Nucleic Acids Research, 2001, Vol. 29, No. 12 **2535–2541**

Hepatitis C virus internal ribosome entry site RNA contains a tertiary structural element in a functional domain of stem–loop II

Alita J. Lyons, J. Robin Lytle, Jordi Gomez[1] and Hugh D. Robertson*

Program in Biochemistry and Structural Biology, Weill Graduate School of Medical Sciences of Cornell University, 1300 York Avenue, New York, NY 10021, USA and [1]Laboratorio Hepatologia-Medicina Interna, Edificio Almacenes Generales, Hospital del Valle d'Hebron, Paseo del Valle de Hebron s/n 08035, Barcelona, Spain

Received February 20, 2001; Revised April 11, 2001; Accepted April 25, 2001

ABSTRACT

The internal ribosome entry site (IRES) of hepatitis C virus (HCV) RNA contains >300 bases of highly conserved 5′-terminal sequence, most of it in the uncapped 5′-untranslated region (5′-UTR) upstream from the single AUG initiator triplet at which translation of the HCV polyprotein begins. Although progress has been made in defining singularities like the RNA pseudoknot near this AUG, the sequence and structural features of the HCV IRES which stimulate accurate and efficient initiation of protein synthesis are only partially defined. Here we report that a region further upstream from the AUG, stem–loop II of the HCV IRES, also contains an element of local tertiary structure which we have detected using RNase H cleavage and have mapped using the singular ability of two bases therein to undergo covalent intra-chain crosslinking stimulated by UV light. This pre-existing element maps to two non-contiguous stretches of the HCV IRES sequence, residues 53–68 and 103–117. Several earlier studies have shown that the correct sequence between bases 45 and 70 of the HCV IRES stem–loop II domain is required for initiation of protein synthesis. Because features of local tertiary structure like the one we report here are often associated with protein binding, we propose that the HCV stem–loop II element is directly involved in IRES action.

INTRODUCTION

Hepatitis C virus (HCV) is an enveloped virus classified in the family Flaviridae (1). As shown in the map in Figure 1, its genome consists of a single unsegmented strand of RNA, with a plus polarity, ~9600 nt in length. The genome consists of a 5′-untranslated region (5′-UTR) of 341 nt, followed by an AUG start triplet and a long open reading frame extending along most of its length, followed by a 3′-untranslated region (3′-UTR). The first 400 bases are critical for viral function because they comprise, over most of their length, the HCV internal ribosome entry site (IRES) (2–5), where protein synthesis starts, eventually leading to production of the polyprotein precursor to the various HCV structural and non-structural proteins (1–6; Fig. 1). The HCV IRES is an example of an uncapped 5′-terminal viral mRNA sequence which nonetheless specifies accurate and efficient ribosome binding (reviewed in 2). While members of the picornavirus and pestivirus groups possess IRES sequences (2,7–9), along with certain rare cellular mRNAs (2,10–12), the HCV IRES has been particularly well studied. While it is not yet known how any IRES sequences specify ribosome binding, certain suggestive features have been discerned in HCV and closely related RNAs (3–5,13–15).

Several laboratories have conducted studies correlating structure and function in the HCV IRES, leading to predictions of its secondary structure such as that shown in the lower drawing of Figure 1 (3,13–21). One of the most important advances has been the prediction and detection of a region of unusual structure in the vicinity of the AUG start triplet at bases 342–344. In particular, a pseudoknot involving secondary structural bonding between non-adjacent regions of HCV IRES sequence was shown to involve bases 320–330, immediately upstream from the initiator AUG (19). It seems certain that this structural element must participate in IRES function. Several other groups (14–16,20) have used directed mutagenesis and UV-dependent protein–RNA crosslinking to suggest that RNA sequences within an upstream region of the HCV IRES, stem–loop II, which spans bases 44–118 as shown in Figure 1, are also likely to be involved in an early step of ribosome attachment.

There is already evidence from other systems that structural elements of a higher order can be protein binding recognition sites (22–25). Thus the detection of a tertiary structural element in a second HCV IRES domain, stem–loop II, would parallel the AUG-proximal pseudoknot in potential significance. We report here that the structured region near the pseudoknot and that within stem–loop II both appear to have unusually stable structure as studied by DNA-stimulated RNase H digestion. We further report that, using the technique of RNA–RNA crosslinking by UV light, we have detected, isolated and mapped an element of local tertiary structure which includes stem–loop II bases required for HCV IRES function (14–16,20).

*To whom correspondence should be addressed. Tel: +1 212 746 6400; Fax: +1 212 746 8144; Email: hdrober@mail.med.cornell.edu

