The Transmembrane Domains of Hepatitis C Virus Envelope Glycoproteins E1 and E2 Play a Major Role in Heterodimerization*

Oligomerization of viral envelope proteins is essential to control virus assembly and fusion. The transmembrane domains (TMDs) of hepatitis C virus envelope glycoproteins E1 and E2 have been shown to play multiple functions during the biogenesis of E1E2 heterodimer. This makes them very unique among known transmembrane sequences. In this report, we used alanine scanning insertion mutagenesis in the TMDs of E1 and E2 to examine their role in the assembly of E1E2 heterodimer. Alanine insertion within the center of the TMDs of E1 or E2 or in the N-terminal part of the TMD of E1 dramatically reduced heterodimerization, demonstrating the essential role played by these domains in the assembly of hepatitis C virus envelope glycoproteins. To better understand the alanine scanning data obtained for the TMD of E1 which contains GXXXG motifs, we analyzed by circular dichroism and nuclear magnetic resonance the three-dimensional structure of the E1-(350–370) peptide encompassing the N-terminal sequence of the TMD of E1 involved in heterodimerization. Alanine scanning results and the three-dimensional molecular model we obtained provide the first framework for a molecular level understanding of the mechanism of hepatitis C virus envelope glycoprotein heterodimerization.

After their synthesis and integration into the membrane, a large number of membrane proteins associate and form homo- or hetero-oligomeric complexes with new functions. To ensure specific assembly, these proteins must present complementary recognition regions to each other. These interacting regions may be located on the ectodomains and/or the transmembrane sequences. The importance of noncovalent interactions between transmembrane α-helices in oligomerization of membrane proteins is becoming increasingly apparent (1, 2). However, interactions of transmembrane α-helices have only been well characterized for a restricted number of proteins and mainly in the context of homo-oligomerization. Recently, we have shown that the transmembrane domains (TMDs)1 of hepatitis C virus (HCV) E1 and E2 envelope glycoproteins are required in heterodimerization (3). Therefore, these proteins constitute an attractive model to study hetero-oligomerization of transmembrane domains. In addition, like other viral envelope proteins, HCV glycoproteins are supposed to play a crucial role in viral entry by binding to a receptor present on the host cell and inducing fusion between the viral envelope and a membrane of the host cell (4). To ensure that fusion does not occur during protein synthesis and folding, nor during budding of the viral particle, this function has to be tightly controlled. It is well known that oligomerization plays a crucial role in regulating the fusogenic function of such proteins (4). Studying the early steps of oligomerization of HCV envelope glycoproteins is essential to understand how the virus controls its fusion activity.

HCV is the causal agent of hepatitis C which is a major health problem worldwide. It is a positive stranded RNA virus and is a member of the hepacivirus genus in the family Flaviviridae (5). Its genome encodes a single polyprotein of about 3000 amino acid residues that is co- and post-translationally cleaved to generate at least 10 polypeptides (C, E1, E2, p7, NS2, NS3, NS3A, NS4B, NS5A, and NS5B) (6). The lack of a cell culture system supporting efficient HCV replication and particle assembly have hampered the characterization of the envelope proteins present on the virion. However, indirect evidences, like viral neutralization by antibodies, support the idea that HCV envelope glycoproteins are present on the surface of the virion (7). Our current understanding of the assembly of HCV envelope glycoproteins is based on cell culture transient-expression assays with viral or nonviral expression vectors. E1 and E2 are obtained after cleavage of the polyprotein by host signal peptidase(s) (8). They are heavily modified by N-linked glycosylation and are transmembrane proteins with a large N-terminal ectodomain and a C-terminal hydrophobic anchor (9). E1 and E2 glycoproteins have been shown to

* This work was supported by the CNRS, Institut Pasteur de Lille, European Regional Development Fund (ERDF), a PRFMMP grant from the French Ministry of Research, European Union Grant 0QL2-1999-00356, and Association pour la Recherche contre le Cancer (ARC) (to A. O. D.). The costs of publication of this article were defrayed in part by the Agence Nationale de la Recherche sur le Sida (ANRS) (to A. O. D. and B. B.). The atomic coordinates for the NMR structure of the E1-(350–370) fragment and the NMR restraints are available in the Research Collaboratory for Structural Bioinformatics Protein Data Bank under accession number 1EMZ (RCSB010735). The proton chemical shifts of all residues may be located on the ectodomains and/or the transmembrane sequences. The importance of noncovalent interactions between transmembrane α-helices in oligomerization of membrane proteins is becoming increasingly apparent (1, 2). However, interactions of transmembrane α-helices have only been well characterized for a restricted number of proteins and mainly in the context of homo-oligomerization. Recently, we have shown that the transmembrane domains (TMDs) of hepatitis C virus (HCV) E1 and E2 envelope glycoproteins are required in heterodimerization (3). Therefore, these proteins constitute an attractive model to study hetero-oligomerization of transmembrane domains. In addition, like other viral envelope proteins, HCV glycoproteins are supposed to play a crucial role in viral entry by binding to a receptor present on the host cell and inducing fusion between the viral envelope and a membrane of the host cell (4). To ensure that fusion does not occur during protein synthesis and folding, nor during budding of the viral particle, this function has to be tightly controlled. It is well known that oligomerization plays a crucial role in regulating the fusogenic function of such proteins (4). Studying the early steps of oligomerization of HCV envelope glycoproteins is essential to understand how the virus controls its fusion activity.

