Unusual multiple recoding events leading to alternative forms of hepatitis C virus core protein from genotype 1b.

Steeve Boulant¹, Michel Becchi², François Penin¹, and Jean-Pierre Lavergne¹* 

¹Laboratoire de Bioinformatique et RMN Structurales and ²Centre commun de séquençage
Institut de Biologie et Chimie des protéines
UMR5086 CNRS – Université Claude Bernard Lyon I
7, Passage du Vercors
69367 Lyon cedex07
France

Running title: Alternative forms of HCV core protein

* To whom correspondence should be addressed
Phone.: 33-(0)-472-722-645
Fax: 33-(0)-472-722-605
Email: jp.lavergne@ibcp.fr
SUMMARY

Besides its involvement in the formation of the capsid shell of the virus particles, the core protein of hepatitis C virus (HCV) is believed to play an important role in the pathogenesis and/or establishment of persistent infection. We describe here alternative forms of genotype 1b HCV core protein identified after purification of various products of core protein segment 1-169 expressed in *Escherichia coli* and their analysis by proteolysis, mass spectrometry and amino acid sequencing. These proteins all result from a +1 frameshifting at codon 42 (a different position than that previously reported in genotype 1a) and, for some of them, from a re-phasing in the normal open reading frame at the termination codon 144 in the +1 open reading frame. To test the relevance of these recoding events in a eukaryotic translational context, the nucleotide sequences surrounding the two shift sites were cloned in the three reading frames into expression vectors allowing the production of a C-terminally fused GFP and expressed both in a reticulocyte lysate transcription/translation assay and in culture cells. Both recoding events were confirmed in these expression systems, strengthening the hypothesis that they might occur in HCV infected cells. Moreover, sera from HCV positive patients of genotype 1a or 1b were shown to react differently against synthetic peptides encoded in the +1 open reading frame. All together, these results indicate the occurrence of distinct recoding events in genotypes 1a and 1b, pointing out genotype-dependent specific features for F protein.
Keywords: HCV core protein/hepatitis C virus/recoding/ribosomal frameshifting/translation.
INTRODUCTION

HCV is estimated to chronically infect roughly 170 million people worldwide (1) and is a major public health problem as chronic infection may lead to severe liver diseases including cirrhosis and hepatocellular carcinoma. HCV has a positive-sense, single-stranded RNA genome of ∼9.6 kb and is a member of the Flaviviridae (2) (3). The genome encodes a polyprotein of some 3,000 amino acids, which is post-translationally cleaved by viral and cellular proteases to generate at least 10 viral proteins identified as structural proteins (C, E1, E2 and P7) and non-structural proteins (NS2, NS3, NS4a, NS4b, NS5a and NS5b) (for a review see (4) (5). The translation of the HCV polyprotein is regulated by the highly structured 5'-non-coding region acting as an internal ribosome entry site (IRES) (for a review see (6)).

The HCV core protein is 191 amino acids in length and consists of three distinct predicted domains: an NH2-terminal two-third domain of highly positively charged amino acids, a C-terminal one-third domain of hydrophobic residues and the last 20 or so residues serving as the signal peptide for the downstream protein E1 (7) (8) (9) (10). The initial polyprotein cleavage generates the immature core protein (P23) that undergoes additional processing by the intramembrane-cleaving protease SPP (Signal Peptide Peptidase) (11). This yields mature core P21 whose C-terminus is not precisely known but lies between residues 172 and 182 (11) (12) (13). It has been shown that alternative form(s) of the HCV core protein could be produced as a result of a -2/+1 ribosomal frameshifting at or near codon 11 in genotype 1a (14) (15) (16) (17). Walewski et al. stated that a cluster of unusually conserved synonymous codons in this core-coding region indicated a potential overlapping open reading frame. Specific IgGs for three of four peptides derived from this alternate reading frame protein were detected in chronic HCV sera (14). Xu et al. reported both the in vitro and the in vivo synthesis of a 17 kDa core protein resulting from a -2/+1 ribosomal frameshifting which
was named “F protein”. Here again, antibodies specific for this protein were detected in sera from HCV infected patients (15). F protein might be related to a 16 to 17 kDa protein (P16) previously observed in mammalian cells expression studies besides P21 and P23 and initially thought to be a truncated form of core protein (18). More recently, Choi et al. (19) reported the possibility of multiple frameshifting events at or around codon 11 in core sequence of genotype 1a. Besides the F protein due to -2/+1 frameshifting, a 1.5-kDa protein could also be produced by -1/+2 frameshifting.

We report here the production in *Escherichia coli* of alternative forms of the HCV core protein from genotype 1b resulting from a +1 ribosomal frameshifting at codon 42 that can be followed by a re-phasing in the 0 frame that bypass the stop codon at position 144. The exact positions of both these recoding events were determined by amino acid sequencing and mass spectrometry. The ability of the corresponding nucleotide sequences surrounding the shift sites to induce recoding in a eukaryotic translational context was demonstrated both by an in vitro transcription/translation assay in a reticulocyte lysate and by expression in culture cells. Finally, immunological analysis using various synthetic peptides revealed the presence of antibodies directed against the +1 core reading frame in several sera of HCV positive patients of genotype 1a or 1b. However, the differences of reactivity against these peptides support our finding that the frameshifting site leading to F protein is different in genotypes 1a and 1b.