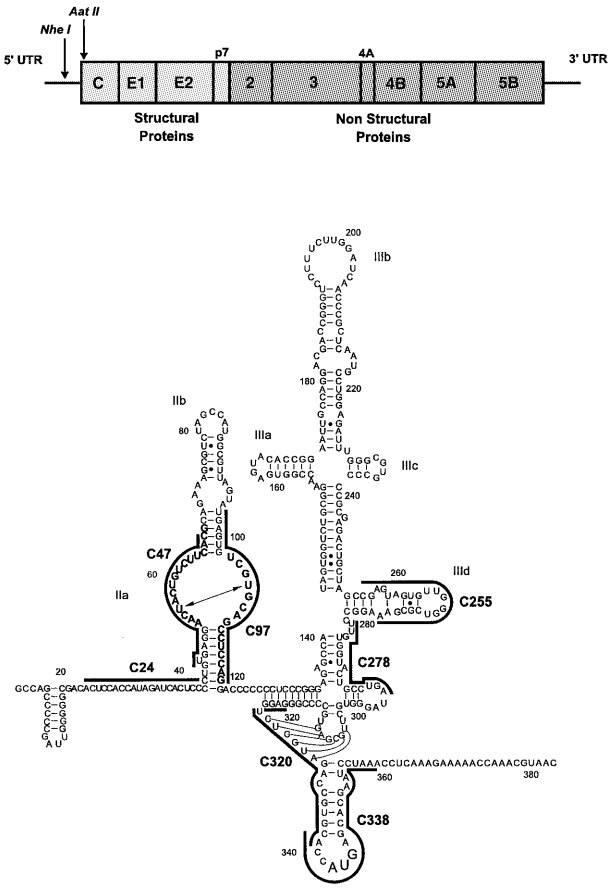

Figure 1. Map of HCV viral RNA and its 5′-UTR. (Top) The 5′-UTR contains the most conserved sequence of the HCV genome. The IRES contains most of the 5′-UTR plus some coding sequence. The two principal 5′-proximal RNAs studied here are created by transcription of DNA templates that have been cleaved with the restriction enzymes *Nhe*I at position 249 and *Aat*II at position 402. (Bottom) We propose that the HCV IRES has an element of local tertiary structure in stem–loop II which places residue U56 in close proximity to U106. The hypothetical folding shown is based on work published by Brown *et al.* (3) and Honda *et al.* (20) with some changes to stem–loop II to accommodate the UV-crosslinkable element described here. Most previous RNA–RNA crosslinked tertiary elements contain 'bubble' regions which, as shown in the NMR study of the loop E region of eukaryotic 5S rRNA (23), are actually filled with kinked and folded elements, so we have drawn the IRES structure in the middle of stem–loop II in this way. The sequences from RNase T1 partial digestion studies (Table 1) are in bold. The positions of the complementary DNA oligonucleotides used in the RNase H experiments depicted in Figure 2 are also indicated.

MATERIALS AND METHODS

Transcriptions

Plasmid pN(1–4728), HCV N-strain, genotype 1b, containing the first 4.7 kb of HCV sequence adjacent to the phage T7 promoter, was a gift from Dr Stanley Lemon (University of Texas, Galveston, TX). This DNA template was cleaved by two different restriction enzymes, *Aat*II and *Nhe*I (Fig. 1). When these templates were transcribed *in vitro* two ^{32}P-labeled RNAs spanning bases 1–402 (encompassing the entire IRES) and bases 1–249 (encompassing a smaller portion of the IRES) were produced. Transcription reaction conditions were based on earlier published work from this laboratory (26). A specific activity of 1.24×10^7 d.p.m./μg was generally used. [α-^{32}P]GTP (NEN Life Science) was used in all labeled transcriptions unless noted otherwise.

DNA-directed RNase H cleavage

Purified RNA transcripts were resuspended at a concentration of 0.6 nM in the reaction mixture. Parallel reactions were run in which a 1.5–2-fold molar excess of each DNA oligonucleotide complementary to a HCV IRES sequence of ~20 bases was either preheated or not preheated with the IRES transcript. In the preheated reactions RNA was incubated for 1 min in water at 90°C and then preheated mixtures of DNA oligonucleotide (synthesized by GeneLink) at ~1 nM and buffer at a final concentration of 20 mM HEPES pH 8.0, 50 mM KCl, 10 mM $MgCl_2$ and 1 mM DTT were added and the reactions allowed to cool slowly to room temperature. In the non-preheated reactions the heating step was omitted and the hybridization of RNA and DNA carried out for 30 min at room temperature. In all cases just before the RNase H digestion was begun, 1 μg carrier tRNA per reaction was added and the mix preheated at 37°C for 1 min. Reactions were started by adding 1 μl RNase H (Promega). Two microliters of a 1:5 RNase H dilution in hybridization buffer were added to control reactions (10 μl final volume) from which DNA oligonucleotide was omitted. Five microliter aliquots were removed from the RNase H digestion reactions at the beginning and at intervals of 5, 15, 30 and 60 min thereafter and electrophoresed on urea-containing 5% polyacrylamide gels. Autoradiography was carried out on Kodak XAR-5 X-ray film.

UV crosslinking

This technique is described by Branch *et al.* (27). After exposure of RNA transcripts suspended in 10 μl drops to UV irradiation (27), control and UV-treated RNAs were separated on 5–8% polyacrylamide gels containing 7 M urea. Autoradiography was carried out on XAR-5 film. A UVP Inc. Mineralight Lamp (model UVG-11) was used for all UV treatments.

RNA fingerprinting

RNA fingerprinting, a technique in which radioactive RNA molecules are digested with RNase T1 (CalBiochem) and the digestion products separated in two dimensions by charge and size, has been described previously (28). Autoradiography of the DEAE–cellulose second dimension was carried out on XAR-5 film.

Partial RNase T1 digestion

Various concentrations of RNase T1 were prepared (0.01, 0.05, 0.1, 0.2 and 1 mg/ml). Analytical amounts of ^{32}P-labeled 249 base HCV RNA with and without previous UV treatment were digested with these RNase fractions and the digestion products electrophoresed on a denaturing 20% polyacrylamide gel. Optimum RNase concentrations were chosen and 2.4×10^7 d.p.m. RNA were digested and again electrophoresed on a denaturing 20% polyacrylamide gel. An RNase T1-resistant band unique to the previously UV-treated RNA was isolated and fingerprinted. The five spots that appeared were eluted and digested with pancreatic RNase (Worthington) and RNase T2 (CalBiochem) to confirm their identity. These techniques have

been described in greater detail by Branch *et al.* (28) and Barrell (29).

