Construction and characterization of chimeric hepatitis C virus E2 glycoproteins: analysis of regions critical for glycoprotein aggregation and CD81 binding

Arvind H. Patel, Jonny Wood, Francois Penin, Jean Dubuisson and J. A. McKeating

1 MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5IR, UK
2 Institut de Biologie et Chimie des Protéines, UPR 412 CNRS, 7 Passage du Vercors, F-69367 Lyon Cedex 07, France
3 CNRS-UMR8526, IBL/Institut Pasteur de Lille, 59021 Lille Cedex, France
4 University of Reading, School of Animal & Microbial Sciences, PO Box 228, Reading, UK

We compared the ability of two closely related truncated E2 glycoproteins (E2\textsubscript{660}) derived from hepatitis C virus (HCV) genotype 1a strains Glasgow (Gla) and H77c to bind a panel of conformation-dependent monoclonal antibodies (MAbs) and CD81. In contrast to H77c, Gla E2\textsubscript{660} formed disulfide-linked high molecular mass aggregates and failed to react with conformation-dependent MAbs and CD81. To delineate amino acid (aa) regions associated with protein aggregation and CD81 binding, several Gla–H77c E2\textsubscript{660} chimeric glycoproteins were constructed. Chimeras C1, C2 and C6, carrying aa 525–660 of Gla E2\textsubscript{660}, produced disulfide-linked aggregates and failed to bind CD81 and conformation-dependent MAbs, suggesting that amino acids within this region are responsible for protein misfolding. The presence of Gla hypervariable region 1 (aa 384–406) on H77 c E2\textsubscript{660}, chimera C4, had no effect on protein folding or CD81 binding. Chimeras C3 and C5, carrying aa 384–524 or 407–524 of Gla E2\textsubscript{660}, respectively, were recognized by conformation-dependent MAbs and yet failed to bind CD81, suggesting that amino acids in region 407–524 are important in modulating CD81 interaction without affecting antigen folding. Comparison of Gla and H77c E2\textsubscript{660} aa sequences with those of genotype 1a and divergent genotypes identified a number of variant amino acids, including two putative N-linked glycosylation sites at positions 476 and 532. However, introduction of G476N–G478S and/or D532N in Gla E2\textsubscript{660} had no effect on antigenicity or aggregation.

Introduction

Hepatitis C virus (HCV) is the major aetiological agent of post-transfusion non-A, non-B hepatitis. Approximately 170 million people worldwide are infected with HCV (Lavanchy et al., 1999). In a large number of individuals, infection leads to virus persistence and chronic disease, often resulting in liver cirrhosis and hepatocellular carcinoma (Houghton, 1996). There is no vaccine or effective antiviral treatment available.

HCV, a member of the \textit{Flaviviridae} family, is an enveloped virus with a positive-strand RNA genome of approximately 9.5 kb (Choo et al., 1989). The viral genome encodes a single polyprotein of approximately 3010 amino acids (aa) that is processed into functional proteins by host and viral proteases (Houghton, 1996; Ryan et al., 1998). The putative HCV structural proteins [core and the two envelope glycoproteins (gp), E1 and E2] are located within the N terminus of the polyprotein, whilst the non-structural proteins reside within the C-terminal part (Clarke, 1997). Gps E1 and E2, when expressed \textit{in vitro}, associate to form two types of complexes: a heterodimer stabilized by non-covalent interactions and high molecular mass disulfide-linked aggregates representing misfolded proteins (Choukhi et al., 1998; Coquerel et al., 1998; Deleersnyder et al., 1997; Dubuisson et al., 1994; Dubuisson & Rice, 1996; Ralston et al., 1993). Both types of complex accumulate within the endoplasmic reticulum (ER), the proposed site for HCV assembly and budding. Due to the lack of a suitable cell culture system for \textit{in vitro} propagation of HCV, it has been impossible to study the nature of the gp
complex present on the virus particle. However, studies using conformation-dependent monoclonal antibodies (MAbs) which specifically recognize nondisulfide-linked E2, either alone or when complexed with E1, strongly suggest that the gp complex stabilized by non-covalent interactions represents the pre-budding form of E1E2 native heterodimers (Deleersnyder et al., 1997; Michalak et al., 1997). However disulfide-bridged gp aggregates may have a role in vivo; indeed, Liberman et al. (1999) reported that ER-retained E2 could activate grp78 (BiP) and grp94 chaperone promoters. Overexpression of grp78 has been reported to decrease the sensitivity of cells to cytotoxic T cell killing, an activity that may be important for HCV persistence. Further work is needed to define possible roles for aggregated E2 in HCV pathogenesis.

