RNA (1995), *1*:526-537. Cambridge University Press. Printed in the USA.

An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5′ noncoding region

CHANGYU WANG,[1,3] SHU-YUN LE,[2] NAUSHAD ALI,[1] and ALEEM SIDDIQUI[1]

[1] Department of Microbiology and Program in Molecular Biology, University of Colorado Health Sciences Center, B172, Denver, Colorado 80262, USA
[2] Laboratory of Mathematical Biology, National Cancer Institute, DCBD, Frederick, Maryland 21701, USA

ABSTRACT

Translation of the human hepatitis C virus (HCV) RNA genome occurs by a mechanism known as "internal ribosome entry." This unusual strategy of translation is employed by naturally uncapped picornaviral genomic RNAs and several cellular mRNAs. A common feature of these RNAs is a relatively long 5′ noncoding region (NCR) that folds into a complex secondary structure harboring an internal ribosome entry site (IRES). Evidence derived from the use of dicistronic expression systems, combined with an extensive mutational analysis, demonstrated the presence of an IRES within the HCV 5′NCR. The results of our continued mutational analysis to map the critical structural elements of the HCV IRES has led to the identification of a pseudoknot structure upstream of the initiator AUG. The evidence presented in this study is based upon the mutational analysis of the putative pseudoknot structure. This is further substantiated by biochemical and enzymatic probing of the wild-type and mutant 5′NCR. Further, the thermodynamic calculations, based upon a modified RNAKNOT program, are consistent with the presence of a pseudoknot structure located upstream of the initiator AUG. Maintenance of this structural element is critical for internal initiation of translation. The pseudoknot structure in the 5′NCR represents a highly conserved feature of all HCV subtypes and members of the pestivirus family, including hog cholera virus and bovine viral diarrhea virus.

Keywords: hepatitis C virus; internal initiation of translation; internal ribosome entry site; pseudoknot

INTRODUCTION

Human hepatitis C virus (HCV) is a newly discovered virus that causes chronic hepatitis (Bradley et al., 1991). HCV-associated liver disease represents a major health problem worldwide (Houghton et al., 1991; Choo et al., 1992). A strong link has been demonstrated between HCV infection and the development of hepatocellular carcinoma (Tsukuma et al., 1993). Based on several criteria, including genome organization and virion properties, HCV has been classified as one of the genera of the family Falviviridae (Kato et al., 1990; Choo et al., 1991; Okamoto et al., 1991; Takamizawa et al., 1991). The 9.4-kb-long RNA genome has a plus polarity and encodes a polyprotein of 3,000 amino acids (Houghton et al., 1991). The polyprotein includes two structural proteins, capsid and envelope, and about five or more nonstructural proteins, NS1-NS5 (Houghton et al., 1991; Grakoui et al., 1993) (Fig. 1A). The 5′ noncoding region of the HCV is about 332-342 nt in length and has been predicted to fold into a complex secondary structure (Fig. 1B) (Brown et al., 1991). This sequence harbors multiple AUG codons upstream of the initiator AUG located at either 332 nt or 342 nt, depending on the HCV subtype (Houghton et al., 1991; Okamoto et al., 1991; Wang et al., 1993). Evidence accumulated largely via the use of dicistronic expression vectors suggests that translation of the HC viral genome occurs by an internal entry of ribosome into the 5′NCR (Tsukiyama-Kohara et al., 1992; Wang et al., 1993; Kettinen et al., 1994; Wang & Siddiqui, 1995). HCV genome employs a translational strategy similar to that known for picornaviruses. Picornaviral RNAs contain a relatively large 5′NCR and a *cis*-acting internal ribosome entry

Reprint requests to: Aleem Siddiqui, Department of Microbiology, B172, University of Colorado Health Sciences Center, 4200 E. 9th Avenue, Denver, Colorado 80262, USA; e-mail: siddiqui_a@defiance.hsc.colorado.edu.
[3] Present address: Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

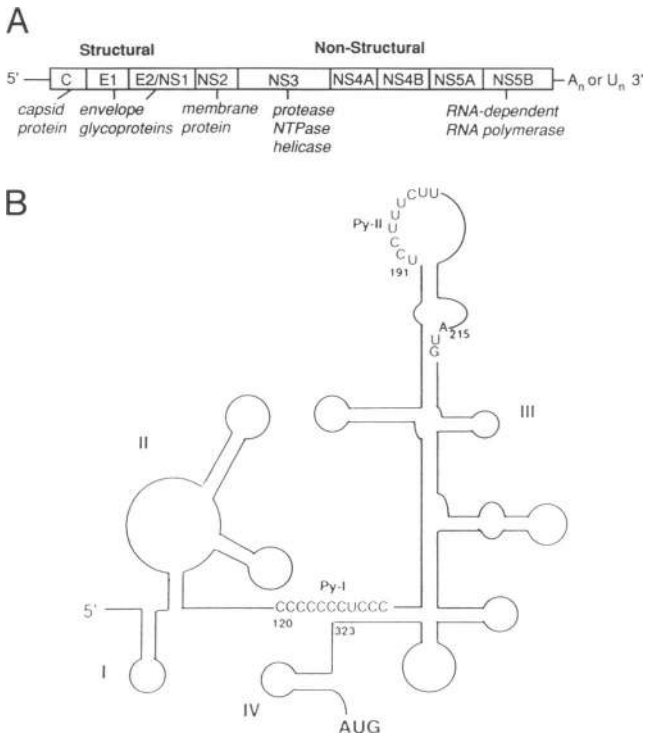

FIGURE 1. A: Genomic organization of the HCV RNA genome. **B:** Schematic map of computer-predicted RNA secondary structure of the 5'NCR (adapted from Brown et al., 1992). Sequences of the two pyrimidine tracts (Py-I, Py-II) are shown.

site (IRES) (Pelletier & Sonenberg, 1988; Jackson et al., 1990; Wimmer et al., 1993). Although the majority of eukaryotic mRNAs are translated by scanning mechanism (Kozak, 1989), IRES-mediated translation has been observed for some cellular mRNA (Macejak & Sarnow, 1991; Oh et al., 1992). Recently another member of the family Falviviridae, bovine viral diarrhea virus (BVDV), a pestivirus, has been shown to contain an IRES element within its 380-nt-long 5' noncoding region preceding the sequences that encode the viral polyprotein (Poole et al., 1995). Thus, HCV and BVDV, the two members of the family Flaviviridae, represent the viruses outside of picornavirus family that follow the novel translational scheme of internal ribosome entry.

Based on the extensive mutational analysis, the HCV IRES appears to include almost the entire 5'NCR. In mapping the limits of 5'NCR, Wang et al. (1994) showed that either insertion or deletion of nucleotides in the 3' boundary of the 5'NCR displayed inhibitory effects on translation initiation. This study further examined the functional significance of the two putative pyrimidine-rich sequence motifs (Py-I, Py-II) of the HCV 5'NCR (Fig. 1B) (Wang et al., 1994). The pyrimidine tract designated as Yn-Xm-AUG motif, represents a principal feature of the picornaviral 5'NCR that has been examined in detail (Jackson et al., 1990; Wimmer et al., 1993). Genetic analysis revealed that replacement of almost 60% of the pyrimidine residues in the Py-1 motif had no effect on translation initiation, provided the helical structure was maintained in that region. Further, this analysis led to the identification of a conserved helical structure associated with the Py-1 motif whose maintenance was shown to be essential for efficient translation. Several mutations in the second pyrimidine motif (Py-II), including the noninitiator AUG codon following the Yn-Xm motif, also had no significant effect on the translational efficiency of the HCV 5'NCR (Wang et al., 1994). In light of these results and those reported for EMCV IRES by Kaminski et al. (1994), the significance of pyrimidine-rich nature of the Yn-Xm-AUG motif in internal initiation needs to be reevaluated.