HCV is the causal agent of hepatitis C which is a major health problem worldwide. It is a positive stranded RNA virus and is a member of the hepacivirus genus in the family Flaviviridae (5). Its genome encodes a single polyprotein of about 3000 amino acid residues that is co- and post-translationally cleaved to generate at least 10 polypeptides (C, E1, E2, p7, NS2, NS3, NS3A, NS4B, NS5A, and NS5B) (6). The lack of a cell culture system supporting efficient HCV replication and particle assembly have hampered the characterization of the envelope proteins present on the virion. However, indirect evidences, like viral neutralization by antibodies, support the idea that HCV envelope glycoproteins are present on the surface of the virion (7). Our current understanding of the assembly of HCV envelope glycoproteins is based on cell culture transient-expression assays with viral or nonviral expression vectors. E1 and E2 are obtained after cleavage of the polyprotein by host signal peptidase(s) (8). They are heavily modified by N-linked glycosylation and are transmembrane proteins with a large N-terminal ectodomain and a C-terminal hydrophobic anchor (9). E1 and E2 glycoproteins have been shown to

1 The abbreviations used are: TMD, transmembrane domain; HCV, hepatitis C virus; mAb, monoclonal antibody; DQF-COSY, double quantum-filtered correlation spectroscopy; LPC, lysophosphatidylcholine; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; RMSD, root mean squared deviation; TFE, 2,2,2-trifluoroethanol; TOCSY, total correlation spectroscopy; HPLC, high-pressure liquid chromatography; ER, endoplasmic reticulum.
assemble as a noncovalent heterodimer (10). However, this process is not efficient, and misfolded heterogeneous disulfide-linked aggregates of E1 and E2 are also produced, at least when using heterologous expression systems (9). Coexpression of E1 and E2 has been shown to be necessary to ensure correct folding of E1 (11). The two glycoproteins likely play separate functions in HCV entry. Indeed, E2 might be involved in receptor binding (12, 13), whereas E1 has been proposed to be the fusion protein (14).

The fact that the envelope proteins of HCV are translated from a single coding region implies that internal signal peptides must be used. The signal sequences of E1 and E2 are present at the C terminus of the immature form of the capsid protein and in the second half of the TMD of E1, respectively (8). In addition, a hydrophobic sequence present in the second half of the TMD of E2 is the signal sequence for a polypeptide called p7. Recently, we showed that the TMDs of E1 and E2 play a major role in the subcellular localization of E1-E2 complex (15–17). The TMDs of E1 and E2 have also been suggested to play a major role in the assembly of the heterodimer. Indeed, deletion of the C-terminal hydrophobic sequence of E2 (11, 18), or its replacement by the membrane anchor of CD4 or a glyco-lysophosphatidylinositol moiety, have been shown to abolish the formation of E1-E2 complexes (15).

The multiple functions played by the TMDs of E1 and E2 glycoproteins are supposed to be essential for the formation of the viral envelope, and most likely impose limitations on the amino acid variability of these domains (3). These TMDs are composed of two stretches of hydrophobic residues separated by a short segment containing at least one fully conserved charged residue (Fig. 1). Replacement of these charged residues by alanine has been shown to alter all the functions of the TMDs of E1 and E2 (3), indicating that charged residues present within these domains are crucial for their multifunctionality.

In this report, we studied the role played by the TMDs of E1 and E2 in heterodimerization. We used alanine scanning insertion mutagenesis (19, 20) in these domains to examine their role in the assembly of the heterodimer. This technique has been shown to be a powerful method to detect dimerization of transmembrane a-helices. Indeed, insertion of a single amino acid into a transmembrane helix displaces the residues on the N-terminal side of the insertion by 110° relative to those on the C-terminal side of the insertion, disrupting a helix-helix packing interface involving residues on both sides of the insertion. Two distinct segments of the TMD of E1 and one of the TMD of E2 were very sensitive to alanine insertion, demonstrating the essential role played by these domains in the assembly of HCV envelope glycoproteins. Neither ER retention nor signal sequence cleavage were affected by these mutations, indicating that assembly can be dissociated from other functions played by the TMDs of HCV envelope glycoproteins. In complement to the alanine scanning data obtained for the N-terminal part of TMD of E1 which contains GXGG motifs, we used circular dichroism (CD) and nuclear magnetic resonance (NMR) to analyze, in the presence of various membrane mimetic solvents (detergents and trifluoroethanol/water mixtures), the three-dimensional structure of an E1 peptide encompassing the sequence involved in heterodimerization (E1-(350–370) segment). Alanine scanning results and the three-dimensional molecular model obtained we provided the first framework for a molecular level understanding of the mechanism of E1E2 heterodimerization. To our knowledge, this is the first report showing that oligomerization of viral envelope proteins can be controlled by transmembrane sequences.
of TFE, the samples were solubilized in 10 mM phosphate (pH 7.4). Micellar solutions were routinely checked for background absorbance, from which it was noted that light scattering was insignificant at the SDS and LPC concentrations used. Peptide concentrations were determined by amino acid analysis.

