Secondary structure of the 5′ nontranslated regions of hepatitis C virus and pestivirus genomic RNAs

Edwin A.Brown, Hangchun Zhang, Li-Hua Ping and Stanley M.Lemon
Department of Medicine, 547 Burnett-Womack CSB, CB # 7030, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7030, USA

Received July 20, 1992; Accepted September 4, 1992   EMBL accession no. Z14099

ABSTRACT

The RNA genomes of human hepatitis C virus (HCV) and the animal pestiviruses responsible for bovine viral diarrhea (BVDV) and hog cholera (HChV) have relatively lengthy 5′ nontranslated regions (5′NTRs) sharing short segments of conserved primary nucleotide sequence. The functions of these 5′NTRs are poorly understood. By comparative sequence analysis and thermodynamic modeling of the 5′NTRs of multiple BVDV and HChV strains, we developed models of the secondary structures of these RNAs. These pestiviral 5′NTRs are highly conserved structurally, despite substantial differences in their primary nucleotide sequences. The assignment of similar structures to conserved segments of primary nucleotide sequence present in the 5′NTR of HCV resulted in a model of the secondary structure of the HCV 5′NTR which was refined by determining sites at which synthetic HCV RNA was cleaved by double- and single-strand specific RNases. These studies indicate the existence of a large conserved stem-loop structure within the 3′ 200 bases of the 5′NTRs of both HCV and pestiviruses which corresponds to the ribosomal landing pad (internal ribosomal entry site) of HCV. This structure shows little relatedness to the ribosomal landing pad of hepatitis A virus, suggesting that these functionally similar structures may have evolved independently.

INTRODUCTION

Hepatitis C virus (HCV) has recently been identified as the cause of most cases of post-transfusion viral hepatitis, and the complete nucleotide sequences of several strains of HCV have been determined (1, and references therein). HCV is distantly related to both human flaviviruses (2,3), such as dengue and yellow fever virus, as well as the animal pestiviruses responsible for bovine viral diarrhea (BVDV) and hog cholera (HChV) (4). These are all relatively small, enveloped viruses containing single-stranded, messenger-sense RNA genomes 10–12 kb in length (5,6) which are similarly organized. It has been suggested that each of these three groups of viruses should be considered a distinct genus within the family Flaviviridae (7). However, the virion RNA of HCV and the pestiviruses differ from that of the flaviviruses with respect to the presence of a relatively lengthy 5′ nontranslated region (5′NTR) (345–385 bases) preceding the single large open reading frame (ORF) encoding the viral polyprotein. The 5′NTRs of HCV and the pestiviruses have been shown to share short segments of conserved primary nucleotide sequence, but the functions served by this region of the genome remain poorly characterized (7).

Lengthy 5′NTRs are also present in the RNA genomes of the picornaviruses, where they have been shown to play a critical role in controlling the initiation of cap-independent translation at 'ribosomal landing pads' (8) or 'internal ribosomal entry sites' (9). These translational control elements are complex RNA structures several hundreds of nucleotides in length, which are located at a distance from the 5′ terminus of the virion RNA. As in the picornaviruses, there are multiple potential initiation codons within the 5′NTRs of HCV and the pestiviruses which precede the AUG codon located at the 5′ end of the large ORF. Recently, Tsukiyama-Kohara et al. (10) demonstrated the presence of a ribosomal landing pad located within the 3′ 230 nucleotides of the 5′NTR of HCV which was active in vitro. This finding was particularly interesting, because there is no significant nucleotide sequence relatedness between the 5′NTR of HCV and the 5′NTR of any picornavirus. In this report, we describe the secondary structure of the 5′NTR of HCV as well as that of two different pestiviruses. We show that the ribosomal landing pad of HCV has a secondary structure which is distinct from that of any picornavirus, although we also show that the putative HCV ribosomal landing pad, like the landing pads of picornaviruses, contains a short single-stranded oligopyrimidine tract which may serve as a potential 18S ribosomal RNA binding site.