Our continued mutational studies of the various structural domains within the HCV 5'NCR has led to the identification of a pseudoknot-like structure in a region upstream of the initiator AUG codon. First, the genetic analysis revealed the possible base pairing between nt 305–311 and nt 325–331. Mutations at either sequence dramatically reduced the ability of 5'NCR to mediate internal initiation of translation. Compensatory mutations that restored base pairing in the region regained the translational efficiency to a considerable level compared to either mutant. Biochemical studies provided further evidence for the base pairing between the nucleotides in that region. Finally, the thermodynamic calculations based on a modified version of RNAKNOT program (Chen et al., 1992) predicted the presence of a significant stable RNA pseudoknot structure forming biologically relevant tertiary interactions. Furthermore, the pseudoknot structure represents a highly conserved feature of all the HCV subtypes and pestiviruses. This is the first report, to our knowledge, that provides experimental evidence for a functional role of a pseudoknot structure in internal initiation of translation.

RESULTS

The secondary structure RNA folding model of the 5'NCR of HCV RNA proposed by Brown et al. (1992) predicted that the nucleotide sequences 305–311 and 325–331 are unstructured (Fig. 2). However, a closer examination of these sequences in the HCV 5'NCR revealed that these two regions are complementary to each other. Base pairing interactions between these nucleotides and those reported by us previously between nt 126–134 and nt 315–323 (Wang et al., 1994) together reveal the potential to form a pseudoknot-like structure (Fig. 2). The helical interaction between nt 126–134 and nt 315–323, shown to be essential for internal initiation by mutational analysis, is associated with the pyrimidine tract (Py-I) (Wang et al., 1994).

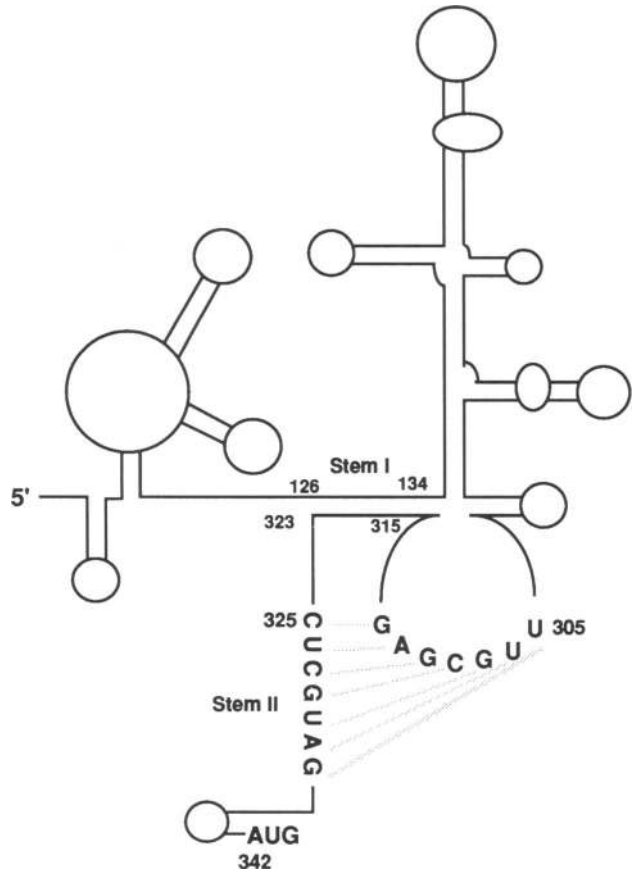

FIGURE 2. **A:** Predicted pseudoknot structure in the 5'NCR of the HCV RNA. Possible base pairing interactions between two previously proposed single-stranded regions are represented by dashed lines. Stems I and II, which constitute the pseudoknot structure, are indicated.

TABLE 1. Expression of luciferase under the control of wild-type and mutant HCV 5'NCR.

Plasmid construct	Wild-type sequence	Mutated sequence	% Luciferase activities	
			In vitro[a]	In vivo[b]
pHC5NC	–	–	100	100
pHC5NC/307	^{307}GCG^{309}	^{307}UGC^{309}	1.2	0.1
pHC5NC/327	^{327}CGU^{329}	^{327}GCA^{329}	2.3	0.5
pHC5NC/307/327	^{307}GCG^{329} ^{327}UGC$_{329}$	^{307}UGC^{309} ^{327}ACG$_{329}$	32.2	9.2

[a] RNA derived from the indicated plasmids was synthesized in vitro and translated in reticulocyte lysates. Percent luciferase activity was measured according to de Wet et al. (1987).

[b] RNA from indicated plasmids was synthesized in vitro and introduced into HepG2 cells by lipofectin method. After 6 h, cell lysates were used to measure luciferase activity according to de Wet et al. (1987). The percentage of LUC activity in the mutant samples was normalized against the wild type, which was set arbitrarily at 100. The average luciferase light units for wild-type RNA translated in vitro were 885,500 and in transfected HepG2 lysates were 30,000 light units/5 $\times$ 10^{-5} cells.

A pseudoknot may result from base pairing of nucleotides of a single-stranded loop region with complementary sequence outside this loop (Pleij, 1990, 1993). This pairing can be between any combination of single-stranded regions in hairpin loops, bulges, internal loops, junctions, or unpaired bases between secondary structure motifs (Wyatt et al., 1990; Pleij, 1993). Pseudoknots are characterized by the presence of at least two stems. In the case of HCV 5'NCR, the Py-I-associated helical structure could form stem I, whereas the suspected base pairings between nt 305–311 and nt 325–331 constitute stem II of the pseudoknot.

Mutational analysis

To provide genetic evidence for the existence of tertiary structural elements within the HCV 5'NCR and examine its functional role in translation initiation, mutagenesis studies were carried out in the predicted stem II region (Fig. 2). The effects of mutations on translation were investigated by both in vitro and in vivo translation studies. In vitro synthesized wild-type and mutant RNAs were both translated in rabbit reticulocyte lysates followed by SDS-PAGE and transfected into HepG2 cells. Luciferase activities were measured both from the reticulocyte lysates and transfected HepG2 cell lysates. First, mutations were introduced in one of the complementary strands of the putative stem II structure. Three nucleotides at nt 307–309 were altered from GCG to UGC to generate pHC5NC/307 (Fig. 3A). This mutation, which should disrupt the suspected base pairing interaction, caused a dramatic reduction in translation of the reporter gene (Fig. 3B; Table 1). Similarly, mutations (pHC5NC/327) in the other strand of the stem II structure (Fig. 3A) abolished translation initiation (Fig. 3B; Table 1). Mutant pHC5NC/327 contains three nucleotide substitutions at nt 327–329 from CGU to GCA. A compensatory mutation was constructed by combining the nucleotide substitutions of pHC5NC/307 and pHC5NC/327, and designated pHC5NC/307/327 (Fig. 3A). As shown in Figure 3 and Table 1, when base pairing interactions were restored by the compensatory mutations, translational efficiency was only partially recovered, but was 20–100-fold higher (comparison of in vitro and in vivo luciferase activity) than either mutant. The partial recovery of the translational efficiency (3–10-fold lower than wild type) by the compensatory mutation suggests that the primary sequences in the pseudoknot region may be an important component of the HCV IRES function.

The present mutational analysis of the putative stem II structure, combined with the mutational analysis of the putative stem I structure described previously (Wang et al., 1994), unambiguously point to the existence of a possible pseudoknot structure in the 3' boundary of the 5'NCR near the initiator AUG. According to this