**CD Measurements** CD spectra were recorded on a Jobin-Yvon CD6 spectrometer calibrated with ammonium d_{22}-camphorsulfonate. Routinely, measurements were done at 298 K in 0.1-cm path length quartz cuvettes (Hellma) with peptide concentrations ranging from 30 to 50 μM. Spectra were recorded in the 190–250 nm wavelength range with 0.2-nm increments and 2-s integration time. The baseline-corrected spectra were smoothed by using a third-order least squared polynomial fit. Assuming that the residue molar ellipticity at 222 nm is exclusively due to α-helix (34), the α-helical content was estimated using the following equation: \( \text{helix} = 100\left[\frac{\theta_{185}}{\theta_{222}}\right] \), where \( \theta_{185} \) is the observed mean residue ellipticity and \( \theta_{222}^\text{helix} \) is the maximum mean residue ellipticity for a helix of finite length. \( \theta_{222}^\text{helix} \) was calculated using the empirical equation for helix-length dependence proposed by Chen et al. (35) \( \theta_{222}^\text{helix} = \left[\theta_{222} - f_{\lambda} \theta_{222}^\text{flat}\right] \), where \( \theta_{222}^\text{flat} \) is the maximum mean residue ellipticity for a helix of infinite length (39,500 for λ = 222 nm), \( f_{\lambda} \) is the fraction of helix, \( i \) is the number of helical segments (1 in this case), \( h \) is a wavelength dependent constant (2.57 for λ = 222 nm), and \( n \) is the number of residues.

**1H NMR Spectroscopy**—Lyophilized peptide was dissolved in an aqueous solution containing 50% TFE-d_{2} (12.5:2.5-trifluoroethyl-1-d_{2}, ethanol >99%) in 0.1 M phosphate buffer. The final peptide concentration was 4 mM and the pH measured was 5.7 (uncorrected). Sodium 2,2-dimethyl-2-silapentane-5-sulfonate was added as an internal reference. All NMR experiments were recorded at 500 MHz on a Varian Unity-plus spectrometer. Spectra were acquired non-spinning at temperatures of 293 and 303 K. Two-dimensional homonuclear 1H experiments (DQF-COSY, Clean-TOCSY, and NOESY) were performed according to the conventional pulse sequences. Water suppression was carried out using selective, low power irradiation during the 1.5-s relaxation delay and conventional pulse sequences. Water suppression was carried out using

**RESULTS**

**Identification of TMD Segments Involved in Heterodimerization**

To confirm the role played by the TMDs of HCV envelope glycoproteins in heterodimerization, we used alanine scanning insertion mutagenesis, a technique which has been shown to disrupt helix-helix interaction in a membrane environment (19, 20). The TMDs of E1 and E2 are present at the C terminus of these proteins (Fig. 1). The N-terminal amino acids of these domains have not been precisely determined but it has been predicted that they could start at position 353 and 718 for E1 and E2, respectively (16, 42). A series of mutants were obtained by introducing a single alanine residue in the TMD of E1 or E2 (Fig. 1). Since HCV envelope glycoproteins are produced after cleavage of a polypeptide, these mutations were introduced in the context of an E1E2 polypeptide. Mutated E1E2 polypeptides were expressed in HepG2 cells by using the vaccinia virus expression system. The ability of the mutants to disrupt heterodimerization was observed when alanine insertion mutations in the TMDs were introduced at positions Ala353, Ala369, and Ala369', of the TMD of E1. Similar results were observed for E2 mutants Ala727', and Ala730' (Fig. 3); however, the effects of alanine insertion were less dramatic than for E1 mutants. Indeed, for E1 mutants Ala353', Ala358', Ala361', and Ala369', disruption of heterodimerization was observed when alanine residues were introduced at positions Ala353', Ala358', Ala361', or Ala369', of the TMD of E1. Similar results were observed for E2 mutants Ala727', and Ala730' (Fig. 3); however, the effects of alanine insertion were less dramatic than for E1 mutants. Indeed, for E1 mutants Ala353', Ala358', Ala361', and Ala369', heterodimerization was reduced to less than 10% of the wild type level (Fig. 2B), whereas about 35% of E1-E2 complex was still detected for the most disruptive mutations in E2 (Fig. 3; Ala727', Ala730'). It is noteworthy that the intensity of E2 pre-
Heterodimerization of HCV Envelope Glycoproteins

Since we have previously shown that E1 needs to be coexpressed with E2 to fold properly (11, 16), we were interested to

locally close to the charged residue which has been reported to be important for the multifunctionality of the TMD of E1 (3). Similarly, the segment identified in the TMD of E2 might include the two charged residues involved in the multifunctionality of this domain (Fig. 1). Interestingly, no change in the efficiency of cleavage of E1E2 polyprotein was observed for E1 mutants (data not shown), indicating that the signal sequence located in the TMD of E1 is not affected by an alanine insertion in this domain.

Assembly of the native E1-E2 complex is not efficient. Indeed, misfolded disulfide-linked aggregates are abundant in cells expressing these proteins (9). We were interested to know whether the mutations introduced in the TMD of E1 or E2 would also affect the formation of these aggregates. For this purpose, we used a mAb that we recently described as specific of E1E2 aggregates (43), and we analyzed the formation of E1E2 aggregates in pulse-chase experiments. Similar levels of aggregates were observed for all the alanine mutants (data not shown), indicating that only noncovalent complex formation is impaired by the insertion of an alanine in segments critical for assembly.

