Acta virologica 56: 217 – 226, 2012 doi:10.4149/av_2012_03_217

The adenine-rich tract in the 5′-end of the hepatitis C virus ORF encodes a peptide regulating the binding of the C protein to RNA

S. CHINNASWAMY[1], W. B. LOTT[2]

[1]National Institute of Biomedical Genomics, P. O.: N. S. S., 2nd Floor, Kalyani, West Bengal-741251, India; [2]Institute for Health and Biomedical Innovation, School of Life Sciences, Queensland University of Technology, Brisbane, QLD 4001, Australia

Received February 3, 2012; accepted August 17, 2012

Summary. – Hepatitis C virus (HCV) core (C) protein is thought to bind to viral RNA before it undergoes oligomerization leading to RNA encapsidation. Details of these events are so far unknown. The 5′-terminal C protein coding sequence that includes an adenine (A)-rich tract is a part of an internal ribosome entry site (IRES). This nucleotide sequence but not the corresponding protein sequence is needed for proper initiation of translation of viral RNA by an IRES-dependent mechanism. In this study, we examined the importance of this sequence for the ability of the C protein to bind to viral RNA. Serially truncated C proteins with deletions from 10 up to 45 N-terminal amino acids were expressed in *Escherichia coli*, purified and tested for binding to viral RNA by a gel shift assay. The results showed that truncation of the C protein from its N-terminus by more than 10 amino acids abolished almost completely its expression in *E. coli*. The latter could be restored by adding a tag to the N-terminus of the protein. The tagged proteins truncated by 15 or more amino acids showed an anomalous migration in SDS-PAGE. Truncation by more than 20 amino acids resulted in a complete loss of ability of tagged C protein to bind to viral RNA. These results provide clues to the early events in the C protein - RNA interactions leading to C protein oligomerization, RNA encapsidation and virion assembly.

Keywords: hepatitis C virus; core protein; RNA binding

Introduction

HCV is a positive-sense single stranded RNA virus in the genus *Hepacivirus* of the family *Flaviviridae* (Robertson *et al.*, 1998). It is responsible for the majority of non-A, non-B viral hepatitis infections. HCV causes persistent infection that can lead to cirrhosis of the liver and hepatocellular carcinoma (Cuthbert, 1994). The approximately 9.6 kb genome contains a single long open reading frame (ORF), which codes for a polyprotein of about 3,000 amino acids, flanked by long and highly structured 5′- and 3′-UTRs. HCV initiates translation via a cap-independent non-scanning mechanism that utilizes an IRES to recruit and assemble the ribosome directly at the start site (Rijnbrand and Lemon, 2000). The HCV IRES is a complex RNA structure located largely in the 5′-UTR but is believed to extend into the ORF (Lu and Wimmer, 1996; Reynolds *et al.*, 1996). The exact 3′ boundary of the HCV IRES is unknown. Since the HCV IRES-ORF sequence is used both for translational control and to code for the amino acid residues in the N-terminus of the HCV polyprotein, the involvement of the ORF in translation initiation suggests a novel dual use of the genome.

IRES-dependent translation initiation requires the 5′-end of the ORF to be free of complex secondary structure, and an adenine (A)-rich sequence downstream of the start codon is incorporated to preclude such structure (Rijnbrand *et al.*, 2001). This (A)-rich region is conserved among HCV geno-

E-mail: sc2@nibmg.ac.in; phone: +91-33-25892151(256).

Abbreviations: AM = anomalous migration; ARF = alternate reading frame; C protein = core protein; C124 = N terminal 124 amino acids of C protein; GEMSA = gel electrophoresis mobility shift assay; HCV = hepatitis C virus; IRES = internal ribosome entry site; IEC = ion exchange chromatography; NLP = nucleocapsid-like particles; MAb = monoclonal antibody; ORF = open reading frame; SEC = size exclusion chromatography

types (Fig. 1a) and is also found in the analogous region of the ORF of viruses in the genus *Pestivirus*, a closely related genus in the family *Flaviridae* that initiates viral translation by an identical IRES mechanism (Rijnbrand *et al.*, 2001; Fletcher and Jackson, 2002). Despite the apparent sequence conservation, the (A)-rich regions of HCV and *Pestivirus* ORFs code for peptide sequences that give rise to functionally distinct proteins. The (A)-rich region of the HCV ORF codes for amino acid residues in the N-terminus of the nucleocapsid (C or core) protein, while the homologous pestivirus region codes for amino acid residues in the N-terminus of N-pro, an autoprotease that is absent in HCV polyprotein. Further, the N-terminus of N-pro that is coded by the (A)-rich region can be truncated without affecting N-pro′s autoproteolytic activity (Rumenapf *et al.*, 1998), suggesting that the (A)-rich region in the pestiviruses may be more important for translation control than for protein coding. In addition to its role in evading the host immune response (Gil *et al.*, 2006), N-pro′s autoproteolytic role removes a potentially superfluous fusion peptide encoded by the ORF portion of the pestivirus IRES and provides the mature N-terminus of the pestivirus C protein. These observations question the role of the N-terminus of the HCV C protein. Is this a true example of dual use of the HCV genome, or is HCV tolerating an unwanted fusion peptide, whose coding sequence is only conserved to allow efficient translation? To begin to answer this question, the role of the N-terminus of the HCV C protein must be characterized.

The full-length HCV C protein (innate form) is composed of ~190 amino acids and incorporates an unusually large percentage of basic residues. The HCV C protein basic residues forming putative RNA binding domains are clustered in the N-terminal two-thirds of the innate form in three discrete regions (Fig. 1b and 1c) (Bukh *et al.*, 1994). Cluster I (CI) spans amino acid residues 1–23, cluster II (CII) spans about residues 39–75 and cluster III (CIII) spans about residues 101–121 (Fig. 1c). Proteolytic processing yields the N-terminal 173 amino acid frag-

