Structural Domains of the 3′-Terminal Sequence of the Hepatitis C Virus Replicative Strand[†]

Mariola Dutkiewicz, Agata Świątkowska, Marek Figlerowicz, and Jerzy Ciesiołka*

Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań 61-704, Poland

Received February 29, 2008; Revised Manuscript Received September 23, 2008

ABSTRACT: Here we present the results of a structural analysis of the 3′-terminal region of the replicative strand of hepatitis C virus (HCV), *IRES(−)*, by the Pb^{2+}-induced cleavage approach and partial digestion with T1 ribonuclease. Oligoribonucleotides that represent selected domains of the earlier proposed in the literature secondary structure models of this region were also synthesized, their structures were analyzed in solution, and the results were compared to those obtained with the full-length molecule. Such "structural fingerprinting" gave better insight into the structure of the *IRES(−)* region. We showed that in the case of the IRES(−) fragment, which consists of 374 nucleotides, its three domains, *D3* (nucleotides 1−104), *DM* (nucleotides 105−222), and *D5* (nucleotides 223−374), independently fold on one another. However, when the IRES(−) molecule is extended by 25 nucleotides of the upstream viral sequence, domains *D3* and *DM* fold autonomously, but a part of domain *D5* interacts with that additional RNA stretch. Analysis *in silico* suggests that similar interactions involving the *IRES(−)* region and upstream sequences are also possible in other fragments of viral RNA, several hundreds of nucleotides in length. The results of experimental probing are supported by secondary structure predictions *in silico* and phylogenetic analysis.

Hepatitis C virus (HCV)[1] is a small, enveloped human virus that belongs to the *Flaviviridae* family. The viral genome consists of a single plus strand RNA molecule, approximately 9600 nucleotides in length, which includes a single large open reading frame and two untranslated regions at the 5′ and 3′ ends (5′UTR and 3′UTR) (*1*−*3*). Both untranslated regions are highly conserved, and they play important roles at the key steps of the viral life cycle: translation and replication. At the 5′UTR, there starts a cap-independent, *IRES*-directed translation that produces one large polyprotein, over 3000 amino acids long, which is processed into at least ten distinct proteins. Replication of the viral RNA consists of two steps. First, a complementary negative strand RNA copy (named replicative strand) is synthesized, starting from the 3′ terminus of the plus strand. It is used as a template in the next replication reaction, during which numerous genomic RNA strands are generated (*1*−*3*).

Replication of the viral RNA is carried out by the RNA-dependent RNA−polymerase complex with its major component being the NS5B protein. Within 3′-terminal regions of both genomic and replicative RNA strands, binding sites for NS5B (*2*) and for some other proteins, viral helicase NS3 (*4*), cellular protein HuR, and hnRNPC (*5*), have been mapped. These proteins usually recognize and bind to specific RNA secondary structures rather than to particular sequences. Therefore, a better knowledge of spatial folding of these regions of viral RNA is important for the understanding of HCV molecular biology. This will also be helpful in designing more effective antiviral agents, such as ribozymes, siRNAs, and antisense oligonucleotides, that might be used to target viral RNAs in these regions.

At the 3′ end of the genomic HCV strand, there is a highly conserved region X. It has been suggested that this region folds into a secondary structure consisting of three hairpin motifs (*6, 7*). Further studies have led to a four-hairpin model with a possibility of formation of a pseudoknot motif (*8*). For the 3′-terminal region of the HCV replicative strand, the *IRES(−)* region, two secondary structure models have been proposed (*9, 10*). In both models, a stretch consisting of ca. 220 nucleotides from the 3′ terminus and the one between nucleotides in positions 320−365 are folded in the same way. However, secondary structure arrangements proposed for the sequence 223−317, which is complementary to domain *IIIcdef* of *IRES(+)* in the genomic HCV strand, are significantly different in both propositions. Moreover, it has been suggested that these nucleotides might be involved in interactions with either nucleotides located close to the 3′ terminus (*10*) or with the region upstream of the *IRES(−)* element (*9*).

The aim of our studies was to better characterize folding of the *IRES(−)* region, especially in its central domain, controversial in the previously proposed models, and to investigate the ability of that domain to interact with distant regions either upstream or downstream in the replicative RNA strand. Results of a structural analysis of large RNA molecules by biochemical mapping methods are usually complex, and in many cases it is not possible to interpret them unambiguously. Therefore, we synthesized oligoribonucleotides that represented selected structural domains of the earlier proposed secondary structure models of the

[†] This work was supported by the Polish Ministry of Science and Higher Education (Grant 2P04A05128 to J.C.).

* To whom correspondence should be addressed. Tel: +48 61 8528503. Fax: +48 61 8520532. E-mail: ciesiolk@ibch.poznan.pl.

[1] Abbreviations: HCV, hepatitis C virus; IRES, internal ribosome entry site; RT, reverse transcription; UTR, untranslated region.

10.1021/bi800348g CCC: $40.75
Published on Web 10/24/2008

Table 1: Oligodeoxyribonucleotides Used in This Study for Construction of DNA Templates

dsDNA template	primers used in PCR reactions
IRES	DNA58 (gccagccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactac); DNA55 (ggtttttctt tgaggtttag gattcgtgct catggtgcac ggtctacgag acctc)
IRES(−)	DNA22a (gccagccccc gattgggggc ga); DNA43a (taatacgact cactataggt ttttctttga ggtttaggat tcg)
IRES(−)ext	DNA62 (taatacgact cactatagtg ggcggcggtt ggtgttacgt ttggtttttc tttgaggttt ag); DNA16 (gccagccccc gattgg)
D3	step I: DNA26 (ggacactcat actaacgcca tggcta); DNA22a step II: DNA43b (taatacgact cactatagga cactcatact aacgccatgg cta); DNA22a
DM	step I: DNA25 (ctccaggcat tgagcgggtt tatcc); DNA24 (gtgcagcctc caggaccccc cctc) step II: DNA42 (taatacgact cactatagtc caggcattga gcgggtttat cc); DNA24
D5	DNA22b (agatttgggc gtgcccccgc ga); DNA43a
D5ext	DNA22b; DNA62

IRES(−) region. Their structures were analyzed in solution, and the results were compared to those obtained with the full-length molecule. On the basis of such "structural fingerprinting" more precise conclusions concerning folding of the *IRES(−)* sequence could be formulated. Experimental structure probing data were supported by secondary structure predictions *in silico* and phylogenetic analysis.

EXPERIMENTAL PROCEDURES

Materials. The materials used in this study were from the following sources: (γ-^{32}P)ATP (4600 Ci/mmol) was from ICN, and all of the chemicals were from Sigma or Serva. Enzymes, Taq polymerase, ribonuclease T1, polynucleotide kinase, T7 RNA polymerase, ribonuclease inhibitor, and T4 RNA ligase, were purchased from MBI Fermentas.

DNA Templates and RNA Synthesis. All oligodeoxyribonucleotides used in the construction of DNA templates (Table 1) were deprotected after synthesis and purified on 12% polyacrylamide gels. DNA bands were excised, eluted with 0.3 M sodium acetate, pH 5.2, and 1 mM EDTA, and precipitated with 3 volumes of ethanol, and DNA was recovered by centrifugation and dissolved in TE buffer.

In order to obtain the dsDNA IRES template (encoding nucleotides 1−374 under the T7 promoter), the shorter fragment dsDNA5′UTR (1−341 nt) was used. It was synthesized by RT-PCR using viral RNA isolated from a patient infected with hepatitis C virus, type 1b. To this end, an approach similar to that earlier described by Chan and co-workers was applied (*11*). First, single-stranded DNA complementary to the 5′ portion of the HCV genome was synthesized using MMLV reverse transcriptase and primer DNA29 (gtgctcatgg tgcacggtct acgagacct). During the next step cDNA was amplified by PCR involving primers DNA22 (ggtgcacggt ctacgagacc tc) and DNA21 (gccagccccc gattgggggc g). As a result the dsDNA5′UTR (1−341 nt) was generated.