HCV gps are thought to initiate infection of the target cells by binding to receptors on the cell plasma membrane followed by membrane fusion and entry. Although the mechanism of HCV entry into cells is unknown, the E2 gp is thought to play a major role in virus attachment to the target cell (Rosa et al., 1996). The E2 gp extends from aa 384–745 of the polyprotein and it carries regions of extreme hypervariability (Kato et al., 1996; Mizushima et al., 1994; Ogata et al., 1991; Weiner et al., 1991). The most variable region (HVR-1) is located within the N-terminal 27 residues (aa 384–411) of E2, while HVR-2 resides in the 476–480 segment. Antibodies specific for epitopes within HVR-1 have been reported to inhibit binding of E2 gp to cells and to block HCV infectivity in vitro and in vivo (Farcì et al., 1996; Habersetzer et al., 1998; Rosa et al., 1996; Shimizu et al., 1996; Zibert et al., 1995).

The E2 gp carries a C-terminal hydrophobic anchor sequence (aa 718–746), removal of which results in secretion of the E2 ectodomain (Matsuura et al., 1994; Michalak et al., 1997; Mizushima et al., 1994; Selby et al., 1994; Spaete et al., 1992). Thus, E2 truncated at either aa 715 and 661 is secreted upon expression in mammalian cells, but only the latter was found to be properly folded (Michalak et al., 1997). A secreted form of E2 was shown to bind to the surface of Molt-4 cells, and this binding could be blocked by sera from HCV-infected chimpanzees (Rosa et al., 1996). Furthermore, in chimpanzees vaccinated with recombinant E1E2 gps, protection from challenge virus correlated with the presence of antibodies capable of inhibiting E2 gp binding to cells. The sera were classified as being able to neutralize E2 binding (NOB), leading the authors to postulate that such activity may be used as a surrogate marker for neutralizing antibodies (Rosa et al., 1996).

Using a truncated form of E2, Pileri et al. (1998) recently identified the cell surface protein, CD81, as a putative receptor for HCV. E2 truncated at aa position 661 (E2_{661}) binds specifically to human but not mouse or African green monkey CD81 (Flint et al., 1999; Higginbottom et al., 2000; Pileri et al., 1998). CD81 is a member of the tetraspan spin superfamily of proteins; it has four transmembrane domains and two extracellular loops (EC1 and EC2) (Levy et al., 1998), the second of which has been shown to bind a truncated form of E2 (Pileri et al., 1998). MAbs specific for CD81 have been reported to have a variety of effects on cellular processes including proliferation, adhesion and motility (Levy et al., 1998). Interestingly, we recently demonstrated that E2 engagement of CD81 also influenced cell proliferation and aggregation (Flint et al., 1999).

There are at least six genotypes of HCV, which differ from each other by 30% over the complete virus genome, with 2–10% variation between subtypes of the same genotype (Simmonds, 1995; Simmonds et al., 1994). In this study, we have used two closely related genotype 1a strains of HCV, namely Glasgow (Gla), cloned directly from an individual infected with a genotype 1a virus (M. McElwee & R. M. Elliott; unpublished), and an infectious cDNA clone of strain H77c (Yanagi et al., 1997). We compared the antigenic and CD81-binding characteristics of E2_{660} derived from strains Gla and H77c. We show that in contrast to H77c E2_{660} Gla protein is predominantly synthesized as misfolded disulfide-linked aggregates and fails to bind CD81. Characterization of several chimeric Gla–H77c E2_{660} gps allowed us to identify regions of E2 important for protein aggregation/folding and for interacting with the putative receptor, CD81.

**Methods**

- **Cell culture.** BHK cells were cultured in Glasgow minimal essential Eagle’s medium supplemented with 10% newborn calf serum, 4% tryptose phosphate broth and penicillin–streptomycin. COS-7 cells were grown in Dulbecco's modified Eagle's medium (GIBCO BRL) supplemented with 10% foetal calf serum, 5% non-essential amino acids and penicillin–streptomycin (EFC10).

- **Plasmid constructs.** The E2-encoding cDNA sequences used in this study were derived from two HCV genotype 1a strains, Glasgow (Gla; kindly provided by M. McElwee and R. M. Elliott) and an infectious cDNA clone (pCV-H77c) of strain H77c (a kind gift from Jens Buhk (Yanagi et al., 1997)). The nucleotide sequences encoding truncated Gla and H77c E2 from aa 364–660 followed by 6 × His residues (E2_{660}) were PCR amplified and cloned into mammalian expression vector pCDNA3.1/Zeo+ (Invitrogen) and the vaccinia virus transfer vector pMJ601 (Davison & Moss, 1990). The Gla–H77c chimeric clones (C1–C6) were constructed into pCDNA3.1/Zeo+ as shown in Fig. 3(A). The E2 sequences expressed in all constructs encode the E2 signal sequence (aa 364–383), which is cleaved during protein translocation, leaving a mature protein carrying aa 384–660 of E2 followed by 6 × His (E2_{660}). Nucleotide sequences of all constructs were determined to confirm the origin of inserts. The Gla and H77c E2_{660} cDNA in pMJ601 was inserted into the thymidine kinase gene of vaccinia virus strain WR by homologous recombination and the recombinant viruses isolated as previously described (Davison & Moss, 1990).