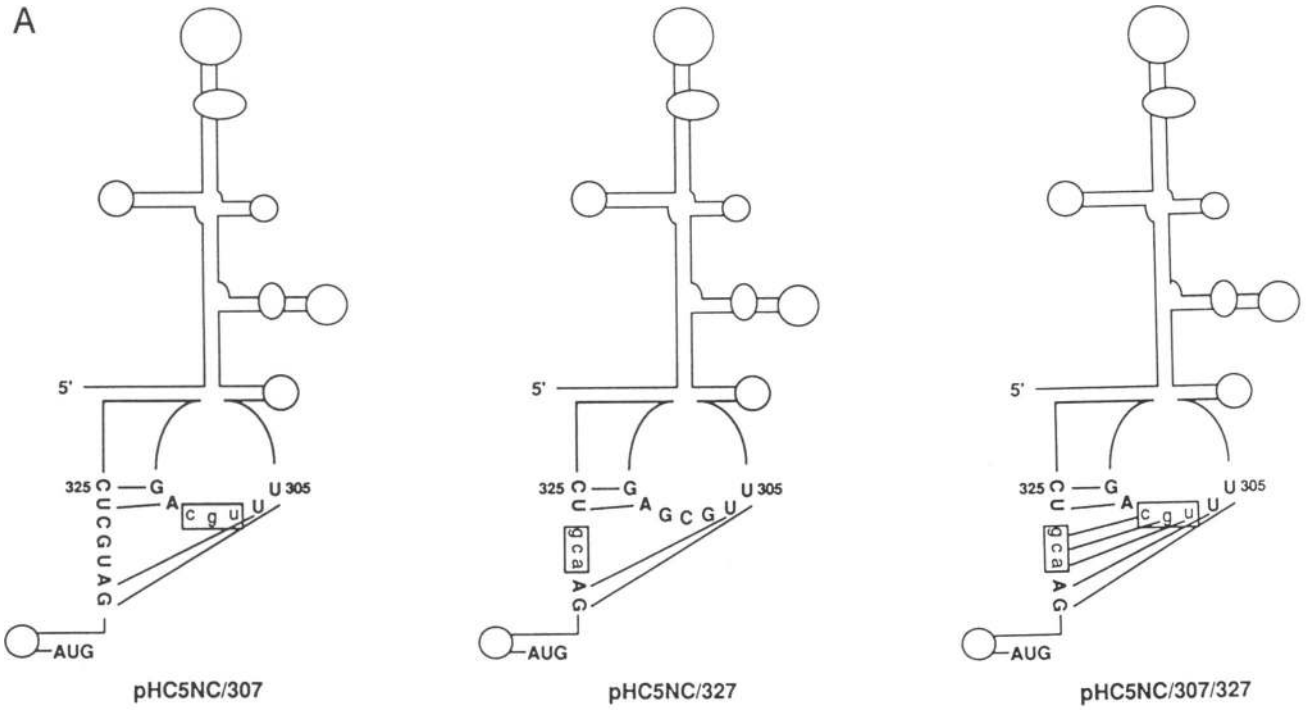

FIGURE 3. **A:** Mutagenesis in stem II structure of the suspected pseudoknot structure of the HCV 5'NCR. Nucleotide sequences of the mutated regions are presented in the context of the predicted secondary structure. Altered nucleotides are marked with open box and represented by lower case. Mutants are named according to the position of the first nucleotide altered and include pHC5NC/307, pHC5NC/327, and pHC5NC/307/327. **B:** Effects of mutations in the proposed pseudoknot structure on translation. SDS-PAGE pattern of the [^{35}S]-labeled translation products of the RNAs in rabbit reticulocyte lysates is shown. Position of luciferase protein (LUC) is indicated.

structural arrangement, the Py-I-associated helical structure (stem I) is an important part of the pseudoknot structure. Additionally, the predicted tertiary structural folding would also be consistent with RNase probing data of these regions in the 5'NCR (Brown et al., 1992). The RNase V1 cleavage patterns, according to the authors, suggest that nt 305–306 are involved in base pairing interactions. However, their thermodynamic calculations did not reveal this tertiary interaction. The proposed pseudoknot model presented here would be consistent with their speculation.

Enzymatic and chemical probing of the HCV RNA

To provide additional biochemical evidence in support of the presence of the predicted pseudoknot structure, RNA structural analysis of the 5'NCR of the HCV RNA was performed by chemical and enzymatic probing of a 350-nt transcript. This transcript contains nt 1–350 of the viral 5'-end sequence. The structural analysis was focused on the 3' boundary of the 5'NCR that contains the nucleotides involved in forming the proposed pseudoknot.

The structure of the transcripts was determined using a chemical and several enzymes that react selectively with the single-stranded or the helical regions in RNA (Christiansen & Garrett, 1988; Stern et al., 1988). In this analysis, the reverse transcriptase activity terminates when it encounters an RNase-directed cleavage or a DMS-directed base modification in the RNA. Single-stranded regions were probed for reactivity to DMS, RNase A, RNase T1, and RNase PhyM. The RNase A cleaves specifically at C and U, RNase T1 at G, RNase PhyM at A and U residues. DMS methylates adenine at the N1 position, cytosine at N3, and guanine at N7, but the latter does not terminate reverse transcription. Helical regions were probed for reactivity to cobra venom nuclease V1. This nuclease predominantly recognizes and cleaves within base paired RNA. Due to the nature of these modifications, primer-extended DNA fragments by reverse transcriptase will stop one base short of the corresponding reactive residue as shown in Figure 4B. However, RNase V1 cleavages yield primer extension products that will comigrate with the same residue as seen on the gel (Fig. 4B). The reactivity of the nucleotides to reagents, specific for single-stranded or helical RNA, is indicated in Figure 4A. The data are consistent with the proposed pseudoknot structure model of the HCV 5'NCR. Residues C327 U326 C325, A310 G309 C308, and U306 were cleaved by the double-stranded-specific RNase V1, indicating that these nucleotides are involved in base pairing interactions, which is predicted from the tertiary structure model (Fig. 2). This stretch of nucleotides constitutes stem II of the pseudoknot. The presence of stem I was confirmed by the V1 cleavages of the residues C317 C318 and by the extensive dissociation of the reverse transcriptase from the templates at residues G323 G322 A321 G320 G319 G318, presumably due to the highly stable helical structure in that region. Of particular interest were the nucleotides in the region of nt 305–311, which participate in the formation of stem II, that were sensitive to both single-stranded- and helix-sensitive probes. Residues A310 G309 C308 and U306 in this region were cleaved by double-stranded-specific RNase V1, whereas residues G307 and G309, and A310 were recognized by single-stranded-specific RNase T1 and DMS, respectively. Reactivity of RNA sequences to both single-stranded- and helix-sensitive probes has been documented in a number of RNA structure studies and was attributed either to base stacking of the reactive nucleotides within an unpaired region or to tertiary interactions of the nucleotides (Puglisi et al., 1988; Wyatt et al., 1990; Wassarman & Steitz, 1991; Jacobson et al., 1993). The dual reactivity of this region may indicate the presence of a tertiary structure that is in equilibrium with the stem-loop secondary structure.

In support of the proposed pseudoknot in the HCV 5'NCR, structural analysis of the HCV 5'NCR with specific mutations, including the compensatory mutations in stem II, was carried out. The cleavage patterns of the mutant 5'NCRs with both double-stranded-specific RNase V1 and single-stranded-specific RNase T1 is presented in Figure 4C. In this analysis, the focus of RNA structure is stem II, in the region of nt 325–331. These nucleotide sequences are proposed to participate in forming the stem II structure of the putative pseudoknot by base pairing with nt 305–311 (Figs. 2, 4A). The presence of the base pairing interactions was again suggested by the helix-specific RNase V1 cleavage at residues U329 A330 G331 in the wild-type HCV 5'NCR (pHC5NC). However, in the mutant pHC5NC/307, the stem II structure is disrupted, as shown by the lack of helix-specific RNase V1 cleavage and the presence of single-stranded-specific RNase T1 cleavage at residue G328 of the nt 325–331 region. This result supports potential base pairing interactions between nt 307–309 and 325–329, as proposed in the current pseudoknot model, because the nucleotide changes at nt 307–309 resulted in the structural alterations at nt 325–329 (Fig. 4C). Furthermore, the compensatory mutation (pHC5NC/307/327) that is expected to restore the proposed stem II structure displayed cleavage patterns in this region (nt 325–331) that were similar to wild-type 5'NCR by both RNase V1 and RNase T1 probing. Surprisingly, the RNase V1 cleavage pattern of the mutant pHC5NC/327 (lane 2), which should disrupt the stem II structure, was similar to the compensatory mutant (lane 4) in the region of 325–331. However, careful examination reveals that the CGU to GCA change at nt 327–329 in the mutant RNA could potentially assume alternate base pairing interactions between nt 326–331 and nt 304–309. This newly generated base pairing interaction could explain the RNase V1 cleavages observed here.

Finally, the structural analyses of the wild-type and mutant 5'NCRs detailed here (Fig. 4B,C) conclusively support the proposed pseudoknot structure in the region.