The dsDNA5′UTR fragment was reamplified by PCR with primers DNA58 and DNA55 (Table 1), and the reaction mixture contained 3 nM dsDNA, 0.25 μM both primers, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl, 200 μM each dNTP, and 1.25 units of Taq polymerase. The reaction was performed on a Biometra UNOII thermocycler for 20 cycles of 30 s at 94 °C, 30 s at 65 °C, and 2 min at 72 °C. The reaction products were purified by phenol/chloroform (1:1) extraction and precipitated with ethanol. The obtained dsDNA IRES template was dissolved in TE buffer. Subsequently, the dsDNA IRES and appropriate primers were used in a series of PCR reactions to obtain the templates used for RNA *in vitro* transcription (Table 1). The PCR reactions were performed as described above. Synthesis of dsDNA templates for D3 and DM RNAs was carried out in two steps. In the first step, dsDNAs encoding the desired RNA sequence was obtained, to which the T7 promoter was attached in the next step. All of the dsDNA templates were dissolved in TE buffer and used in transcription reactions.

The transcription reactions contained 0.4 μM dsDNA template, 40 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, 1 mM each NTP, 2.5 mM guanosine, 750 units/mL ribonuclease inhibitor, and 2000 units/mL T7 RNA polymerase (*8, 12*). Following 4 h incubation at 37 °C the RNA transcripts were purified on denaturing 8% polyacrylamide gels and labeled with ^{32}P at their 5′ or 3′ ends with polynucleotide kinase or RNA ligase according to standard procedures.

Metal Ion-Induced Cleavage. Prior to the cleavage reaction with Pb^{2+} ions the ^{32}P-end-labeled RNAs were renatured in the buffer, 40 mM NaCl, 10 mM Tris-HCl, pH 7.2, and 10 mM $MgCl_2$, by heating at 65 °C for 5 min and slow cooling to 37 °C (*13, 14*). The samples were then supplemented with tRNA carrier (Boehringer) to a final RNA concentration of 8 μM. Subsequently, lead acetate solution was added, and the reactions proceeded at 37 °C for 10 min. The reactions were terminated by mixing their aliquots with 8 M urea/dyes/20 mM EDTA solution, and samples were loaded on 12% polyacrylamide, 0.75% bisacrylamide, and 7 M urea gels. Electrophoresis was performed at 2000 V for 2−3 h, followed by autoradiography at −70 °C with an intensifying screen or phosphorimaging with a Typhoon 8600 analyzer (Molecular Dynamics).

Nuclease Mapping. The ^{32}P-end-labeled RNA transcripts were renatured and supplemented with tRNA carrier as described above. Limited digestions with ribonuclease T1 were carried out in the buffer: 10 mM Tris-HCl, pH 7.2, 40 mM NaCl, and 10 mM $MgCl_2$. Reactions were performed at 37 °C for different time intervals with 50 units/mL concentration of ribonuclease T1. The reactions were terminated by adding 8 M urea/dyes/20 mM EDTA solution and freezing samples on dry ice. The reaction products were analyzed by polyacrylamide gel electrophoresis and visualized by autoradiography or phosphorimaging.

Identification of Cleavage Sites. In order to assign the cleavage sites, products of metal ion-induced cleavage or nuclease digestion were run along with the products of alkaline RNA hydrolysis and limited T1 nuclease digestion of the same RNA under denaturing conditions. An alkaline hydrolysis ladder was generated by incubation of 5′- or 3′-end-^{32}P-labeled RNA with 5 volumes of formamide at 100 °C for 10 min. Partial T1 nuclease digestion was performed in 50 mM sodium citrate, pH 5.3, and 7 M urea with 0.2 unit of the enzyme at 55 °C for 10 min.

Secondary Structure Predictions in Silico. The mfold 3.2 computer program was used to predict RNA secondary structures (*15, 16*). The detailed description of the program and an algorithm used in thermodynamic calculations are

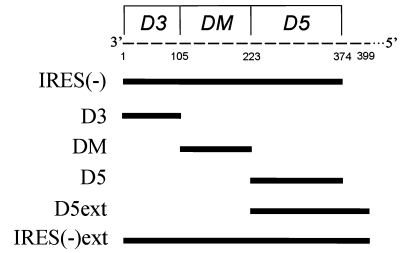

FIGURE 1: Fragments of the 3′-terminal region of the HCV RNA replicative strand. The IRES(−) RNA represents an oligomer that is complementary to the *IRES* region of the HCV RNA genome, and IRES(−)ext is its derivative with an additional 25 nucleotides of the upstream viral sequence. Their subfragments, D3, DM, D5, and D5ext, correspond to selected structural motifs of the earlier proposed secondary structure models of the *IRES(−)* region.

placed on the server http://www.fronted.bioinfo.rpi.edu/applications/mfold/rna/. One of the applications offered by the program is calculating ss-count values, which compute the probability of given nucleotides to appear in single-stranded RNA regions. To calculate the ss-count values, all of the secondary structures of the analyzed RNA sequence predicted *in silico* within 10% of free energy of the most stable variants are considered. The number of the predicted structures could vary depending on the folding parameters; default values proposed by the program were used.

Phylogenetic Analysis. In order to compare particular isolates of hepatitis C virus, type 1b, the Los Alamos HCV sequence database was used (*17*). It is available online on the Internet site http//hcv.lanl.gov/content/hcv-db. Sequences from the database shorter than the entire viral genome (but longer than 100 nucleotides) were also taken into account in this analysis. A total of 434 sequences were compared in the case of fragment *D5*, and 175 sequences were analyzed in the case of fragment *D5ext*.

RESULTS

Synthesis of RNA Fragments and Their Preparation for Structural Probing. Several RNA molecules, derivatives of the *IRES(−)* region of the HCV replicative strand, were synthesized by *in vitro* transcription using the T7 RNA polymerase system (Figure 1). In addition to the full-length IRES(−) (nucleotides 1−374; numbered consecutively from the 3′ end (position 1) to the 5′ end) and its extended derivative IRES(−)ext (nucleotides 1−399), their subfragments, D3, DM, D5, and D5ext, were also obtained. The subfragments corresponded to selected structural domains of the secondary structure models that have been proposed earlier for the *IRES(−)* region (*9, 10*).

Prior to structural probing, the ^{32}P-end-labeled RNA oligomers were subjected to a standard denaturation−renaturation procedure in the buffer 40 mM NaCl, 10 mM Tris-HCl, pH 7.2, and 10 mM $MgCl_2$ (*13, 14*). Although the concentration of Mg^{2+} ions was higher than 1−5 mM, which is thought to occur *in vivo* (*18*), such conditions are believed to better stabilize the spatial RNA folds and have been often used in RNA structural and functional studies. Electrophoresis of the renatured RNA molecules, IRES(−), IRES(−)ext, D5, and D5ext, in nondenaturing polyacrylamide gels showed major compact bands in each case suggesting structural homogeneity of the RNA samples (data not shown). The renatured RNAs were subjected to structural probing by the Pb^{2+}-induced cleavage method and limited digestion with ribonuclease T1.

Pb^{2+} Probing. In order to reveal the similarities and differences in the folding of IRES(−) and IRES(−)ext RNAs and their subfragments, the Pb^{2+}-induced cleavage method was used. This method has been shown to be very informative in detecting not only substantial differences in Watson−Crick base pairing but even subtle differences in the flexibility of RNA polynucleotide chains resulting from changes in stacking or tertiary interactions (*13, 14, 19, 20*). In particular, this approach has been used in structural studies of several RNAs, in which their subfragments were also employed, for example, 5S rRNAs (*21*), genomic and antigenomic HDV ribozymes (ref *22* and unpublished results of our laboratory), the 3′-terminal region of the genomic HCV RNA, and the 5′-terminal region of its replicative strand (*8*).

