 4. Gels and plots are shown in Fig. S5; reaction conditions are given in the Fig. S5 legend, and kinetic values are summarized in Table S1. *B*, plots showing global data fitting of consecutive dATP addition (Fig. S5) to a blunt-end starter duplex at 4 mM dATP. Each color represents a unique species in the reaction pathway: red, blunt-end starter duplex; green, product after first NTA; blue, product after second NTA; black, product after third NTA. *C*, global model used for NTA reactions with dATP is shown with the parameters obtained from global fitting. Analogous schemes for NTA of the other dNTPs are shown in Fig. S6.

global simulation and fitting (Fig. 5, B and C, and Figs. S5 and S6) (47). For each dNTP, the NTA progress curves using four concentrations of nucleotide were fit globally to reaction

schemes that include up to three rapid-equilibrium nucleotide-binding steps, the second and third of which require translocation, and up to three steps of nucleotide addition (Fig. 5C and

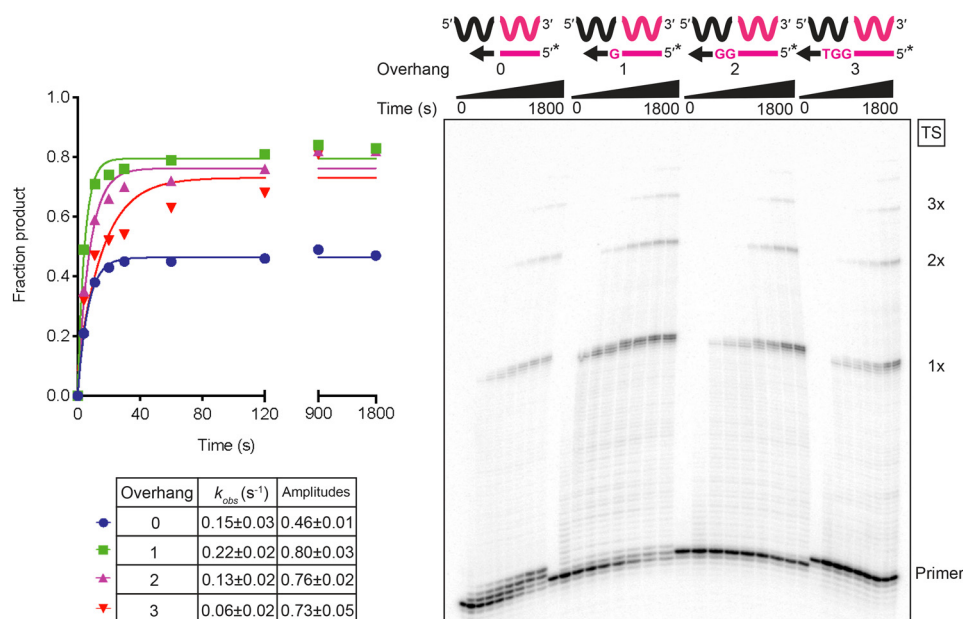


Figure 6. Template switching is favored by a single nucleotide 3' overhang. Reactions used 50 nM starter duplexes with a blunt end (0 overhang) or 1-, 2-, or 3-nt 3' overhangs, 500 mM Gsl-IIC RT, 100 nM 50-nt acceptor RNA in reaction medium containing 200 mM NaCl and were done at 60 °C. The schematics above the gel diagrams the reactions with starter duplexes having a 5'-³²P-labeled (*) DNA primer and different numbers of 3' overhang nucleotides. Reactions were stopped after times ranging from 5 to 1,800 s, and the products were analyzed by denaturing PAGE and quantified from phosphorimager scans of the dried gel. The plots show the data fit by a single-exponential function to calculate the k_{obs} and amplitude for each time course, and the values are summarized in the table below together with the standard error of the fit. The gel is labeled as in Fig. 1.

Fig. S6). Supporting and extending the conclusions from the mixed dNTP reactions (Fig. 4), for each dNTP the first addition is the most efficient and the second and third additions are progressively less efficient. Interestingly, the best-fit models also include both paused and stopped states, as suggested by the multiphase kinetics and lack of complete decay of intermediates (Fig. 5B and Fig. S6). Paused states have been observed for translesion polymerases, with pausing occurring when the 3' end of the primer is in a pretranslocation state occupying the nucleotide-binding pocket (48, 49). The stopped state may reflect a further inhibited enzyme state that occurs during NTA or dissociation of the enzyme.

We also explored the starter duplex requirements for NTA (Fig. S7). NTA product formation was slightly higher for a blunt-end RNA/DNA starter duplex than for a blunt-end DNA/DNA starter duplex of the same nucleotide sequence. However, in contrast to terminal transferase activity (50), NTA by Gsl-IIC RT did not occur efficiently to a ssDNA primer under our experimental conditions (Fig. S7), likely reflecting that an annealed template strand is required for productive binding of the enzyme.

Overall, our findings show that purines are preferred for the first NTA, and dATP is strongly preferred for the second NTA step. The findings that the first nucleotide addition from a blunt-end duplex is faster and more efficient than subsequent additions may be relevant to the physiological functions of NTA and template switching by group II intron RTs (see "Discussion").

Template switching is favored by a single nucleotide 3' overhang

The difference in efficiency between the first and subsequent NTAs by Gsl-IIC RT prompted us to investigate template-

switching activity from different lengths of 3'-DNA overhangs. To this end, we compared the template-switching activity from a blunt-end starter duplex to that from starter duplexes with 1-, 2-, or 3-nt 3'-DNA overhangs complementary to the 3' end of the 50-nt RNA or DNA acceptor oligonucleotides (Fig. 6 and Fig. S8, respectively). In both cases, the reaction occurred at the highest rate and amplitude with a 1-nt 3' overhang, and the rate decreased with longer 3' overhangs, despite the complete complementarity of the overhang with the 3' end of the acceptor. Template switching from the blunt-end starter duplex to the acceptor RNA (0.15 s^{-1}) was somewhat slower than from a 1-nt overhang (0.22 s^{-1}), but faster than NTA to the same blunt-end duplex at the same dNTP concentration (0.064 s^{-1} ; Fig. 4B). Similar trends were seen for template switching to a DNA acceptor, including that template switching from a blunt-end duplex is slower than that from a duplex with a 1-nt overhang but faster than NTA under the same conditions (Fig. S8).

Interestingly, the amplitude of template switching from the blunt-end duplex was substantially lower than those for starter duplexes with complementary 1-, 2-, or 3-nt 3' overhangs, and there was a prominent accumulation of products extended by just 1–3 nucleotides, most likely by NTA (see Fig. 6 and Fig. S8 for RNA and DNA acceptors, respectively). The increased NTA observed in the reaction with the blunt-end duplex likely reflects that the absence of a complementary 3' overhang nucleotide hinders productive binding of the acceptor for template switching, thereby favoring the competing NTA reaction. The lower rates of template switching from starter duplexes with complementary 2- and 3-nt 3' overhangs could reflect that the template-switching pocket is optimized for the formation of a single bp that

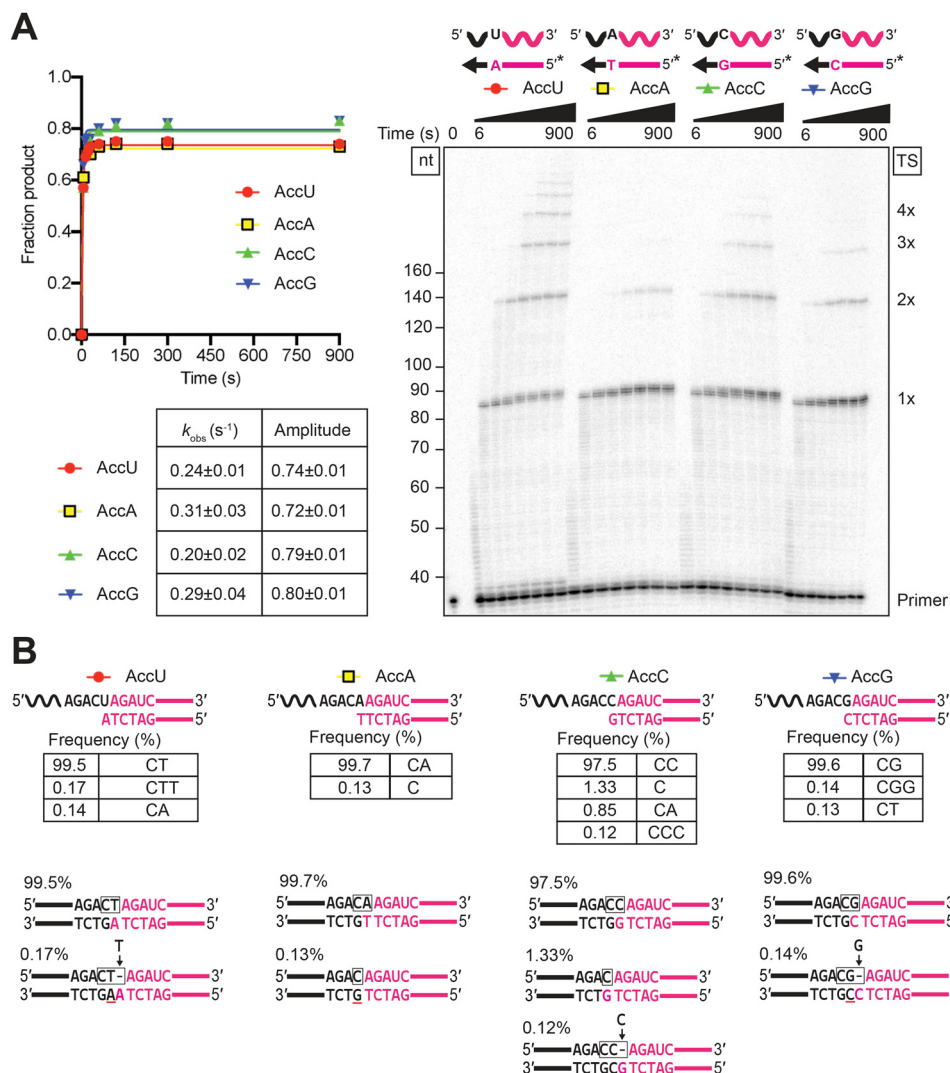


Figure 7. Template switching is directed by a single base pair between the 1-nt 3' DNA overhang and the 3' nucleotide of the acceptor RNA. Template-switching reactions with 50-nt acceptor RNAs (100 nM) differing only in their 3' nucleotide and starter duplexes having a complementary 1-nt 3' DNA (50 nM) were done with 500 nM GsI-III RT at 60 °C in reaction medium containing 200 mM NaCl. **A**, time courses. Template-switching reactions with ³²P-labeled (*) starter duplexes were stopped after times ranging from 6 to 900 s, and the products were analyzed by denaturing PAGE and quantified from a phosphorimager scan of the dried gel. The gel (right) is labeled as in Fig. 1. The plots (left) show the time course data fit by a single-exponential function to calculate the k_{obs} and amplitude for each time course, with the values and standard error of the fit summarized in the table below. **B**, RNA-Seq analysis of template-switching reactions. Template-switching reactions with unlabeled starter duplexes were done under the same conditions as in **A** with the reaction stopped after 15 min. RNA-Seq analysis was done as described in Fig. 1 and under "Experimental procedures." The figure diagrams the template-switching reaction for each acceptor/starter duplex combination above, with the sequences and percentages of the most frequent template-switching junctions ($\geq 0.1\%$) for each combination listed below. Nucleotides derived from the acceptor are in black, and nucleotides derived from the starter duplex are in red, with the box indicating the junction sequence. A black letter with red underline indicates a nucleotide inferred to result from NTA to the 3' end of the DNA primer. A gap in the top strand due to NTA is shown as a dash, and nucleotides inferred to fill the gap after PCR to add RNA-Seq adapters (Fig. 1A) are shown above the line with an arrow pointing to the gap. The low frequency junctions containing an extra nucleotide (CTT for the 3'-U acceptor (0.17%), CCC for the C acceptor (0.12%), and CGG for the 3'-G acceptor (0.14%)) can be explained by template switching from donors that have undergone an NTA of a complementary nucleotide resulting in a 2'-nt 3' overhang that leaves a gap in the top strand, which is filled by a complementary nucleotide during the PCR used to add RNA-Seq adapters. Other aberrant products may reflect heterogeneity or resections at the 3' ends of the synthesized oligonucleotides (e.g. the 3'-C junction for the 3'-C acceptor (1.33%) can be explained by template switching to a mis-synthesized or resected acceptor RNA lacking the terminal C residue). Complete data for junction sequences are shown in Table S3.