Primer extension inhibition

Primer extension inhibition assays were performed using two primers of 18 deoxynucleotides (GeneLink). Primer C65 hybridizes to the HCV 5′-UTR with its 3′-end annealing to base 65. Primer C128 base pairs to the 5′-UTR of HCV with its 3′-end annealing to base 128. Both primers (50 pmol) were treated with 40 U T4 polynucleotide kinase (New England Biolabs) and 100 µCi [γ-^{32}P]ATP (NEN Life Science) in buffer supplied by the manufacturer at 37°C for 30 min and purified on a denaturing 15% polyacrylamide gel. Aliquots of 50 000 d.p.m. ^{32}P-labeled primer, 2 pmol unlabeled primer and 100 ng unlabeled, UV-treated 249 base RNA were dried and resuspended in 1 µl of hybridization buffer (80% formamide, 0.4 M NaCl, 40 mM PIPES pH 6.8, and 1 mM EDTA), heated to 45°C for 15 min in a sealed glass capillary tube and cooled to room temperature. To the hybridization mixture were then added 10 µl of reaction buffer (Promega) containing 10 mM dNTPs and 50 U RNasin (Promega), followed by 10 U AMV reverse transcriptase (Promega) and a 1 h incubation at 42°C. Four sequencing reactions, each containing one of the four dideoxynucleoside triphosphates, were carried out with each primer. The extension reactions were then electrophoresed on an 8% denaturing polyacrylamide gel and autoradiographed on Kodak XAR-5 X-ray film.

RESULTS

Initial experiments with HCV IRES RNA transcripts involved the 402 base RNA (Fig. 1) and seven complementary 20 base DNA oligonucleotides. At the top of Figure 2 is a map showing a linear schematic drawing of the HCV 5′-terminal RNA domain, below which appear boxes indicating the 20 base DNAs, each synthesized to be complementary to an HCV IRES region beginning with the residue number shown (thus C24 is complementary to bases 24–43, and so forth). Preliminary results (not shown) demonstrated that radioactive DNA oligonucleotides C24, C255 and C338 could bind to unlabeled 402 base HCV IRES RNA and undergo a gel mobility shift. Three other DNAs, C47, C278 and C320, could not be bound to their corresponding RNA regions in this assay, even if the DNA and RNA were preheated to 90°C in water and slowly cooled under physiological salt conditions.

RNase H cleavage reaction kinetics of the DNA:RNA hybrid regions formed with various DNAs and the HCV IRES are also shown in Figure 2. Column A shows the results of mixing the RNA, DNA and RNase H and incubating at 37°C under physiological conditions; column B shows reactions in which substrate and DNA were preheated to 90°C and cooled before RNase H treatment. Interestingly, a comparison of C24 versus C47 cleavage in stem–loop II and C320 versus C338 in stem–loop III cleavage shows that there are two pairs of adjacent locations on the HCV IRES where complete resistance to RNase H cleavage gives way abruptly to complete sensitivity. Furthermore, C97 was also unable to promote RNase H cleavage, despite the relatively 'open' structure depicted in Figure 1, lower drawing. DNA C255, which spans the entire stem–loop IIId domain of the HCV IRES, stimulates moderate cleavage with or without preheating, suggesting a relatively open structure, in agreement with the gel shift results.

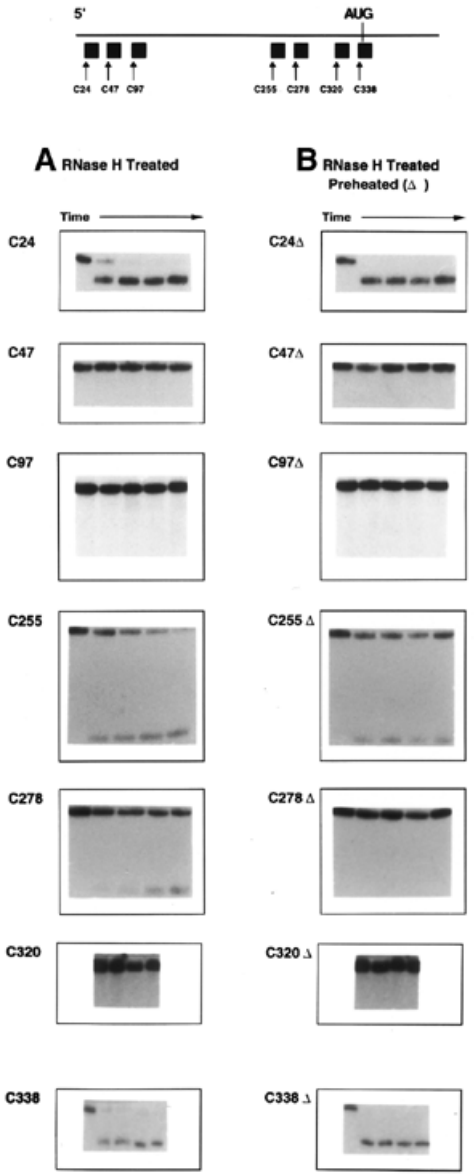

Figure 2. RNase H treatment of HCV IRES RNA following hybridization with 20 base DNA oligonucleotides. The ^{32}P-labeled 402 base transcript spanning the HCV IRES (Fig. 1) was prepared as described in Materials and Methods and treated with RNase H in the presence of various DNA oligonucleotides complementary to different regions of the IRES sequence as shown on the map at the top of the figure: C24, bases 24–43; C47, bases 47–66; C97, bases 97–116; C255, bases 255–274; C278, bases 278–297; C320, bases 320–339; C338, bases 338–360. In control reactions in the absence of DNA no RNase H cleavage occurs; similarly, in the presence of each DNA but without RNase H no cleavage is observed. (**A**) In the left column the results of mixing RNA substrate, a 1.5–2-fold molar excess of DNA and RNase H and incubating as described in Materials and Methods are shown. In each individual reaction profile the left-most lane represents time 0, while each lane to the right represents sequentially 5, 15, 30 and 60 min incubation. (**B**) The right column represents reactions in which substrate and DNA were preheated separately to 90°C, buffer added, the reactions mixed, cooled and RNase H added, as described in Materials and Methods. Again the left-most lane represents time 0, with additional lanes representing increments as in (A). Reactions involving C320 omitted the 5 min time point.

