JMB

ELSEVIER

Multiple Nucleotide Preferences Determine Cleavage-Site Recognition by the HIV-1 and M-MuLV RNases H

Sharon J. Schultz, Miaohua Zhang and James J. Champoux*

Department of Microbiology, School of Medicine, Box 357242, University of Washington, Seattle, WA 98195, USA

The RNase H activity of reverse transcriptase is required during retroviral replication and represents a potential target in antiviral drug therapies. Sequence features flanking a cleavage site influence the three types of retroviral RNase H activity: internal, DNA 3′-end-directed, and RNA 5′-end-directed. Using the reverse transcriptases of HIV-1 (human immunodeficiency virus type 1) and Moloney murine leukemia virus (M-MuLV), we evaluated how individual base preferences at a cleavage site direct retroviral RNase H specificity. Strong test cleavage sites (designated as between nucleotide positions −1 and +1) for the HIV-1 and M-MuLV enzymes were introduced into model hybrid substrates designed to assay internal or DNA 3′-end-directed cleavage, and base substitutions were tested at specific nucleotide positions. For internal cleavage, positions +1, −2, −4, −5, −10, and −14 for HIV-1 and positions +1, −2, −6, and −7 for M-MuLV significantly affected RNase H cleavage efficiency, while positions −7 and −12 for HIV-1 and positions −4, −9, and −11 for M-MuLV had more modest effects. DNA 3′-end-directed cleavage was influenced substantially by positions +1, −2, −4, and −5 for HIV-1 and positions +1, −2, −6, and −7 for M-MuLV. Cleavage-site distance from the recessed end did not affect sequence preferences for M-MuLV reverse transcriptase. Based on the identified sequence preferences, a cleavage site recognized by both HIV-1 and M-MuLV enzymes was introduced into a sequence that was otherwise resistant to RNase H. The isolated RNase H domain of M-MuLV reverse transcriptase retained sequence preferences at positions +1 and −2 despite prolific cleavage in the absence of the polymerase domain. The sequence preferences of retroviral RNase H likely reflect structural features in the substrate that favor cleavage and represent a novel specificity determinant to consider in drug design.

© 2010 Elsevier Ltd. All rights reserved.

Keywords: RNase H; reverse transcription; human immunodeficiency virus type 1, HIV-1; Moloney murine leukemia virus, M-MuLV; reverse transcriptase

Edited by J. Karn

Introduction

In reverse transcription, the single-stranded RNA genome of a retrovirus is copied into a linear double-stranded DNA that integrates into the host cell chromosome.[1,2] During synthesis, a host cell tRNA is used to initiate minus-strand DNA synthesis, while a sequence in the viral genome termed the polypurine tract (PPT) is cleaved and subsequently used as the primer to initiate plus-strand DNA synthesis. Unique sequences at the 5′ and 3′ ends of the viral genome are duplicated through template switches to generate long terminal repeats at both ends of the unintegrated viral DNA. The virally encoded reverse transcriptase contains an RNase H activity and an RNA- and DNA-dependent DNA polymerase activity, and both activities are required for viral replication.[3–6] During reverse transcription, reverse transcriptase

*Corresponding author. E-mail address: champoux@u.washington.edu.

Abbreviations used: M-MuLV, Moloney murine leukemia virus; PPT, polypurine tract; HBD, hybrid binding domain.

carries out DNA polymerization, strand transfers, and cleavage of the viral RNA genome once it has been used as a template to produce RNA/DNA hybrids.

The reverse transcriptase of HIV-1 (human immunodeficiency virus type 1) is a heterodimer containing 66-kDa (p66) and 51-kDa (p51) subunits.[7] The p66 subunit has an amino-terminal DNA polymerase domain that is followed by the connection and RNase H domains, while the p51 subunit is a proteolytic fragment of p66 that lacks the carboxy-terminal RNase H domain. The reverse transcriptase of Moloney murine leukemia virus (M-MuLV) is a 76-kDa monomer that contains an amino-terminal DNA polymerase domain, a connection domain, and a carboxy-terminal RNase H domain.[8] Crystal structures of the human and murine reverse transcriptases indicate similar nucleic acid binding clefts and a comparable tertiary structure for the RNase H domains.[9–13] In addition, co-crystal structures of the HIV-1 enzyme indicate that the distance between the active sites of the polymerase and RNase H domains is approximately 17–18 nt as measured on a template strand and that the nucleic acid binding cleft has numerous and varied contacts with the substrate, the majority of which are observed between the polymerase domain and the DNA region proximal to the 3′-primer terminus.[11,12,14]

The RNase H activity of reverse transcriptase facilitates reverse transcription in several ways.[15,16] RNase H degrades the viral genome in the RNA/DNA hybrids formed during minus-strand synthesis. Genome degradation assists strand transfers and plus-strand synthesis and represents the majority of RNase H cleavages required during reverse transcription. RNase H carries out a sequence-specific cleavage in the viral genome that generates the PPT primer required for plus-strand initiation. Also, with sequence specificity, RNase H removes the tRNA and PPT primers used to initiate minus-strand synthesis and plus-strand DNA synthesis, respectively. Correct generation of the PPT primer and proper removal of the PPT and tRNA primers create the long terminal repeat ends required for integration. Despite these essential roles in reverse transcription, no clinically relevant drug has yet been developed against the RNase H activity of reverse transcriptase (reviewed by Klumpp and Mirzadegan[17]).

Based upon the interactions of reverse transcriptase with the substrate, RNase H cleavages have been classified into three distinct types (for a review, see Refs. [15] and [18]). Internal RNase H cleavages occur when reverse transcriptase binds an RNA/DNA hybrid without associating with the end of an RNA or DNA strand.[19–22] Interestingly, internal cleavage sites are not selected at random. Through an analysis of base frequencies found at the nucleotide positions surrounding internal cleavage sites, we showed that the HIV-1 and M-MuLV reverse transcriptases have a statistically significant preference for certain nucleotides at specific positions.[23] With the scissile phosphate defined as occurring between positions −1 and +1 and strong or moderate preferences indicated by uppercase or lowercase letters, respectively, HIV-1 reverse transcriptase has nucleotide preferences at positions +1 (A/U), −2 (C/G), −4 (C/g/u), −7 (G/C), −12 (A/g/u), and −14 (A/g). The M-MuLV reverse transcriptase has similar preferences at +1 (A/U) and −2 (C/G) but shows unique preferences at positions −6 (C/G/u) and −11 (A).

The two other types of RNase H cleavage result from the preferential association of the polymerase domain of reverse transcriptase to recessed DNA 3′ or RNA 5′ ends in a hybrid.[24–27] Both the distance from and the type of recessed end are important determinants for end-directed cleavages. DNA 3′-end-directed cleavages occur ∼15–20 nt away from the recessed terminus for HIV-1 reverse transcriptase and ∼17–20 nt away for the M-MuLV enzyme, while RNA 5′-end-directed cleavages are ∼13–20 nt away from the recessed RNA 5′ end for both enzymes.[28] Based on a statistical analysis, sequence is also an important determinant of end-directed cleavages, and DNA 3′-end-directed and RNA 5′-end-directed types of cleavage share the same nucleotide preferences at specific positions flanking a cleavage site.[28,29] Both HIV-1 and M-MuLV enzymes have strong preferences at positions +1 (A/U/C) and −2 (C/G), as well as at position −4 (U/C/G) for the human enzyme.

At present, the relative importance of each preferred position as inferred from the statistical analyses is unknown. It is also possible that some positions that emerged from the statistical analyses reflect a chance association in the particular sequences chosen and were not related to enzyme specificity. To distinguish whether all of the positions identified by the statistical analyses affect cleavage equally or whether some positions are more important than others, we tested the effects of base substitutions at several nucleotide positions near cleavage sites recognized by the HIV-1 and M-MuLV reverse transcriptases. Our analyses show that some positions are more influential than others in promoting cleavage and identify additional position preferences for each enzyme. Also, we tested if any sequence preferences are displayed by the isolated RNase H domain of M-MuLV, which retains enzymatic activity but is unable to perform sequence-specific cleavages, such as PPT primer generation or removal.[30,31] These data offer insights into the sequence preferences that influence general genome degradation by the HIV-1 and M-MuLV reverse transcriptases and may suggest new opportunities for targeting drugs specific to the retroviral RNases H.

Results

It was important to identify a strong and relatively isolated cleavage site that was recognized by each enzyme to examine base preferences at individual positions flanking an RNase H cleavage site

recognized by HIV-1 or M-MuLV reverse transcriptase. From previous analyses of strong internal cleavage sites, we chose one RNA sequence with a strong site for the HIV-1 reverse transcriptase located 3 nt downstream of a strong site for the M-MuLV reverse transcriptase[23] and generated the 49mer RNA termed Rmzc that contained both sites (see Materials and Methods). From the 5′ end of Rmzc,

the HIV-1 site is located between nucleotides 33 and 34 (site 33/34) and the M-MuLV site is between nucleotides 30 and 31 (site 30/31) (Figs. 1a and 4a). The convenient proximity of these sites allowed a single base substitution in Rmzc to test different nucleotide positions relative to each cleavage site. For example, changing the base at position −5 for site 33/34 also changes position −2 for site 30/31.
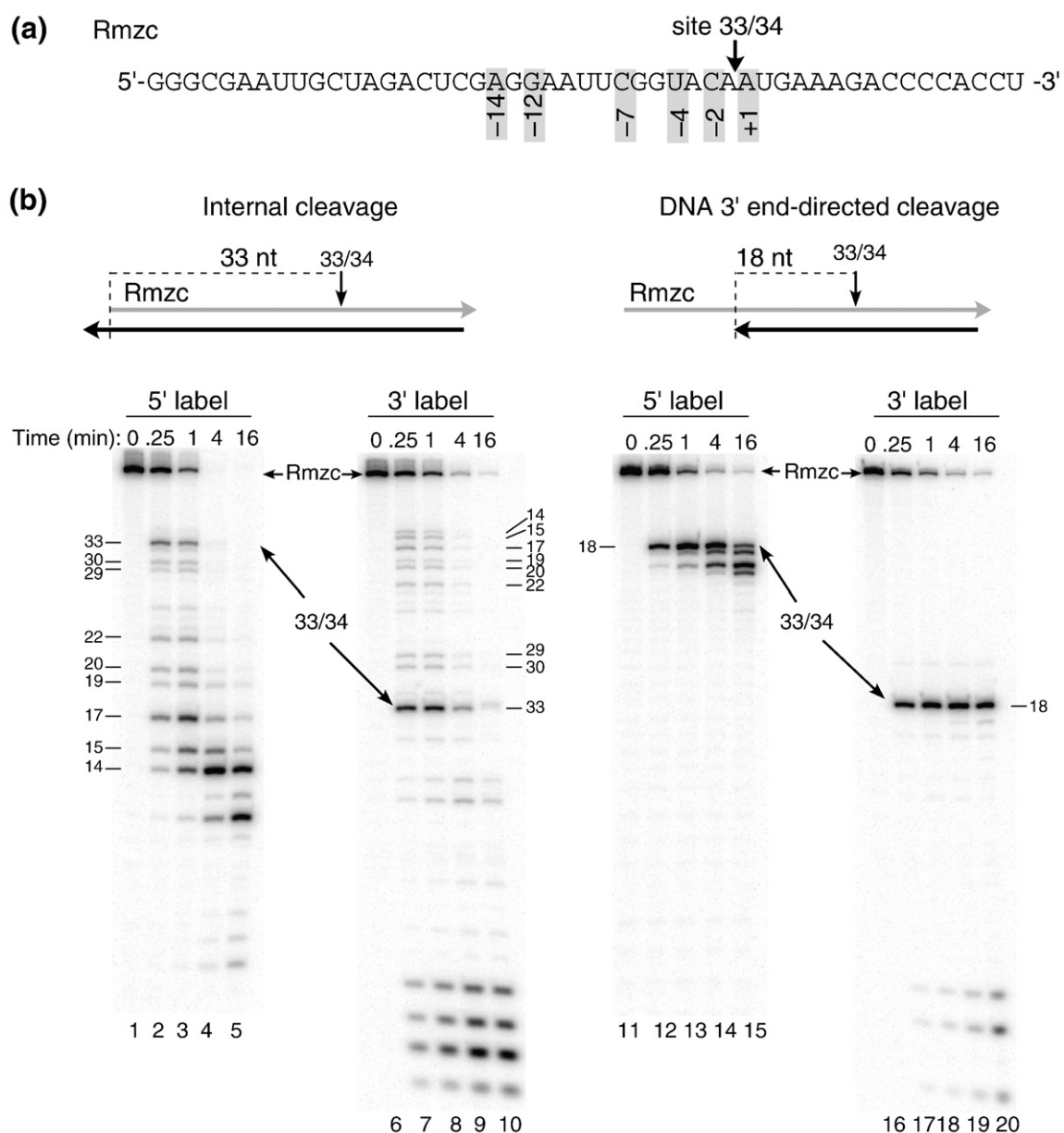
Fig. 1. Model substrates to analyze internal and DNA 3′-end-directed cleavages by HIV-1 reverse transcriptase. (a) The sequence of 49mer RNA Rmzc is shown from 5′ to 3′, and site 33/34, the strong cleavage site recognized by HIV-1 reverse transcriptase, is indicated with an arrow. For site 33/34, the positions of preferred nucleotides as determined from an earlier statistical analysis are indicated by the numbers in shaded boxes. (b) Cleavage of site 33/34 in Rmzc hybrids designed to test internal (left panel) or DNA 3′-end-directed (right panel) cleavage. Above each panel is a schematic of the hybrid substrate containing Rmzc (gray) and a DNA strand (black). HIV-1 reverse transcriptase was incubated with hybrids containing 5′-end-labeled Rmzc (5′ label; lanes 1–5 and 11–15) or 3′-end-labeled Rmzc (3′ label; lanes 6–10 and 16–20) in time course assays. Samples were either left untreated (0 min) or removed at 0.25, 1, 4, or 16 min, analyzed in denaturing 20% polyacrylamide gels, and visualized using the PhosphorImager. The positions of Rmzc and the products resulting from cleavage at site 33/34 are indicated with arrows, and the sizes of cleavage products are indicated in nucleotides.