An oral presentation of part of this work was done at the 9th International Meeting on Hepatitis C and related viruses (San Diego, U.S.A. - July 2002).
EPERIMENTAL PROCEDURES

Plasmid construction and protein purification- A 507-bp fragment corresponding to amino acids 1 to 169 of the HCV core protein (C_{HCV}1-169) was amplified by PCR from sequence with EMBL accession number D89872 encoded by the plasmid PCMV-C980 (a gift from Dr Shimotohno) and two specific primers containing a Ndel site or a PstI site, respectively. The Ndel- PstI fragment was cloned into the expression vector pT7-7 (6xHis) (20). C_{HCV}1-169 carrying a polyhistidine fused to the C terminus [C_{HCV}1-169(6xHis)] was expressed from the resulting plasmid after transformation in the *E. coli* strain BL21 SI (Life Technologies) producing T7 RNA polymerase.

*E. coli* BL21 SI was transformed with the plasmid and transformants were grown at 37°C in LB medium without NaCl until the culture reached an OD_{600} of 0.7. Expression was induced by adding NaCl to a final concentration of 200mM. Incubation was continued for a further 3h. Cells were harvested by centrifugation at 5,500 x g for 10 min at 4°C and then re-suspended in 25mM Tris-HCl pH 7.5, 5mM MgCl₂, DTT 1mM, PMSF 1mM and 100 units/mL benzonase. The cells were lysed using an SLM-Aminco French press at 1,200 p.s.i. followed by centrifugation at 30,000 x g for 30 min. The pellet was resuspended in 20mM Tris-HCl pH 8.0, 500mM NaCl, 6M Urea, 10mM β-Mercaptoethanol and 0.1 % dodecylmaltoside (Buffer A), then homogenized by sonication and centrifuged at 24,000 x g for 20 min. The supernatant was collected and the pellet submitted to a second urea extraction as described above. Both supernatants were pooled and loaded over a Ni-NTA agarose column (Qiagen) previously equilibrated with buffer A. The column was washed with 3 volumes of buffer A, then with 3 volumes of buffer A containing 10mM imidazole and the proteins were eluted with buffer A containing 250mM imidazole. The fractions containing [C_{HCV}1-169(6xHis)] were pooled and subjected to reversed-phase HPLC on a VYDAC C8
column (300Å, 10 µm, 10 X 250 mm) equipped with a C8 Aquapore guard column (Brownlee, 4.6 x 30mm) using a linear gradient of acetonitrile in 10% trifluoroacetic acid (TFA) at 1.5 mL/min flow rate. Linear gradient steps were performed using Waters 510 HPLC pumps as follows: 0 min, 25% acetonitrile; 0-5 min, 35% acetonitrile; 5-20 min, 50% acetonitrile; 20-50 min, 60% acetonitrile; 50-60 min, 100% acetonitrile. Chromatography was monitored at 220 nm and 280 nm using a Waters 991 photodiode array detector. Proteins corresponding to the main peaks were lyophilized and identified by mass spectrometry and N-terminal sequencing.

Mass spectrometry and peptide sequencing- All Liquid Chromatography/Mass Spectrometry (LC/MS) analysis were carried out using a Sciex API 165 quadrupole mass spectrometer coupled to an Applied Biosystem ABI 140D capillary LC system. The mass spectrometer was operated using two Electrospray ionization (ESI) sources (microspray and ionspray) in the positive ion mode. The microspray source was used for the direct infusion of protein solutions with 0.2 µL/min flow rate in a CH₃OH/H₂O (50/50, v/v) mixture containing 0.1% of HCOOH. LC/MS was carried out on a C18 HPLC micro-column (Brownlee, 150 x 0.5 mm i.d., 5 µm particle size, 300Å pore) at a flow rate of 10 µL/min connected to a 785A Absorbance Detector (Applied Biosystems) and an ionspray source. The V8 digested peptides were separated using mobile phases A and B with a four-step linear gradient of 10% B in the first 5 min, followed by 10 to 70% B in the next 60 min then 70 to 95% B for 10 min and hold at 95% B in the last 10 min (mobile phase A: 0.05% TFA in H₂O; mobile phase B: 0.04% TFA in CH₃CN/H₂O, (90/10, v/v)). Absorbance detection was set at 214 nm. The scan range was set at m/z 700-2200.

The peptides were sequenced by automatic Edman degradation using a Procise 492A liquid-phase sequencer (Applied Biosystems).
In Vitro transcription and translation- A DNA fragment from nucleotide 100 to nucleotide 200 of the HCV core protein coding sequence was inserted into the unique NheI cloning site of the plasmid pQBI T7-GFP (Quantum Biotechnologies). This construct allows the expression, under the control of a T7 promoter, of the green fluorescent protein (GFP) fused to the C-terminus of the inserted sequence (Figure 1). The resulting plasmids were used for in vitro transcription/translation assays in reticulocyte lysates (TNT® T7 reticulocyte lysate, Promega) containing 20 µCi of 35S Methionine (>1000 Ci/mmol) according to manufacturer’s instructions. The same constructions were made with a DNA sequence from nucleotide 412 to nucleotide 480 (Figure 1). The resulting plasmids were used for in vitro transcription/translation assays in a reticulocyte lysate as described above.