Altogether, our data indicate that segments located close to the charged residue(s) of the TMDs of E1 and E2 are directly involved in heterodimerization. In addition, a second segment playing a role in E1E2 assembly is present near residue 358 of the TMD of E1.

Mutants Impaired in Heterodimerization Retain Their ER Retention Function

Since the TMDs of E1 and E2 are ER retention signals (15, 16), we wanted to determine whether this function would be maintained despite defective assembly of some mutants. To answer this question, we determined the subcellular localization of some of the mutants showing the most disruptive effect on assembly (Ala358 and Ala727) by analyzing the type of glycans associated with these proteins. We have shown previously that the type of glycans associated with HCV envelope glycoproteins, or chimeras having the TMD of E1 or E2, is a good indicator of their subcellular localization (16, 17). To analyze the glycans associated with an E2 mutant, we used a vaccinia virus recombinant expressing E2-Ala727 alone. Since E1 does not fold properly in the absence of E2 (11), we analyzed the subcellular localization of a chimeric protein CD4-E1-Ala358, comprising the ectodomain of CD4 fused to the TMD of E1-Ala358, and expressed by using a vaccinia vector. Such a strategy has been used previously to demonstrate that the TMD of E1 is an ER retention signal (16).

To evaluate the level of processing of the glycans associated with E2-Ala727 or CD4-E1-Ala358, HepG2 cells expressing these proteins were pulse-labeled with [2-3H]mannose, chased for 4 h, and used for immunoprecipitation with an anti-E2 (H53) or anti-CD4 (OKT4) mAb. Glycans associated with these proteins were removed by peptide N-glycanase F treatment (9). We were interested to know whether the mutations introduced in the TMD of E1 or E2 impaired the folding of E1 (11, 16), we were interested to
or 7 mannose residues, respectively. species possessing two GlcNAc residues at their reducing end and 9, 8, 31432 E1-Ala358 of E1 by E2. HCV envelope glycoproteins is important for assisted folding intensity of the bands observed during the chase (Fig. 5). conditions formed high molecular weight aggregates as a has to be noted that part of E1 separated under nonreducing quantitative analysis showed a 40% reduction of this form. It takes advantage of an increase in mobility as a protein acquires residential disulfide bonds. An oxidized form of E1, which appeared slowly, was clearly detected in the context of the wild type M9, M8, and M7 indicate the oligosaccharide species possessing two GlcNAc residues at their reducing end and 9, 8, or 7 mannose residues, respectively.

know whether impairment in heterodimerization would lead to less efficient folding of E1. E1 folding was analyzed for one of the most affected mutants (Ala358) expressed in the context of an E1E2 polyprotein. Disulfide bond formation in E1 was monitored by SDS-polyacrylamide gel electrophoresis under nonreducing conditions as described previously (31). This method takes advantage of an increase in mobility as a protein acquires a compact conformation stabilized by the formation of intramolecular disulfide bonds. An oxidized form of E1, which appeared slowly, was clearly detected in the context of the wild type E1E2 (Fig. 5) as previously observed (31). In contrast, the intensity of the oxidized form of E1-Ala358 was lower, and a quantitative analysis showed a 40% reduction of this form. It has to be noted that part of E1 separated under nonreducing conditions formed high molecular weight aggregates as a function of time (data not shown), which explains the lower intensity of the bands observed during the chase (Fig. 5). Together, these data indicate that interaction of the TMDs of HCV envelope glycoproteins is important for assisted folding of E1 by E2.

FIG. 4. HPLC analysis of the oligosaccharides bound to CD4-E1-Ala358 (A) and E2-Ala727 (B). HepG2 cells were co-infected with vTF7-3 and the appropriate vaccinia virus recombinant at a multiplicity of 5 plaque forming units/cell. At 4 h post-infection, cells were metabolically labeled for 5 min and chased for indicated times (min). Cell lysates were immunoprecipitated with mAb A4 (anti-E1), and proteins were separated by SDS-polyacrylamide gel electrophoresis (10% acrylamide) under nonreducing conditions. wt E1, wild type; E1red, reduced E1; E1ox, oxidized E1.

Structural Characterization of the E1 Region Involved in Heterodimerization

Previous reports have shown that alanine insertions disrupt interactions when introduced near the center of transmembrane helices (19, 20). Our data indicate that segments located close to the charged residue(s) present in the TMDs of E1 and E2 are directly involved in heterodimerization, suggesting that, despite their charge, these residues might be located near the center of the membrane spanning sequence (see “Discussion”). However, the presence of a second heterodimerization segment in the N-terminal region of the TMD of E1 is a novel finding which is not readily understood. To better understand the results observed for E1 mutants, the structure of a synthetic peptide identical to the sequence of the N-terminal part of the TMD of E1 (denoted E1-(350–370); see Fig. 1) was studied by CD and NMR. Indeed, CD and NMR spectroscopies can provide conformational information at the residue level on membrane proteins or isolated fragments, especially transmembrane segments incorporated into membrane environment or organic solvent (see Refs. 44–46 and references therein; for a review, see Ref. 47). For example, Shon et al. (48) have shown that the conformation of bacteriophage P1 coat protein with a single transmembrane segment is very similar in detergent micelles and phospholipid bilayer using solution and solid-state NMR spectroscopies, respectively. For larger membrane proteins Arseniev and colleagues (e.g., Refs. 49–51) have shown that the structure of isolated transmembrane spans of bacteriorhodopsin in the presence of organic solvents or detergent micelles is quite consistent with the structure of the corresponding regions in the whole protein determined by crystallography (52).