MATERIALS AND METHODS

Approach to RNA secondary structure determination

Because predictions of RNA secondary structure which are based exclusively on thermodynamic considerations have significant limitations (11), we examined the phylogenetically-related 5′NTR sequences of HChV (6,12) and BVDV (5,13) for the presence of covariant nucleotide substitutions predictive of conserved, base-paired helical RNA structures. Subsequent folding of the RNA in a computer program was then constrained to maintain the helical stuctures identified in this fashion, resulting in a model structure based on both thermodynamic as well as phylogenetic considerations. We applied this approach first to the pestiviruses, because the very high degree of nucleotide conservation within the 5′NTRs of various HCV strains (greater than 91%) does not permit the identification of sufficient numbers of covariant nucleotide substitutions, and because the extent of sequence relatedness within the 5′NTRs of HCV and the animal pestiviruses does not permit unequivocal alignment of these sequences. Once a model of the pestiviral secondary structure was obtained, this was employed as a 'scaffold' for the subsequent folding of the HCV 5′NTR sequence. This approach assumes that homologous sequences have a common function, and are likely to preserve similar higher ordered structures. The final predicted structure of the HCV 5′NTR was then validated by double- and single-strand specific nuclease analysis of synthetic HCV RNA.

Comparative sequence analysis of pestiviral 5′NTRs

The nucleotide sequences of the NADL strain of BVDV (5) and the Albert (6) and Brecia (12) strains of HChV were obtained from GenBank (accession numbers M31182, J04358, and M31768 respectively), while that of the Osloss strain of BVDV was obtained from the European patent application (13). These four sequences were aligned (Fig. 1) with the program PILEUP, a multiple sequence alignment program included in the University of Wisconsin Genetic Computer Group (UWGCG) sequence analysis software package (14). The alignment was carried out using a gap weight of 5.00 and a gap length penalty of 0.30, with minor manual adjustments. Covariant substitutions were found by a manual search through potential helical structures, looking for substitutions of one Watson–Crick base pair for another (e.g., an AU base pair in one virus strain substituting for a GC base pair in another). The presence of two or more covariant substitutions within the same potential helical segment was accepted as proof of the existence of the helix (11).

Thermodynamic model of pestivirus 5′NTR structures

Conserved helical structures identified by the comparative sequence analysis were included as folding constraints in the FOLD program of Zucker and Stiegler, which is included in the UWGCG sequence analysis programs (14). The folding energies were those of Freier (15). GU base pairs were the only non-canonical base pairs permitted in this analysis. The output of the initial RNA folding revealed other previously unrecognized covariant substitutions, thus confirming additional helical structures. Conflicts in the predicted structures of the BVDV and HChV 5′NTRs were resolved by finding the structure most compatible with all four sequences. Where no conserved structure was possible, the RNA was left single-stranded.

Modeling of the HCV 5′NTR secondary structure

The nucleotide sequence of the 5′NTR of the AG94 strain of HCV was obtained by dideoxynucleotide sequencing of polymerase chain reaction (PCR) products and cDNA clones derived from reverse transcription/PCR cloning of virus present in the serum of a seropositive hemophilic child (Zhang et al., unpublished data). Like most North American HCV strains, the 5′NTR of this strain has a very high degree of sequence identity (99.4%) with HCV-1 (7). The AG94 5′NTR was aligned with the pestiviral 5′NTRs by adding it to the above multiple sequence alignment and adjusting alignment to maximize regions of sequence identity. Regions of the AG94 5′NTR sharing significant primary nucleotide sequence identity with pestiviral 5′NTRs were assigned base-paired configurations within the HCV structure which were identical to those in the pestiviral structure.

