Template Switches during Plus-Strand DNA Synthesis of Duck Hepatitis B Virus Are Influenced by the Base Composition of the Minus-Strand Terminal Redundancy

Jeffrey W. Habig and Daniel D. Loeb*

McArthur Laboratory for Cancer Research, University of Wisconsin Medical School, Madison, Wisconsin 53706

Received 22 July 2003/Accepted 25 August 2003

The hepadnaviruses are a family of DNA viruses constrained by a small genome size of ~3 kbp. Members of this family include the human pathogen hepatitis B virus and the avian duck hepatitis B virus (DHBV). DHBV has served as a useful model for understanding many aspects of hepadnavirus biology, including replication (6). Hepatitis B virus and DHBV share very little, if any, nucleotide sequence identity; yet, they maintain a very similar genomic organization and basic replication strategy (6). Their lack of nucleotide conservation is intriguing in light of the important contributions made by the cation strategy (6). Their lack of nucleotide conservation is important for both plus-strand template switches in DHBV.

Two template switches are necessary during plus-strand DNA synthesis of the relaxed circular (RC) form of the hepadnavirus genome. The 3' end of the minus-strand DNA makes important contributions to both of these template switches. It acts as the donor site for the first template switch, called primer translocation, and subsequently acts as the acceptor site for the second template switch, termed circularization. Circularization involves transfer of the nascent 3' end of the plus strand from the 5' end of the minus-strand DNA to the 3' end, where further elongation can lead to production of RC DNA. In duck hepatitis B virus (DHBV), a small terminal redundancy (5'r and 3'r) on the ends of the minus-strand DNA has been shown to be important, but not sufficient, for circularization. We investigated what contribution, if any, the base composition of the terminal redundancy made to the circularization process. Using a genetic approach, we found a strong positive correlation between the fraction of A and T residues within the terminal redundancy and the efficiency of the circularization process in those variants. Additionally, we found that the level of in situ priming increases, at the expense of primer translocation, as the fraction of A and T residues in the 3'r decreases. Thus, a terminal redundancy rich in A and T residues is important for both plus-strand template switches in DHBV.

The infection process leads to the formation of a genomic-length covalently closed circular (ccc) DNA species, which is maintained as a plasmid in the nucleus (29). The cccDNA species serves as the transcriptional template for production of all viral RNA species, including the pregenomic (pg) RNA (Fig. 1A). The pgRNA is a longer-than-genome-length mRNA, which is converted into either a relaxed circular (RC) or duplex linear (DL) form of the DNA genome via reverse transcription (Fig. 1B to I).

Synthesis of RC DNA requires two template switches during plus-strand replication, primer translocation (Fig. 1F) and circularization (Fig. 1H). There is accumulating evidence that these two processes share mechanisms. Interactions between two sets of distally located cis-acting elements in the minus-strand DNA (3E-M3 and 5E-M5) have been shown to contribute to both processes, presumably by base pairing to juxtapose the donor and acceptor sites for the two template switches (9, 15). Additionally, a small DNA hairpin near the 3' end of the minus-strand DNA has been shown to promote primer translocation by inhibiting in situ priming (8), as well as contributing to the circularization process (7). In this study, we provide evidence of another example of a cis-acting element that contributes to both plus-strand template switches. This cis-acting element is the small terminal redundancy (termed r) at the 5' and 3' ends (5'r and 3'r, respectively) of the full-length minus-strand DNA (Fig. 1E to I). It was previously shown that sequence identity between 5'r and 3'r is necessary, but not sufficient, for efficient circularization during DHBV replication (16). The aim of the present study was to understand why identity of r was not sufficient and to understand what additional contributions were made by the sequence in r to promote circularization.

