Two Alternative Translation Mechanisms Are Responsible for the Expression of the HCV ARFP/F/Core+1 Coding Open Reading Frame*

Niki Vassilaki and Penelope Mavromara‡

From the Molecular Virology Laboratory, Hellenic Pasteur Institute, 127 Vas. Sophias Avenue, Athens, Greece 11521

HCV-1 produces a novel protein, known as ARFP, F, or core+1. This protein is encoded by an open reading frame (ORF) that overlaps the core gene in the +1 frame (core+1 ORF). In vitro this protein is produced by a ribosomal frameshift mechanism. However, similar studies failed to detect the ARFP/F/core+1 protein in the HCV-1a (H) isolate. To clarify this issue and to elucidate the functions of this protein, we examined the expression of the core+1 ORF by the HCV-1 and HCV-1a (H) isolates in vivo, in transfected cells. For this purpose, we carried out luciferase (LUC) tagging experiments combined with site-directed mutagenesis studies. Our results showed that the core+1-LUC chimeric protein was efficiently produced in vivo by both isolates. More importantly, neither changes in the specific 10-A region of HCV-1 (codons 8–11), the proposed frameshift site for the production of the ARFP/F/core+1 protein in vitro, nor the alteration of the ATG start site of the HCV polyprotein to a stop codon significantly affected the in vivo expression of the core+1 ORF. Furthermore, we showed that efficient translation initiation of the core+1 ORF is mediated by internal initiation codon(s) within the core/core+1-coding sequence, located between nucleotides 583 and 606. Collectively, our data suggest the existence of an alternative translation initiation mechanism that may result in the synthesis of a shorter form of the core+1 protein in transfected cells.

Hepatitis C virus (HCV) is the main etiologic agent of post-transfusion and sporadic non-A, non-B hepatitis in the world. This virus establishes chronic infection in most acutely infected individuals, frequently leading to liver cirrhosis and hepatocellular carcinoma. HCV is an enveloped, single-stranded, positive sense RNA virus. It is a member of the Hepacivirus genus within the Flaviviridae family. The viral genome is ~10-kb long and encodes a 3011-amino acid polyprotein precursor. This polyprotein is co- and post-translationally processed by cellular and viral proteases giving rise to three structural (core, E1, and E2) and at least six non-structural (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) proteins. Initiation of translation of the HCV genome is controlled by an internal ribosome entry site (IRES) located mainly within the 5′-non-coding region of the viral RNA, between nucleotides 42 and 341 or 356, the 3′-limit being controversial. The core protein, which forms the viral nucleocapsid, is predicted to be 191 amino acids long and to have a molecular mass of 23 kDa (p23). Further processing of p23 produces the mature core protein (p21), consisting of 173–182 amino acids. An additional HCV polypeptide of 16/17 kDa (p16/p17) has recently been discovered. For this, we carried out transient transfection experiments, both HCV-1 and HCV-1a (H) efficiently express the core+1-LUC chimeric protein. This form of the core+1-LUC chimeric protein...

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Molecular Virology Laboratory, Hellenic Pasteur Institute, 127 Vas. Sophias Ave., Athens, Greece 11521. Tel.: 210-6478875; Fax: 210-6478877; E-mail: penelopm@scl.gr.

‡ The abbreviations used are: HCV, hepatitis C virus; ORF, open reading frame; nt, nucleotides; IRES, internal ribosome entry site; LUC, luciferase; CAT, chloramphenicol acetyltransferase.
Expression of the HCV Core +1 ORF

ARFP/F/core +1 chimeric protein is synthesized even when the core is not produced.

EXPERIMENTAL PROCEDURES

Plasmid Construction and Site-directed Mutagenesis—Site-directed mutagenesis was carried out using the Quikchange™ kit (Stratagene). The templates and oligonucleotides used in the mutational analysis and the corresponding mutants are listed in Table I. All mutations were confirmed by sequence analysis. The HCV-1 cDNA sequences were obtained from plasmids path 10/17-38 and path 11/36-27, kindly provided by Dr. M. Beach (CDC). The HCV-1 (H) cDNA sequence was obtained from pDNA-C1, kindly provided by Dr. G. Inchauspe.

The dicistronic constructs pHPI-1391, -1333, and -1352 contain the chloramphenicol acetyltransferase (CAT) gene as the first cistron followed by the entire IRES and part of the wild-type core-coding sequences (nt 9–630) from the prototype HCV-1 isolate fused to the firefly LUC gene in the 0, +1, and −1 frames, respectively. They were produced by site-directed mutagenesis from dicistronic pHPI-1311, −1313 and −1312, respectively, using primers 5′-GGATCCCAACCGATGATTTAC1312 (sense) and 5′-GGCGTCTTCCCCTTTGGATCCA-3′ (antisense). This set of primers converts the start codon of the luciferase-coding region into a glycine codon (ATG GGG).