Substitutions that affect HIV-1 RNase H cleavage

The sequence surrounding site 33/34 in Rmzc matched the preferred nucleotides previously identified at positions +1, −2, −4, −7, −12, and −14 for internal cleavage by HIV-1 RNase H (Fig. 1a).[23] Hybrid substrates were created to promote internal or DNA 3′-end-directed cleavage at site 33/34 (Fig. 1b, top illustrations) by choosing the appropriate length of a complementary DNA strand. In the internal cleavage hybrid, site 33/34 was located 33 nt from the RNA 5′ end, while in the DNA 3′-end-directed cleavage hybrid, site 33/34 was positioned 18 nt from the recessed DNA 3′ end.

To evaluate cleavage of site 33/34 in the internal and DNA 3′-end-directed cleavage substrates, we used both 5′-end-labeled and 3′-end-labeled Rmzc RNAs. When the internal cleavage substrate containing 5′-end-labeled Rmzc was incubated with HIV-1 reverse transcriptase, internal cleavage at site 33/34 was exceptionally strong as the amount of 33mer product was similar to that of the 14mer to 17mer products generated by RNA 5′-end-directed cleavages in the 0.25- and 1-min time points (Fig. 1b, compare lanes 1–5). At later time points, additional cleavage at site 33/34 was masked by cleavages proximal to the labeled 5′ end. When the internal cleavage substrate contained 3′-end-labeled Rmzc, exceptionally strong cleavage at site 33/34 was again evident (Fig. 1b, lanes 6–10). With the use of the DNA 3′-end-directed cleavage substrate with 5′-end-labeled or 3′-end-labeled Rmzc, cleavage at site 33/34 primarily generated the 18mer product in the 0.25- and 1-min time points (Fig. 1b, lanes 11–13 and 16–18, respectively). In the following experiments that tested the effects of sequence changes on cleavage, the 1-min time points were used as an indicator of the initial cleavage events.

In the internal cleavage assay, changing the preferred +1A to a disfavored G significantly decreased cleavage (Fig. 2a, compare lanes 2 and 4). Similarly, changing the −2C to U or the −4U to A substantially decreased cleavage (Fig. 2a, lanes 6 and 10). While changing the −7C to U or −14A to C decreased cleavage, changing the −12G to C did not seem to affect internal cleavage (Fig. 2a, lanes 14, 18, and 16). Combining −7U or −12C with −14C decreased cleavage slightly more than the −14C substitution alone (Fig. 2a, compare lanes 20 and 22 with lane 18). Substitutions were also tested at some positions not previously associated with nucleotide preferences by the statistical tests. While a −3G did not alter cleavage, a −5U decreased internal cleavage (Fig. 2a, lanes 8 and 12). Also, replacing the −9U with A or G seemed to slightly reduce cleavage (Fig. 2a, lanes 24 and 26). Identical time course assays were repeated in multiple experiments, and the amount of the 33mer product observed in the 1-min time points was quantified to better evaluate the effects of these substitutions (Fig. 2b). The analyses showed that internal cleavage at site 33/34 was reduced by

more than 60% with +1G or −14C, by ∼70% for −5U, and by more than 90% for −2U or −4A. It was also evident that the substitutions of −7U and −12C were comparable with the −9A and −9G changes that had no substantial effect on internal cleavage. However, the reduction in cleavage by −14C was greater when combined with −7U or −12C.

To further evaluate the contributions of select nucleotide positions to internal cleavage, we compared the effects of all three base substitutions at a given position (Fig. 2c). Consistent with the strong cleavage observed at site 33/34, the original Rmzc sequence typically generated the highest levels of internal cleavage (left bar in each graph). Substitutions at position −2 indicated that U and A were deleterious and that C and G were very favorable. Substitutions at position −4 showed that all bases but A favored cleavage. This analysis also revealed new strong preferences at position −5 (G>A>C/U) and position −10 (A/U>G/C) and only slight preferences at position −3 (favors G) and position −9 (disfavors C).

To test how base substitutions affected DNA 3′-end-directed cleavage at site 33/34, we used substrates for DNA 3′-end-directed cleavage in RNase H assays with HIV-1 reverse transcriptase. As compared with the original Rmzc sequence, DNA 3′-end-directed cleavage was reduced by substitutions +1G, −2U, −4A, and −5U (Fig. 3a, compare lanes 2, 4, 6, 8, and 10). The substitution −7U appeared to slightly reduce cleavage, and single or even double substitutions located farther from site 33/34 did not decrease cleavage significantly (Fig. 3a, lanes 12, 14, 16, 18, and 20). DNA 3′-end-directed cleavage was substantially reduced when disfavored bases were combined at positions +1 and −2 (Fig. 3a, lane 24) and eliminated by additionally changing the −4 or −5 position (Fig. 3a, lanes 26 and 28, respectively). Quantitation of results from repeated experiments showed that the individual substitutions at position +1, −2, −4, or −5 and the combined −12C/−14C substitution reduced DNA 3′-end-directed cleavage by ∼50% or more (Fig. 3b). A slight reduction in cleavage occurred with the single substitutions at positions −7, −12, or −14 and the combined −7U/−14C substitution. DNA 3′-end-directed cleavages were decreased by more than 90% by the combined +1G/−2U change and barely detectable with the additional −4A or −5U change.

All three base substitutions were carried out at positions −2, −3, −4, −5, −9, and −10 using the DNA 3′-end-directed cleavage substrates (Fig. 3c). Cleavage at site 33/34 was highest in hybrids containing the original Rmzc sequence, and, in general, the results were similar to those observed for internal cleavage (compare Figs. 2c and 3c). Notably, at position −4, a G was favored over C in DNA 3′-end-directed cleavage, while the converse was true for internal cleavage. Consistent with the base preferences observed earlier, none of the substitutions at position −9 or −10 significantly affected DNA 3′-end-directed cleavage.
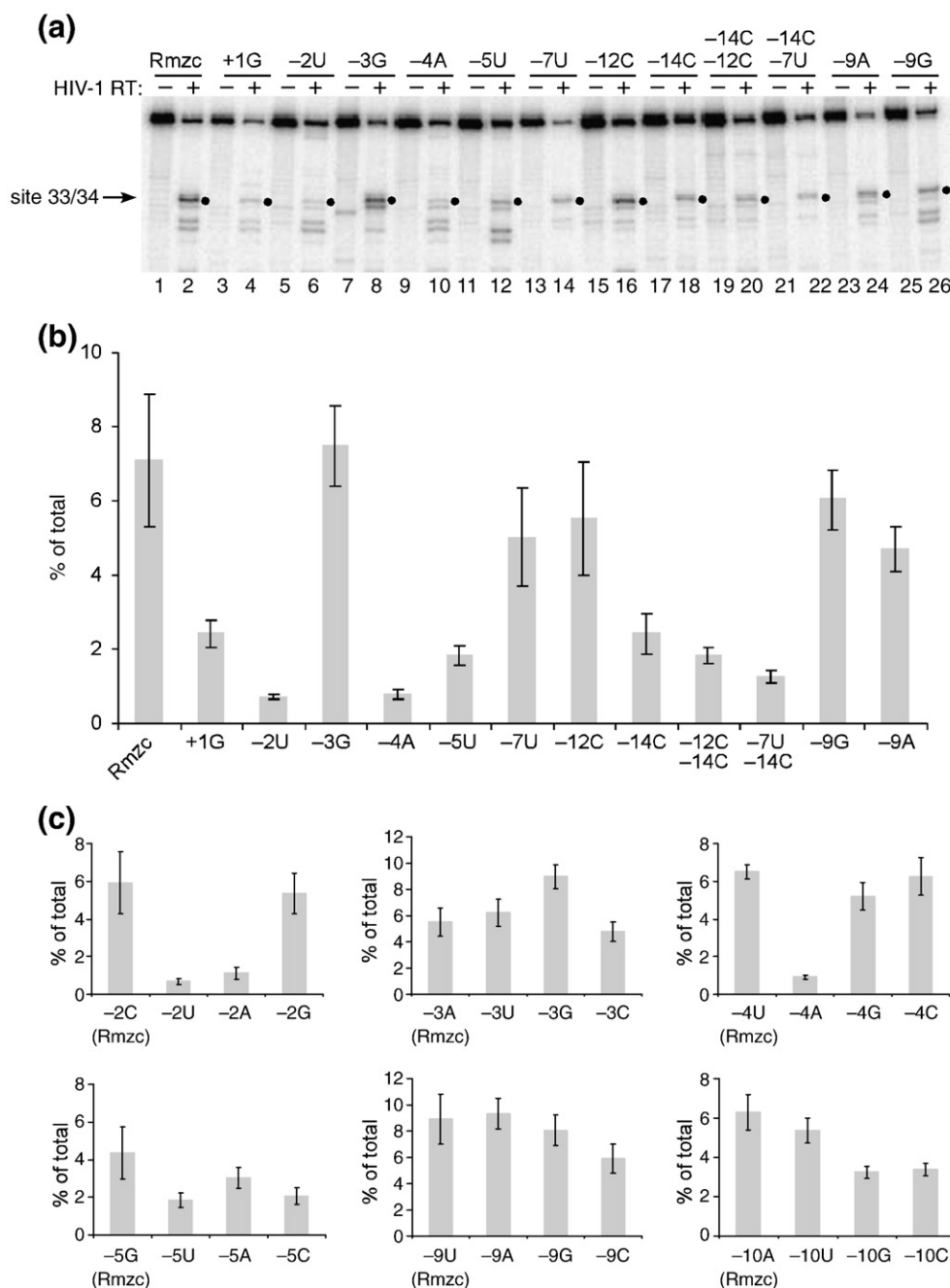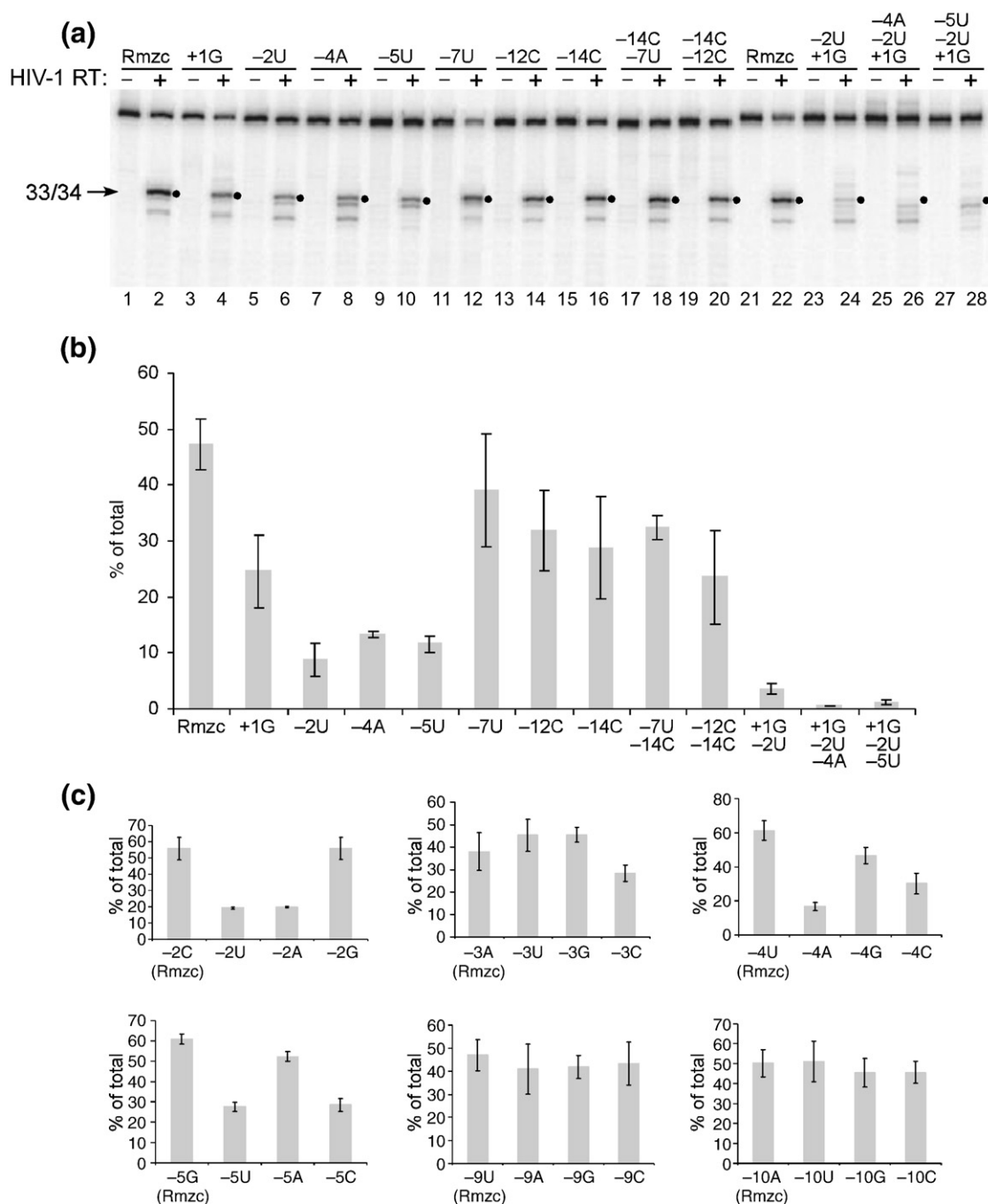
Fig. 2. Effects of nucleotide substitutions on internal RNase H cleavage by HIV-1 reverse transcriptase. (a) 5′-End-labeled Rmzc or Rmzc RNAs with the indicated substitutions relative to site 33/34 were used to generate internal cleavage substrates as shown in Fig. 1b. The positions and changes introduced into each substrate relative to site 33/34 are indicated above each lane by position number and base. These substrates were incubated with HIV-1 reverse transcriptase in time course experiments, and the 1-min time points (even-numbered lanes) or the untreated substrates (odd-numbered lanes) were analyzed as described in Fig. 1b. The relevant portion of the resulting gel image is shown, and filled circles indicate the position of the 33mer product resulting from cleavage at site 33/34. (b) Bar graph depicting the total amount of 33mer product (% of total) generated by internal cleavage at site 33/34 using data from the 1-min time point of five experiments with the substrates described in (a). (c) Bar graphs depicting the total amount of 33mer product (% of total) generated by internal cleavage at site 33/34 using substrates comparing cleavage of Rmzc (leftmost bar in each graph) with all three nucleotide substitutions at the indicated positions. Data from 1-min time points generated in five experiments are shown for all positions. For (b) and (c), error bars represent ±1 SD.