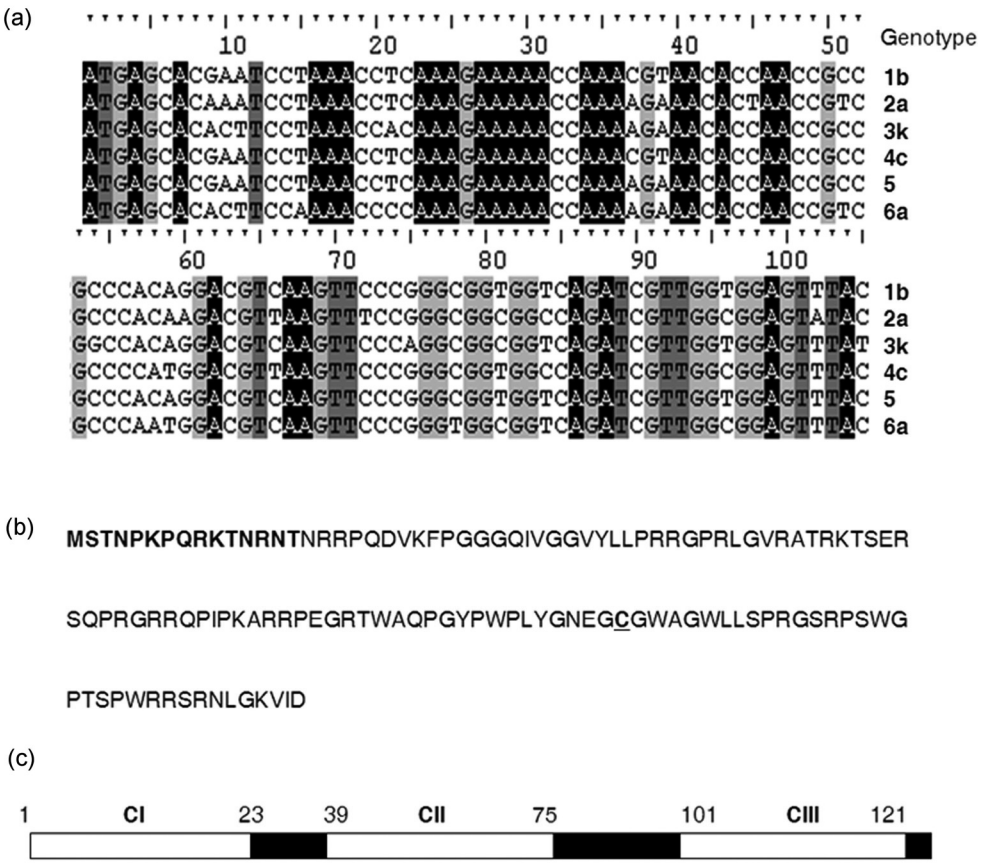

Fig. 1

The N-terminal HCV C protein coding region

(a) Nucleotide sequence alignment for different HCV genotypes. (b) The amino acid sequence of the C124 protein of HCV genotype 1a. The amino acids corresponding to (A)-rich sequences are in bold, the single cysteine is underlined. (c) The map showing the three basic amino acid clusters (CI, CII, and CIII).

ment comprising the "mature" HCV C protein (Liu *et al.*, 1997) that is implicated in nucleocapsid formation. C124 protein, a fragment of the mature C protein representing the N-terminal 124 amino acids, contains all three putative RNA binding domains, is sufficient to bind RNA, and forms virus-like particles *in vitro* upon the addition of highly structured RNA (Kunkel *et al.*, 2001). C124 protein has unusual amino acid composition (Fig. 1b). A total of twenty one arginines and six lysines accounts for ~22% of the amino acid composition along with unusually high number of prolines (~14%) and glycines (~14%). These observations strongly suggest that the HCV C protein may be natively unfolded at the N-terminal region (Uversky, 2002a,b). This observation was confirmed by a recent study (Duvignaud *et al.*, 2009). The details of C protein-RNA interactions leading to oligomerization of C protein and RNA encapsidation are not clearly understood. In this study, we examined the importance of the N-terminal 124 aa-long sequence of the C protein (C124) for its ability to bind to viral RNA by constructing its mutants with serial deletions of 10-45 N-terminal amino acids. The truncated as well as native C protein was expressed in *E. coli*, purified and tested for binding to viral RNA. The results showed that the CI cluster in the N-terminus of the C protein determines the folding/conformation of the rest of polypeptide and also affects the overall RNA binding ability of the C protein.

Materials and Methods

Plasmids constructs. The pET30a+ (Novagen, USA) plasmid was used for expression. The parent construct C124 (a gift from D. S. Peabody) contains the HCV genotype 1a coding sequence of the C protein from 1–124 amino acids. All constructs were confirmed by DNA sequencing. The primers used for constructing different N-terminal truncated forms of the C protein, were:

HCV124F: 5′p-GGGAAATCCATATGAGCACGAATCCTAAACCTCAAAG- 3′ OH

HCVC114F: 5′p-GGGAAATCCATATGACCAAACGTAACACCAACCGTCGC-3′OH

HCVC109F(*Nco*I): 5′p-GGGAAATGCCATGGGTAACCGTCGCCCACAGGACGTC-3′OH

HCVC104F(*Nco*I): 5′p-GGGAAATGCCATGGGTGACGTCAAGTTCCCGGGTGGC-3′OH

HCVC89F(*Nco*I): 5′p-GGGAAATGCCATGGGTTTGTTGCCGCGCAGGGGC-3′OH

HCVC79F(*Nco*I): 5′p-GGGAAATGCCATGGGTGGTGTGCGCGCGACGAGG-3′OH

HCVC124R: 5′p- AGACCGGAATTCTCATTAATCGATGACCTTACCCAAATTG-3′OH

All oligos were custom synthesized by IDT technologies (USA). *Pfu* TURBO polymerase (Stratagene, USA) and restriction enzymes (Promega, USA) were used for clone construction.

Expression of recombinant proteins in E. coli. BL21DE3 (pLysS) colonies harboring the clones were grown in LB medium under antibiotic selection at 37°C in a orbital shaker at 250 rpm until the A_{600} reached 1.0. Expression was induced by adding IPTG to a final concentration of 0.5 mmol/l. The culture was further grown for 4 hrs; the bacteria were pelleted at 6,000 x g for 10 min and the pellet was stored at -70°C. Cells were lysed in a single freeze-thaw step in presence of ice-cold urea buffer (25 mmol/l Na_2HPO_4 (pH 7.0), 8 mol/l urea, 250 mmol/l NaCl, 2 mmol/l EDTA, 1 mmol/l DTT). The lysate was clarified by filtering through 0.25 micron sterile filters (Millipore, USA) before purification. A protease inhibitor cocktail (Sigma-Aldrich, USA) was added to the clarified lysate and also to the pooled fractions after each chromatography procedure.

Ion exchange chromatography (IEC). All purification steps were carried out with an ACTA-FPLC (GE Healthcare, UK). Mono-S 5/50 GL high performance column (GE healthcare) was pre-equilibrated with 5 volumes of buffer A (0.25 mol/l HEPES (pH 7.0), 8 mol/l urea). The sample (1.5 ml) diluted in buffer A was applied to the column and unbound proteins were washed with 2 ml of buffer A. The column was eluted with a linear gradient of NaCl (buffer B: 0.25 mol/l HEPES (pH 7.0), 8 mol/l urea and 1.5 mol/l NaCl). The elution conditions were 0%–100% buffer B in 25 volumes at 0.75 ml/min flow rate. 1 ml fractions were collected and assayed for the presence of C protein by SDS-PAGE and western blot.