- **Site-directed mutagenesis of E2.** The GeneEditor in vitro site-directed mutagenesis system (Promega) was used to generate glycosylation site mutants of Gla E2_{660}. The plasmid encoding Gla E2_{660} was used as a template together with the primer 5′ ATG GTC GGG GCC GCT TCC GTT GGC ATA ACT GAT 3′ or 5′ GAC GTC CGT ATC ATT TGC ACC CCA GCT 3′, or both (mutated nucleotides are shown in bold), to convert GGA at aa position 476–478 to NGS (called
Expression of parental and chimeric E2<sub>660</sub>. COS-7 cells were transfected with pcDNA3.1/Zeol (+) alone or with pcDNA3.1/Zeol (+) carrying E2<sub>660</sub> sequences using a liposome-mediated method (Rose et al., 1991) for 5 h at 37 °C. Cells were washed and incubated in EFC10 for 72 h, after which the extracellular medium was collected and clarified by centrifugation at 3000 r.p.m. for 10 min at 4 °C. For radiolabelling of proteins, the transfected COS-7 cells were incubated at 37 °C for 24 h, washed with PBS, and incubated in methionine-free medium containing 50 µCi/ml <sup>35</sup>S-methionine for a further 24 h. The medium of transfected cells was harvested and clarified as described above. To radiolabel E2<sub>660</sub> expressed by recombinant vaccinia virus, BHK cells were incubated at a multiplicity of 10 p.f.u. per cell and incubated for 5 h at 37 °C. Infected cells were washed with PBS, incubated in methionine-free medium containing 50 µCi/ml <sup>35</sup>S-methionine for 18 h and the medium harvested. The E2<sub>660</sub> secreted into the medium of transfected or infected cells was quantified using an ELISA-based GNA lectin capture assay as reported previously (Flint et al., 2000) and described briefly below.

Affinity purification of E2<sub>660</sub>. The medium of transfected or infected cells containing <sup>35</sup>S-methionine-labelled proteins was equilibrated to 20 mM Tris–HCl, pH 7.4, 20 mM iodoacetamide, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100 (final concentrations) and incubated with Ni–NTA (Qiagen) resin for 1 h at 4 °C. The resin was pelleted for 30 s in a microcentrifuge and washed four times with the above buffer. The bound proteins were eluted in wash buffer containing 100 mM EDTA and analysed by SDS–PAGE under reducing and non-reducing conditions.

Antibodies. MAb s AP213 and AP320 were generated in mice immunized with a recombinant form of HCV strain Gla E1E2 expressed in mammalian cells and were epitope mapped by PEPSCAN (unpublished). MAb s H31, H44, H50, H60, H61, H53 and H33 were generated as previously described (Deleersnyder et al., 1997). The antibody × His MAb (6His) was purchased from Qiagen and used according to the manufacturer’s instructions.

Flow cytometric analysis of E2-cell binding and NOB. The interaction of E2<sub>660</sub> with cells was quantified using a FACS-based assay. In brief, cells under test were washed twice in PBS–1% FCS–0.05% sodium azide (WB) and resuspended at 2 x 10<sup>6</sup>/ml; 2 x 10<sup>5</sup> cells were incubated with E2 at room temperature for 1 h and unbound antigen was removed by washing twice in WB. Cells were incubated with MAb 6His (10 µg/ml) for 1 h at room temperature. Finally, cell-bound MAb was visualized with an anti-mouse IgG–PE conjugate (Seralabs) and analysed by FACS (Becton Dickinson). Median fluorescence intensities (MFI) were determined using Cellquest software (Becton Dickinson).

MAbs were evaluated for their ability to inhibit E2 binding to cells as previously reported (Flint et al., 1999). Briefly, E2 (5 µg/ml) was incubated with the MAb under test (10 µg/ml) for 1 h at room temperature and antigen–antibody complex formation was then verified by GNA-capture EIA. Complexes were allowed to bind to cells for 1 h at room temperature, unbound complex was removed by washing twice with WB, and the cell-bound complex was visualized with an anti-species IgG–PE conjugate (Seralabs). MFI were determined using Cellquest software and the percentage inhibition determined.

