Origin of Hepatitis Delta Virus mRNA

SEVERIN GUDIMA,1 SHWU-YUAN WU,2 CHENG-MING CHIANG,2 GLORIA MORALEDÁ,1 AND JOHN TAYLOR1*

Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111-2497,1 and Department of Biochemistry, University of Illinois, Urbana, Illinois 618012

Received 15 March 2000/Accepted 15 May 2000

Hepatitis delta virus (HDV) is unique relative to all known animal viruses, especially in terms of its ability to redirect host RNA polymerase(s) to transcribe its 1,679-nucleotide (nt) circular RNA genome. During replication there accumulates not only more molecules of the genome but also its exact complement, the antigenome. In addition, there are relatively smaller amounts of an 800-nt RNA of antigenomic polarity that is polyadenylated and considered to act as mRNA for transcription of the single and essential HDV protein, the delta antigen. Characterization of this mRNA could provide insights into the in vivo mechanism of HDV RNA-directed RNA transcription and processing. Previously, we showed that the 5′ end of this RNA was located in the majority of species, at nt 1630. The present studies show that (i) at least some of this RNA, as extracted from the liver of an HDV-infected woodchuck, behaved as if it contained a 5′-cap structure; (ii) in the infected liver there were additional polyadenylated antigenomic HDV RNA species with 5′ ends located at least 202 nt and even 335 nt beyond the nt 1630 site, (iii) the 5′ end at nt 1630 was not detected in transfected cells, following DNA-directed HDV RNA transcription, in the absence of genome replication, and (iv) nevertheless, using in vitro transcription with purified human RNA polymerase II holoenzyme and genomic RNA template, we did not detect initiation of template-dependent RNA synthesis; we observed only low levels of 3′-end addition to the template. These new findings support the interpretation that the 5′ end detected at nt 1630 during HDV replication represents a specific site for the initiation of an RNA-directed RNA synthesis, which is then modified by capping.

The two major species of RNA that accumulate during hepatitis delta virus (HDV) replication are the unit-length genome and its exact complement, the antigenome (Fig. 1, panels 1 and 2) (2). These 1,679-nucleotide (nt) RNAs are detected in both circular and linear forms. In addition, there are relatively smaller amounts of an 800-nt polyadenylated RNA of antigenomic polarity (Fig. 1, panel 3). This putative mRNA contains the open reading frame for the small delta protein, a 195-amino-acid species that is essential for HDV genome replication. The first study addresses whether the mRNA has a 5′-cap structure at nt 1630. The second considers whether in vivo HDV RNA-directed RNA transcription is an attempt to reconstitute initiation of transcription in vitro. The third examines the initiation of transcription and processing. In a recent study we used 5′-RACE to determine the location of the 5′ ends of the mRNA under several different conditions of HDV replication. The final study examines the consequences of some form of endonucleolytic processing. In the latter possibility there are precedents of stable polyadenylated RNA species whose 5′ ends are uncapped and arise via site-specific posttranscriptional endonucleolytic cleavage (17, 25).

In the present paper, we describe four studies that allow us to distinguish between some of the hypotheses mentioned above. The first study addresses whether the mRNA has a 5′-cap structure at nt 1630. The second considers whether there are also 5′ ends beyond nt 1630. The third examines the role of genome replication in 5′-end formation. The final study is an attempt to reconstitute initiation of transcription in vitro using Pol II.

MATERIALS AND METHODS

Plasmids. Most plasmids used, such as pTW101, have been previously described (30). pDL553 is a construct based on pSVL (Pharmacia), which expresses 1.2 copies of HDV genomic RNA (30). pDL554 is similar to pDL553 except for a 2-nt deletion at position 1434 to 1435 (12). pDL444 expresses the small form of the delta protein (13). pSG211, which expresses almost-full-length antigeno-

*Corresponding author. Mailing address: Fox Chase Cancer Center, 7701 Burholme Ave., Philadelphia, PA 19111-2497. Phone: (215) 728-2436, Fax: (215) 728-3105. E-mail: jm_taylor@fccc.edu.
mic HDV RNA was constructed by insertion of the large PshAl-Xhol fragment of HDV at the Smal and Xhol sites of pSVL.

