Identification of a Conserved RNA Replication Element (cre) within the 3Dpol-Coding Sequence of Hepatoviruses

Citation for published version:

Digital Object Identifier (DOI):
10.1128/JVI.00787-08

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Virology

Publisher Rights Statement:
Copyright © 2008, American Society for Microbiology. All Rights Reserved.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Identification of a Conserved RNA Replication Element (cre) within the 3DPol-Coding Sequence of Hepatoviruses

Yan Yang,†‡ MinKyung Yi,†‡ David J. Evans,2 Peter Simmonds,3 and Stanley M. Lemon1*

Center for Hepatitis Research, Institute for Human Infections and Immunity, and the Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, Texas 77555-1073; Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom; and Virus Evolution Group, Centre for Infectious Diseases, University of Edinburgh, Summerhall, Edinburgh EH9 1QH, United Kingdom

Received 12 April 2008/Accepted 29 July 2008

Hepatitis A virus (HAV), a member of the family Picomavirusidae, is a common etiological agent of acute hepatitis in humans. Several characteristics of HAV distinguish it from other picornaviruses, justifying its classification within a separate genus in this family, the genus Hepatovirus (10). First, the primary sequence of the positive-sense, single-stranded RNA genome of HAV shares little homology with that of other picornaviruses. Avian encephalitis virus (AEV) is the only exception to this statement, as it bears considerable sequence similarity to HAV, despite its significantly larger size and lower free folding energy. However, a region of markedly suppressed synonymous codon variability was identified in alignments of HAV sequences near the 5′ end of the 3DPol-coding sequence of HAV, consistent with noncoding constraints imposed by an underlying RNA secondary structure. Subsequent MFOLD predictions identified a 110-nucleotide (nt) complex stem-loop in this region with a typical AAACA/G cre motif in its top loop. A potentially homologous RNA structure was identified in this region of the avian encephalitis virus genome, despite little nucleotide sequence relatedness between it and HAV. Mutations that disrupted secondary RNA structure or the AAACA/G cre motif, without altering the amino acid sequence of 3Dpol, ablated replication of a subgenomic HAV replicon in transfected human hepatoma cells. Replication competence could be rescued by reinsertion of the native 110-nt stem-loop structure (but not an abbreviated 45-nt stem-loop) upstream of the HAV coding sequence in the replicon. These results suggest that this stem-loop is functionally similar to cre elements of other picornaviruses and likely involved in templating Vpg uridylylation as in other picornaviruses, despite the virus' unique biological attributes. Strongly hepatotropic in vivo and restricted in its host range to humans, chimpanzees, and certain New World primates, wild-type (wt) HAV replicates very poorly in cultured cells, demonstrating a protracted replication cycle and typically no evidence of a cytopathic effect in infected cells (10). Recent evidence suggests that the 3A protein of HAV is specifically directed to the outer mitochondrial membrane (28), rather than membranes of the endoplasmic reticulum, as is the case with other picornaviruses, possibly contributing to the low efficiency of replication that typifies this virus.

Despite these striking distinguishing characteristics, HAV shares many features in common with other picornaviruses, particularly in terms of its overall genome organization and apparent replication strategy (10). Like the genomes of other picornaviruses, the HAV genome contains a lengthy 5′ untranslated region (5′ UTR), followed by the single long open reading frame encoding the polyprotein, a short 3′ UTR, and a 3′-polyadenylated tail. The 5′ UTR lacks a 5′-terminal m7G cap structure and in genomic RNA is covalently linked to a small virus-encoded peptide (3B or Vpg) (27). The internal ribosome entry site (IRES) located within the 5′ UTR directs the cap-independent translation of the polyprotein, which is cotranslationally processed by the 3Cpro proteinase to produce structural proteins that comprise the viral capsid and nonstruc-
ural proteins involved in replication of the viral genome (2). Like other picornaviruses, genome replication is a two-stage process, with the input RNA genome first transcribed to produce antisense RNA, which then functions as template for the synthesis of positive-sense progeny genomes (10). 3D\textsuperscript{NS}-, the RNA-dependent RNA polymerase and catalytic core of the viral replicase, directs the synthesis of both positive- and negative-strand RNA. Although far less well studied than poliovirus (PV), the synthesis of HAV RNA is thought to be primed by VPg-pUpU, the product of 3D\textsuperscript{NS}-mediated uridylylation of VPg.