Based upon sequence comparisons between the avian and mammalian hepadnaviruses and genetic evidence in DHBV, it has been hypothesized that the circularization process may be more efficient in the presence of a terminal redundancy with an increased fraction of A and T residues (16, 23). In this report, we present evidence in support of this hypothesis by correlating the ability of DHBV to perform the circularization process with the fraction of A and T residues in the sequence of r. Additionally, we present evidence that primer translocation occurs more readily, at the expense of in situ priming, when the fraction of A and T residues in the sequence of r is high. We discuss the implications of these findings to the processes of primer translocation and circularization.
FIG. 1. Model for transcription and replication of DHBV. (A) cccDNA or in vitro-synthesized circular DNA monomer (gray circle) serves as the transcriptional template for synthesis of the terminally redundant pgRNA (black line). The genomic and pgRNA sequence that serves as the template for r is denoted, as are the locations of DR1 and DR2 (boxes). The pgRNA is capped and polyadenylated. (B) The pgRNA serves as a template for synthesis of minus-strand DNA (27). The P protein (oval) binds to the stem-loop structure near the 5′ end of the pgRNA to facilitate encapsidation and protein-primed initiation of minus-strand DNA synthesis (31). The first 4 nt of minus-strand DNA and the 5′ r are synthesized using the bulge of the stem-loop as template (thin vertical lines represent base pairing) (20, 28, 30). The 12-nt direct repeats, DR1 and DR2 (indicated by boxes), are shown with DR1 present twice in the pgRNA. (C) Upon switching templates, minus-strand DNA synthesis (thick black line) continues from a region, complementary to the nascent minus-strand, overlapping the 3′ copy of DR1 (28, 30). (D and E) During minus-strand DNA synthesis, the RNase H activity of the P protein degrades the pgRNA (27) (D). Minus-strand DNA synthesis continues to the 5′ end of the pgRNA, where the final RNase H cleavage creates a capped 18- or 19-nt oligoribonucleotide that serves as the primer for plus-strand DNA synthesis (13, 17, 26). The 3′ end of the primer lies within the DR1 sequence. The full-length minus-strand DNA is shown as a solid black line, with the P protein (circle) covalently linked to the 5′ end. (F) The first template switch of plus-strand DNA replication, primer translocation, involves moving at least the 3′ end of the primer from the donor site (DR1) to the distally located acceptor site (DR2), where initiation of plus-strand DNA synthesis occurs (13). (G) DNA synthesis proceeds to the 5′ end of the minus-strand DNA template, synthesizing approximately 50 nt of plus-strand DNA. (H) The second template switch, circularization, permits the 3′ end of the nascent plus-strand DNA to anneal to the 3′ end of the minus-strand DNA. This template switch is facilitated, in part, by a small terminal redundancy of 7 or 8 nt (5). (I) A small fraction of plus-strand DNA synthesis (ca. 5%) initiates from DR1 rather than DR2, a process termed in situ priming. Plus strands initiating from DR1 result in a DL form of the viral genome (26). In situ priming is inhibited, in part, by a small DNA hairpin overlapping the 5′ end of DR1 (8).

MATERIALS AND METHODS

Molecular clones. All molecular clones were derived from DHBV type 3 (25). A population of primers consisting of either five random nucleotides or five directed nucleotides (TTCTT and TCCGG variants) covering the coordinates 2529 to 2933 was used to create molecular clones. These primers, along with two additional primers, one upstream of the 5′ end of the primer lies within the DR1 sequence. The full-length minus-strand DNA is shown as a solid black line, with the P protein (circle) covalently linked to the 5′ end. (F) The first template switch of plus-strand DNA replication, primer translocation, involves moving at least the 3′ end of the primer from the donor site (DR1) to the distally located acceptor site (DR2), where initiation of plus-strand DNA synthesis occurs (13). (G) DNA synthesis proceeds to the 5′ end of the minus-strand DNA template, synthesizing approximately 50 nt of plus-strand DNA. (H) The second template switch, circularization, permits the 3′ end of the nascent plus-strand DNA to anneal to the 3′ end of the minus-strand DNA. This template switch is facilitated, in part, by a small terminal redundancy of 7 or 8 nt (5). (I) A small fraction of plus-strand DNA synthesis (ca. 5%) initiates from DR1 rather than DR2, a process termed in situ priming. Plus strands initiating from DR1 result in a DL form of the viral genome (26). In situ priming is inhibited, in part, by a small DNA hairpin overlapping the 5′ end of DR1 (8).

Generation of circular monomers for transfection. The pD3 and p813-1 plasmids were generated by inserting a full-length genomic monomer into the EcoRI site of the plasmid. Circular genomic-length monomers, akin to cccDNA, were generated for each variant independently prior to transfection. Each plasmid (10 μg) was digested at 37°C overnight with EcoRI in the recommended buffer. The restriction enzyme activity was inactivated by treatment at 65°C for 20 min. Digestions were diluted to large volumes (4 ml) to select for intramolecular (self-ligation) reactions using T4 DNA ligase (NEB) at room temperature overnight. The DNA ligase activity was then inactivated by treatment at 65°C for 20 min. Centrifugation 100 filters were used to concentrate the DNA from 4 ml to ≈100 μl (following the manufacturer’s recommendations; 1,000 × g). A buffer exchange was performed by adding 2 ml of H2O and reconcentrating to a final volume of 50 to 200 μl. The concentrated circular monomers were used for transient transfection of LMH cells. Analysis using ethidium bromide-stained agarose gels indicated the ligations were successful.