Size-exclusion chromatography (SEC). Refolded proteins after ion exchange chromatography were concentrated by ultrafiltration (Centricon filtration units, Millipore). SUPERDEX-75 10/300 GL column (GE Healthcare) was pre-equilibrated with 5 volumes of denaturing buffer (20 mmol/l Tris (pH 7.0) 100 mmol/l NaCl and 8 mmol/l urea) and 0.5 ml of the concentrated sample was applied to the column. The flow rate was maintained at 0.5 ml/min and 1 ml fractions were collected and assayed for presence of C proteins.

Affinity chromatography. HisTrap HP columns (GE Healthcare) were pre-equilibrated with 10 volumes of sample application buffer (10 mmol/l Tris (pH 8.0), 1 mol/l NaCl, 20% glycerol, 0.25% Triton-X-100, 8 mol/l urea and 20 mmol/l imidazole). Cell lysate prepared in sample application buffer was applied to the column and the column was washed with 10 volumes of Wash buffer (10 mmol/l Tris (pH5.9), 1 mol/l NaCl, 20% glycerol, 0.25% Triton-X-100, 8 mol/l urea and 20 mmol/l imidazole) to remove nonspecific proteins. Proteins were eluted by single step elution (100% Elution buffer) and assayed on SDS-PAGE gels. Sample application buffer supplemented with 120 mmol/l imidazole was used as Elution buffer.

After each chromatography step the fractions containing the HCV C proteins were pooled and protease inhibitor cocktail (Sigma-Aldrich) was added, and the fractions were dialyzed overnight against a buffer containing 20 mmol/l Tris (pH 7.4), 100 mmol/l NaCl and 1 mmol/l DTT at 4°C with constant stirring. All proteins were concentrated using centricon filtering units (Millipore).

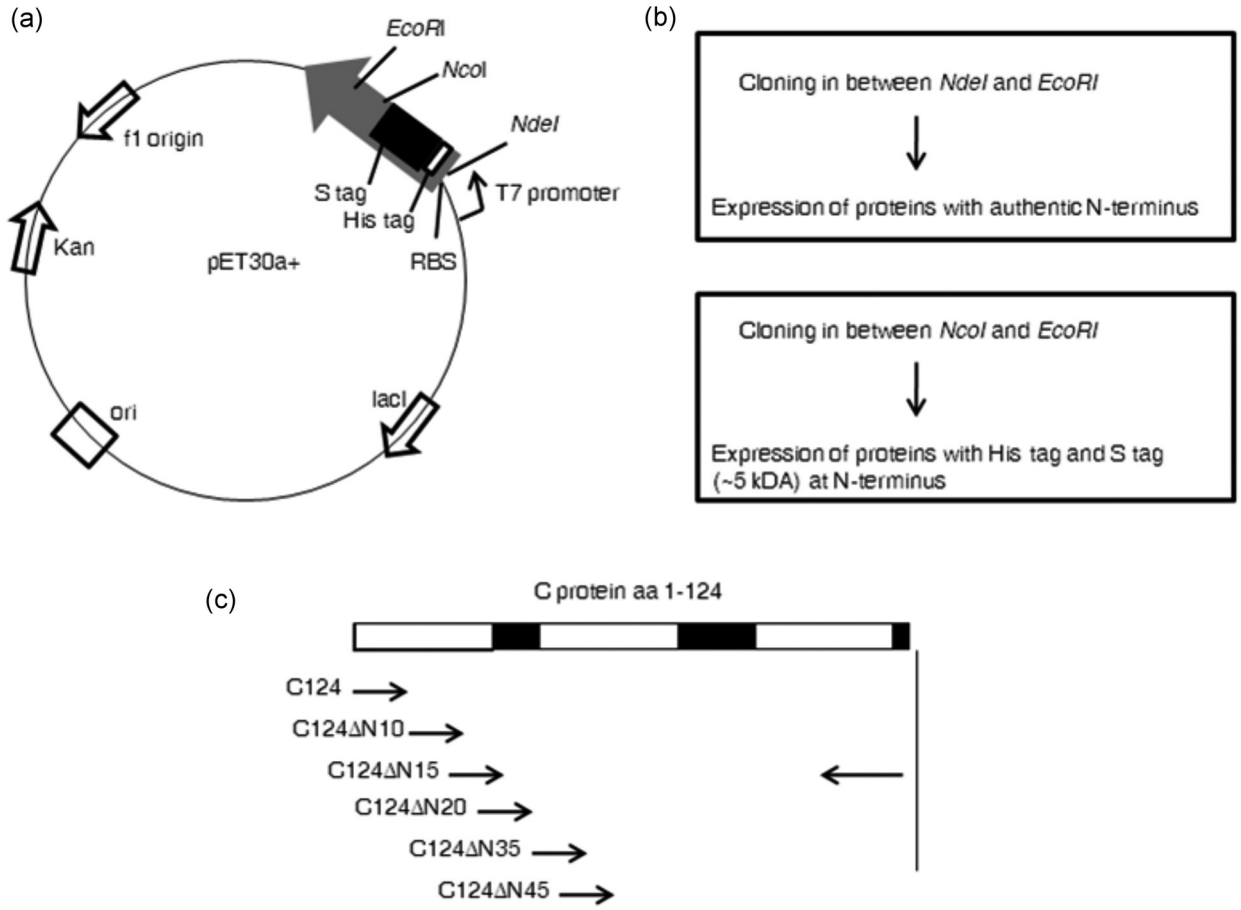

Fig. 2

Cloning strategy for expression of C124 protein and its N-terminal truncated forms

(a) pET30A+ vector. (b) The cloning strategy for non-tagged (upper box) and tagged proteins (lower box). (c) The map with C124 protein and its N-terminal truncated forms. The arrows indicate forward and reverse primers used in amplification.

Proteins were quantified by Bradford assay (Bio-Rad) following manufacturer's instructions.

Western blot analysis. After SDS-PAGE separation, the proteins were blotted onto a 0.45 micron nitrocellulose membrane (NCM) (Bio-Rad, USA) using Bio-Rad's Trans-Blot electrophoretic cell according to the manufacturer's instructions. We used MAb against aa 21–40 of the HCV C protein (Thermo Scientific Pierce, USA). The NCM was developed using Bio-Rad's alkaline-phosphatase immuno-blot kit following the instructions provided in the kit.