The genomes of all picornaviruses contain RNA replication signals within both the 5' and 3'-terminal domains. However, the genomes of many other picornaviruses also have been found to contain internally located stem-loop structures that are essential for viral RNA synthesis. First recognized within the capsid coding region of human rhinovirus 14 (HRV-14), its function in viral RNA replication in vivo could not be complemented in trans, leading to its designation as a cis-acting replication element (cre) (12). The ability of the cre to support viral RNA synthesis is dependent upon both specific RNA structure and certain nucleotides within the loop region (14, 16, 25, 29). On the other hand, it is independent of its position within the genome and of whether its sequence is translated into protein (12). Similar RNA elements have subsequently been identified within the P1 sequence of cardioviruses, the 2C-coding region of PV, the 2A-coding sequence of HRV-2, and the 5' UTR of an aphthovirus (4, 5, 9, 11).

Studies by Paul and colleagues (14, 16, 17) have shown that the PV cre functions as the template for VPg (3B) uridylylation through a "slide-back" mechanism catalyzed by 3D\textsuperscript{NS} in association with 3CD. The uridylylation of VPg, possibly in the context of 3AB, leads to the production of VPg-pUpU, which serves as the protein primer for new RNA synthesis (8, 15). Extensive mutagenesis of the HRV-14 and PV cre revealed a critical conserved AAACA/G motif in the 5' half of the loop sequence that is essential for cre function (29). Similar conserved AAACA motifs are present within the loops of the cre elements of other picornaviruses and are important for RNA replication (17, 31). Evidence suggests that a cre is likely to be present in all picornaviruses but at different positions within the genome in different picornaviruses and with substantial variation in primary nucleotide sequences. To date, however, searches for such an element in the HAV genome have not been productive.

Internal base pairing that creates stem-loops and other RNA structures places constraints on sequence variability in bases required for structure formation. In the hepatitis C virus (HCV) genome, this constraint is manifested by a marked suppression of synonymous codon variability within several evolutionarily conserved stem-loops in the core and NS5B-coding regions that have demonstrated roles in viral replication (13, 26, 32). Discrete RNA structures such as the cre in the coding region of human enteroviruses (HEVs) and other viruses that lack other large-scale RNA secondary structures (22) should also create characteristic suppression of synonymous site variability (SSSV), similar to that observed in HCV. Here, we describe the use of independent phylogenetic and thermodynamic methods to scan HAV sequences for covariant sites and associated RNA secondary structures, leading to the prediction of a conserved stem-loop structure within RNA encoding the 3D\textsuperscript{NS} RNA-dependent RNA polymerase. We confirm the functional importance of this structure by mutagenesis and reverse molecular genetics. We show that this RNA element shares several common features with other picornavirus cre elements but is unique in both size and location.

MATERIALS AND METHODS

Nucleotide sequences. The following sequences of human and simian HAV were used for analysis of variability at synonymous sites and for RNA structure determination:

- NC_001489, DQ646426, AB258387, AB020657, HPAACG, AB20566, AB20564, AB20565, AB20568, AB20569, AF314208, HPAACG, AB30205, AF512536, AF357222, AF485328, HAVCOMPL, HPA158, AY644670, AY644676, HAVRNNHS, EF207320, EF406357, EF406359, SHVAGM2, AJ299464, AY971410, and HEA299464. Sequences showing <1% sequence divergence from other published sequences were excluded from analysis. The following nonidentical sequences for AEV were also analyzed for RNA secondary structure: NC_003990, AJ225173, AY517471, and AJ225539. Alignments of HAV and AEV sequences were carried out by identification and alignment of conserved amino acid sequence motifs in the coding region and automated alignment of intervening regions using ClustalW with default settings. Alignments of HEV species A and B viruses corresponded to those used in previous analyses (23).