GNA-capture EIA and CD81-capture EIA. Briefly, GNA (Galanthea niveola) lectin (Boehringer Mannheim) was used to coat Immulon II EIA plates (Dynal) at 1 µg/ml overnight at 4 °C. After washing in Tris-buffered saline, the plates were blocked with 4% milk powder (Cadbury’s) and E2 allowed to bind for 2 h at room temperature. Bound antigen was visualized with MAb s specific for E2, an anti-species IgG–HRP (Seralabs) and TMB (3,3’5,5’-tetramethylbenzidine; Sigma) substrate. Absorbance values were determined at 450 nm (Dynatech). Purified E2 (gift of M. Houghton, Chiron) was used as a calibrator to enable quantification of E2 levels in transient transfection samples.

GST–hCD81 and –mCD81 fusion proteins expressing the second extracellular region (EC2) of human or mouse CD81, respectively, were used to coat Immulon II EIA plates (Dynal) at 0.5 µg/ml for 4 h at 37 °C as previously described (Flint et al., 1999, 2000). The ability of E2 to specifically bind hCD81 was followed as above for the GNA EIA, using MAb s H53 or 6His to detect CD81-bound E2.

SDS–PAGE analysis. The <sup>35</sup>S-methionine-labelled proteins eluted from Ni–NTA resins were mixed with SDS–PAGE denaturant buffer (200 mM Tris–HCl, pH 6.7; 0.5% SDS, 10% glycerol) and boiled for 3 min in the presence (reducing conditions) or absence (non-reducing conditions) of β-mercaptoethanol. Samples were subjected to SDS–PAGE (10% polyacrylamide) and the labelled proteins detected by fluorography.

Sequence analysis. The protein sequence of the E2 (aa 410–660) segment of H77c strain (EMBL accession no. AF011751) was used to search all E2 HCV variants in the EMBL database using the FASTA homology program (Pearson & Lipman, 1988). Incomplete sequences were removed from the list of matching sequences and multiple sequence alignment carried out with the CLUSTAL W program (Thompson et al., 1994). For polymorphic analysis, sequences of isolates known to belong to genotype 1a (i.e. HCV-1, HCV-J1, HCV-H) were used to search the EMBL database. The 1a subtype determination was confirmed by following the recommended standard procedure (Robertson et al., 1998) and using the representative genotype and subtype sequences reported (Chamberlain et al., 1997). The EMBL accession numbers for the final set of 11 sequences classified in genotype 1a was as follows: M62382, D10749, M62381, X84079, M63232, M67463, G252549, AF011751, AF011752, AF011753 and AF009606. Sequence analyses were made using the Network Protein Sequence @analysis facilities (NPS@ ; Combet et al., 2000) through the IBCP server (http://ibip.ibcp.fr/NPSA). Visualization and calculation of most represented amino acid at each position were done with the MPSA program (Blanchet et al., 2000).

Results

Characterization of Gla and H77c E2<sub>660</sub>. Truncated soluble forms of Gla and H77c E2<sub>660</sub> were expressed at comparable levels (data not shown) and purified preparations were analysed by SDS–PAGE under reducing and non-reducing conditions (Fig. 1). Reducing conditions showed both proteins to migrate as monomers of expected molecular mass. The relative faster mobility of Gla E2<sub>660</sub> is likely due to differences in glycosylation arising as a result of the absence of two putative N-linked glycosylation sites (see below). Non-reducing SDS–PAGE demonstrated that a significant amount of H77c E2<sub>660</sub> migrated as a monomeric (native) form, but higher molecular mass products (aggregates) were also seen (Fig. 1). In contrast, the majority of Gla E2<sub>660</sub> migrated as high molecular mass aggregates suggesting that Gla E2 gp may misfold (Fig. 1).

Strain H E2, which is closely related to strain H77c (Kolykhov et al., 1996), has been shown to bind specifically to human CD81 (hCD81) on the cell surface (Flint et al.,
A. H. Patel and others

Fig. 1. SDS–PAGE analysis of soluble Gla and H77c E2660. BHK cells were infected with the wild-type (wt) vaccinia virus strain WR, or recombinant vaccinia viruses expressing Gla or H77 E2660, in the presence of [35S]methionine. The radiolabelled proteins from the medium were purified, subjected to 10% SDS–PAGE under reducing and non-reducing conditions and detected by fluorography. The positions of the aggregates (Agg) and native E2660, and of protein size markers (kDa), are shown.

To test whether Gla and H77c E2660 can bind CD81, we performed flow-cytometric analysis using a rat granulocyte cell line, RBL, stably expressing hCD81 (Flint et al., 1999). The H77c E2660 bound specifically to RBL cells expressing hCD81 (giving a 3–5-fold increase in MFI binding to RBL-CD81 compared to RBL alone), whereas Gla E2660 failed to interact with CD81 (1–2-fold increase in MFI between RBL-CD81 and RBL) (data not shown).