In vitro transcription and labeling. A plasmid encoding the HCV IRES-CAT gene (a gift from R. Rijnbrand) was transcribed *in vitro* using the MAXI-SCRIPT (Ambion, USA) transcription kit following the manufacturer's instructions. End-labeling of the transcript was accomplished using the KINASE-MAX kit (Ambion). The RNA was first dephosphorylated at the 5′-end using calf intestinal alkaline phosphatase and then radio-labeled with ^{2}P γ-ATP using polynucleotide kinase. The unincorporated nucleotides were removed using NUC-AWAY spin columns (Ambion), and the labeled RNA was quantified spectroscopically.

Gel electrophoresis mobility shift assay (GEMSA). One ng of ^{32}P radio-labeled HCV IRES RNA was mixed with: 5 μmol/l of purified C124 protein, C124NΔ10 truncated protein; 20 μmol/l of purified C124NΔ15, C124NΔ20, C124NΔ35, and C124NΔ45 truncated proteins in binding buffer (20 mmol/l HEPES pH 7.0, 100 mmol/l NaCl, 1 mmol/l $MgCl_2$, 0.2 mmol/l DTT, 5% glycerol, and 20 U of RNase inhibitor (Promega) in a final reaction volume of 20 μl, and incubated at 37°C for 20 min. 2 μl of sample loading buffer (Bio-Rad) was added to each reaction. The reactions were loaded onto a 1% agarose gel, separated by electrophoresis in TAE buffer (Bio-Rad) at 120 V (constant voltage) for 1.5 hr and visualized by autoradiography.

Results and Discussion

Construction, expression and purification of C124 protein and its N-terminal truncated forms

The 50 C-terminal amino acids of the HCV C protein are hydrophobic and interfere with expression in bacteria (Kunkel *et al*., 2001). C124 protein was used as the parental reference protein for the N-terminal truncated proteins used in this study, because it has been expressed in bacteria, it is sufficient to bind RNA with wild type affinity, and forms nucleocapsid-like particles (NLPs) *in vitro* (Kunkel *et al*., 2001). pET30a+ plasmid was used to express the C124 protein and its derivatives (Fig. 2a). To study the effect of the N-terminus of C124 protein on its function, we originally intended to express the proteins as discrete fusion-less proteins with an authentic N-terminus. For this purpose, constructs were cloned at the *Nde*I site of the plasmid (Fig. 2b). C124 cDNA was used as a template to construct HCV C protein expression constructs that produce C proteins with 5 (C124NΔ5), 10 (C124NΔ10), 15 (C124NΔ15), 20 (C124NΔ20), 35 (C124NΔ35), and 45 (C124NΔ45) amino acid deletions from the N-terminus beginning from the first amino acid (methionine) (Fig. 2c). Since this strategy did not include an affinity tag that would aid in purification, the proteins were purified in the following way. Firstly the lysates were subjected to IEC and then cleaned up by SEC. 8 M urea was maintained as a denaturant throughout the purification steps for all the proteins in this study to prevent the association of the proteins with nucleic acids. Previous studies showed that refolding urea-denatured C124 protein by dialysis yields functionally active protein, indicating that the protein is assembly-competent (Kunkel *et al*., 2001). After each purification step, the urea was removed by dialysis at 4°C. This strategy yielded robust levels of expression of the parental C124 protein and of C124NΔ5 and C124NΔ10 truncated proteins, which were subsequently purified to near homogeneity (Fig. 3a, data not shown for C124NΔ5 truncated protein). However, no expression was detected by SDS-PAGE for the C124NΔ15, C124NΔ20, C124NΔ35, and C124NΔ45 truncated proteins, and no appreciable amounts of the proteins were isolated from these constructs using the above purification strategies (data not shown). However, very low levels of expression were confirmed in the lysates by Western blot analysis using an anti-HCV C protein MAb that recognizes aa 21–40 of the C protein (data not shown). Because of these difficulties, an alternate strategy was used to express and purify the remaining truncated proteins by incorporating a His-tag and S-tag at the N-termini of the C124 protein deletion mutants (Fig. 2b). Efficient expression of the poorly expressing C124 protein deletion mutants was achieved by placing the 5 kDa fusion protein coding region at the 5′-ends of the C124 derived coding sequences (Fig. 3b).

C124 protein has a high net charge and low mean hydrophobicity (Kunkel *et al*., 2001), which are characteristics of proteins that are natively unfolded or intrinsically disordered (Uversky, 2002a,b). Recent biophysical studies placed the HCV C protein into the family of intrinsically unstructured proteins (Duvignaud *et al*., 2009). Free unstructured C124 protein adopts a stable conformation only upon interacting with RNA (Kim *et al*., 2006). Recombinant C124 protein′s weak dimerizing tendency in SDS-PAGE has been previously reported (Kunkel *et al*., 2001; Acosta-Rivero *et al*., 2005; Kota *et al*., 2009) and the dimers are biophysically characterized (Boulant *et al*., 2005). We also observed weak dimerization for the C124 protein and the C124ΔN10 truncated protein in several of our preparations, which typically have low concentrations of DTT (1 mmol/l) (Fig. 3a, note that a small fraction of the C124ΔN10 truncated protein migrates as a putative dimer). In this context, it is interesting to note that the mobility of C124ΔN15, C124ΔN20, C124ΔN35, and C124ΔN45 truncated proteins on SDS-PAGE gels is dependent on the presence of 8 mol/l urea in the storage buffer (Fig. 3c and data not shown for C124ΔN15). After removal of urea from the buffer by dialysis, these proteins migrated on SDS-PAGE as higher molecular mass proteins. The calculated molecular masses based on the amino acid composition of the respective fusion protein monomers (including the N-terminal tags) are: 16.4 kDa (C124ΔN20 truncated protein), 14.9 kDa (C124ΔN35 truncated protein), and 13.8 kDa (C124ΔN45 truncated protein). The newly formed protein species deviated drastically from the expected mobility of their monomeric forms on SDS-PAGE. We call these new species as anomalously migrating (AM) proteins. These AM protein bands were observed only after removing the urea from the buffer and represented close to 100% of the total protein isolated from C124ΔN20, C124ΔN35, and C124ΔN45 truncated protein preparations after dialysis (Fig. 3c). This contrasts with the behavior of the C124 protein and the C124ΔN10 truncated protein, which migrated in a manner consistent with their calculated molecular mass regardless of the presence of urea (Fig. 3a).

Western blot analysis of C124 protein and its N-terminal truncated forms

The HCV C protein fragments present in the lysates of IPTG induced *E. coli* cells were confirmed by Western blot analysis (Fig. 4), except for the H124ΔN35 and H124ΔN45 truncated protein. The MAb used in the procedure recognizes aa 21–40 of the HCV C protein, which is absent in these mutants. Surprisingly the AM bands of the renatured C124ΔN15 and C124ΔN20 truncated C proteins were undetectable by Western blot analysis using this MAb (not shown), whereas a small fraction of the faster migrating proteins could still be detected (at similar positions to that in Fig. 4, lanes 3 and 4 where urea-denatured whole cell lysates were used for blotting).