CD Analyses—The secondary structure of E1-(350–370) was readily examined by CD spectroscopy after solubilization in various solvent systems which provide a membrane-like environment. Similar approaches have been used in related studies (see above). Similar results were obtained in three distinct solvent systems in the presence of 10 mM sodium phosphate (pH 7.4): 50% TFE, 1% lyso phosphatidyl choline micelles, and 200 mM SDS micelles. In all cases, the CD spectra of E1-(350–370) exhibited two minima around 208 and 222 nm, and a maximum at 192 nm. This is typical of a peptide in α-helix conformation. Similar results were obtained when using dodecylmaltoside or dodecylphosphocholine as membrane mimetic systems (data not shown). Increasing the TFE concentration above 50% resulted in only minor changes in the amplitude of the spectrum (data not shown), indicating that the maximum α-helical folding is reached in 50% TFE. This is consistent with the suggestion of Jaspanoff and Fersht that, for a peptide with high helical propensity, the helicity is complete by 50% TFE. The maximal molar ellipticity per residue observed at 222 nm corresponded to an α-helical content of 48, 44, and 52% in TFE, SDS, and LPC, respectively. These values
are similar and clearly indicate that E1-(350–370) is only partly folded into α-helix. Moreover, these results indicate that 50% TFE is an appropriate medium to mimic the membrane-like environment of E1-(350–370) and validate its use to analyze the conformation of this peptide by NMR (for an overview about TFE, see Refs. 54 and 55).

NMR Spectroscopy—Deuterated micellar SDS and dodecylphosphocholine are popular membrane mimetic solvents for structure analysis of membrane peptides by liquid NMR (47). Unfortunately, E1-(350–370) appeared to be poorly soluble in SDS or dodecylphosphocholine in the millimolar concentration ranges required for NMR experiments. Consequently, we studied the three-dimensional structure of E1-(350–370) dissolved in 50% TFE (v/v) which yielded well resolved spectra as illustrated by the extract of NOESY (Fig. 7A). The spectra were assigned using the classical method (38): the spin systems were identified with DQF-COSY and TOCSY spectra. The sequential assignment was performed with the help of the NOESY spectrum obtained at a mixing time of 150 ms. Despite the poor dispersion of the NH and Hα resonances (all NH resonances are in a range of 0.9 ppm, and 17 of 21 Hα resonances are in a range of 0.5 ppm), the sequential attribution of all spin systems was completed (see Fig. 7B). No sign of any other stable peptide conformation was observed in the NMR spectra. The 1H chemical shifts are available under the BMRB accession number 4699.

Secondary Structure—Fig. 7B shows an overview of the sequential and medium range NOE connectivities and the chemical shift analysis for 1Hα. Despite the lack of data due to numerous overlapping cross-peaks (indicated by asterisks on Fig. 7B), the analysis of NOEs allows the distinction of several structural regions. The main body of the peptide (Ile365–Lys370 segment) displays most of the characteristics of an α-helix conformation, such as strong dNN(\(i, i+1\)) and medium doN(\(i, i+1\)) sequential connectivities, and weak doN(\(i, i+2\)), medium doN(\(i, i+3\)), medium or strong daβi(\(i, i+3\)), and weak daN(\(i, i+4\)) medium range connectivities. The C-terminal part of the peptide (Ala365 and Lys370) remains more flexible with less medium range connectivities, which is the sign of a fraying end often encountered in peptide termini. The second structural unit (from Gly354 to Gly358) is characterized by stronger doN(\(i, i+2\)) constraints and the lack of daN(\(i, i+4\)) connectivities present in more flexible helical structures or that observed in a 3/10 helix. The N-terminal part of the peptide (Gly350–Trp353) is devoid of medium range NOEs and remains unstructured.

The 1Hα chemical shift difference (\(\Delta \delta \)) is another NMR-related parameter used to analyze the protein secondary structure, and it is negligibly affected by addition of up to 50% TFE (56). Relative to a random conformation, an increase of helicity...
TABLE I
Statistics on the final simulated annealing structures of E1-(350–370) in 50% TFE

| A. Constraints used | Distance restraints |
|---------------------|---------------------|
| Intra-residue        | 132                 |
| Sequential          | 97                  |
| Medium range        | 105                 |
| Long range          | 3                   |
| Total distance restraints | 337               |

| B. Statistics for the final X-PLOR structures |
|----------------------------------------------|
| Number of structures in the final set | 24 |
| X-PLOR energy (kcal mol⁻¹) | −68.9 ± 11.6 |

| NOE violations |
|----------------|
| Number > 0.5 Å | None |
| Root mean square deviation (Å) | 0.037 ± 0.005 |
| Deviation from idealized covalent geometry |
| Angles (degree) | 1.13 ± 0.11 |
| Improper (degree) | 0.27 ± 0.02 |
| Bonds (Å) | 0.0052 ± 0.0003 |

| Root mean square deviation (Å) |
|--------------------------------|
| Backbone (Cα, Cβ, N) | 0.31 |
| Helix (359–367) | 2.87 |
| All heavy atoms | 1.63 |
| All residues | 3.99 |
| Ramachandran data* |
| Residues in most favoured regions (%) | 77.3 |
| Residues in allowed regions (%) | 19.8 |
| Residues in generously allowed regions (%) | 2.9 |

* From PROCHECK-NMR (41) excluding glycine residues.