Analysis of SSSV. Synonymous sequence variability was determined by measurement of mean pairwise distances at each codon position in the open reading frame of hepatoviruses and enteroviruses. Variability at each codon was calculated using the program Sequence Scan in the Simmonic sequence editor (21). Mean pairwise synonymous variability was restricted to aligned codons where the translated amino acid was the same. Each pairwise value was normalized by dividing by the degeneracy of the codon, with normalization factors for twofold degenerate sites of 0.5, for threefold degenerate sites of 0.6666, for fourfold degenerate sites of 0.75, and for sixfold degenerate sites of 0.8333. This takes into account the different sequence distances achievable at maximally diverged sites. Variability at each codon position was averaged over a sliding window of 35 codons.

RNA secondary structure prediction. Base pairing in the region of the genome showing SSSV was predicted using MFOLD using default settings through the web interface at http://www.bioinfo.rpi.edu/applications/mfold. PFOLD analysis was used the web interface at http://www.daimi.au.dk/~joncomp/bioenfold. All programs were run with default settings. Thermodynamic structure predictions were carried out using the program ZIPFOLD on the MFOLD server. Minimum folding energies (MFES) of native HAV and HEV-B sequences were compared with the mean values generated for 50 control sequences in which sequence order had been scrambled using an algorithm (NDR) that preserves the dinucleotide frequencies of the native sequence, implemented in the Simmonic sequence editor package. MFE differences (MFEDs) were calculated as MFED = MFE\textsubscript{native} - MFE\textsubscript{scrambled}, which are the MFEs of native sequences and mean MFEs of the 50 sequence-order-randomized controls, respectively.

Plasmids. The contribution of RNA structures to replication of the HAV genome was assessed by creating mutations within pHAVLuc, which contains the cDNA of a replication-competent, subgenomic RNA replicon, HAVLuc, in which an in-frame fusion of the firefly luciferase coding sequence replaces all but the 5' 150 and 3' 90 nucleotides of the P1 region of the HM175/18f genome (30). pHAVLuc-\Delta3D is a related replication-incompetent mutant of pHAVLuc, which contains a single base substitution creating a premature termination codon within the 3D\textsuperscript{NS} sequence; it is referred to here simply as \Delta3D. Mutations disrupting the native RNA sequence of the 3D\textsuperscript{NS}-coding region were introduced by QuikChange site-directed mutagenesis (Stratagene). The following oligonucleotides were used for construction of mutations: for MutA, GTTCATGAA GTTGGCTGTAAGAACACTTTCCTACCTAAAAAGAAGGT (+) and GACTTTTC TAAAAGGTGTCCTTGCAGACGACATTTGAGAC (+); for MutB, GCTT TTAGAGAAAACTCTCAATCTACCATGTTGAAAC (+) and GTT TTATGAGCGAAGTAAATGACATTACGTTTTT (+); and for MutC, GACCTGACATGTGCCTCCACAGACCTTTTGAAGAAC (+) and GACTTTTC TAAAAGGTGTCCTTGCAGACGACATTTGAGAC (+). The MFE\textsubscript{target} and MFE\textsubscript{random} are the MFES of native sequences and mean MFES of the 50 sequence-order-randomized controls, respectively.

To construct HAV replicons with reinsertions of the putative wt and mutated cre sequences immediately downstream of the luciferase coding sequence, the SacI site (nt 3006 in the HM175/18f sequence [7, 33], which was used to fuse the luciferase sequence to sequence encoding the C terminus of VP1 in pHAVLuc,
was eliminated by mutating $T^{500}$ to G (silent base change) to create pHAVLuc_v.2. A new SacI site was then placed at the 3' end of the full-length cre sequence by introducing $A^{507c}$G and $A^{507c}$G mutations to create pHAVLuc_v.3. The sequence between nt 3010 and 5955 (HM175/18f sequence) was deleted by QuikChange mutagenesis to fuse the full-length cre sequence in frame to the 3' end of the luciferase sequence in pHAVLuc_v.3, resulting in pHAVLuc_v.3. Next, pHAVLuc was digested with SacI (nt 3006 in the HM175/18f sequence) and XhoI (nt 7013), and the small fragment was ligated into the SacI/Xhol sites in pHAVLuc_v.3 to create pwt/wt. A similar strategy was used to construct pMutA/MutA, containing two mutated cre elements derived from MutA, and pwt/MutA and pMutA/wt. For construction of ps-cre/ps and ps-cre/MutA, the 45-nucleotide cre RNA segment was introduced into pHAVLuc and pMutA by QuikChange site-directed mutagenesis using the oligonucleotides GTTCAATG GAGCTCTAAGCTTATAAATGGGACTCTTTCTAAAAAGCGTTTTGGA GACCACATTCATGTACCCATTTGGACTTTCC (+) and GTTCAATG GAGCTCTAAGCTTATAAATGGGACTCTTTCTAAAAAGCGTTTTGGA GACCACATTCATGTACCCATTTGGACTTTCC (-). All plasmid regions subjected to PCR mutagenesis were sequenced to ensure that no adventitious mutations were introduced.