To further quantify the E2–CD81 interaction using a more sensitive assay, we studied the ability of Gla and H77c E2660 to interact with a recombinant form of human CD81, GST–hCD81, using an ELISA-based binding assay (Flint et al., 1999, 2000). Gla and H77c E2660 gps were tested for their ability to bind GNA lectin (Fig. 2A), GST–murineCD81 (mCD81) or GST–hCD81 (Fig. 2B); bound antigen was detected using MAb 6His and anti-mouse IgG–HRP. Consistent with our cell-binding data, H77c E2660 bound specifically to GST–hCD81 and not GST–mCD81, whereas Gla E2660 failed to interact with either of the CD81 proteins (Fig. 2B). Similar results were obtained using a number of linear and conformational anti-E2 MAbs to detect the CD81-captured E2660 gps (data not shown).

Antigenic characterization of Gla–H77c E2660 gp chimeras

Since Gla and H77c E2660 proteins display such distinct antigenic phenotypes and yet share 88.8% aa identity (see Fig. 5) we constructed a number of chimeric gps to help delineate the regions affecting E2 aggregation and CD81 binding (Fig. 3A). BHK cells expressing H77 E2660 were incubated with ELISA plates coated with (A) GNA, or (B) GST–hCD81 or GST–mCD81, as described in Methods. Bound E2 was detected using MAb 6His and anti-mouse IgG–HRP.
HCV E2 aggregation and CD81 binding

Fig. 3. Construction and characterization of Gla and H77c E2\textsubscript{660} chimeras. (A) The structure of cDNA encoding full-length E2 (aa 384–746) is shown at the top. Amino acids 364–383 and 384–414 represent the secretory signal sequence and HVR-1, respectively. The E2 transmembrane domain sequence (aa 718–746) is shown. (B) Schematic representation of HCV strains Gla and H77c E2\textsubscript{660} and their chimeras (C3–C6). The chimeric proteins were generated by swapping aa regions 384–406 (containing most of HVR-1), 406–524, and 524–660 of Gla and H77c E2\textsubscript{660} in various combinations as indicated. All constructs have a C-terminal 6\textsuperscript{His} tag (dark box). A summary of MAb and CD81 reactivity of the E2\textsubscript{660} proteins is shown, where + indicates the relative affinities of interaction. (C) SDS–PAGE analysis of chimeric E2\textsubscript{660} gps. COS-7 cells were transfected with pCDNA3.1Zeo\textsuperscript{[a]} (control, lane C) or pCDNA3.1Zeo\textsuperscript{[b]} expressing Gla, H77c or the chimeric E2\textsubscript{660} and the cells allowed to express proteins in the presence of \[^{35}\text{S}\]methionine. The labelled proteins secreted into the medium of transfected cells were Ni–NTA purified, subjected to 10\% SDS–PAGE under reducing and non-reducing conditions, and detected by fluorography. The positions of the aggregates (Agg) and native E2\textsubscript{660} and of protein size markers (kDa), are shown.

performed for all proteins – data not shown). Similarly, a MAb 6His (specific for the C-terminal 6\times His tag) also recognized all the gps (data not shown). MAb AP213, specific for Gla HVR-1, bound chimeras C3, C4 and C6 as well as Gla E2\textsubscript{660}, as expected since all of these gps contain Gla HVR-1 sequence (Table 1, Fig. 3B). In contrast, the conformation-dependent MAbs H31, H44, H50, H60, H61, H53 and H33, which are specific for epitopes independent of the HVR, were able to recognize H77c E2 but not Gla E2\textsubscript{660} gp (Table 1). Substitution of aa regions 407–660 and 525–660 of the H77c E2\textsubscript{660} with the corresponding domains from Gla (chimeras C1 and C2, respectively; Fig. 3B) resulted in proteins that were not recognized by any conformation-dependent MAbs (Table 1). Similarly, substitution of aa 406–524 in Gla E2\textsubscript{660} with the corresponding region of H77c gp (chimera C6) generated a protein not recognized by the conformational MAbs tested. Conversely, when aa regions 407–524 of H77c E2\textsubscript{660} (chimera C5), or 525–660 or 407–660 of Gla E2\textsubscript{660} (chimeras C3 and C4, respectively) were swapped for the appropriate Gla or H77c sequences, the proteins were recognized by all of the MAbs. It is interesting to note that MAb H60 recognized chimeric gps C3, C4 and C5 with reduced affinities compared to H77c E2\textsubscript{660} (Table 1; dose-response curves were performed – data not shown). These results are consistent with the data shown in.
**Table 1. Reactivity of the Gla-H77c E2\textsubscript{660} chimeras with MAbs specific for linear and conformation-dependent epitopes**

Culture medium of COS-7 cells transfected with constructs shown in Fig. 3 was analysed by GNA-capture ELISA using various MAbs as described in Methods. Readings in bold and italic represent MAb reactivities of high and low affinity, respectively. NOB, neutralization of E2\textsubscript{660} binding to CD81 (see Methods).