Substitutions that affect M-MuLV RNase H cleavage

Site 30/31 in the RNA Rmzc is three bases upstream of site 33/34 recognized by the HIV-1 RNase H and matches the preferred nucleotides previously identified at positions +1, −2, −6, and −11 for internal cleavage by M-MuLV RNase H (Fig. 4a).[23] When an internal cleavage substrate containing 5′-end-labeled

Fig. 3. Effects of nucleotide substitutions on DNA 3′-end-directed RNase H cleavage by HIV-1 reverse transcriptase. (a) DNA 3′-end-directed cleavage substrates as shown in Fig. 1b were made using 5′-end-labeled Rmzc or Rmzc RNAs with the indicated nucleotide substitutions, and the resulting substrates were analyzed as described in Fig. 2a. (b) A bar graph depicting the total amount of 33mer product (% of total) generated by DNA 3′-end-directed cleavage at site 33/34 from 1-min time points generated in five experiments using the substrates shown in (a). (c) Bar graphs depicting the total amount of 33mer product (% of total) generated by DNA 3′-end-directed cleavage at site 33/34 using substrates comparing cleavage of Rmzc (left) with substrates containing the three other nucleotides at the indicated positions. Data represent the 1-min time points generated in five (positions −2, −9, and −10) or four (positions −3, −4, and −5) experiments.

Rmzc was treated with M-MuLV reverse transcriptase, the amount of 30mer product resulting from cleavage at site 30/31 was only slightly less at the 0.25- and 1-min time points than those generated by RNA 5′-end-directed cleavages (15mer to 17mer products) (Fig. 4b, left panel, lanes 1–5). Results

with the substrate containing 3′-end-labeled Rmzc confirmed that site 30/31 was a strong and relatively isolated internal cleavage site (Fig. 4b, lanes 6–10). When positioned 18 nt from the DNA 3′ primer terminus in the DNA 3′-end-directed cleavage substrate, site 30/31 was cleaved exceptionally well with
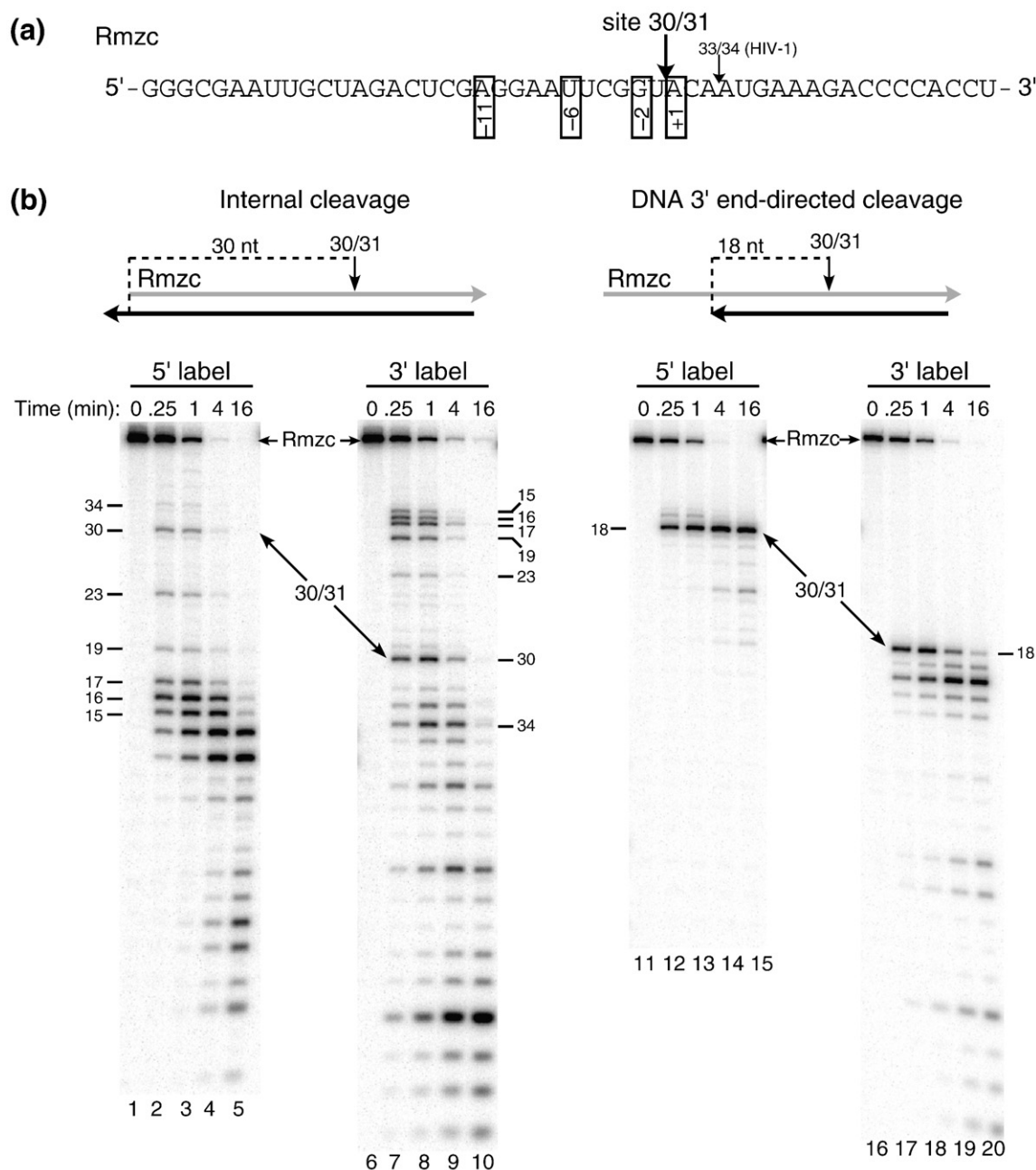
Fig. 4. Model substrates to analyze internal and DNA 3′-end-directed cleavages by M-MuLV reverse transcriptase. (a) The sequence of Rmzc is shown from 5′ to 3′, and site 30/31, the strong cleavage site recognized by M-MuLV reverse transcriptase, is indicated just upstream of site 33/34 (used for HIV-1 RNase H assays; see Fig. 1). The positions of preferred nucleotides as determined previously are indicated by the numbers in open boxes. (b) Cleavage of site 30/31 in Rmzc hybrids designed to test internal (left panel) or DNA 3′-end-directed (right panel) cleavage. The hybrid substrates are shown above each panel and configured as described in Fig. 1b. Hybrids containing Rmzc labeled at the 5′ end (lanes 1–5 and 11–15) or the 3′ end (lanes 6–10 and 16–20) were incubated with M-MuLV reverse transcriptase in time course assays, and samples were analyzed as described in Fig. 1b. The positions of Rmzc and the 30mer product resulting from cleavage at site 30/31 are indicated with arrows, and the sizes of cleavage products are indicated in nucleotides.

either end-labeled substrate (Fig. 4b, right panel, lanes 11–15 and 16–20).