Cell culture and isolation of viral DNA. The chicken hepatoma cell line LMH was used to replicate DHBV (3, 11). Cells were cultured as previously described (19). Calcium phosphate transfections were performed. Viral DNA was isolated from cytoplasmic capsids 3 days posttransfection as previously described (8).
RNA primer extension analyses. Isolation of poly(A) RNA was performed as previously described (18). Using RNase-free conditions, approximately one-fifth to one-half of the total fraction of poly(A) isolated RNA was used in each primer extension reaction. The minus-sense DNA primer used in these reactions hybridized to coordinates 2700 to 2678. The sample was ethanol precipitated, washed in 70% ethanol, dried, and resuspended in 13 μl (2 μl of 5% avian myeloblastosis virus [AMV] reverse transcriptase [RT] buffer, 1 pmol of end-labeled primer). The mixture was denatured at 80°C for 3 min, followed by incubation at 45°C for 15 min to allow primer-template annealing. Primer extension was performed for 90 min at 45°C in a 30-μl reaction volume (1× AMV RT buffer [Promega], deoxynucleoside triphosphates, RNasin [Promega], and AMV RT [Promega]). Next, alkaline hydrolysis (0.2 N NaOH and 25 mM EDTA) was performed at 95°C for 15 min followed by incubation at 65°C for 15 min. Samples were neutralized with Tris-HCl, ethanol precipitated, washed in 70% ethanol, dried, and resuspended in 2.5 μl of loading dye. Samples were heated denatured and electrophoresed through a 5% polyacrylamide gel containing 7.6 M urea at 50 W for ~3 h. Gels were dried, exposed, and visualized using Molecular Dynamics phosphorimaging cassettes and the Storm PhosphorImager.

DNA primer extension analyses. Primer extension was performed essentially as previously described (8). The use of three primers enabled us to measure four events during replication. The primer used to detect plus-strand DNA that had initiated at DR2 and elongated to the 5’ end of the minus-strand DNA or beyond (total DR2 priming) had sequence complementarity to nucleotide coordinates 2528 to 2511 (primer 1; annealing temperature, 37°C). A primer with sequence complementarity to nucleotide coordinates 2629 to 2598 (primer 2; annealing temperature, 58°C) measured two distinct events: the relative amount of plus-strand DNA that had primed and elongated from direct repeat 1 (DR1; in situ priming) and the fraction of the plus-strand DNA detected by primer 1 that had circularized and elongated ~90 nt (postcircularization). A primer with sequence complementarity to nucleotide coordinates 2425 to 2447 (primer 3; annealing temperature, 58°C) was used to measure the 5’ termini of minus-strand DNA that had synthesized at least 90 nt. An internal standard (i.s.) was added to each viral DNA sample to account for variation between the independent primer extension reactions.

Primer extension reactions mapping the 5’ termini of minus-strand DNA (Fig. 2) were performed in the absence of an i.s. and without alkaline treatment of the viral DNA. Reactions were performed using either primer 3 or primer 4 (sequence complementarity to nucleotide coordinates 2417 to 2433), both of which anneal to the minus-strand DNA. Visualization and quantification were performed with Molecular Dynamics phosphorimaging cassettes and the Storm PhosphorImager.

Southern blotting analyses. Southern blotting was performed as previously described for DNA replication intermediates using 1.25% agarose gels containing 1× Tris-borate-EDTA (2). Blots were probed with genomic-length, plus-strand RNA probes. Visualization and quantification was performed with Molecular Dynamics phosphorimaging cassettes and the Storm PhosphorImager.

Calculations. Southern blotting was used to measure DL DNA and RC DNA synthesis relative to the full-length, minus-strand DNA, as follows: RC DNA synthesis = 100 × (RC DNA/total full-length minus-strand DNA); DL DNA synthesis = 100 × (DL DNA/total full-length minus-strand DNA). The quantitative primer extension assay depicted below in Fig. 4 was used to measure the relative efficiencies of the template switches for each variant: circularization efficiency = 100 × [(postcircularization/i.s.)/(total priming from DR2/i.s.)]; in situ priming = 100 × ([(priming from DR1/i.s.)/(minus-strand DNA 5’ termini/i.s.)]; primer utilization = 100 × ([(priming from DR1/i.s.) + (total priming from DR2/i.s.)/(minus-strand DNA 5’ termini/i.s.)]; primer translocation = 100 × ([(total priming from DR2/i.s.)/(minus-strand DNA 5’ termini/i.s.)]. All measurements were normalized to a wild-type reference sample (wild type = 1 for in situ priming; wild type = 100 for all other measurements).

Statistical analyses. A two-sided Kendall’s rank correlation test was used to test for a correlation between the fraction of residues in the mutated sequence of r and either primer translocation, circularization, in situ priming, or RC DNA synthesis. A one-sided test was used to test for a negative correlation between in situ priming and primer translocation. The test statistic for the Kendall rank correlation (τa) was used to compare the strength of the correlations between A and T residues and A residues alone. All statistical analyses were done using Mstat version 3.21 (kindly provided by N. Drinkwater, McArdle Laboratory for Cancer Research).