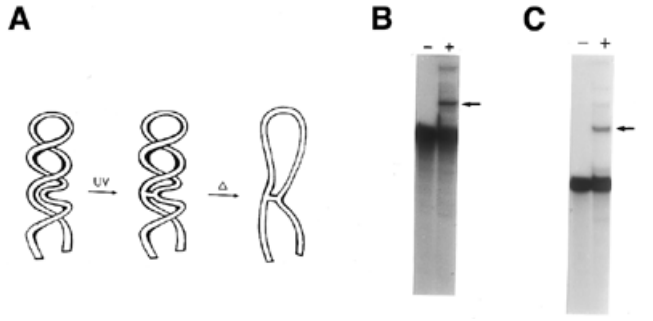

Figure 3. Higher order structure mapping of the HCV 5′-UTR by UV crosslinking. (**A**) The kind of RNA elements susceptible to UV-induced crosslinking. The left drawing shows a self-structured RNA with both Watson–Crick base pairing (upper) and a non-Watson–Crick tertiary element (kink in lower right strand) which pre-exists in the RNA before any treatment. UV light (peak 254 nm) then causes covalent base–base crosslinks to arise in the pre-existing structural element between the most closely situated pair of residues which are not involved in conventional base pairing (central drawing). This process is directly comparable to the UV-induced formation of thymine dimers in DNA. Heat treatment denatures the Watson–Crick secondary structure, leaving the covalent link to confer altered properties on the RNA, which is now partially circular (right drawing). (**B**) PAGE of a transcript of *Aat*II-cleaved DNA template (bases 1–402) without (–) and with (+) previous UV treatment. The arrow indicates the UV-crosslinked species. (**C**) PAGE of a transcript of *Nhe*I-cleaved DNA template (bases 1–249) without (–) and with (+) UV treatment. The arrow indicates the UV-crosslinked species.

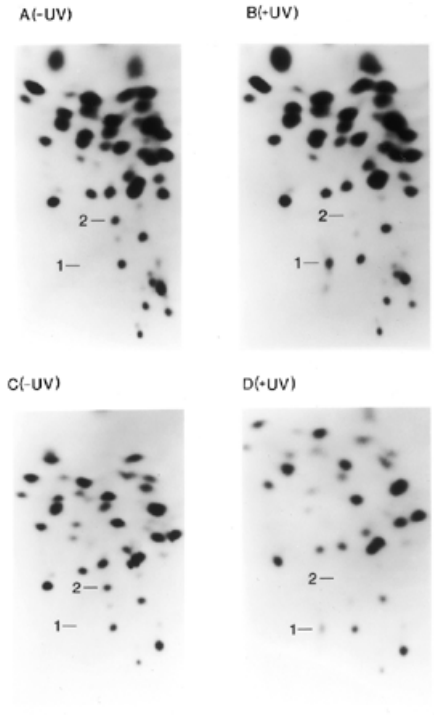

Figure 4. RNA fingerprints of control and UV-treated RNAs. (**A**) Fingerprint of the control (–UV) 402 base RNA. (**B**) Fingerprint of the UV-treated (+UV) 402 base RNA. (**C**) Fingerprint of the control (–UV) 249 base RNA. (**D**) Fingerprint of the UV-treated (+UV) 249 base RNA. The origin of all fingerprints is located in the bottom right corner. The digestion products are separated in the first (leftward) dimension by charge and in the second (upward) dimension by size.

Thus the RNase H cleavage data confirm and extend the preliminary gel shift results. In particular, the stable structure between C320 and C338 corresponds almost exactly to the element of pseudoknot structure postulated to exist in that region. Surprisingly, the other domain of stable structure, which begins somewhere between bases 43 and ~50, has not been previously identified, although various folding schemes have been proposed.

Since such stability suggests a pre-existing tertiary structural element in stem–loop II of the HCV IRES, both the 402 and the 249 base transcripts were synthesized and purified. Both transcripts were exposed to UV light for a total of 3 min. The RNAs were electrophoresed on either 4 or 8% polyacrylamide gels under denaturing conditions. The covalent linkage of two nucleotides induced by UV irradiation caused a change in the conformation of the molecule (Fig. 3A), which then migrated more slowly than the non-crosslinked molecule on a denaturing gel, allowing separation of the two species (Fig. 3B and C). Both crosslinked and non-crosslinked RNAs were eluted from the gel and used to isolate the bases taking part in the crosslink.