For immunoprecipitation, monoclonal anti-GFP antibody (mAb 3E6, Q.BIOgene) was conjugated to protein A-sepharose for 1h at 4°C in PBS buffer. Antibody conjugated beads were then equilibrated in non-denaturing lysis buffer (1% triton (w/v), 50 mM Tris-HCl pH 7.4, 300 mM NaCl, 5 mM EDTA, 1 mM PMSF) and added to 10 µl reticulocyte lysate in 500 µl of non-denaturing lysis buffer for 1h at 4°C. After repeated washing with wash buffer (lysis buffer containing 0.1% Triton), proteins were eluted with Laemmli sample buffer. Immunoprecipitated proteins were analysed by SDS-PAGE, autoradiographed and scanned using a STORM 860 (Molecular Dynamics) Phosphorimager.

Expression in culture cells- The DNA fragments described above and used to test the ability of the corresponding RNA sequences to direct frameshifting in a reticulocyte lysate assay were used for cellular expression. They were cloned into the unique NheI site of the plasmid pQBI25-fPA (Quantum Biotechnologies). These constructs allow the expression, under the control of a CMV promoter, of GFP fused to the C-terminus of the inserted sequence (Figure 1). Hela cells were grown and maintained in Glasgow minimal Eagle’s medium supplemented with 10% fetal calf serum, 100 IU/mL penicillin/streptomycin (Sigma). For transfection, cells
were washed, treated with trypsin and plated in an 8 chambers culture slide (Falcon) at the required density in a humidified CO2 incubator (5%) at 37°C overnight. Cells were transfected by calcium phosphate according to the manufacturer’s instructions (Invitrogen). 16 hours after transfection, the cells were fixed by 4% paraformaldehyde in PBS at 4°C for 30 min and nuclei were stained in PBS containing 5 µg/mL Hoechst 33258 (Sigma). Fluorescence microscopy was performed using a Zeiss Axioplan 2 microscope.

Immunoblotting- Samples were separated by SDS-PAGE and electrotransferred to nitrocellulose membranes with a blotting apparatus (BioRad laboratories). Membranes were blocked with 5% non-fat dry milk in phosphate-buffered saline at room temperature for 1 h. Monoclonal anti-core antibody (AbCys) (1/5,000 dilution) and phosphatase alkaline-labeled goat anti-mouse immunoglobulin G IgG (H+L) (1/1,000 dilution; BioRad laboratories) were used to detect the expression of HCV core protein. SuperSignal®West HisProbe™Kit (Pierce) was used to detect 6xHis tagged proteins according to the manufacturer’s instructions.

EIA- Three synthetic peptides encoded in the +1 ORF of core and predicted to be antigenic were obtained by chemical synthesis:. Peptide F1 (core[11-25], NVTPTAAHRTLSSRA), peptide F2 (Core[46-60], ARLGRLPSGRNLVEG) and peptide F3 (Core[106-120], GAPQTPGVGRVIWVR). For the enzyme immunoassay (EIA) the wells of a microtiter plate were coated with 0.5 µg of peptide and incubated with 10 µl of human serum and 90 µl of diluent at room temperature for 1h. The wells were washed and subsequently incubated with 100 µl of a 1 :3,000 dilution of a horserasish peroxydase-conjugated goat anti-human antibody (Pierce). The wells were washed again and allowed to react with 2,2’-Azino-bis(3-ethylbenz-thiazoline-6-sulfonic acid) and H2O2 for color development. The reaction was analysed at 405 nm in a Dynex MRX microplate reader.
RESULTS

Identification of alternative forms of HCV core protein

Expression of a plasmid coding for the HCV core protein fragment \([C_{HCV1-169}(6xHis)]\) led to the production of protein found predominantly in the inclusion bodies. The proteins were extracted with 6M urea in the presence of 10mM \(\beta\)-Mercaptoethanol and the presence of polyhistidine fused to the C-terminus permitted its purification on Ni-NTA agarose. A second purification step on reverse phase HPLC permitted the separation of \(C_{HCV1-169}(6xHis)\) (peak 5) from minor peaks (peaks 1 to 4) that represented about 35% of total protein (Figs. 2A and 2B). Mass spectrometry analysis of peak 5 gave a molecular weight of 19,421 Da that corresponded to \(C_{HCV2-169}(6xHis)\) (Table I) and thus indicated the removal of the N-terminal methionine residue. Analysis of minor peaks by immuno-blotting showed that they all reacted with an antibody directed against the N-terminus of the core protein (Fig. 2C) and with an anti-(6xHis) antibody (Fig. 2D). Therefore, the fractions 1 to 4 eluting earlier than \(C_{HCV2-169}(6xHis)\) by reverse phase HPLC contained proteins harboring both the N- and C-termini of \(C_{HCV2-169}(6xHis)\) but exhibiting a lower molecular mass. Mass spectrometry analysis gave molecular weights of 18,381 Da (peak 1), 18,282 Da (peak 2), 18,395 Da (peak 3) and 18,335 Da (peak 4) that were incompatible with proteolytic clivages at the N-terminus of \(C_{HCV2-169}(6xHis)\) and represent alternative forms of \(C_{HCV2-169}(6xHis)\) as demonstrated below.

SDS-PAGE and immuno-blott analysis also revealed that some of these proteins formed dimers that were resistant to SDS-PAGE (Fig. 2B, C and D, lanes 1 to 3).

