Expression of Unmodified Hepatitis C Virus Envelope Glycoprotein-Coding Sequences Leads to Cryptic Intron Excision and Cell Surface Expression of E1/E2 Heterodimers Comprising Full-Length and Partially Deleted E1

Julie Dumonceaux,†‡ Emmanuel G. Cormier,†‡ Francis Kajumo,† Gerald P. Donovan,‡ Jayanta Roy-Chowdhury,† Ira J. Fox,§ Jason P. Gardner,‡ and Tatjana Dragic†*

Albert Einstein College of Medicine, Bronx, New York 10461; Progenics Pharmaceuticals, Inc., Tarrytown, New York 10591; and Department of Surgery, Nebraska Medical Center, Omaha, Nebraska 98198

Received 22 July 2003/Accepted 12 September 2003

Hepatitis C virus (HCV) is a positive-strand RNA virus that replicates exclusively in the cytoplasm of infected cells. The viral envelope glycoproteins, E1 and E2, appear to be retained in the endoplasmic reticulum, where viral budding is thought to occur. Surprisingly, we found that the expression system used to generate HCV envelope glycoproteins influences their subcellular localization and processing. These findings have important implications for optimizing novel HCV fusion and entry assays as well as for budding and virus particle formation.

It is estimated that 170 million people worldwide are infected with the hepatitis C virus (HCV) and are at risk of developing chronic hepatitis or cirrhosis, the latter often leading to hepatocellular carcinoma (1, 11, 29, 30). Basic understanding of HCV replication and pathogenesis remains poor, primarily due to a lack of experimental models and key reagents. The HCV genome is a 9.4-kb single-stranded positive RNA that replicates exclusively in the cytoplasm of infected cells (47). The genomic RNA encodes a ~3,000-amino-acid polyprotein that is processed to generate at least 10 proteins, termed C, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B (47). The C protein constitutes the nucleocapsid; E1 and E2 are transmembrane envelope glycoproteins; p7 is a membrane-spanning protein of unknown function; and the various nonstructural (NS) proteins have replication functions (2, 37).

Translocation of the HCV structural proteins, C-E1-E2-p7, into the endoplasmic reticulum (ER) is accompanied by cleavage of internal signal sequences by ER-resident signal peptidases. Proper folding of the envelope glycoproteins, E1 and E2, is dependent on their cotranslation and processing within the context of the HCV polyprotein (7, 13, 14, 16, 34, 36, 40). The transmembrane (TM) domains of E1 and E2 are also the signal sequences (SS) of E2 and p7, respectively, and topological reorientation of cleaved E1 and E2 C termini toward the cytosol converts these regions into single membrane-spanning domains (5, 9). The TM domains of E1 and E2 mediate their noncovalent heterodimerization (13, 15, 38, 39, 45, 51). The apparent absence of E1/E2 heterodimers on the cell surface as well as the lack of N-glycan modifications by Golgi enzymes suggest that the HCV envelope glycoproteins are retained in the ER (16, 32, 36, 40, 51). The TM domains of E1 and E2 mediate ER retention as well as E1/E2 heterodimerization (6, 8, 10, 19, 20), which has made it difficult to generate cell surface-expressed E1/E2 heterodimers for use in cell fusion and viral pseudotype assays.

The initial goal of our studies was to create recombinant cell surface-expressed HCV envelope glycoproteins that would be incorporated onto pseudovirions and mediate entry into HCV target cells. We chose a strategy wherein the ectodomains of HCV E1 and E2 were fused to the TM domains of E1 and E2 from a related alphavirus, the Semliki Forest virus (SFV). The SFV envelope glycoproteins form cell surface-associated heterodimers that efficiently pseudotype heterologous viral nucleocapsids in order to mediate their entry into host cells. Chimeric HCV-SFV envelope glycoproteins were similar to unmodified HCV envelope glycoproteins in size and posttranslational processing and were expressed on the cell surface. During these experiments, however, we detected unmodified, wild-type HCV E1 and E2 on the cell surface, and this unexpected observation was investigated further.