RESULTS

Hypothesis, rationale, and experimental design. To test the hypothesis that the circularization process occurs more efficiently in the presence of a minus-strand DNA terminal re-
dundancy with a high fraction of A and T residues rather than G and C residues, variants of DHBV were made with mutations in the sequence of r and tested for their ability to perform the circularization process. Four of the nucleotides within 3’r (coordinates 2534 to 2537) were not easily amenable to mutation, as they served as the acceptor site for the minus-strand template switch (Fig. 1C). Altering those nucleotides may lead to reduced or aberrant template switches, even if compensatory changes were introduced into epsilon (Fig. 1B) (19). Therefore, mutagenesis was limited to the five adjacent nucleotides from coordinate 2529 through 2533 in both the 5’r and 3’r. The molecular variants used in this study are described according to their unique sequences spanning these nucleotides (plus-strand polarity), where wild type is AAGAA.

Mutations were introduced into a plasmid containing a monomer (1.0-mer) of the DHBV genome, which is unable to support replication upon transfection into LMH cell cultures. Instead, circular monomers were generated for transient transfection of LMH cell cultures by excising the DHBV insert from their respective plasmids followed by a treatment designed to favor intramolecular ligations and, thus, the formation of DNA circles. Upon transfection, the circular monomers expressed pgRNA, similar to native cccDNA, by bypassing the polyadenylation signal on the first pass, allowing for the synthesis of a terminally redundant (termed R) pgRNA (Fig. 1A). The consequence was that mutations introduced into the genomic sequence, serving as template for the terminally redundant portion of the pgRNA, were synthesized into both copies. Included in the terminally redundant portion of the pgRNA are the sequences that serve as the templates for the terminal redundancies of the minus-strand DNA. Thus, this expression strategy ensured that all of our variants had sequence identity between the 5’r and 3’r sequences.

Altered 5’ termini of minus-strand DNA are a rare event following changes to the sequence of r. The 5’ end of the minus-strand DNA is synthesized using two templates: epsilon and the sequence adjacent to the minus-strand acceptor site. Following synthesis of the first 4 nt of minus-strand DNA using epsilon (Fig. 1B), the minus-strand template switch transfers the nascent minus strand to the acceptor site, which overlaps the 5’ end of DR1 near the 3’ end of the pgRNA, prior to elongation (Fig. 1C). Previously, variants were shown to contain new 5’ minus-strand DNA termini upon deletion of the 3’ copy of DR1, which results in a loss of the minus-strand DNA acceptor site (4). Because the mutations introduced into the r sequence are adjacent to the minus-strand acceptor site, we did not want to assume that the accuracy of the minus-strand template switch was unaltered by our mutations. If new acceptor sites were unmasked, it would result in minus-strand DNA with unique 5’ termini and, thus, minus-strand DNA with an altered terminal redundancy.

To verify the integrity of the position of minus-strand DNA 5’ termini, two methods were used: primer extension and Southern blotting. Both analyses were performed using viral DNA isolated from transiently transfected LMH cells. The 5’ termini of the minus-strand DNA were initially investigated using primer extension with primer 3, which hybridized 90 nt from the 5’ end of minus-strand DNA and 30 nt from DR2 (see Fig. 4B). As the circularization process requires plus-strand priming from DR2, minus strands with 5’ termini not detected by this primer are unlikely to influence the circularization process. Primer extension on viral DNA isolated from cells transiently transfected with the wild-type circular monomer had 5’ termini that mapped to coordinate 2537, as expected (Fig. 2A, lane 1). In general, the primer extension products on viral DNA isolated from the variants with changes in the r sequence indicated the 5’ termini of their minus-strand DNA mapped to the same position as the wild-type reference (Fig. 2A). However, a faint band was occasionally detected at coordinate 2576, which corresponds to the UUAC motif within the 3’ copy of the epsilon sequence. As shown in lane 15, primer extension on one variant, GGTGC, led to a significant increase in the intensity of the band at coordinate 2576. Additionally, primer extension on viral DNA from two other variants (GAGTT and AGGGT) also revealed bands consistent with alternate 5’ termini using primer 3 (data not shown). The primer extension reactions for the GAGTT and AGGGT variants were repeated using a different primer (primer 4) to ensure the products were independent of the primer. As shown in Fig. 2B, no alternate 5’ termini were detected in any of the wild-type reference samples from independent transfections (lanes 1, 4, 7, 10, 12, and 15). In addition to coordinate 2537, 5’ termini for the AGGGT variants were mapped to coordinates 2576, 2644 and, inconsistently, 2587 (Fig. 2B, lanes 6, 9, 14, and 16). Interestingly, bands at coordinates 2576 and 2644 were also detected in a variant containing a deletion of the acceptor site for the minus-strand template switch (4). Conversely, the GAGTT variant had 5’ minus-strand termini that mapped to coordinate 2603, in addition to coordinate 2537 (Fig. 2B, lanes 2, 3, 5, 8, 11, and 13). There is some uncertainty of the precise 5’ coordinate of these extra bands, as their migration appeared to vary slightly in some gels (data not shown). The mechanism leading to these minus-strand DNA species with altered 5’ ends is unclear and was not pursued further. However, the occurrence of alternate 5’ termini in the GAGTT, AGGGT, and GGTGC variants led to their exclusion from the analysis.