Alternative core proteins result from multiple recoding events

In order to obtain the full-length amino acid sequences, the protein eluting at peaks 1 to 4 from the HPLC chromatography (Fig. 2A) and \(C_{HCV2-169}(6xHis)\) (peak 5, Fig. 2A) were
submitted to V8 proteolysis. The patterns of V8 proteolysis of these proteins were very similar, but very different from that of C_{HCV}2-169(6xHis) (result not shown). LC/MS analysis revealed that V8 fragments obtained from C_{HCV}2-169(6xHis) had molecular masses identical to those predicted from their amino acid sequences (Table I). In the case of the alternative protein, the presence of a 1,974 Da fragment corresponding to the C terminus of C_{HCV}2-169(6xHis) confirmed the previous finding concerning the reactivity of this protein with an anti-(6xHis) antibody (Fig. 2D). In contrast, fragments of 6,539 Da and 9,920 Da were not consistent with V8 proteolysis fragments of C_{HCV}2-169(6xHis). Each of these two fragments was submitted to chemical amino acid sequencing. As shown in Table I, the 6,539 Da fragment corresponded to the first 40 amino acids of C_{HCV}2-169(6xHis) followed by 18 amino acids resulting from a +1 ribosomal frameshifting in the core protein coding sequence. This frameshifting occurred at codon 42 in the core protein coding sequence and resulted in the reading of a GGU codon coding for Gly in the +1 ORF instead of the reading of an AGG codon coding for Arg in the 0 ORF (Figure 3). The 9,920 Da fragment was too long for complete sequencing but sequencing of its N-terminus gave the sequence GDNLSPRL, corresponding to the continuation of translation in the +1 ORF (Table I). The presence of the 1,974 Da fragment corresponding to the C-terminus in the 0 ORF strongly suggested the occurrence, in the 9,920 Da fragment, of a translational event leading to re-phasing in the 0 ORF. This re-phasing event should have occurred before or at the termination codon at position 144 in the +1 ORF. In order to test this hypothesis, extensive tryptic cleavage of the 9,920 Da fragment was carried out and the peptide mixture was analyzed by LC/MS. All fragments could be attributed to sequences in the 0 or +1 ORF except for a fragment at m/z 1,009.6 for [M+H]^+ corresponding to the peptide straddling the re-phasing site in the 0 ORF. Looking at the collision source fragmentation, we observed abundant and characteristic y-ion fragments at m/z 896.5 (y_{10}), 809.4 (y_9), 738.5 (y_8), 641.5 (y_7), 544.4 (y_6) and 431.4 (y_5)
according to the notation proposed by Roepstorff and Fohlman (21). These fragment ions corresponded to the amino acid sequence LSAPPL, which matched the predicted sequence LSAPPLGGAAR resulting from a –1 frameshifting. The molecular mass of 9,920 Da is consistent with the occurrence of a –1 frameshift at the UAG termination codon in the +1 ORF. This frameshift results in the reading of a CUA codon coding for Leu and allows translation to continue in the 0 ORF, leading to the recurrence of the usual C-terminal core protein sequence in the alternative protein (Figure 3). In conclusion, the protein eluted at peak 3 results both from a +1 frameshift at codon 42 and a –1 frameshift at codon 144, leading to an alternative core protein containing the first 42 amino acids of the core protein, 101 residues coded by the +1 ORF, and then residues 144 to the 6xHis tag of the core protein. This protein is thus an alternative form of core and was named DFC_{HCV} protein for “Double Frameshifted Core” protein.

The other minor protein fractions eluting from HPLC chromatography (peaks 1, 2 and 4 in figure 2A) were also submitted to extensive tryptic cleavage and peptide mixtures were submitted to LC/MS. All fragments, including that corresponding to the +1 frameshift site at codon 42 could, in all cases, be attributed to sequences present in the DFC protein except for the fragment straddling the –1 frameshift site. As reported in Table II, the protein from peak 1 contained a tryptic peptide that could be attributed to two +1 frameshift events at codons 144 and 145 in the +1 ORF. The protein from peak 2 contained a tryptic peptide that could be attributed to a +2 frameshift event at codon 144 in the +1 ORF. Finally, the fraction from peak 4 contained a tryptic peptide that could be attributed to the same –1 frameshift, as observed in DFC protein (at codon 144) together with the bypass of the codon 145 or 146 (both of them coding for Gly). All these events lead to a rephasing in the 0 ORF of the core coding sequence. Integration of the peaks from the RP-HPLC (Figure 1) gave the following ratios: 68% for C_{HCV}2-169(6xHis), 9% for peak 1, 3% for peak 2, 18% for peak 3 and 2% for
peak 4. It is worth mentioning that the alternative core protein resulting from the +1 frameshift only (i.e. ending at the termination codon in the +1 ORF without occurrence of the –1 frameshift) was found in the flowthrough fraction of the Ni-NTA agarose chromatography as expected since it did not harbor the C-terminal 6xHis tag. It was characterized by immunoprecipitation (using an anti-core antibody directed against the N-terminus of Core protein) followed by RP-HPLC purification and mass spectroscopy (data not shown). The ratio of this protein fraction was estimated to be equivalent to the sum of fractions from peaks 1 to 4.