positions the 3' end of the acceptor for continued cDNA synthesis at the RT active site (see "Discussion"). NTA products were also lower with these longer 3' overhangs, as expected for the lower efficiency of NTA to longer 3' overhangs (see Fig. S5). The finding that the rate of template switching was highest for a complementary 1-nt 3' overhang indicates that a single bp is optimal for binding the acceptor in a productive conformation for template switching.

Template-switching fidelity is dictated by a single base pair between the 3'-DNA overhang and the 3' nucleotide of the acceptor RNA

After determining that a 1-nt 3' overhang is optimal for template switching, we wanted to compare the kinetics and fidelity of template switching for all four possible bp combinations between the 1-nt 3' overhang and the 3' nucleotide of the acceptor RNA. As shown in Fig. 7A, template switching

occurred efficiently for all four combinations, with indistinguishable rates but slightly higher amplitudes for rC/dG and rG/dC than for rA/dT or rU/dA. The differences in amplitude may be due to a kinetic competition between template switching and the competing NTA reaction, which is sensitive to small differences in the rate of template switching that are not readily detectable in the observed rate constants because of the rapid time scale of the reactions.

Sequencing of the template-switching junctions between the starter duplex and the acceptor RNA showed accurate extension of all four combinations, with 97.5–99.7% of the products having the seamless junctions expected for extension of the single bp (Fig. 7B). The sequencing also showed a smattering of aberrant products due to the competing NTA reaction or impurities in the synthetic oligonucleotides, as described in the legend of Fig. 7B. The high efficiency of seamless template switching dictated by a single bp between the 1-nt 3' overhang and the 3' nucleotide of the acceptor presumably reflects a very tight and specific interaction of complementary 3' termini within the binding pocket for the acceptor RNA.

Nucleotide sequence biases in TGIRT-seq reflect the efficiency of template switching to different 3' nucleotides of acceptor RNAs at high-salt concentrations

Previous TGIRT-seq analysis of miRNA reference sets showed that 3'-sequence biases in TGIRT-seq are almost completely confined to the 3'-terminal nucleotide of the acceptor, with a preference for miRNAs with a 3'-G residue and against those with a 3'-U residue (51). In light of our finding that template switching is less efficient at high-salt concentrations (Fig. 2), we wondered whether the high-salt concentration used in TGIRT-seq to suppress multiple template switches (450 mM NaCl (8, 9)) might exacerbate differences between acceptor RNAs with different 3' nucleotides, which are muted at 200 mM NaCl (Fig. 7).

To test this idea, we compared template switching to acceptor RNAs with all four Watson-Crick base-pairing combinations between the single-nucleotide 3' overhang and 3' nucleotide of the acceptor at 450 mM NaCl. Under these conditions, we found larger differences in the amplitude of the reaction for acceptors with different 3' nucleotides in order $G > C \approx A > U$ (Fig. 8A), which match the previously determined nucleotide sequence biases for TGIRT-seq (51). Additionally, when the reactions were carried out with lower concentrations of acceptor RNAs (10 nM) to more closely resemble the conditions during TGIRT-seq, the difference in amplitude for template switching to the most and least favored 3' nucleotides (G and U, respectively) increased (Fig. 8B), as expected if productive binding of the acceptor became limiting under these conditions. The more uniform rates and amplitudes for template switching to acceptor RNAs with different 3' nucleotides at 200 mM NaCl (Fig. 7) suggest that the nucleotide sequence biases for 3' adapter in TGIRT-seq might be ameliorated by carrying out the initial template-switching step at a lower salt concentration.

Template switching from a 1-nt overhang duplex disfavors the extension of mismatches

To further investigate how GsI–IIC RT template switching discriminates between complementary and mismatched acceptors, we carried out parallel reactions comparing a complementary 3'-C acceptor and 3'-G overhang combination to a mismatched 3'-C acceptor and 3'-C overhang combination (Fig. 9). Both the rate and amplitude of product formation were substantially decreased for the mismatched combination, with the rate ~ 8 -fold lower ($k_{\text{obs}} = 0.03$ compared with 0.23 s^{-1}) and the amplitude $\sim 50\%$ lower than those for the complementary combination (0.43 compared with 0.88, although the slow phase appeared to be continuing past 900 s for the mismatched combination; Fig. 9A).

Sequencing of the reaction products again showed high fidelity for the complementary combination, with 97.8% of the product having the sequence expected for base pairing of the 3'-G overhang nucleotide to the 3'-C of the acceptor (Fig. 9B). By contrast, the most abundant template-switching junction for the mismatched combination was CCG (45.6%), which can be explained by template switching from a donor that has undergone NTA to add a complementary G overhang, leaving a gap (*dash*) in the template strand, which is filled by a complementary nucleotide (indicated by an *arrow* pointing to the gap) during the PCR used to add RNA-Seq adapters (Fig. 9B). Surprisingly, the second most abundant junction (27.5%) contained a relatively long insertion that appears to reflect an intermediate template switch to a low-level contaminating oligonucleotide (*italics*) with a 3'-G that could base pair to the 3'-C overhang of the starter duplex (Fig. 9B). Another aberrant junction (CCGG; 3.73%) required two NTAs before adding a complementary G overhang, and another (CC, 3.66%) can be explained by NTA of a complementary G overhang to a blunt-end version of the starter duplex that may be present at low levels due to mis-synthesis or resection of the mismatched 3' overhang (Fig. 9B). By contrast, the junction expected for extension of the mismatch (CG; with the G in the top strand incorporated opposite the mismatched C in the bottom strand after PCR) was detected at lower frequency (2.91%) (Fig. 9B). The preference for inefficient alternative reactions, which revert to a complementary 3' overhang nucleotide for template switching rather than use a mismatched acceptor RNA added in excess, indicates that GsI–IIC RT has a strong aversion to extending mismatches at the template-switching junction.

Mechanisms of template switching from a blunt-end duplex

Finally, to explore the mechanism of template switching from a blunt-end duplex, we performed similar template-switching reactions from the blunt-end version of the starter duplex to acceptor RNAs ending in U, A, C, G, and N, where N is an equimolar mixture of the four acceptor RNAs.

As shown in Fig. 10A, the rate and amplitude of template switching starting from an artificial blunt-end duplex were highest for the 3'-N acceptor, which can base pair with any 3' overhang nucleotide added by NTA, and for the 3'-U acceptor, which can base pair with a 3' overhang A, the preferred nucleotide for NTA to the same blunt-end duplex (see Fig. 5 and Fig.

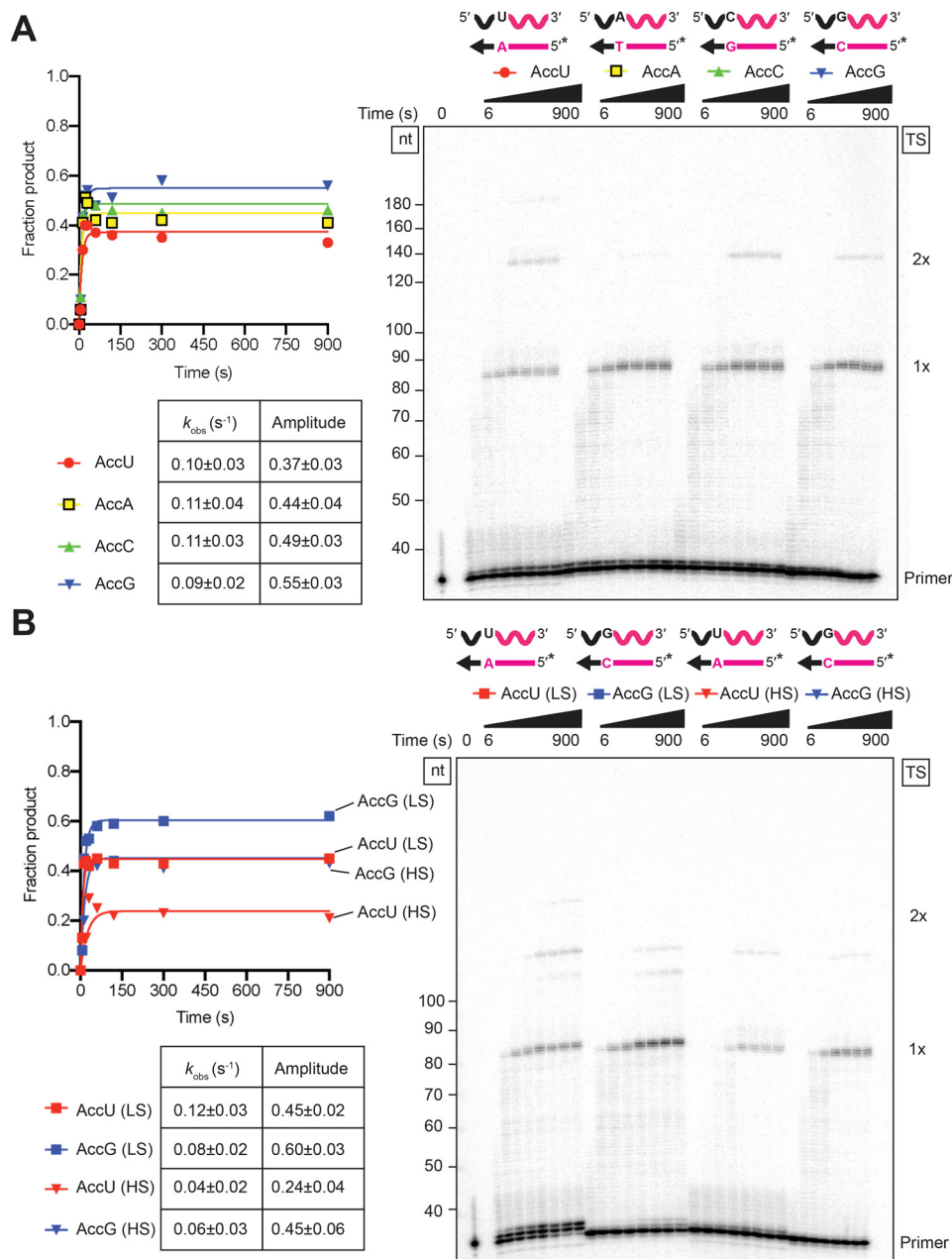


Figure 8. Biases in 3'-adapter addition in TGIRT-seq reflect the efficiency of template switching to acceptor RNAs with different 3' nucleotides at a high-salt concentration. *A*, template-switching reactions with 50-nt acceptor RNAs differing only in their 3' nucleotide and ^{32}P -labeled (*) starter duplexes having a complementary 1-nt 3' DNA overhang were done as in Fig. 7, but in reaction medium containing 450 mM instead of 200 mM NaCl. *B*, template-switching reactions were done as in *A*, but with 10 nM instead of 100 nM acceptor RNA in reaction medium containing either 200 mM NaCl (LS) or 450 mM NaCl (HS). The gels are labeled as in Fig. 1. The plots to the left of the gel show the time course data fit by a single-exponential function to calculate the k_{obs} and amplitude for each time course, and the values and standard errors of the fit are summarized in the tables below the plots.