Isolation of polyadenylated RNA isolated from liver of HDV-infected woodchuck. A woodchuck chronically infected with woodchuck hepatitis virus was superinfected with HDV, and at 25 days, around the time of peak acute HDV infection, the animal was sacrificed (29). Total RNA was extracted from samples of liver using a guanidium isothiocyanate procedure involving a final centrifugation to equilibrium in the presence of added cesium chloride (3). Polyadenylated RNA was then selected by two passages over an oligo(dT)-cellulose column (6).

RNA transcription in vitro. Large-scale transcription of 1.2× genomic RNA from the liver of an HDV-infected woodchuck. In step 1, this RNA was treated with calf intestinal alkaline phosphatase (CIP) to remove all exposed 5′-phosphates, and then in step 5, the cDNA product was used as template for a nested PCR, with primers as indicated in Fig. 2B. A gel analysis of the final inner PCR products is shown in Fig. 2C. Lane 1 represents a reaction performed in the absence of added template. Lane 2 corresponds to product made from RNA using steps 1 to 5. In lane 2, we detected a major band of about 180 bp. Such a band was not detected in the absence of the TAP treatment or in the absence of both CIP and TAP treatments (data not shown). Next, the total product in lane 2 was cloned and sequenced. The sequences of 18 clones are summarized in Fig. 2D. After allowances for repeat isolations of the same sequence, there are actually 10 unique sequences. These differences reflect either variations in the 5′ end of the RNA target or the 3′ end of the adapter. (For more details, see the legend to Fig. 2). For each sequence, we show only one interpretation of the adapter-target junction, although in some cases there are other interpretations. Collectively, these data indicate that the capped 5′ ends of the target RNAs must be located in the region of nt 1631 to 1628. However, we note that a more specific interpretation is that the majority of the sequenced junctions are consistent with a capped target RNA beginning at nt 1630, along with a less frequent site at nt 1628.

In a modification of this RLM-RACE procedure, we omit- ted the CIP and TAP treatments for the target RNA. This variation would be able to detect a target RNA that already had a 5′ monophosphate, but in fact, none were found (data not shown). Neither the original procedure nor this modified procedure would have detected RNAs whose 5′-end group was a hydroxyl, diprophosphate, or triphosphate. Overall, our inter-
pretation is that the polyadenylated antigenomic RNAs, at least for 5' ends in the vicinity of nt 1630, have primarily a cap structure rather than a 5' monophosphate. 5' ends at sites beyond nt 1630. The following experiments were undertaken to determine whether there might be at least a minor population of polyadenylated RNAs with 5' ends much beyond nt 1630 (Fig. 1, panel 4). Just the detection of such species would be proof that there are sites beyond nt 1630 on the genomic RNA template at which polyadenylated RNAs are initiated.

We used a novel RT-PCR strategy, other than RLM-RACE, to search for such species. The polyadenylated RNA from the infected liver was first isolated and then reverse transcribed using an oligo(dT) primer to obtain cDNA. Then, for PCR, we used three pairs of primers, P-Q, P-R, and P-S, as shown in Fig. 3A. With this strategy, the detection of a PCR product is an indication of polyadenylated RNAs with 5' ends at or beyond positions nt 1630, 153, and 286, respectively. The results are shown in Fig. 3B, lanes 1, 3, and 5, respectively. As a positive control, we carried out amplification of HDV cDNA in parallel reactions using the same primer pairs, as shown in lanes 2, 4, and 6, respectively, and detected products consistent with the expectations of 100, 302, and 435 bp, respectively. The polyadenylated RNA from infected liver produced bands corresponding to 100 bp (lane 1) and 302 bp (lane 3) and barely detectable amounts of 435 bp (lane 5). The HDV-specific identities of both the 302- and 435-bp products were confirmed by cloning and nucleotide sequencing (data not shown).