In Vitro Transcription and Translation—In Vitro transcription and translation experiments were carried out on uncapped RNAs in a total volume of 25 µl using [35S]methionine (Amersham Biosciences). The translation products (5 µl) were analyzed by 12% SDS-PAGE, transferred onto nitrocellulose membranes, and visualized by autoradiography.

Cells and DNA Transfection Experiments—BHK-21 and Huh-7 cells were maintained in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal bovine serum (DMEM/FBS) at 37 °C in a 5% CO2 incubator. Cells seeded in 6-well plates (60% confluence) were transfected with 1 µg of plasmid DNA in the presence of Lipofectamine plus plus reagent (Invitrogen) according to the manufacturer’s protocol. The medium was replaced with new DMEM/FBS 24 h post-transfection. The cells were washed twice with phosphate-buffered saline 48 h post-transfection and lysed in 260 µl of 1× luciferase lysis buffer (Promega). Firefly LUC was quantified by mixing 20 µl of extract with 100 µl of luciferase assay reagent (Promega) and measuring the luminescence directly with a Turner TD-20/20 luminometer. In the case of the dicistronic constructs, CAT was quantified with the CAT-ELISA kit (Roche Applied Science) according to the manufacturer’s instructions.

Antibodies—The goat polyclonal antibody against the firefly luciferase protein was obtained from Promega Corporation at a concentration of 1 mg/ml.

Immunoprecipitation Analysis—Thirty-six hours after transfection with pHPI-1362 or pHPI-1363, monolayers of BHK-21 cells (∼106 cells) were metabolically labeled for 12 h with 20 µCi of [35S]methionine (Amersham Biosciences) per ml of methionine-free medium supplemented with 1% fetal bovine serum. The labeled cells were rinsed with phosphate-buffered saline and lysed in 500 µl of total volume of triple detergent buffer consisting of 50 µM Tris (pH 8), 150 mM NaCl, 0.1% SDS, 1% Nonidet P-40, 0.5% sodium deoxycholate, and 100 µg/ml phenylmethylsulfonyl fluoride. Cell lysates were mixed by vortexing and centrifuged at 14,000 × g for 10 min at 4 °C. Clarified lysates were incubated with 10 µl of anti-LUC polyclonal antibody on a rocker overnight at 4 °C. Protein G PLUS-agarose (Santa Cruz Biotechnology) was added (20 µl) to this mixture and the reactions were incubated in the same conditions for an additional 2 h. After microcentrifugation, the agarose beads were washed three times with a buffer containing 50 mM Tris (pH 8), 150 mM NaCl, 0.1% Nonidet P-40, and 1 mM EDTA. The immunoprecipitates were subsequently resolved by 10% SDS-PAGE, transferred onto nitrocellulose membranes, and detected by autoradiography.
Expression of the HCV Core+1 ORF

RESULTS

The Core+1 ORF Is Efficiently Expressed in Transfected Cells—As all previous studies concerning the expression of the core+1 ORF have been carried out primarily in *in vitro* systems based on rabbit reticulocyte lysates, we examined core+1 translation from the HCV-1 and HCV-1a (H) isolates in mammalian cells. For this, we carried out transient transfection assays, initially in BHK-21 cells, and monitored the expression of core+1 ORFs by tagging the ORFs with the luciferase cistron, GGG, derived from the ATG initiator by site-directed mutagenesis, is boxed. The LUC gene was fused in the +1 frame relative to the preceding core-encoding sequence in pHPI-1331 (HCV-1) and pHPI-1334 (HCV-1a (H)), in the +1 frame in pHPI-1333 (HCV-1) and pHPI-1335 (HCV-1a (H)), and in the −1 frame in pHPI-1332 (HCV-1) and pHPI-1336 (HCV-1a (H)). The *underlined* nucleotide indicates an insertion of a thymidine residue, and the inverted triangle indicates a deletion of an adenine residue. Panels B and C, in *vivo* (a) and *in vitro* (b) expression of the HCV-1 (B) and HCV-1a (H) (C) fusion constructs. *a*, duplicate cultures of BHK-21 cells were transfected with each construct and the relative ratio of LUC activity to CAT quantity was determined. *Bars* represent the means obtained in two separate experiments in duplicate. *Error bars* represent the standard deviation. *b*, each construct was transcribed *in vitro* and equal amounts of all RNAs were translated in Flexi rabbit reticulocyte lysates. Translation products were directly separated by SDS-PAGE and analyzed by autoradiography. Fusion proteins are indicated by filled arrowheads. Open arrowheads show the CAT protein. NC, negative control.