Constructs were generated for expression of full-length unmodified HCV E1 (E1), E2 (E2), and E1/E2 (E1-E2). DNA encoding HCV envelope glycoproteins was derived by PCR amplification of p90/HCV FL-long pU carrying the full-length genome of an infectious HCV isolate H77 (26). The first nucleotide of the C start codon is defined as position +1 in the HCV coding sequence. E1 therefore comprises nucleotides 511 to 1149, E2 comprises nucleotides 1111 to 2238, and E1-E2 includes nucleotides 511 to 2238 of HCV. Start and stop translation codons were introduced at the beginning and end of every construct, which were cloned into the pcDNA3.1+ expression vector (Invitrogen) and verified by DNA sequencing. Transient expression of HCV envelope glycoproteins was achieved by lipofection of the different expression plasmids into HeLa cells. Alternatively, HeLa cells were infected with a
vaccinia virus vector expressing the T7 polymerase (vTF7.3 [17]) followed by lipofection with HCV E1/E2-expression plasmids (resulting in cytoplasmic transcription from the T7 promoter in pcDNA3.1+).

Intracellular but not cell surface-associated E1 and E2 were detected by immunofluorescence after vaccinia virus-driven expression of E1-E2 (Fig. 1a to d). In contrast, E1 and E2 were detected both intracellularly and on the cell surface following plasmid-based expression of the E1-E2 construct (Fig. 1f to h). Intracellular staining of envelope glycoproteins was comparable in the two expression systems (Fig. 1a, b, e and f), indicating that similar levels of E1 and E2 were being generated but transport to the cell surface was not occurring in the vaccinia virus-based expression system. Differences in expression levels were not observed between envelope glycoproteins expressed as single proteins (E1 and E2) or as part of an E1-E2 polypeptide (E1-E2) (data not shown). Similar expression patterns were obtained after transient expression of E1 and E2 by vaccinia- and plasmid-based systems in a hepatoma cell line, HepG2 (data not shown).

Vaccinia vector- and plasmid vector-based expression of unmodified HCV envelope glycoproteins invariably generated a major E2 protein species of 62-kDa apparent molecular size (Fig. 2a and b). Vaccinia expression of unmodified E1 and E1-E2 constructs generated four E1 protein species with apparent molecular sizes of 18, 21, 24, and 27 kDa (Fig. 2c). Treatment of cell lysates by endoglycosidase H (endo H) generated a single small-molecular-size band corresponding to the deglycosylated E1 protein core (Fig. 2c). The 27-kDa species therefore corresponds to the fully glycosylated E1 protein, whereas the lower-molecular-size species correspond to incomplete glycosylation products. In contrast, plasmid-based expression of E1-E2 generated E1 proteins of 27 and 20 kDa apparent molecular sizes (Fig. 2d). The 20-kDa E1 species was not the result of hypoglycosylation, because two E1 protein species were still present after endo H treatment of cell lysates (Fig. 2d) (see below).

Vaccinia virus-based expression of heterologous genes results in transcription in the cytoplasm. In contrast, transient and stable expression by DNA vectors results in nuclear transcription followed by mRNA maturation and transport to the cytoplasm. We hypothesized that mRNA modifications may account for the different E1 species expressed by plasmid vectors. We therefore performed reverse transcription-PCR on RNA extracts of HeLa cells stably expressing E1. Sequence analyses of the PCR products detected a deletion between nucleotides 675 and 887 (inclusive) which preserves the E1 reading frame and would encode a protein with a 71-amino-acid deletion (Fig. 3a and data not shown). This deletion was not present in PCR products derived by amplification of genomic DNA containing the integrated E1 construct (data not shown).

Analysis of the entire HCV sequence by a splice site prediction neural network (http://www.fruitfly.org/seq_tools/splice.html) revealed the presence of a putative splice donor site in position 675 of E1, whereas putative splice acceptor sites were found in position 887 of E1 and 2183 of E2 (Fig. 3a). Based on these findings, we assumed that putative intron splicing between positions 675 and 887 resulted in expression of a partially deleted E1 protein, corresponding to the 20-kDa species observed after transient expression of E1 and E1-E2 constructs. The splice acceptor at position 2183 does not appear to be functional.