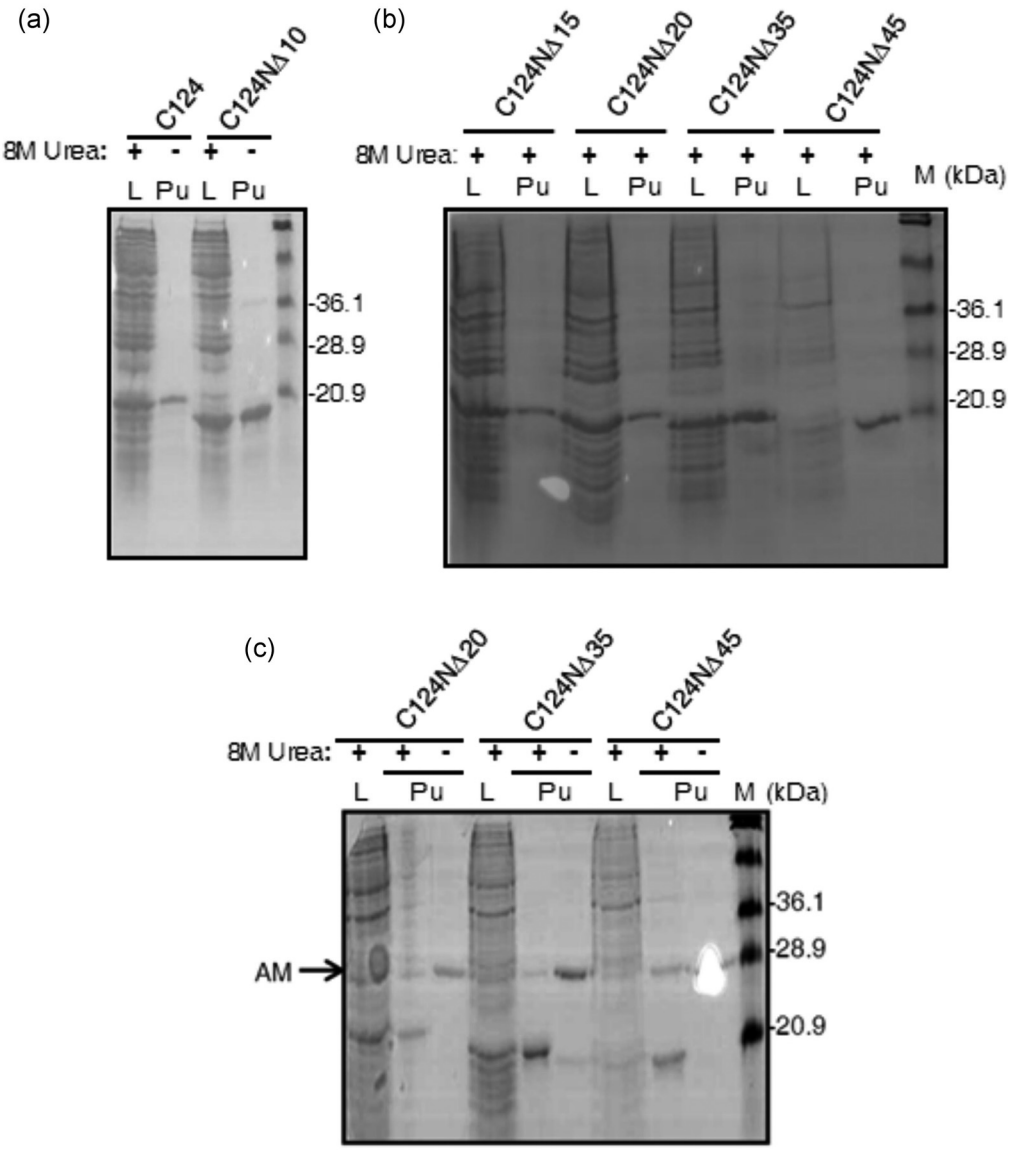

Fig. 3

Expression and purification of C124 protein and its N-terminal truncated forms

SDS-PAGE in the presence (+) or absence (-) of 8 mol/l urea. (a) Purificates (Pu) obtained by IEC and SEC. (b) Purificates (Pu) obtained by affinity chromatography. (c) Purificates (Pu) obtained by affinity chromatography. L = lysates; Pu = purificates; AM = anamolous migration; M = protein size markers.

RNA binding ability of C124 protein and its N-terminal truncated forms

The purified proteins were refolded at 4°C in a buffer containing 1 mmol/l DTT as described in the methods, and GEMSA was used to characterize their ability to bind radiolabeled RNA. The C124 protein and C124ΔN10 truncated protein were completely retained the probe completely in the wells (Fig. 5), suggesting a high-affinity RNA-protein interaction that induced multimerization or aggregation of the protein-RNA complexes. The radiation in the wells of the lanes corresponding to the C124 protein and C124ΔN10 truncated protein complexes were of similar intensity (Fig. 5), with very little radiation entering the gel. Deleting five more amino acids to produce the C124ΔN15 truncated protein decreased the RNA binding efficiency from 100% to ~10%. These five deleted amino acids included only one arginine residue out of the twenty-three combined arginine and lysine residues that were present in the C124ΔN10 truncated protein. This dramatic decrease in RNA binding efficiency is unlikely to be due to a simple reduction in

protein basicity, but more likely results from a conformational change in the protein. Deleting the N terminal 20 amino acids to produce the C124ΔN20 truncated protein did not further reduce the RNA binding efficiency. However, the C124ΔN35 and C124ΔN45 truncated proteins showed no detectable retention of radio-labeled RNA in the wells, suggesting that the linker region between CI and CII is important for the RNA binding ability of the protein (Fig. 1c). We further concluded that the major portion of the peptide encoded by the (A)-rich tract of the HCV ORF (HCV nt 342–389, encoding the N terminal 15 amino acids residues of the C protein) is relatively unimportant for RNA binding, and is presumably also unimportant for nucleocapsid formation, in line with a previous study in cell-free lysates (Klein *et al.*, 2004).