A. Fig. 1. Analysis of the expression of the core+1-LUC chimeric gene. Panel A, schematic representation of the CAT-LUC dicistronic constructs used for the tagging experiments. The entire IRES (nt 9–341) and part of the core-coding sequence (nt 342–630) from HCV-1 and HCV-1a (H) were fused with the LUC gene under the control of both CMV and T7 promoters of pHPI-1048 (14). The nucleotide sequences of the junction between the core and luciferase-coding regions are illustrated. The first codon of luciferase cistron, GGG, derived from the ATG initiator by site-directed mutagenesis, is boxed. The LUC gene was fused in the +1 frame relative to the preceding core-encoding sequence in pHPI-1331 (HCV-1) and pHPI-1334 (HCV-1a (H)), in the +1 frame in pHPI-1333 (HCV-1) and pHPI-1335 (HCV-1a (H)), and in the −1 frame in pHPI-1332 (HCV-1) and pHPI-1336 (HCV-1a (H)). The underlined nucleotide indicates an insertion of a thymidine residue, and the inverted triangle indicates a deletion of an adenine residue. Panels B and C, in *vivo* (a) and *in vitro* (b) expression of the HCV-1 (B) and HCV-1a (H) (C) fusion constructs. *a*, duplicate cultures of BHK-21 cells were transfected with each construct and the relative ratio of LUC activity to CAT quantity was determined. *Bars* represent the means obtained in two separate experiments in duplicate. *Error bars* represent the standard deviation. *b*, each construct was transcribed *in vitro* and equal amounts of all RNAs were translated in Flexi rabbit reticulocyte lysates. Translation products were directly separated by SDS-PAGE and analyzed by autoradiography. Fusion proteins are indicated by filled arrowheads. Open arrowheads show the CAT protein. NC, negative control.

B. HCV-1

C. HCV-1a (H)

Directly related to the expression of the fused core or core+1-coding sequences, and CAT activity serves as an internal control to standardize transfection efficiencies *in vivo* or potential variations in the transcript abundance *in vitro*.

BH21 cells were transfected with each construct, and 48 h later LUC and CAT activities were measured. Fig. 1, B and C show the level of LUC expression relative to CAT in the transfected cells. In the case of HCV-1, substantial amounts of luciferase were expressed from the core+1-LUC cassette of pHPI-1333, as the levels of the luciferase activity were similar to that of the core-LUC fusion protein derived from pHPI-1331 (Fig. 1Ba). Only background levels of luciferase activity were detected following the expression of the corresponding negative control core-1-LUC construct (pHPI-1332). Surprisingly, in contrast to *in vitro* (9), very high levels of luciferase activity were observed from construct pHPI-1335, which contains the core+1 ORF from HCV-1a (H) fused to the LUC gene. The levels were about 200% of that of the HCV-1a (H) core-LUC hybrid protein yielded from pHPI-1334 (Fig. 1Ca). Again, the corresponding negative control plasmid (pHPI-1336) resulted in background levels of luciferase expression. Thus, the HCV-1 core+1 LUC-tagged protein is efficiently produced *in vivo*, with
FIG. 2. Effect of mutations within codons 8–11 of the HCV-1 (N18, N19) and HCV-1a (H) (N15, N16) core-coding sequence on the expression of core+1-LUC chimeric gene. Panels A and B, the core nucleotide sequences in the region of codons 8–11 of the wild-type HCV-1 (A) and HCV-1a (H) (B) plasmids, as well as of the corresponding mutant variants N18, N19 (HCV-1) (A) and N15, N16 (HCV-1a (H)) (B). The
Expression of the HCV Core+1 ORF

similar translation levels as the core-coding sequence. More importantly, unlike in rabbit reticulocyte lysates, our data indicate that the HCV-1a (H) isolate efficiently expresses the core+1 ORF in transiently transfected BHK-21 cells.

To confirm the divergence between the previous in vitro and our current in vivo studies, the same DNA plasmids were tested for expression in vitro in rabbit reticulocyte lysates (Fig. 1, B and C). As anticipated, the in vitro translation of the core-LUC construct from HCV-1 (pHPI-1331) or HCV-1a (H) (pHPI-1334) resulted in the synthesis of a chimeric core-LUC protein with an apparent molecular mass of 72 kDa (Fig. 1, B and Ch, lane 1). Furthermore, in agreement with previous data (9), when pHPI-1333, containing the core+1-LUC cassette from HCV-1, was used, a hybrid protein with an apparent molecular mass of 72 kDa was produced (Fig. 1Bh, lane 2). On the contrary, pHPI-1335, carrying the equivalent core+1-LUC cassette from HCV-1a (H), exhibited no luciferase expression (Fig. 1Cb, lane 2). The 72-kDa protein was not produced by pHPI-1332 (HCV-1) or pHPI-1336 (HCV-1a (H)), which carry the LUC gene fused to the –1 frame of the core-coding region (Fig. 1, B and Ch, lane 3). These data confirm previous in vitro expression studies from our laboratory showing that the core+1 ORF was expressed from the HCV-1 core-coding region but not from the HCV-1a (H) equivalent. They also reveal that the predominant translation mechanism that directs the expression of ARFP/F/core+1 ORF differs in vitro and in transfected cells.