**Sequences including the shift sites direct recoding both in a eukaryotic in vitro transcription/translation assay and in culture cells**

To test the ability of the sequence including codon 42 (nucleotides 126 to 128) to direct +1 frameshifting in a eukaryotic translational context, the core DNA sequence from nucleotides 100 to 200 was inserted into a plasmid upstream from the gene coding for GFP used as a reporter protein. Three constructs were made for which the DNA coding for GFP was fused in the three reading frames with regard to the upstream inserted sequence (Figure 1). As can be seen in Figure 4A, expression of the resulting plasmids in a reticulocyte lysate yielded a large amount of fused GFP cloned in the 0 frame (lane 4). Interestingly, a low but detectable amount of fused GFP cloned in the –1 frame attests for the occurrence of a +1 frameshifting (lane 3). In contrast, no fused GFP cloned in the +1 frame was detected (lane 5). These results are to be compared to a blank without plasmid (lane 1) and to the control GFP plasmid (lane 2). The amount of +1 frameshifting was determined to be 1.9% when compared to the product of the control GFP plasmid taken as 100% (Figure 4B lanes 3 and 2 respectively). The core DNA sequence used in the *in vitro* reticulocyte lysate assay described above were cloned into a plasmid permitting the production of fused GFP under the control of a CMV promoter in
eukaryotic cells (Figure 1). As can be seen in Fig. 5, the results obtained confirmed those obtained in the reticulocyte lysate assay. The sequence used to test the occurrence of a +1 frameshifting led to the production of GFP in the cells when the protein was cloned either in the −1 frame (A) or in the 0 frame (B), but did not lead to the production of GFP when it was cloned in the +1 frame (C). Hence this sequence is able to direct +1 ribosomal frameshifting in a cellular context.

Concerning the −1 frameshift at codon 144 (nucleotides 432 to 434) in the +1 ORF, a DNA sequence from nucleotides 412 to 480 was inserted into a plasmid upstream of the gene coding for GFP used as a reporter protein. Here again, three constructs were made for which GFP was fused in the three reading frames (Figure 1) and tested in the reticulocyte lysate assay. As can be seen in figure 4A, when GFP was fused in the +1 frame (lane 6), a fusion product was observed attesting for the occurrence of a −1 frameshifting at a rate of 2.18% (Figure 4B, lane6). When GFP was fused in the 0 frame (lane 7), the production of a GFP fusion product (at a rate of 1.70%; Figure 4B, lane 7) of nearly the same size than GFP alone (compare to control GFP, lane 2) attest for the bypass of the STOP codon in the 0 frame and might be explained by an internal initiation at a downstream ATG codon. When GFP was fused in the −1 frame (lane 8), the mechanism leading to the production of a GFP fusion product (at a rate of 1.95%; Figure 4B, lane8) is not clear but it might arise from a +1 frameshifting leading to a fusion protein with a slightly higher apparent molecular mass due to the presence of numerous arginine residues. When this sequence, used to test the occurrence of a re-phasing at the level of the STOP codon in the +1 ORF, was assayed in a cellular context it led to the production of GFP in the cells whatever the frame of the GFP cloning (Figure 5D, E, F). All together, these results reflect the ability of the core nucleotide sequence lying between nucleotides 412 to 480 to induce unusual recoding events in a
eukaryotic translational context leading to the bypass of the termination codon 144 in the +1 ORF.

**Sera from HCV positive patients are reactive against synthetic peptides encoded in the +1 open reading frame.**

Three synthetic peptides encoded in the +1 ORF of the core protein of genotype 1b were chemically synthetized and used in EIA to test the reactivity of sera from HCV positive patients of genotype 1a and 1b (Figure 6). As shown in this Figure, when the peptide used belongs to the amino acid sequence in the +1 ORF located before amino acid 42 (Core[11-25], namely peptide F1), 2 out of 10 sera from genotype 1a are reactive whereas no serum from genotype 1b is reactive. In contrast, when the peptides used belong to the amino acid sequence in the +1 ORF located after amino acid 42 (namely peptides F2 (Core[46-60]) and F3 (Core[106-120])), 3 out of 10 sera from genotype 1a and 6 out of 10 sera from genotype 1b react with peptide F2, and 3 out of 10 sera from genotype 1a and 1 out of 10 sera from genotype 1b react with peptide F3. These results strengthen the hypothesis that an alternative core protein is expressed *in vivo* in genotype 1b as a result of a +1 frameshifting. In addition, the absence of reactivity of sera of genotype 1b against peptide F1 (Core[11-25]) indicate that the shift site is downstream of position 25. This support our finding that the shift site leading to F protein is different in genotypes 1a and 1b (shift site at or near codon 11 for the former, and at codon 42 for the latter).
DISCUSSION

In this report we demonstrate that alternative HCV core proteins from the genotype 1b are expressed in *Escherichia coli* as a result of two different recoding events, that are a +1 frameshifting at codon 42, that can be followed by a re-phasing in the normal open reading frame at stop codon 144 using multiple mechanisms. These recoding events can be reproduced both in an *in vitro* eukaryotic translational context and in culture cells.