S5 for NTA preferences). The efficiency of template switching to the remaining acceptors followed the order $C > G > A$, roughly matching the order expected from the efficiencies of NTA to the blunt-end duplex ($G > C \approx T$; Fig. 5A). The efficient use of the 3'-U acceptor for template switching from a blunt-end duplex contrasts with its least efficient use for template switching from a complementary 1-nt 3' overhang duplex (Figs. 7 and 8) and indicates that NTA of a complementary A overhang that can base pair with the 3'-U is rate-limiting for template switching from the blunt-end duplex.

Additional experiments showed that template switching to a 3'-C acceptor had a higher $K_{1/2}$ value for dNTP from a blunt-end

duplex than from a 1-nt overhang 3'-G duplex (0.92 and 0.03 mM, respectively; Figs. S9, A and B and S10A), and similar results were obtained for template switching to a 3'-C acceptor from a blunt-end or 3'-G-overhang duplex with varying concentrations of dGTP, the nucleotide complementary to the first and second base in the acceptor RNA ($K_{1/2} = 3$ and 0.03 mM, respectively; Figs. S9, C and D and S10B). This >30-fold higher $K_{1/2}$ value for dNTP addition for template switching from a blunt-end duplex is similar to the $K_{1/2}$ value for NTA (Figs. 4 and 5). Although this finding is consistent with a requirement for NTA to produce a complementary 1-nt 3' overhang prior to template-switching, it could also reflect template switching

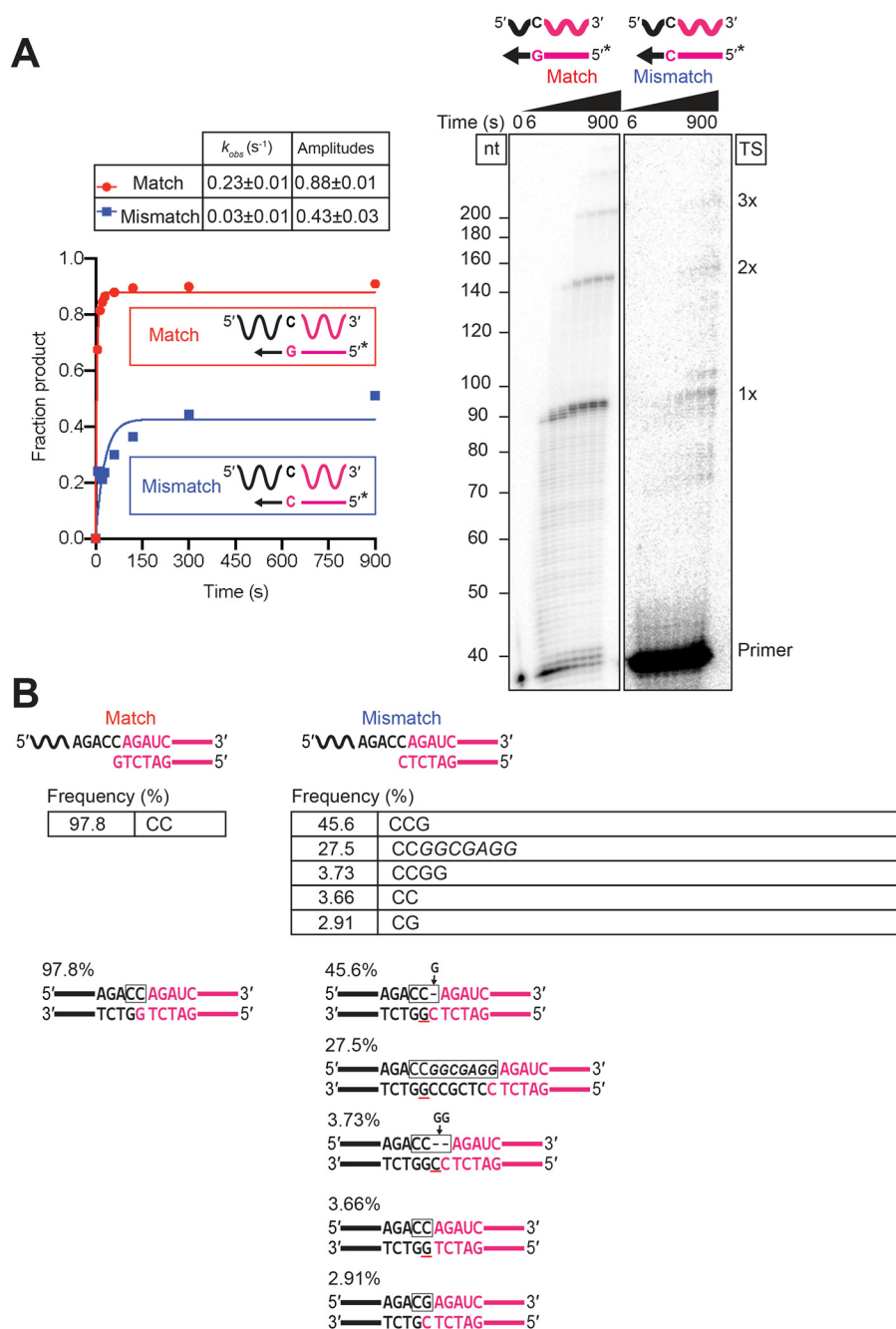


Figure 9. Template switching by Gsl-IIC RT disfavors the extension of mismatches. Template-switching reactions to a 50-nt acceptor RNA with a 3'-C residue using starter duplexes having either a complementary 1-nt 3'-G overhang (matched) or a non-complementary 1-nt 3'-C overhang (mismatched) were done as described in Fig. 7. A, time courses. Template-switching reactions with ³²P-labeled (*) starter duplexes were stopped after times ranging from 6 to 900 s, and the products were analyzed by denaturing PAGE and quantified from phosphorimager scans of the dried gel. The gels (different exposures for the matched and mismatched configurations) are labeled as in Fig. 1. The plots to the left of the gel show the data fit by a single-exponential function to calculate the k_{obs} and amplitude for each time course, and the values together with the standard error of the fit are summarized in the table above. B, RNA-Seq analysis of template-switching junctions. Template-switching reactions with unlabeled starter duplexes were done under the same conditions as in A with the reaction stopped after 15 min. RNA-Seq analysis was done as described in Fig. 1 and "Experimental procedures." The figure diagrams the template-switching reaction for each acceptor/starter duplex combination above, with the sequences and percentages of the most frequent template-switching junctions for the mismatched combination shown below. Nucleotides that were derived from the acceptor are in black, and nucleotides derived from the starter duplex are in red, with the box indicating the junction sequence. A black letter with red underline indicates a nucleotide residue inferred to result from NTA. A gap in the top strand due to NTA is shown as a dash, and nucleotides inferred to fill the gap after PCR to add RNA-Seq adapters are indicated above the line with arrows pointing to the gap. Nucleotides in italics are putatively derived from an intermediate template switch to contaminating oligonucleotides present in the enzyme preparation or reagents. Complete data for junction sequences are shown in Table S3.

directly from a blunt-end duplex, with weaker dNTP binding reflecting that the templating 3' nucleotide of the acceptor is not anchored by a phosphodiester bond to the downstream template strand as it would ordinarily be for cDNA synthesis.

To attempt to distinguish these possibilities, we analyzed the junctions resulting from template switching from the blunt-end duplex to the 3'-N acceptor by RNA-Seq (Fig. 10B). For this analysis, we separately analyzed the two types of template-