Next, several substitutions were introduced at positions around site 30/31 and tested using internal cleavage substrates. Consistent with nucleotide preferences previously identified for positions +1 (A/U), −2 (C/G), and −6 (C/G/u), internal

cleavage at site 30/31 was reduced by ~80%–90% by the introduction of +1G, −2U, or −6A but was unaffected by the −6G change (Fig. 5a, compare lanes 2, 8, 12, 24, and 26; Fig. 5b). Changing the preferred −11A to C had a moderate effect, reducing cleavage by only ~30% (Fig. 5b). Substitutions at the non-preferred positions +4, +2, −1, and −4 did not
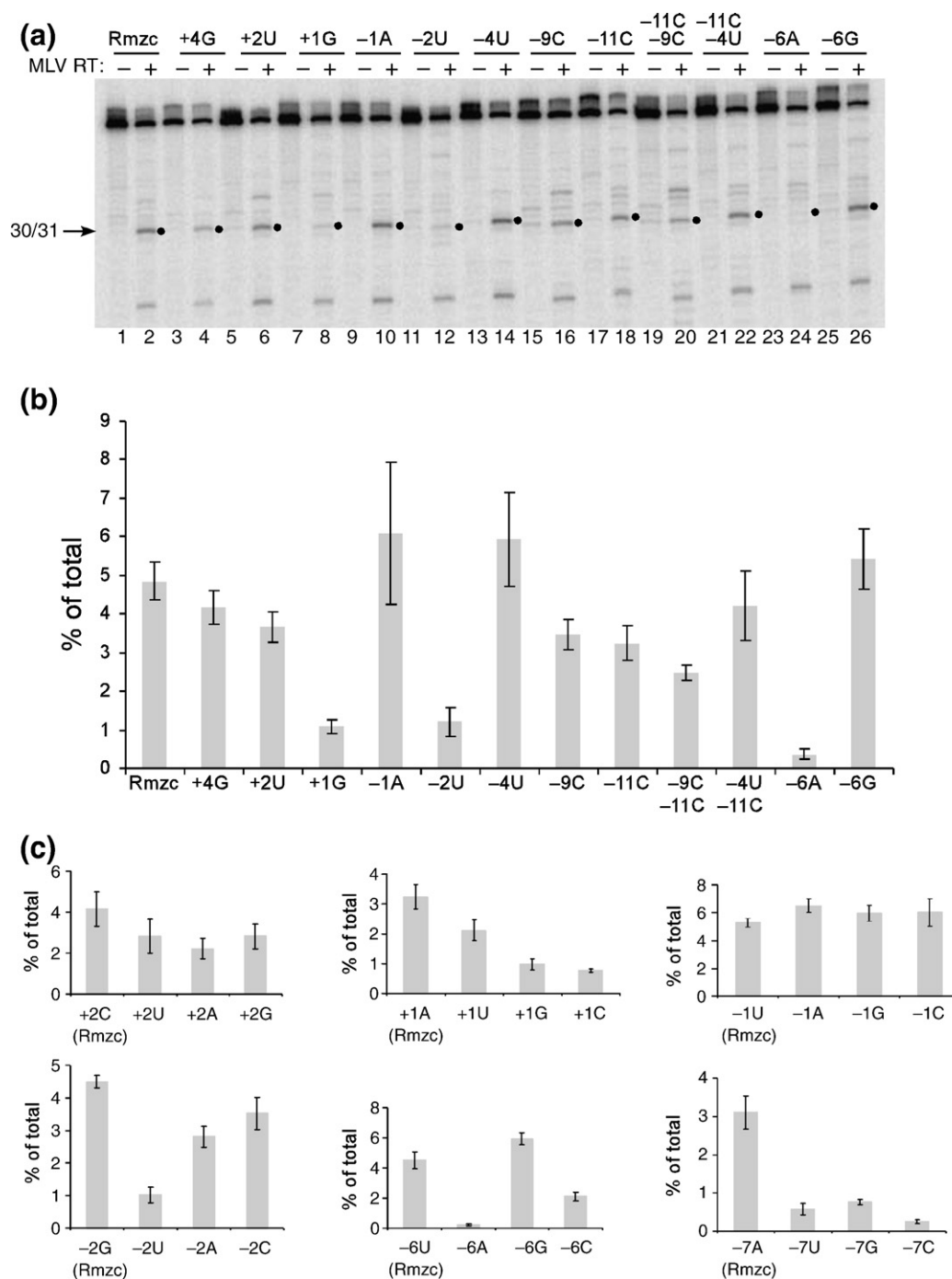
Fig. 5. Effects of nucleotide substitutions on internal RNase H cleavage by M-MuLV reverse transcriptase. (a) 5′-End-labeled Rmzc or Rmzc RNAs with the indicated substitutions relative to site 30/31 were used to generate internal cleavage substrates as described in Fig. 2. The positions and changes introduced into each substrate relative to site 30/31 are indicated above each lane by position number and base. These substrates were incubated with M-MuLV reverse transcriptase in time course experiments, and the 0.25-min time points (even-numbered lanes) or the untreated substrates (odd-numbered lanes) were analyzed. Filled circles indicate the position of the 30mer product resulting from cleavage at the 30/31 site. (b) A bar graph depicting the amount of 30mer product (% of total) generated by internal cleavage at site 30/31 from 0.25-min time points generated in five experiments using the substrates described in (a). (c) Bar graphs depicting the total amount of 30mer product (% of total) generated by internal cleavage at site 30/31 using substrates comparing cleavage of Rmzc (left) with substrates containing the indicated substitutions. Data represent the 0.25-min time points generated in five experiments.

significantly affect cleavage (Fig. 5a, lanes 4, 6, 10, and 14; Fig. 5b). As compared with the original Rmzc sequence, the −9C change reduced cleavage

almost equal to that of the −11C substitution and the combined −9C/−11C change reduced cleavage at site 30/31 by ~50% (Fig. 5b).

All three base substitutions were tested at several positions near site 30/31, including the preferred positions +1, −2, and −6 (Fig. 5c). Consistent with previously identified base preferences, internal cleavage was favored by A and U at position +1, decreased by U at position −2, and substantially reduced by A at position −6. Interestingly, a new preference was observed at position −7, where A was significantly favored over the other base choices. Also, a C was slightly preferred at position +2, and no base was favored at position −1.

Similar assays with M-MuLV reverse transcriptase were used to evaluate how base substitutions near site 30/31 in Rmzc influenced DNA 3′-end-directed cleavage. In the course of this analysis, we also asked if the sequence preferences that promote cleavage were affected by the distance of a cleavage

site from the DNA 3′ end by positioning site 30/31 at both 18 and 15 nt from the recessed 3′ end (Fig. 6). All three base substitutions were carried out at positions +1, −1, −2, −6, and −7. For the 18-nt substrate, strong preferences were observed at positions +1 (A/U), −2 (A/C/G), −6 (C/G/U), and −7 (A), but substitutions at position −1 were not significantly different from the control (Fig. 6a). Due to the optimal positioning of site 30/31 in the cleavage window,[28] site 30/31 was cleaved much more efficiently in nearly all of the 18-nt substrates as compared with the 15-nt substrates (compare Fig. 6a and b). Most importantly, the base preferences for the 18- and 15-nt substrates were essentially identical and were very similar to those observed for the internal cleavage substrates (compare Fig. 6 with Fig. 5c).
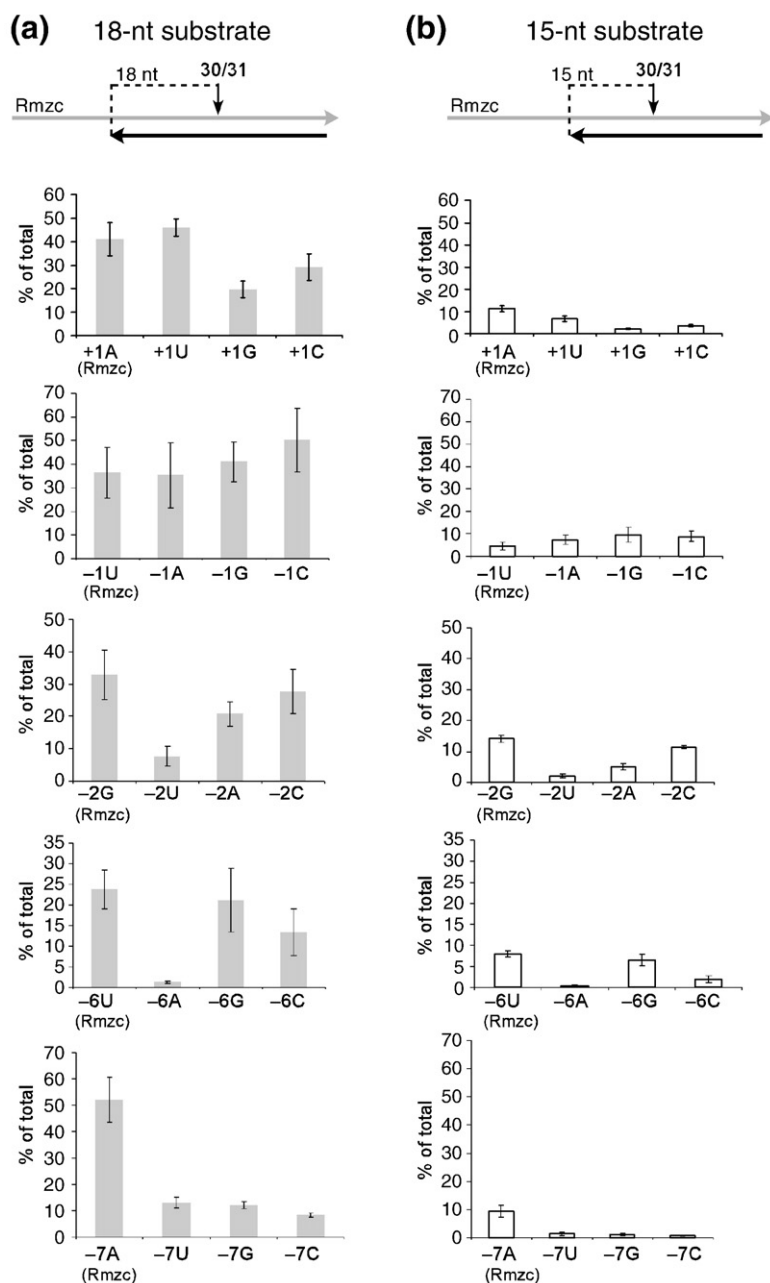

Fig. 6. Effects of nucleotide substitutions on DNA 3′-end-directed RNase H cleavage by M-MuLV reverse transcriptase. (a) DNA 3′-end-directed cleavage substrates that positioned site 30/31 at 18 nt from the recessed DNA 3′ end (18-nt substrate) were made using Rmzc or Rmzc RNAs with the indicated nucleotide substitutions, and the resulting substrates were analyzed as described in Fig. 3. The bar graphs depict the amount of 30mer product (% of total) generated by DNA 3′-end-directed cleavage at site 30/31 as described in Fig. 5c. Data represent the 0.25-min time points generated in five (positions −1, −6, and −7), four (position +1), or three (position −2) experiments. (b) DNA 3′-end-directed cleavage substrates that positioned site 30/31 at 15 nt from the recessed DNA 3′ end (15-nt substrate). Graphs are as described for (a).