Several methods were employed to map the crosslinked nucleotides, including RNA fingerprinting analyses, partial RNase T1 digestion studies, 'nearest neighbor' secondary digestion analyses and primer extension inhibition assays. For RNA fingerprinting both crosslinked and non-crosslinked species were separated by gel electrophoresis, eluted from the gel, digested with RNase T1 and then fingerprinted (Fig. 4). The fingerprint pattern of the crosslinked RNA contains an extra species as a result of a new covalent link between two specific bases in the RNA (Fig. 4B and D, spot 1). Also, the crosslinked RNA fingerprints lack an oligonucleotide present in the control fingerprints (spot 2) because that species has been crosslinked to another oligonucleotide, thus acquiring a different mobility (Fig. 4A and C). Spot 2 was found to contain bases 53–60 of the HCV IRES RNA. Spot 1 was also recovered, subjected to further analyses and found to contain bases 53–60 plus additional unknown bases.

To define further the crosslink and its context a partial RNase T1 digestion assay was carried out. Briefly, both crosslinked and non-crosslinked HCV RNA transcripts were digested with 0.25 µg/µl RNase T1 and incubated at 37°C for 40 min. This digestion allowed a T1-resistant product containing the crosslinked bases and surrounding sequences to be purified from the rest of the molecule via electrophoresis on a 20% denaturing polyacrylamide gel (data not shown). This RNase T1-resistant product was then recovered from the gel and subjected to RNA fingerprinting. The oligonucleotides recovered from this fingerprint were found to contain the sequences shown in Table 1. The protected region runs from base A53 to G68 and U103 to G117. The lack of free $_{106}$UG$_{107}$ strongly suggests that this dinucleotide is involved in the crosslink, and the proposed composition for spot 1 is listed in Table 1.

In order to discover exactly which bases from the isolated sequences were crosslinked, secondary digestions with pancreatic RNase, which cleaves after pyrimidines, and RNase T2, which cleaves after every base, were carried out on spot 1 isolated from fingerprints of crosslinked RNA. Spot 1, from transcripts labeled either with [α-^{32}P]UTP, ATP or GTP, was digested

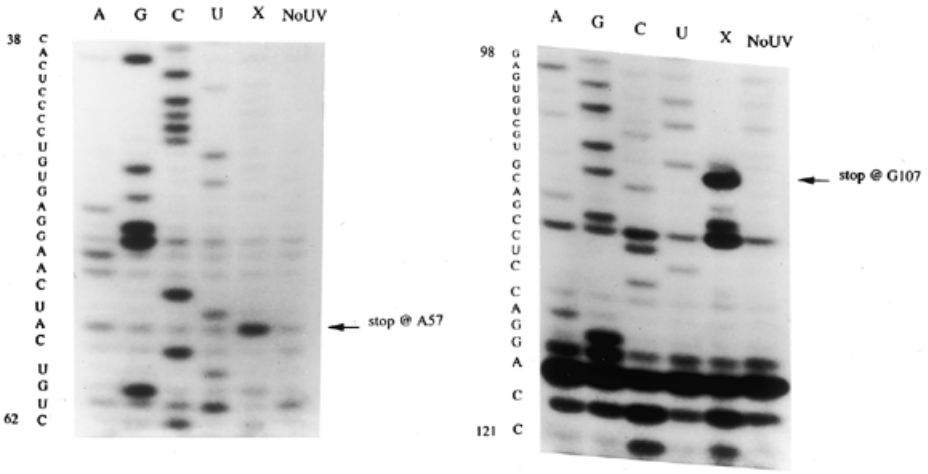

Figure 5. Primer extension on UV-crosslinked 249 base RNA. (Left) PAGE of primer extension reactions using the 249 base RNA and primer C65. Left lanes contain dideoxynucleotides causing stops to occur at the indicated template bases: A, G, C and U. Lane X contains the gel-purified, UV-crosslinked 249 base RNA. The arrow on the right indicates that primer extension stops at base A57. Lane 'No UV' contains gel-purified 249 base RNA that has not been exposed to UV light. (Right) Lanes A, G, C and U indicate the sequence of the 249 base RNA template using primer C128. Lane X contains the gel-purified, UV-crosslinked 249 base RNA. The arrow on the right indicates a stop at base G107. Lane 'No UV' contains a similar reaction using 249 base control RNA as template.

Table 1. Oligonucleotides in the HCV UV-crosslinked domain

Spot no.	Sequence
1	$_{53}$AACUACUG$_{60}$[U] | $_{106}$UG$_{107}$[C]
2	$_{61}$UCUUCACG$_{68}$[C]
3	$_{111}$CCUCCAG$_{117}$[G]
4	$_{103}$UCG$_{105}$[U]
5	$_{108}$CAG$_{110}$[C]

RNase T1-resistant oligonucleotides were eluted from the DEAE–cellulose thin layer plate comprising the second dimension of the RNA fingerprint of the UV-specific partial digestion product from HCV bases 1–249, as described in the text. The oligonucleotides were subjected to further analysis using pancreatic RNase and RNase T2 as described in Materials and Methods and their sequences assigned as indicated. In each case the oligonucleotides are depicted in the 5′→3′ orientation. The subscript numerals indicate the positions of the first and last bases of each oligonucleotide according to the numbering system used in Figure 1. The nearest neighbor to each 3′-terminal base residue appears in brackets.

with pancreatic RNase and yielded the sequences $_{53}$AAC$_{55}$, $_{57}$AC$_{58}$, U59, G60 and a new species that is most likely a more completely digested crosslinked species. These results strongly suggest that U56 from the octanucleotide 53–60 is involved in the crosslink. To determine whether U106 or G107 was the other base involved in the crosslink [α-^{32}P]CTP-labeled spot 1 was digested with pancreatic RNase which yielded $_{53}$AAC$_{55}$, $_{57}$AC$_{58}$ and G107 (labeled by C108). Since all bases except free U56 and U106 were liberated from spot 1, these results strongly suggest that bases U56 and U106 are covalently crosslinked to each other.