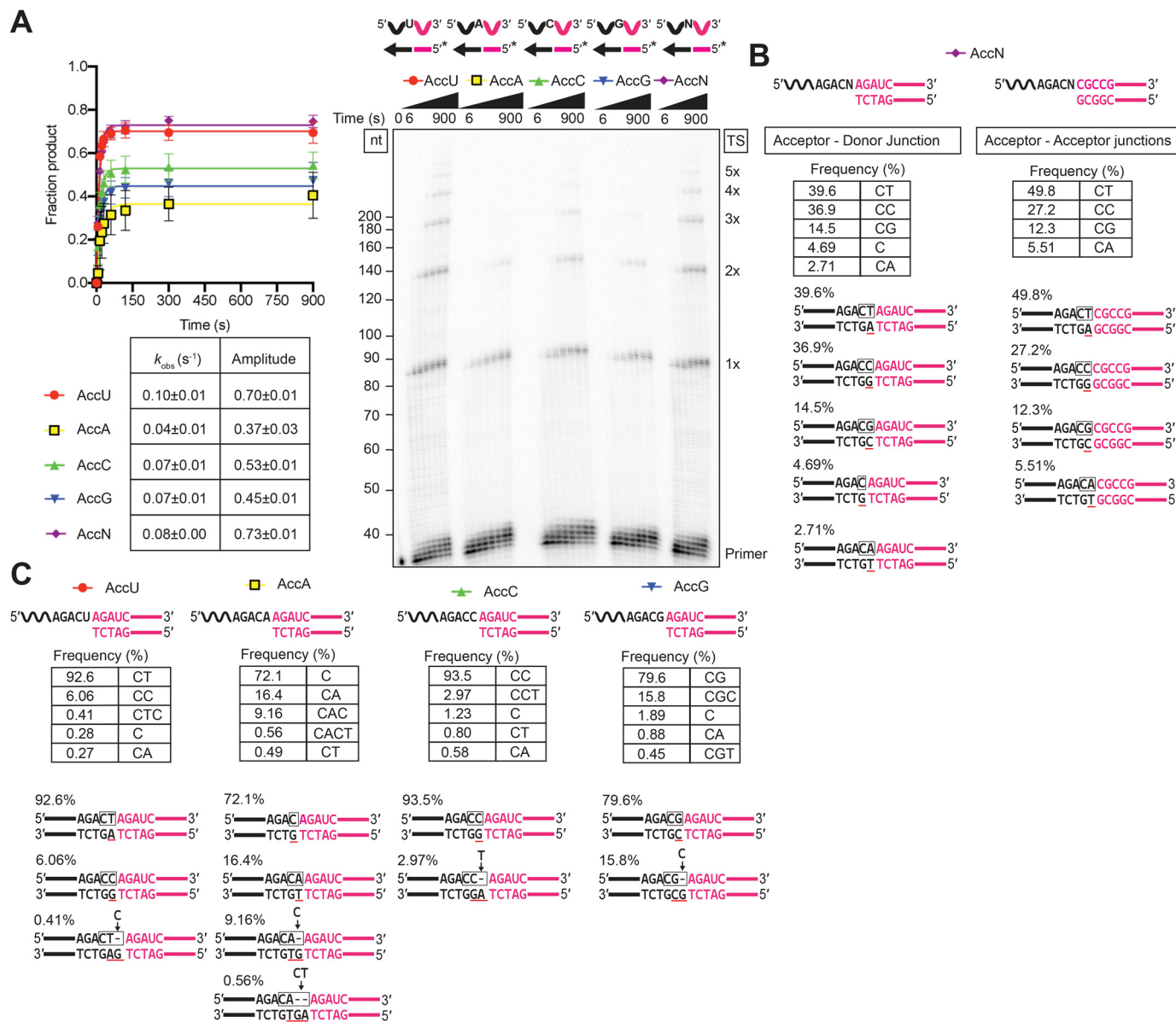


Figure 10. Template switching from a blunt-end starter duplex is inefficient and yields heterologous junction sequences. Template-switching reactions with 50-nt acceptor RNAs (100 nM) differing only in their 3'-nucleotide residue and a blunt end-starter R2 RNA/R2R DNA starter duplex (50 nM) were done as described in Fig. 7. The 3'-N acceptor is an equimolar mix of acceptor RNAs having 3' A, C, G, or U residues added at the same total concentration (100 nM) as each individual acceptor. *A*, time courses. Template-switching reactions with ^{32}P -labeled (*) starter duplexes were stopped after times ranging from 6 to 900 s, and the products were analyzed by denaturing PAGE and quantified from phosphorimager scans of the dried gel. The gel is labeled as in Fig. 1. The plot to the left of the gel shows the data for each acceptor fit by a single-exponential function, with the values and standard error of the fit summarized in the table below the plot. *B* and *C*, RNA-Seq analysis of template-switching junctions between the acceptor RNAs and unlabeled blunt-end duplex were done for 15 min under the same conditions as in *A*. The figure diagrams the template-switching reaction for each combination, with the sequences and frequencies of the most frequent template-switching junctions shown below. *B* shows sequences and frequencies for both the initial Acceptor-Donor junctions and subsequent Acceptor-Acceptor junctions, and *C* shows only those for the initial Acceptor-Donor junctions. Nucleotides that were derived from the acceptor RNA are in black, and nucleotides derived from the synthetic blunt-end duplex or blunt-end duplexes formed after completion of cDNA synthesis are in red, with the box indicating the junction sequence. A black letter with red underline indicates a nucleotide inferred to result from NTA to a blunt-end duplex. Gaps in the top strand due to NTA are shown as dashes, and nucleotides inferred to fill those gaps after PCR to add RNA-Seq adapters are indicated above the line with arrows pointing to the gap. Complete data for junction sequences are shown in Table S3.

switching junctions formed in the reaction: those resulting from template switching from the starter duplex to the 3' end of the acceptor RNA (acceptor-donor junctions) and those resulting from secondary template switches from the 5' end of an acceptor RNA after completion of cDNA synthesis to the 3' end of another acceptor RNA (acceptor-acceptor junctions). The expectation was that template switching directly from a blunt-end duplex would result in roughly equal proportions of

junctions for 3' A, C, G, and U acceptors, whereas template switching via NTA would parallel the efficiency of NTA of the complementary 3' overhang nucleotide to the blunt-end duplex. For both acceptor-donor and acceptor-acceptor junctions, the proportion of junctions for different acceptor 3' nucleotides was $U > C \gg G > A$ (Fig. 10B), roughly paralleling that for NTA of the complementary nucleotide ($A > G \gg C \approx T$) to the same blunt-end duplex (Fig. 5A). The proportion of

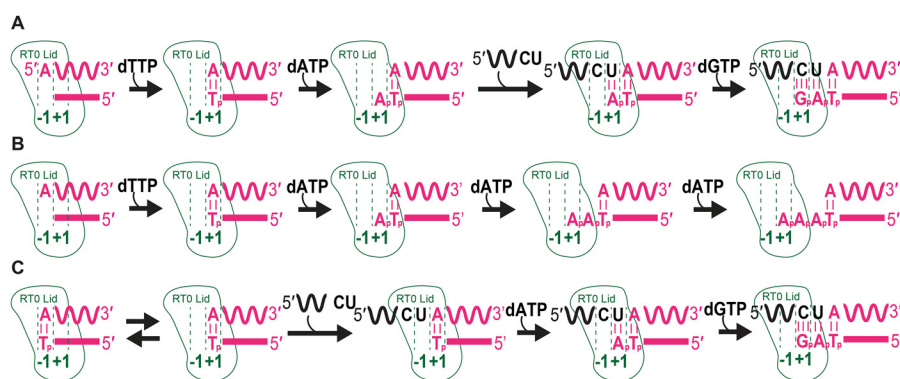


Figure 11. Models of template switching and non-templated nucleotide addition reactions. *A*, template switching to an acceptor RNA from an RNA template/DNA primer heteroduplex with a 1-nt 3' overhang added by NTA after completion of cDNA synthesis or as an artificial starter duplex. *B*, NTA to a blunt-end RNA/DNA heteroduplex in the absence of an acceptor nucleic acid. *C*, template-switching to an acceptor RNA from a blunt-end RNA/DNA heteroduplex without NTA. See under "Discussion" for details.

CT junctions, requiring NTA of a 3' overhang A that can base pair to the 3'-U acceptor, was somewhat higher for the secondary template switches from the 5' end of the acceptor than those from the artificial starter duplex. Additionally, template switching from the blunt-end starter duplex yielded a significant proportion of single C junctions (4.69%), possibly via NTA of a G residue to a population of acceptors ending with a single 3'-C instead of a 3'-CN (also seen in the experiments below). Together, these findings suggest that template switching from a blunt-end duplex occurs largely via NTA to produce a 3' overhang nucleotide that can base pair with the 3' nucleotide of the acceptor.

Additional support for this conclusion came from template-switching reactions from a blunt-end duplex to individual acceptor RNAs with A, C, G or U 3' nucleotides (Fig. 10C). In this case, the expectation for the NTA-mediated mechanism was that the highest proportion of seamless junctions would be found for the 3'-U and -C acceptors, which require NTA of a complementary A or G overhang, the two residues most favored for NTA to the blunt-end duplex. By contrast, the lowest proportion would be expected for the 3'-G and -A acceptors, which require NTA of a complementary C or U overhang, the two residues least favored for NTA to the blunt-end duplex. In agreement with these expectations, the 3'-U and -C acceptors yielding seamless CT or CG junctions at frequencies of 92.6 and 93.5%, respectively, and the 3'-G and 3'-A acceptors yielding seamless CG and CA junctions at frequencies of 79.6 and 16.4%, respectively (Fig. 9C). The CC junction seen at 6.06% for the 3'-U acceptor could also reflect NTA of a G overhang, enabling a G:U wobble bp with the 3'-U (52), with the subsequent PCR inserting a C opposite the G added by NTA.

All four of the acceptors yielded some junctions in which one or two NTAs occurred prior to the template switch. Additionally, all four acceptors yielded some junctions with a single C residue, suggesting either that acceptor RNAs lacking the 3'-terminal nucleotide are present at low concentrations in all four acceptor RNA preparations or that some mechanism exists for resection of the 3'-terminal residue of the acceptor. In the case of the 3'-A acceptor, the single C junction was the most abundant (72.1%), possibly reflecting that NTA of a G overhang complementary to the 3'-C of this rare acceptor is much more

efficient than NTA of a 3'-U overhang complementary to the 3'-A of the predominant acceptor.

Considered together, our results indicate that template switching from a blunt-end duplex occurs largely via NTA to add a complementary 3' overhang nucleotide to the starter duplex. However, they also leave open the possibility that some template switching can occur directly from a blunt-end duplex without NTA, particularly in those cases for which NTA of a complementary 3' overhang nucleotide to the blunt-end starter duplex is inefficient (see "Discussion").

Discussion

Here, we analyzed the end-to-end template switching and NTA activities of the thermostable GsI-IIC group II intron RT. Our results indicate that template switching by this enzyme is favored by a single bp between the 3' nucleotide of the acceptor nucleic acid and a 1-nt 3'-DNA overhang of an RNA/DNA heteroduplex, with the 1-nt overhang provided either as part of an artificial starter duplex or by NTA to the 3' end of a cDNA after the completion of cDNA synthesis. Mechanistically, there are both shared features and differences with the end-to-end template-switching activities of other non-LTR retroelement RTs, as discussed further below. The finding that a single bp is sufficient to faithfully direct template switching at 60 °C, the operational temperature for GsI-IIC RT, indicates a very-high degree of specificity, and presumably affinity, for this bp within the template-switching pocket, and it has implications for both the biological function and biotechnological applications of template switching (see below).

Template switching and non-templated nucleotide addition are related reactions

Fig. 11A shows a model of the template-switching reaction after cDNA synthesis from a template RNA with a 5'-A residue. After addition of a deoxythymidine to complete cDNA synthesis, the enzyme releases pyrophosphate and undergoes translocation (53), moving the terminal rA/dT bp out of the RT active site (*i.e.* from position -1 to +1). This translocation leaves the GsI-IIC RT active site (position -1) free to bind and add a nucleotide by NTA. Any of the four deoxynucleotides can be added by NTA, but our data reveal a preference for dATP.

Upon incorporation of a nucleotide and pyrophosphate release, translocation of the added A residue from -1 to $+1$ creates a binding pocket that favors base pairing of an acceptor RNA with a complementary $3'$ -U residue at $+1$, resulting in positioning of the penultimate C residue of the acceptor as the templating RNA base at the RT active site (position -1) for continued cDNA synthesis.