To further evaluate the integrity of the minus-strand DNA, Southern blotting was performed to look for the accumulation of shorter-than-genome-length, single-stranded DNA species that would not have been detected in the primer extension analyses. This was primarily a concern for analyses of in situ priming levels (see below), as a minus-strand DNA that does not contain DR2 may lead to increased in situ priming. However, the results indicated that aberrant minus-strand DNA species were not apparent for any of the variants (see Fig. 5C). In summary, all but three of the variants tested appeared to have the correct 5’ termini and, thus, the appropriate 5’r sequence in the minus-strand DNA.

Changes in the sequence of r can affect the transcriptional initiation site for pgRNA. The 3’r sequence is generated by elongation of the minus-strand DNA to the 5’ end of the pgRNA (Fig. 1D and E). Thus, mutations that alter the transcriptional start site would change the size of the terminal redundancy (r) in the minus-strand DNA by altering the 5’ end of the pgRNA template. The 5’ end of the pgRNA for each variant was compared with the wild-type reference in a primer extension analysis on poly(A)-selected RNA isolated from transiently transfected LMH cells. In this analysis, the P protein expression plasmid was not included in the transfection.
The primer used in these reactions had complementarity to coordinates 2700 to 2678, which lie within the redundant region (R) of the pgRNA. Primer extension on RNA isolated from cells transfected with the wild-type circular monomer produced a doublet at coordinates 2529 and 2530 (Fig. 3, lane 1), consistent with previous reports that map the 5′ end of the DHBV pgRNA (1, 13, 22). This validated the use of circular monomers as transcriptional templates for pgRNA synthesis. Next, the primer extension products for each of the variants were compared with those of a wild-type reference. In most cases, similar band patterns were obtained with the variants and the wild-type reference, although there were exceptions. Primer extension of some of the variants resulted in reduced heterogeneity in the doublet (Fig. 3, lanes 2 and 4). These variants were not removed from further analyses based solely upon this factor. However, variants with primer extension products repeatedly containing additional bands, such as the TTCTT variant (Fig. 3, lane 3), were excluded from the statistical analyses that follow. In total, nine variants were excluded based upon the results of this assay.

The circularization process is more efficient with a terminal redundant sequence containing a higher fraction of A and T residues. After validating that the minus-strand terminal redundancies had the correct size, the remaining variants were analyzed for their ability to perform the circularization process. DNA replication intermediates were isolated from LMH cells 3 days posttransfection. A quantitative primer extension analysis was used to measure the circularization efficiency for each variant (Fig. 4A). Primer 1 was used to measure the total amount of plus-strand DNA that had initiated from DR2 and elongated to the 5′ end of the minus-strand DNA template or beyond (total DR2 priming) (Fig. 5A). By detecting total plus-strand DNA primed from DR2 using a primer that anneals to the location just prior to circularization, the analysis was not influenced by defects at prior steps in replication. Primer 2 was used to measure the fraction of plus-strand DNA that initiated from DR2, circularized, and elongated at least 90 nt (Fig. 5A, postcirc). All variants were normalized to the wild-type reference virus within a given transfection. Circularization efficiency was calculated as described in Materials and Methods.

Two specific mutations were introduced into the sequence of r, TTCTT and TCCGG. The TTCTT variant was the complementary sequence of the wild-type r. Based upon the hypothesis, it was predicted that this variant would circularize with high efficiency, akin to wild type. Similarly, the TCCGG variant was the complementary sequence of the AGGCC variant previously shown to be defective for circularization (see the 326 variant in reference 16). Thus, the TCCGG variant was also predicted to be defective for circularization. Results indicated the TCCGG variant was defective for circularization (55% of the wild-type reference), although it circularized about twofold better than the AGGCC variant (Table 1). The circularization efficiency of the TTCTT variant was calculated as ~80% of the wild-type reference. However, interpretation of the circularization phenotype for this variant was confounded by the fact that the transcriptional start site appeared to be heterogeneous (Fig. 3, lane 3).

A second, more general mutagenesis approach involved generating DHBV variants with randomly introduced sequence in
r (coordinates 2529 to 2533). The circularization efficiency for each variant was then measured using the primer extension assay. The measurements for the variants that were not excluded from the analysis based on altered ends of minus-strand DNA are listed in Table 1. Next, the mean circularization efficiencies for the variants were determined as a function of the fraction of A and T residues within the mutated sequence of r chosen for mutagenesis (Fig. 5B). A statistical correlation existed (P < 0.003) between circularization efficiency and the total number of A and T residues within the 5 nt of r chosen for mutagenesis (Fig. 5B). As circularization is a prerequisite for synthesis of RC DNA, Southern blotting (Fig. 5C) was used to corroborate the primer extension results. The proportion of RC DNA synthesized by each variant was measured as described in Materials and Methods. A statistical correlation (P < 0.0003) was also detected between the RC DNA levels and the number of A and T residues in the mutated region of r (Fig. 5D). Thus, there was a positive correlation between the fraction of A and T residues in r and the ability to both perform the circularization process and synthesize RC DNA.