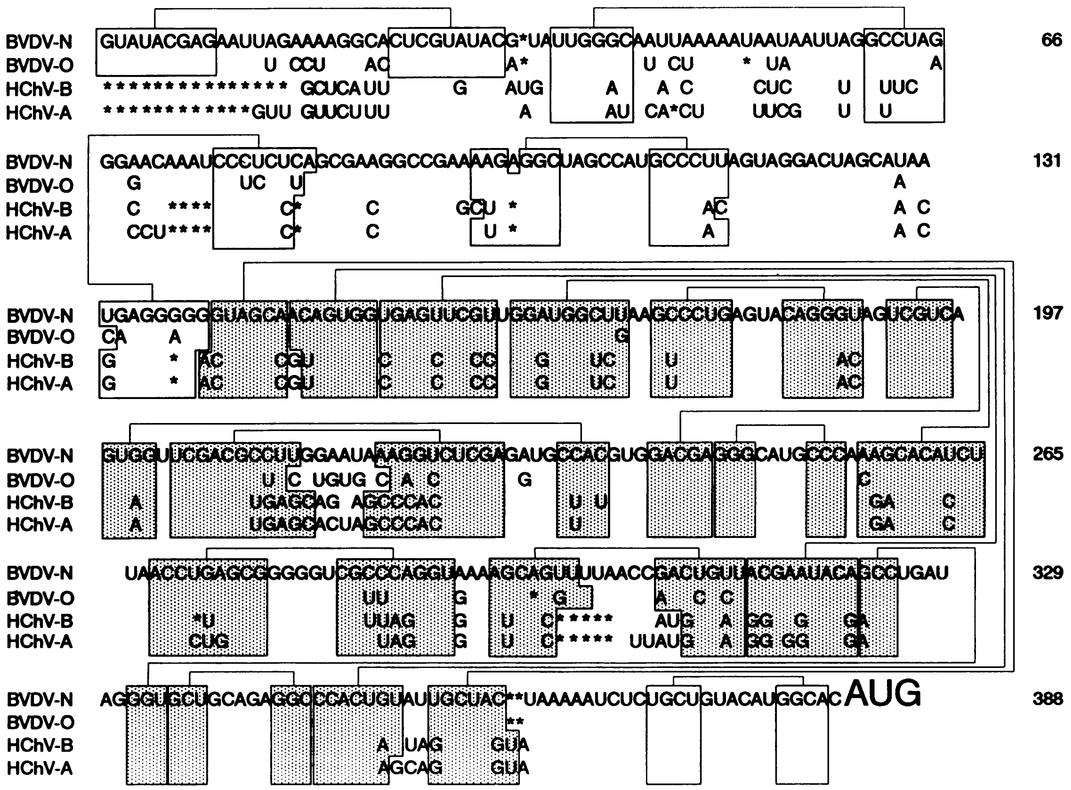
BVDV-N GUAUACGAGAAUUAGAAAAGGCACUCGUAUACG*UAUUGGGCAAUUAAAAAUAAUAAUUAGGCCUAG 66
BVDV-O U CCU AC A* U CU * UA A
HChV-B *************** GCUCA UU G AUG A A C CUC U UUC
HChV-A ************GUU GUUCU UU A AU CA*CU UUCG U U
BVDV-N GGAACAAAUCCCUCUCAGCGAAGGCCGAAAAGAGGCUAGCCAUGCCCUUAGUAGGACUAGCAUAA 131
BVDV-O G UC U A
HChV-B C **** C* C GCU * AC A C
HChV-A CCU**** C* C U * A A C
BVDV-N UGAGGGGGGUAGCAACAGUGGUGAGUUCGUUGGAUGGCUUAAGCCCUGAGUACAGGGUAGUCGUCA 197
BVDV-O CA A G
HChV-B G * AC CGU C C CC G UC U AC
HChV-A G * AC CGU C C CC G UC U AC
BVDV-N GUGGUUCGACGCCUUGGAAUAAAGGUCUCGAGAUGCCACGUGGACGAGGGCAUGCCCAAAGCACAUCU 265
BVDV-O U C UGUG C A C G C
HChV-B A UGAGCAG AGCCCAC U U GA C
HChV-A A UGAGCACUAGCCCAC U GA C
BVDV-N UAACCUGAGCGGGGGUCGCCCAGGUAAAAGCAGUUUUAACCGACUGUUACGAAUACAGCCUGAU 329
BVDV-O UU G * G A C C
HChV-B *U UUAG G U C***** AUG A GG G GA
HChV-A CUG UAG G U C***** UUAUG A GG GG GA
BVDV-N AGGGUGCUGCAGAGGCCCACUGUAUUGCUAC**UAAAAAUCUCUGCUGUACAUGGCACAUG 388
BVDV-O **
HChV-B A UAG GUA
HChV-A AGCAG GUA