However, some helical folding was observed on the C-terminal part of the peptide (Trp368-Lys370 segment) in calculated structures (Fig. 8A). This is in agreement with the presence of some medium range NOE constraints in this region (see Fig. 7). Concerning the N-terminal part, it should be noted that the Gly354-Gly358 segment was mostly observed as helical in each calculated structure (Figs. 8, B and C), but failed to superimpose in all structures. This could be related to the weak intensities of medium range α-helix NOEs in this region (see Fig. 7). This structural instability of the N terminus of the peptide is very likely due to the presence of glycine residues that are well known to be helix breakers. However, in the full-length E1 glycoprotein context and/or E1E2 heterodimer context, it is very likely that the stabilization of the conformation of this segment arises from its interaction with other segments of these proteins. In such a context, the presence of an helical folding of the Gly350-Gly358 segment is possible. Fig. 8D shows a theoretical α-helix projection of the Ala349-Ile359 segment assuming that such a conformation might occur in the full-length protein context. This representation highlights that the three glycine residues 350, 354, and 358 are located on the same side of the putative α-helix, suggesting a role for these glycines in E1E2 heterodimerization.

**DISCUSSION**

The TMDs of HCV envelope glycoproteins are extreme examples of multifunctionality of membrane spanning sequences. Indeed, besides their role as membrane anchor, they possess a signal sequence function in their C-terminal half, are responsible for ER localization of E1 and E2, and play a role in the assembly of these proteins. In this work, we used alanine scanning insertion mutagenesis to study the oligomerization function of these domains. We demonstrated that the TMDs of E1 and E2 play a direct role in heterodimerization and that E1E2 assembly can be disrupted without affecting the other functions. Moreover, whereas one segment was found to be critical for heterodimerization in the TMD of E2, two structurally different segments were identified for E1. In addition, we determined the three-dimensional structure of E1-(350–370) peptide in the presence of membrane mimetic solvents by NMR to provide a structural framework for a better analysis of the mechanism of E1E2 heterodimerization.

The TMDs of HCV envelope glycoproteins play a direct role in heterodimerization. Deletion or replacement of the C-terminal hydrophobic sequence of E2, or mutation of the charged residues in the TMDs of E1 or E2 have been shown to dramatically reduce the formation of E1-E2 complexes (3, 11, 15, 18). Although these data indicate that the TMDs are involved in the assembly of HCV envelope glycoproteins, the effect of the mutations of charged residues might be indirect. Indeed, these mutations lead to secretion or transport of the mutated protein out of the ER compartment, and the lack of complex formation could be due to the impossibility for E1 and E2 to find each other at the time of assembly. However, E1 and E2 interact before completion of their folding (9), and this process is supposed to occur in a specific compartment, the ER (58). The impairment in heterodimerization following alanine scanning insertion mutagenesis clearly shows that the TMDs of E1 and E2 are directly involved in assembly. Indeed, this technique has been shown to be very useful to identify critical segments of transmembrane α-helices involved in helix-helix interactions (19, 20). In addition, data of alanine scanning insertion mutagenesis (20, 59) fit very well with a recent model predicting the stability of transmembrane helix-helix interactions in mutants of glycoporphin A (60). Interestingly, with the exception of the impairment in assisted folding of E1 by E2, no other functions of the TMDs of HCV envelope glycoproteins seems to be
Fig. 8. Structure of E1-(350–370) calculated from NMR-derived constraints measured in 50% TFE. A, superimposition of the α-carbon chains of the 24 final structures aligned for the best overlap of residues Ile350 to Asn370. B, backbone and Cα–Cβ bonds of the best representative structure of the set of 24 structures. Color code of residue conformation: dark gray, α-helix (Ile350–Asn370); medium gray, frayed helix (Gly354–Gly358 and Trp353–Trp355); light gray, random (Gly350–Trp355, Ala356–Lys370). C, axial helix projection of Gly350–Ile359 segment of the best NMR structure. D, ideal α-helix projection of the Ala354–Ile359 segment showing that glycine residues 350, 354, and 358 are localized on the same side of the putative helix (see the text for further explanations). The letters in parentheses indicate possible amino acid changes observed at the indicated positions (reported in Ref. 3). The figures were drawn with Rasmol 2.6 (R. Sayle program).

altered. This is clearly different from replacement mutagenesis of the charged residues which leads to an alteration of all the functions played by these TMDs (3).