This finding was confirmed using primer extension inhibition assays. Having isolated the sequences in the crosslinked regions, primers C128 and C65 were designed to hybridize upstream of the bases believed to be involved in the crosslink. In primer extension reactions strong stops were observed at bases A57 (Fig. 5, left) and G107 (Fig. 5, right), indicating an element at bases U56 and U106 which inhibits primer extension. The strong stops observed at C111/G110 in Figure 5, right, are present in control lanes as well and do not indicate a second photocrosslink involving either of these bases. This conclusion is also supported by the recovery of spots $_{111}$CCUCCAG$_{117}$ and $_{108}$CAG$_{110}$ (Table 1) free of any crosslinks. The primer extension inhibition data thus support the conclusion reached above that bases U56 and U106 become crosslinked.

In Figure 1 the crosslink that forms upon UV treatment between bases U56 and U106 of the IRES of HCV is indicated by an arrow. The secondary structure has been drawn with a 'bubble' region to represent the crosslinkable tertiary element in the IRES (1). The bold sequences surrounding the crosslink are the sequences protected during partial RNase T1 digestion. These products are described in Table 1.

DISCUSSION

Using hybridization to complementary DNA oligonucleotides and RNase H cleavage of DNA:RNA hybrids thus formed we have discovered an element of stable local structure in a 5′-proximal region of the HCV IRES called stem–loop II. Employing the method of RNA–RNA crosslinking by UV irradiation (27), which has identified important tertiary structural elements in a number of RNA molecules (22–26), we have mapped and sequenced this domain of HCV stem–loop II (Fig. 1), which lacks conventional Watson–Crick helical structure and promotes UV crosslinking between HCV bases U56 and U106. While most of our detailed sequence analysis and mapping of this feature and its surroundings (Figs 2–4) have been carried out on the 249 base 5′-terminal transcript of the HCV IRES, we confirmed that the identical element is also present in the 402 base 5′-terminal HCV RNA depicted in Figure 1 (see also Fig. 4A and B). This latter result indicates that stem–loop II of the full-size HCV IRES contains this

element of local tertiary structure which may therefore be involved in ribosome binding or recognition of HCV mRNA.

Structural requirements for IRES ribosome binding

The way in which IRES elements bind ribosomes and specify translation initiation with high precision remains unknown. Sequence and secondary structural comparisons between IRES domains from different mRNAs have shown few systematic resemblances. Jackson (2) has proposed that higher order structure along with limited and widely separated regions of base pairing may be the only common IRES features related to ribosome recognition. In this regard, while most of the HCV genome sequence is highly variable the IRES region is highly conserved among isolates, an observation which could be explained by a need to conserve the structural features proposed by Jackson (2). One of these features is presumably the pseudoknot structure found near the HCV AUG initiator codon. Another could well be the tertiary structural element mapped here.

UV-crosslinkable RNA tertiary structural elements often occur at sites of protein binding (22–26). For instance, the E loop of eukaryotic 5S rRNA (22,23) contains a tertiary element to which transcription factor TFIIIA and ribosomal protein L5 bind (24,25). Because of these prior examples, and in the light of the functional involvement of HCV IRES stem–loop II, we propose that the element characterized here is likely to be involved in IRES action. Several recent studies have, in fact, indicated that at least one tract of bases in HCV IRES stem–loop II, which overlaps with the structural domain mapped here, is critical for the initial steps of ribosome binding. Using both mutational analysis and RNA–protein crosslinking, Fukushi *et al.* (15) and Buratti *et al.* (16) have demonstrated that a number of bases in stem–loop II must retain their identity and/or position to maintain IRES function. Furthermore, Fukushi *et al.* (15) have shown that a protein called p25 associates with the stem–loop II domain of the HCV IRES. Pestova *et al.* (14), using RNA–protein crosslinking, have identified a region of stem–loop II which somehow interacts with 40S ribosomal subunits and contains bases from the UV-crosslinkable domain. They also confirmed a role for protein p25 and suggested that it is identical with 40S ribosomal subunit protein S9 (14). Thus the HCV IRES element of local tertiary structure which we have characterized is located at a key functional site and is therefore likely to be involved in IRES action and may do so by acting as a binding signal for ribosomal protein S9.

A number of proposals for the secondary structure of the HCV IRES have appeared over the years. Two studies in particular have attempted to predict the structure of stem–loop II. One of these was in 1999 by Honda *et al.* (20), who used phylogenetic comparisons and directed mutagenesis to choose between two proposed HCV IRES stem–loop II structures. Although the structure favored by Honda *et al.* (20) contains data-supported RNA helices above and below the part of stem–loop II which contains the UV-crosslinkable bases U56 and U106 reported here, base paired structure is also placed in the immediate vicinity of the UV-crosslinkable residues by the MFOLD program, unsupported by experimental data. From what we now know of this local tertiary structure from our work this part of the Honda *et al.* stem–loop II structure (20) should be redrawn without secondary structure in this vicinity since the two U residues which undergo efficient UV crosslinking would then no longer be shown as members of different Watson–Crick base pairs. Such base paired residues in RNA have never been known to undergo the kind of UV crosslinking observed here.