**In vitro RNA transcription.** To produce replicon RNA transcripts, plasmids were linearized at the unique XmaI restriction site located at the 3' end of the HAV sequence (30). RNA transcripts were synthesized by T7 polymerase-mediated transcription (T7 MEGAscript; Ambion). The integrity and yield of the transcribed RNAs were determined by agarose gel electrophoresis.

**Cell culture.** Human hepatoma cells were grown in Dulbecco's modified Eagle's medium (Gibco/BRL) with 10% fetal bovine serum.

**HAV RNA replication assay.** Human cells were transduced with replicon RNA transcripts. Briefly, 5 x 10⁶ cells were electroporated with 10 μg of RNA using a GenePulser II electroporation apparatus (Bio-Rad) with the pulse controller set at 1,400 V and 25 μF and maximum resistance. The cells were subsequently seeded into a 12-well plate and cultured in Dulbecco's modified Eagle's medium (Gibco/BRL) with 10% fetal bovine serum. HAV replication was quantified using the luciferase assay system (Promega) set at 1,400 V and 25 μF and maximum resistance. The cells were subse- quently seeded into a 12-well plate and cultured in Dulbecco's modified Eagle's medium with 10% fetal bovine serum at 37°C until processing for luciferase assay. Cell lysates were harvested by the addition of 100 μl of luciferase assay buffer (Promega) to each well and stored at -20°C until assayed for enzymatic activity.

**RESULTS**

**Prediction of an RNA structure within the N-terminal region of the HAV 3Dpol-coding sequence.** To investigate whether the RNA secondary structure of a previously defined cre in the 2C region of HEVs (5) could be identified by suppression of variability at synonymous sites, alignments of complete coding sequences of species A and B isolates were scanned for variability in flanking regions (14 invariant, 25 minimally variable from a total of 107 codons; 13% and 23%, respectively, $P < 0.0000013$ by the $x^2$ test). Importantly, the codons comprising the AAACA/G motif and adjacent sequence within the 5' half of the loop that has been shown to play a templating role in uridylylation of VPg (16, 29). Importantly, the region of predicted internal base pairing showed a higher frequency of invariant or minimally variable codons than did flanking regions (Fig. 3). For the 35 codons in the stem-loop for which there are potential synonymous alternatives, 20 were invariant (57%) and nine showed variability of $<0.33$ (26%). These frequencies are significantly higher than variability in flanking regions (14 invariant, 25 minimally variable from a total of 107 codons; 13% and 23%, respectively, $P < 0.0000013$ by the $x^2$ test). Important for the presence of detectable stability differences between native sequence of HAV and controls are discussed below.

 Although substantially larger, the predicted stem-loop structure in the HAV 3Dpol region showed several features that are common to cre elements present in other picornavirus genera. These include a large open top loop (18 bases, larger than the 14 bases in HRV-14, HRV-2, and HEV [4, 5, 12] and 15 bases in foot-and-mouth disease virus [11]), as well as the presence of the AAACA/G motif in the 5' half of the loop that has been shown to play a templating role in uridylylation of VPg (16, 29). Importantly, the region of predicted internal base pairing showed a higher frequency of invariant or minimally variable codons than did flanking regions (Fig. 3). For the 35 codons in the stem-loop for which there are potential synonymous alternatives, 20 were invariant (57%) and nine showed variability of $<0.33$ (26%). These frequencies are significantly higher than variability in flanking regions (14 invariant, 25 minimally variable from a total of 107 codons; 13% and 23%, respectively, $P < 0.0000013$ by the $x^2$ test). Importantly, the codons comprising the AAACA/G motif and adjacent sequence within the 5' half of the loop segment were absolutely invariant in their sequence (Fig. 3).