Thus, we interpret that in the liver during a natural infection, some of the polyadenylated antigenomic RNAs extend to at least nt 153 and 286, even though we know from the previous 5'-RACE analyses that the major fraction extends only to nt 1630 (6). We also carried out experiments using not liver RNA but RNA from cells in which genome replication was initiated by transfection with either cDNA constructs or HDV RNA transcribed in vitro. Using primers P and R, we were again able to detect the correct HDV-specific PCR product (data not shown).

Although RT-PCR results can be subject to artifacts, there are several aspects of our experimental design which reduce this possibility. (i) The RNA used was initially selected twice by oligo(dT)-cellulose. (ii) The RT was primed by oligo(dT). (iii) The reverse transcriptase used was a deleted form, lacking RNase H activity. (iv) We were looking for species whose 5' ends were actually beyond, rather than before, nt 1630. (v) Finally, we have confirmed the identity of the PCR products by cloning and nucleotide sequencing. One limitation, however, is that these RT-PCR assays did not give us independent quantitation of the fraction of molecules that proceed beyond nt 1630 relative to those that do not.

In summary, these results lead us to conclude that at least some of the polyadenylated antigenomic HDV RNAs are initiated at sites beyond nt 1630. While more experiments are needed to clarify the mechanism of such synthesis, our findings also open the possibility that species with 5' ends beyond nt 1630 could also function as mRNA.
cells, this DNA construct should be used by Pol II to transcribe (pSG211) by using the expression vector pSVL. In transfectedated antigenomic RNA. We designed the following construct whether, in the absence of HDV RNA-directed RNA synthe-

9 to produce the observed 5'

one or more upstream locations and subsequently were cleaved

That is, they are RNA species that somehow were initiated at

quences of posttranscriptional endonucleolytic RNA cleavage.

second, quite different possibility is that they are the conse-

also sites for the initiation of RNA-directed RNA synthesis. A

9 the origin of the 5'

nt 1642 and 1 was at nt 1656.

in reaction mixtures containing human TBP, a preassembled human Pol II

gene(dT)-cellulose before reverse transcription was carried out on RNAs immobilized to superparamagnetic beads via oli-
go(dT); lanes 2, 4, and 6, shown is a positive control of plasmid DNA (pDL553)
containing HDV sequences that did not need to be reverse transcribed. The
PCRs were carried out with primer pairs P-Q (lanes 1 and 2), P-R (lanes 3 and
4), and P-S (lanes 5 and 6). The products were separated by electrophoresis into

FIG. 3. RT-PCR analysis of polyadenylated HDV RNAs. (A) Shown is the location, relative to the genomic sequence of HDV using the numbering of Kuo et al. (11), of 4 primers used in RT-PCR assays. (B) Lanes 1, 3, and 5, the RNA used was extracted from the liver of an HDV-infected woodchuck, and it was then selected twice by binding to oligo(dT)-cellulose before reverse transcription was carried out on RNAs immobilized to superparamagnetic beads via oli-

FIG. 4. Ability of HDV genomic RNA to act in vitro as a template for RNA Pol II transcription. (A) HDV genomic RNA was transcribed from construct pTW101, which contains 1.2× HDV cDNA, inserted between a phage T7 RNA polymerase promoter and terminator. The numbering for HDV sites (thinner line) is from the sequence described by Kuo et al. (11). (B) The linear HDV genomic RNA was gel purified and used as a template for in vitro transcription in reaction mixtures containing human TBP, a preassembled human Pol II holoenzyme (f:pol II), [a-32P]CTP, and unlabeled (or cold) ribonucleoside triphosphates (cNTPs), in the absence (−) or presence (+) of actinomycin D (actD), δ-amanitin (δ-ama), or DNase I, as previously described (28). Individual components were then selectively omitted from the complete reaction mixture (all) as indicated (lanes 2 to 5 and 9). Reaction products were extracted and subjected to polyacrylamide gel electrophoresis under denaturing conditions, followed by autoradiography (28). For size markers, we used MspI fragments of pHBR322 DNA that were 5’ labeled using T4 polynucleotide kinase. The migra-

minor amounts of other ends in the vicinity were also detected

Furthermore, when studying the replication of certain mu-
tated genomes, we in some cases observed other 5’ ends; in

fact, for one mutant, the majority of the 5’ ends was at nt 1642
(6). Because of these results we undertook even further exper-
iments to determine the basis for the variety of detected 5’ ends.