Protein sequencing results shown in Table I demonstrate that the +1 frameshift occurs at codon 42 in the 0 ORF for all forms of alternative core protein. This result is different from previous reports for HCV genotype 1a, showing that the -2/+1 ribosomal frameshifting requires only codons 8-14 of the core protein-coding sequence and that the shift junction is located at or near codon 11 (14). This -2/+1 frameshifting at codon 11 is linked to an Arg to Lys codon mutation that generates a region of 10 consecutive adenines (16) (17) and leads to the synthesis of a 16 kDa protein, named F protein that might be related to the previously identified p16 protein (22). Discrepancy with our data likely arises from the use of HCV core proteins from different genotypes, i.e. genotype 1a in the previous reports and genotype 1b in this report. This is strengthened by the absence of the Arg to Lys mutation at codon 11 in any of the HCV core RNA sequences of genotype 1b as well as of genotypes other than 1a reported in the sequence databases (i.e. HCVDB database (http://hepatitis.ibcp.fr, data not shown). However, +1 encoded core protein was detected in clinical isolates from genotype 1b containing mutations in codons 9-11 which failed to reproduce the 10 adenine region (23), and antibodies against peptides in the +1 ORF of core were detected in sera of infected patients of different genotypes (14) (15). Then, our results strongly indicate that in the absence of the 10 adenines region, a +1 frameshifting can occur at codon 42, at least in genotype 1b. The cis-acting elements responsible for the +1 frameshift mechanism are less
well defined than those responsible for the –1 frameshift mechanism (see below). A recent study shows that some underrepresented heptanucleotides in S. cerevisiae supported notable +1 frameshifting and that there was nothing about those sequences to suggest that they would do so (24). It is worth to mention that several authors predicted RNA secondary structures downstream codon 42 (25) (26). One can suppose that such structural elements might play a role in the +1 frameshifting event. The details of the molecular mechanisms responsible for this +1 ribosomal frameshifting are currently under investigation.

Our sequencing results (Tables I and II) indicate that the termination codon UAG in the +1 ORF (codon 144) could be bypassed by recoding events leading to the continuation of translation in the 0 ORF. The most frequently observed event is a –1 frameshifting occurring at the termination codon UAG in the sequence G CCC CCC UAG and leading to the reading of a CUA codon coding for a Leu. In canonical –1 frameshifting, first described by Jacks and collaborators studying Rous sarcoma virus (RSV) (27) (28), frameshifting occurs at a slippery heptamer with the sequence X-XXY-YYZ. However, some frameshift sites do not fit this canonical description, for example, G-UUA-AAC for the equine arthritis virus (29) or G-GAU-UUA at the pro-pol junction in mouse mammary tumor virus (MMTV) (30). In potato virus M, frameshifting occurs at A-AAA-UGA, stimulated by the UGA termination codon that can be replaced by UAA or UAG without effect on the frameshifting event (31). Efficient –1 frameshifting usually requires an RNA secondary structure downstream to the slippery heptamer (32). Two recent reports account for the presence of RNA secondary structures in this region of HCV core RNA: phylogenetic sequence analysis suggests the presence of a stem-loop structure between nucleotides 438 and 516 (25), and a thermodynamic and phylogenetic analysis suggests an equivalent structure between nucleotides 443 and 475 (26). It is thus reasonable to speculate that frameshifting at termination codon is due to the presence of such RNA structures.
The question arises whether or not the two frameshift events observed in bacteria for the HCV core protein occurs in eukaryotes and especially in human cells infected by the virus. The –1 ribosomal frameshifting in prokaryotes differs in some ways from the eukaryotic paradigm described above (for review, see (33) (34)). Using an appropriate reporter gene, both the MMTV frameshift (35) and the HIV-1 frameshift (36) have been reproduced in bacteria at rates ranging from 2% to 50%, demonstrating that *Escherichia coli* ribosomes are able to shift frame in the –1 direction in the same manner as their eukaryotic counterparts. This is in keeping with our observations showing that the cloning of a sequence surrounding the –1 shift site (codon 144) in fusion with GFP permitted the expression of the fusion protein both in reticulocyte lysate transcription/translation assays (Fig. 4) and in Hela cells (Fig. 5) as a result of –1 frameshifting. However, this sequence is also able to direct recoding events other than –1 frameshifting, as also observed in bacteria (Table II). This indicates the existence of a multiple frameshift site and supports the presence of shifty elements such as a predicted secondary structure downstream of the RNA. Concerning the +1 shift site identified at codon 42 in bacteria, a sequence surrounding this shift site (but excluding codon 11 identified as the +1 shift site in genotype 1a) is also shown to be effective both in reticulocyte lysate (Fig. 4) and in Hela cells (Fig. 5). Thus, the +1 and –1 frameshifts identified in bacteria for the HCV core protein are likely relevant in eukaryotic systems and probably reflect translational events that might occur in HCV infected cells. We have tried to determine directly the amino acid sequence of the frameshift products synthesized in the reticulocyte lysate system by Edman’s degradation after labeling of expressed proteins with [35S Met] and [3H] amino acids (e.g., [3H Gly]). However, due to the very low abundance of the frameshift products (2%) and the limited [3H] amino acid specific radioactivity, it was not technically possible to get any result by this approach to date.
IgG’s specific for peptides derived from a HCV core protein encoded in the +1 ORF were detected in chronic HCV sera of various genotypes including 1b (14) (15). We have also found numerous sera from HCV positive patients from genotypes 1a and 1b reactive against synthetic peptides encoded in the +1 ORF (Figure 6). The reactivity of sera from genotype 1b is not consistent with the production of a F protein due to a +1 frameshift at codon 11 since the 10 adenines cluster required for the frameshift event is absent from all the genotype 1b RNA sequences present in the HCVDB database. Moreover, although the panel of sera tested is quite reduced, it appeared that no serum from genotype 1b reacted with peptide F1 located before the identified shifty codon 42 while two sera of genotype 1a reacted against this peptide. Hence, the reactivity of these genotype 1b sera likely arise from the production of a F protein with a different +1 frameshift site when compared to genotype 1a. Interestingly, using overlapping synthetic peptides encoded in the +1 ORF of core protein, we detected T cell responses strikingly biased toward the production of interleukin 10 in 7/25 patients infected either with genotypes 1a, 1b or 3 viruses (C. Bain et al., submitted). These data also support the production of +1 encoded core proteins in other genotypes than 1a.