To further investigate the phylogenetic conservation of the 3Dpol stem-loop in HAV, we examined sequences from the distantly related AEV. AEV is tentatively classified within the genus *Hepatovirus*, but recent work has revealed that its IRES structure differs significantly from that of HAV and is actually closer to that of HCV (1). This and other considerations have prompted a reconsideration of its taxonomic classification. The lack of sequence diversity in published complete genome sequences of AEV precluded a meaningful analysis of SSVV for this virus. However, while coding sequences from AEV are highly divergent from those of HAV, a definable alignment of HAV and AEV sequences was possible in the region of the genome around the start of the 3Dpol-coding sequence (Fig. 4). MFOLD analysis of the region homologous to positions 5501 to 6500 of HAV demonstrated a stem-loop structure in a position equivalent to that in HAV, in the virtual absence of nucleotide sequence similarity between viruses (Fig. 2). Although the actual pairings in the stem
differed considerably between HAV and AEV, the AEV structure formed a very similar unpaired terminal loop of 16 bases, with the AAACG sequence placed in a position homologous to that in the HAV loop.

RNA structure within the 3Dpol-coding region is essential for efficient HAV RNA replication. To assess whether the putative stem-loop structure identified within the hepatovirus 3Dpol-coding region has functional importance, possibly equiv-
alent to that of other picornaviruses, we constructed two mutants (MutA and MutB) within the background of a replication-competent, subgenomic RNA replicon, HAVLuc, in which an in-frame fusion of the firefly luciferase coding sequence replaces most of the P1, capsid coding region of the genome of a rapidly replicating, cell culture-adapted variant of HAV (Fig. 5A) (30). To create MutA, we introduced five silent, third-base mutations in the 5’/H11032 half of the stem-loop, including two that disrupted predicted base pairs near the top of the stem and two within the AAACG sequence in the loop (Fig. 5B). MutB contained four silent mutations within the 3’/H11032 sequence of the upper part of the stem (Fig. 5B). Thus, the multiple nucleotide substitutions created in these mutants disrupt the AAACG/A motif and/or predicted base pairing within the helix, without altering the amino acid sequence of the 3Dpol RNA-dependent RNA polymerase. The MFOLD algorithm predicts that the mutations in either of these mutants significantly disrupt the RNA structure predicted within this segment of the wt genome (data not shown).

To evaluate the impact of these mutations on HAV RNA replication, we transcribed RNA in vitro from both wt and mutant constructs and transfected these RNAs into human hepatoma (Huh7) cells that support HAV replication. As a negative control, we also transfected Huh7 cells with a related, replication-defective RNA, HAVLuc-Δ3D (hereafter referred to as Δ3D), in which translation of 3Dpol is prematurely terminated (30). Transfection of this RNA leads to an initial burst of luciferase expression, detectable at 24 h posttransfection, but no subsequent increase in luciferase activity at 48 or 72 h. This early luciferase expression represents translation of the input RNA, which is degraded following transfection. In contrast, the replication-competent wt RNA (HAVLuc), while generating a similar level of luciferase activity at 24 h, shows a sustained, almost 10-fold increase at later time points (Fig. 5C). Interestingly, although none of the nucleotide substitutions that we engineered into MutA or MutB altered the amino acid sequence of the 3Dpol RNA-dependent RNA polymerase, the MFOLD algorithm predicts that the mutations in either of these mutants significantly disrupt the RNA structure predicted within this segment of the wt genome (data not shown).