An unexpected result was obtained when we examined the 5’ ends formed when genome replication was carried out in the presence of controlled amounts of delta protein. To achieve this, cells were cotransfected with a construct (pDL542) and mutated at a distant site (deletion of nt 1434 to 1435) that cannot make delta protein, along with a construct (pDL444) which acts as a controlled source of this protein. After 6 days, the total RNA was extracted and 5’-RACE was used, as previously reported (6), to determine the 5’ end of the polyadenylated antigenomic RNA. We found that 12 out of 23 se-

sequenced clones had a 5’ end at nt 1630. However, 10 were at

nt 1642 and 1 was at nt 1656.

These new results, along with the previous results for mu-
tated HDV genomes (6), force us to address the question of the origin of the 5’ ends at locations other than nt 1630. One possibility is that, as we have concluded for nt 1630, they are also sites for the initiation of RNA-directed RNA synthesis. A second, quite different possibility is that they are the conse-

quences of posttranscriptional endonucleolytic RNA cleavage. That is, they are RNA species that somehow were initiated at one or more upstream locations and subsequently were cleaved to produce the observed 5’ ends.

The following experiment was undertaken to determine whether, in the absence of HDV RNA-directed RNA synthe-
sis, we could detect endonucleolytic cleavage of a polyadenyl-
ated antigenomic RNA. We designed the following construct (pSG211) by using the expression vector pSVL. In transfected
cells, this DNA construct should be used by Pol II to transcribe

RNAs which, after processing, should produce a hybrid mRNA that contains an simian virus 40-specific 5’ end followed by almost the complete antigenome of HDV and ending with polyadenylation directed by the HDV polyadenylation signal. Even in the presence of delta protein, this RNA cannot repli-
cate but it should be able to fold into the rod-like structure typical of full-length antigenomic and genomic RNAs. The HDV sequences (without the vector sequences) of the pro-
cessed RNA are schematically represented in Fig. 1, panel 5.

We waited until 9 days after transfection to increase the opportunity for endonucleolytic cleavage of the hybrid mRNA. Then, we extracted the RNAs and used the 5’-RACE procedure to examine the 5’ ends of the polyadenylated RNAs containing HDV antigenomic sequences. Eight out of eight sequenced clones indicated 5’ ends within the HDV sequences; two of these ends were at nt 1642, two were at nt 1654, 1 was at nt 1657, 1 was at nt 1657, 1 was at nt 1632, 1 was at nt 1638, and 1 was at nt 117. Since one might have expected the 5’ ends to be predominantly in the vector sequences, we tested, in paral-
el, RNAs from cells transfected with other constructs. For two constructs that contained the open reading frame for delta
protein but no additional HDV sequences that would allow formation of rod-like folding (in the vicinity of the top of the rod), 15 out of 15 sequences indicated 5′ ends within vector sequences. As positive controls, two constructs capable of initiating genome replication indicated that 10 out of 10 5′ ends were at nt 1630.

From this experiment, it is important to note that of these 5′ ends determined for nonreplicating rod-like RNA, nt 1642 was found in the case of the mutant (pDL542), mentioned earlier. Furthermore, even for replication-competent mutant genomes previously reported (6), we have detected nt 1632, 1638, 1654, and 1657, along with nt 1642. Thus, we now have to allow that some or all of these 5′ ends may have arisen by endonucleolytic cleavage.

The present experiments have revealed that initiation is not the only way to get 5′ ends located other than at nt 1630. They can arise in the absence of any HDV RNA-directed RNA synthesis. We favor the interpretation that, at least in this particular case, they arose by posttranscriptional endonucleolytic cleavage of those DNA-directed RNA transcripts that were able to fold into the rod-like structure. However, we cannot exclude the possibility that they arose via DNA-directed initiation from what has been reported as a cryptic promoter in the double-stranded cDNA copy of this region of the HDV genome (1, 15).