A similar mechanism likely applies for artificial starter duplexes with a 1-nt $3'$ overhang. In this case, it is unclear whether the $3'$ overhang nucleotide is initially bound at position -1 (the RT active site) or position $+1$ (the position occupied by a 1-nt $3'$ overhang after NTA; see Fig. 11). As above, a 1-nt $3'$ overhang that is initially in the $+1$ position would create a binding pocket for an acceptor RNA with a complementary $3'$ nucleotide, leaving the penultimate $3'$ residue of the acceptor in the -1 position to serve as the templating base for cDNA synthesis. Alternatively, a 1-nt $3'$ overhang that is initially bound in the -1 position would have to translocate to the $+1$ position before polymerization. Mechanistically, the two positions of the $3'$ overhang may be in an equilibrium that is shifted by base pairing of the acceptor and the continuation of cDNA synthesis, a feature that has been described for Y-family DNA polymerases (discussed further below) (49, 54, 55). A comparison of artificial starter duplexes with 1-nt $3'$ overhangs that can or cannot base pair to the $3'$ nucleotide of the acceptor indicates that binding of a mismatched acceptor at the $+1$ position and/or extension from a mismatch at the -1 position are strongly disfavored (Fig. 9).

In the absence of an acceptor template with complementarity to the 1-nt $3'$ overhang of the primer, the RT can add additional nucleotides by NTA (Fig. 11B). The second and third NTA reactions are progressively slower than the first (Figs. 4 and 5), increasing the time window for sampling potential acceptor templates for complementarity. Longer base-pairing interactions assist template switching by retroviral and LTR-retrotransposon RTs (18), but the opposite is true for GsI–IIC RT, with overhang lengths beyond 1-nt decreasing the efficiency of template switching (Fig. 6 and Fig. S8). This difference may reflect that template-switching reactions involving extended base pairing by retroviral RTs require partial or complete dissociation of the enzyme from the initial template, enabling the completed cDNA to base pair to the new template for reinitiation of cDNA synthesis by the same or another RT. By contrast, group II intron RTs remain tightly bound to the donor RNA template/DNA primer duplex, leaving only a narrow binding pocket that may sterically hinder access of the $3'$ end of the acceptor to multiple overhang nucleotide bases (41).

Non-templated nucleotide addition in the absence of an acceptor nucleic acid

In the absence of an acceptor with a complementary $3'$ nucleotide, NTA continues for up to three (and occasionally four) nucleotides, progressively translocating the completed cDNA/RNA template duplex away from the active site (Fig. 11B). Because group II intron RTs lack RNase H activity and bind RNA/DNA heteroduplexes tightly, such translocation via NTA may facilitate dissociation of the enzyme from the completed cDNA

product when an acceptor nucleic acid for template switching is unavailable (34).

Our biochemical analysis shows that in the absence of an acceptor nucleic acid, NTA by GsI–IIC RT favors purine residues, particularly an A residue, at all steps (Fig. 5). A similar nucleotide preference, termed the “A-rule,” has been found for many other polymerases, albeit with some notable exceptions (16, 43–45, 56, 57), and is thought to reflect more stable base-stacking of purines on the terminal bp of the completed duplex (58). The higher $K_{1/2}$ value of dNTP addition for NTA compared with reverse transcription suggests a low affinity interaction when compared with the extension reaction (compare Fig. 4 and Fig. S10, A and B, right panels), presumably due to the absence of base pairing with a templating base. The pausing step we elucidated by modeling of processive NTA addition could indicate the following: (i) the presence of a pre-insertion step, during which the blunt-end duplex occupies the -1 nucleotide-binding site, effectively inhibiting nucleotides from binding; (ii) backtracking of the RT, a process documented for retroviral RTs (59, 60); or (iii) enzyme dissociation from the template–primer heteroduplex. Structural and kinetic analysis of Y-family DNA polymerases showed that an analogous pausing step reflects a dynamic equilibrium between the $3'$ nucleotide of primer occupying the -1 position in the pre-insertion state and the nucleotide-binding competent $+1$ position (49, 54, 55). It is possible that NTA preferences for different nucleotides may be influenced by the sequence of the starter duplex, as a similar trend but with a stronger preference for A over G residues was found for NTA to the $3'$ end of completed cDNAs in TGIRT-seq of miRNA reference sets containing 962 human miRNAs (A, 86.6%; G, 9.9%; C, 2.0%; and U, 1.4% for the first NTA, with 83.3% of the NTAs being a single nucleotide (SRA accession number NTT1, SRX5005854)).

Template switching from a blunt-end duplex

Our analysis of junction sequences suggests that template switching of GsI–IIC RT from a blunt-end duplex occurs largely via NTA of a $3'$ overhang nucleotide that can base pair to the $3'$ nucleotide of the acceptor (Fig. 10). However, the finding that the rate of template switching from a blunt-end duplex to both RNA and DNA acceptor is faster than the rate of NTA to same duplex (Fig. 6 and Fig. S8) could reflect that some template switching can also occur directly from a blunt end duplex without NTA.

As shown in Fig. 11C, template switching directly from a blunt duplex would require that the terminal bp of the duplex be located at the $+1$ position, thereby allowing binding of the terminal nucleotide of the acceptor at the RT active site (position -1). The latter then forms a binding pocket for a complementary dNTP, which can be polymerized onto the DNA primer in a normal manner (Fig. 11C). Unlike the situation after the completion of cDNA synthesis (Fig. 11A), the binding or movement of the terminal bp of an artificial blunt-end duplex to the $+1$ position is not linked to NTA and could again reflect a natural equilibrium between the -1 and $+1$ position, which is pushed forward by binding of the acceptor and the continuation of cDNA synthesis.

We note that the junction sequences formed by direct template switching from a blunt-end duplex would be indistinguishable from those formed via NTA of a nucleotide complementary to the 3' end of the acceptor, the only difference being whether or not the 3' nucleotide of the acceptor served as the templating base. The finding that the $K_{1/2}$ value for dNTP for template switching from a blunt-end duplex is >30-fold higher than that for template switching from a 1-nt 3' overhang duplex (Fig. S10) could reflect either a requirement for NTA or weaker dNTP binding opposite the 3' nucleotide of the acceptor RNA that is not fixed in position by a bp to a 1-nt overhang or by a phosphodiester bond to the downstream RNA template strand as it would ordinarily be during cDNA synthesis.

Comparison of GsI-IIC RT to other non-LTR retroelement RTs

Non-LTR retroelement RTs typically have an NTE with a conserved RT0 motif that likely contributes to a structurally-similar binding pocket for the 3' end of the acceptor nucleic acid during template switching (5, 19, 41, 42). Thus, it is unsurprising that many features of the template switching and NTA mechanisms of Gs-IIC RT are similar to those found previously for the MRP and R2 element RTs, including a prominent role for NTA in adding nucleotides to the 3' end of the cDNA that can base pair to the 3' end of acceptor RNAs (25, 33, 34). For the R2 RT, biochemical experiments examining NTA after the completion of cDNA synthesis found a similar nucleotide preference $A > G \gg C \approx T$, with NTA limited to 1–4 nt and with longer 3' overhangs disfavoring template switching (33, 34), all similar to our findings for GsI-IIC RT.

Despite these similarities, there are significant differences in template switching by the three RTs that are relevant to biotechnological applications and suggest how template-switching activity may be adapted to fit the replication cycle of different non-LTR retroelements. One important difference is that the R2 RT appears to require a “running start” to template-switch and is unable to rebind to a completed cDNA with a 3' overhang added by NTA (33, 34). In contrast, GsI-IIC RT can bind to and template-switch efficiently from synthetic RNA template/DNA primer duplexes with a 1-nt 3' overhang, an ability that enables precise RNA-Seq adapter addition in TGIRT-seq (6). Furthermore, unlike GsI-IIC RT, which requires an RNA/DNA heteroduplex with a 3' overhang for efficient template switching, the R2 element RT can initiate reverse transcription near the 3' end of an acceptor RNA by using an RNA primer (33), and the MRP RT can do so using an ssDNA primer, albeit under somewhat different reaction conditions (25). Finally, unlike GsI-IIC RT template switching, the specificity of which is dictated by a single bp to the 3' nucleotide of the acceptor, the MRP RT has a strong preference for template switching to RNAs with a 3'-CCA and to initiate cDNA synthesis opposite the penultimate C residue (C-2), even in the absence of complementarity to the DNA primer (25). These features likely reflect adaptation of the MRP RT template-switching activity to initiate at the 3' tRNA-like structure of the plasmid transcript for plasmid replication.

Possible biological functions of non-LTR retroelement RT template switching

The high efficiency of end-to-end template switching by non-LTR retroelement RTs and its dependence upon a conserved binding pocket that is not present in retroviral RTs suggest that this activity has an important biological function. In the case of insect R2 and mammalian LINE-1 elements, template switching has been proposed to function in the attachment of the 3' end of a cDNA that was initiated by TPRT to upstream host DNA sequences (35–37). Although this remains an attractive hypothesis, other modes of DNA attachment have not been excluded, and experiments investigating the retrohoming of a linear group II intron RNA in *Drosophila melanogaster*, an analogous situation requiring attachment of a free 3' cDNA end to upstream host sequences, indicated involvement of nonhomologous end-joining enzymes rather than template switching (39). Additionally, recent findings indicate that chimeric integration products of human LINE-1 elements can result from RNA ligation rather than template switching (61).

A different and not mutually exclusive hypothesis is suggested by the common requirement of all non-LTR retroelement RTs to synthesize full-length cDNAs of long RNA templates that are vulnerable to oxidative damage and nicking by host endonucleases (e.g. RNase E in the case of bacterial group II intron RTs (62)). Template switching may provide a means of reverse transcribing through nicked or gapped RNA templates *in vivo* enabling the synthesis of continuous cDNAs from discontinuous RNA templates, a role previously suggested for the clamping activity of retroviral and LTR retroelement RTs (17). The finding that the GsI-IIC RT template switching and NTA activities are tuned to favor formation and extension of a single bp could reflect that group II intron RTs are under stronger selective pressure than retroviral RTs to faithfully replicate their RNA templates, which in the case of group II RTs are highly-structured, catalytically active intron RNAs required for retrohoming. We note that template switching may play a similar role for RNA viruses, whose RdRPs are closely related to group II intron RTs, including a structurally homologous acceptor RNA-binding pocket, termed motif G in viral RdRPs (41, 63). This central function may be combined with additional functions specific to the replication cycle of different non-LTR retroelements and viruses, such as preferential template switching to a 3'-tRNA-like structure in the cases of MRP RTs (25) or higher rates of internal template switches by RdRPs to generate recombinant RNA viruses to evade host defenses or repair defective genomes (64).