Figure 1. Alignment of BVDV strains NADL (BVDV-N) and Osloss (BVDV-O) and HChV strains Brecia (HChV-B) and Albert (HChV-A) inclusive of nucleotides 1 through 388 (BVDV-N numbering). Boxes and connecting lines indicate proposed helical structures; shaded boxes are those helices contributing to domain III (see Fig. 2). (*) indicates base deletions. The initiator AUG is in large-face type.

Thus, helices sharing largely conserved primary sequences between these two viral genera were used to constrain the folding of the HCV 5′NTR by the FOLD program.

Nuclease analysis of the HCV 5′NTR secondary structure

The plasmid pHCVNTR, containing cDNA representing the 5′NTR of AG94 HCV (bases 1–399) inserted into the multiple cloning site of pGEM3/zf(−) at engineered *EcoRI* and *XbaI* sites, was restricted with *XbaI* and runoff RNA transcripts made under direction of T7 RNA polymerase (Promega). This synthetic RNA was subjected to digestion with double-strand specific V_1 (cobra venom) RNase, or single-strand specific RNases S_1, T_1 and T_2, as described previously (16). The location of specific nuclease cleavage sites was determined by polyacrylamide gel electrophoresis of primer extension products (16).

RESULTS

Secondary structure of pestiviral 5′NTRs

As the nucleotide sequences of the 5′NTR of BVDV and HChV have multiple highly conserved regions, the alignment of the pestiviral 5′NTR sequences was straightforward except at the extreme 5′ terminus (Fig. 1). However, these sequences are sufficiently divergent to permit the ready identification of numerous covariant nucleotide substitutions indicative of conserved helical structures, which were included as contraints in the computer folding of the 5′NTR by the FOLD program. The predicted secondary structures of the BVDV and HChV 5′NTRs are depicted in Fig. 2. Although the BVDV and HChV 5′NTR sequences differ at approximately 26% of base positions, the predicted secondary structures are almost identical, with each containing 4 major structural domains (Fig. 2). At the 5′ terminus, the 5′NTR of BVDV contains a conserved helical structure with an apical loop that is variable in both sequence and length between the two BVDV strains (domain I) (see Fig. 1). As suggested in Fig. 2B, it is likely that the reported HChV sequences do not include the extreme 5′ terminal nucleotides of the viral RNA (6,12), as the sequences available for the HChV strains extend only to the downstream half of this helix.

While most regions of these predicted structures are well supported by the presence of covariant nucleotide substitutions (Fig. 2A), this is not the case for all regions. Thus, although the apical segments of the second stem-loop of domain I in Figs. 2A and 2B are likely to contain additional helical structure, it