Thermodynamic calculations

The unusual folding region (UFR) in the 5'NCR of HCV and two related pestiviruses, hog cholera virus (HoCV) and BVDV, were detected by extensive Monte Carlo simulations. The thermodynamically favored helical stems in these UFRs were computed by EFFOLD (Le et al., 1993). These computed helical stems were compiled and verified by inspecting their conservation and compensatory base changes among divergent HCV subtypes, HoCV and BVDV (Figs. 5, 6). Based on these conserved helical stems in HCV and pestiviruses, the possible pseudoknotting interactions were searched thoroughly by the RNAKNOT program (Chen et al., 1992). The proposed RNA pseudoknot in HCV-1 (Fig. 5A) is highly significant. The three significance scores, n_1, n_2, and z in the predicted pseudoknot are 0, 42, and

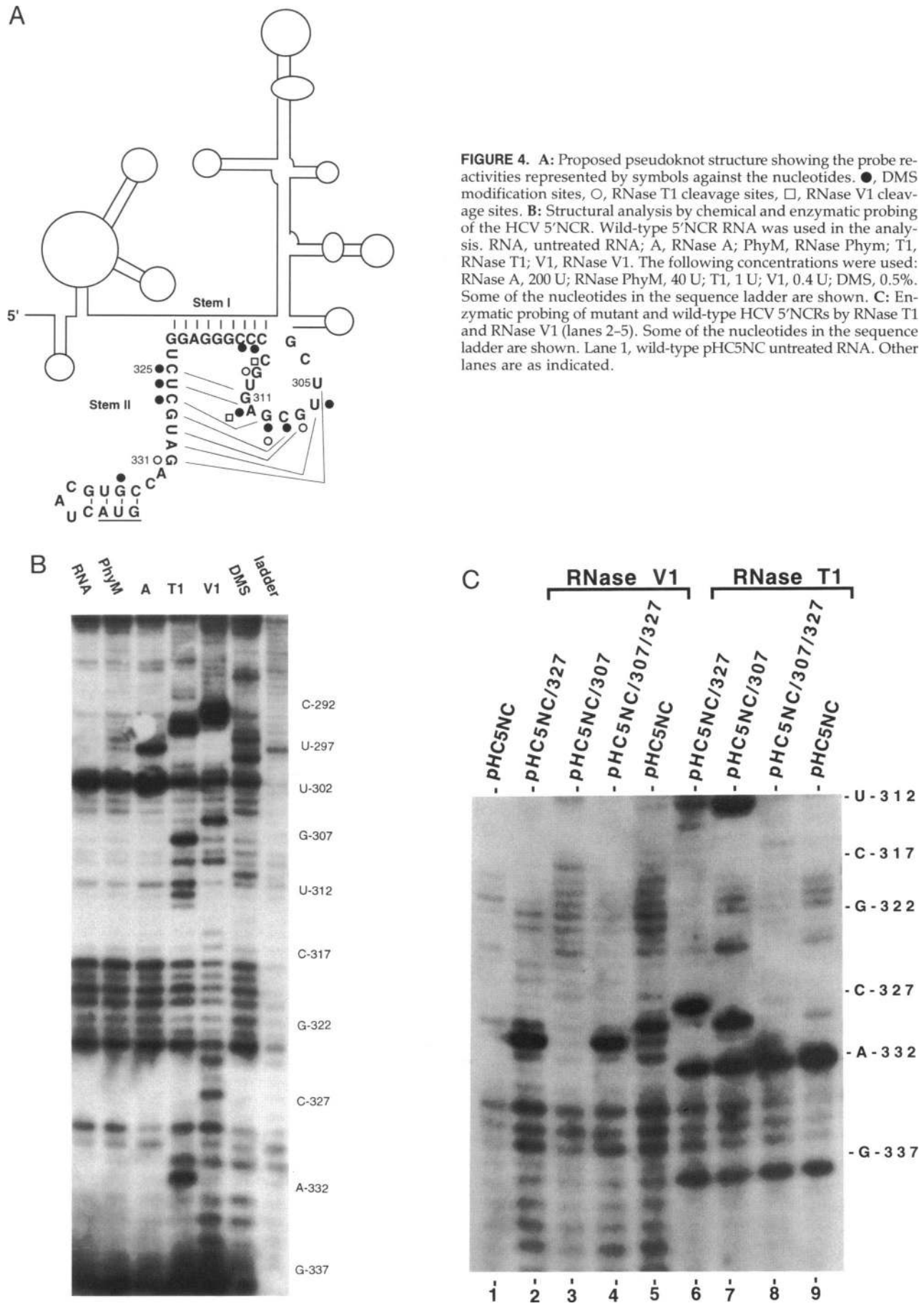

FIGURE 4. **A:** Proposed pseudoknot structure showing the probe reactivities represented by symbols against the nucleotides. ●, DMS modification sites, ○, RNase T1 cleavage sites, □, RNase V1 cleavage sites. **B:** Structural analysis by chemical and enzymatic probing of the HCV 5′NCR. Wild-type 5′NCR RNA was used in the analysis. RNA, untreated RNA; A, RNase A; PhyM, RNase Phym; T1, RNase T1; V1, RNase V1. The following concentrations were used: RNase A, 200 U; RNase PhyM, 40 U; T1, 1 U; V1, 0.4 U; DMS, 0.5%. Some of the nucleotides in the sequence ladder are shown. **C:** Enzymatic probing of mutant and wild-type HCV 5′NCRs by RNase T1 and RNase V1 (lanes 2–5). Some of the nucleotides in the sequence ladder are shown. Lane 1, wild-type pHC5NC untreated RNA. Other lanes are as indicated.