At first, we analyzed the IRES(−) RNA (nucleotides 1−374) and its subfragments, D3, DM, and D5, which were 105, 120, and 152 nucleotides in length, respectively (Figure 2). The length of the IRES(−) molecule was similar to the length of the secondary structure model of *IRES(−)* published earlier, consisting of 365 nucleotides, in which possible long-distance interactions between hairpin *I′* and domain *IIIcdef′* were proposed (*10*). In order to find out whether such interactions take place, we compared patterns of cleavages induced by Pb^{2+} ions in the full-length IRES(−) as well as in its subfragments. Representative autoradiograms from Pb^{2+}-induced cleavage experiments are shown in Figures S1, S2, and S3 (Supporting Information), and the cleavage sites and their relative intensities are displayed in Figure 2. Subsequently, the IRES(−)ext RNA and its fragment D5ext were also analyzed by the Pb^{2+} cleavage approach (Figure 3 and Figure S4 in Supporting Information). These RNA molecules were extended by 25 nucleotides upstream of the replicative strand, which might potentially base pair with nucleotides of domain *IIIcdef′*, similarly as in the model proposed by Schuster et al. (*9*).

(A) Full-Length IRES(−) RNA. All very strong cleavages induced in IRES(−) RNA in the presence of Pb^{2+} ions are mapped to single-stranded regions in the model shown in Figure 2, which is similar to the one proposed by Smith et al. (*10*). These are mainly cleavages in the bulge and internal loop regions of hairpin *IIz′*, in the junction region between hairpins *IIz′* and *IIy′*, and between *IIIa′* and *IIIb′*, in the apical and internal loops of *IIIb′*, several cleavages occur in the *IIIcdef′* region, in the junction between *IIIf′* and *IV′*, in the internal and apical loops of hairpin *IV′*, and, finally, in the 5′-terminal single-stranded stretch. Regions of IRES(−) resistant to Pb^{2+} degradation comprise mostly helical stems. The lack of cleavages in the stable tetraloop of *IIy′* and only weak cleavages in the apical loops of *I′*, *IIz′*, and *IIIa′* are consistent with the earlier observations concerning specificity of Pb^{2+}-induced degradation (*13, 14*) and the proposed secondary structure arrangement of these regions. Within the largest internal loop and bulge regions of hairpin *IIz′* several cleavages of moderate intensity were observed, with weak cleavages in the adjacent double-stranded stems. In general, hairpins *I′*, *IIy′*, and *IIIa′* are cleaved very weakly, presumably because of their stable secondary structures. Cleavages within the *IIIb′* and *IV′* regions support their secondary structures of imperfect hairpin type. The pattern of Pb^{2+}-induced cleavage of IRES(−) is essentially consistent with

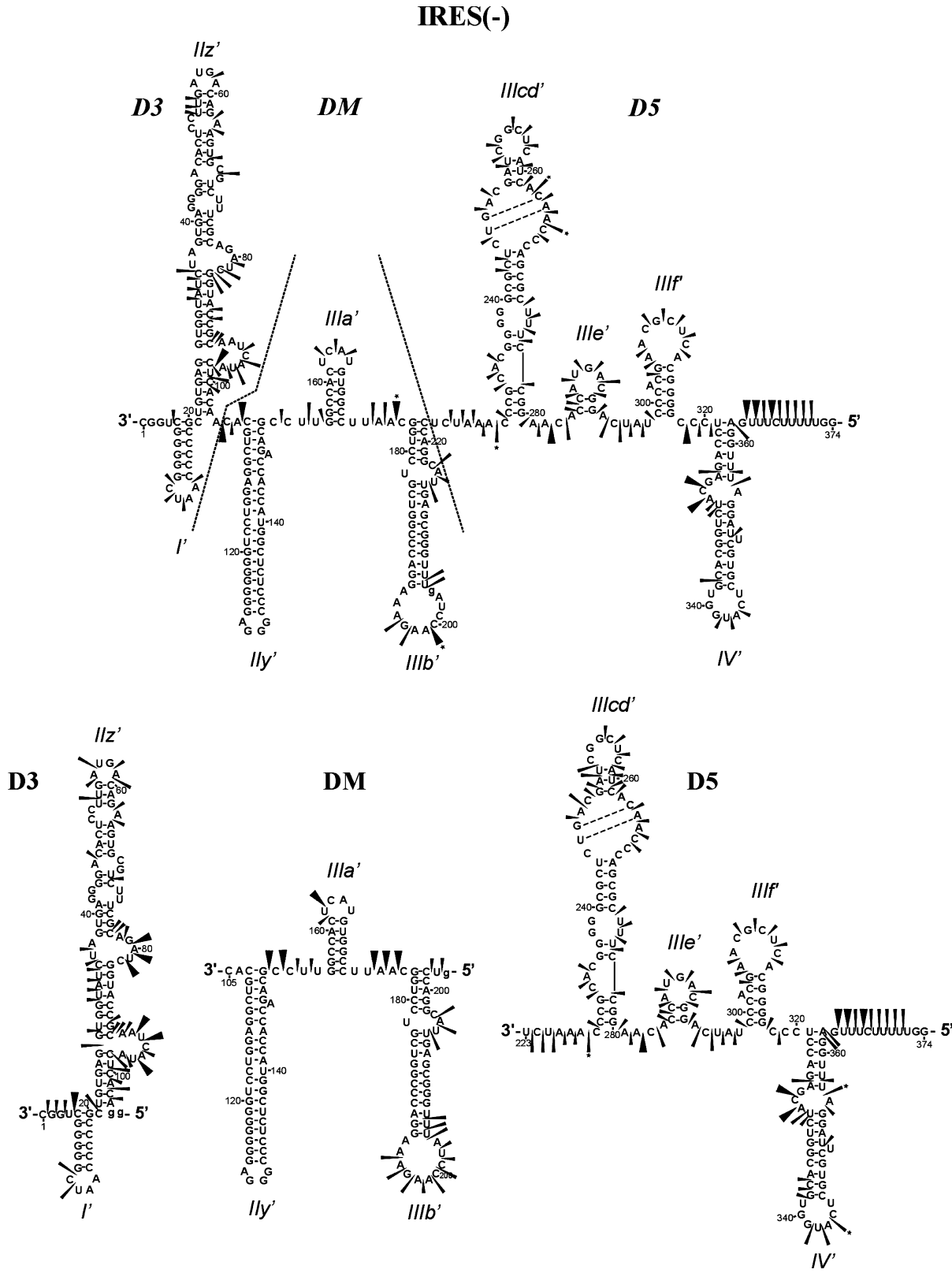

FIGURE 2: Structural analysis of the IRES(−) RNA and its subfragments, D3, DM, and D5, by the Pb^{2+}-induced cleavage method. Cleavages are displayed on the secondary structure models of the analyzed RNA molecules, and their relative intensities are correlated with the size of black triangles: small, weak; medium, strong; large, very strong cleavages.

the secondary structure model of this region proposed by Smith et al. (*10*) with the exception of domain *IIIcdef'*. Several cleavages are mapped to the double-stranded region of domain *IIIcdef'*, and only weak cleavages occur within internal loops of that domain. This prompted us to propose a new secondary structure arrangement of that domain, which is more compatible with our results of Pb^{2+} probing and computer predictions (Figure 2 and further discussion in the text).