<table>
<thead>
<tr>
<th>E2\textsubscript{660} Epitope</th>
<th>Binding of MAbs to antigens (A\textsubscript{660})</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP213</td>
<td>Not tested</td>
</tr>
<tr>
<td>H77c</td>
<td>0.05</td>
</tr>
<tr>
<td>Gla</td>
<td>2.00</td>
</tr>
<tr>
<td>C1</td>
<td>0.04</td>
</tr>
<tr>
<td>C2</td>
<td>0.04</td>
</tr>
<tr>
<td>C3</td>
<td>1.97</td>
</tr>
<tr>
<td>C4</td>
<td>2.08</td>
</tr>
<tr>
<td>C5</td>
<td>0.05</td>
</tr>
<tr>
<td>C6</td>
<td>2.35</td>
</tr>
</tbody>
</table>

**Fig. 4.** Interaction of chimeric H77c-Gla gps with CD81. The secreted proteins from transfected COS-7 cells were incubated with GNA, GST–hCD81 or GST–mCD81 and bound proteins detected with MAb 6His, as described in Methods.

Figs 1 and 3 (C), suggesting that Gla E2\textsubscript{660} and chimeras C1, C2 and C6 misfold. Furthermore, they suggest that variant amino acids within region 525–660 of Gla are responsible for the misfolding. Since the epitopes recognized by our conformation-dependent MAbs are unknown we cannot exclude the possibility that the lack of recognition by these MAbs is due to amino acid variation within Gla region 525–660. However, this is unlikely since we tested a large panel of MAbs which do not cross-compete with each other for E2 recognition and display different NOB activities (data not shown; Table 1).

The chimeric glycoproteins were tested for their ability to interact with GST–hCD81. Saturating amounts of the glycoproteins were captured by GNA lectin, GST–hCD81 or GST–mCD81 (Fig. 4). H77c E2\textsubscript{660} and chimera C4 bound CD81 with similar relative affinities (Fig. 4), whereas Gla and the remaining chimeras failed to interact with GST–hCD81 (Fig. 4). Since chimeras C1, C2 and C6 misfold, it was not surprising to observe their negligible binding to CD81. However, chimeras C3, C4 and C5 all fold correctly (Fig. 3 B, Table 1) and yet C3 and C5 fail to show detectable levels of CD81 binding. Chimeras C3 and C5 both encode aa region 407–524 of Gla, suggesting that residues within this region may be critical for interacting with CD81.

**Effect of aa changes at positions 476, 478 and 532 of E2 Gla on protein aggregation and CD81 binding**

Studies on the chimeric proteins suggested that aa regions 407–524 and 525–660 are important in determining E2 interaction with CD81 and ‘correct’ protein folding and aggregation, respectively. Amino acid sequence comparison between Gla and H77c in region 407–524 shows 15 residue changes (Fig. 5), five of which are unlikely to have any effect on CD81 recognition since they are either conservative hydrophobic polymorphic residues, not likely to be exposed at the surface of E2 (positions 411, 422 and 438), or conservative charged residues (positions 431 and 492) (Fig. 6). Of the remaining aa changes, six are polymorphic in genotype 1a (positions 424, 444, 446, 466, 480 and 482). In addition, aa 453 could also be considered polymorphic due to the functional similarity between Ser and Thr residues (Fig. 6). It is reasonable therefore to assume that such polymorphic changes are not involved in CD81 recognition. Although not observed in genotype 1a sequences analysed to date, the R461S change was observed in several divergent sequences and hence its
HCV E2 aggregation and CD81 binding

Fig. 5. Comparison of Gla and H77c aa 384–660. The sequence alignment was performed using GCG FASTA (Pearson & Lipman, 1988). Cysteine residues are indicated by an asterisk, and the predicted N-linked glycosylation sites (NXS/T) by shaded boxes. The HVR-1 and HVR-2 regions are in open boxes.

Involvement in CD81 binding is questionable. The two remaining changes, N476G and S478G, abolish a conservative putative glycosylation site (Figs 5 and 6). We therefore mutated aa 476 and 478 within Gla E2460 to the residues found in H77, GGG→NGS (G476N–G478S), and analysed the protein for CD81 binding and aggregation. Non-reducing SDS–PAGE analysis of Gla G476N–G478S E2 showed the gp to migrate as a high molecular mass aggregate (Fig. 7) which failed to show detectable binding to CD81 (data not shown).