We therefore generated new E1 expression constructs wherein the splice acceptor site in E1 was removed by conser-
Cell surface-associated E1 protein could not be detected in any of the stable HeLa clones by flow cytometry using two different anti-E1 monoclonal antibodies (MAbs), A4 (15) and 081-5 (Austral Biologicals) (data not shown). We reason that their epitopes may not be accessible in the full-length protein. Cell surface-associated E1 was readily detected, however, by cell surface biotinylation followed by streptavidin capture and immunoblotting of E1*-E2* and E1*-E2*-p7-expressing cells with an anti-E1 MAb (Fig. 5a). When E1 was expressed alone it was not detectable on the cell surface, suggesting that coexpression of E2 is required for efficient transport of E1 to the plasma membrane (Fig. 5a).

Cell surface-associated E2 was detected in stable HeLa clones by flow cytometry after labeling with four different anti-E2 MAbs of E2*, E1*-E2*, and E1*-E2*-p7-expressing cells (Fig. 5b and data not shown). MAb H2, which has been reported to recognize E1/E2 heterodimers, did not recognize cell surface E2, but this antibody is also not reactive with HCV particles in patient sera (13). E1*-E2* and E1*-E2*-p7 also were stably expressed in hepatic NKNT3 cells, which display morphological characteristics of liver parenchyma cells, express key genes of liver metabolism, and are not tumorigenic in SCID mice (24, 25). E2 was readily detected on the surface of NKNT3 cells, suggesting that plasma membrane localization is an inherent property of HCV envelope glycoproteins rather than of the cell line that they are expressed in (Fig. 5c). Coexpression of E1/E2 with p7 does not appear to influence the processing and cell surface localization of the envelope glycoproteins.

Finally, cell surface-associated E1 and E2 were analyzed for their ability to form noncovalent heterodimers. HeLa cells stably expressing different combination of E1, E2, and p7 were preincubated with an anti-E2 MAb, and protein-antibody complexes were recovered by immunoprecipitation of cell lysates with G protein-coupled agorose beads. In this manner, only cell surface-associated envelope glycoproteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and immunoblotting with an anti-E1 MAb. E1 readily immunoprecipitated with E2 only in cells expressing E1*-E2*, and only if the cells were preincubated with an anti-E2 MAb (Fig. 6a). Similarly, E2 was detected in cells expressing E2*, E1*-E2*, or E1*-E2*-p7 only if the cells were preincubated with an anti-E2 MAb (Fig. 6b). Therefore, E1 and E2 proteins associated with the plasma membrane also form noncovalent heterodimers.

HCV envelope glycoproteins E1 and E2 have been described to form membrane-anchored, noncovalent heterodimers that are retained in the ER, where HCV budding is believed to occur (37). The signals for heterodimerization and ER retention colocalize to residues in the TM domains of E1 and E2, hence it seems the two functions cannot be dissociated.
This has made it difficult to generate cell surface-associated variants of E1/E2 heterodimers, which would be invaluable for the development of cell fusion assays and virus pseudotypes. Attempts to create such variants have thus far focused on fusing E1 and E2 ectodomains to the TM domains of vesicular stomatitis virus (VSV) G or influenza hemagglutinin envelope glycoproteins, which have no known dimerization function (20, 27, 55). Additionally, in all of these studies, chimeric E1 and E2 proteins were translated from separate mRNAs, which may have further minimized their potential to form native heterodimers. Even though E2-FA chimeras underwent pH-dependent conformational changes and were incorporated into influenza virus particles, they did not induce fusion with target cells (20). HCV-VSV chimeric envelope glycoproteins also did not appear to reproducibly model HCV fusion and entry (4, 27, 28, 33, 35, 55).

The initial goal of our study was to create chimeric HCV envelope glycoproteins that would be expressed on the cell surface as E1/E2 heterodimers. We therefore chose a strategy wherein the ectodomains of HCV E1 and E2 would be fused to the TM domains of E1 and E2 from a related alphavirus, SFV. Chimeric HCV-SFV envelope glycoproteins were expressed on the cell surface and resembled unmodified HCV envelope glycoproteins (37). The initial goal of our study was to create chimeric HCV envelope glycoproteins that would be expressed on the cell surface as E1/E2 heterodimers. We therefore chose a strategy wherein the ectodomains of HCV E1 and E2 would be fused to the TM domains of E1 and E2 from a related alphavirus, SFV. Chimeric HCV-SFV envelope glycoproteins were expressed on the cell surface and resembled unmodified HCV envelope glycoproteins (37). This has made it difficult to generate cell surface-associated variants of E1/E2 heterodimers, which would be invaluable for the development of cell fusion assays and virus pseudotypes. Attempts to create such variants have thus far focused on fusing E1 and E2 ectodomains to the TM domains of vesicular stomatitis virus (VSV) G or influenza hemagglutinin envelope glycoproteins, which have no known dimerization function (20, 27, 55). Additionally, in all of these studies, chimeric E1 and E2 proteins were translated from separate mRNAs, which may have further minimized their potential to form native heterodimers. Even though E2-FA chimeras underwent pH-dependent conformational changes and were incorporated into influenza virus particles, they did not induce fusion with target cells (20). HCV-VSV chimeric envelope glycoproteins also did not appear to reproducibly model HCV fusion and entry (4, 27, 28, 33, 35, 55).