(B) Isolated Domains D3, DM, and D5. The Pb^{2+}-induced cleavage pattern of D3 RNA is consistent with the secondary structure model of this fragment shown in Figure 2 taking

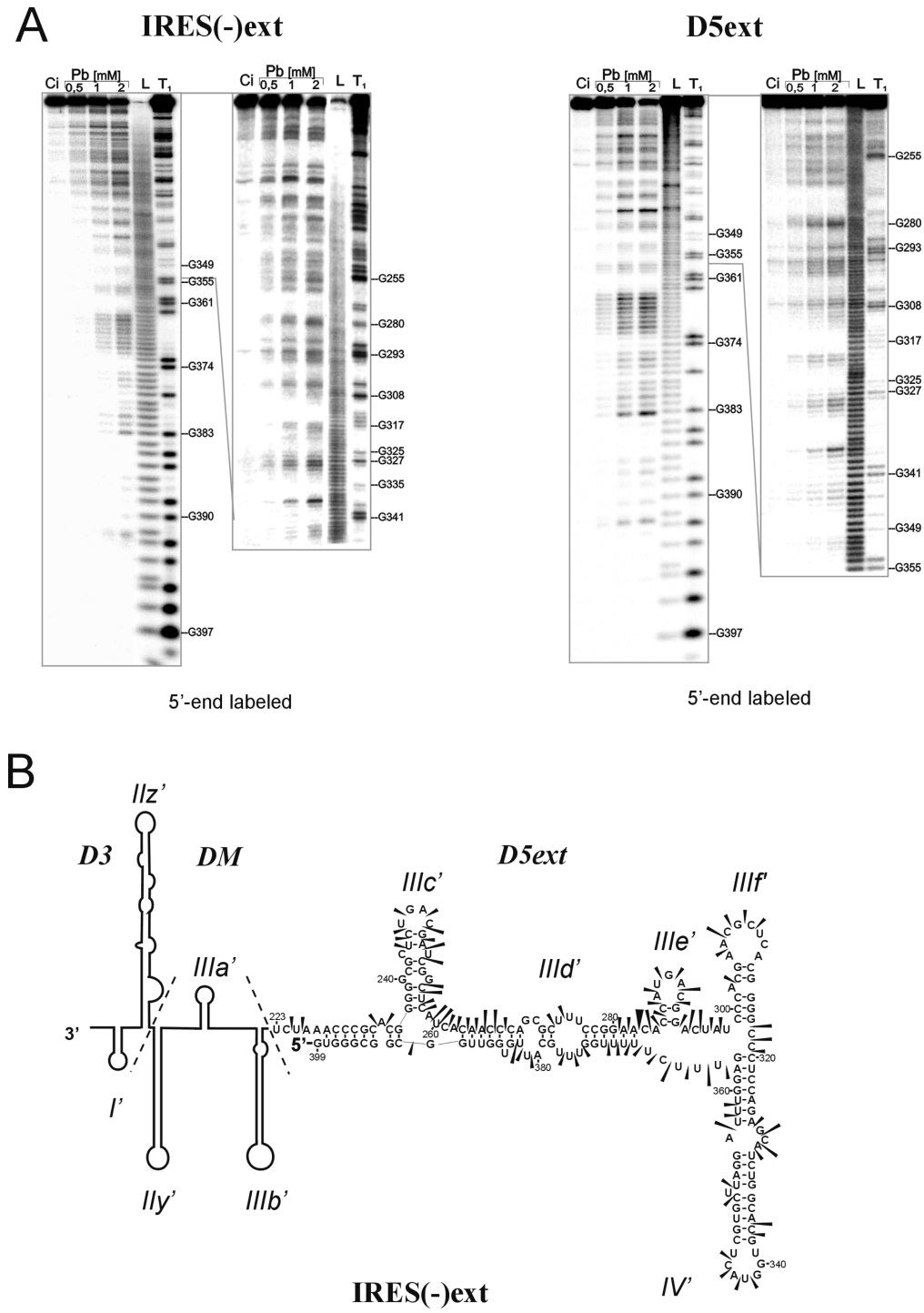

FIGURE 3: Probing of the structure of RNA oligomers IRES(−)ext and D5ext. (A) Autoradiograms show the products of Pb^{2+}-induced reactions analyzed on 12% polyacrylamide gels. Lanes: Ci, reaction control; L, formamide ladder; T_1, limited hydrolysis by RNase T1 in denaturing conditions. Selected guanine residues are labeled on the right. The reaction was performed with 5′-end-^{32}P-labeled RNA, and the short and long run of the gel is shown in the figure. (B) Cleavages induced in the *D5ext* region by Pb^{2+} ions are displayed on the secondary structure model of the IRES(−)ext RNA. Domains *D3* and *DM* are shown schematically. Relative intensities of Pb^{2+} cleavages are marked with black triangles as follows: small, weak; medium, strong; large, very strong cleavages.

into account our knowledge of Pb^{2+} cleavage specificity (*13, 14*). The strongest cleavages appear in internal loop and bulge regions. Weak cleavages appear in some double-stranded regions, for example, in the region between U32 and C35. However, these cleavages can be explained by relatively low stability of the secondary structure of that region consisting of AU and GU base pairs. The double-stranded stem of hairpin *I*′ is entirely resistant to cleavage, as expected. Most importantly, the cleavage pattern of D3 is very similar to the one characteristic of the corresponding region of the IRES(−) molecule.

The Pb^{2+}-induced cleavage pattern of DM RNA supports very nicely the presence of hairpins *IIy*′, *IIIa*′, and *IIIb*′. Their base-paired regions are generally not cleaved. Cleavages appear mainly in single-stranded junctions and apical loops with the exception of the purine-rich, stable tetraloop *IIy*′. Similarly, as in the case of D3, the cleavage pattern of DM

closely resembles the pattern observed for the corresponding region of the IRES(−) RNA.

Within D5 RNA numerous cleavages are observed mainly in the internal and apical loop regions, nearby the bulge nucleotides, between hairpins *IIIcd′* and *IIIf′* where a small hairpin *IIIe′* might form, and within the unstructured regions at both the 3′ and 5′ ends of the molecule. The secondary structure of D5, which is shown in Figure 2, differs from both the earlier propositions (*9, 10*) for arrangement of this region within the IRES(−) sequence. The structure was generated by the mfold3.2 program (*15, 16*) and was among those that exceeded the free energy of the most stable variant by less than 10%. Most of the observed Pb^{2+} cleavages are consistent with this arrangement of D5 RNA. The lack of cleavages at two GG dinucleotides, in positions 238−239 and 254−255, was expected since such stretches are usually not cleaved by Pb^{2+} ions (*13, 14*). Short double-stranded segments, those of hairpin *IIIe′* and adjacent to the apical loop of *IIIcd′*, are susceptible to cleavage, most likely because of their thermodynamic instability. Despite general correspondence of the experimental data and the secondary structure arrangement of D5, some inconsistencies still exist, mainly in the region of hairpin *IIIcd′*.

It is of importance that the Pb^{2+}-induced cleavage patterns found for oligomers DM, D3, and D5 are similar to those characteristic for the corresponding regions of IRES(−). It suggests that these sequences are folded into closely related secondary structures. It also means that these sequences fold independently within the IRES(−) molecule. As far as the possibility of interaction of domain *D5* with hairpin *I′*, which has been proposed earlier (*10*), is concerned, the comparison of Pb^{2+}-induced cleavage patterns of IRES(−) and D5 shows that such interactions do not occur. However, there were some difficulties in the interpretation of the autoradiograms of both the oligomers because of band compression effects observed for the 5′ side of hairpin *I′*. Thus we decided to use RNase T1 digestion in order to confirm the conclusions derived from Pb^{2+}-induced cleavage data (later in the text).

(C) Extended Domains: IRES(−)ext and D5ext. In order to find out whether folding of domain *D5* within the *IRES(−)* region is influenced by upstream sequences of viral RNA and, in particular, whether these sequences interact with nucleotides of region *IIIcdef′*, we synthesized IRES(−)ext and D5ext RNAs. These fragments were derivatives of IRES(−) and D5 molecules extended on their 5′ ends by a stretch of 25 nucleotides. The stretch was predicted *in silico* to be able to influence folding of the *IRES(−)* region (described further in the text). The IRES(−)ext and D5ext fragments were subjected to a structural analysis with the Pb^{2+}-induced cleavage approach, and the results are shown in Figure 3 and Figure S4 (Supporting Information).