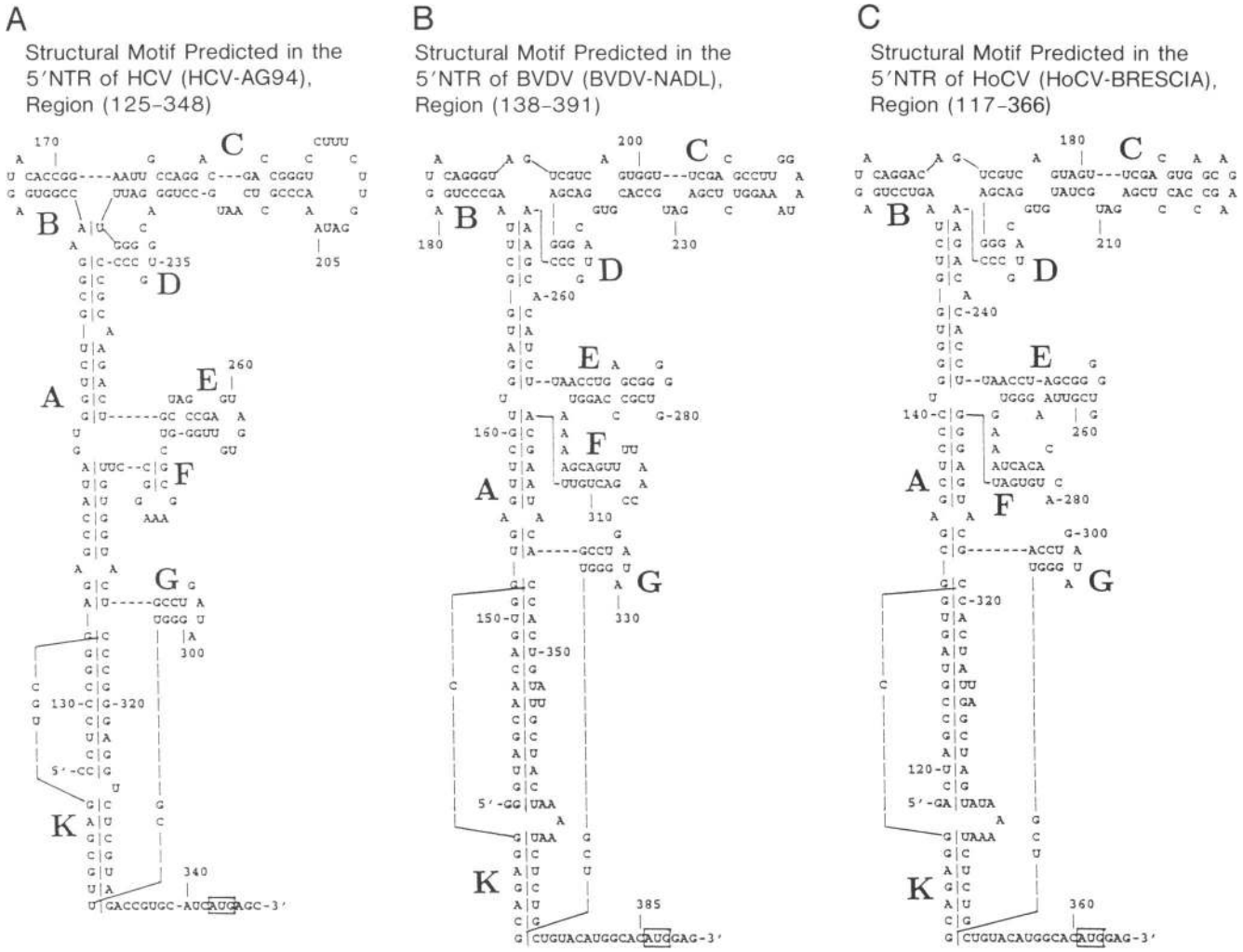

FIGURE 5. Predicted tertiary structures of the 3' border of the 5'NCR of HCV, HoCV, and BVDV. **A:** HCV-AG94 (nt 125–348) (Brown et al., 1992). **B:** BVDV-NADL (nt 138–391). **C:** HoCV-Brescia (nt 117–366). The predicted pseudoknot in the model is denoted by the letter K and stem-loops are indicated from A to G. The AUG start codon of the viral ORF is indicated in a box. Note that the HCV-AG94 contains the AUG codon at nt 343 as opposed to 342 for the HCV-1 5'NCR.

18.22, respectively, using 1,000 randomly shuffled sequences. The extremely large value of 18.22 (z) and/or small values of n_1 or n_2 indicates the nonrandomness of the occurrence of the pseudoknot in the HCV 5'NCR. The predicted pseudoknot in HCV-1 is conserved in the divergent HCV subtypes and two related pestiviruses, HoCV and BVDV. Three compensatory base changes in the pseudoknotting were observed among HCV, HoCV, and BVDV (Fig. 5). Among them, U306–G332, U307–A331, and C309–G329 of HCV-AG94 of HCV are covariantly changed to G-C, C-G, and G-C in HoCV and BVDV, respectively. Moreover, all helical stems in the proposed RNA higher-order structural models of HCV, HoCV, and BVDV are supported by compensatory base changes, except for two small stems D and G, which are highly conserved among these viruses (Fig. 5). These measures provide strong evidence for the existence of the pseudoknot in the RNA molecules of these viruses.

Thermodynamic calculations based on EFFOLD, SEGFOLD, and RNAKNOT (Chen et al., 1992; Le et al., 1992, 1993, 1994) show that the predicted pseudoknot represents a highly conserved feature among divergent HCV subtypes and other related pestiviruses such as HoCV and BVDV (Figs. 5, 6).

DISCUSSION

Translation of the HCV RNA genome is regulated by an IRES element that resides in the 5'NCR. This study has identified an RNA pseudoknot structure as a functionally important determinant of the HCV IRES. The experimental evidence presented here is based on genetic, biochemical, and thermodynamic calculations. These analyses together provide the evidence for the existence of an RNA pseudoknot structure within the HCV 5'NCR and implicate a functional role of this tertiary interaction in internal initiation of trans-

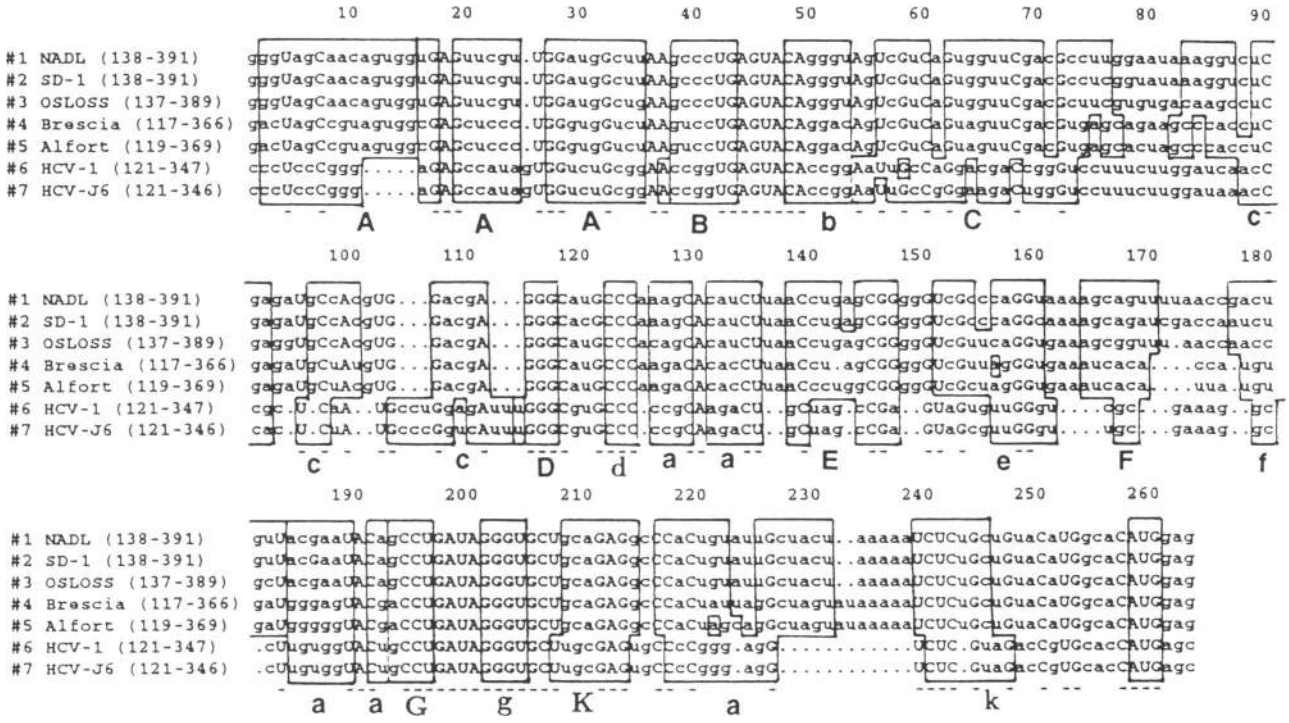

FIGURE 6. Alignment of unusual folding region (UFR) in the 3' border of the 5'NCR of several HCV subtypes and pestivirus RNAs. Because the nucleotide sequences of all the 5'NCR of HCV are highly conserved, HCV-1 sequences were chosen to represent class I and HCV-J6, class II. Deletions are denoted by dots. Conserved base pairings are boxed. Nucleotides shown within a box that are enclosed by another box represent nucleotides that fail to form an appropriate base pair. Conserved and nonconserved nucleotides in the alignment are indented by upper and lower case letters, respectively. Conserved elements within pestiviruses and HCV are indicated by boxes A–G and a–g. The proposed pseudoknot in the RNAs are indicated by K and k. Boxed AUG indicates the initiator codon.

lation. The pseudoknotting in the region includes a highly conserved helical structure associated with the pyrimidine tract (Py-I) described previously (Wang et al., 1994). This structure constitutes stem I of the pseudoknot. Point mutations in each of the complementary strands of the second putative stem (stem II) within the probable pseudoknot structure dramatically reduced translation initiation (Fig. 3; Table 1). Although the compensatory mutation in stem II only partially restored the translational efficiency, this value was considerably higher than either mutant (Fig. 3B). The partial recovery of the translation (10–30% of the wild type) by the compensatory mutation indicates that the primary sequences in the mutated regions are important. The lower level of translational recovery could have also resulted from an A-U base pair in the compensatory mutant instead of the G-U base pair found in the wild-type 5'NCR. This would be consistent with the notion that primary sequences may play an important functional role in the translation initiation. Further, the close proximity of this region to the initiator AUG may make this region highly sensitive to base substitutions. Thus, it is likely that the compensatory mutant may have assumed an alternate conformation unable to fully restore wild-type levels of translation initiation.

The results of the chemical and enzymatic probing of the RNA structure, together with the genetic data, point to the existence of an RNA pseudoknot structure within the 5'NCR near the initiator AUG. Similar pseudoknot structure can be predicted in the 5'NCR of all the HCV subtypes thus far sequenced and members of the pestiviruses (Figs. 5, 6). Pestivirus BVDV 5'NCR has been recently shown to contain an IRES element (Poole et al., 1995).

