The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation

Nathalie Majeau, Valérie Gagné, Annie Boivin, Mariélène Bolduc, Josée-Anne Majeau, Dominique Ouellet and Denis Leclerc

Centre de Recherche en Infectiologie, Pav. CHUL, U. Laval, 2705 boul. Laurier, Québec (Québec), Canada G1V 4G2

INTRODUCTION

Hepatitis C virus (HCV) is a plus-strand RNA virus that causes acute and chronic liver disease. More than 170 million people worldwide are infected, which is four times as many as for HIV (Lauer & Walker, 2001). In the next few years, the number of deaths from HCV-associated diseases may even surpass the death rate caused by AIDS (Cohen, 1999). Current therapies against HCV are unsatisfactory (Koshy et al., 2002; Toyoda et al., 2000). Therefore, there is an urgent need to identify targets for the development of novel drugs to prevent the spread of the disease. Inhibition of HCV assembly is an excellent target because it is unique to the virus and thus could lead to the discovery of a specific drug with minimal secondary effects.

The HCV core (C) protein is involved in the assembly and packaging of the viral plus-strand RNA genome (Lin et al., 1994). The HCV C gene encodes a protein of 191 amino acids (P23), which is located at the N terminus of the 3011 amino acid-long viral polyprotein (Choo et al., 1994; Lauer & Walker, 2001). The C protein is cleaved from the protein precursor by a host signal peptidase located in the endoplasmic reticulum (ER) (Hijikata et al., 1993; Hüsey et al., 1996; Lin et al., 1994). Following cleavage, the C protein remains anchored to the ER through a C-terminal hydrophobic region (Grakoui et al., 1993) and gets further processed at its C terminus by a signal peptide peptidase leading to a mature protein composed of the N-terminal 179 amino acid residues, referred to as P21 (Liu et al., 1997; McLauchlan et al., 2002; Yasui et al., 1998). Analysis of the C protein produced in mammalian expression systems and the C protein isolated from infectious sera suggest that P21 is the major form of the C protein in viral particles (Ishida et al., 2001; Yasui et al., 1998).

The diameters of non-enveloped HCV core particles collected in the 1.22–1.25 g ml⁻¹ fractions from sucrose gradients were shown to be 33–40 nm (Ishida et al., 2001). HCV nucleocapsid particles ranging from 25 to 35 nm were also extracted by ether/butanol treatment from the very-low-density lipoprotein (VLDL) fraction of infected plasma (Andre et al., 2002). Optical rotational EM techniques have revealed that the structure of the HCV capsid exhibits sixfold symmetry with a regular hexagon side of 20 nm (Ishida et al., 2001).

The mechanisms of HCV assembly and formation of the nucleocapsid structure are not well understood. As for many viruses, assembly is believed to be initiated by the binding of the C protein to a defined structure, often a stem–loop, in the nucleic acid sequence. This interaction facilitates nucleation of the C protein, and oligomerization occurs via specific protein–protein interactions. Several reports have described the formation of nucleocapsid-like particles (NLPs) when HCV C protein is expressed in E. coli (Lorenzo et al., 2001), yeast (Acosta-Rivero et al., 2002), insect cells (Baumert et al., 1998) or mammalian systems (Blanchard et al., 2002; Shimizu et al., 1996). The most recent development in assembly came from an in vitro system developed with purified recombinant proteins from E. coli (Kunkel et al., 2001). It was shown that either the
first 600 nt of the HCV genome or structured tRNA could trigger the formation of NLPs when mixed with purified C protein. Also, it was shown that the N-terminal 120 aa of the C protein were sufficient for self-assembly into particles (Kunkel et al., 2001; Lorenzo et al., 2001).

In this study, we investigated the essential domain of the C protein involved in viral assembly. Different forms of HCV C protein produced in E. coli were purified under native conditions and their capacity to self-assemble in vitro was evaluated. We determined that the 75 N-terminal residues of the C protein could generate NLPs. An extension to 79 aa generated more uniform and packed particles. Finally, we showed that an HCV nucleocapsid made of the 82 N-terminal aa of the C protein could also be formed in vivo if the protein was targeted to the ER membranes of the cell.

**METHODS**

**Cloning and expression of HCV C proteins in E. coli.** The plasmid pCV-H77c (generously provided by J. Bukh, NIH) was used to generate the HCV C constructs presented in Fig. 1. C 1–170, C 1–112 and C 1–71 were amplified by PCR with primers C 170-6h (5′-CATGGGATCCCTTACTAATGGTGATGGTGATGGTGACGCGTGGTACTAGTAGGAAGGTTCCCTGTTGCATAGTTCACGCC-3′), C 112-6h (5′-CATGGGATCCCTTACTAATGGTGATGGTGATGGTGACGCGTGGTACTAGTAGGAAGGTTCCCTGTTGCATAGTTCACGCC-3′) and C 71-6h (5′-CATGGGATCCCTTACTAATGGTGATGGTGACGCGTGGTACTAGTAGGAAGGTTCCCTGTTGCATAGTTCACGCC-3′), respectively, together with primer C N (5′-CATGAACCATGGCGAGCACGAATCTGCAAAGACCCCCCAGCTACCCTC-3′). PCR products were digested with restriction enzymes NcoI and BamHI and cloned into a pET3d expression vector (New England Biolabs). Core C-terminal deletion constructs C 1–92, C 1–82, C 1–79 and C 1–75 were generated by PCR using the C 1–170 clone as template DNA and the