Although the results strongly supported the hypothesis, further statistical tests were done to examine other possibilities. First, the possibility that a correlation existed between circularization efficiency and the purine or pyrimidine content of the sequence in r was considered. This possibility was not supported using a two-sided test (P > 0.14), nor was the pos-
sibility of a correlation between circularization efficiency and either A and C residues or G and T residues ($P > 0.16$). Next, the correlation between A and T residue composition and circularization efficiency was separated into either A or T residues. A significant $P$ value was obtained for the A residues alone ($P < 0.004$), but not for the T residues alone ($P > 0.7$). A correlation was found between either the G or C residues and circularization ($P < 0.04$ for each), although in both cases the correlation was negative. Finally, when the data set was ordered and classified into groups solely by differences in their circularization phenotype, no clear alternative relationship emerged, although it is possible the data set was not large enough to be informative for this type of analysis (data not shown).

A high fraction of A and T residues in the sequence of $r$ promotes primer translocation over in situ priming. Southern blotting of the DNA replication intermediates indicated that many of the variants with mutations in the sequence of $r$ not only affected RC DNA synthesis but also affected the amount of DL DNA synthesis, which is the end product of in situ priming (Fig. 5C). Thus, the mutations introduced into the sequence of $r$ were able to influence the discrimination between in situ priming (Fig. 1F) and primer translocation (Fig. 1G). The contribution of the $r$ sequence to this discrimination is most likely attributed to the interaction of the 3' end of the minus-strand DNA and the RNA primer (Fig. 1E). Sequence changes in the 5' of $r$ did not appear to contribute to the discrimination, as introduction of the AGGCC sequence into either the 3', the 5', or both led to increased DL DNA only when the 3' sequence was changed (16). We hypothesized that the increase in the level of in situ priming was related to the increase in the fraction of G and C residues in the mutated sequence, perhaps leading to increased stability between the RNA primer and DR1. The proportion of DL DNA measured by Southern blotting was determined for the variants. There was a marginal statistical correlation between the level of DL DNA synthesis and the fraction of A and T residues ($P < 0.06$; data not shown).

The quantitative primer extension assay (Fig. 4B) was used as a second assay to measure in situ priming (Fig. 5A, primer 2 image; results with primer 3 not shown). Primer 2 measured the amount of plus-strand DNA that initiated from DR1, while primer 3 measured the amount of minus-strand DNA 5' termini. In situ priming for each variant (Table 1) was calculated as described in Materials and Methods. Using this assay, a strong correlation was detected ($P < 0.003$) between the level of in situ priming and the fraction of A and T residues in the mutated region of $r$ (Fig. 6A). A strong correlation ($P < 0.008$) was also detected between primer translocation and the fraction of A and T residues in the mutated region of $r$ (Fig. 6B). In fact, there was a negative correlation between in situ priming and primer translocation ($P < 0.008$; one-sided analysis). Combined, these findings indicate that when the fraction of G