### Table 1. MFEs of sequences flanking the HAV versus HEV-B cre

<table>
<thead>
<tr>
<th>cre region</th>
<th>Position (bp)</th>
<th>Length (bp)</th>
<th>MFE</th>
<th>MFED (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAV</td>
<td>5952-6002</td>
<td>100</td>
<td>−15.4, −15.7, −1.9</td>
<td></td>
</tr>
<tr>
<td>HEV-B</td>
<td>4401-4451</td>
<td>300</td>
<td>−56.7, −57.2, −0.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4305-4455</td>
<td>300</td>
<td>−71.8, −67.2, 6.8</td>
<td></td>
</tr>
</tbody>
</table>

*Comparison of MFEs of 100- and 200-nt segments centered on the cre in the human HAV and HEV-B genomes with those of sequence order-scrambled controls. The MFED quantifies the contribution of sequence order to RNA secondary structure.
RNA replication. Since the amino acid sequences of the polyproteins of MutA and MutB were unchanged from those of the wt genome, these results suggest that the predicted RNA structure in the 3Dpol-coding sequence is critically important for HAV RNA replication and that it may in fact function as a cre.

To further evaluate this possibility, we created a third mutant, MutC, containing only a single base alteration, A5999 to G (Fig. 5B). This mutation does not alter the MFOLD-predicted secondary structure in this region of the genome, nor the amino acid sequence of 3Dpol, but changes the AAACG sequence to AGACG, thus knocking out the AAACA/G motif. Since the second adenosine in this motif is critically important to the slide-back mechanism underlying the 3Dpol-mediated uridylylation of VPg (16, 17), this mutation would be expected to be lethal to replication, if in fact the predicted stem-loop functions as a cre. Consistent with this hypothesis, the MutC RNA also failed to replicate following transfection into Huh7 cells (Fig. 5C).

Functional rescue of MutA by the full-length but not a short, 45-nt RNA element. An interesting characteristic of the picornavirus cre is its ability to function in supporting viral RNA replication in a manner that is independent of its position within the genome (5, 12). We thus attempted to rescue the replication competence of MutA by inserting the entire 110-nt stem-loop sequence (Fig. 6A, f-cre, nt 5946 to 6059) downstream of the HAV coding sequence, at unique restriction sites engineered into MutA between the stop codon terminating translation of 3Dpol and the 3′/H11032 UTR. This replicon RNA showed no evidence of replication following transfection into Huh7 cells (data not shown). However, since an RNA pseudoknot has been proposed previously to form around the junction of the 3Dpol-coding sequence and the 3′ UTR (6), it is likely that the insertion of the f-cre sequence at this position

FIG. 3. Distribution of synonymous site variability among codons forming the 3Dpol stem-loop and flanking sequences of HAV. The 38 codons contributing to the complex stem-loop structure shown in Fig. 2 (nt 5946 to 6059) are separated from flanking codons in the center of the figure.

FIG. 4. Alignment of nucleotide sequences based on inferred amino acid sequences of human and simian HAVs and of AEV. Nucleotides within the 3Dpol stem-loops of HAV and AEV are shown in bold, with bases within the stem regions of the structure shown in inverse. The AAACA/G motif in the loop region (AAACG in HAV) is boxed. Sequences used are those shown in Fig. 2.
We next attempted to rescue the replication competence of MutA by inserting the f-cre sequence (Fig. 6A) immediately downstream of the luciferase sequence, fused in frame with and upstream of the residual VP1 coding sequence in HAVLuc (Fig. 6B, wt/MutA). To control for the possible loss of replication efficiency due to the insertion of this added sequence, we created a parallel construct containing the native wt stem-loop structure (Fig. 5B), the native stem-loop presumably retains its structure and possibly its protein-binding activities in wt/MutC. Despite this, the replication phenotype of wt/MutC was indistinguishable from that of HAVLuc or wt/wt, indicating that these redundant stem-loops neither compete with each other for limited amounts of essential binding proteins nor otherwise impede RNA synthesis. Taken together, these data show that the wt stem-loop structure is capable of supporting viral RNA replication when placed within the genome at a location several thousand nucleotides upstream of its native location, consistent with the location-independent nature of cre elements identified previously in other picornaviruses.