Our results in E.coli show the occurrence of the two recoding events on the same sequence leading to an alternative core protein harboring N- and C-terminal domains identical to those of the core protein encoded in the 0 ORF but with a different 101 amino acid central domain defined by the two frameshift sites (Table I). However, in the eucaryotic context, the frequency of each frameshift events is quite low (about 2%). Consequently, the probability for the production of this double frameshifted core protein in very low (about 0.04%) and its biological relevance is thus questionable. The single occurrence of +1 frameshift at position 42 leads to a F protein harboring the 42 first aa of core protein followed by 101 residue coded by the +1 ORF. This F protein of 143 aa is different and shorter than that of genotype 1a which exhibits 160 aa or so, including only the 8 to 11 first aa of core protein followed by
residues coded by the +1 ORF. However, both proteins harbor a large common and conserved region coded by +1 ORF, explaining the immunological cross reactivities for them. These F protein discrepancies between genotypes are enigmatic. It should be point out that in the case of genotype 1b, F protein include the so-called immunodominant antigenic domain of core protein (37) that have been shown to bear at least one conformational epitope involving an helix-loop-helix structure, as determined by NMR (38).

Although the primary function of the core protein is the formation of the viral nucleocapsid, numerous functional analyses have shown that the core protein can modulate gene transcription, cell proliferation and cell death, interfere with lipid metabolism, and suppress host immune responses (reviewed in (9) (10) (39)). Co-expression of F and/or alternative forms, together with core protein, renders the various activities attributed to the core protein worthy of being carefully revisited. Indeed, even if frameshifts efficiency seems to be low in a eukaryotic expression system (estimated to be about 2% by the scanning of the gels in the reticulocyte lysate assays), alternative core proteins might regulate cellular functions that are important for the viral life cycle or might play a role in viral morphogenesis or viral entry. In addition, F protein of genotype 1b might have previously escaped attention as it also possesses the immunodominant epitopes located in the N-terminus, in common with the core protein.
2. Choo, Q. L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C.,
   2451-2455.
3. Robertson, B., Myers, G., Howard, C., Brettin, T., Bukh, J., Gaschen, B., Gojobori, T.,
   Maertens, G., Mizokami, M., Nainan, O., Netesov, S., Nishioka, K., Shin i, T.,
   2503
   116
   **67**, 1385-1395.
8. Harada, S., Watanabe, Y., Takeuchi, K., Suzuki, T., Katayama, T., Takebe, Y., Saito, 
    3980-3988.
    721.


**FOOT NOTES**

This work was supported by research grants from Université Claude Bernard Lyon I, from CNRS « Programme Physique et Chimie du Vivant », from ARC (n° 4301 to J.P.L.) and from “Programme de recherche fondamentale en microbiologie et maladies infectieuses et parasitaires” from MENRT (n° 1A031G to J.P.L.).

**Acknowledgements**

We wish to thank Dominique Mazzocut and Madeleine Courteau for their skilled technical assistance in protein sequencing, Anne Laure Nouvion for her skilled technical assistance in cell culture, and Dr G. Inchauspe for the gift of the 15 mer synthetic peptides.
FIGURE LEGENDS

Fig.1: **Construction of GFP fusion expression vectors.** The inserted sequences are described in “Experimental procedures”. In pQBI-42(0), the GFP reporter was fused to the same reading frame as the core protein sequence. In pQBI-42(+1), the GFP reporter was fused to the +1 reading frame. In pQBI-42(-1), the GFP reporter was fused to the –1 reading frame.

In pQBI-144(0), the GFP reporter was fused to the same reading frame as the core protein sequence. In pQBI-144(+1), the GFP reporter was fused to the +1 reading frame. In pQBI-144(-1), the GFP reporter was fused to the –1 reading frame. The same constructs were made in the plasmid pQBI25-fPA for cell expression under the control of a CMV promoter, leading to the pQBI25-42 and pQBI25-144 series. Small letters: GFP gene; capital letters: Core gene.

The first codon of GFP is boxed.

Fig.2: **Purification and characterization of the various forms of HCV core protein.**

A: Proteins retained on Ni-NTA agarose were submitted to RP-HPLC as described in “Experimental procedures”.

B: Coomassie blue staining following SDS-PAGE of protein fractions separated by RP-HPLC. Western-blot analysis of the different fractions with an anti-core antibody (C) or with an anti-His6 antibody (D) were carried out as described in “Experimental procedures”.

Fig 3: **Amino acid sequences encoded in the three ORF of core.** The shifty codons 42 and 144 are boxed and the corresponding amino acids in the 0 and +1 ORFs are underlined.