The A-rich Sequence at Codons 8–11 of the HCV Core-coding Region Is Not Essential for the Production of the ARFP/F/Core+1 Protein in Transfected Cells—The core-coding region of HCV-1 (H) lacks the 10 consecutive A residues found at codons 8–11 (nt 364–373) of the HCV-1 genome, a known slippery site prone to ribosomal frameshift events. Thus, we directly assessed the importance of the 10-A residue region for the expression of the core+1 ORF in transfected cells as compared with that in rabbit reticulocyte lysates. For this purpose, mutational studies were performed based on previously described naturally occurring mutations at the proposed frameshifting site (codons 8–14). Specifically, Yeh et al. (15) provided evidence that a p16 protein from the core-coding region of clinical isolates from patients with hepatocellular carcinoma was produced in vitro in rabbit reticulocyte lysates. The detected protein was reported as being a short form of the core protein even though the protein was immunologically distinct from the p21 core. The presence of this protein was associated with specific missense mutations within codons 9–11, none of which generate a 10-A residue sequence. To determine the effect of these naturally occurring mutations on the production of the ARFP/F/core+1 protein in vitro, the dicistronic constructs pHPI-1333 (HCV-1) and pHPI-1335 (HCV-1a (H)) were used as templates. In the case of HCV-1 (Fig. 2A), insertion of mutation N18, which contains two A to G substitutions and an A to C substitution at nt 366, 367, and 373 respectively (codons 9 and 11), gave rise to pHPI-1382, whereas mutation N19, which carries an A to G substitution and two A to C substitutions at nt 367, 369, and 373 respectively (codons 9, 10, and 11), resulted in pHPI-1383. For HCV-1a (H) (Fig. 2B), insertion of mutation N15, which contains an A to G change at position 366 (codon 9), gave rise to pHPI-1395, and insertion of N16, which carries an A to C substitution at nt 369 (codon 10), resulted in pHPI-1396. Both N15 and N16 mutations contain a single substitution as the HCV-1a (H) isolate already carries a G and a C at positions 367 and 373 respectively. None of these mutations had a significant effect on luciferase activity in vitro (Fig. 2, C and Dd), indicating that the presence of the 10 consecutive adenines at codons 8–11, as found in the HCV-1 isolate, is not critical for core+1 ORF expression in vivo. However, in vitro none of the above mutants produced detectable levels of the core+1-LUC hybrid protein (Fig. 2, C and Dh, lanes 2 and 3). Thus, mutations that disrupt the 10-A residue sequence within codons 8–11 of the core-coding region fail to support the production of the core+1-LUC protein in vitro, whereas the same mutations have no effect on the expression of the protein in vivo. Thus, our data do not appear to be consistent with those previously reported (15). The reason for this is not clear at the moment, but it is likely to be due to differences in the core-coding sequence beyond codons 9–11 between the HCV isolates used in the two studies or to differences between the two rabbit reticulocyte lysate expression systems. Overall, our results indicate that the A-rich region at codons 8–11 is critical for the expression of the ARFP/F/core+1 protein only in rabbit reticulocyte lysates.

The ATG Initiator Codon of the HCV Core-coding Region Is Not Essential for the Expression of the ARFP/F/Core+1 Protein in Transfected Cells—To investigate further the molecular mechanism(s) implicated in the in vivo expression of the core+1 ORF, two mutations were introduced into the core-coding region of the core+1-LUC-tagged constructs of both the HCV-1 (pHPI-1333) and HCV-1a (H) (pHPI-1335) isolates (Fig. 3A). Mutation N3 converted the ATG initiator codon of the core ORF into a terminator codon and mutation N6 introduced a stop codon at the 25th position of the core-coding sequence at nt 414 (P25, CCG). The resulting plasmids were named pHPI-1343 and pHPI-1344 for HCV-1, and pHPI-1346 and pHPI-1347 for HCV-1a (H), respectively. If a ribosomal frameshift mechanism is indeed responsible for the production of the ARFP/F/core+1 protein in vivo, then the N3 mutation should abrogate the production of the chimeric protein, whereas N6 should have no effect. However, transfection studies showed that not only the N3 mutation failed to block core+1 expression but also the resulting mutant exhibited increased levels of luciferase activity (Fig. 4, A and B) compared with the wild-type core+1-LUC (positive control). Similar data were obtained for both HCV-1 (Fig. 4Aa) and HCV-1a (H) (Fig. 4Ba). Additionally, the N6 mutation resulted in increased levels of luciferase activity in both isolates; the levels of luciferase activity detected in these constructs were similar to those observed in constructs carrying the N3 mutation (Fig. 4, A and Ba). These results suggest that, in contrast to in vitro (9), the in vivo translation of the core+1 ORF does not require core expression. Instead, our data suggest that blocking the translation of the core ORF has a positive effect on the translation of the core+1 ORF.