The f-cre sequence inserted in wt/MutA (Fig. 6A) is substantially longer than the minimal 33-nt rhinovirus cre sequence that we previously found to be sufficient for efficient replication of HRV-14 RNA (29). Thus, it was of interest to determine if the entire 110-nt f-cre sequence is required for replication of HAV RNA. To address this question, we constructed an additional mutant, s-cre/MutA (Fig. 7A), in which a 45-nt segment representing the HAV f-cre loop sequence and the adjacent 12-bp upper helical segment containing a 2-nt internal loop and 1-nt bulge (Fig. 7B, s-cre, nt 5982 to 6026) was inserted in frame between the luciferase and residual VP1 sequences of MutA. cre sequences of this length are capable of rescuing replication of other HRV-14 cre mutants and also support 3Dpol-mediated uridylylation of VPg in cell-free reactions (29). This mutant failed to replicate following transfection into Huh7 cells (Fig. 7C, s-cre/MutA). The lack of replication competence could not be attributed to the addition of the s-cre sequence upstream of the HAV coding sequence, as
FIG. 6. Replication competence is rescued by reinsertion of the $3D_{pol}$ stem-loop fused downstream of the luciferase sequence in the MutA replicon. (A) The full-length, wt 110-nt $f$-cre sequence inserted downstream of the luciferase sequence in the constructs shown in panel B. (B) Schematic showing the insertion of the wt $f$-cre stem-loop downstream of the luciferase sequence in the wt/wt, wt/MutA, and wt/MutC replicons. MutA/wt and MutA/MutA contain a similarly inserted, mutated $f$-cre sequence (mutations identical to MutA as in Fig. 5B) in the background of HAVLuc and MutA, respectively. (C) Relative luciferase activities expressed by HAVLuc, $\Delta 3D$, MutC (Fig. 5), wt/wt, wt/MutA, MutA/wt, MutA/MutA, and wt/MutC at 24, 48, 84, and 106 h (left to right, respectively, within each set of bars) post-transfection of Huh7 cells. The results shown are the averages of nine independent cultures of transfected cells; error bars indicate the standard deviations.
the insertion was well tolerated in the background of HAVLuc (Fig. 7C, s-cre/wt). Although further studies were not undertaken, these data indicate that HAV RNA replication requires a substantially lengthier cre than does that of other picornaviruses, possibly as long as the entire f-cre sequence (Fig. 6A). This may be related to the relatively low free folding energy of the HAV cre compared to that of other picornavirus stem-loop structures, as discussed below.

**DISCUSSION**

Conserved internal stem-loop structures that function as templates for the uridylylation of VPg and that are thus essential for RNA replication have been identified previously in representative viruses from most of the major pathogenic picornavirus genera. Such a cre element has not, however, been identified previously within the genome of hepatoviruses. Previous studies have shown that the sequence spanning the VP4 to 2A coding regions of HAV can be deleted without loss of RNA replication competence, indicating that there is not an essential cre element residing within the P1 sequence of HAV. Thus, it was of interest to find a large, stable RNA structure existing within the P3 segment encoding the viral RNA-dependent RNA polymerase, 3D\textsuperscript{pol}. This structure conforms to the general requirements of a picornavirus cre element, including an AAACA/G motif appropriately placed within the top loop (5, 12, 17, 29). The AAACA/G motif is particularly important for cre function, as tandem adenosine residues within it template the uridylylation of VPg by a slide-back mechanism (16, 17, 29). Thus, the fact of the ablation of the critical second adenosine residue within the motif in MutC is particularly informative. This single base change altered neither MFOLD-predicted secondary structure nor amino acid sequence within 3Dpol but nonetheless completely eliminated replication of the HAVLuc replicon (Fig. 5C). Further biochemical studies will be required to confirm a
role in VPg uridylylation for this novel HAV stem-loop structure.