RNA secondary structure at the 5′-end of the C protein coding sequence

Intrigued by the loss of expression of truncated forms of C124 protein beyond 10 amino acids from the N-terminus (without a fusion tag), we analyzed the possible reasons for this result. The HCV C protein coding sequence contains extensive RNA secondary structure (Tuplin *et al.*, 2004). Secondary structure in mRNA near the start codon can interfere with translation initiation in bacteria (De Smit and van Duin, 1994). HCV IRES-dependent translation initiation and *E. coli* translation initiation are mechanistically similar (Terenin *et al.*, 2008), since both use a non-scanning mode to assemble the initiation complex at the start site. Therefore both mechanisms are sensitive to secondary structure in the RNA immediately downstream from the start codon (Kozak, 2005; Pandey *et al.*, 2009; Punginelli *et al.*, 2004; Rijnbrand *et al.*, 2001). Intrinsic distal downstream secondary structures in the RNA encoding the C124 protein would be pushed upstream when the 5′ coding sequences are deleted (Fig. 6). Therefore, RNA secondary structures that are inconsequential in C124 mRNA might interfere with translation initiation or elongation when deletions bring them nearer to the start codon. Alternatively, the amino acids coded by the (A)-rich tract at the 5′-end of the HCV ORF may provide stability to the C protein and their deletion might have resulted in unstable proteins that were not sufficiently abundant to be detected by coomassie staining.

Conclusions

Our results suggest that a major conformational change occurred in the renatured C proteins when 15 or more amino acids were deleted from the N-terminus. This conformational change dramatically changed the protein′s electrophoretic mobility in SDS-PAGE and prevented the proteins from interacting with an antibody that recognizes these proteins in their denatured states (Fig. 4, lanes 3 and 4). It is known that recombinant

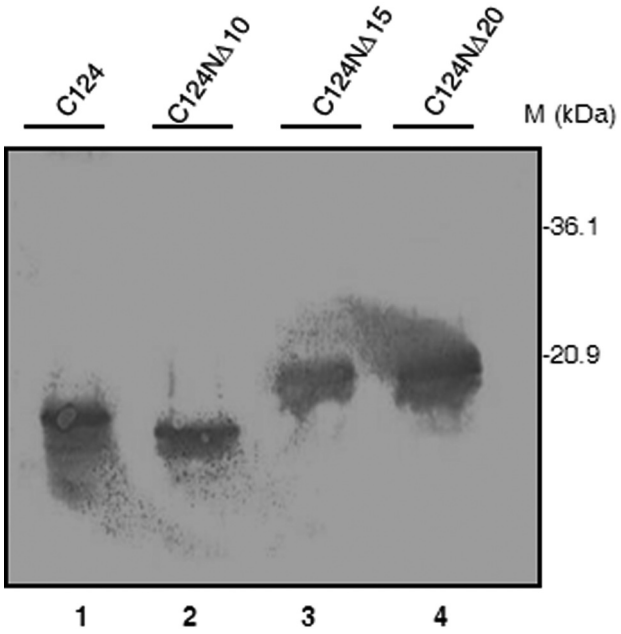

Fig. 4

Western blot analysis of C124 protein and its N-terminal truncated forms

M = protein size markers.

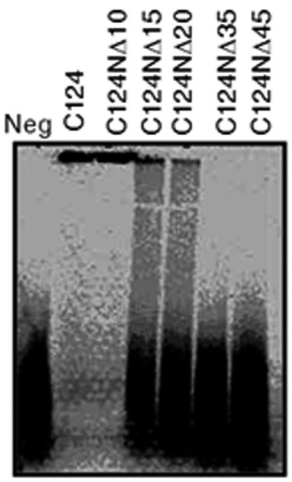

Fig. 5

RNA binding ability of C124 protein and its N-terminal truncated forms

Autoradiograph of GEMSA products separated by native agarose gel electrophoresis. Neg = negative control.

C106 protein (N-terminal 106 amino acids of HCV C protein) trimers or higher order oligomers do not react with a polyclonal antibody that recognizes its monomeric or dimeric forms

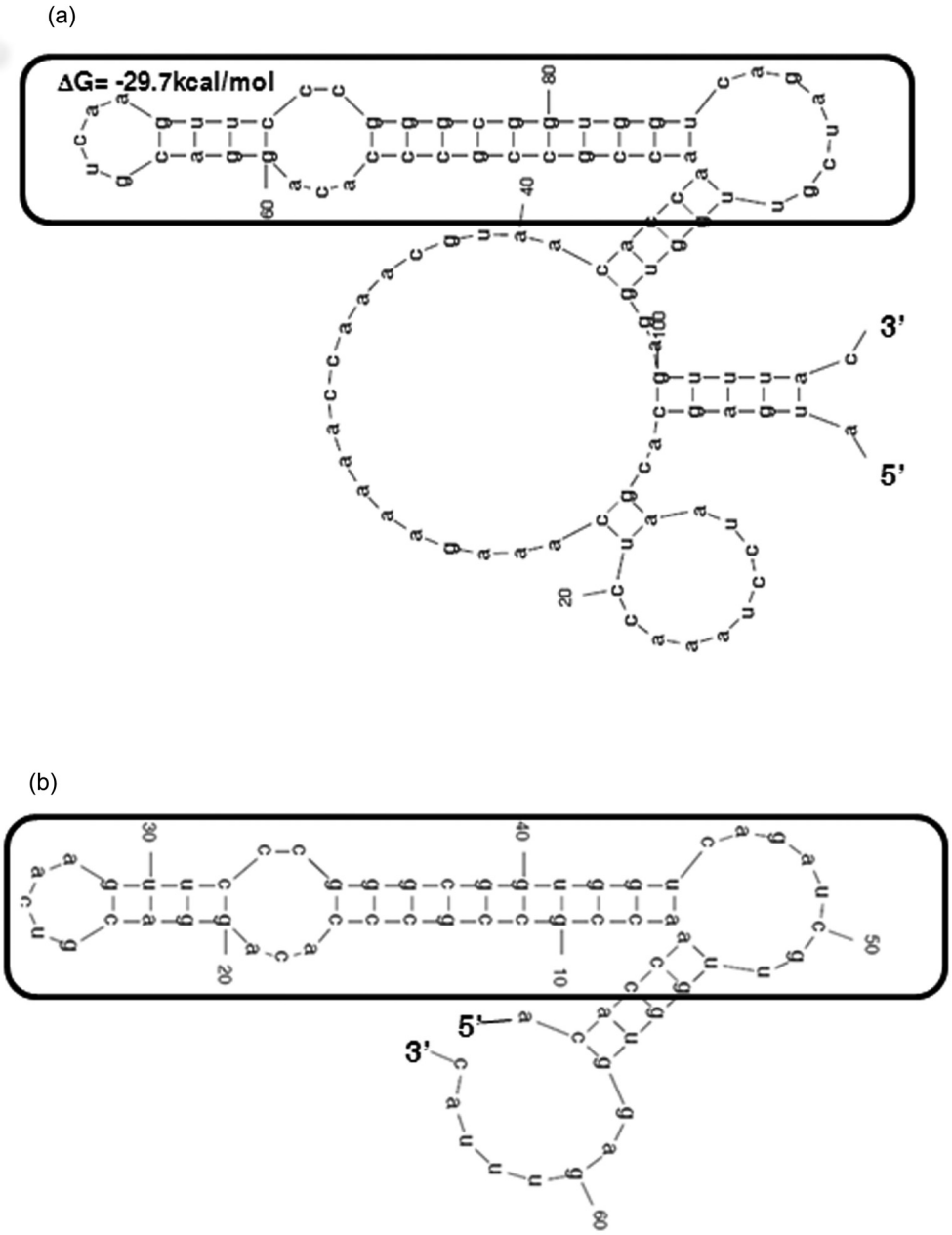

Fig. 6

Secondary structure of nucleotide sequences encoding C124 protein and its N-terminal truncated form

(a) C124. (b) C124NΔ40. The stably folded part is boxed. The structures were obtained using MFOLD.