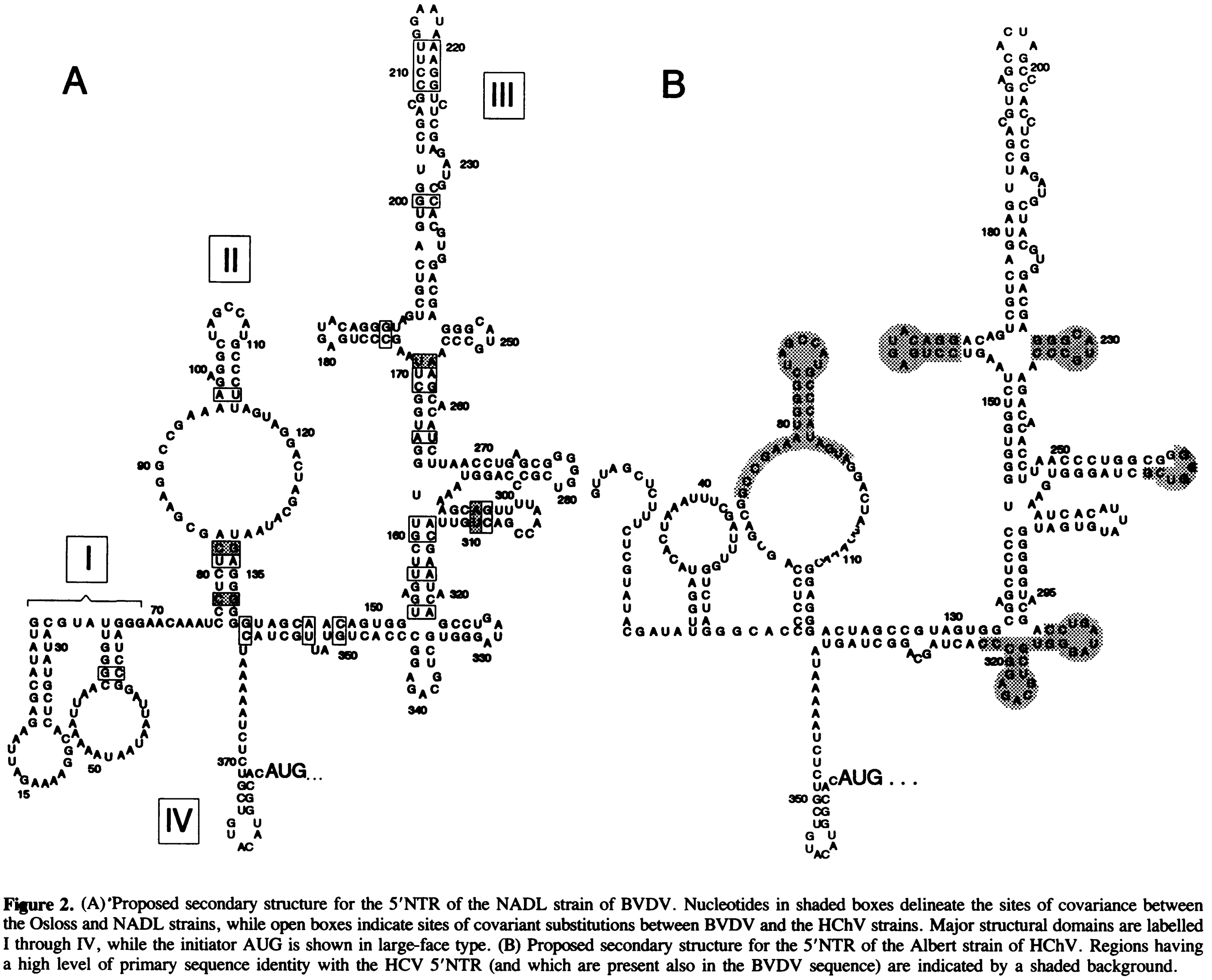

Figure 2. (A) Proposed secondary structure for the 5′NTR of the NADL strain of BVDV. Nucleotides in shaded boxes delineate the sites of covariance between the Osloss and NADL strains, while open boxes indicate sites of covariant substitutions between BVDV and the HChV strains. Major structural domains are labelled I through IV, while the initiator AUG is shown in large-face type. (B) Proposed secondary structure for the 5′NTR of the Albert strain of HChV. Regions having a high level of primary sequence identity with the HCV 5′NTR (and which are present also in the BVDV sequence) are indicated by a shaded background.