The initial goal of our study was to create chimeric HCV envelope glycoproteins that would be expressed on the cell surface as E1/E2 heterodimers. We therefore chose a strategy wherein the ectodomains of HCV E1 and E2 would be fused to the TM domains of E1 and E2 from a related alphavirus, SFV. Chimeric HCV-SFV envelope glycoproteins were expressed on the cell surface and resembled unmodified HCV envelope glycoproteins (37). This has made it difficult to generate cell surface-associated variants of E1/E2 heterodimers, which would be invaluable for the development of cell fusion assays and virus pseudotypes. Attempts to create such variants have thus far focused on fusing E1 and E2 ectodomains to the TM domains of vesicular stomatitis virus (VSV) G or influenza hemagglutinin envelope glycoproteins, which have no known dimerization function (20, 27, 55). Additionally, in all of these studies, chimeric E1 and E2 proteins were translated from separate mRNAs, which may have further minimized their potential to form native heterodimers. Even though E2-FA chimeras underwent pH-dependent conformational changes and were incorporated into influenza virus particles, they did not induce fusion with target cells (20). HCV-VSV chimeric envelope glycoproteins also did not appear to reproducibly model HCV fusion and entry (4, 27, 28, 33, 35, 55).

The initial goal of our study was to create chimeric HCV envelope glycoproteins that would be expressed on the cell surface as E1/E2 heterodimers. We therefore chose a strategy wherein the ectodomains of HCV E1 and E2 would be fused to the TM domains of E1 and E2 from a related alphavirus, SFV. Chimeric HCV-SFV envelope glycoproteins were expressed on the cell surface and resembled unmodified HCV envelope glycoproteins (37). This has made it difficult to generate cell surface-associated variants of E1/E2 heterodimers, which would be invaluable for the development of cell fusion assays and virus pseudotypes. Attempts to create such variants have thus far focused on fusing E1 and E2 ectodomains to the TM domains of vesicular stomatitis virus (VSV) G or influenza hemagglutinin envelope glycoproteins, which have no known dimerization function (20, 27, 55). Additionally, in all of these studies, chimeric E1 and E2 proteins were translated from separate mRNAs, which may have further minimized their potential to form native heterodimers. Even though E2-FA chimeras underwent pH-dependent conformational changes and were incorporated into influenza virus particles, they did not induce fusion with target cells (20). HCV-VSV chimeric envelope glycoproteins also did not appear to reproducibly model HCV fusion and entry (4, 27, 28, 33, 35, 55).
coproteins in size and processing. The most surprising finding, and one that changed the focus of our studies, was the expression of unmodified HCV E1 and E2 on the cell surface. E2 was detected on the cell surface by flow cytometry with four different anti-E2 MAbs. Cell surface associated E2 expression was also detected in a hepatic cell line and was not influenced by the presence of p7. By biotin-tagging cell surface proteins we demonstrated that full-length E1 was also associated with the plasma membrane. Most importantly, we were able to specifically coimmunoprecipitate E1 protein with an anti-E2 MAb, thus demonstrating that cell surface-associated E1 and E2 form noncovalent heterodimers.