Pseudoknot structures are recognized in diverse RNA molecules and play different roles. These functions include: translation regulation (Tang & Draper, 1989; Leathers et al., 1993); ribosomal frameshifting, producing efficient cleavage sites in group I introns and hepatitis delta virus (Pleij, 1990; Brierley et al., 1991; Perrota & Bean, 1991; Willis et al., 1991; Chamorro et al., 1992; Somogyi et al., 1993); and protein recognition (Shamoo et al., 1993). During ribosomal frameshifting, a pseudoknot structure functions to stall ribosomes engaged in translating the RNA (Brierley et al., 1991; Willis et al., 1991; Chamorro et al., 1992; Tu et al., 1992; Somogyi et al., 1993). This pausing increases the efficiency of ribosomal frameshifting. In the case of bacteriophage T4 gene 32 mRNA, the pseudoknot serves an autoregulatory function, being recognized by gene 32 protein product, and interferes with

further translation of that gene (Shamoo et al., 1993). These observations suggest that pseudoknots are recognized by the components of translational apparatus (Schimmel, 1989). In light of these roles, one possible function for the pseudoknot structures in the HCV 5'NCR could be to serve as efficient binding sites for ribosomes and/or relevant initiation factors involved in internal initiation of translation.

The common RNA-folding higher-order structures, including pseudoknots, among all the HCV subtypes and related pestiviruses were computed by a combination of thermodynamic, phylogenetic, and statistical methods. The predicted conserved RNA pseudoknots among these divergent viruses strengthen the genetic and biochemical evidence presented in this study. The proposed RNA pseudoknot is not a typical H-type pseudoknot structure. One of the loops is too large to cross over any RNA groove. Most of the nucleotide sequences in the loop are folded to form a highly stable and conserved RNA secondary structure and are outside the groove. The extensive Monte Carlo simulations suggest that the proposed common RNA structural motif in the 5'NCR of these viruses is a highly stable and statistically significant structural element, which implies a structural role for the primary sequence.

Interestingly, there is a striking resemblance between the proposed structural motif of the UFR in the HCV/pestiviral 5'NCRs in this region and the structure of the group I intron (helical stem P3 to stem P7) (Cech, 1993). The HCV pseudoknot K corresponds to P7, stem A to P4 and P5, stem G to P6, and the long stem folded between the nucleotides of 5' and 3' end of HCV to helical stem P3 of group I intron (Fig. 5).

From the genetic and biochemical data presented here, it may be concluded that these tertiary structural elements of the 5'NCR may constitute an important component of the HCV IRES. However, it should be emphasized that the maintenance of the pseudoknot structure alone may not be sufficient for HCV IRES function, but other sequences and relevant structures of the 5'NCR may play an equally important role in translation. For instance, the functional role of domain III structure was demonstrated by the deletion mutation (T7CΔ152–278), which abolished translation (Wang et al., 1994). In this mutant, the pseudoknot structure should be intact, whereas the major portion of the domain III has been deleted.

Based on the secondary structural models, it was predicted that the region surrounding the pyrimidine tract is unstructured in picornavirus 5'NCR (Pilipenko et al., 1989; Skinner et al., 1989; Brown et al., 1991; Le at al., 1992). However, it has been proposed recently that the pyrimidine tract could be involved in forming some tertiary structures such as pseudoknots (Le et al., 1992, 1993). Such higher-order structures may be important components of the picornaviral IRES elements. Of interest in this respect is the mutational study of the poliovirus IRES by Pilipenko et al. (1992), in which deletion of most of the sequences in the pyrimidine tract, while retaining only five nucleotides, had no major effect on translation. Of the five nucleotides, three uridine residues were predicted to be involved in base pairing interactions that have been implicated in pseudoknot structure according to the new RNA folding model of the poliovirus 5'NCR (Le et al., 1992). Thus, it is possible that the inhibitory effect of mutations in the pyrimidine tract of picornavirus 5'NCR may be due to structural perturbances caused by the mutations rather than changes in the primary sequence. These mutations in the pyrimidine tract could have disrupted the tertiary structure, leading to the loss of IRES function. Further mutational studies in the relevant sequences, predicted to be involved in the tertiary structures, are needed to substantiate the model. Tertiary structure has also been predicted in the 5'NCR of the infectious bronchitis virus RNA 3, which is capable of directing internal initiation of translation (Le et al., 1994). A pseudoknot-like structure was identified in the 3' NCR of poliovirus (Jacobson et al., 1993). The present functional analysis of the HCV 5'NCR supported by the genetic and biochemical data and substantiated by thermodynamic calculations provides the first demonstration of the role of an RNA pseudoknot in internal initiation of translation. This superstructure, which also includes the pyrimidine tract-I-related helical secondary structure, constitutes the functional component of the HCV IRES. An alternate model of the HCV and pestiviral IRES elements has been predicted recently by Le et al. (1995), which appears to be slightly different from that proposed here. Although those predictions were based only on thermodynamic calculations, the present data is supported by experimental evidence and substantiated by UFR RNA folding computed by a modified version of RNAKNOT (Chen et al., 1992). At present we have no explanation for this discrepancy.

Based on several criteria, the picornavirus IRES elements are classified into two groups (Jackson et al., 1990; Wimmer et al., 1993). In group I IRES (enteroviruses, rhinoviruses), the ribosomes bind to the IRES element, which includes a noninitiator AUG triplet, and scan until a proper AUG triplet is encountered to initiate translation. In the case of group II IRES (cardioviruses and aphthoviruses), the ribosome binds to the IRES element, which includes an initiator AUG triplet, and translation is initiated without scanning. In both cases, a proper spacing between an AUG triplet and certain upstream motifs is essential for the IRES function. Similarly, in the case of HCV IRES, a spacing requirement between the initiator AUG and an upstream motif in the 5'NCR appears to be essential, as suggested by the mutagenesis studies (Wang et al., 1993, 1994; Wang & Siddiqui, 1995). It is likely, although not yet proven, that the initiator AUG may be involved in

the HCV IRES function, similar to the group II IRES elements. The HCV IRES element shares yet another feature with the group II IRES elements in that both can direct internal initiation of translation efficiently in vitro in reticulocyte lysates.

Finally, the present study identifies a higher-order superstructure in the 5'NCR in the vicinity of the initiator AUG codon, which by functional analysis appears to constitute an essential component of HCV IRES. Conservation of this superstructure in other members of pestiviruses lends support to the importance of these interactions in internal initiation of translation.

MATERIALS AND METHODS

Recombinant plasmid construction and mutagenesis

The construction, manipulation, and growth of recombinant plasmids was carried out by standard methods. The 5' noncoding region of an HCV isolate from Denver patient was cloned in pGEM-4 immediately preceding the luciferase reporter gene and designated T7C1-341luc (Wang et al., 1993, 1994). Site-directed mutagenesis was carried out by Altered Sites kit (Promega) at defined sites within the 5'NCR. Plasmid pHC5NC was developed by cloning a large *Hin*d III-*Hpa* I fragment, including the 5'NCR linked to luciferase gene at *Hin*d III-*Sma* I-sites in the phagemid vector, pAlter-1. Mutations in the 5'NCR were confirmed by double-stranded DNA sequencing by dideoxynucleotide method. Mutants are named according to the first nucleotide position of each mutated sequence. For example, in plasmid PHC5NC/327, the mutated sequence starts at nt 327. The map positions of the mutants are also described in the context of secondary structure (Fig. 3A).

In vitro transcription and translation

Plasmid DNAs were purified by a standard CsCl density gradient centrifugation method. Plasmid pHC5NCluc and its mutated derivatives were linearized by *Bam*H I digestion and transcribed in vitro with SP6 polymerase (Promega). All the RNAs used for translation were uncapped. In vitro translations were carried out in rabbit reticulocyte lysates (Promega). Proteins were labeled with *trans*^{35}S-methionine (ICN Biomedicals) and fractionated by 10% SDS-PAGE.