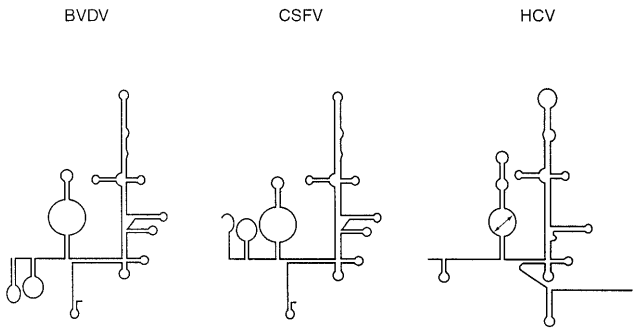

Figure 6. Comparison of the refined HCV IRES structure and similar regions from pestivirus RNAs. The right panel, showing a line drawing of the newly refined HCV IRES structure proposed in Figure 1, can be compared to line drawings of the 5′-proximal regions of BVDV and CSFV RNAs. It is clear that our redrawn HCV IRES has several structural similarities to the pestivirus RNAs in addition to the sequence homologies to HCV already pointed out by Brown *et al.* (3), from which the BVDV and CSFV folding schemes have been taken.

Another proposal for secondary structure in the 5′-proximal domain of the HCV IRES was published by Le *et al.* (21) in 1998. This model proposes two separate, smaller stem–loops in place of the larger stem–loop II discussed above. In this structure one of the UV-crosslinkable bases (U56) is in a loop portion while the other is part of the base paired stem of an adjacent structure. These structures should also be redrawn as a single stem with an open loop since the two UV-crosslinkable bases in analogous tertiary structural elements have been shown by NMR and other techniques (23) to be the two bases closest to each other in space, a very unlikely outcome with the structure depicted by Le *et al.* (21).

Implications of the stem–loop II IRES structural element for HCV therapy

The conserved sequences in the IRES domain of the HCV 5′-UTR make this region, in principle, a good target for therapeutic approaches utilizing antisense or targeted ribozyme technology. Studies by others (30) and ourselves (Fig. 2 and additional unpublished experiments) have focused on attempts to find conserved sequences in the HCV IRES which could be attacked by these approaches. Highly structured regions, such as the one mapped here, prove to be resistant to such cleavage. In this regard we have discovered several additional UV-crosslinkable elements of local tertiary structure within the HCV IRES domain, the mapping of which is ongoing (data not shown). It is likely that the immediate vicinities of these elements will also thwart therapeutic cleavage attempts. More fine structure mapping will be necessary, along with comparative studies to define the best therapeutic targets.

IRES comparisons

The HCV IRES is thought to be most closely related to those contained in the RNA genomes of several pestiviruses, including classical swine fever virus (CSFV, previously called hog cholera virus, HChV), bovine viral diarrhea virus (BVDV)

and border disease virus (BDV). Figure 6 shows line drawings of the HCV structure, compared to those of CSFV and BVDV proposed earlier (3). When we redraw the HCV stem–loop II structure as an open 'bubble' to accommodate our new structural data we see that it closely resembles the CSFV and BVDV stem–loop II structures proposed by Brown *et al.* in 1992 (3; see Fig. 6). Thus the lower parts of this HCV stem–loop II and its equivalent structures in the pestivirus RNAs can all be represented as open bubble regions with full potential to form non-Watson–Crick local tertiary structure. In addition, while the upper portions of these structures have extensive sequence homology (3,20), there is little such homology between the UV-crosslinkable region of HCV stem–loop II and the analogous BVDV and CSFV open regions shown in Figure 6. Nonetheless, they may all contain similar structural elements which could specify ribosome function and which could all be UV crosslinkable.

It will require direct analysis of these pestivirus RNAs by UV crosslinking and mapping in comparison to the HCV IRES to test this hypothesis. If UV-crosslinkable elements in CSFV and BVDV turn out to map in the domains analogous to HCV stem–loop II, as predicted, such findings would be a first step in the understanding of IRES structural evolution. A systematic approach involving structural, as well as sequence, comparisons should then lead to a fuller understanding of how such highly dissimilar IRES elements can retain such specific ribosome binding behavior.

ACKNOWLEDGEMENTS

We thank Sylvia Genus for excellent technical assistance, Dr Olivia D.Neel for helpful advice and discussions and Dr Stanley Lemon for a gift of cloned DNA. This work was supported in part by grant no. NYS U35-8010 to H.D.R. from the Center for Advanced Technology Program of the New York State Science and Technology Foundation, grant no. DK 56424 to H.D.R. from the NIH and grant no. F.I.S.S. (B.A.E.) 96/5701 to J.G.