Implications for RNA-Seq

Finally, our results have significant implications for the use of group II intron RT template switching in RNA-Seq. First, we found that the configuration of starter RNA template/DNA primer duplex used in TGIRT-seq in which a 1-nt 3' DNA overhang base pairs to the 3' nucleotide of target RNAs (Fig. 1A) is in fact the most accurate and efficient configuration for GsI-IIC RT template switching. The thermostable TeI4c group II intron RT has also been shown to template-switch efficiently from an artificial starter duplex with a 1-nt 3' overhang, and this

is likely to be a common characteristic of group II intron RTs (6). Second, we found that the previously-determined 3'-sequence biases in TGIRT-seq reflect the efficiency of template switching to acceptor RNAs with different 3' nucleotides, particularly under the high-salt conditions used for TGIRT-seq (Fig. 8). These 3' biases, which are restricted to the 3'-terminal nucleotide of the acceptor, account for about half of the sequence bias in TGIRT-seq, the remainder coming from the Thermostable 5' App RNA/DNA Ligase (New England Biolabs) used for R1R adapter ligation (see Fig. 1A) (51). Although we showed previously that these 3' biases could be remediated computationally or by changing the ratio of 3' DNA overhangs in the starter duplex (51), we found here that they can be more simply moderated by carrying out the template-switching reaction at lower salt concentrations (200 mM NaCl; Fig. 8). It should also be noted that the 3'-end biases for template switching by GsI-IIC RT (for acceptor RNAs with a 3'-G and against those with a 3'-U) are not universal, as TeI4c RT preferentially template switches to acceptor RNAs with a 3'-A residue,³ implying that these template-switching biases can be modified and perhaps eliminated by protein engineering of the acceptor-binding pocket. Finally, in addition to reducing template-switching biases, we found that lower salt concentrations substantially increase the efficiency of TGIRT template switching (2–4-fold increase in rate; 1.7-fold increase in amplitude; Fig. 2), perhaps by stabilizing a productive enzyme–substrate conformation. The higher efficiency of template switching at lower salt concentrations enables more efficient capture of RNA templates (Fig. S3), as well as more efficient use of RNAs with modified 3' ends, including 3'-phosphate and 2'-O-Me groups (Fig. 3), which are present at the 3' ends of PIWI-interacting RNAs and plant miRNAs (65, 66). Although these benefits for different applications must be weighed against the effect of increased multiple template switching at lower salt concentrations, recent TGIRT-seq of human plasma RNA at 200 mM instead of 450 mM NaCl substantially increased the efficiency of TGIRT-seq library construction, without unacceptably increasing the proportion of multiple template switches (0.5–4% fusion reads, which include multiple-template switches).⁴

Experimental procedures

Protein purification

GsI-IIC RT with an N-terminal maltose-binding protein tag to keep the protein soluble in the absence of bound nucleic acids was expressed from plasmid pMRF-GsI-IIC and purified with minor modifications of a procedure described previously (6). A freshly transformed colony of Rosetta 2 (DE3) cells (EMD Millipore) was inoculated into 1,000 ml of LB medium containing ampicillin (50 µg/ml) and chloramphenicol (25 µg/ml) in a 4,000-ml Erlenmeyer flask and grown overnight with shaking at 37 °C. 50 ml of the starter culture was then added to 1,000 ml of LB medium containing ampicillin (50 µg/ml) in one to six 4,000-ml Erlenmeyer flasks and grown at 37 °C to an OD₆₀₀ = 0.6–0.7, at which time protein expression was induced by add-

ing 1 mM isopropyl 1-thio-β-D-galactopyranoside and incubating overnight with shaking at 19 °C. The cells were pelleted by centrifugation and stored at –80 °C overnight. After thawing on ice, the cells were lysed by sonication in 500 mM NaCl, 20 mM Tris-HCl, pH 7.5, 20% glycerol, 1 mg/ml lysozyme, 0.2 mM phenylmethylsulfonyl fluoride (Roche Applied Science), and the lysate was clarified by centrifugation at 30,000 × g for 60 min at 4 °C in a JA 25.50 rotor (Beckman). Nucleic acids in the clarified lysate were precipitated by slowly adding polyethyleneimine to the lysate with constant stirring in an ice bath to a final concentration of 0.4%, and then centrifuging at 30,000 × g for 25 min at 4 °C in a JA 25.50 rotor (Beckman). GsI-IIC RT and other cellular proteins were then precipitated from the supernatant with 60% saturating ammonium sulfate, pelleted at 30,000 × g for 25 min at 4 °C in a JA 25.50 rotor (Beckman), and resuspended in 25 ml of A1 buffer (300 mM NaCl, 25 mM Tris-HCl, pH 7.5, 10% glycerol). The protein mixture was then purified through two tandem 5-ml MBPTrap HP columns (GE Healthcare). After loading, the tandem column was washed with 5 column volumes (CVs) of A1 buffer, and the maltose-binding protein–tagged GsI-IIC RT was eluted with 10 CVs of 500 mM NaCl, 25 mM Tris-HCl, pH 7.5, 10% glycerol containing 10 mM maltose. The final fractions containing the RT were diluted to 200 mM NaCl, 20 mM Tris-HCl, pH 7.5, 10% glycerol and loaded onto a 5-ml HiTrap Heparin HP column (GE Healthcare) and eluted with a 12-CV gradient from buffer A1 to A2 taking 0.5-ml fractions. Fractions containing GsI-IIC RT were identified by SDS-PAGE, pooled, and concentrated using a 30K centrifugation filter (Amicon). The concentrated protein was then dialyzed into 500 mM NaCl, 20 mM Tris-HCl, pH 7.5, 50% glycerol, and 10-µl aliquots were flash-frozen using liquid nitrogen and stored at –80 °C. Thawed aliquots were only used once.

DNA and RNA oligonucleotides

The DNA and RNA oligonucleotides used in this work are listed in Table S1. Most were purchased in RNase-free HPLC-purified form from Integrated DNA Technologies. Exceptions were the R2R primers and the cDNA mimic used in Fig. S3, which were 5'-end-labeled and purified by electrophoresis in a denaturing 8% polyacrylamide gel. For biochemical assays, the R2R and primer extension (PE) primer (DNA primers) (100 pmol) were labeled with [γ -³²P]ATP (125 pmol; 6,000 Ci/mmol; 150 µCi/µl; PerkinElmer Life Sciences) by incubating DNA with T4 polynucleotide kinase (10 units; New England Biolabs) for 30 min at 37 °C. A typical 10-µl reaction was then diluted to 40 µl with double-distilled H₂O and extracted with an equal volume of phenol/chloroform/isomyl alcohol (25:24:1). Unincorporated nucleotides were removed from the aqueous phase by using a P-30 Microspin Column RNase Free (Bio-Rad), and the oligonucleotide concentration was measured by using the Qubit ssDNA assay kit (Thermo Fisher Scientific). RNA acceptors used in Fig. S3 were labeled by the same method, and their concentrations were measured using a Qubit RNA HS Assay kit (Thermo Fisher Scientific). The concentration of unlabeled oligonucleotides was determined by spectrophotometry using a NanoDrop 1000 (Thermo Fisher Scientific).

³ Y. Qin and A. M. Lambowitz, unpublished data.

⁴ D. C. Wu and A. M. Lambowitz, unpublished data.

Template-switching and NTA assays

Template-switching reactions with GsI–IIC RT were carried out by using the previously described protocol with minor modifications (6). The RNA template/DNA primer starter duplex was based on that used for 3'-RNA-Seq adapter addition in TGIRT-seq (8) and consists of a 34-nt RNA oligonucleotide containing an Illumina R2 sequence (R2 RNA) with a 3-blocking group (3SpC3, Integrated DNA Technologies; Table S1) annealed to a complementary DNA primer (R2R) that leaves either a blunt end, a single nucleotide 3'-DNA overhang end, or in some experiments longer 3'-DNA overhang ends (Table S1). The oligonucleotides were annealed at a ratio of 1:1.2 to yield a final duplex concentration of 250 nM by heating to 82 °C for 2 min and then slowly cooling to room temperature. Unless specified otherwise, reactions were done with 500 nM GsI–IIC RT, 50 nM R2 RNA/R2R DNA starter duplex, and 100 nM acceptor oligonucleotide in 25 μ l of medium containing 200 mM NaCl, 5 mM MgCl₂, 20 mM Tris-HCl, pH 7.5, 5 mM fresh DTT, and an equimolar mix of 1 mM each of dATP, dCTP, dGTP, and dTTP (Promega) to give 4 mM total dNTP concentration. Reactions were set up with all components except dNTPs and preincubated for 30 min at room temperature and then initiated by adding dNTPs. Reactions were incubated at 60 °C for times indicated in the figure legends and stopped by adding 2.5- μ l portions to 7.5 μ l of 0.25 M EDTA. The products were further processed by adding 0.5 μ l of 5 N NaOH and heating to 95 °C for 3 min to degrade RNA and remove tightly-bound GsI–IIC RT, followed by a cooling to room temperature and neutralization with 0.5 μ l of 5 N HCl. After adding formamide loading dye (5 μ l; 95% formamide, 0.025% xylene cyanol, 0.025% bromophenol blue, 10 mM Tris-HCl, pH 7.5, 6.25 mM EDTA), the products were denatured by heating to 99 °C for 10 min and placed on ice prior to electrophoresis in a denaturing 6% polyacrylamide gel containing 7 M urea, 89 mM Tris borate, and 2 mM EDTA, pH 8 at 65 watts for 1.5–2 h. A 5'-end-labeled, ssDNA ladder (10–200 nts: ss20 DNA Ladder, Simplex Sciences) was run in a parallel lane in most experiments. After we used the molecular weight markers to establish the identity of the major product bands (Figs. 1, 3, and 7–10 and Fig. S1), markers were omitted from some experiments (Figs. 2 and 6 and Figs. S8 and S9, A and B). The gels were dried, exposed to an Imaging screen-K (Bio-Rad), and scanned using a Typhoon FLA 9500 (GE Healthcare). The Typhoon laser scanner was used for all other experiments where “phosphor-imager” is indicated. NTA reactions were done as described for template-switching reactions but in the absence of acceptor oligonucleotide.

Analysis of template-switching reactions using nondenaturing PAGE

The 50-nt acceptor RNA (100 pmol) was 5'-labeled with [γ -³²P]ATP (125 pmol; 6,000 Ci/mmol; 150 μ Ci/ μ l; PerkinElmer Life Sciences) by incubating with T4 polynucleotide kinase (10 units; New England Biolabs) for 30 min at 37 °C.

Template-switching reactions were done as described above using 5'-³²P-labeled 50-nt acceptor RNA diluted with unlabeled 50-nt acceptor to reach the desired final concentration

and 50 nM starter duplex. Reactions were incubated at 60 °C for 15 min and stopped by adding 2.5- μ l portions to 7.5 μ l of 0.25 M EDTA and 1 μ l of proteinase K (0.8 units, New England Biolabs) and incubating at room temperature for 30 min. After addition of 5 μ l of 3 \times loading buffer (30% glycerol, 0.025% xylene cyanol, 0.025% bromophenol blue, 30 mM Tris, 150 mM HEPES, pH 7.5), 5 μ l of the sample was loaded onto a 8% polyacrylamide gel containing 34 mM Tris, 66 mM HEPES, pH 7.5, 0.1 mM EDTA, and 3 mM MgCl₂. The gel was then run at 4 °C for 4 h at a constant power of 3 watts and then dried and imaged as described above for denaturing polyacrylamide gels. The cDNA mimic oligonucleotide used in the native PAGE experiment (Fig. S3) consists of the R2R-G sequence fused to the reverse complement of the 50-nt acceptor lacking the terminal C residue. The cDNA mimic was annealed to R2 RNA and 50-nt acceptor C oligonucleotides to provide a marker for template-switching product bands in the gel.