Creation of a site recognized by internal cleavage and DNA 3′-end-directed cleavage

We next asked if a cleavage site could be introduced into a sequence that was resistant to RNase H activity. We were especially curious to test the newly identified preferences of the HIV-1 enzyme at position −5 (G/a) and the M-MuLV enzyme at position −7 (A). An AU-rich sequence that is not cleaved by either the HIV-1 RNase H or the M-MuLV RNase H is found in the region upstream of the −1/+1 cleavage site that generates the PPT primer in

the M-MuLV genome.[21,23,32] This sequence was used to generate RNA R−54/−14, where the proposed Site T (for target site) was within the resistant region between the 34th and 35th nucleotides from the RNA 5′ end (Fig. 7a). For Site T, preferred bases were already present at positions +1, −2, and −11 for M-MuLV and at positions +1, −2, −12, and −14 for HIV-1, but disfavored bases were present at positions −4 through −7 for both enzymes. A variety of base substitutions were introduced at positions −4 to −7, and cleavage was assayed using internal and DNA 3′-end-directed cleavage substrates.
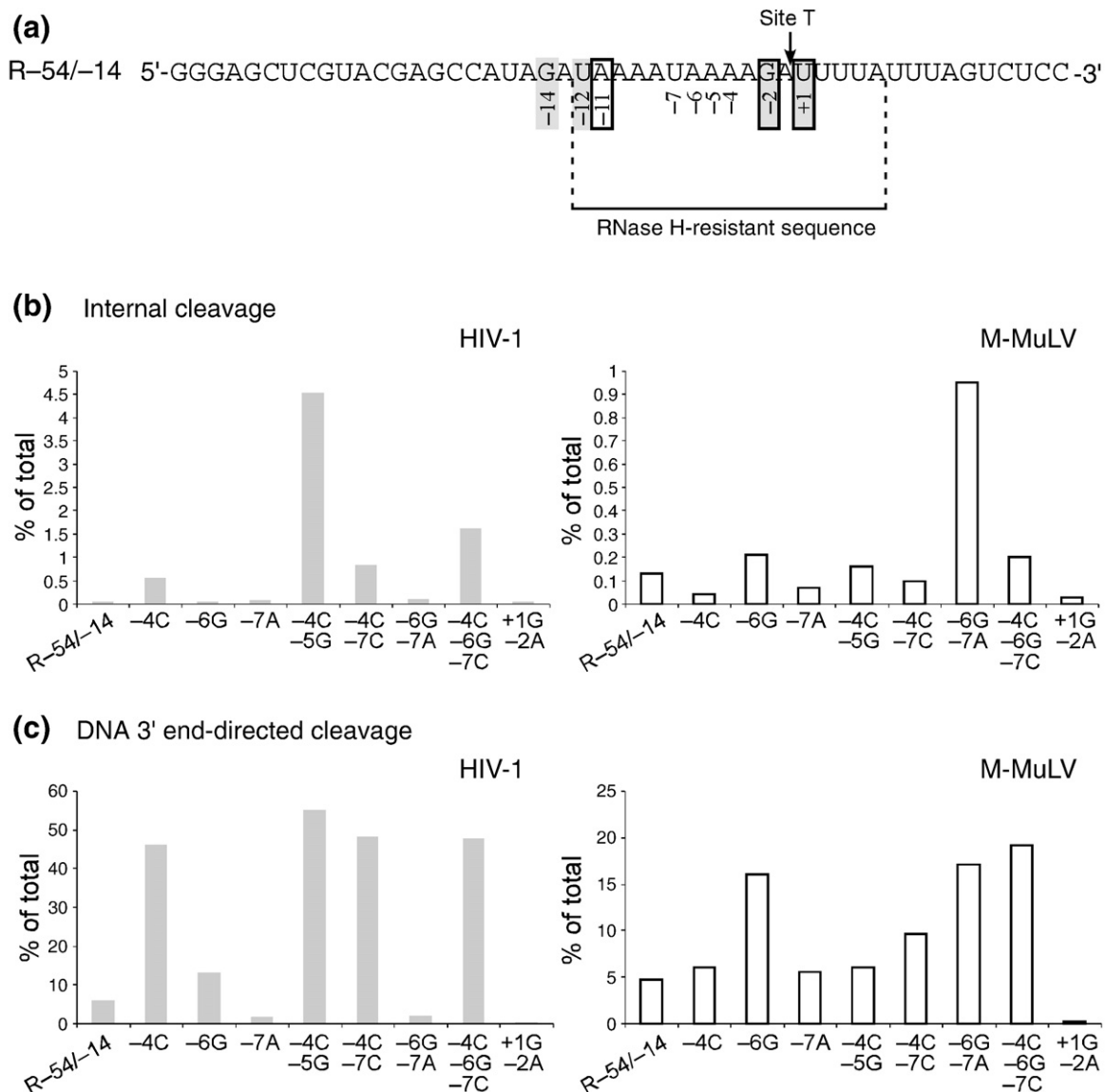
Fig. 7. Creation of Site T recognized by the HIV-1 and M-MuLV RNases H. (a) The sequence of RNA R−54/−14 is shown with the targeted location of a new cleavage site (Site T) in an RNase-H-resistant sequence. Site T has preferred nucleotides for HIV-1 at positions +1, −2, −12, and −14 (shadowed boxes) and for M-MuLV at positions +1, −2, and −11 (open boxes) but has disfavored nucleotides at positions −4 through −7. (b) Internal cleavage substrates containing R−54/−14 or substituted R−54/−14 RNAs were analyzed in time course assays and analyzed as described in Fig. 1. Graphs show the total amount of 34mer product (% of total) generated by internal cleavage at Site T using representative data from 1-min time points for HIV-1 reverse transcriptase and 0.25-min time points for M-MuLV reverse transcriptase. (c) DNA 3′-end-directed cleavage substrates containing R−54/−14 or substituted R−54/−14 RNAs with the DNA 3′ primer terminus located at 18 nt from Site T were analyzed, and the data are shown as described for (b). For (b) and (c), the data shown are from representative experiments.

Internal cleavage substrates containing RNA R −54/−14 or RNAs with the indicated substitutions were treated with HIV-1 reverse transcriptase, and the 1-min time points from time course assays were analyzed (Fig. 7b, HIV-1). A single substitution of −4A to C increased internal cleavage by 14-fold at Site T, while the combined −4C/−5G increased internal cleavage by more than 100-fold. Also for HIV-1, the combined −4C/−7C or −4C/−6G/−7C increased cleavage by 20-fold or 40-fold, respectively, but other changes had little or no effect on internal cleavage. For M-MuLV, internal cleavage at Site T increased by 7-fold with the combined −6G/−7A change, but only 2-fold or lower effects were observed with other substitutions.

These substitutions were also tested in a DNA 3′-end-directed cleavage assay with Site T positioned at 18 nt from the DNA 3′ end (Fig. 7c). Similar to internal cleavage, all changes that included the −4C increased DNA 3′-end-directed cleavage for HIV-1 reverse transcriptase by 8- to 9-fold and a −6G caused a 2-fold increase. For M-MuLV reverse transcriptase, cleavage was increased 3- to 4-fold by all

changes that included the −6G and 2-fold by the combined −4C/−7C. Finally, the low but detectable level of DNA 3′-end-directed cleavage observed at Site T for both enzymes was eliminated by the disfavored combined +1G/−2A change.

Nucleotide preferences of the isolated M-MuLV RNase H domain

In an earlier study, we showed that the isolated RNase H domain derived from the M-MuLV reverse transcriptase has enzymatic activity but was unable to generate the PPT primer or remove the extended PPT and tRNA primers.[30] Here, we asked if the isolated RNase H domain retained any of the sequence preferences involved in cleavage-site recognition.

First, multiple cleavage sites recognized by the isolated RNase H domain were mapped. Five RNAs derived both from viral sequences and from *in vitro* transcription plasmids were used to generate RNA/DNA hybrids that were treated with the isolated RNase H domain or the intact M-
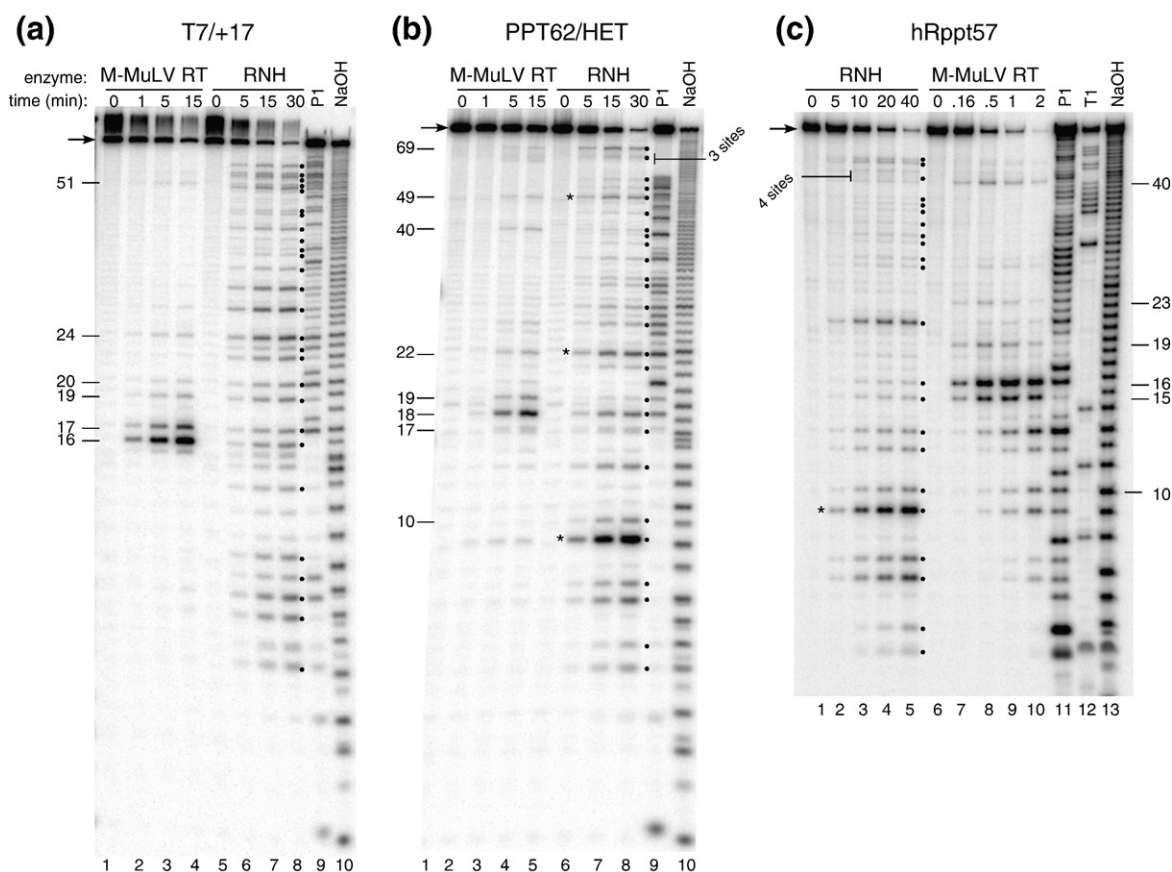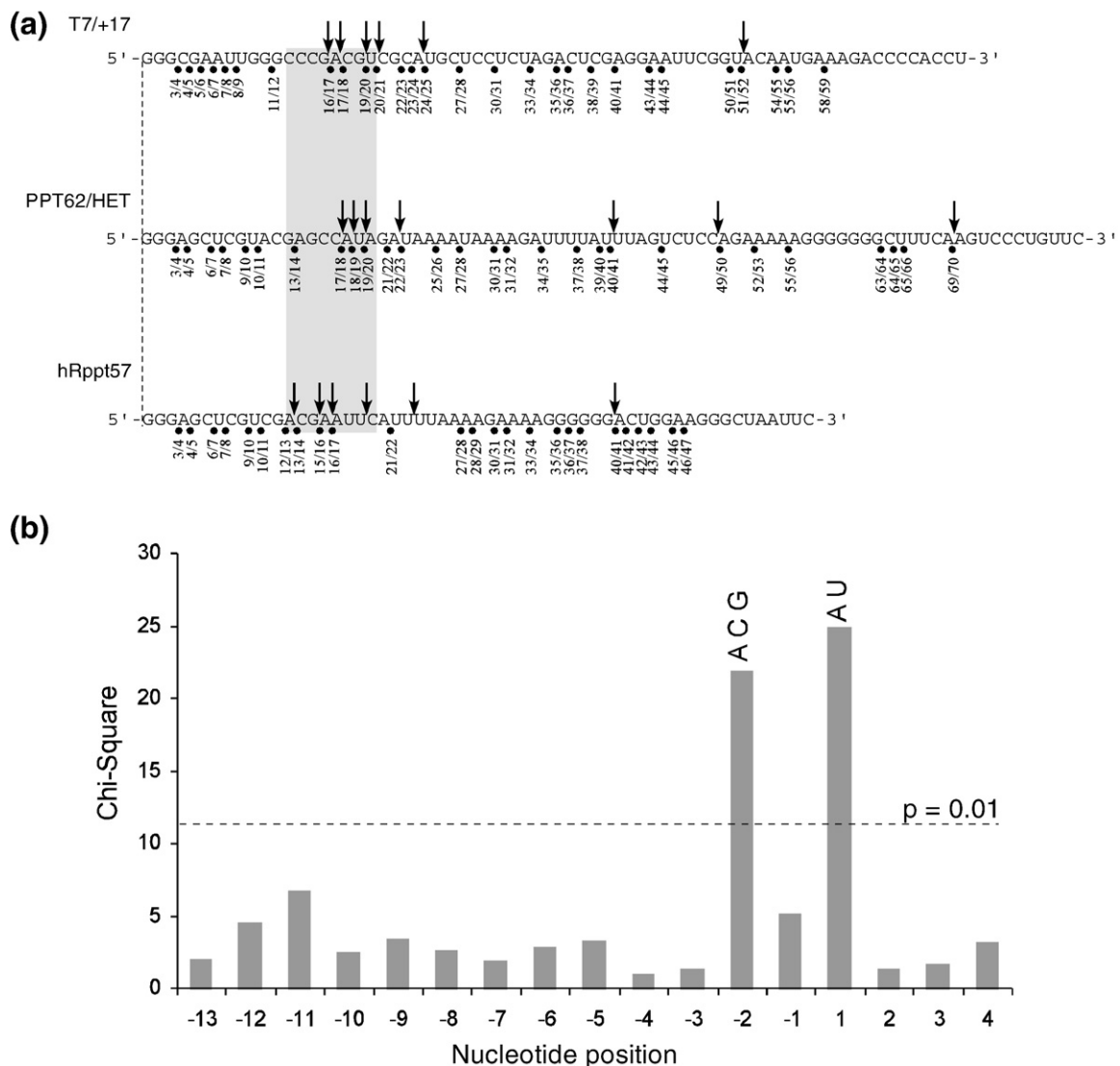