is not possible to definitively assign base pairings without confirmatory nuclease or chemical sensitivity studies. Similarly, while the base of the large stem-loop of domain II has a well defined helix containing multiple covariant mutations, the central region of this structure contains conserved sequences within which we were unable to establish a secondary structure by identification of covariant substitutions. Domain III (nucleotides 142−358) is a large complex structure consisting of a long irregular helix with multiple branching stem-loops which is well supported by numerous covariant substitutions evident in comparisons of the HChV and BVDV sequences (Fig. 2A). The 3′ terminus of the NTR (domain IV) may contain a short single stranded region followed by a small stem loop structure leading to the AUG start codon, but this region is highly conserved in all four viruses and devoid of any covariant substitutions.

Secondary structure of the 5′NTR of HCV

The secondary structure of the 5′NTR of the AG94 strain of HCV was predicted by assigning similar structural contexts to regions of the HCV 5′NTR which show significant primary nucleotide sequence identity to the pestiviral sequences (Fig. 2B) The secondary structure generated by the FOLD program under these constraints was substantially confirmed by analysis of the susceptibility of synthetic RNA to cleavage with single- and double-strand specific RNases (Fig. 3). Each site of cleavage with the single-strand specific RNases T_1, T_2 and S_2 was located either within or immediately adjacent to a predicted single-stranded region, while almost all V_1 cleavage sites were located within predicted base-paired helical structures. However, the double-strand specific RNase V_1 did cleave within a predicted single-stranded regions at bases 307−308 (Fig.3), suggesting that this region might be involved in additional secondary structure that was not identified by the combined phylogenetic/thermodynamic approach.

To further assess the validity of the HCV secondary structure model, we searched GenBank for 5′NTR sequences of HCV strains, and determined the impact of nucleotide substitutions present in these different strains on the structural model. Altogether, the partial or complete 5′NTR sequences of 81 strains of HCV were considered. Among these strains, nucleotide substitutions were identified at 59 (17.3%) base positions (data not shown). Over half of these base positions were located within predicted single-stranded regions. Substitutions occuring within predicted helices were either permissive for, or would have a minimal effect on the predicted secondary structure. Nucleotide substitutions which were located within predicted base-paired stuctures include multiple covariant substitutions within the helix formed by residues 172−191, and 206−227, some of which were noted previously by Tsukiyama-Kohara et al. (10). Thus, both nuclease digestion studies and comparative sequence analysis of a large number of HCV strains support the structural model shown in Fig. 3.

The most striking feature of this model is the conservation of the pestiviral structural domain III within the HCV structure. Although this region of the HCV 5′NTR (bases 125−323) contains several short segments that are identical in primary sequence to similar segments in the pestiviral 5′NTRs (Fig. 2A), the AG94 strain of HCV and the NADL strain of BVDV have only 47% sequence identity in this region. Despite this low level of sequence relatedness, the predicted secondary structures are virtually superimposable on each other in this region (compare Figs. 2 and 3). In contrast, the predicted secondary structures of the 5′ domains of the 5′NTRs of these viruses show little similarity, with the exception of a conserved stem-loop stucture located within domain II.