In the IRES(−)ext, the Pb^{2+} pattern obtained for the 3′-terminal region comprising nucleotides 1−222 turns out to be very similar to that characteristic of the same region of IRES(−). Thus the additional stretch of 25 nucleotides does not seem to disturb folding of domains *D3* and *DM*. Moreover, in both the IRES(−)ext and D5ext molecules, the patterns obtained for the region 281−361 closely resemble each other and those patterns of the two types of RNAs within the corresponding regions of IRES(−) and D5. This suggests that these regions fold similarly and form imperfect hairpins *IIIf′* and *IV′*. Differences in the cleavage patterns of the four studied RNAs within their domains *D5* are illustrated in Figure 4 and Figure S5 (Supporting Information). They concern the stretches 223−280 and those located upstream of A362. It turns out that these differences are already revealed when the patterns of D5 and the corresponding region of D5ext are compared. Thus the additional stretch of 25 nucleotides extending domain D5 must be involved in its folding. This stretch itself is not evenly cleaved by Pb^{2+} ions, which may suggest an involvement of some nucleotides in base pairing. Similar observations can be made while comparing the cleavage patterns of D5 and IRES(−)ext (Figure 4). The 5′-terminal nucleotides of IRES(−)ext seem to participate in folding of domain *D5* in this RNA molecule (Figure 3B).

Digestion of IRES(−) and D5 with Ribonuclease T1. In order to verify the possibility of base pairing between nucleotides of hairpin *I′* and domain *IIIcdef′* (*10*), we compared the susceptibility of IRES(−) and D5 to digestion with ribonuclease T1. The enzymatic probing data are shown in Figure S6 (Supporting Information). The enzyme cleaves the *IIIcdef′* domain of D5 very similarly to that seen in the corresponding region of IRES(−), at G247, G250, G254, and G255. Two of these nucleotides, G250 and G254, have been earlier proposed to base pair with nucleotides of hairpin *I′* (*10*). In both D5 and IRES(−) RNAs the enzyme does not recognize the residues G232, G236−240, and G242. However, G232 is unpaired in the earlier suggested model of interactions, and this residue should have been susceptible to digestion. Thus our results do not support the proposed scheme of interactions between hairpin *I′* and domain *IIIcdef′*. Moreover, they also show that folding of the *IIIcdef′* region is not influenced by hairpin *I′* since the digestion patterns are essentially identical in both IRES and D5 RNAs, although only in the former molecule hairpin *I′* is present. The results of T1 probing are consistent with the new secondary structure model of domain *D5* (see the next chapter and Figure 2). Strong T1 cleavages are observed in the apical loops of hairpins: *IIIcd′*, *IIIe′*, *IIIf′*, and *IV′* at guanosine residues in positions 254, 289, 308, and 341 (Figure S6 in Supporting Information).

Predicted in Silico Secondary Structure of Domain D5 and a Possibility of Its Interaction with Upstream Regions of Viral RNA. We analyzed *in silico* the secondary structure folding of three fragments of IRES(−): the one corresponding to domain *IIIcdef′* (nucleotides 223−320), the entire domain *D5* deprived of two nucleotides at its 5′ end (nucleotides 223−372), and the same domain but extended in the 5′ direction by an additional 25 nucleotides of the viral RNA sequence (nucleotides 223−399). The mfold 3.2 version of Zuker's program was used (*15, 16*), and all secondary structures deviating by less than 10% from the free energy of the most stable structure were considered. The results of computer predictions are illustrated in Figure S7 (Supporting Information) as ss-count values, which show the calculated propensity of given nucleotides to be single-stranded. It turns out that stable hairpin *IIIf′* and imperfect hairpin *IV′* are well predicted by the program. There is also a double-stranded element formed by some nucleotides of domain *IIIcde′* and nucleotides located upstream of hairpin *IV′*. This element appears only in the longest analyzed RNA fragment, and it has been earlier proposed in Schuster's model of the *IRES(−)* region in a similar but not identical form. Hairpin *IV′* exists

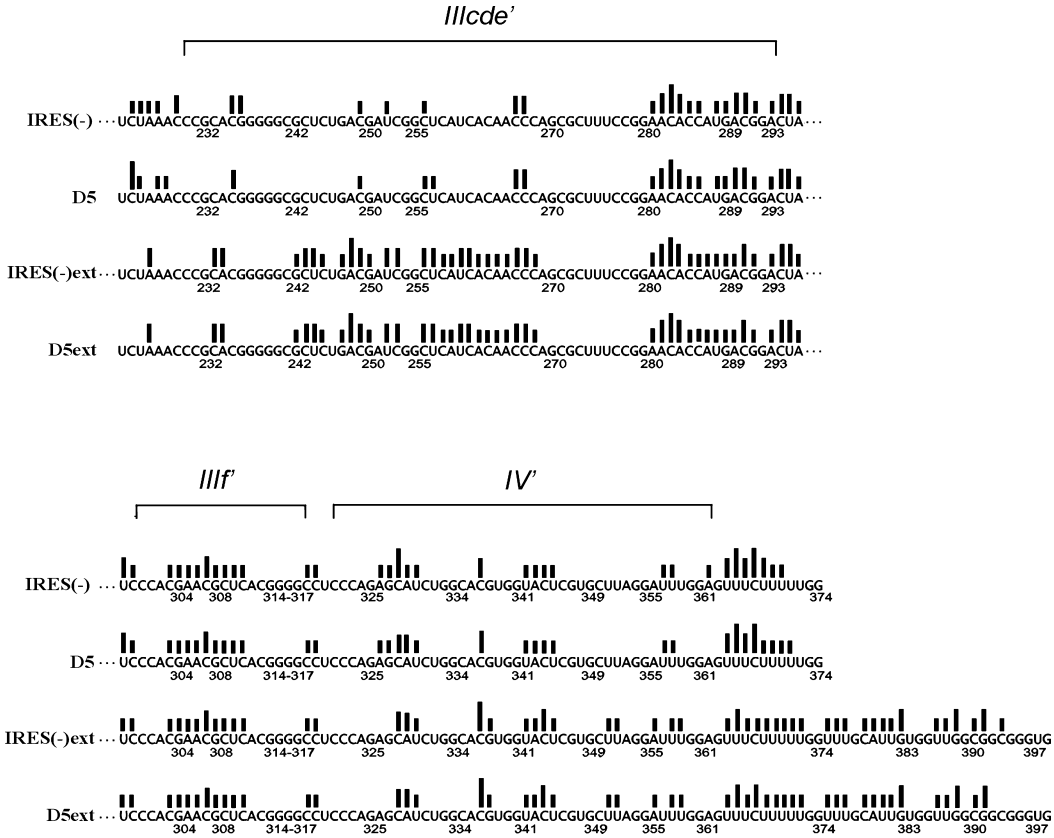

FIGURE 4: Similarities and differences in the Pb^{2+}-induced cleavage patterns, which were obtained with IRES(−), IRES(−)ext, D5, and D5ext, in the regions corresponding to their domains *D5/D5ext* (nucleotides 223−374/399). Cleavages induced by Pb^{2+} ions in the regions between nucleotides 223 and 374/399 are displayed on linear sequences of the examined RNAs. Relative intensities of cleavages are correlated with heights of black bars.

in both Schuster's and Smith's models, and hairpin *IIIf′* is present in Schuster's model. Although the results of ss-count analysis give only a rough illustration of the tendency of particular RNA regions to form base-pairing interactions, they are very helpful in interpretation of experimental data.

In order to identify possible secondary interactions between the *IIIcdef′* region of *IRES(−)* and other regions located in the 5′ direction of viral RNA, we performed computer simulations using the mfold 3.2 program (*15, 16*). In this analysis the consensus sequence of HCV RNA genotype 1b was used. We examined the predicted secondary structures of several RNA fragments: 200, 300, 400, 500, 600, 700, and 800 nucleotides in length. These fragments corresponded to the following regions of viral RNA: 223−423, 223−523, 223−623, 223−723, 223−823, 223−923, and 223−1023. The analysis revealed that in each fragment there occur interactions, predicted with high probability, between a part of domain *IIIcdef′* (nucleotides 263−282) and the region spanning nucleotides at positions 369−389. These interactions were found in RNA secondary structures with free energies differing by less than 5% from the free energy of the most stable variant (data not shown).