To confirm again the differences between the in vitro and in vivo results, the same DNA plasmids were tested in vitro in weak-type sequences of codons 8–11 are shown in bold. The arrows indicate the inserted mutations and the bold characters indicate the mutated nucleotides and affected amino acids. The numbers in brackets indicate the number of the mutated codons. Panels C and D, the HCV-1 (C) and HCV-1a (H) (D) wild-type (pHPI-1333 and pHPI-1335, respectively) and corresponding mutants (pHPI-1382 (N18), –1383 (N19), and pHPI-1395 (N15), –1396 (N16), respectively) were used to transfect BHK-21 cells (a) or transcribed in vitro and equal amounts of RNAs were translated in Flexi rabbit reticulocyte lysates (b). a. duplicate cultures of BHK-21 cells were transfected with the wild-type or the mutated constructs. The activity of each mutant was calculated by determining the ratio of LUC activity to CAT quantity and is expressed as a percentage of that of the wild type. Bars represent the means observed for three separate experiments each carried out in duplicate. Error bars correspond to the standard deviation. b, 5A of the [35S]methionine-labeled in vitro translation products were separated by 12% SDS-PAGE and analyzed by autoradiography. The core+1-LUC fusion protein is indicated by the filled arrowhead. Open arrowheads show the CAT protein. WT, wild type; NC, negative control.
rabbit reticulocyte lysates. Consistent with previous in vitro studies (9), the N3 mutation abrogated the synthesis of the 72-kDa core/H110011-LUC protein from HCV-1 (Fig. 4Ab, lane 2), whereas N6 had no effect on the production of the core/H110011-LUC chimeric protein (Fig. 4Ab, lane 3). Furthermore, as expected, the core/H110011-LUC constructs (WT, N3, N6) from HCV-1a (H) failed to produce detectable levels of the chimeric protein (Fig. 4Bb, lanes 1, 2, 3). Thus, there appears to be a difference between the predominant translation mechanism for ARFP/F/core/H110011 expression between rabbit reticulocyte lysates and transfected cells, as expression of the core/H110011 ORF in vivo does not require the ATG start codon of the HCV polyprotein. Taken together these results indicate that ribosomal frameshift events are not the predominant mechanism directing core/H110011 expression in transfected cells.

**Efficient Translation of the Core+1 ORF in Transfected Cells Is Mediated by Internal Initiation Codon(s)---**As the expression of the core+1 ORF in vivo is not suppressed by changes in the initiator ATG or the A-rich region, we hypothesized that downstream codon(s) may function as translation initiation sites for the expression of the core+1 ORF. To examine this hypothesis two series of mutagenesis experiments were carried out.

Firstly, three nonsense mutations were separately inserted into the core+1-coding sequences of HCV-1 and HCV-1a (H), in pHPI-1333 (HCV-1) and pHPI-1335 (HCV-1a (H)), which carry the dicistronic CAT-IRES-core/H110011-LUC cassette. To facilitate the description of the mutations affecting the core/H110011 ORF, we arbitrarily defined the GCA alanine codon at nt 346 as the first codon of the core/H110011 ORF (Fig. 3B). The N1 mutation introduced a TAG stop codon into the core/H110011 ORF at nt 472 (Trp43, TGG), resulting in pHPI-1342 and pHPI-1345, respectively for HCV-1 and HCV-1a (H). Mutation N21 changed the 79th codon of the core/H110011 ORF at nt 580 (Gly79, GGT) to a TAG stop codon, yielding pHPI-1380 (HCV-1), and pHPI-1398 (HCV-1a (H)). Finally, mutation N22 introduced a TAG terminator codon eight codons downstream of mutation N21 at nt 604 (Met87), giving rise to pHPI-1381 (HCV-1) and pHPI-1397 (HCV-1a (H)). As expected, in vitro expression from the plasmids carrying the N1, N21, and N22 mutations failed to support the production of detectable levels of the core/H110011-LUC gene as shown by...
the fact that the level of luciferase activity was about the same as in the wild-type constructs in both BHK-21 and Huh-7 cells (Fig. 5, A-D, a and b). No differences were observed between HCV-1 (pHPI-1342, -1380), and HCV-1a (pHPI-1345, -1398). On the contrary, the N22 mutation almost completely abolished the synthesis of the core+1-LUC protein from both HCV-1 and HCV-1a (H), in both BHK-21 and Huh-7 cell lines, as the levels of luciferase activity were similar to those of the negative control (Fig. 5, C and D, a and b). Only background levels of luciferase activity were detected from the negative controls pHPI-1332 (HCV-1) and pHPI-1336 (HCV-1a (H)). Therefore, these data support the conclusion that efficient translational initiation of the core+1 ORF is mediated from internal initiation codons that may be located between nt 583 and 606 (codons 80–87).