(Kota *et al*., 2009). His-tag-S-tag fusions are widely used for expression of recombinant proteins in *E. coli* using the pET30 vectors and, to our knowledge, have not been reported to interfere with the folding and/or conformation of their partner proteins. Alternately, the AM proteins could be authentic C protein dimers, in which case the N-terminal 10–15 amino acids of C124 protein might be responsible for preventing C protein dimerization in absence of RNA. Deleting these amino acids may have induced the C protein to dimerize in RNA-free solution to produce the AM protein bands observed in our PAGE gels. Furthermore, these dimers could have resulted from spontaneous transglutamination (Lu *et al*., 2001), or could be disulphide bond-mediated (through a lone cysteine residue at position 91, Fig. 1b). A disulphide bond mediated C dimer was

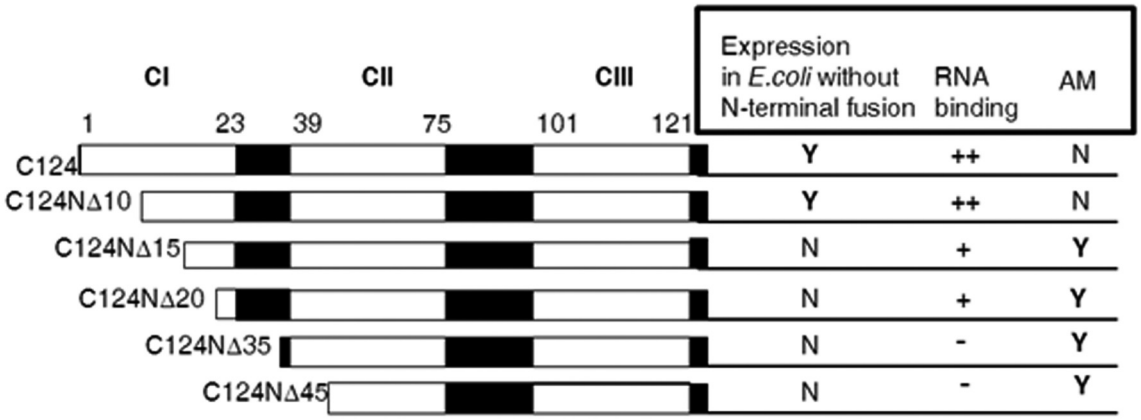

Fig. 7

Expression, RNA binding and AM forms of C124 protein and its N-terminal truncated forms

Yes (Y), no (N), strong (++), weak (+), no binding (–).

recently reported to be important for nucleocapsid formation (Kushima *et al.*, 2010). More experiments are needed to fully understand this phenomenon better. The basic amino acids in the CI cluster do not affect virus replication or infectivity when mutated to alanines one at a time or in groups of two or three, suggesting a structural role for this cluster in nucleocapsid formation (Alsaleh *et al.*, 2010). The composition of the (A)-rich tract in the HCV ORF determines the production of the alternate reading frame (ARF) of C protein which is being implicated in HCV pathogenesis (Boumlic *et al.*, 2011). In summary, deletion of the 15 N-terminal amino acids from the C124 protein (which are encoded by 45 bases spanning the (A)-rich tract) bring about three important changes in the physical and/or biochemical characteristics of the protein (Fig. 7): 1) The expression of the protein in *E. coli* is almost abolished, 2) The protein′s RNA binding ability decreases from 100% to ~10%, and 3) the folding and/or conformation of the protein appears to be drastically altered affecting its mobility on SDS-PAGE gels. Deleting 35 or more amino acids from the N-terminus completely abolishes the RNA binding ability of the C124 protein, suggesting the importance of the N-terminus in RNA binding. Klein *et al.* (2004), also showed that deleting 20 amino acids at the N-terminus of the full length "mature" C protein (C173 protein) decreased the efficiency of nucleocapsid formation to ~30 % of the wild type levels, and deleting 42 or more amino acids completely abolished nucleocapsid formation in cell free lysates. This further suggests that, despite using a C-terminal truncated C protein, our findings may be relevant to the full length C protein, and that they may be biologically significant. In conclusion, the (A)-rich tract of the C protein coding region is important at the RNA level to preclude secondary structure during translation initiation (Rijnbrand *et al.*, 2001), and at the protein level it encodes an amino acid sequence that may be critical for the conformation of C protein, and therefore critical to its ability to bind to RNA.

Acknowledgements. This work was carried out in the Dept. of Chemistry and Biochemistry, New Mexico State University, Las Cruces, NM 88003-8001 during the authors′ previous appointment and was supported by grant number RR-16480 from the NIH through the New Mexico IDeA Networks of Biomedical Research Excellence (NM-INBRE). The authors thank D. S. Peabody, University of New Mexico at Albuquerque for the parent C124 plasmid construct and for critical review of this manuscript and Rene Rijnbrand, University of Texas Medical Branch, Galveston, TX for HCV IRES-CAT plasmid.

References

Acosta-Rivero N, Rodriguez A, Mussachio A, Poutu J, Falcon V, Torres D, Aguilar JC, Linares M, Alonso M, Perez A, Menendez I, Morales-Grillo J, Marquez G, Duenas-Carrera S (2005): A C-terminal truncated hepatitis C virus core protein variant assembles in vitro into virus-like particles in the absence of structured nucleic acids. Biochem. Biophys. Res. Commun. 334, 901–906. http://dx.doi.org/10.1016/j.bbrc.2005.06.185

Alsaleh K, Delavalle P-Y, Pillez A, Duverlie G, Descamps V, Roullie Y, Dubuisson J, Wychowski C (2010): Identification of basic amino acids at the N-terminal end of the core protein that are crucial for hepatitis C virus infectivity. J. Virol. 84, 12515–12528. http://dx.doi.org/10.1128/JVI.01393-10

Boulant S, Vanbelle C, Ebel C, Penin F, Lavergne JP (2005): Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. J. Virol. 79, 11353–11365. http://dx.doi.org/10.1128/JVI.79.17.11353-11365.2005

Boumlic A, Vassilaki N, Dalagiorgou G, Kochlios E, Kakkanas A, Georgoupoulou U, Markoulatos P, Orfanoudakis G, Mavromara P (2011): Internal translation initiation stimulates expression of the ARF/core+1 open reading frame of HCV genotype 1b. Virus Res. 155, 213–220. http://dx.doi.org/10.1016/j.virusres.2010.10.007

Bukh J, Purcell RH, Miller RH (1994): Sequence analysis of the core gene of 14 hepatitis C virus genotypes. Proc. Natl. Acad. Sci. USA 91, 8239–8243. http://dx.doi.org/10.1073/pnas.91.17.8239

Cuthbert JA (1994): Hepatitis C: progress and problems. Clin. Microbiol. Rev. 7, 505–532.