Fig. 8. Comparison of cleavage sites recognized by the isolated M-MuLV RNase H domain *versus* the M-MuLV reverse transcriptase. Substrates containing the indicated 5′-end-labeled RNAs were incubated with M-MuLV reverse transcriptase (M-MuLV RT) or the isolated RNase H domain for the indicated times, and samples were analyzed as described in Fig. 1. As markers, single-strand 5′-end-labeled RNAs were treated as indicated with nuclease P1 (P1), sodium hydroxide (NaOH), or ribonuclease T1 (T1) to enable a correlation of the cleavage sites with the RNA sequence. Numbers on the side indicate the size in nucleotides of products generated by M-MuLV RT cleavages, filled circles indicate products resulting from the isolated RNase domain cleavage, and asterisks indicate exceptionally strong cleavage sites that are discussed in the text. Representative experiments are shown with substrates containing RNAs T7/+17 (a), PPT62/HET (b), and hRppt57 (c).

MuLV reverse transcriptase (see Fig. 9a and Materials and Methods). With the use of the substrate containing RNA T7/+17, the isolated RNase H domain cleaved at multiple internal sites over the entire length of the RNA (Fig. 8a, lanes 5–8, cleavage products indicated by filled circles). Some products were clearly more abundant than others, suggesting some differences in cleavage preference by the isolated RNase H domain. In contrast, the intact M-MuLV reverse transcriptase preferentially cleaved sites (positions 16–20) found within the RNA 5′-end-directed cleavage window and recognized

only a few internal sites (Fig. 8a, lanes 1–4). Similar results were observed using substrates that contained RNAs PPT62/HET and hRppt57, as sites recognized by the RNase H domain were abundant and fairly evenly distributed, while those recognized by the M-MuLV reverse transcriptase were much less frequent (Fig. 8b and c).

The isolated RNase H domain strongly recognized some sites (Fig. 8, indicated with asterisks). One such site was between the 9th and 10th nucleotides from the RNA 5′ end of the PPT62/HET and hRppt57 RNAs, which share the same first 10 nt of

Fig. 9. Analysis of cleavage sites recognized by the isolated RNase H domain of M-MuLV reverse transcriptase. (a) The sequences of the RNAs used in the hybrids of Fig. 8 are shown with the cleavage sites recognized by the M-MuLV reverse transcriptase (arrows) and the isolated RNase H domain (filled circles). The RNA 5′-end-directed cleavage window is indicated by a shadowed box that falls from positions 13 through 20 in the aligned RNA 5′ ends. (b) Determination of the base preferences at nucleotide positions surrounding the RNase H cleavage sites. The chi-square values of the nucleotide distribution from positions -13 to $+4$ were determined by comparison with overall base frequencies and plotted as a function of nucleotide position relative to the scissile phosphate located between positions -1 and $+1$. This analysis is based on a total of 88 cleavage sites recognized by the isolated RNase H domain. The chi-square value of 11.34, which has a *p* value of 0.01, is indicated by a dashed line. Nucleotide positions with scores greater than 11.34 are considered significant. The preferred nucleotides for each significant position are indicated above the corresponding bar in uppercase (strongly preferred) letters. Counting from the RNA 5′ end, the position of each cleavage site is indicated below the RNA sequence by the 5′ and 3′ nucleotide positions that border the scissile phosphate.

5′ sequence due to *in vitro* transcription (Fig. 8b, lanes 6–8, and Fig. 8c, lanes 2–5). We interpret this cleavage to represent a sequence preference at this site rather than an end-directed cleavage preference since no exceptionally strong cleavage site was observed at the 5′ ends of other RNAs used in this analysis (Fig. 8a, T7/+17, asterisks, and other data not shown). Similar examples of equally strong cleavage sites were found between the 22nd and 23rd nucleotides and the 49th and 50th nucleotides in PPT62/HET (Fig. 8b, lanes 2–5).

In a comparison of the cleavage sites recognized by the M-MuLV reverse transcriptase *versus* the isolated RNase H domain, the intact enzyme recognized far fewer sites, and half of these were clustered in the RNA 5′-end-directed cleavage window (Fig. 9a). In contrast, the RNase H domain cleaved at approximately four times as many positions and the sites were distributed more evenly throughout the RNAs. Notably, the isolated RNase H domain recognized all sites cleaved by the intact M-MuLV reverse transcriptase, with the exception of sites 19/20 and 23/24 in the substrate containing RNA hRppt57. To determine if the isolated RNase H domain displayed any base preferences near a cleavage site, we aligned the sequences from positions −13 to +4 for 88 cleavage sites in several RNAs, including those in Fig. 8 (see Materials and Methods). The base distribution at each position was compared with the expected base distributions using the chi-square method, and the resulting chi-square values were plotted as a function of nucleotide positions (Fig. 9b), to determine the significance of any deviations from random nucleotide frequencies. For the isolated RNase H domain, two nucleotide positions showed strong deviations from random, with p values less than 0.01. The first was position +1, which showed a strong preference for A or U, and the second was position −2, which strongly disfavored U.

Discussion

In previous work with the HIV-1 and M-MuLV reverse transcriptases, a statistical analysis of sequence alignments from RNase H cleavage sites revealed that internal and end-directed cleavage sites were flanked by base preferences at specific positions.[23,29] For an internal cleavage site, HIV-1 RNase H had six positions with nucleotide preferences, while M-MuLV had four. However, this analysis could not discern the relative contribution of any single position to cleavage and additional preferences that were not revealed by the statistical study as a result of the particular distribution of bases in the chosen sequences could exist. In this study, a hybrid substrate containing a strong, isolated cleavage site recognized by each enzyme (site 30/31 for M-MuLV; site 33/34 for HIV-1) was used to ask if the preferred positions around a cleavage site are equally important or if the preferred positions have distinct effects on the extent of cleavage. Our results indicate that the nucleotide preferences at most positions contribute individually to cleavage-site recognition and that there are large variations in the extent that each position affects cleavage. Some positions have very pronounced effects on cleavage, while other positions have moderate or slight effects. These observations are summarized in Fig. 10 and discussed below.

Many of the preferred positions have a substantial and independent effect on retroviral RNase H cleavage (Fig. 10, black boxes). For HIV-1 RNase H, the previously identified preferences at positions +1 (A/U), −2 (C/G), −4 (C/G/U), and −14 (A/g) and the newly identified preferences at positions −5 (G/a) and −10 (A/U) significantly affected internal cleavage (Fig. 10). The introduction of a −2U, −2A, or −4A had particularly negative consequences on cleavage. The newly identified preferences at positions −5 and −10 were not observed in our earlier studies, perhaps because the nucleotides favored at these positions relative to the cleavage sites were not present at frequencies high enough in the RNA sequences used. For M-MuLV RNase H, three of the previously identified positions, +1 (A/U/c), −2 (A/C/G), and −6 (C/G/U), and a newly identified preference at position −7 (A) significantly affected internal cleavage. The loss of an A at position −7 or the introduction of a −2U or a −6A was especially deleterious for M-MuLV RNase H cleavage. Thus, many preferred positions around an RNase H cleavage site can individually have a large effect on the cleavage efficiency for both HIV-1 and M-MuLV enzymes.

Another class of preferred positions had only a moderate effect on cleavage when tested alone but a more substantial effect when combined with a
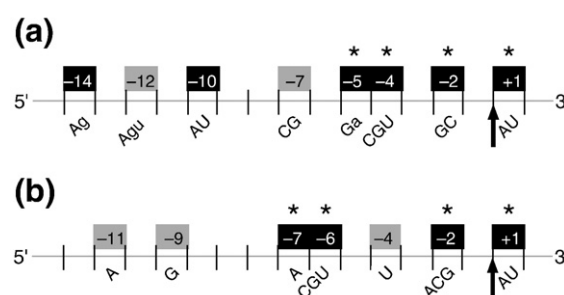
Fig. 10. Summary of base preferences at nucleotide positions surrounding an RNase H cleavage site recognized by HIV-1 or M-MuLV reverse transcriptase. The scissile phosphate (indicated by an arrow) is designated between the −1 and +1 positions below an RNA strand (gray) that would be in an RNA/DNA hybrid. The nucleotide positions that have a significant (black box) or moderate (gray box) effect on the efficiency of internal cleavage by HIV-1 RNase H (a) or M-MuLV RNase H (b) are shown with the positions numbered (above) and the base preferences (below) indicated. Positions important for DNA 3′-end-directed cleavage are indicated with an asterisk. Uppercase or lowercase letters indicate a strong or moderate preference, respectively, for specific nucleotides at the indicated positions.

second change (Fig. 10, gray boxes). For HIV-1 RNase H, positions −7 (C/G) and −12 (A/G/u) moderately influenced internal cleavage when a disfavored nucleotide was introduced, but cleavage was substantially reduced with the combined change of −7U/−14C or −12C/−14C. Similarly, for M-MuLV RNase H, only a moderate reduction in cleavage was observed when a disfavored C was introduced at the preferred position −11. This moderate effect was somewhat surprising because the preference for an A at position −11 had the highest statistical significance of any of the preferred positions previously shown for M-MuLV internal cleavage.[23] Also, for M-MuLV, position −9 (G) was not identified earlier as a preferred position, but the −9C change reduced cleavage to a level comparable with that of the −11C change, and the combined change −9C/−11C decreased cleavage more than either change independently. Another position with a moderate effect on cleavage for M-MuLV was position −4, since a −4U slightly increased cleavage and when combined with a −11C generated a cleavage level that was intermediate to either change alone. These observations suggest that positions −4, −9, and −11 for M-MuLV and positions −7 and −12 for HIV-1 have moderate effects on RNase H cleavage, and when combined with other weak or moderate preferences, the effects are additive.

Finally, there were some positions that had very modest effects on cleavage. Internal cleavage by HIV-1 RNase H had slight preferences at positions −3 (favors G) and −9 (disfavors C), while that by M-MuLV showed a slight preference at position +2 (favors C). Most likely, the bases favored at these positions contribute to an optimal combination for cleavage that is unique to the test cleavage sites. These slightly favored bases are not indicated in Fig. 10, and further experiments are required to establish whether such subtle effects are important in cleavage preferences.

DNA 3′-end-directed cleavage was significantly affected by positions +1, −2, −4, and −5 for HIV-1 and by positions +1, −2, −6, and −7 for M-MuLV (Fig. 10, asterisks). While the total amount of product generated by DNA 3′-end-directed cleavage was significantly greater than that generated by internal cleavage, most base preferences were the same for internal and DNA 3′-end-directed cleavages, and the effects of base substitutions were proportionally equivalent. For example, substitutions at position −2 affected both types of cleavage very similarly for HIV-1 and M-MuLV RNases H. Also, most single substitutions at positions more distal from the cleavage site did not significantly influence DNA 3′-end-directed cleavage. An interesting exception for HIV-1 was the single −14C change, which slightly reduced DNA 3′-end-directed cleavage, and when −14C was combined with a −7U or a −12C, end-directed cleavage was reduced further. This observation suggests that while one disfavored nucleotide at a position distal from the cleavage site appears unlikely to affect the DNA 3′-end-directed cleavage efficiency, two or more

disfavored nucleotides can significantly affect such cleavage.