Secondly, as the region between nt 583 and 606 (codons 80–87) contains two ATGs (nt 598-ATGNNNATG-606), we assessed the functional importance of these ATGs as initiation sites for the translation of the core+1 protein in vivo. For this purpose, we changed ATG598 (85th codon) and ATG604 (87th codon) of the dicistronic HCV-1 construct pHPI-1333 simultaneously and separately to the non-initiator codon GGG. Mutation N25 (Fig. 3B) converted both methionines at positions 85 and 87 to glycines, resulting in pHPI-1401, whereas mutation N23 (Fig. 3B) altered only Met85 and mutation N24 (Fig. 3B) altered only Met87 giving rise to pHPI-1399 and pHPI-1400, respectively. The transfection of BHK-21 (A) and Huh-7 (B) with mutants pHPI-1399 (N23) and pHPI-1400 (N24) yielded similar levels of luciferase translation as the wild-type construct (Fig. 6). In contrast, mutation N25 severely affected the production of the chimeric core+1-LUC protein, which was about 23% of the wild-type level in BHK-21 cells (Fig. 6A) and about 26% in Huh-7 (Fig. 6B). These results suggest that the two methionines (Met85 and Met87) of the core+1-coding region are involved in core+1 expression, as conversion of both of them to glycine significantly reduced the levels of luciferase activity. Interestingly, however, the conversion of each of the two methionines separately to glycine had no effect on the expression of the core+1 ORF.  

Finally, to confirm the above data, we compared the size of the core+1-LUC proteins produced in vitro and in vivo. Taking into account the above results, we expected that the protein predominantly produced in transfected cells would have a molecular mass of about 62 kDa, that is 10 kDa shorter that the 72-kDa core+1-LUC protein synthesized in vitro. To test this hypothesis, the IRES-core+1-LUC cassette contained in the dicistronic construct pHPI-1333 (HCV-1), as well as the corresponding negative control IRES-core-1-LUC cassette of pHPI-1332 were transferred into a monocistronic expression vector under the control of a CMV promoter, resulting in pHPI-1362 and pHPI-1363 respectively (Fig. 7A). This system improves the detection of the luciferase protein, as HCV IRES is more active in monocistronic constructs. Specifically, the luciferase activity exhibited by the monocistronic IRES-core+1-LUC construct pHPI-1362 in BHK-21 cells forty-eight hours post-transfection was about 9-fold higher than that yielded from the dicistronic pHPI-1333 (Fig. 7B). Only background levels of luciferase activity were exhibited from the negative control pHPI-1363. Thus, immunoprecipitation experiments were carried out with extracts of BHK-21 cells transfected with pHPI-1362, using a goat polyclonal antibody raised against luciferase. As predicted from our mutational analysis, a protein with an apparent molecular mass of around 62 kDa reacted strongly with the polyclonal antibody (Fig. 7C, lane 2). This protein was clearly smaller than the chimeric core+1-LUC protein produced in vitro from the pHPI-1333 construct (Fig. 7C, lane 3). No protein was produced by the negative control pHPI-1363 (Fig. 7C, lane 1). These results are consistent with our mutagenesis data (Figs. 5 and 6) and support the hypothesis that an internal translational initiation site is being used during the core+1 synthesis in vivo.  

Overall, these data provide strong evidence that a shorter
Fig. 5. Mutational analysis within the core+1-coding sequence of HCV-1 and HCV-1a (H) isolates. The HCV-1 (A and C) and HCV-1a (H) (B and D) wild type (pHPI-1333 and pHPI-1335, respectively), and mutated plasmids (pHPI-1342 (N1), pHPI-1380 (N21), pHPI-1381 (N22), and pHPI-1345 (N1), pHPI-1398 (N21), pHPI-1397 (N22), respectively) were expressed in BHK-21 (a) and Huh-7 (b) cells or in Flexi rabbit reticulocyte lysates (c).
form of core+1 protein is produced in transfected cells by an alternative translational mechanism that directs efficient translational initiation from internal codons within the core+1-coding sequence.

**DISCUSSION**

In this study, we provide several lines of evidence supporting an alternative translation mechanism for the expression of the HCV core+1 ORF in transfected cells. This alternative mechanism is predicted to direct the synthesis of a shorter form of the ARFP/F/core+1 protein.

First, the core+1 ORF from both HCV-1 and HCV-1a (H) isolates is efficiently translated in transfected cells as shown by the high levels of luciferase activity produced from dicistronic isolates is efficiently translated in transfected cells as shown by the high levels of luciferase activity produced from dicistronic constructs containing the HCV-1 or HCV-1a (H) core cDNA sequence fused with the luciferase gene in the +1 frame. This is in contrast to the results of expression studies in rabbit reticulocyte lysates (Fig. 1C), which reproducibly failed to detect expression of the core+1 ORF from the HCV-1a (H) isolate (9).

Second, naturally occurring mutations identified within codons 9–11 of the core-coding sequence (N18 and N19 for HCV-1 or N15 and N16 for HCV-1a (H)) in clinical isolates from cancer patients had no effect on the translation of the core+1-LUC in transfected cells. This indicates that the expression of the core+1 ORF is independent on the A-rich nature of codons 8–11, the region proposed to be the frameshift site for the core+1 translation in vitro. In contrast, these mutations abolished the production of the hybrid protein in rabbit reticulocyte lysates, supporting the critical role of this region for the in vitro production of the ARFP/F/core+1 protein (p16/p17).