The TMDs of E1 and E2 are probably not the sole determinants for heterodimerization. N-terminal sequences in E2 and also in E1 have been suggested to be important for HCV envelope glycoprotein assembly (61), but deletion mutant analysis has failed to identify any single region which is required for noncovalent interaction (62). Interestingly, assisted folding of the ectodomain of E1 by E2 suggests that regions other than the TMDs might enter into contact. For many viral envelope proteins, the ectodomains have been shown to be involved in oligomerization (63), a feature which is important to regulate the fusogenic function of these proteins (64). Why should there be different regions involved in heterodimerization of HCV envelope glycoproteins? Assembly and folding of HCV envelope glycoproteins seem to be interconnected events. It is likely that an early contact between these proteins, probably initiated by their TMDs, is necessary to bring their ectodomains into contact, which seems to be necessary for the formation of a native complex. This could explain why an alanine insertion mutation which is disruptive for assembly is also disruptive for the assisted folding of E1 by E2.

The data of alanine scanning insertion mutagenesis in the TMD of E2 suggest that this domain forms a single transmembrane segment with the charged residues located in the middle of the membrane spanning sequence. This fits well with the prediction of a single transmembrane helical segment from residue 718 to 742. Indeed, a minimum of 16 leucines are required to form a transmembrane α-helix (65), but stretches of 20–25 hydrophobic amino acids are generally observed in integral membrane proteins. Although the presence of charged residues within the center of this transmembrane helix seems to be surprising a priori, it is very unlikely that segment 718–742 adopts a stable conformation allowing two transmembrane passages. In such eventuality, each of the two small hydrophobic stretches (718–727 and 731–742) would have to adopt an extended structure. Such an extended structure is very unfavorable in a membrane environment because of the absence of stabilization by hydrogen bonds, while the hydrophobic bonds network of an α-helix offers the most stable solution from a thermodynamic standpoint. The formation of transmembrane β-sheets as observed in the porin family of membrane proteins also yields a stable hydrogen-bond network compatible with the membranous environment. However, examination of the nature and position of residues in the TMD of E2 does not provide any sign that such a stable conformation might take place. Concerning the potential problem of the 2 charged residues within the membrane (Asp and Arg), it is possible that they could form ion pairs as already described in the cases of several membrane proteins (66). Taken together, these data clearly support the hypothesis that the TMD of E2 is a single transmembrane helix centered near amino acid 730.

Concerning the TMD of E1, the data of alanine scanning insertion mutagenesis allowed us to identify two distinct segments involved in heterodimerization with the TMD of E2. First, similarly to what was observed for the mutations in E2, the disruptive effect of Ala369 inserted very close to Lys370 suggests that this charged residue might also have a central position in the membrane. Second, the N-terminal region of the TMD of E1 appears to be very sensitive to alanine insertions. This is likely due to the presence of GXXCG motifs which have been well documented to ensure specific homodimerization of transmembrane segments in membrane proteins (see below). These features, together with the thermodynamic reasons exposed above for the TMD of E2, strongly suggest that E1 is anchored in the membrane by a single transmembrane α-helix. This is also supported by the fact that in the E1-(350–370) model (see Fig. 8), the Ile359–Asn367 segment was found to form a stable α-helix and the Gly354–Gly358 segment has a clear tendency to adopt an helical fold (see Fig. 8). In addition, taking into account that aromatic residues and positively charged residues are often encountered at the membrane interface, it might be predicted that His356 and Trp353 residues form the N terminus of the TMD of E1. In summary, we thus assume that the 25-amino acid long segment Gly354–Phe379 forms a single transmembrane helix and is the counterpart of the E2 transmembrane helix. It should be underlined that a single membrane spanning topology for the TMDs of E1 and E2 would also fit very well with the current model of ER retention by membrane determinants. A usual feature of such signals is the presence of one or several hydrophilic residues within the hydrophobic TMD (67–70), and their effect is strongest when they are localized toward the middle of the membrane (67).

To sequentially ensure some of their functions, the TMDs of E1 and E2 might adopt different topologies. Besides their role in heterodimerization, these TMDs also have a signal sequence function in their C-terminal half (3). Since the ectodomains of
E1 and E2 are translocated into the lumen of the ER, the N terminus of their TMD should also be oriented toward the luminal side of the ER. The signal sequence function present in the second half of these TMDs suggests that the C terminus of these domains should also be oriented, at least transiently, toward the luminal side of the ER. These domains are therefore likely to form a hairpin structure with the charged residues close to the cytosolic face of the membrane. It is reasonable to think that this hairpin structure might be transient in the environment of the translocon before signal sequence cleavage has occurred. A reorientation of the second stretch of hydrophobic residues leading to a single transmembrane segment would occur immediately after signal sequence cleavage at its C terminus and before membrane integration and heterodimerization. Assuming that the TMDs of E1 and E2 are both composed of a single transmembrane α-helix, it is reasonable to postulate that these 2 helices interact together to ensure E1E2 heterodimer formation. However, the putative existence of a single transmembrane segment between Gly350 and Phe379 suggests that part of the Gly350-Gly354 segment might lie outside of the membrane while it is sensitive to alamino insertion. From a structural point of view, this segment appears to be unfolded (see Fig. 8), but this is likely due to the absence of conformational stabilization by tertiary interactions in the isolated E1-(350–370) peptide context used here. In the E1 and/or E2 context, it is thus possible that the Gly350-Gly354 segment folds into an α-helix and that the whole E1-(350–379) segment forms a single long α-helix upon E1E2 heterodimerization. Interestingly, the simulation of such a conformation for the Gly350-Gly354 segment clearly shows that the three well conserved glycine residues (Gly350, Gly354, and Gly358) lie on the same side of the putative helix (see helix projection, Fig. 8D) and form 2 consecutive putative GXXG motifs (Gly350-Gly354 and Gly354-Gly358). The presence of such a glycine motif has been reported to be essential to ensure the specific helix to helix interaction for several transmembrane proteins such as glycoporin A (71) and phage M13 coat protein (72). Moreover, the importance of the GXXG motif has been recently highlighted by two articles of Engelmann and colleagues (73, 74): (i) in one experimental study, they have selected TMDs exhibiting high affinity homooligomerization from a randomized sequence library based on the right-handed dimerization motif of glycoporin A and this GXXG motif was the most frequent isolated using the TOXCAT system which measures transmembrane helix-helix association in the Escherichia coli inner membrane (73); (ii) in a theoretical study, they have analyzed frequently occurring combinations of residues in a data base of putative TMDs and the main theme observed is patterns of small residues (Gly, Ala, and Ser) at i and i+4 found in association with large aliphatic residues (Ile, Val, and Leu) at neighboring positions (i.e. i+1/−/− and i+1/−/−) (74). A very suggestive correlation can be made with our studies; coinciding with the most dramatic effects of Ala insertion, the TMD of E1 displays in its N-terminal part two successive GXXG motifs, and the second one is GVXXG. Consequently, it is tempting to speculate that the 2 glycine motifs of E1 play a major role for E1E2-specific heterodimerization and are located at the interface of E1 and E2 dimer.