Fig. 4: **In vitro transcription/translation assays.**

A: Templates for the recoding assays were constructed as described in “Experimental procedures”. (see also Figure 1) and tested in a rabbit reticulocyte lysate *in vitro* transcription/translation assay. The products were immunoprecipitated with anti-GFP antibody, separated by SDS-PAGE, dried and autoradiographed. Blank without plasmid (lane 1), control GFP (lane 2), pQBI-42(-1) (lane 3), pQBI-42(0) (lane 4), pQBI-42(+1) (lane 5), pQBI-144(-1) (lane 6), pQBI-144(0) (lane 7)
and pQBI-144(+1) (lane 8). **B**: Relative percentages of the transcription/translation products determined using a Phosphorimager as described in “Experimental procedures”.

Fig. 5: **Cellular expression of the GFP fusion constructs.** Templates for the recoding assays were constructed as described in “Experimental procedures”. (see also Figure 2). Cells were transfected with these constructs and observed by microscopy for the fluorescence of the GFP fusion proteins. **A**: pQBI25-42(-1), **B**: pQBI25-42(0), **C**: pQBI25-42(+1), **D**: pQBI25-144(-1), **E**: pQBI25-144(0) and **F**: pQBI25-144(+1).

Fig 6: **Reactivity of sera from genotypes 1a and 1b HCV infected patients against synthetic peptides encoded in the +1 ORF of core protein.**

The sera from HCV infected patients were tested by EIA as described in “Experimental procedures” using synthetic peptides encoded in the +1 ORF of core protein (namely peptides F1, F2 and F3). The blank was obtained with the sera of an HCV negative patient. The sera were considered to be positive when the OD$_{405}$ had a positive-to-negative ratio $\geq 2$. OD$_{405}$, optical density at 405 nm.
| Fragment of CHCV (2-169)6His (peak 5) and of peak 3 (i.e., DFCHCV protein) resulting from 3 h endoproteinase Glu-C proteolysis were submitted to LC/MS analysis and peptide sequencing as described in “Materials and Methods”.

| a | b |

| Table I: Mass determination and amino acid sequencing of endoproteinase Glu-C cleavage products. |

| a | b |

<table>
<thead>
<tr>
<th>C&lt;sub&gt;HCV&lt;/sub&gt;(2-169)6His (MW 19,421 Da)</th>
<th>DFC&lt;sub&gt;HCV&lt;/sub&gt; (peak 3) (MW 18,395 Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mass (Da) measured</td>
<td>mass (Da) calculated</td>
</tr>
<tr>
<td>5,929.03±0.56</td>
<td>5,927.83</td>
</tr>
<tr>
<td>2,185.81±0.82</td>
<td>2,185.53</td>
</tr>
<tr>
<td>1,992.42±0.58</td>
<td>1,992.18</td>
</tr>
<tr>
<td>7,413.77±1.50</td>
<td>7,413.64</td>
</tr>
<tr>
<td>1,974.22±0.27</td>
<td>1,975.04</td>
</tr>
<tr>
<td>HPLC fraction</td>
<td>Sequence of tryptic segment</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Peak 1</td>
<td>LSAPPRA&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>139 140 141 142 143 144</td>
</tr>
<tr>
<td></td>
<td>UUG UCG GCG CCC CCC</td>
</tr>
<tr>
<td></td>
<td>Arg  Gly Ala Ala Arg&lt;sup&gt;d,c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Peak 2</td>
<td>LSAPPGGAAR</td>
</tr>
<tr>
<td></td>
<td>139 140 141 142 143 144</td>
</tr>
<tr>
<td></td>
<td>UUG UCG GCG CCC CCC</td>
</tr>
<tr>
<td></td>
<td>Gly Gly Ala Ala Arg</td>
</tr>
<tr>
<td>Peak 3</td>
<td>LSAPLPGAAR</td>
</tr>
<tr>
<td></td>
<td>139 140 141 142 143 144</td>
</tr>
<tr>
<td></td>
<td>UUG UCG GCG CCC CCC</td>
</tr>
<tr>
<td></td>
<td>Leu Gly Gly Ala Ala Arg</td>
</tr>
<tr>
<td>Peak 4</td>
<td>LSAPLPGAAR</td>
</tr>
<tr>
<td></td>
<td>139 140 141 142 143 144</td>
</tr>
<tr>
<td></td>
<td>UUG UCG GCG CCC CCC</td>
</tr>
<tr>
<td></td>
<td>Leu Gly Gly Ala Ala Arg</td>
</tr>
</tbody>
</table>

**Table II:** Amino acid sequences observed in alternative core proteins at the second frameshift site (termination codon 144 in the +1 ORF).

<sup>a</sup>Numbering of codons is according to the +1 ORF. The termination codon 144 in the +1 ORF is indicated in bold. Codons and amino acids in the 0 ORF are underlined.

<sup>b</sup>The codon and amino acid (Arg) in the +2 ORF are indicated by a wavy line.

<sup>c</sup>Amino acids (and corresponding codons) belonging to the next tryptic fragment are in italics.
Boulant et al. Fig 1
Boulant et al. Fig 2
Fig 3
Boulant et al. Fig 4
Boulant et al. Fig 5
Boulant et al. Fig 6.
Unusual multiple recoding events leading to alternative forms of hepatitis C virus core protein from genotype 1b

Steeve Boulant, Michel Becchi, François Penin and Jean-Pierre Lavergne

J. Biol. Chem. published online September 2, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M307174200

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2003/09/02/jbc.M307174200.citation.full.html#ref-list-1