Nevertheless, our experiments support two important conclusions regarding those 5′ ends that are at nt 1630. Such ends were not produced by posttranscriptional cleavage and were not DNA-directed initiations produced by the above-mentioned cryptic promoter.

Transcription by Pol II in vitro, using HDV genomic RNA as template. The final experimental approach we used was to reexamine in vitro transcription by Pol II of HDV RNA species. We reasoned that if nt 1630 corresponds to an initiation site and if Pol II is responsible for such RNA-directed transcription, then it might be possible to recreate such synthesis in vitro, using a Pol II-dependent cell-free transcription system with HDV genomic RNA as template. An in vitro transcription system reconstituted with a preassembled human Pol II holoenzyme (Epol II; see references 28 and 29) and a TATA-binding factor (TBP) was used, as it has been shown previously that this two-component system can efficiently and specifically produce RNA from DNA templates that contain promoters, such as those from adenovirus (28), human immunodeficiency virus (28), and human papillomavirus (8).

For these studies, we prepared a greater-than-unit-length genomic RNA template via transcription with T7 RNA polymerase utilizing the plasmid DNA diagrammed in Fig. 4A. This RNA was gel purified and then was used as a template for transcription by Pol II. As shown in Fig. 4B, an HDV-specific product of approximately 220 nt was detected in the presence of Pol II holoenzyme and TBP (compare lanes 1 and 4). Formation of this HDV product was dependent on Pol II holoenzyme, as omitting Pol II holoenzyme, but not TBP, from the complete reaction abolished the signals (lanes 1 to 3). That inclusion of α-amanitin, but not actinomycin D or DNase I, during the transcription reaction eliminated the signals (lanes 6 to 8) further indicates that this HDV-specific product was indeed generated by Pol II from HDV genomic RNA and was not derived from any DNA molecules, if present, in the reactions. When cold ribonucleoside triphosphates were left out of the complete reaction mixture, the signals from all the labeled species were enhanced (compare lanes 1 and 5), which suggested to us that these products were generated by the 3′-end-labeling activity of Pol II. Consistent with this interpretation, the size of the HDV-specific product, which was still detected in the reactions without cold ribonucleoside triphosphates (compare lanes 5 and 9), was slightly reduced (compare lanes 1 and 5). The HDV-specific product, therefore, likely represents a 3′-end-labeled fragment released from the 3′ end of the 2-kb RNA template by prior action of the HDV ribozyme.

It should be noted that even the HDV-specific end-labeled product was of very low abundance. It took 4 days to detect by autoradiography, which should be contrasted to transcription from a Pol II-directed DNA promoter, the products of which were usually detected in several hours (data not shown).

We also carried out Pol II transcription experiments in the absence of radioactive label, using glyoxalation, agarose gels, and Northern analysis to assay for reaction products. Other modifications included the use of trimers of HDV RNA and also with antigenomic rather than genomic RNA templates. None of these studies detected discrete-sized run-off RNA-directed species consistent with initiation at nt 1630 or any other location (data not shown).

In summary, in these studies using an in vitro Pol II system with clear competence for DNA-directed RNA synthesis, we were unable to detect initiation of RNA-directed RNA synthesis on an HDV RNA template. All that we detected could be described as short additions to the 3′ end of the template molecules.

**DISCUSSION**

One of the major questions for HDV replication has concerned whether the 5′ end detected at nt 1630 on the antigenic polyadenylated RNA represents a site of initiation of RNA-directed RNA synthesis or a site of posttranscriptional endonucleolytic RNA processing (2, 6, 9, 14, 18). Here we provide three lines of evidence, which taken together, lead us to favor the first possibility. (i) By RLM-RACE, we obtained evidence that the 5′ end contains a cap structure (Fig. 2D). All known examples of RNA species that are capped in vivo, involve initiation to create 5′ ends with a triphosphate that is then converted to a cap structure. RLM-RACE is a very sensitive method but because it is indirect, we cannot comment on the presence or absence of cap-associated methylation modifications. It should be added that our studies did not detect 5′ ends at nt 1630 that contained a 5′ monophosphate and that the strategy used would not have been able to detect species with a 5′ hydroxyl or triphosphate. (ii) By 5′ RACE, we found that 5′ ends at nt 1630 were not just preferred but actually only detectable when genome replication was also occurring. (iii) For cells transfected with a DNA construct that expresses almost the entire antigenomic sequence as a nonreplicating polyadenylated RNA, we were unable to detect by 5′ RACE any ends at nt 1630. This also argues against an endonucleolytic cleavage mechanism.