Primer extension assays

Primer extension reactions with GsI–IIC RT were carried out similarly to template-switching reactions by using a 50-nt, 5'-end-labeled DNA primer (PE primer) annealed near the 3' end of a 1.1-kb *in vitro*-transcribed RNA. The transcript was generated by T3 runoff transcription (T3 MEGAscript kit, Thermo Fisher Scientific) of pBluescript KS(+) (Agilent) linearized using XmnI (New England Biolabs) and cleaned up using a MEGAclean kit (Thermo Fisher Scientific). The labeled DNA primer was annealed to the RNA template at a ratio of 1:1.2 to a yield a final duplex concentration of 250 nM by heating to 82 °C for 2 min followed by slowly cooling to room temperature. GsI–IIC RT (500 nM) was pre-incubated with 50 nM of the annealed template–primer in 25 μ l of reaction medium containing 200 mM NaCl, 5 mM MgCl₂, 20 mM Tris-HCl, pH 7.5, 5 mM DTT for 30 min at room temperature, and reverse transcription was initiated by adding 1 μ l of the 25 mM dNTP mix to give a final dNTP concentration of 1 mM for each dNTP. After incubating at 60 °C for times indicated in the figure legend, the reaction was terminated, processed, and analyzed by electrophoresis in a denaturing 8% polyacrylamide gel, as described above for template-switching reactions.

RNA sequencing

RNA-Seq libraries were prepared as described previously (8, 9) by using TGIRT-III (InGex), a commercial version of GsI–IIC RT, with different R2 RNA/R2R starter duplexes and acceptor nucleic acids as indicated in the text. The initial template-switching reactions for addition of the R2R adapter to the 5' end of the cDNA were done as described above with 500 nM TGIRT-III enzyme, 50 nM unlabeled starter duplex, and 100 nM acceptor RNA for 15 min at 60 °C. After terminating the reactions with NaOH and neutralization with HCl as described above for template-switching reactions, the volume was raised to 100 μ l with H₂O, and cDNA products containing the R2R adapter attached to their 5' end were cleaned-up by using a MinElute column (Qiagen) to remove unused primer. A 5'-adenylated R1R adapter was then ligated to the 3' end of the cDNA using Thermostable 5' App DNA/RNA Ligase (New England Biolabs) for 1 h at 65 °C. After another MinElute clean

up, the entirety of the eluent was used for a 12-cycle PCR using Phusion polymerase (Thermo Fisher Scientific), and the resulting libraries were cleaned up by using 1.4× Ampure XP beads to remove residual primers, primer dimers, salts, and enzymes. The quality of the libraries was assessed by using a 2100 Bioanalyzer Instrument with a High Sensitivity DNA chip (Agilent). The libraries were sequenced on an Illumina NextSeq instrument to obtain 1–2 million 75-nt paired-end reads. Read 1 was used for analysis.

To analyze template-switching junctions, the 75-nt reads 1 from each dataset were trimmed from the 5' end using Cutadapt version 2.5 (67) to remove all but the last three nucleotides (GAC) of the adapter proximal acceptor by using the following parameters: cutadapt -O 10 -nextseq-trim = 20 -trim-n -q 20 -discard-untrimmed -g CGCCGGACCGTG-CACCATCTGGAGTTATAGAGATGAGTCTCACATA -j 8 -e 0.1 -o {Step1 trimmed reads} {Read 1 file}. The reads were then trimmed from the 3' end to leave the junction sequences flanked by GAC at the 5' end and CGC (acceptor) or AGA (donor) at the 3' end by using the following parameters: cutadapt -O 10 -nextseq-trim = 20 -trim-n -q 20 -discard-untrimmed -a CGGACCGTGCACCAT -a TCGGAAGAGC-ACACG -j 8 -e 0.1 -o {Step2 trimmed reads} {Step1 trimmed reads}. Finally, the trimmed reads containing either the acceptor-donor (5'-GAC(N)_nAGA-3') or acceptor-acceptor (5'-GAC(N)_nCGC-3') junction were binned and counted by using lab scripts to determine the frequencies of different junctions.

Analysis of kinetics experiments

Phosphorimager scans of reaction time courses were quantified with ImageQuant TL 8.1 (GE Healthcare) by generating rectangular boxes around the unextended primer and each reaction product. Background was subtracted by using analogous rectangles on a portion of the screen that did not correspond to a gel lane. Fractions of product were plotted *versus* time and fit by a single-exponential function. For analyses of concentration dependence, rate constants were plotted against the concentration of the species being varied (enzyme or dNTP) and fit by a hyperbolic equation to obtain values of the maximal rate constant and half-maximal concentration of the varied species. Data fitting was done using Prism 8 (GraphPad). Unless otherwise indicated, the reported uncertainty values reflect the standard error obtained from these fits. For reactions that were performed in triplicate (*e.g.* Fig. 1B), fitting each data set individually gave rate constants with standard error values after averaging of <25% and amplitudes with standard error values after averaging of <5% (data not shown), reflecting the level of day-to-day variability in these reactions.

Reactions monitoring NTA of individual nucleotides were analyzed by global modeling using KinTek Explorer (47). Data were fit by a model that included up to three equilibrium binding steps, three irreversible extensions, and paused and stopped states at each cycle. To obtain an initial fit, we set the reaction rates and equilibrium binding constants to those derived from conventional fitting of the data to analytic functions. The fit was further refined by adjusting the reaction rates to approximate the data and then using the data fit editor to fit globally to data from reactions at four different dNTP concentrations. This pro-

cess was repeated iteratively to obtain the final fits. Plots of normalized χ^2 values *versus* the rate constant parameter (1D Fitspace) are shown in Fig. S6.

Author contributions—A. M. Lentzsch, J. Y., R. R., and A. M. Lambowitz conceptualization; A. M. Lentzsch and J. Y. data curation; A. M. Lentzsch, J. Y., R. R., and A. M. Lambowitz formal analysis; A. M. Lentzsch, R. R., and A. M. Lambowitz validation; A. M. Lentzsch and J. Y. investigation; A. M. Lentzsch, J. Y., and R. R. methodology; A. M. Lentzsch, R. R., and A. M. Lambowitz writing-original draft; A. M. Lentzsch, J. Y., R. R., and A. M. Lambowitz writing-review and editing; R. R. and A. M. Lambowitz resources; R. R. and A. M. Lambowitz supervision; R. R. and A. M. Lambowitz funding acquisition; A. M. Lambowitz project administration.

Acknowledgment—We thank Dr. Jennifer Stamos for comments on the manuscript.

References

- Coffin, J. M. (1979) Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J. Gen. Virol.* **42**, 1–26 [CrossRef Medline](#)
- Gilboa, E., Mitra, S. W., Goff, S., and Baltimore, D. (1979) A detailed model of reverse transcription and tests of crucial aspects. *Cell* **18**, 93–100 [CrossRef Medline](#)
- Peliska, J. A., and Benkovic, S. J. (1992) Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science* **258**, 1112–1118 [CrossRef Medline](#)
- Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R., and Siebert, P. D. (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *BioTechniques* **30**, 892–897 [CrossRef Medline](#)
- Eickbush, T. H., and Jamburuthugoda, V. K. (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* **134**, 221–234 [CrossRef Medline](#)
- Mohr, S., Ghanem, E., Smith, W., Sheeter, D., Qin, Y., King, O., Polioudakis, D., Iyer, V. R., Hunicke-Smith, S., Swamy, S., Kuersten, S., and Lambowitz, A. M. (2013) Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**, 958–970 [CrossRef Medline](#)
- Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., and Sandberg, R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 [CrossRef Medline](#)
- Nottingham, R. M., Wu, D. C., Qin, Y., Yao, J., Hunicke-Smith, S., and Lambowitz, A. M. (2016) RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA* **22**, 597–613 [CrossRef Medline](#)
- Qin, Y., Yao, J., Wu, D. C., Nottingham, R. M., Mohr, S., Hunicke-Smith, S., and Lambowitz, A. M. (2016) High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. *RNA* **22**, 111–128 [CrossRef Medline](#)
- Basu, V. P., Song, M., Gao, L., Rigby, S. T., Hanson, M. N., and Bambara, R. A. (2008) Strand transfer events during HIV-1 reverse transcription. *Virus Res.* **134**, 19–38 [CrossRef Medline](#)
- Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R., and Mak, J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.* **2**, 461–472 [CrossRef Medline](#)
- Onafuwa-Nuga, A., and Telesnitsky, A. (2009) The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol. Mol. Biol. Rev.* **73**, 451–480 [CrossRef Medline](#)
- Streeck, H., Li, B., Poon, A. F., Schneidewind, A., Gladden, A. D., Power, K. A., Daskalakis, D., Bazner, S., Zuniga, R., Brander, C., Rosenberg, E. S., Frost, S. D., Altfield, M., and Allen, T. M. (2008) Immune-driven recombination and loss of control after HIV superinfection. *J. Exp. Med.* **205**, 1789–1796 [CrossRef Medline](#)