RNA transfection

A liver-derived human hepatoma cell line HepG2 was used in this study for direct RNA transfection experiments. In vitro-synthesized RNAs were introduced into cultured cells with Lipofectin (GIBCO/BRL) as described previously (Wang et al., 1993, 1994). Nearly confluent cell monolayers in 60-mm-diameter dishes were transfected with 5–10 μg of RNA-mixed with 30 μg of Lipofectin in 3 mL of Opti-MEM I (GIBCO/BRL). Five to six hours after transfection, cell lysates were prepared and assayed for luciferase activity as described by de Wet et al. (1987) with a luminometer (Mode Optocomp I; MGM Instruments, Inc.).

Chemical and enzymatic probing of RNA

RNA structure analysis was carried out according to Krol and Carbon (1989), using both single- and double-stranded-specific RNases and single-stranded-specific chemical reagent dimethyl sulfate (DMS). The full-length 5'NCR RNA was generated by in vitro transcription of the plasmid DNA pGEM5NC using T7 RNA polymerase after it was linearized with *Eco*R I. Plasmid pGEM5NC contains the RT-PCR generated full-length 5'NCR of HCV (HCV-1 strain) that was cloned into *Sma* I and *Sac* I sites of the pGEM4 vector (Wang et al., 1993). For enzymatic probing, 50 μL reaction mixture contained 1–2 μg RNA, 10 μg yeast tRNA, and following enzymes at the indicated concentrations, respectively: 40–100 U RNase PhyM, or 100–400 U RNase A, or 0.4–1 U RNase T1, or 0.1–0.4 U RNase V1 (all the enzymes were purchased from USB) in buffer A (10 mM Tris-HCl, pH 7.5, 10 mM MgCl2, 50 mM KCl). The RNA was digested for 15 min at room temperature for RNase A, T1, V1, or 30 min for RNase PhyM. Reactions were stopped by addition of phenol/chloroform. The mixture was extracted with phenol twice and precipitated in ethanol in the presence of 0.3 M sodium acetate and 10 μg of tRNA. The pellet was dissolved in 10 μL dH2O.

For chemical probing of the RNA, 1 μL DMS (Aldrich) was added to 200 μL buffer B (100 mM Hepes, pH 7.5, 10 mM MgCl2, 50 mM KCl) containing RNA and incubated at room temperature for 5 min. The reaction was precipitated in ethanol in the presence of 0.3 M sodium acetate and 10 mg tRNA. RNA pellet was dissolved in 10 μL dH_2O.

RNase cleavage and DMS modification sites were identified by primer extension with a 5' end-labeled antisense deoxyoligonucleotide primer, 5'-CGAGCTCATGATG-3', which is complementary to nt 350–338 of the HCV RNA genome. The labeling of the primer was performed in a total volume of 10 μL containing standard kinase buffer, 100 pM primer, 5 μL [γ-^{32}P]-dATP (166 mCi/mL) (ICN Biomedicals), 10 U T4 polynucleotide kinase (Promega). The reaction mixtures were incubated at 37 °C for 1 h. After the addition of 10 mg tRNA, the reaction mixture was extracted with phenol/chloroform and precipitated in ethanol. Final pellet of RNA was dissolved in 100 μL dH2O. The hybridization of a ^{32}P-labeled primer with the RNA samples was performed in a total volume of 5 μL in a hybridization buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl), including 1 μL RNA and 1 μL primer. The oligonucleotide primer described above was used for structural probing and represented sequences beginning at nt 342 (5') and extended upstream 3' in the antisense orientation up to nt 328. The solution was heated for 4 min at 85 °C, slowly cooled to 35 °C over 30–40 min, and then placed on ice for 5 min. The annealed primer was extended by the addition of 2 μL 5× RT buffer, 10 U RNasin, 0.125 mM dNTPs, 10 U reverse transcriptase (Promega), and incubated at 37 C for 30 min. Reaction was stopped by the addition of 7 μL sequencing loading buffer, and an aliquot of 4 μL was electrophoresed in an 8% acrylamide/7 M urea gel. Gels were dried and autoradiographed.

Structural analysis of the mutant HCV 5'NCRs was also carried out (Fig. 4C). Mutated HCV 5'NCR-containing con-

structs, pHC5NC/307, pHC5NC/327, and pHC5NC/307/327, were digested with *Xba* I within the luciferase-encoding region (at +48 position) and RNAs were generated subsequently by in vitro transcription with SP6 RNA polymerase. As a control, WILd-type HCV 5'NCR-containing construct pHC5NC was also digested with *Xba* I and used for an RNA control without enzymatic treatment. Structural probing of these RNAs was performed with both the double-stranded-specific RNase V1 and the single-stranded-specific RNase T1, respectively, as described above. RNase cleavage sites were identified by primer extension with 5' end-labeled antisense oligonucleotide primer, 5'-TATGTTTTTGGCGTC, which is complementary to nt +21 to +7 of the luciferase-encoding region.

Thermodynamic calculations of RNA higher-order structure

Prediction of RNA structures included the use of the following methods, EFFOLD, SEGFOLD, and RNAKNOT. First, the UFR in the 5'NCR of HCV, HoCV, and BVDV were searched for a 500-nt sequence (including partial coding sequences) from the 5' end by a statistical method, SEGFOLD (Le et al., 1993). In SEGFOLD, the significance of an RNA folding is evaluated by comparing the predicted thermodynamic stability of the actual segment with those of its random permutations and other possible foldings in the sequence. A region that is both highly stable (lower stability score as compared to other possible foldings) and more significant (lower significance score related to the random shuffling segments) is referred to as a UFR and is assumed to have a significant folding form implying a structural role for its sequence information.

The higher-order RNA structures, folded in the UFRs detected in the HCV and related pestiviruses, were then computed by EFFOLD (Le et al., 1992, 1993, 1994). The program EFFOLD enables the prediction of alternative RNA secondary structures by fluctuating thermodynamic energy parameters in the calculation of RNA folding. If a potential pseudoknotting or tertiary interaction was detected, it was assessed by a modified version of the original RNAKNOT program (Chen et al., 1992) by means of three significance scores, n_1, n_2, and z (Le et al., 1992). The scores n_1 and n_2 were defined as the numbers of randomized sequences that had a pseudoknot thermodynamically more stable than the real sequence. The smaller the n_1 and n_2, the more significant the tertiary interaction. The z score was defined as $z = (n\text{obs} - r\text{mean})/\text{SD}$ in the test, where $n$obs is the number of times the tertiary interaction occurs in the real sequence, $r$mean is the average number of times the occurrence of the folding pattern in the random permutation sequences, and SD is the standard deviation. Finally, the predicted RNA folding was verified by examining their conservation and compensatory base changes among HCV subtypes and other pestiviruses (Fig. 6).

ACKNOWLEDGMENTS

A.S. received grant support from Lucille P. Markey Charitable trust and from Colorado Institute for Research in Biotechnology. C.W. was supported by PHS grant 5 P30 DK34914. N.A. received support from Colorado Advanced Technology Institute.

Received May 8, 1995; returned for revision June 28, 1995; revised manuscript received July 14, 1995

REFERENCES

Bradley DW, Krawczynski K, Beach MJ, Purdy MA. 1991. Non-A, non-B hepatitis: Toward the discovery of hepatitis C and E viruses. *Semin Liver Dis 11*:128–146.

Brierley IN, Rolley J, Jenner AJ, Inglis SC. 1991. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J Mol Biol 220*:889–902.

Brown EA, Day SP, Jansen RW, Lemon SM. 1991. The 5' nontranslated region of hepatitis A virus RNA: Secondary structure and elements required for translation in vitro. *J Virol 65*:5825–5838.

Brown EA, Zhang H, Ping LH, Lemon SM. 1992. Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res 20*:5041–5045.

Cech TR. 1993. Structure and mechanism of the large catalytic RNAs: Group I and group II introns and ribonuclease P. In: Gesteland RF, Atkins JF eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 239–270.

Chamorro M, Parkin N, Varmus HE. 1992. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc Natl Acad Sci USA 89*:713–717.

Chen JH, Le SY, Maizel JV Jr. 1992. A procedure for RNA pseudoknot prediction. *Comput Appl Biosci 8*:243–248.

Choo QL, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina-Selby A, Barr PJ, Weiner AJ, Bradley DW, Kuo G, Houghton M. 1991. Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci USA 88*:2451–2455.