Third, the N3 mutation, which converted the initiator ATG codon of the HCV polyprotein into a stop codon, did not abolish the production of the core+1-LUC fusion protein in vivo, indicating that efficient translation initiation of the core+1 ORF does not require the polyprotein initiator codon. In contrast, as expected from previous studies, the same mutation failed to initiate the translation of the core+1-LUC protein in vitro.

Fourth, further mutational studies suggested that the translational initiation site for core+1 is located between nt 583 and 606 within the core-coding sequence, as the nonsense mutation of the 43rd codon, at nt 472 (mutation N1), or of the 79th codon, at nt 580 (mutation N21) of the core+1 ORF did not affect core+1-LUC translation, whereas the nonsense mutation of codon 87, at nt 604 (mutation N22), abolished the production of the core+1-LUC fusion protein (arbitrarily starting measurement from the GCA alanine codon of the core+1 ORF). Additionally, we showed that the efficiency of core+1 expression is dependent on the presence of ATG598 or/and ATG604, the two ATGs contained in the region of nt 583–606, as mutation N25, which converted both of these ATGs to GGG, reduced core+1 expression to about 25% of that in the wild-type.

Fifth, immunoprecipitation analysis of the chimeric core+1-LUC protein produced in mammalian cells indicated that the immunoprecipitated protein was smaller (by about 10 kDa) than the core+1-LUC hybrid protein produced in vitro (Fig. 7).

These data suggest that in transfected cells, efficient translation initiation of the core+1 ORF is mediated from internal codon(s) located between nt 583 and 606, which may coincide with the ATG598 and ATG604 of the core+1 ORF. Consequently, the predominant form of ARFP/F/core+1 protein produced in vivo in these conditions should be smaller than the 16/17-kDa product synthesized in vitro, as it is predicted to lack the first 85 amino acids. Notably, the shorter form of the ARFP/F/core+1 protein is still a very basic protein (pl~12) and contains one of the two previously predicted hydrophobic domains in its N-terminal half.

It should be noted that the production of the ARFP/F/core+1 protein was not affected by the conversion of only one of the two ATGs at positions 598 and 604 of the core+1 ORF to GGG (mutations N23, N24), whereas it was significantly reduced by the mutation of both of these ATGs to GGG (mutation N25). This suggests that both ATGs are involved in the initiation of **A.**

**Fig. 6. Effect of mutations targeting codons ATG598 and ATG604 of the core+1-coding sequence.** Duplicate cultures of BHK-21 (A) and Huh-7 (B) cells were transfected with the dicistronic HCV-1 wild-type (pHPI-1333) and mutated constructs: pHPI-1399 (N23), pHPI-1400 (N24), and pHPI-1401 (N25). The relative activity of each mutant variant was calculated as described in the legend of Fig. 2. Bars represent the means from two separate experiments each carried out in duplicate. Error bars represent the S.D. WT, wild type; NC, negative control.

---

*a and b, the experiments were carried out in duplicate and repeated at least twice. The relative activity of each mutant variant was determined as described in the legend of Fig. 2. Bars represent the means. Error bars correspond to the S.D. c, translation products were separated by SDS-PAGE and analyzed by autoradiography. The positions of the hybrid core+1-LUC and the CAT proteins are indicated by filled and open arrowheads, respectively. WT, wild type; NC, negative control.*
core+1 translation and are able to substitute for each other. It is noteworthy that the ATG\textsuperscript{598} and ATG\textsuperscript{604} codons are well conserved among HCV isolates. A comparative analysis of 117 HCV core sequences from the genotypes 1a, 1b, 2a-c, 2k, 3a, 3b, 4a-f, 5a, 6a, 6d, 6e, and 6k, from the GenBank\textsuperscript{TM} data base, revealed that 66 variants contain both ATG\textsuperscript{598} and ATG\textsuperscript{604}, 23 carry only the ATG\textsuperscript{598} or the ATG\textsuperscript{604}, and 5 lack both ATGs. This high level of conservation is consistent with these ATGs having a functional role during core\textsuperscript{1} translational initiation.

Notably, the dicistronic constructs pHPI-1312 (HCV-1) and pHPI-1330 (HCV-1a (H)), which contain the LUC gene fused in the frame of the core-coding sequence at nt 630, displayed high levels of luciferase activity when the initiator ATG codon of the LUC gene was included (data not shown). This expression was abolished only after the conversion of the ATG start codon of LUC to GGG. This finding suggests that any ATG start codon located in the vicinity of nt 583 and 606 is functional.

It should also be emphasized that the conversion of both ATGs to GGG did not completely abolish the expression of the core\textsuperscript{1}-LUC protein (25% of the wild-type level), which is consistent with the concomitant expression of the larger (16/17 kDa) form of the core\textsuperscript{1} protein. Indeed, we have preliminary evidence supporting the expression of both forms of the ARFP/F/core\textsuperscript{1} protein (data not shown). Clearly, however, in the conditions used in our experiments translation of the core\textsuperscript{1} ORF from internal initiation codons was very efficient.