14. Ritchie, A. J., Cai, F., Smith, N. M., Chen, S., Song, H., Brackenridge, S., Abdoal Karim, S. S., Korber, B. T., McMichael, A. J., Gao, F., and Goonetilleke, N. (2014) Recombination-mediated escape from primary CD8+ T cells in acute HIV-1 infection. *Retrovirology* **11**, 69 [CrossRef Medline](#)
15. Song, H., Giorgi, E. E., Ganusov, V. V., Cai, F., Athreya, G., Yoon, H., Carja, O., Hora, B., Hraber, P., Romero-Severson, E., Jiang, C., Li, X., Wang, S., Li, H., Salazar-Gonzalez, J. F., et al. (2018) Tracking HIV-1 recombination to resolve its contribution to HIV-1 evolution in natural infection. *Nat. Commun.* **9**, 1928 [CrossRef Medline](#)
16. Zajac, P., Islam, S., Hochgerner, H., Lönnerberg, P., and Linnarsson, S. (2013) Base preferences in non-templated nucleotide incorporation by MMLV-derived reverse transcriptases. *PLoS ONE* **8**, e85270 [CrossRef Medline](#)
17. Oz-Gleenberg, I., Herschhorn, A., and Hizi, A. (2011) Reverse transcriptases can clamp together nucleic acids strands with two complementary bases at their 3'-termini for initiating DNA synthesis. *Nucleic Acids Res.* **39**, 1042–1053 [CrossRef Medline](#)
18. Oz-Gleenberg, I., Herzig, E., Voronin, N., and Hizi, A. (2012) Substrate variations that affect the nucleic acid clamp activity of reverse transcriptases. *FEBS J.* **279**, 1894–1903 [CrossRef Medline](#)
19. Xiong, Y., and Eickbush, T. H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**, 3353–3362 [CrossRef Medline](#)
20. Lambowitz, A. M., and Belfort, M. (2015) Mobile bacterial Group II introns at the crux of eukaryotic evolution. *Microbiol. Spectr.* **3**, MDNA3-0050-2014 [CrossRef Medline](#)
21. Blocker, F. J., Mohr, G., Conlan, L. H., Qi, L., Belfort, M., and Lambowitz, A. M. (2005) Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA* **11**, 14–28 [CrossRef Medline](#)
22. Collins, R. A., Stohl, L. L., Cole, M. D., and Lambowitz, A. M. (1981) Characterization of a novel plasmid DNA found in mitochondria of *N. crassa*. *Cell* **24**, 443–452 [CrossRef Medline](#)
23. Kuiper, M. T., and Lambowitz, A. M. (1988) A novel reverse transcriptase activity associated with mitochondrial plasmids of *Neurospora*. *Cell* **55**, 693–704 [CrossRef Medline](#)
24. Wang, H., and Lambowitz, A. M. (1993) The Mauriceville plasmid reverse transcriptase can initiate cDNA synthesis *de novo* and may be related to reverse transcriptase and DNA polymerase progenitor. *Cell* **75**, 1071–1081 [CrossRef Medline](#)
25. Chen, B., and Lambowitz, A. M. (1997) *De novo* and DNA primer-mediated initiation of cDNA synthesis by the Mauriceville retroplasmid reverse transcriptase involve recognition of a 3'-CCA sequence. *J. Mol. Biol.* **271**, 311–332 [CrossRef Medline](#)
26. Kennell, J. C., Wang, H., and Lambowitz, A. M. (1994) The Mauriceville plasmid of *Neurospora* spp. uses novel mechanisms for initiating reverse transcription *in vivo*. *Mol. Cell. Biol.* **14**, 3094–3107 [CrossRef Medline](#)
27. Chiang, C. C., and Lambowitz, A. M. (1997) The Mauriceville retroplasmid reverse transcriptase initiates cDNA synthesis *de novo* at the 3' end of tRNAs. *Mol. Cell. Biol.* **17**, 4526–4535 [CrossRef Medline](#)
28. Galligan, J. T., and Kennell, J. C. (2007) *Microbial Linear Plasmids*, (Meinhardt, F., Klassen, R., eds) Vol. 7, pp. 163–185. Springer, Berlin
29. Baidyaroy, D., Hausner, G., and Bertrand, H. (2012) *In vivo* conformation and replication intermediates of circular mitochondrial plasmids in *Neurospora* and *Cryphonectria parasitica*. *Fungal Biol.* **116**, 919–931 [CrossRef Medline](#)
30. Akins, R. A., Kelley, R. L., and Lambowitz, A. M. (1986) Mitochondrial plasmids of *Neurospora*: integration into mitochondrial DNA and evidence for reverse transcription in mitochondria. *Cell* **47**, 505–516 [CrossRef Medline](#)
31. Chiang, C. C., Kennell, J. C., Wanner, L. A., and Lambowitz, A. M. (1994) A mitochondrial retroplasmid integrates into mitochondrial DNA by a novel mechanism involving the synthesis of a hybrid cDNA and homologous recombination. *Mol. Cell. Biol.* **14**, 6419–6432 [CrossRef Medline](#)
32. Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 [CrossRef Medline](#)
33. Bibillo, A., and Eickbush, T. H. (2002) The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. *J. Mol. Biol.* **316**, 459–473 [CrossRef Medline](#)
34. Bibillo, A., and Eickbush, T. H. (2004) End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. *J. Biol. Chem.* **279**, 14945–14953 [CrossRef Medline](#)
35. George, J. A., Burke, W. D., and Eickbush, T. H. (1996) Analysis of the 5' junctions of R2 insertions with the 28S gene: implications for non-LTR retrotransposition. *Genetics* **142**, 853–863 [Medline](#)
36. Buzdin, A. A. (2004) Retroelements and formation of chimeric retrogenes. *Cell. Mol. Life Sci.* **61**, 2046–2059 [CrossRef Medline](#)
37. Khadgi, B. B., Govindaraju, A., and Christensen, S. M. (2019) Completion of LINE integration involves an open “4-way” branched DNA intermediate. *Nucleic Acids Res.* **47**, 8708–8719 [CrossRef Medline](#)
38. Mohr, G., Kang, S. Y., Park, S. K., Qin, Y., Grohman, J., Yao, J., Stamos, J. L., and Lambowitz, A. M. (2018) A highly proliferative group IIC intron from *Geobacillus stearothermophilus* reveals new features of group II intron mobility and splicing. *J. Mol. Biol.* **430**, 2760–2783 [CrossRef Medline](#)
39. White, T. B., and Lambowitz, A. M. (2012) The retrohoming of linear group II intron RNAs in *Drosophila melanogaster* occurs by both DNA ligase 4-dependent and -independent mechanisms. *PLoS Genet.* **8**, e1002534 [CrossRef Medline](#)
40. Wu, D. C., and Lambowitz, A. M. (2017) Facile single-stranded DNA sequencing of human plasma DNA via thermostable group II intron reverse transcriptase template switching. *Sci. Rep.* **7**, 8421 [CrossRef Medline](#)
41. Stamos, J. L., Lentzsch, A. M., and Lambowitz, A. M. (2017) Structure of a thermostable group II intron reverse transcriptase with template-primer and its functional and evolutionary implications. *Mol. Cell* **68**, 926–939. [e4 CrossRef Medline](#)
42. Jamburuthugoda, V. K., and Eickbush, T. H. (2014) Identification of RNA binding motifs in the R2 retrotransposon-encoded reverse transcriptase. *Nucleic Acids Res.* **42**, 8405–8415 [CrossRef Medline](#)
43. Shibutani, S., Takeshita, M., and Grollman, A. P. (1997) Translesional synthesis on DNA templates containing a single abasic site. A mechanistic study of the “A rule”. *J. Biol. Chem.* **272**, 13916–13922 [CrossRef Medline](#)
44. Golinelli, M.-P., and Hughes, S. H. (2002) Nontemplated nucleotide addition by HIV-1 reverse transcriptase. *Biochemistry* **41**, 5894–5906 [CrossRef Medline](#)
45. Golinelli, M.-P., and Hughes, S. H. (2002) Nontemplated base addition by HIV-1 RT can induce nonspecific strand transfer *in vitro*. *Virology* **294**, 122–134 [CrossRef Medline](#)
46. Luczkowiak, J., Matamoros, T., and Menéndez-Arias, L. (2018) Template-primer binding affinity and RNase H cleavage specificity contribute to the strand transfer efficiency of HIV-1 reverse transcriptase. *J. Biol. Chem.* **293**, 13351–13363 [CrossRef Medline](#)
47. Johnson, K. A., Simpson, Z. B., and Blom, T. (2009) Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem.* **387**, 20–29 [CrossRef Medline](#)
48. Fiala, K. A., Brown, J. A., Ling, H., Kshetry, A. K., Zhang, J., Taylor, J.-S., Yang, W., and Suo, Z. (2007) Mechanism of template-independent nucleotide incorporation catalyzed by a template-dependent DNA polymerase. *J. Mol. Biol.* **365**, 590–602 [CrossRef Medline](#)
49. Brenlla, A., Markiewicz, R. P., Rueda, D., and Romano, L. J. (2014) Nucleotide selection by the Y-family DNA polymerase Dpo4 involves template translocation and misalignment. *Nucleic Acids Res.* **42**, 2555–2563 [CrossRef Medline](#)
50. Motea, E. A., and Berdis, A. J. (2010) Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase. *Biochim. Biophys. Acta.* **1804**, 1151–1166 [CrossRef Medline](#)
51. Xu, H., Yao, J., Wu, D. C., and Lambowitz, A. M. (2019) Improved TGIRT-seq methods for comprehensive transcriptome profiling with decreased adapter dimer formation and bias correction. *Sci. Rep.* **9**, 7953 [CrossRef Medline](#)
52. Varani, G., and McClain, W. H. (2000) The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* **1**, 18–23 [CrossRef Medline](#)

EDITORS' PICK: Group II intron RT template switching

53. Raper, A. T., Reed, A. J., and Suo, Z. (2018) Kinetic mechanism of DNA polymerases: contributions of conformational dynamics and a third divalent metal ion. *Chem. Rev.* **118**, 6000–6025 [CrossRef Medline](#)
54. Washington, M. T., Prakash, L., and Prakash, S. (2001) Yeast DNA polymerase η utilizes an induced-fit mechanism of nucleotide incorporation. *Cell* **107**, 917–927 [CrossRef Medline](#)
55. Rechkoblit, O., Malinina, L., Cheng, Y., Kuryavyi, V., Broyde, S., Geacintov, N. E., and Patel, D. J. (2006) Stepwise translocation of Dpo4 polymerase during error-free bypass of an oxoG lesion. *PLoS Biol.* **4**, e11 [CrossRef Medline](#)
56. Clark, J. M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* **16**, 9677–9686 [CrossRef Medline](#)
57. Oz-Gleenberg, I., Herzig, E., and Hizi, A. (2012) Template-independent DNA synthesis activity associated with the reverse transcriptase of the long terminal repeat retrotransposon Tf1. *FEBS J.* **279**, 142–153 [CrossRef Medline](#)
58. Obeid, S., Blatter, N., Kranaster, R., Schnur, A., Diederichs, K., Welte, W., and Marx, A. (2010) Replication through an abasic DNA lesion: structural basis for adenine selectivity. *EMBO J.* **29**, 1738–1747 [CrossRef Medline](#)
59. Liu, S., Abbondanzieri, E. A., Rausch, J. W., Le Grice, S. F., and Zhuang, X. (2008) Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. *Science* **322**, 1092–1097 [CrossRef Medline](#)
60. Malik, O., Khamis, H., Rudnizky, S., Marx, A., and Kaplan, A. (2017) Pausing kinetics dominates strand-displacement polymerization by reverse transcriptase. *Nucleic Acids Res.* **45**, 10190–10205 [CrossRef Medline](#)
61. Moldovan, J. B., Wang, Y., Shuman, S., Mills, R. E., and Moran, J. V. (2019) RNA ligation precedes the retrotransposition of U6/LINE-1 chimeric RNA. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20612–20622 [CrossRef Medline](#)
62. Coros, C. J., Piazza, C. L., Chalamcharla, V. R., and Belfort, M. (2008) A mutant screen reveals RNase E as a silencer of group II intron retromobility in *Escherichia coli*. *RNA* **14**, 2634–2644 [CrossRef Medline](#)
63. Wu, J., Liu, W., and Gong, P. (2015) A structural overview of RNA-dependent RNA polymerases from the Flaviviridae family. *Int. J. Mol. Sci.* **16**, 12943–12957 [CrossRef Medline](#)
64. Simon-Loriere, E., and Holmes, E. C. (2011) Why do RNA viruses recombine? *Nat. Rev. Microbiol.* **9**, 617–626 [CrossRef Medline](#)
65. Iwasaki, Y. W., Siomi, M. C., and Siomi, H. (2015) PIWI-Interacting RNA: its biogenesis and functions. *Annu. Rev. Biochem.* **84**, 405–433 [CrossRef Medline](#)
66. Yu, Y., Jia, T., and Chen, X. (2017) The 'how' and "where" of plant microRNAs. *New Phytol.* **216**, 1002–1017 [CrossRef Medline](#)
67. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 [CrossRef](#)