If nt 1630 is a site of initiation, it seems reasonable to suggest that nearby some sequence and/or structural feature serves as a promoter. This idea has been the rationale behind several mutagenesis studies (1, 6, 30). However, some studies have gone so far as to examine double-stranded cDNA versions of HDV sequences from this region for such a putative promoter and have even reported such activity in transfected cells (1, 15). In contrast, our 5′ RACE data for antigenomic RNA expressed from a nonreplicating form of HDV cDNA did not detect nt 1630 as a 5′ end. Thus, we argue against the biological relevance to HDV of the cDNA promoter and consider the question of the mechanism of initiation of RNA-directed RNA synthesis still unresolved.

Our studies also show that there are also real but relatively minor amounts of polyadenylated antigenomic RNAs with 5′
ends that are located 202 nt and even 335 nt beyond 1630 (Fig. 3B). This allows us to conclude that, in a natural HDV infection, there are at least some sites other than nt 1630 for the initiation of polyadenylated HDV RNAs, even though we cannot say whether the RNAs we detected are capped or whether such RNAs have undergone an endonucleolytic cleavage.

As mentioned above, we were unable to detect any evidence for endonucleolytic cleavage at nt 1630. However, under conditions in which HDV replication was not occurring, we were able to obtain in vivo data consistent with endonucleolytic cleavage at other sites. Specifically, we used 5'-RACE to examine RNA from cells transfected with a cDNA construct designed to direct the synthesis of DNA-directed RNA transcripts that (i) were initiated in vector sequences, (ii) were followed by the almost-unit-length antigenome, and finally (iii) contained signals for polyadenylation. At 9 days after transfection, we detected polyadenylated RNAs with 5' ends not as expected, in the vector sequences, but at a variety of locations within the HDV sequence. In some cases, the location was near to, but never at, nt 1630. Since in these experiments there is not even the genomic RNA template for RNA-directed RNA synthesis, our interpretation is that these particular 5' ends arose via posttranscriptional endonucleolytic cleavage. This interpretation has implications for other studies in which RNA-directed RNA synthesis is occurring for either wild-type or mutated HDV, and we detected rare 5' ends that are not at nt 1630. Previously we have pointed out that such ends have a preference for the 5' nucleotide being adenosine or guanosine, indicative of sites of initiation by a host RNA polymerase (6). However, such indications are not sufficient, and maybe now we must allow that for each non-1630 end, additional data are needed to distinguish whether it arose via endonucleolytic action or initiation.

Returning to the question of nt 1630 as a site for the initiation of RNA-directed RNA synthesis, our in vivo data indicative of capping provoked us to attempt an in vitro reconstruction experiment. We used a system of proven competence for DNA-directed RNA synthesis by RNA Pol II. However, all we were able to detect in the presence of a genomic HDV RNA template was small amounts of 3' ends created by initiation and/or endonucleolytic cleavage, but which because they are not protected, are less stable and thus less readily detected.

ACKNOWLEDGMENTS

This work was supported by NIH grants AI-26522 and CA-06927 for J.T. and CA-81017 and GM-59643 for C.-M.C. Additional support was from an appropriation by the Commonwealth of Pennsylvania. At Fox Chase, we thank the DNA Synthesis Facility for providing the necessary oligonucleotides and Aneta Cywinski of the DNA Sequencing Facility for the automated sequencing. Thanks also to Preecha Biswas, Jinzhong Chang, Huan Zhou, and Vadim Bichko for their earlier efforts directed at in vitro transcription of HDV RNA templates. William Mason, Glenn Rall, Richard Katz, and Jinzhong Chang gave constructive comments on the manuscript.

REFERENCES

hepatitis delta virus. R.G. Landes Co., Austin, Tex.