Choo QL, Kuo G, Weiner AJ, Wang KS, Overby LR, Bradley DW, Houghton M. 1992. Identification of the major, parenteral non-A, non-B hepatitis agent (hepatitis C virus) using a recombinant cDNA approach. *Semin Liver Dis 12*:279–288.

Christiansen J, Garrett R. 1988. Enzymatic and chemical probing of ribosomal RNA-protein interactions. *Methods Enzymol 164*:457–469.

de Wet JR, Wood KV, DeLuca M, Helinski DR, Subramani S. 1987. Firefly luciferase gene: Structure and expression in mammalian cells. *Mol Cell Biol 7*:725–737.

Grakoui A, McCourt DW, Wychowski C, Feinstone SM, Rice CM. 1993. Characterization of the hepatitis C virus-encoded serine proteinase: Determination of proteinase-dependent polyprotein cleavage sites. *J Virol 67*:2832–2843.

Houghton M, Weiner A, Han J, Kuo G, Choo QL. 1991. Molecular biology of the hepatitis C viruses: Implications for diagnosis, development and control of viral disease. *Hepatology 14*:381–388.

Jackson RJ, Howell MT, Kaminski A. 1990. The novel mechanism of initiation of picornavirus RNA translation. *Trends Biochem Sci 15*:477–483.

Jacobson SJ, Konings DAM, Sarnow P. 1993. Biochemical and genetic evidence for a pseudoknot structure at the 3' terminus of the poliovirus RNA genome and its role in viral RNA amplification. *J Virol 67*:2961–2971.

Kaminski A, Belsham GJ, Jackson RJ. 1994. Translation of encephalomyocarditis virus RNA: Parameters influencing the selection of the internal initiation site. *EMBO J 13*:1673–1681.

Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, Shimotohno K. 1990. Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc Natl Acad Sci USA 87*:9524–9528.

Kettinen H, Grace K, Grunert S, Clarke B, Rowlands D, Jackson R. 1994. Mapping of the internal ribosome entry site at the 5' end of the hepatitis C virus genome. *Proceedings of the International Symposium on Viral Hepatitis and Liver Disease*. Forthcoming.

Kozak M. 1989. The scanning model for translation: An update. *J Cell Biol 108*:229–241.

Krol A, Carbon P. 1989. A guide for probing native small nuclear RNA and ribonucleoprotein structure. *Methods Enzymol 180*:212–227.

Le SY, Chen JH, Sonenberg N, Maizel JV. 1992. Conserved tertiary structure elements in the 5' untranslated region of human enteroviruses and rhinoviruses. *Virology 191*:858–866.

Le SY, Chen JH, Sonenberg N, Maizel JV Jr. 1993. Conserved tertiary structural elements in the 5' nontranslated region of cardiovirus, aphthovirus and hepatitis A virus RNAs. *Nucleic Acids Res 21*:2445–2451.

Le SY, Sonenberg N, Maizel JV Jr. 1994. Distinct structural elements and internal entry of ribosomes in mRNA3 encoded by infectious bronchitis virus. *Virology 198*:405–411.

Le SY, Sonenberg N, Maizel JV Jr. 1995. Unusual folding regions and ribosome landing pad within hepatitis C virus and pestiviruses. *Gene 154*:137–143.

Leathers V, Tanguay R, Kobayashi M, Gallie DR. 1993. A phylogenetically conserved sequence within viral 3' untranslated RNA pseudoknots regulates translation. *Mol Cell Biol 13*:5331–5347.

Macejak D, Sarnow P. 1991. Internal initiation of translation mediated by 5' leader of a cellular mRNA. *Nature 353*:90–94.

Oh S, Scott M, Sarnow P. 1992. Homeotic gene antennapedia mRNA contains 5'-noncoding sequences that confer translation initiation by internal ribosome binding. *Genes & Dev 6*:1643–1653.

Okamoto H, Okada S, Sugiyama Y, Kurai K, Iizuka H, Machida A, Miyakawa Y, Mayumi M. 1991. Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: Comparison with reported isolates for conserved and divergent regions. *J Gen Virol 72*:2697–2704.

Pelletier J, Sonenberg N. 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature 334*:320–325.

Perrota AT, Bean MD. 1991. A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature 350*: 434–436.

Pleij CWA. 1990. Pseudoknots: A new motif in the RNA game. *Trends Biochem Sci 15*:143–147.

Pleij CWA. 1993. RNA pseudoknots. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 609–613.

Pilipenko EV, Blinov VM, Ramanova LI, Sinyakov AN, Maslova SV, Agol VI. 1989. Conserved structural domains in the 5'-untranslated region of picornaviral genomes: An analysis of the segment controlling translation and neurovirulence. *Virology 168*:201–209.

Pilipenko EV, Gmyl AP, Maslova SV, Svitkin YV, Sinyakov AN, Agol VI. 1992. Prokaryotic-like *cis* elements in the cap-independent internal initiation of translation on picornavirus RNA. *Cell 68*:119–131.

Poole TL, Wang C, Popp RA, Potgieter LND, Siddiqui A, Collett MS. 1995. Pestivirus translation initiation by internal ribosome entry. *Virology 206*:750–754.

Puglisi JD, Wyatt JR, Tinoco I Jr. 1988. A pseudoknotted RNA oligonucleotide. *Nature 331*:283–286.

Schimmel P. 1989. RNA pseudoknot that interacts with components of the translation apparatus. *Cell 58*:9–12

Shamoo Y, Tarn A, Konigsberg WH, Williams KR. 1993. Translational repression by the bacteriophage T4 gene 32 protein involves specific recognition of an RNA pseudoknot structure. *J Mol Biol 232*: 89–104.

Skinner MA, Racaniello VR, Dunn G, Cooper J, Minor PD, Almond JW. 1989. New model for the secondary structure of the 5' noncoding RNA of poliovirus is supported by biochemical and genetic data that also show that RNA secondary structure is important in neurovirulence. *J Mol Biol 207*:379–392.

Somogyi P, Jenner AJ, Brierley I, Inglis SC. 1993. Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol 13*: 6931–6940.

Stern S, Moazed D, Noller HF. 1988. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol 164*:481–489.

Takamizawa A, Mori C, Fuke I, Manabe S, Murakami S, Fujita J, Onishi E, Andoh T, Yoshida I, Okayama H. 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. *J Virol 65*:1105–1113.

Tang CK, Draper DE. 1989. Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell 57*:531–536.

Tsukiyama-Kohara K, Iizuka N, Kohara M, Nomoto A. 1992. Internal ribosome entry site within hepatitis C virus RNA. *J Virol 66*: 1476–1483.

Tsukuma H, Hiyama T, Tanaka S, Nakao M, Yabuuchi T, Kitamura T, Nakanishi K, Fujimoto I, Inoue A, Yamazaki H, Kawashima T. 1993. Risk factors for hepatocellular carcinoma among patients with chronic liver disease. *N Engl J Med 328*:1797–1801.

Tu C, Tzeng TH, Bruenn JH. 1992. Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc Natl Acad Sci USA 89*:8636–8640.

Wang C, Sarnow P, Siddiqui A. 1993. Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *J Virol 67*:3338–3344.

Wang C, Sarnow P, Siddiqui A. 1994. A conserved helical element is essential for internal initiation of translation of hepatitis C virus RNA. *J Virol 68*:7301–7307.

Wang C, Siddiqui A. 1995. Structure and function of the hepatitis C virus internal ribosome entry site. In: Sarnow P, ed. Cap independent translation. *Curr Topics Microbiol Immunol 203*:99–115.

Wassarman D, Steitz JA. 1991. Structural analysis of the 7S ribonucleoprotein (RNP), the most abundant human small RNA of unknown function. *Mol Cell Biol 11*:3432–3445.

Willis NM, Gesteland RF, Atkins JF. 1991. Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus *gag* stop codon. *Proc Natl Acad Sci USA 88*:6991–6995.

Wimmer E, Hellen CUT, Cao X. 1993. Genetics of poliovirus. *Annu Rev Genet 27*:353–436.

Wyatt JR, Puglisi JD, Tinoco I Jr. 1990. RNA pseudoknots; stability and loop size requirements. *J Mol Biol 214*:455–470.