Thus, taken together, our experimental data strongly suggest that the conserved RNA structure that we have identified near the 5’ end of the 3D<sup>pol</sup>-coding sequence is in fact the HAV cre. Importantly, it appears to be present in all hepatoviruses, including both human and simian HAVs, as well as the distantly related avian virus AEV (Fig. 2). However, in contrast to other recognized picornavirus cre elements, such as the enterovirus and rhinovirus cre elements, it is a significantly larger structure. The minimal functional HRV-14 cre requires no more than a top loop of 14 nt extending from a helical segment of only 9 bp in order to support RNA replication in cell-free reactions (29). Thus, only 33 nt of sequence is required to support RNA replication. In contrast, the structure that we identified in the HAV 3D<sup>pol</sup>-coding region is 110 nt in length and contains a top loop of 18 nt with a much longer stem segment comprising 35 bp, interrupted by four internal loops and two 1-nt bulges (Fig. 2). The high-resolution nuclear magnetic resonance structure of the HRV-14 cre indicates that the 14-nt loop segment adopts a specific fold that derives stability from base stacking interactions (25). It is not possible to predict whether a similar structure might be adopted by the larger loop segment of the HAV cre. While further studies will be needed to determine the minimal functional HAV cre, a 45-nt segment containing the top loop and immediately adjacent 12-bp helical segment was inadequate to support HAV RNA replication (Fig. 7). To some extent, this may reflect the low free energy on folding associated with this RNA structure and the low G+C content of hepatovirus/AEV genome sequences (38% and 42%, respectively).

We identified this RNA structure by combined phylogenetic and thermodynamic predictive strategies. We first identified a region within the protein coding sequence of the HAV genome that displayed marked suppression of synonymous variability among codons (Fig. 1B). Since a similar SSSV scan of HEV species revealed a unique low-variability signal that aligned precisely with the previously identified enterovirus cre (Fig. 1A), the low synonymous site variability signature identified within the hepatovirus genome may be similarly indicative of an equivalent cre element. This low synonymous variation within base-paired regions of the cre elements arises from the need to conserve nucleotide sequence in order to maintain both regional RNA secondary structure and the amino acid sequence of the protein encoded by the RNA. MFOLD analyses of the region identified by the SSSV scan revealed a large stem-loop structure that is conserved across members of the genus Hepatovirus and led to the mutational analysis described above. Interestingly, the HAV cre was “invisible” using standard MFOLD folding free energy scanning. There was no detectable difference between the folding free energy of the HAV coding segment containing the cre and that of sequence content order-scrambled controls (Table 1), even though it was not difficult to identify the enterovirus cre with this method. Both MFEs (per base) and the arithmetical difference between the MFEs of native and scrambled control sequences (MFEDs) are much lower for 100- and 300-nt segments centered on the HAV than for the HEV species B cre.

The reasons underlying these differences in the HAV cre and the cre elements of other picornaviruses are not clear. One consideration is that there may be specific differences between functional RNA structures in viral genomes with different G+C contents and that the HAV cre may have to be larger (compared to the G+C-rich enterovirus and rhinovirus cre elements) because of a larger proportion of A:U pairings within the duplex region of the A+U-rich HAV structure. Twenty-five of the 35 bp in the cre sequence of HAV are either A/U or G/U (Fig. 2). However, only four of the nine base pairs in the minimal HRV-14 cre are G:C pairings (29), compared to 5 of 12 in the top part of the stem in the HAV s-cre sequence that was not capable of supporting RNA replication (Fig. 7). This region of the HRV-14 and PV cre appears to be important for recognition by 3CD and 3D<sup>pol</sup>, which is important in establishing the complex that leads to uridylylation of VPg (19). The fact that a cre appears to exist within the HAV genome, as shown here, suggests that the general scheme of VPg uridylylation resulting in the protein primer for RNA synthesis that has been established for PV most likely holds true for HAV as well. However, there are likely to be important differences in the details of this process, given the remarkably different size and folding free energy of the HAV cre stem-loop structure as revealed in the studies described here, the apparent targeting of the HAV 3A (and likely 3AB) protein to mitochondrial rather than endoplasmic reticulum membranes, and differences in the biophysical properties of the VPg protein of HAV from those of other picornaviruses (27, 28).

Whatever the basis may be for the longer cre loop and overall greater size of the HAV cre, the identification of this essential replication element enhances our understanding of the molecular virology of HAV, an important but relatively ignored human pathogen. Like almost all other aspects of this fascinating virus, the HAV cre shows both significant similarities and substantial important differences in comparison to related aspects of other well-studied picornavirus genera.

ACKNOWLEDGMENTS

We are grateful for expert assistance provided by Yinghong Ma and Jeremy Yates. This work was supported in part by a grant from the National Institutes of Health, U19-AI40035. Yan Yang was supported by a McLaughlin Postdoctoral Fellowship.

REFERENCES

10128 YANG ET AL. J. VIROL.