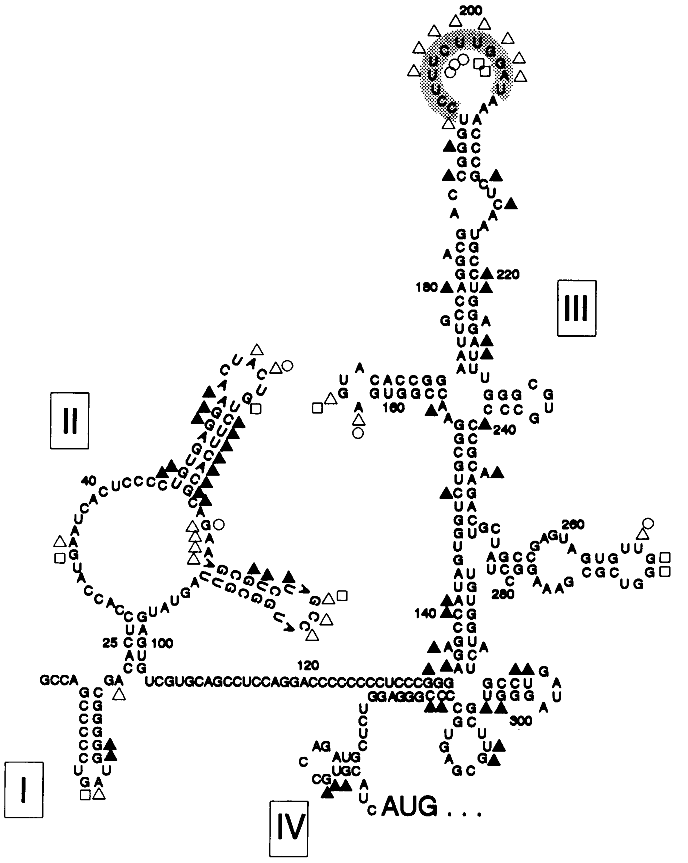

Figure 3. Proposed secondary structure of the 5′NTR of HCV (strain AG94). Sites of nuclease cleavages are indicated by symbols adjacent to individual nucleotides: □ = T_1, ○ = T_2, △ = S_1 (single-strand specific) and ▲ = V_1 (double-stranded specific). The pyrimidine-rich tract within the apical loop of domain III which is complementary to 18S ribosomal RNA is indicated by a shaded background.

DISCUSSION

In this report, we present a model of the secondary structure of the 5′NTR of HCV which is derived from a combination of phylogenetic, thermodynamic and biochemical approaches. We show that the HCV 5′NTR shares a large, conserved stem-loop structure (domain III) with the 5′NTRs of the pestiviruses, BVDV and HChV. This conserved secondary structure element is located between bases 125−323 of the HCV 5′NTR, a region approximating that shown by Tsukiyama-Kohara et al. (10) to be essential for internal ribosomal entry in bicistronic constructs containing reporter genes separated by HCV 5′NTR sequences. Although it is uncertain whether bases 103−124 and/or 324−341 are required for internal initiation, domain III appears to represent much if not all of the ribosomal landing pad of HCV. The fact that very similar structures are present in the 5′NTRs of BVDV and HChV suggests that these viruses may also initiate translation by internal ribosomal entry, although this has yet to be demonstrated. Such an hypothesis is consistent with the fact that these pestiviral 5′NTRs have multiple AUG codons preceding the initiation codon located at the 5′ end of the large ORF (in fact there are 6 such triplets in the NADL strain of BVDV).

Ribosomal landing pads that are capable of initiating translation internally were first identified within the 5′NTRs of the picornaviruses (8,9). These RNA structures have been shown to interact with a variety of cellular proteins, and specific features of these interactions appear to be important in determining the host range of individual viruses (17). Although HCV and hepatitis A virus (HAV), a picornavirus, both cause human disease due to their replication within a common cell type, the hepatocyte, the RNA sequences comprising the ribosomal landing pads of these two viruses are remarkably different. The HAV landing pad involves a large, complex series of stem-loop structures involving over 575 nucleotides (Brown et al., unpublished data), while the HCV landing pad is only approximately 200 bases in length. Furthermore, there is no structural homology evident between these two functionally similar translational control elements suggesting that each may have evolved independently of the other.