De Smit MH, van Duin J (1994): Control of translation by mRNA secondary structure in E. coli. A quantitative analysis of literature data. J. Mol. Biol. 244, 144–150. http://dx.doi.org/10.1006/jmbi.1994.1714

Duvignaud J-B, Savard C, Fromentin R, Majeau N, Leclerc D, Gagne SM (2009): Structure and dynamics of the N-terminal half of hepatitis c core protein: An intrinsically unstructured protein, Biochem. Biophys. Res. Comm. 378, 27–31. http://dx.doi.org/10.1016/j.bbrc.2008.10.141

Fletcher SP, Jackson RJ (2002): Pestivirus internal ribosome entry site (IRES) structure and function: elements in the 5' untranslated region important for IRES function. J. Virol. 76, 5024–5033. http://dx.doi.org/10.1128/JVI.76.10.5024-5033.2002

Gil LH, Ansari IH, Vassilev V, Liang D, Lai VC, Zhong W, Hong Z, Dubovi EJ, Donis RO (2006): The amino-terminal domain of bovine viral diarrhea virus Npro protein is necessary for alpha/beta interferon antagonism. J. Virol. 80, 900–911. http://dx.doi.org/10.1128/JVI.80.2.900-911.2006

Kim M, Ha Y, Park HJ (2006): Structural requirements for assembly and homotypic interactions of the hepatitis C virus core protein. Virus Res. 122,137–143. http://dx.doi.org/10.1016/j.virusres.2006.07.008

Klein KC, Polyak SJ, Lingappa JR (2004): Unique Features of Hepatitis C Virus Capsid Formation Revealed by De Novo Cell-Free Assembly. J. Virol. 78, 9257–9269. http://dx.doi.org/10.1128/JVI.78.17.9257-9269.2004

Kota S, Coito C, Mousseau G, Lavergne JP, Strosberg AD (2009): Peptide inhibitors of hepatitis C virus core oligomerization and virus production. J. Gen. Virol. 90, 1319–1328. http://dx.doi.org/10.1099/vir.0.008565-0

Kunkel M, Lorinczi M, Rijnbrand R, Lemon SM, Watowich SJ (2001): Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein. J. Virol. 75, 2119–2129. http://dx.doi.org/10.1128/JVI.75.5.2119-2129.2001

Kozak M (2005): Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361, 13–37. http://dx.doi.org/10.1016/j.gene.2005.06.037

Kushima Y, Wakita T, Hijikata M (2010): A disulphide-bonded dimer of the core protein of the hepatitis C virus is important for virus-like particle production. J. Virol. 84, 9118–9127. http://dx.doi.org/10.1128/JVI.00402-10

Liu Q, Tackney C, Bhat RA, Prince AM, Zhang P (1997): Regulated processing of hepatitis C virus core protein is linked to subcellular localization, J. Virol. 71, 657–662.

Lu HH, 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. http://dx.doi.org/10.1073/pnas.93.4.1412

Lu W, Strohecker A, Ou J-H (2001): Post-translational modification of the hepatitis C virus core protein by tissue transglutaminase. J. Biol. Chem. 276, 47993–47999.

Pandey JP, Gorla P, Manavathi B, Siddavattam D (2009): mRNA secondary structure modulates the translation of organophosphate hydrolase (OPH) in E. coli. Mol. Biol. Rep. 36, 449–454. http://dx.doi.org/10.1007/s11033-007-9200-5

Punginelli C, Ize B, Stanely NR, Stewart V, Sawers G, Berks BC, Palmer T (2004): mRNA secondary structure modulates translation of Tat-dependent formate dehyrogenase N. J. Bacteriol. 186, 6311–6315. http://dx.doi.org/10.1128/JB.186.18.6311-6315.2004

Reynolds JE, Kaminski A, Carroll AR, Clarke BE, Rowlands DJ, Jackson RJ (1996): Internal initiation of translation of hepatitis C virus RNA: the ribosome entry site is at the authentic initiation codon. RNA 2, 867–878.

Rijnbrand RC, Lemon SM (2000): Internal ribosome entry site-mediated translation in hepatitis C virus replication. Curr. Top. Microbiol. Immunol. 242, 85–116. http://dx.doi.org/10.1007/978-3-642-59605-6_5

Rijnbrand R, Bredenbeek PJ, Haasnoot PC, Kieft JS, Spaan WJ, Lemon SM (2001): The influence of downstream protein-coding sequence on internal ribosome entry on hepatitis C virus and other flavivirus RNAs. RNA 7, 585–597. http://dx.doi.org/10.1017/S1355838201000589

Robertson B, Myers G, Howard C, Brettin T, Bukh J, Gaschen B, Gojobori T, Maertens M, Nainan O, Netesov S, Nishioka K, Shini T, Simmonds P, Smith D, Stuyver L, Weiner A (1998): Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Virus Taxonomy. Arch. Virol. 143, 2493–2503. http://dx.doi.org/10.1007/s007050050479

Rumenapf T, Stark R, Heimann M, Thiel HJ (1998): N-terminal protease of pestiviruses: identification of putative catalytic residues by site-directed mutagenesis. J. Virol. 72, 2544–2547.

Terenin I M, Dmitriev SE, Andreev DE, Shatsky IN (2008): Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. Nat. Struc. Mol. Biol. 15, 836–841. http://dx.doi.org/10.1038/nsmb.1445

Tuplin A, Evans DJ, Simmonds P (2004): Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. J. Gen. Virol. 85, 3037–3047. http://dx.doi.org/10.1099/vir.0.80141-0

Uversky VN (2002a): What does it mean to be natively unfolded? Eur. J. Biochem. 269, 2–12. http://dx.doi.org/10.1046/j.0014-2956.2001.02649.x

Uversky VN (2002b): Cracking the folding code. Why do some proteins adopt partially folded conformations, whereas other don't? FEBS Lett. 514, 181–183. http://dx.doi.org/10.1016/S0014-5793(02)02359-1