Phylogenetic Analysis. In the analysis of variability of the *IRES(−)* region, over 400 (434) viral sequences deposited in the Los Alamos HCV sequence database (*17*) were considered. In the region corresponding to domain *IIIcdef′* single mutations were found in 82 sequences, including insertions, deletions, and substitutions. The results are displayed on the secondary structure model of that domain, which is most consistent with the results of our structural analysis, in the upper panel of Figure 5. In most cases, cytosine is replaced with uridine. A hot spot of this type of transition mutation is observed at nt 243 (32 ×, in 7% of the sequences), and uridine at this position changes a C-G pair to a U-G wobble pair in this stem region. This is an example of a mutation in the *IIIcd′* region that basically preserves the secondary structure, which also holds for mutations C271g and at C273g. In the considered region several other mutations occur in single-stranded RNA stretches (for example, A248g in 17 sequences). Such changes do not confirm (but also do not rule out) the newly proposed model of the secondary structure of domain *IIIcdef′*.

In order to evaluate the possibility of base pairing between domain *IIIcde′* and sequences located in the 5′ direction of the replicative strand, we compared 175 different viral RNAs between nucleotides 223 and 399 (Figure 5, bottom panel). The most frequent mutations found in the *IIIcde′* region include the following substitutions, C243u (in 8 sequences, i.e., in 5%) and A248g (in 7 sequences, 4%), both located at unpaired positions in hairpin *IIIc′*. Other mutations occur sporadically, one to three times (0.6−1.7%). In the 5′-proximal part between positions 363 and 399, mutations are found with frequencies of 1 to 11 times per nucleotide position (0.6−6%). Here, changes are most frequently observed at positions 370 (5 ×) and 392 (11 ×), but they disrupt Watson−Crick base pairing.

In the hairpin *IIIf′* region only two single mutations are found: A312u or A312c in the apical loop. Finally, several mutations found in hairpin *IV′* are consistent with the secondary structure of this region: C350u (occurring in 30%

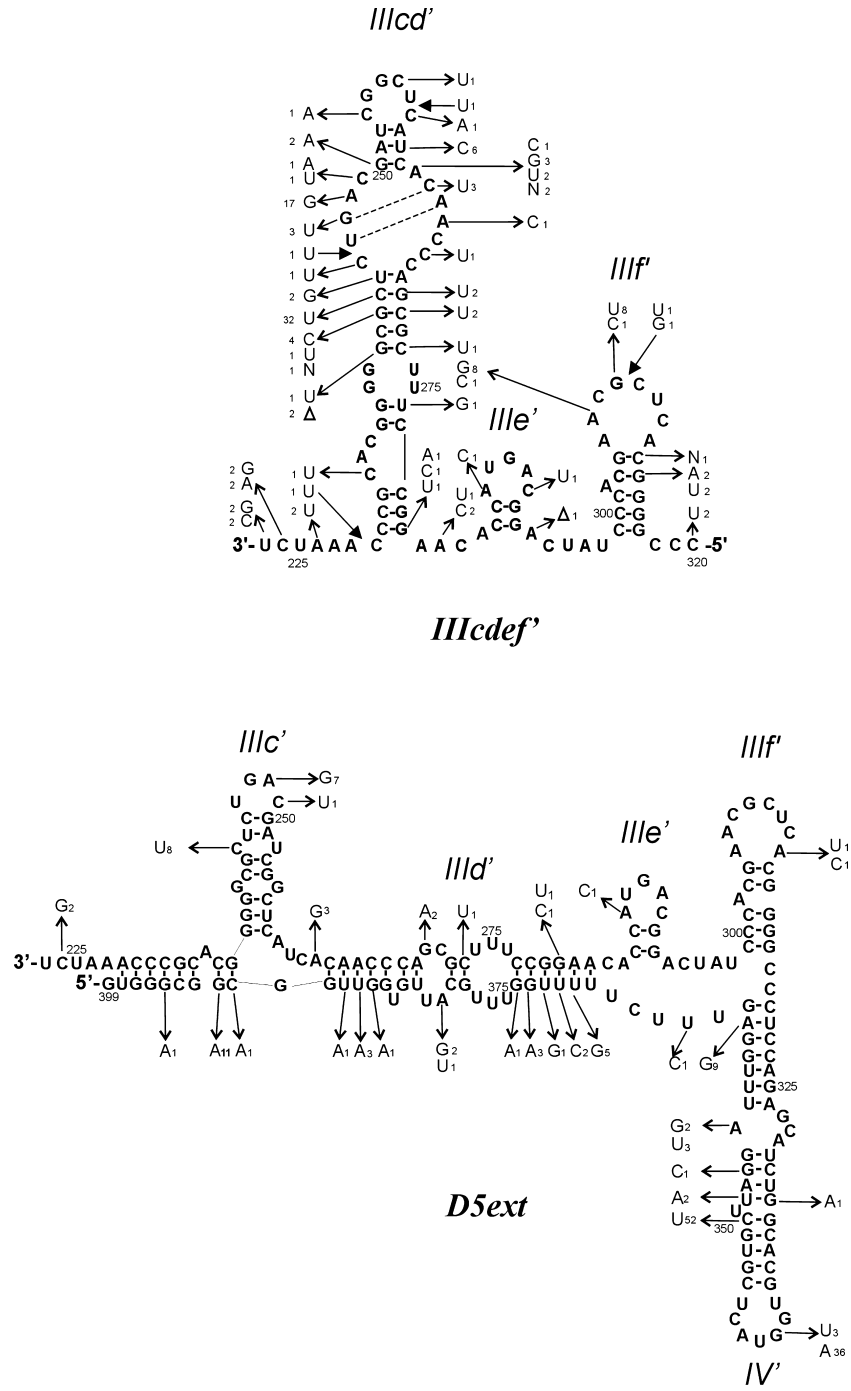

FIGURE 5: Phylogenetic analysis of the sequences of the *IRES(−)* region of the HCV RNA replicative strand corresponding to the *IIIcdef'* region of domain *D5* (nucleotides 223−320) and to domain *D5ext* (nucleotides 223−399). Different numbers of sequences were analyzed, 434 and 175, respectively. Thin arrowheads indicate substitutions, filled arrowheads indicate insertions, and triangles indicate deletions. The numbers, which accompany the nucleotide symbols, indicate in how many isolates the particular mutation was found. The nucleotide exchanges/ insertions/deletions differ in the upper and lower panel as a consequence of two different sets of sequences taken for analysis (see also the Experimental Procedures section).

of the sequences), A362g (in 5% of the sequences), in the apical loop at G341 (36 × adenosine, i.e., in 21% of the sequences, 3 × uridine), in the internal loop at A356 (guanosine or uridine, 2 or 3 times, respectively), one compensative mutation G333a, and U352a (2 ×) and G354c (1 ×) both disrupting base pairing.

DISCUSSION

Three Structural Domains, D3, DM, and D5, Fold Independently within the IRES(−) Region. For the 3′-terminal region of the replicative strand of HCV RNA, the *IRES(−)* region, two secondary structure models have been proposed by Schuster et al. (*9*) and Smith et al. (*10*) (see the introduction). In order to solve some uncertainties concerning their propositions, we analyzed folding of three RNA oligomers, D3, DM, and D5, which represented selected structural domains of the earlier proposed secondary structure models of that region (Figure 1). Their structures were probed by Pb^{2+}-induced cleavage approach, which has been shown to be able to reveal even subtle differences in conformation of an RNA chain (*14*). The results of structural mapping obtained for the D3, DM, and D5 fragments were compared

to those obtained with the IRES(−) RNA (Figure 2). It turned out that the patterns of cleavages induced in the RNA isolated domains were very similar to those obtained for the corresponding regions of the full-length molecule. Therefore, it seems that all three domains fold independently in the structural context of the *IRES(−)* region. This conclusion contradicts an earlier suggestion of Smith et al. (*10*) that domain *D5* and hairpin *I′* may interact with each other.