However, it is interesting to note that the 5′NTRs of HCV and HAV have two features in common. Short, single-stranded, pyrimidine-rich domains are present in both RNA structures. In the HAV 5′NTR, a large pyrimidine-rich tract extends from base 99–151, immediately upstream of the landing pad, while a shorter oligopyrimidine tract is located within the landing pad just upstream of the initiator AUG. Similarly, in the HCV 5′NTR, there is a short pyrimidine-rich region located immediately upstream of the domain III stem-loop structure, and a shorter oligopyrimidine tract located within the loop at the top of this structure. These pyrimidine-rich tracts are notably absent in the pestivirus 5′NTRs. Neither of the HCV pyrimidine-rich tracts contain the 'UUUCC' motif corresponding to the 'box A' which Pilipenko et al. (18) recently noted precedes an AUG triplet (either cryptic or initiator) by approximately 25 bases in picornaviral ribosomal landing pads. However, the apical loop of domain III contains a 'UUUCU' sequence which is located 17 nucleotides from a downstream cryptic AUG, an arrangement which is not unlike that present in the coxsackievirus A9 and rhinovirus 14 landing pads (18). Alternatively, base-pairing present near the base of the domain III structure brings the first of the HCV pyrimidine-rich tracts into close proximity with the initiator AUG, approximately 25 bases upstream of it. These regions are thus likely to be fruitful sites for future mutagenesis experiments.

Finally, short conserved regions that are complementary to 18S ribosomal RNA have been identified within the pyrimidine-rich regions of picornaviral landing pads (18). Such a potential 18S ribosomal RNA binding site is also present within the HCV structure, located within the loop at the top of domain III. This sequence, CCUUUCUUGGA (HCV bases 192–203), is conserved in all but one available HCV sequence, and is complementary to bases 461–471 of human 18S RNA. In contrast, the apical loop of domain III of the pestiviruses is highly variable in its sequence (Fig. 1). This region of the HCV RNA is single-stranded and, moreover, is very likely to be located on the surface of the three dimensional structure of the folded 5′NTR, given the fact that it was a prominent site for cleavage by each of the three single-strand specific RNases tested (Fig. 3).

REFERENCES

1. Okamoto, H., Kurai, K., Okada, S.-I., Yamamoto, K., Lizuka, H., Tanaka, T., Fukuda, S., Tsuda, F. & Mishiro, S. (1992) *Virology* **188**, 331–341.
2. Miller, R.H. & Purcell, R.H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2057–2061.
3. Choo, Q.-L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P.J., Weiner, A.J., Bradley, D.W., Kuo, G. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2451–2455.
4. Moennig, V. (1990) *Vet. Microbiol.* **23**, 35–54.
5. Collett, M.S., Larson, R., Gold, C., Strick, D., Anderson, D.K. & Purchio, A.F. (1988) *Virology* **165**, 191–199.
6. Meyers, G., Rumenapf, T. & Thiel, H.-J. (1989) *Virology* **171**, 555–567.
7. Han, J.H., Shyamala, V., Richman, K.H., Brauer, M.J., Irvine, B., Urdea, M.S., Tekamp-Olson, P., Kuo, G., Choo, Q.-L. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1711–1715.
8. Pelletier, J. & Sonenberg, N. (1988) *Nature* **334**, 320–325.
9. Jang, S.K. & Wimmer, E. (1990) *Genes Dev.* **4**, 1560–1572.
10. Tsukiyama-Kohara, K., Iizuka, N., Kohara, M. & Nomoto, A. (1992) *J. Virol.* **66**, 1476–1483.
11. James, B.D., Olsen, G.J. & Pace, N.R. (1989) *Meth. Enzymol.* **180**, 227–239.
12. Moormann, R.J.M., Warmerdam, P.A.M., van der Meer, B., Schaaper, W.M.M., Wensvoort, G. & Hulst, M.M. (1990) *Virology* **177**, 184–198.
13. Renard, A., Dina, D. & Martial, J. (1987) *European Patent Application 86870095. 6 Publication number 0208672*
14. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
15. Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. & Turner, D.H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9373–9377.
16. Brown, E.A., Day, S.P., Jansen, R.W. & Lemon, S.M. (1991) *J. Virol.* **65**, 5828–5838.
17. Agol, V.I. (1991) *Adv. Virus Res.* **40**, 103–180.
18. Pilipenko, E.V., Gmyl, A.P., Maslova, S.V., Svitkin, Y.V., Sinyakov, A.N. & Agol, V.I. (1992) *Cell* **68**, 119–131.