REFERENCES

1. Houghton,M. (1996) Hepatitis C viruses. In Fields,B.N., Knipe,D.M., Howley,P.M. *et al.* (eds) *Fields Virology*, 3rd Edn. Lippincott-Raven, Philadelphia, PA, pp. 1035–1058.
2. Jackson,R.J. (1996) A comparative view of initiation site selection mechanisms. In *Translational Control*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 71–111.
3. Brown,E.A., Zhang,H., Ping,L. and Lemon,S.M. (1992) Secondary structure of the 5′ nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res.* , **20**, 5041–5045.
4. Tsukiyama-Kohara,K., Iizuka,N., Kohara,M. and Nomoto,A. (1992) Internal ribosome entry site within hepatitis C virus RNA. *J. Virol.*, **66**, 1476–1483.
5. Reynolds,J.E., Kaminski,A., Kettinen,H.J., Grace,K., Clarke,B.E., Carroll,A.R., Rowlands,D.J. and Jackson,R.J. (1995) Unique features of internal initiation of hepatitis C virus RNA translation. *EMBO J.*, **14**, 6010–6020.
6. Bukh,J., Purcell,R.H. and Miller,R.H. (1992) Sequence analysis of the 5′ noncoding region of hepatitis C virus. *Proc. Natl Acad. Sci. USA*, **89**, 4942–4946.
7. Honda,M., Ping,L.H., Rijnbrand,R.C.A., Amphlett,E., Clarke,B., Rowlands,D. and Lemon,S.M. (1996) Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. *Virology*, **222**, 31–42.
8. Reynolds,J.E., Kaminski,A., Carroll,A.R., Clarke,B.E., Rowlands,D.J. and Jackson,R.J. (1996) Internal initiation of translation of hepatitis C virus RNA: the ribosome entry site is at the authentic initiation codon. *RNA*, **2**, 867–878.
9. Rijnbrand,R., van der Straaten,T., van Rijn,P.A., Spaan,W.J.M. and Bredenbeek,P.J. (1997) Internal entry of ribosomes is directed by the 5′ noncoding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot upstream of the initiation codon. *J. Virol.*, **71**, 451–457.
10. Macejak,D.G. and Sarnow,P. (1991) Internal initiation of translation mediated by the 5′ leader of a cellular mRNA. *Nature*, **353**, 90–94.
11. Oh,S.K., Scott,M.P. and Sarnow,P. (1992) Homeotic gene *Antennapedia* mRNA contains 5′-noncoding sequences that confer translational initiation by internal ribosome binding. *Genes Dev.*, **6**, 1643–1653.
12. Johannes,G. and Sarnow,P. (1998) Cap-independent polysomal association of natural mRNAs encoding c-myc, BiP and eIF4G conferred by internal ribosome entry sites. *RNA*, **4**, 1500–1513.
13. Honda,M., Brown,E.A. and Lemon,S.M. (1996) Stability of a stem–loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA*, **2**, 955–968.
14. Pestova,T.V., Shatsky,I.N., Fletcher,S.P., Jackson,R.J. and Hellen,C.U.T. (1998) A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev.*, **12**, 67–83.
15. Fukushi,S., Kurihara,C., Ishiyama,N., Hoshino,F., Oya,A. and Katayama,K. (1997) The sequence element of the internal ribosome entry site and a 25-kilodalton cellular protein contribute to efficient internal initiation of translation of hepatitis C virus RNA. *J. Virol.*, **71**, 1662–1666.
16. Buratti,E., Tisminetzky,S., Zotti,M. and Baralle,F.E. (1998) Functional analysis of the interaction between HCV 5′ UTR and putative subunits of eukaryotic translation initiation factor eIF3. *Nucleic Acids Res.*, **26**, 3179–3187.
17. Lu,H.H. and Wimmer E. (1996) Poliovirus chimeras replicating under the translational control of genetic elements of hepatitis C virus reveal unusual properties of the internal ribosomal entry site of hepatitis C virus. *Proc. Natl Acad. Sci. USA*, **93**, 1412–1417.
18. Wang,C., Sarnow,P. and 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.
19. Wang,C., Le,S.Y. and Siddiqui,A. (1995) An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5′ noncoding region. *RNA*, **1**, 526–537.
20. Honda,M., Beard,M.R., Ping,L.H. and Lemon,S.M. (1999) A phylogenetically conserved stem–loop structure at the 5′ border of the interal ribosome entry site of hepatits C virus is required for cap-independent viral translation. *J. Virol.*, **73**, 1165–1174.
21. Le,S.-Y., Liu,W.-M. and Maizel,J.V.Jr (1998). Phylogenetic evidence for the improved RNA higher-order structure in internal ribosome entry sequences of HCV and pestiviruses. *Virus Genes*, **17**, 279–295.
22. Branch,A.D., Benenfeld,B.J. and Robertson,H.D. (1985) An ultraviolet-sensitive RNA structural element in a viroid-like domain of the hepatitis delta virus. *Science*, **243**, 652–659.
23. Wimberly,B., Varani,G. and Tinoco,J.I. (1993) The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry*, **32**, 1078–1087.
24. Romaniuk,P.J. (1989) The role of highly conserved single-stranded nucleotides of *Xenopus* 5S RNA in the binding of transcription factor IIIA. *Biochemistry*, **28**, 1388–1395.
25. Allison,L.A., Romaniuk,P.J. and Bakken,A.H. (1991) RNA–protein interactions of stored 5S RNA with TFIIIA and ribosomal protein L5 during *Xenopus oogenesis*. *Dev. Biol.*, **144**, 129–144.
26. Circle,D.A., Neel,O.D., Robertson,H.D., Clarke,P.A. and Mathews,M.B. (1997) Surprising specificity of PKR binding to delta agent genomic RNA. *RNA*, **3**, 438–448.
27. Branch,A.D., Benenfeld,B.J., Paul,C.P. and Robertson,H.D. (1990) Analysis of ultraviolet-induced RNA-RNA cross-links: a means for probing RNA structure–function relationships. *Methods Enzymol.*, **180**, 418–442.
28. Branch,A.D., Benenfeld,B.J. and Robertson,H.D. (1989) RNA fingerprinting. *Methods Enzymol.*, **180**, 130–154.
29. Barrell,B.G. (1971) Fractionation and sequence analysis of radioactive nucleotides. In Cantoni,D.L. and Davies,D.R. (eds) *Procedures in Nucleic Acid Research*. Harper and Row, New York, NY, Vol. 2, pp. 751–779.
30. Lima,W.F., Brown-Driver,V., Fox,M., Hanecak,R. and Bruice,T.W. (1997) Combinatorial screening and rational optimization for hybridization to folded hepatitis C virus RNA of oligonucleotides with biological antisense activity. *J. Biol. Chem.*, **272**, 626–638.