Similar aa sequence comparison between Gla and H77c in region 525–660 shows eight changes in the Gla sequence (Figs 5 and 6). Comparison of this region in strains Gla and H77c to those of 11 genotype 1a and/or 140 cross-genotype sequences showed that five of the eight variant amino acids are conservative and polymorphic (positions 603, 608, 617, 625 and 641; Fig. 6) and unlikely to be responsible for gp folding. Although the polymorphism at position 580 suggests minimal effect(s) on protein folding, such a non-conservative change (L580H) may modify the accessibility of the adjacent cysteine at position 581 which in turn may affect its ability to form aberrant disulfide bridges. The two remaining changes, N532D and Y624H, are unique to Gla and may be involved in the protein misfolding process. Since glycosylation is known to contribute to both protein solubility and stability (Gahmberg & Tolvanen, 1996), we mutated aa 532 within Gla E2 to that found in H77, D→N, both alone and in combination with G476N–G478S mutation (Fig. 7). Non-reducing SDS–PAGE analysis of Gla D532N and Gla D532N/G476N–G478S showed both gps to migrate predominantly as high molecular mass aggregates (Fig. 7). Under non-reducing conditions, a small proportion of the double mutant was produced in a monomeric native form which migrated more slowly, consistent with the predicted increase in molecular mass (Fig. 7). Not surprisingly, both Gla D532N and Gla D532N/G476N–G478S E2 gps failed to show detectable binding to CD81 (data not shown).

Discussion

In the absence of a suitable HCV-permissive cell culture system and patient samples containing high levels of virus particles, it is difficult to assess the nature of the gp complexes present on the virion surface and their role in cell attachment and entry. Studies using both human and rodent conformation-dependent MAbs to E2 have been useful in characterizing the native complex of E1E2, which has been proposed to be the major viral envelope component (Deleersnyder et al., 1997; Michalak et al., 1997). In a previous report, it was shown that the full-length E1E2 glycoproteins encoded by HCV genotype 1a strain Gla form misfolded, disulfide-bridged aggregated complexes (Patel et al., 1999). In contrast, E1E2...
glycoproteins derived from an infectious clone of genotype 1a strain H77c produce both native and aggregated complexes. A secretory form of strain H E2, truncated at aa 661, has been shown to fold in a manner comparable to the full-length gp and has been used as a soluble mimic for the native gp complex (Flint et al., 1999; Michalak et al., 1997). In this paper, we analysed and compared the antigenic characteristics of E2 from strains Gla and H77c and showed that Gla E2 formed predominantly misfolded disulfide-bonded aggregates which failed to bind CD81, the putative receptor for HCV (Pileri et al., 1998). We were interested in delineating the regions of E2 affecting aggregation and CD81 interaction and since Gla and H77c E2 differ in these properties, they are ideal candidates for the generation of chimeric proteins for analysing structure–function relationships.

Several chimeric Gla–H77c E2 gps were generated: all chimeras carrying the aa 525–660 region of Gla E2 were produced as high molecular mass aggregates and failed to be recognized by a panel of MAbs. Similar results were obtained with a number of conformation-sensitive E2 specific human MAbs (data not shown). Taken together, these data suggest that amino acids within region 525–660 may be important in

### A. Region 406-524

<table>
<thead>
<tr>
<th>Position</th>
<th>411</th>
<th>422</th>
<th>424</th>
<th>431</th>
<th>438</th>
<th>444</th>
<th>446</th>
<th>453</th>
<th>461</th>
<th>466</th>
<th>476</th>
<th>478</th>
<th>480</th>
<th>482</th>
<th>492</th>
</tr>
</thead>
<tbody>
<tr>
<td>H77c</td>
<td>I</td>
<td>I</td>
<td>S</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>K</td>
<td>P</td>
<td>R</td>
<td>A</td>
<td>N</td>
<td>S</td>
<td>L</td>
<td>E</td>
<td>R</td>
</tr>
<tr>
<td>Gla</td>
<td>V</td>
<td>V</td>
<td>R</td>
<td>D</td>
<td>I</td>
<td>Y</td>
<td>G</td>
<td>S</td>
<td>S</td>
<td>D</td>
<td>G</td>
<td>G</td>
<td>P</td>
<td>H</td>
<td>K</td>
</tr>
<tr>
<td>Polymorphism</td>
<td>I</td>
<td>I</td>
<td>S</td>
<td>E</td>
<td>L</td>
<td>Q</td>
<td>K</td>
<td>P</td>
<td>R</td>
<td>A</td>
<td>N</td>
<td>S</td>
<td>L</td>
<td>E</td>
<td>R</td>
</tr>
<tr>
<td>(Genotype 1a)</td>
<td>L</td>
<td>R</td>
<td>D</td>
<td>I</td>
<td>H</td>
<td>G</td>
<td>T</td>
<td>P</td>
<td>D</td>
<td>T</td>
<td>P</td>
<td>H</td>
<td>K</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### B. Region 525-660