The physiological implications of these findings on the expression of core\textsuperscript{1} in \textit{vivo} remain to be elucidated. However, the short form of the ARFP/F/core\textsuperscript{1} protein is probably synthesized in conditions that restrict the expression of the viral polyprotein, inasmuch as suppression of core expression in the presence of mutation N3 or N6 failed to abolish the production of core\textsuperscript{1}-LUC. This suggests that the ARFP/F/core\textsuperscript{1} protein may play a critical role in controlling the life cycle of the virus.

RNA and to a lesser extent DNA viruses that are subjected to genome size constraints have devised strategies to expand their coding capacity, such as ribosomal frameshifting (16) and internal translational initiation. Different mechanisms that allow escape from an upstream initiator codon and direct initiation from internal codons have been described; these include context-dependent leaky-scanning (13, 17, 18, 19, 20, 21), ribosome shunting (17, 22, 23, 24, 25, 26, 27), and IRES-mediated initiation (28). The internal initiation of ARFP/F/core\textsuperscript{1} translation \textit{in vivo} might include features of one of these mechanisms. If the expression of core\textsuperscript{1} is indeed mediated by an IRES element, this element is predicted to be a different IRES from that located at the 5'-end of the viral RNA, as we have data suggesting that core\textsuperscript{1} translation \textit{in vivo} does not require the HCV IRES (data not shown).

The reason for the divergence between our \textit{in vitro} and \textit{in vivo} results is not clear. It is possible that the ARFP/F/core\textsuperscript{1} protein is more efficiently expressed \textit{in vitro} because of the presence of additional factors such as ribosomes, tRNAs, and translation initiation factors. Alternatively, the ARFP/F/core\textsuperscript{1} protein may be unstable \textit{in vivo} due to degradation by cellular proteases or other factors. Further studies are needed to resolve these issues and to gain a better understanding of the physiological role of the core\textsuperscript{1} protein in the life cycle of the virus.

**Fig. 7. Expression of the chimeric core\textsuperscript{1}-LUC protein in transfected cells.** Panel A, schematic representation of the monocistronic constructs pHPI-1362 (core\textsuperscript{1}-LUC) and pHPI-1363 (core\textsuperscript{1}-LUC). Panel B, duplicate cultures of BHK-21 cells were transfected with the monocistronic core\textsuperscript{1}-LUC construct pHPI-1362 or the dicistronic core\textsuperscript{1}-LUC pHPI-1333 and the relative luciferase activity was determined. Bars represent the means from two separate experiments. Error bars represent the S.D. Panel C, immunoprecipitation of [\textsuperscript{35}S]methionine-labeled translation products of the core\textsuperscript{1}-LUC and core\textsuperscript{1}-LUC containing monocistronic constructs from transiently transfected BHK-21 cells using an anti-LUC goat polyclonal antibody. The immunoprecipitates were analyzed by SDS-PAGE followed by autoradiography. The hybrid core\textsuperscript{1}-LUC protein produced \textit{in vivo} is marked by a dot. The open arrowhead shows the [\textsuperscript{35}S]methionine-labeled core\textsuperscript{1}-LUC protein synthesized in rabbit reticulocyte lysates. NC, negative control.
vivo results for the principal mechanism involved in the core +1 translation remains unclear. Recent data have shown that another mechanism related to in vivo translational initiation, the interaction between the cap and poly(A) tail of eukaryotic mRNAs, does not occur in Flexi rabbit reticulocyte lysates. Instead, the mechanism operates in a modified system based on ribosome-depleted rabbit reticulocyte lysates (29, 30). Alternatively, the translation of the ARFP/F/core protein may be regulated by different mechanisms depending on the cellular conditions and the ways in which the translation machinery is modified during HCV infection.

Acknowledgments—Plasmids path 10/17–38 and path 11/36–27 were obtained from the Centers for Disease Control and Prevention, National Center for Infectious Diseases. Plasmid pDNA-C1 was kindly provided by Dr. G. Inchauspe. We thank N. Michaelidou for excellent technical assistance with tissue culture and our colleagues from the Molecular Center for Infectious Diseases. Plasmid pDNA-C1 was kindly provided by Dr. G. Inchauspe. We thank N. Michaelidou for excellent technical assistance with tissue culture and our colleagues from the Molecular Virology laboratory for helpful discussions.

REFERENCES
Two Alternative Translation Mechanisms Are Responsible for the Expression of the HCV ARFP/F/Core+1 Coding Open Reading Frame
Niki Vassilaki and Penelope Mavromara

doi: 10.1074/jbc.M305504200 originally published online July 21, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M305504200

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 29 references, 17 of which can be accessed free at http://www.jbc.org/content/278/42/40503.full.html#ref-list-1