### Table 1. Replication measurements for each $r$ variant

<table>
<thead>
<tr>
<th>Variant</th>
<th>RC DNA synthesis$^a$</th>
<th>Circularization$^a$</th>
<th>In situ priming$^b$</th>
<th>Primer utilization$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AATAA</td>
<td>110 ± 9</td>
<td>91 ± 13</td>
<td>0.6 ± 0.05</td>
<td>115 ± 12</td>
</tr>
<tr>
<td>AAGAA</td>
<td>100</td>
<td>100</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>GAACA</td>
<td>73 ± 6</td>
<td>90 ± 9</td>
<td>1.5 ± 0.3</td>
<td>88 ± 13</td>
</tr>
<tr>
<td>GACA</td>
<td>77 ± 9</td>
<td>106 ± 18</td>
<td>3.1 ± 0.8</td>
<td>75 ± 5</td>
</tr>
<tr>
<td>AGACA</td>
<td>73 ± 5</td>
<td>96 ± 5</td>
<td>1.4 ± 0.1</td>
<td>75 ± 1</td>
</tr>
<tr>
<td>AGGAA</td>
<td>72 ± 6</td>
<td>64 ± 9</td>
<td>2.6 ± 0.7</td>
<td>106 ± 19</td>
</tr>
<tr>
<td>AGAGT</td>
<td>83 ± 16</td>
<td>70 ± 10</td>
<td>1.9 ± 0.5</td>
<td>126 ± 22</td>
</tr>
<tr>
<td>TGGGTT</td>
<td>35 ± 7</td>
<td>55 ± 7</td>
<td>3.0 ± 0.2</td>
<td>92 ± 26</td>
</tr>
<tr>
<td>AGGCT</td>
<td>54 ± 6</td>
<td>54 ± 7</td>
<td>2.2 ± 0.2</td>
<td>100 ± 11</td>
</tr>
<tr>
<td>AGACC</td>
<td>55 ± 6</td>
<td>86 ± 8</td>
<td>2.6 ± 0.4</td>
<td>81 ± 9</td>
</tr>
<tr>
<td>AGGTC</td>
<td>47 ± 4</td>
<td>54 ± 8</td>
<td>3.7 ± 0.9</td>
<td>105 ± 11</td>
</tr>
<tr>
<td>TGTCTG</td>
<td>61 ± 8</td>
<td>87 ± 21</td>
<td>2.5 ± 0.3</td>
<td>73 ± 13</td>
</tr>
<tr>
<td>TGGAC</td>
<td>45 ± 4</td>
<td>64 ± 8</td>
<td>2.8 ± 0.5</td>
<td>70 ± 10</td>
</tr>
<tr>
<td>TGGGA</td>
<td>20 ± 2</td>
<td>38 ± 8</td>
<td>2.7 ± 0.6</td>
<td>52 ± 6</td>
</tr>
<tr>
<td>GGCAA</td>
<td>28 ± 4</td>
<td>62 ± 4</td>
<td>3.1 ± 0.2</td>
<td>51 ± 3</td>
</tr>
<tr>
<td>CGGAC</td>
<td>39 ± 6</td>
<td>61 ± 3</td>
<td>3.5 ± 0.6</td>
<td>66 ± 5</td>
</tr>
<tr>
<td>TCCGG</td>
<td>26 ± 7</td>
<td>55 ± 8</td>
<td>4.9 ± 1.6</td>
<td>49 ± 12</td>
</tr>
<tr>
<td>AGGCC</td>
<td>14 ± 4</td>
<td>27 ± 7</td>
<td>3.5 ± 0.6</td>
<td>68 ± 24</td>
</tr>
<tr>
<td>GGCC</td>
<td>15 ± 11</td>
<td>27 ± 4</td>
<td>3.7 ± 0.3</td>
<td>46 ± 12</td>
</tr>
</tbody>
</table>

$^a$ Variants are ordered according to the number of A and T residues. Calculations were performed as described in Materials and Methods. The values for wild type (AAGAA) were set to either 100 (RC DNA synthesis, circularization, and primer utilization) or 1 (in situ priming). Values reported are the means ± standard deviations.

$^b$ Measured using primer extension.

$^c$ Measured using Southern blotting.

**FIG. 6.** Correlation between the fraction of A and T residues in the sequence of r and in situ priming and primer translocation. (A) Graphical analysis of the negative correlation between the fraction of A and T residues in the sequence of r and the level of in situ priming of those variants ($P < 0.003$) as determined by primer extension. (B) Graphical analysis of the correlation between the fraction of A and T residues in the sequence of r and primer translocation of those variants ($P < 0.0003$) as determined by primer extension. Values represented in panels A and B are the means of independent analyses ($n = 2$ to 7) for each variant after normalization to a wild-type reference sample. Table 1 reports the measurements for each variant.
and C residues within the mutated sequence of r is increased, priming from DR1 increases at the expense of priming from DR2.

DISCUSSION

A previous study showed that sequence identity between 5′r and 3′r helps facilitate the circularization reaction (16), although the mechanistic contribution made by the sequence identity was not established. In this study, we show evidence for a strong, positive correlation between the fraction of A and T residues in the terminal redundancy and the efficiency of the circularization process \( (P < 0.003) \) (Fig. 5B), although statistical significance was also obtained when testing for a correlation between circularization efficiency and the number of A residues alone \( (P < 0.004) \). The test statistic tau increased, but only slightly, when a correlation between circularization efficiency and both A and T residues was considered \( (\text{tau} = 0.50) \) compared with A residues alone \( (\text{tau} = 0.49) \). The inability to discriminate between these two possibilities is likely due to an underrepresentation of T residues in the population \( (13\% \text{ T}, 34\% \text{ A}, 35\% \text{ G}, \text{ and } 19\% \text{ C}) \) (Table 1). One model consistent with our findings involves melting of the duplex between the 3′ end of the nascent plus-strand DNA and the 5′ end of the minus-strand DNA (Fig. 1G) prior to duplex formation between the plus-strand complement of 5′r and the 3′r in the minus strand (Fig. 1H). Increasing the fraction of A and T residues in the sequence of r is expected to facilitate the dissociation of the two strands in the 5′r duplex, as the stability of that duplex should be lower than if the sequence contained a higher fraction of G and C residues. This model indicates the need for a precise juxtaposition of the 5′r and 3′r regions to facilitate the transfer. Although evidence exists for cis-acting elements that contribute to the circularization process (10, 15), there is no evidence of either cis- or trans-acting factors known to contribute to a precise juxtaposition of these regions. An example of a possible candidate would be an interaction between the P protein, which is covalently linked to the 5′r donor site, and a cis-acting element near the 3′ end of the minus-strand DNA.