One of the complicating factors in identifying properly folded and functional E1 and E2 has been the host of expression systems used to study these proteins, and a careful survey of the literature reveals significant diversity in the number and size of protein species corresponding to E1 and E2. In this study, unmodified and chimeric envelope glycoproteins were generated by using two different expression systems. The use of vaccinia virus-based expression is justifiable on the premise that it circumvents the nucleus, just as HCV replication does. In this expression system, E1 and E2 remain intracellular. Vaccinia replication, however, modifies internal cellular membranes as well as the translation machinery (41, 44, 46, 48–50), and the apparent trapping of HCV envelope glycoproteins inside the cell may be an artifact. Indeed, vaccinia virus-based expression has been shown to cause ER retention of other viral envelope glycoproteins (50, 54). The observation that vaccinia virus-based expression generates hypoglycosylated E1 proteins led us to explore plasmid delivery of HCV envelope glycoproteins. Plasmid-based expression of proteins typically does not adversely affect cellular protein synthesis but does involve nuclear transcription, which is not a natural part of HCV replication. Indeed, we clearly show that plasmid-based expression of HCV envelope glycoproteins results in putative intron excision in E1 mRNA that translates into a protein with a deletion. Other modifications such as mRNA editing were not observed (unpublished results). Our finding highlights an inherent complication in expressing RNA virus proteins by DNA-based expression systems.

Recently it was reported that HCV envelope glycoproteins are able to pseudotype retroviral particles and mediate their entry into target cells (3, 22). Both studies used plasmid vectors to express unmodified E1/E2, meaning that pseudoviral envelopes contained both full-length proteins and those with E1 deleted. We have confirmed that unmodified HCV envelope glycoproteins are able to mediate entry of retroviral pseudotypes into several hepatic and nonhepatic cell lines as well as primary hepatocytes (unpublished results). We are presently determining how the presence of partially deleted E1 species in pseudoviral envelopes affects entry into different target cells. These studies will allow us to further optimize pseudovirion entry mediated by HCV envelope glycoproteins, which will facilitate structure/function studies of HCV envelope glycoproteins as well as the identification of HCV receptors and target cells.

It remains to be determined whether cell surface-associated E1/E2 heterodimers have any physiological relevance in the viral replication cycle. Our observation that HCV envelope glycoproteins are expressed on the surface of cells that closely resemble primary hepatocytes implies that there is no specific retention mechanism for HCV envelope glycoproteins in liver cells. The postulated HCV replication cycle is based on analogies to the closely related flaviviruses and pestiviruses and it is generally assumed that flaviviruses bud into the ER and mature by passage into cytoplasmic vesicles (42). Thus far the cellular localization of HCV envelope glycoproteins and particles has mostly been studied in cells transfected or infected in vitro. Virus-like particles mostly occurred in cytoplasmic vesicles, suggesting vesicle-based morphogenesis of HCV (12, 18,
FIG. 6. Heterodimerization of E1 and E2 on the cell surface. Intact HeLa cells stably expressing modified HCV envelope glycoproteins were incubated with the anti-E2 MAB H53 (1:100, lyed, and incubated with protein G-coupled agarose beads overnight at 4°C (Oncogene Research Products) followed by three washes with the lysis buffer. As a control, lysates from cells that had not been treated with H53 were used. The presence of E1 was detected by immunoblotting with anti-E1 MAB A4 (a), whereas the presence of E2 was detected by immunoblotting with anti-E2 MAB A11 (b). Arrowheads indicate the positions of full-length E1* and E2* proteins. M, molecular size markers; IP, immunoprecipitation; IgG, immunoglobulin G.

21, 23, 43, 52, 53). No study, however, has clearly documented the HCV budding and maturation process, perhaps because it does not occur in these experimental systems or perhaps because it is an extremely rare event that is difficult to detect by standard methods. We are presently addressing these questions by expressing E1/E2 envelope glycoproteins in human primary hepatocytes.

We are especially thankful to Jean Dubuisson (Institut Pasteur, Lilles, France) for repeatedly and generously providing MABs A4, A11, H2, H52, H53, and H60. We also thank Charles Rice (Rockefeller University, New York, N.Y.) for providing the p90/HCV FL-A11, H2, H52, H53, and H60. We also thank Charles Rice (Rockefeller University, New York, N.Y.) for providing the p90/HCV FL-long clone and Margaret Kielian (Albert Einstein College of Medicine, Bronx, N.Y.) for the pSP6SFV-4 clone.

This work was supported by The Speaker’s Fund for Biomedical Research Award to Young Investigators to T.D., by grant AI051134 to Fordham, N.Y.) for the pSP6SFV-4 clone.

REFERENCES


20. Flint, M., J. M. Thomas, C. M. Maidens, C. Shotton, S. Levy, W. S. Barclay,