In conclusion, the structural and functional analyses reported here allowed us to show for the first time that oligomerization of viral envelope proteins can be controlled by transmembrane sequences and to propose a molecular framework for the understanding of the mechanism of E1E2 heterodimerization. Besides the comprehension of the assembly of HCV envelope glycoproteins, the multifunctionality of the TMDs of E1 and E2 provides a very interesting model to study the structure/function relationship of transmembrane sequences.
Mizushima, H., Hijikata, M., Asabe, S.-I., Hirota, M., Kimura, K., and Shimotohno, K. (1994) J. Virol. 68, 6215–6222

Soulie, S., Neumann, J. M., Berthoumes, C., Moller, J. V., le Maire, M., and Forge, V. (1999) Biochemistry 38, 5813–58121

Opella, S. J. (1997) Nat. Struct. Biol. 4, (suppl.) 845–848

Shon, K. J., Kim, Y., Colnago, L. A., and Opella, S. J. (1991) Science 252, 1303–1305

Barsukov, I. L., Abdulaeva, G. V., Arseniev, A. S., and Bystrov, V. F. (1990) Eur. J. Biochem. 192, 321–327

Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P., and Landau, E. M. (1997) Science 277, 1676–1681

Janin, A., and Fersht, A. R. (1994) Biochemistry 33, 2129–2135

Buck, M. (1998) Q. Rev. Biophys. 31, 297–355

Monterrey, R., Aubert-Foucher, E., McLeish, M. J., Hill, J. M., Ficheux, D., Jaquinod, M., van der Rest, M., Deleage, G., and Penin, F. (1999) Biochemistry 38, 6479–6488

Merutka, G., Dyson, H. J., and Wright, P. E. (1995) J. Biol. Chem. 270, 119–129

Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) Biochemistry 31, 1647–1651

Hammond, C., and Helenius, A. (1995) Curr. Opin. Cell Biol. 7, 523–529

Mingarro, I., Elofsson, A., and von Heijne, G. (1997) J. Mol. Biol. 272, 633–641

McKenzie, K. R., and Engelman, D. M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 3583–3590

Yi, Y., Nakamoto, Y., Kaneko, S., Yamashita, T., and Murakami, S. (1997) Virology 231, 119–129

Patel, J., Patel, A. H., and McLauchlan, J. (1999) J. Gen. Virol. 80, 1681–1690

Doms, R. W. (1990) Methods Enzymol. 191, 841–854

Hernandez, L. D., Hoffman, L. E., Wolsberg, T. G., and White, J. M. (1996) Annu. Rev. Cell Dev. Biol. 12, 627–661

Monne, M., Nilsson, I., Elofsson, A., and von Heijne, G. (1999) J. Mol. Biol. 293, 807–814

Kaback, H. R., Voss, J., and Wu, J. (1997) Curr. Opin. Struct. Biol. 7, 537–542

Bonifacino, J. S., Cosson, P., Shah, N., and Klausner, R. D. (1991) EMBO J. 10, 2783–2793

Bonifacino, J. S., Suzuki, C. K., and Klausner, R. D. (1990) Science 247, 79–82

Lettourneur, F., and Cosson, P. (1998) J. Biol. Chem. 273, 33273–33278

Yang, M., Ellenberg, J., Bonifacino, J. S., and Weissman, A. M. (1997) J. Biol. Chem. 272, 1970–1975

Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J., and Engelman, D. M. (1992) Biochemistry 31, 12719–12725

Li, Z., Gibowicka, M., Joensson, C., and Deber, C. M. (1993) J. Biol. Chem. 268, 4584–4587

Russ, W. P., and Engelman, D. M. (2000) J. Mol. Biol. 296, 911–919

Senes, A., Gerstein, M., and Engelman, D. M. (2000) J. Mol. Biol. 296, 921–936