In order to support the above conclusion, the digestion patterns of D5 and IRES(−) RNAs generated by ribonuclease T1 were compared (Figure S6 in Supporting Information). The comparison of T1 cleavage patterns did not support the proposed interactions between domain *D5* and hairpin *I′*. The positions and intensities of cleavages at the corresponding positions of D5 and IRES(−) were identical. One could expect some differences if the suggested interactions exist since they would only be possible in a full-length molecule.

The IIIcd′ Region of Domain D5 Interacts with the Neighboring Upstream Sequence of Viral RNA. In Schuster's model of *IRES(−)* secondary structure (*9*) nucleotides of domain *IIIcd′* interact with the neighboring nucleotides located upstream in the replicative RNA strand. We showed that such interactions could also be predicted with high probability by the mfold computer program. Subsequently, the Pb^{2+}-induced cleavage method was used for structural mapping of two RNAs, IRES(−)ext and D5ext, which were derivatives of IRES(−) and D5 oligomers but were extended by 25 nucleotides in the 5′ direction of viral RNA sequence. Direct comparison of Pb^{2+}-induced cleavage patterns of IRES(−)ext and D5ext with those characteristic to IRES(−) and D5 oligomers shows that region *IIIcd′* interacts with the 25 nucleotide long stretch located at the 5′ terminus of extended RNA fragments (Figures 3 and 4). On the other hand, very similar Pb^{2+}-induced cleavage patterns were observed for the 3′-terminal 222 nucleotides, which correspond to domains *D3* and *DM*. Thus clearly only domain *IIIcd′* of domain *D5* is involved in interactions with sequences located upstream of the *IRES(−)* element in the replicative strand of HCV RNA.

The secondary structure of the *IRES(−)* region has also been mapped by Pb^{2+}-induced cleavage method by Schuster et al. (*9*). It has to be noted that IRES(−)ext used in our studies comprised 3′-terminal 399 nucleotides whereas their fragment had 416 nucleotides to which four and five extra nucleotides were attached at the 5′ and 3′ ends. In spite of these differences, the results of our studies were very similar to the earlier published data. The only region for which clearly distinct Pb^{2+}-induced cleavage patterns were observed was hairpin *I′*. Schuster and co-workers (*9*) observed several strong cleavages in the apical loop of the hairpin and on the 5′ side of the double-stranded stem while in our studies this region was hardly susceptible to digestion. This might be the result of slightly different reaction buffers or conditions of electrophoretic separation. Although unlikely, it cannot be excluded that additional sequences present at the 3′ and 5′ ends in the construct of Schuster et al. (*9*) might have influenced folding of hairpin *I′*, resulting in a different cleavage pattern.

The Secondary Structure of Domain D5. The secondary structure of 3′-terminal 222 nucleotides of the *IRES(−)* region, which correspond to domains *D3* and *DM*, is well characterized (refs (*9*) and (*10*) and this study). This region does not interact with the neighboring upstream sequences of the viral RNA. Therefore, folding of domain *D5* seems to be responsible for the controversies concerning the structure of *IRES(−)*. Predicted *in silico* secondary structure models of three fragments of viral RNA, which began with U223 but had increased in length toward the 5′ direction, showed that in all the fragments stable hairpin *IIIf′* is formed (Figure S7 in Supporting Information). In two longer fragments hairpin *IV′* also exists. In two earlier proposed secondary structures of the *IRES(−)* region hairpin *IV′* is formed in both models, Schuster's and Smith's, while in the former one hairpin *IIIf′* is also present.

Our results of the structural analysis of IRES(−), IRES(−)ext, and their isolated fragments D5 and D5ext confirmed the presence of two imperfect hairpins, *IIIf′* and *IV′*, in all of these RNAs. Folding of the 3′ part of domain *D5* (nucleotides 223−280) depends on a possibility of its interaction with the upstream viral sequence. In the short IRES(−) construct such interactions cannot be formed. We propose a new structural arrangement of that region (Figure 2), which differs from that shown in the model of Smith et al. (*10*) for the 3′-terminal 365 nucleotides of the viral replicative strand. Our proposition is more consistent with the experimental data and computer predictions. In the IRES(−)ext molecule the 3′ part of domain *D5* (nucleotides 223−280) interacts with the upstream viral stretch spanning nucleotides 367 and 399. Our results of experimental probing of IRES(−)ext support the secondary structure arrangement proposed by Schuster et al. (*9*) for 3′-terminal 416 nucleotides of the RNA replicative strand. It is of importance that in both the extended RNA molecules (the one described by Schuster et al. and IRES(−)ext used in our studies) domains *D3* and *DM* as well as hairpins *IIIf′* and *IV′* of domain *D5* are folded in the same way as in the short *IRES(−)* region.

Phylogenetic analysis of the sequence between nucleotides 223 and 374 in various RNA isolates of HCV genotype 1b was also performed. However, not too many mutations were found in this highly conserved region (Figure 5). One frequent mutation that preserves base pairing occurs in domain *IIIcdef′* at position 243, thus supporting our new proposition of the secondary structure arrangement for this region. In the longer RNA stretch spanning nucleotides 223−399, in which domain *IIIcdef′* interacts with sequences upstream the IRES(−) element, nucleotide 243 occurs in a non-W−C base pair position. Two mutations are also present in the hairpin *IV′* region: C350u and A362g. Both of these mutations preserve the proposed secondary structure folding of that region into an imperfect hairpin motif with an asymmetric internal loop (Figure 5).

Involvement of the IRES(−) Region in the Replication Process. Several experiments have been performed which aimed at the localization of the promoter region in the replicative RNA strand of HCV and determination of the minimal length of RNA able to act in transcription. Most information has been obtained *in vitro* with the use of recombinant protein NS5B and fragments of viral RNA of different length. Initially, it has been shown that the 3′-terminal 122 nucleotides do not allow initiation of transcription with NS5B *in vitro*, which suggested that the promoter sequence is localized between nucleotides 122 and 239 (*23*). According to the results of our structural studies this sequence is a part of domains *DM* and *D5*. On the other

hand, Kashiwagi et al. have shown that the region 40−130 acts as a positive transcription control and possibly a place of polymerase binding (*24*). However, much higher transcription efficiency occurs with longer templates consisting of nucleotides 1−341 (*23*) or 40−341 (*24*). An experiment performed *in vitro* with purified enzyme, replicative RNA strand, and several antisense DNA oligomers has shown that the highest inhibition of initiation of RNA synthesis occurred with an oligomer hybridizing to the region 85−103. Most likely, the oligomer does not block a polymerase binding site directly but acts indirectly by causing structural rearrangements of hairpins *I′* and *IIz′* that interfere with initiation of RNA synthesis (*25*). Interesting results on binding of purified polymerase NS5B to different fragments of the *IRES(−)* region and their use as RNA templates have been described by Astier-Gin et al. (*26*). It turns out that within the region of the 3′-terminal 219 nucleotides there is no strong polymerase-binding site whereas deletion of a stretch between nucleotides 219 and 239 severely disturbs enzyme binding and RNA synthesis (*26*). The significant role of the region 219−239 has been confirmed in the study of Reigadas et al. (*25*). Three out of four antisense oligomers, which showed the highest inhibitory properties, were in part complementary to this region. The region 219−239 is a fragment of domain *D5* that was extensively studied in this report. In RNA molecules D5ext and IRES(−)ext the stretch spanning nucleotides 228 and 236 is involved in interactions with the 5′-terminal sequence 391−398 forming a double-stranded stem composed of eight base pairs (Figure 3). This segment might be of great importance for the binding of such viral proteins like NS3 and NS5B that are the major components of the replication complex, and both have affinity to double-stranded RNA stems. For viral protein NS3, that acts as helicase, a binding site within the IRES(−) region has been already proposed (*27*). It seems that the protein requires the integrity of the double-stranded stem of the hairpin *I′* to bind (*27*). The autonomous folding of hairpin *I′*, evidenced in our study, corresponds well to the functional role of this region. Structural properties of the *D3* and *D5* domains seem to be important for the explanation of their role in viral replication. Since folding of these domains turned out to be responsible for earlier controversies concerning the secondary structure of the 3′ terminus of HCV replicative strand, our results should be helpful in further studies of the role of these regions in the functioning of the virus as well as in finding nucleic acid-based therapeutics against HCV infections.