An alternative model for the contribution of sequence identity may be to promote a triple-stranded DNA complex between the 5′r, 3′r, and the 3′ end of the nascent plus strand. This is an intriguing possibility, as forming a triple-stranded complex could either be an intermediate state in the model described above, or it could facilitate the entire reaction with elongation occurring directly from the triple-stranded complex, which has been shown for some DNA polymerases (21). Many of the triple helices studied to date involve polypurine- or polypyrimidine-rich sequences where the orientation of the strands is dependent upon the base composition (5). Although we did not find a statistical correlation between the fraction of purine or pyrimidine residues and circularization efficiency \( (P > 0.14; \text{two-sided analysis}) \), it remains possible a nonconventional complex involving the three strands of nucleic acid contributes to circularization.

We previously identified a small DNA hairpin in the minus-strand DNA that regulates in situ priming in DHBV (8). Here, we present evidence that an additional criterion that can modulate the level of in situ priming is the fraction of A and T versus G and C residues in the sequence of r. In particular, higher levels of in situ priming correlated with an increased fraction of G and C residues in the sequence of r \( (\text{Fig. 6A}) \). As mentioned earlier, we think this phenotype is independent of the sequence in 5′r; rather, it is related either to changes in the sequence of the RNA primer alone or to the duplex between the RNA primer and the 3′r sequence. In one model, the increased G and C residues in the sequence of the 3′r and the 5′ end of the primer would increase the stability between the primer and the 3′ end of minus-strand DNA (Fig. 1E). One consequence of higher stability may be an increase in the time the primer resides at DR1 at the expense of primer translocation, which could enable more of the primers to overcome the inhibition of priming at DR1 by the small DNA hairpin (8), leading to increased in situ priming (Fig. 1I). Other models can be invoked where the sequence of either 3′r or the 5′ end of the primer contributes to primer translocation independently of base pairing, such as contributing to the juxtaposition of DR1 and DR2 in some manner.

A previous study has drawn into question whether the small terminal redundancy includes the nucleotide at position 2529 (14). That study showed credible evidence that minus-strand DNA synthesis ends at the penultimate nucleotide on the pgRNA template, as drawn in our replication model (Fig. 1). While it is possible the 5-nt mutation used in our analysis includes a single nucleotide outside of the terminal redundancy, we do not think it impacts the outcome or our interpretations. If the increased A and T residues enhance dissociation of the 3′ end of the nascent plus-strand DNA from the 5′ end of the minus-strand DNA, then the adjacent base pair or base pairs may contribute to the process, although this conclusion cannot be drawn from our results. Similarly, there may be one less base pair between the RNA primer and the 3′ end of the minus-strand DNA prior to translocation. However, the mutated nucleotide does exist in the primer and may make a contribution to stability or otherwise. Furthermore, if we only consider the fraction of A and T residues in the 4 nt from coordinates 2530 to 2533, our in situ priming and primer translocation correlation tests are still statistically significant \( (P < 0.002 \text{ and } P < 0.02, \text{respectively}) \).

The plus-strand template switches, primer translocation and circularization, share their dependence upon the 3′ end of the minus-strand DNA. The 3′ end of the minus-strand DNA serves as the donor site for primer translocation and, subsequently, serves as the acceptor site for circularization. A number of cis-acting elements have been reported that contribute to both processes (8, 9). Here we report evidence of another. The sequence of r was able to modulate both primer translocation and circularization (Fig. 5B and 6A). Additionally, the sequence changes in r appear to influence both the transcriptional start site, possibly by mutating the initiator (Inr) element (24), and the 5′ termini of minus-strand DNA, likely by influencing the minus-strand template switch (Fig. 2B and 3). The extent to which the corresponding region of the genome in mammalian viruses contributes to replication is unknown. However, the mammalian viruses also contain sequence in r with a high fraction of A and T residues, with apparent conservation of the relative positioning of the G or C residues and of the A or T residues as noted by Seeger and colleagues (23). This conservation of position is interesting and, if important,
may explain why there appears to be some variance between the circularization efficiencies of viruses containing the same number of A and T residues.

ACKNOWLEDGMENTS

We are grateful to the members of the Loeb lab for helpful discussions and to Turkan Arca for her technical assistance in generating molecular clones. We thank Jesse Summers, Bill Sugden, Norman Drinkwater, and Ning Liu for critically reviewing the manuscript. This work was supported by National Institutes of Health grants PO1 CA22443, P30 CA07175, and T32 CA9135.

REFERENCES