Previous studies have shown that the distance between the recessed DNA 3′ end and the cleavage site is also an important determinant of how efficiently a site is cleaved.[26,33–36] DNA 3′-end-directed cleavage is greatest when this distance is ∼15–20 nt for HIV-1 RNase H and is ∼17–20 nt for M-MuLV RNase H.[28] In this study, we tested whether distance from the recessed primer terminus might influence the base preferences at several positions for M-MuLV RNase H. Despite severely reducing the cleavage efficiency by positioning the cleavage site at 15 nt from the DNA 3′ end, the base preferences for cleavage were essentially the same as when the cleavage site was positioned at 18 nt. We have observed similar results with other sites located at different distances from the recessed DNA end for both HIV-1 and M-MuLV RNases H (data not shown). These observations support the conclusion that the distance and sequence preferences of reverse transcriptase are independent determinants of RNase H cleavage.

In the co-crystal structure of HIV-1 reverse transcriptase with a PPT-containing RNA/DNA hybrid, contacts with the substrate appear to cluster in three regions of the protein.[12] First, there are residues near the RNase H active site that contact the RNA strand from positions +2 to −2, including base-specific contacts between Gln475 and the −2G in the substrate and between Arg448 and the +1A in the substrate. Second, there are multiple hydrogen-bonding contacts between the phosphate backbone in the DNA primer strand that is opposite to RNA positions −4 to −9 and a region of reverse transcriptase found close to the RNase H active site. This region has been termed the RNase H primer grip and is found in both HIV-1 and M-MuLV RNases H.[12,37] The RNase H primer grip influences substrate binding and positioning, and point mutations in this region of HIV-1 RNase H can decrease RNase H activity and alter recognition of the PPT.[38–42] Third, residues closer to the polymerase active site have multiple contacts (including base-specific interactions), with both strands of the substrate beginning at position −11 and continuing further upstream on the RNA strand relative to the potential cleavage site. For the preferred positions that promote RNase H cleavage, positions +1 and −2 would be near the RNase H active site, positions −4 through −7 would be in the RNase H primer grip region, and position −11 for M-MuLV and positions −12 and −14 for HIV-1 would be near the polymerase active site. At present, it remains ambiguous where position −9 or −10 would interact with either the M-MuLV reverse transcriptase or the HIV-1 reverse transcriptase.

While the isolated RNase H domain of HIV-1 is inactive,[43,44] the isolated M-MuLV RNase H domain retains enzymatic activity but is unable to carry out specific cleavages, such as PPT primer generation or the removal of the extended tRNA or PPT primers.[30,31,45] Here, we tested if the isolated M-MuLV RNase H domain cleaved randomly in an RNA/DNA hybrid or exhibited any sequence

preferences near a cleavage site. After the mapping of multiple cleavage sites recognized by the M-MuLV RNase H domain, an analysis of the base frequencies at the flanking nucleotide positions showed that the preferences of the full-length enzyme at positions -2 (A/C/G) and $+1$ (A/U) were retained. Thus, the isolated RNase H domain cleaves sites with limited selectivity based upon base preferences at positions -2 and $+1$. From the HIV-1 co-crystal structure,[12] these preferences likely reflect base-specific contacts with residues near the RNase H active site. The lack of preferences for nucleotides at the more distal positions is consistent with the absence of the connection and polymerase domains in RNase H.

It is informative to consider whether the sequence preferences identified for cleavage sites recognized by the retroviral RNases H extend to other members of the family of RNases H. Unlike retroviral reverse transcriptases, the prokaryotic and most eukaryotic RNases H typically lack long N-terminal or C-terminal extensions associated with the RNase H domain, and so far no sequence specificity has been identified near cleavage sites.[46,47] Thus, while interactions between the substrate and the protein in the vicinity of the active site can result in some sequence preferences for retroviral RNase H cleavage, the non-retroviral enzymes apparently lack this level of specificity. Higher eukaryotic type I RNases H typically have a short N-terminal hybrid binding domain (HBD) that promotes binding of hybrids in a non-sequence-specific manner and is associated with enhancing the specific activity and processivity of the enzyme.[47] Recent co-crystal structures of the HBD from human RNase H1 complexed with 12-bp RNA/DNA hybrids have revealed that the HBD interacts with the backbones of the hybrid substrate along the minor groove, contacting 2′-OH groups on the RNA strand and both phosphate groups and deoxyribose moieties on the DNA strand.[48] These non-specific contacts account for the enhancement of RNase H1 activity by the HBD and are consistent with the lack of any sequence specificity at cleavage sites. Thus, sequence preferences at specific nucleotide positions around a cleavage site may be a determinant that is unique to the retroviral RNases H.

We were able to create a cleavage site in an RNase H-resistant 17-nt sequence that contained two stretches of $(A)_4$ and one of $(U)_4$ (Site T, Fig. 7). An r(A):d(T) tract of four or more bases has distinct structural features, including a straight trajectory and a narrow minor groove,[49] and may influence the specificity of the PPT cleavage by RNase H.[12] In the RNase-H-resistant sequence, substitutions were introduced in the middle $(A)_4$ tract, which corresponded to positions -4 through -7 relative to Site T. Interestingly, very few changes were needed to promote cleavage at Site T. Only a single change of -4C for HIV-1 or a -6G for M-MuLV substantially increased DNA 3′-end-directed cleavage, and internal cleavage required two changes for each enzyme, a -4C/-5C for HIV-1 and a -6G/-7A for M-MuLV. Our findings are consistent with the possibility that

cleavage efficiency is affected by the substrate structure as determined by sequence. Also, the minimal changes required to introduce cleavage at a site within the RNase-H-resistant sequence are consistent with the base preferences presented in Fig. 10.

Reverse transcriptase is a crucial target in anti-retroviral therapy of HIV-1-infected patients. While the vast majority of clinical antivirals are currently specific for the polymerase domain, the RNase H activity of reverse transcriptase represents an excellent target for antiviral therapy as well.[17] Importantly, an anti-RNase H drug must be selectively directed against the viral activity without interfering with that of the endogenous human RNase H1. From the sequence preferences involved in cleavage-site selection by the retroviral RNases H, there may be several ways to interfere with proper viral RNase H activity. For example, drugs that interfere with the interactions between the HIV-1 enzyme and positions -4 to -7 or further upstream at positions -10 and -14 would be predicted to substantially interfere with retroviral RNase H activity but would be unlikely to affect the endogenous RNase H1.

Materials and Methods

Enzymes and reagents

DNA oligonucleotides were purchased from Eurofins MWG Operon, Invitrogen, and Bioneer, Inc. RNAs were made by *in vitro* transcription with a MEGAshortscript™ Kit from Ambion. Recombinant HIV-1 and M-MuLV reverse transcriptases were purchased from Worthington Biochemicals and Amersham Pharmacia Biotech, respectively. The recombinant RNase H domain of M-MuLV reverse transcriptase was prepared as described previously.[30] T7 DNA polymerase and calf intestinal alkaline phosphatase were obtained from USB/Affymetrix. To generate RNA ladders for mapping cleavage sites, we obtained nuclease P1 (cleaves at all positions) and RNase T_1 (cleaves 3′ of G positions) from New England Biolabs and BRL (now Invitrogen), respectively. Other DNA-modifying and restriction enzymes were purchased from New England Biolabs.

Preparation of RNAs

RNAs Rmzc and R$-54/-14$ were prepared by *in vitro* transcription of annealed DNAs as described previously.[28,29] For RNA Rmzc, a template DNA (5′-AGG-TGGGGTCTTTCA<u>TTGTA</u>CCGAATTCCTCGAGTCTAG-CAATTCGCCCTA<u>TAGTG</u>AGTCGTATTAAATT-3′) was annealed to a primer DNA (5′-AATTTAATACGACTCAC-TATAGG-3′). Site 33/34 (used for HIV-1; Fig. 1) and site 30/31 (used for M-MuLV; Fig. 4) are located between the 16th and 17th and the 19th and 20th nucleotides, respectively, from the template DNA 5′ end (underlined). For RNA R$-54/-14$, a template DNA (5′-GGAGACTA-AATAAA<u>A</u>TCTTTTAT<u>TTT</u>ATCTATGGCTCGTAC-GAGCTC<u>CC</u>TATAGTGAGTCGTATTAAATT-3′) was annealed to a primer DNA (5′-AATTTAATACGACTCAC-TATAGG-3′). Site T (Fig. 7) is located between the 15th and 16th nucleotides from the template DNA 5′ end

(underlined). Base substitutions were carried out for each site as described in Results by changing the appropriate bases in the template DNA. The 49mer RNA R–63/–23 was prepared as described previously.[28]

For RNA T7/+17, two 29mer DNAs (5′-GTACAATGA-AAGACCCCACCTCGAGACGA-3′ and 5′-AGCTTCGTCTCGAGGTGGGGTCTTTCATT-3′) were annealed and placed in Acc65I and HindIII-linearized pGEM7Zf(+), and the resulting plasmid was linearized with BsmBI and transcribed *in vitro*. For RNA hRppt57, two 50mer DNAs (5′-AGCTTCGTCTCCGAATTAGCCCTTC-CAGTCCCCCCTTTTCTTTTAAAATG-3′ and 5′-AATT-CATTTTAAAAGAAAAGGGGGGACTGGAAGGGC-TAATTCGGAGACGA-3′) were annealed and placed in HindIII and EcoRI-linearized pGEM9Zf(−), and the resulting plasmid was linearized with BsmBI for *in vitro* transcription. For RNA PPT62/HET, a 95-bp DNA was prepared by PCR of template 5[22] using a 19mer DNA (5′-TCGGGAGCTCGTACGAGCC-3′) and a 33mer DNA (5′-GGAATTCGTCTCGGAACAGGGACTTGAAAGCCC-3′). After digesting with EcoRI and SacI, the PCR product was introduced into EcoRI and SacI-linearized pGEM9Zf(−), and the resulting plasmid was linearized with BsmBI for *in vitro* transcription.

Preparation of hybrid substrates

RNAs were labeled at the 5′ or 3′ end and typically annealed to DNA strands using ratios of RNA/DNA at 1:1 to make internal cleavage substrates for Rmzc RNAs, at 1:10 to generate DNA 3′-end-directed cleavage substrates for R−54/−14 RNAs, and at 1:2 for all other RNAs as described previously.[28] Hybrids with RNAs containing base substitutions were prepared as described for the original RNA, as detailed below.

Internal cleavage assays with Rmzc used a 49mer DNA (5′-GTGGGGTCTTTCATTGTACCGAATTCCTCGA-GTCTAGCAATTCGCCCAT-3′), and the resulting hybrid 5′ ends were recessed by 2 nt. DNA 3′-end-directed cleavage assays with Rmzc used 30mer DNAs to position site 33/34 and site 30/31 at 18 and 15 nt, respectively, from the DNA 3′ end (5′-GGGGTCTTTCATTGTACCGAAT-TCCTCGAG-3′) or to position site 30/31 at 18 nt from the DNA 3′ end (5′-GTCTTTCATTGTACCGAATTCCTC-GAGTCT-3′). Internal cleavage assays with R−54/−14 similarly used a 49mer DNA (5′-AGACTAAATAAAAT-CTTTTATTTTATCTATGGCTCGTACGAGCTCCCTA-3′), and DNA 3′-end-directed cleavage assays used a 30mer DNA (5′-GACTAAATAAAATCTTTTATTTTATC-TATG-3′). Hybrids were generated with T7/+17 using a 70mer DNA (5′-GTGGGGTCTTTCATTGTACCGAA-TTCCTCGAGTCTAGAGGAGCATGCGACGTCG-GGCCCAATTCGCCCGA-3′), with PPT62/HET using template 5[22] and with hRppt57 using a 57mer DNA (5′-ATTAGCCCTTCCAGTCCCCCCTTTTCTTTTAA-AATGAATTCGTCGACGAGCTCCCTA-3′).