ACKNOWLEDGMENT

We thank Barbara Smólska for expert technical assistance and Jan Wrzesinski for critical review of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Probing of the structure of RNA IRES(−) (Figure S1), RNA fragments D3 and DM (Figure S2), RNA fragment D5 (Figure S3), and RNA fragment D5ext (Figure S4) by Pb^{2+}-induced cleavage method, comparison of the Pb^{2+}-induced cleavage patterns obtained for IRES(−), IRES(−)ext, D5, and D5ext in the regions corresponding to their domains *D5/D5ext* (Figure S5), comparison of the cleavage patterns induced by RNase T1 in the corresponding regions of RNA IRES(−) and D5 (Figure S6), and secondary structure models of fragments of the *IRES(−)* region predicted *in silico* (Figure S7). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Penin, F., Dubuisson, J., Rey, F. A., Moradpour, D., and Pawlotsky, J. M. (2004) Structural biology of hepatitis C virus. *Hepatology 39*, 5–19.
2. Shi, S. T., and Lai, M. M. (2006) HCV 5′ and 3′UTR: When Translation meets Replication, in *Hepatitis C Viruses* (Seng-Lai, T., Ed.) pp 49−88, Horizon Bioscience, Norfolk.
3. Chevaliez, S., and Pawlotsky, J. M. (2006) HCV Genome and Life Cycle, in *Hepatitis C Viruses* (Seng-Lai, T., Ed.) pp 5−48, Horizon Bioscience, Norfolk.
4. Banerjee, R., and Dasgupta, A. (2001) Specific interaction of hepatitis C virus protease/helicase NS3 with the 3′-terminal sequences of viral positive- and negative-strand RNA. *J. Virol. 75*, 1708–1721.
5. Spangberg, K., Wiklund, L., and Schwartz, S. (2000) HuR, a protein implicated in oncogene and growth factor mRNA decay, binds to the 3′ ends of hepatitis C virus RNA of both polarities. *Virology 274*, 378–390.
6. Ito, T., and Lai, M. M. (1997) Determination of the secondary structure of and cellular protein binding to the 3′-untranslated region of the hepatitis C virus RNA genome. *J. Virol. 71*, 8698–8706.
7. Blight, K. J., and Rice, C. M. (1997) Secondary structure determination of the conserved 98-base sequence at the 3′ terminus of hepatitis C virus genome RNA. *J. Virol. 71*, 7345–7352.
8. Dutkiewicz, M., and Ciesiołka, J. (2005) Structural characterization of the highly conserved 98-base sequence at the 3′ end of HCV RNA genome and the complementary sequence located at the 5′ end of the replicative viral strand. *Nucleic Acids Res. 33*, 693–703.
9. Schuster, C., Isel, C., Imbert, I., Ehresmann, C., Marquet, R., and Kieny, M. P. (2002) Secondary structure of the 3′ terminus of hepatitis C virus minus-strand RNA. *J. Virol. 76*, 8058–8068.
10. Smith, R. M., Walton, C. M., Wu, C. H., and Wu, G. Y. (2002) Secondary structure and hybridization accessibility of hepatitis C virus 3′-terminal sequences. *J. Virol. 76*, 9563–9574.
11. Chan, S.-W., McOmish, F., Holmes, E. C., Dow, B., Peutherer, J. F., Follett, E., Yap, P. L., and Simmonds, P. (1992) Analysis of a new hepatitis C virus type and its phylogenetic relationship existing variants. *J. Gen. Virol. 73*, 1131–1141.
12. Łęgiewicz, M., Wichłacz, A., Brzezicha, B., and Ciesiołka, J. (2006) Antigenomic delta ribozyme variants with mutations in the catalytic core obtained by the in vitro selection method. *Nucleic Acids Res. 34*, 1270–1280.
13. Ciesiołka, J., Michałowski, D., Wrzesiński, J., Krajewski, J., and Krzyżosiak, W. J. (1998) Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. *J. Mol. Biol. 275*, 211–220.
14. Kirsebom, L. A. and Ciesiołka, J. (2005) Pb^{2+}- induced Cleavage of RNA, in *Handbook of RNA Biochemistry* (Hartmann, R. K., Bindereif, A., Schon, A., and Westhof, E., Eds.) pp 214−228, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
15. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res. 31*, 3406–3415.
16. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol. 288*, 911–940.
17. Kuiken, C., Hraber, P., Thurmond, J., and Yusim, K. (2008) The hepatitis C sequence database in Los Alamos. *Nucleic Acids Res. 36*, D512–D516.
18. Traut, T. W. (1994) Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem. 140*, 1–22.
19. Ciesiołka, J., Lorenz, S., and Erdmann, V. A. (1992) Structural analysis of three prokaryotic 5S rRNA species and selected 5S rRNA-ribosomal-protein complexes by means of Pb(II)-induced hydrolysis. *Eur. J. Biochem. 204*, 575–581.
20. Ciesiołka, J., Lorenz, S., and Erdmann, V. A. (1992) Different conformational forms of *Escherichia coli* and rat liver 5S rRNA revealed by Pb(II)-induced hydrolysis. *Eur. J. Biochem. 204*, 583–589.

21. Ciesiołka, J., and Krzyżosiak, W. J. (1996) Structural analysis of two plant 5S rRNA species and fragments thereof by lead-induced hydrolysis. *Biochem. Mol. Biol. Int. 39*, 319–328.
22. Matysiak, M., Wrzesiński, J., and Ciesiołka, J. (1999) Sequential folding of the genomic ribozyme of the hepatitis delta virus: structural analysis of RNA transcription intermediates. *J. Mol. Biol. 291*, 283–294.
23. Oh, J. W., Ito, T., and Lai, M. M. (1999) A recombinant hepatitis C virus RNA-dependent RNA polymerase capable of copying the full-length viral RNA. *J. Virol. 73*, 7694–7702.
24. Kashiwagi, T., Hara, K., Kohara, M., Iwahashi, J., Hamada, N., Honda-Yoshino, H., and Toyoda, T. (2002) Promoter/origin structure of the complementary strand of hepatitis C virus genome. *J. Biol. Chem. 277*, 28700–28705.
25. Reigadas, S., Ventura, M., Andeola, M. L., Michel, J., Gryaznov, S., Tarrago-Litvak, L., Litvak, S., and Astier-Gin, T. (2003) An oligonucleotide complementary to the SL-B1 domain in the 3′-end of the minus-strand RNA of the hepatitis C virus inhibits in vitro initiation of RNA synthesis by the viral polymerase. *Virology 314*, 206–220.
26. Astier-Gin, T., Bellecave, P., Litvak, S., and Ventura, M. (2005) Template requirements and binding of hepatitis C virus NS5B polymerase during in vitro RNA synthesis from the 3′-end of virus minus-strand RNA. *FEBS J. 272*, 3872–3886.
27. Banerjee, R., and Dasgupta, A. (2001) Specific interaction of hepatitis C virus protease/helicase NS3 with the 3′-terminal sequences of viral positive- and negative-strand RNA. *J. Virol. 75*, 1708–1721.

BI800348G