<table>
<thead>
<tr>
<th>Position</th>
<th>532</th>
<th>580</th>
<th>603</th>
<th>608</th>
<th>617</th>
<th>624</th>
<th>625</th>
<th>641</th>
</tr>
</thead>
<tbody>
<tr>
<td>H77c</td>
<td>N</td>
<td>L</td>
<td>I</td>
<td>M</td>
<td>H</td>
<td>Y</td>
<td>T</td>
<td>E</td>
</tr>
<tr>
<td>Gla</td>
<td>D</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>R</td>
<td>H</td>
<td>S</td>
<td>D</td>
</tr>
<tr>
<td>Polymorphism</td>
<td>D</td>
<td>L</td>
<td>I</td>
<td>M</td>
<td>H</td>
<td>Y</td>
<td>T</td>
<td>E</td>
</tr>
<tr>
<td>(Genotype 1a)</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>D</td>
<td>R</td>
<td>Q</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Polymorphism | N | T | L | L | H | F | T | N |
| (all genotypes) | I | M | Y | S | S |
| L | I | L | E | F |
| H | K | V | Y | R |

Fig. 6. Analysis of polymorphic residues between H77c and Gla E2. Polymorphic residues in regions 406–524 (A) and 525–660 (B) between H77c, Gla, genotype 1a variants (11 sequences) and all genotype sequences available (140 sequences, including H77c and Gla). Comparison to the polymorphism observed for natural variants of genotype 1a (11 sequences) and for all genotypes (140 sequences, including that of H77c and Gla). The observed residues at each position are indicated in decreasing order of frequency from top to bottom. The less frequently observed residues (< 5%) are indicated in italic. Residues that were observed only once at a given position were not reported because they can be due to PCR and/or sequencing error and/or sequencing of defective virus. #. Residues involved in HVR-2; *, polymorphic residues were not reported because of ambiguities in multi-sequence alignment due to gaps or additional residues in this region. Particular positions are indicated in bold (see explanations in the text).
determining the ‘correct’ folding and subunit aggregation of the gp. We are currently assessing the effect of mutations within this region in the native E1E2 complex.

Data obtained with chimeric gps which resulted in disulfide-bridged aggregates and which failed to bind conformation-dependent MAbs could not be interpreted further with respect to analysing the regions involved in the CD81 interaction. Substitution of most of the Gla HVR-1 (aa 364–406) within H77c (chimera C4) had no significant effect on folding or CD81 binding. However, chimera C3, carrying region 384–524 of Gla, demonstrated minimal CD81 binding and yet was recognized by all the conformation-dependent MAbs analysed, suggesting that region 384–524 may be important in modulating CD81 interaction without affecting protein folding. This conclusion is supported by data from chimera C5, which carries Gla aa 407–524 within H77c E2, and which shows no change(s) in antigenic conformation yet fails to show a detectable interaction with CD81 (Fig. 4). Previously, we reported that the CD81-binding site within E2 necessary is of a conformational nature (Flint et al., 1999). In addition, MAbs specific for epitopes 480–493 and 544–551 inhibit the E2–CD81 interaction, suggesting that these regions play a direct role in CD81 interaction (Flint et al., 1999). Our data with Gla–H77c E2 chimeras implicating amino acids in region 407–524 in CD81 binding are consistent with and support these observations.

Amino acid sequence comparison analysis between Gla and H77c E2 shows 88–89% identity (Fig. 5). Since the cysteine residues in Gla and H77c E2 are completely conserved, it is unlikely that Gla E2 adopts a conformation that is similar to that found in the native E1E2 complex. We have recently generated a number of Gla–H77c chimeras in the context of full-length E1E2 glycoproteins to further delineate the role of amino acids involved in E1E2 complex formation and its subsequent interaction with CD81. Site-directed mutagenesis of critical amino acid residues identified in this study will increase our understanding of the mechanisms involved in these processes.

We wish to thank D. McGeoch for critical reading of this manuscript; R. M. Elliott and J. Bukh for Gla and H77c cDNA constructs, respectively; Mike Flint and Catherine Maidens for E2 protein and helpful discussions; S. Fong for monoclonal antibodies; S. Levy for GST fusion proteins expressing hCD81 and mCD81 and Larry Loomis-Price for epitope mapping MAb AP320. J.D. was supported by grant 9736 from the Association pour la Recherche sur le Cancer. J.M. acknowledges the financial support of The Wellcome Trust.

References


Received 31 May 2000; Accepted 11 August 2000