RNase H cleavage assays

A 10 nM concentration of hybrid substrates in 20-μl reactions was treated with 0.2 pmol HIV-1 reverse transcriptase for all assays and 6–10 or 0.6–1 pmol M-MuLV reverse transcriptase in internal or DNA 3′-end-directed assay, respectively, as described previously.[28] With the use of denaturing 20% polyacrylamide gels, cleavage products were visualized by PhosphorImager analysis and analyzed using ImageQuant software

(Molecular Dynamics).[28] Cleavage assays with the isolated RNase H domain of M-MuLV reverse transcriptase were carried out essentially as previously described.[30]

Statistical analysis for nucleotide preferences of the isolated RNase H domain

The chi-square one-dimensional test was used to calculate the deviation from the random distribution of bases at each position of the aligned sequences.[23,29] The expected frequency of each nucleotide was determined by summing the nucleotide frequencies in the unique sequences of Rmzc, T7/+17, PPT62/HET, hRppt57, and R+1/+29. The expected nucleotide frequencies were A = 0.286, C = 0.197, G = 0.296, and U = 0.220.

References

1. Arts, E. J. & LeGrice, S. F. J. (1998). Interaction of retroviral reverse transcriptase with template–primer duplexes during replication. *Prog. Nucleic Acid Res. Mol. Biol.* **58**, 339–393.
2. Telesnitsky, A. & Goff, S. P. (1993). Strong-stop strand transfer during reverse transcription. In *Reverse Transcriptase* (Skalka, A. M. & Goff, S. P., eds), pp. 49–83, Cold Spring Harbor Laboratory Press, Plainview, NY.
3. Tisdale, M., Schulze, T., Larder, B. A. & Moelling, K. (1991). Mutations within the RNase H domain of human immunodeficiency virus type 1 reverse transcriptase abolish virus infectivity. *J. Gen. Virol.* **72**, 59–66.
4. Tanese, N., Telesnitsky, A. & Goff, S. P. (1991). Abortive reverse transcription by mutants of Moloney murine leukemia virus deficient in the reverse transcriptase-associated RNase H function. *J. Virol.* **65**, 4387–4397.
5. Repaske, R., Hartley, J. W., Kavlick, M. F., O'Neill, R. R. & Austin, J. B. (1989). Inhibition of RNase H activity and viral replication by single mutations in the 3′ region of Moloney murine leukemia virus reverse transcriptase. *J. Virol.* **63**, 1460–1464.
6. Nikolenko, G. N., Svarovskaia, E. S., Delviks, K. A. & Pathak, V. K. (2004). Antiretroviral drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase increase template-switching frequency. *J. Virol.* **78**, 8761–8770.
7. di Marzo Veronese, F., Copeland, T. D., DeVico, A. L., Rahman, R., Oroszlan, S., Gallo, R. C. & Sarngadharan, M. G. (1986). Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. *Science*, **231**, 1289–1291.
8. Telesnitsky, A. & Goff, S. P. (1993). RNase H domain mutations affect the interaction between Moloney murine leukemia virus reverse transcriptase and its primer–template. *Proc. Natl Acad. Sci. USA*, **90**, 1276–1280.
9. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992). Crystal structure at 3.5 Å

resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science*, **256**, 1783–1790.

10. Ding, J., Das, K., Hsiou, Y., Sarafianos, S. G., Clark, A. D., Jr., Jacobo-Molina, A. *et al.* (1998). Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template–primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* **284**, 1095–1111.

11. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669–1675.

12. Sarafianos, S. G., Das, K., Tantillo, C., Clark, A. D., Jr., Ding, J., Whitcomb, J. M. *et al.* (2001). Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J.* **20**, 1449–1461.

13. Das, D. & Georgiadis, M. M. (2004). The crystal structure of the monomeric reverse transcriptase from Moloney murine leukemia virus. *Structure (Camb.)*, **12**, 819–829.

14. Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C. *et al.* (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.

15. Champoux, J. J. & Schultz, S. J. (2009). Ribonuclease H: properties, substrate specificity and roles in retroviral reverse transcription. *FEBS J.* **276**, 1506–1516.

16. Champoux, J. J. (1993). Roles of ribonuclease H in reverse transcription. In *Reverse Transcriptase* (Skalka, A. M. & Goff, S. P., eds), pp. 103–118, Cold Spring Harbor Press, Cold Spring Harbor, NY.

17. Klumpp, K. & Mirzadegan, T. (2006). Recent progress in the design of small molecule inhibitors of HIV RNase H. *Curr. Pharm. Des.* **12**, 1909–1922.

18. Schultz, S. J. & Champoux, J. J. (2008). RNase H activity: structure, specificity, and function in reverse transcription. *Virus Res.* **134**, 86–103.

19. Fuentes, G. M., Rodriguez-Rodriguez, L., Fay, P. J. & Bambara, R. A. (1995). Use of an oligoribonucleotide containing the polypurine tract sequence as a primer by HIV reverse transcriptase. *J. Biol. Chem.* **270**, 28169–28176.

20. Kelleher, C. D. & Champoux, J. J. (2000). RNA degradation and primer selection by Moloney murine leukemia virus reverse transcriptase contribute to the accuracy of plus strand initiation. *J. Biol. Chem.* **275**, 13061–13070.

21. Schultz, S. J., Zhang, M., Kelleher, C. D. & Champoux, J. J. (2000). Analysis of plus-strand primer selection, removal, and reutilization by retroviral reverse transcriptases. *J. Biol. Chem.* **275**, 32299–32309.

22. Schultz, S. J., Zhang, M. & Champoux, J. J. (2003). Specific cleavages by RNase H facilitate initiation of plus-strand RNA synthesis by Moloney murine leukemia virus. *J. Virol.* **77**, 5275–5285.

23. Schultz, S. J., Zhang, M. & Champoux, J. J. (2004). Recognition of internal cleavage sites by retroviral RNases H. *J. Mol. Biol.* **344**, 635–652.

24. DeStefano, J. J., Cristofaro, J. V., Derebail, S., Bohlayer, W. P. & Fitzgerald-Heath, M. J. (2001). Physical mapping of HIV reverse transcriptase to the 5′ end of RNA primers. *J. Biol. Chem.* **276**, 32515–32521.

25. DeStefano, J. J., Bambara, R. A. & Fay, P. J. (1993). Parameters that influence the binding of human immunodeficiency virus reverse transcriptase to nucleic acid structures. *Biochemistry*, **32**, 6908–6915.

26. DeStefano, J. J., Mallaber, L. M., Fay, P. J. & Bambara, R. A. (1993). Determinants of the RNase H cleavage specificity of human immunodeficiency virus reverse transcriptase. *Nucleic Acids Res.* **21**, 4330–4338.

27. Palaniappan, C., Fuentes, G. M., Rodriguez-Rodriguez, L., Fay, P. J. & Bambara, R. A. (1996). Helix structure and ends of RNA/DNA hybrids direct the cleavage specificity of HIV-1 reverse transcriptase RNase H. *J. Biol. Chem.* **271**, 2063–2070.

28. Schultz, S. J., Zhang, M. & Champoux, J. J. (2009). Preferred sequences within a defined cleavage window specify DNA 3′ end-directed cleavages by retroviral RNases H. *J. Biol. Chem.* **284**, 32225–32238.

29. Schultz, S. J., Zhang, M. & Champoux, J. J. (2006). Sequence, distance, and accessibility are determinants of 5′-end-directed cleavages by retroviral RNases H. *J. Biol. Chem.* **281**, 1943–1955.

30. Schultz, S. J. & Champoux, J. J. (1996). RNase H domain of Moloney murine leukemia virus reverse transcriptase retains activity but requires the polymerase domain for specificity. *J. Virol.* **70**, 8630–8638.

31. Zhan, X. & Crouch, R. J. (1997). The isolated RNase H domain of murine leukemia virus reverse transcriptase. Retention of activity with concomitant loss of specificity. *J. Biol. Chem.* **272**, 22023–22029.

32. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981). Nucleotide sequence of Moloney murine leukaemia virus. *Nature*, **293**, 543–548.

33. DeStefano, J. J. (1995). The orientation of binding of human immunodeficiency virus reverse transcriptase on nucleic acid hybrids. *Nucleic Acids Res.* **23**, 3901–3908.

34. Furfine, E. S. & Reardon, J. E. (1991). Reverse transcriptase.RNase H from the human immunodeficiency virus. Relationship of the DNA polymerase and RNA hydrolysis activities. *J. Biol. Chem.* **266**, 406–412.

35. Gopalakrishnan, V., Peliska, J. A. & Benkovic, S. J. (1992). Human immunodeficiency virus type 1 reverse transcriptase: spatial and temporal relationship between the polymerase and RNase H activities. *Proc. Natl Acad. Sci. USA*, **89**, 10763–10767.

36. Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. (1992). Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* **267**, 25988–25997.

37. Lim, D., Gregorio, G. G., Bingman, C., Martinez-Hackert, E., Hendrickson, W. A. & Goff, S. P. (2006). Crystal structure of the Moloney murine leukemia virus RNase H domain. *J. Virol.* **80**, 8379–8389.

38. Julias, J. G., McWilliams, M. J., Sarafianos, S. G., Arnold, E. & Hughes, S. H. (2002). Mutations in the RNase H domain of HIV-1 reverse transcriptase affect the initiation of DNA synthesis and the specificity of RNase H cleavage *in vivo*. *Proc. Natl Acad. Sci. USA*, **99**, 9515–9520.

39. Julias, J. G., McWilliams, M. J., Sarafianos, S. G., Alvord, W. G., Arnold, E. & Hughes, S. H. (2003). Mutation of amino acids in the connection domain of human immunodeficiency virus type 1 reverse transcriptase that contact the template–primer affects RNase H activity. *J. Virol.* **77**, 8548–8554.

40. Rausch, J. W., Lener, D., Miller, J. T., Julias, J. G., Hughes, S. H. & Le Grice, S. F. (2002). Altering the RNase H primer grip of human immunodeficiency virus reverse transcriptase modifies cleavage specificity. *Biochemistry*, **41**, 4856–4865.

41. McWilliams, M. J., Julias, J. G., Sarafianos, S. G., Alvord, W. G., Arnold, E. & Hughes, S. H. (2003).

Mutations in the 5′ end of the human immunodeficiency virus type 1 polypurine tract affect RNase H cleavage specificity and virus titer. *J. Virol.* **77**, 11150–11157.

42. McWilliams, M. J., Julias, J. G., Sarafianos, S. G., Alvord, W. G., Arnold, E. & Hughes, S. H. (2006). Combining mutations in HIV-1 reverse transcriptase with mutations in the HIV-1 polypurine tract affects RNase H cleavages involved in PPT utilization. *Virology*, **348**, 378–388.

43. Hostomsky, Z., Hostomska, Z., Hudson, G. O., Moomaw, E. W. & Nodes, B. R. (1991). Reconstitution *in vitro* of RNase H activity by using purified N-terminal and C-terminal domains of human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. USA*, **88**, 1148–1152.

44. Smith, J. S. & Roth, M. J. (1993). Purification and characterization of an active human immunodeficiency virus type 1 RNase H domain. *J. Virol.* **67**, 4037–4049.

45. Tanese, N. & Goff, S. P. (1988). Domain structure of the Moloney murine leukemia virus reverse transcriptase: mutational analysis and separate expression of the DNA polymerase and RNase H activities. *Proc. Natl Acad. Sci. USA*, **85**, 1777–1781.

46. Tadokoro, T. & Kanaya, S. (2009). Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J.* **276**, 1482–1493.

47. Cerritelli, S. M. & Crouch, R. J. (2009). Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* **276**, 1494–1505.

48. Nowotny, M., Gaidamakov, S. A., Ghirlando, R., Cerritelli, S. M., Crouch, R. J. & Yang, W. (2007). Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol. Cell*, **28**, 264–276.

49. Han, G. W., Kopka, M. L., Cascio, D., Grzeskowiak, K. & Dickerson, R. E. (1997). Structure of a DNA analog of the primer for HIV-1 RT second strand synthesis. *J. Mol. Biol.* **269**, 811–826.