![Fig. 1. Schematic representation of the HCV C protein. (A) Amino acid sequence of the HCV C protein and mapping of the important domains: domain 1, which is rich in basic residues (labelled with asterisks), domain 2, which includes the lipid-droplets-association motif with P-138 and P-143 (Hope et al., 2002), and domain 3, the hydrophobic region including the ER-anchoring domain. The HCV C protein constructs expressed in E. coli and P. pastoris are shown in (B) and (C), respectively. The coding frame of the polyprotein is represented by a pale grey box. The 6-His tag located at the C terminus of all the constructs is shown in dark grey. The Spel (S) and Mulf (M) restriction sites are shown in white. N, Ncol site; B, BamHI site; E, EcoRI site; ER, ER-anchoring domain.](image-url)
primers C 92 (5′-CATGACTAGTGCACCTCATTGTCG-3′), C 82 (5′-CATGACTAGTGGTACCCCGTGGTACC-3′), C 79 (5′-ACCTGACTAGTGGTCGACCCGCTCTG-3′) and C 75 (5′-ACCTGACTAGTGGTATCCTCCGCGCCGAC-3′), respectively, together with primer c′-6h-Pet (5′-CATGACTAGTCCACCGGT-CAC-3′). The clones were circularized by ligation after digestion of the DNA with SspI. The sequences of the HCV clones were confirmed by DNA sequencing. The E. coli expression strain BL21(DE3) RIL (Stratagene) was transformed with C protein-expressing pET3d expression strain BL21(DE3) RIL (Stratagene) was transformed with C protein-expressing pET3d vector (Fig. 1B). The proteins were separated under native conditions using an affinity column specific for the His vector (Fig. 1B). The proteins were separated under native conditions using an affinity column specific for the His tag located at the C terminus of each of the mutant proteins. 

To improve our understanding of HCV assembly, we generated several C-terminally deleted forms of the HCV C protein that excluded the C-terminal hydrophobic ER-anchoring domain (Fig. 1A) and different sections of the HCV C protein (see below).

**Rabbit polyclonal anti-HCV C protein antibodies.** The rabbit polyclonal anti-HCV C protein antibody against C 1–170 was prepared using an HCV C protein of 170 aa expressed in E. coli BL21(DE3) RIL cells as described above. The C protein was purified under denaturing conditions, according to Kunkel et al. (2001). Purified protein (25 µg in 50 µl elution buffer) was mixed with 250 ng tRNA in the presence of 1× protease inhibitor cocktail and 0.5 U RNase inhibitor (Roche). The reactions were incubated at 37°C for 10 min followed by 15 min on ice.

**Expression and purification of different forms of the HCV C protein**

To improve our understanding of HCV assembly, we generated several C-terminally deleted forms of the HCV C protein that excluded the C-terminal hydrophobic ER-anchoring domain (Fig. 1A) and different sections of the HCV C protein (Fig. 1B). The proteins were purified under native conditions using an affinity column specific for the His tag located at the C terminus of each of the mutant proteins. An additional 5 aa were also present at the C terminus of all the constructs because of the presence of the SspI and MluI sites (Fig. 1B). The purified proteins were separated by SDS-PAGE (Fig. 2A) and were detected with an anti-C protein antibody raised against an N-terminal C protein peptide (Fig. 2B) to confirm the presence of the recombinant proteins. The yields were between 1 and 10 mg protein (1 bacterial culture)−1, but the C 1–170 clone...
produced less protein than the other constructs. We observed a few contaminant bands on the gel, which co-purified with the recombinant proteins. These proteins (at 25 kDa and 38 kDa) did not react with the anti-C protein antibodies in Western blots. However, a signal corresponding to a dimer of the HCV C protein could be detected in some experiments (data not shown).

In vitro assembly of the recombinant HCV C proteins into NLPs

The HCV C protein can form nucleocapsid-like particles in solution if provided with RNAs containing stem–loop secondary structures (Kunkel et al., 2001). We investigated the minimal length of C protein that would allow NLP formation. Purified proteins were used at concentrations ranging from 0.1 to 1 mg ml⁻¹ in an in vitro assembly assay. Routinely, assembly was initiated by adding tRNA to the proteins in acetate buffer solution as described in Methods. The proteins C 1–170, C 1–112, C 1–92, C 1–82, C 1–79 and C 1–75 (Fig. 3A–F, respectively) formed N LPs as revealed by EM after negative staining with 1% uranyl acetate. However, even at higher protein concentrations, samples containing C 1–71 did not generate particles (Fig. 3G). Protein integrity, including that of C 1–71, was not affected by the in vitro assembly conditions, as verified by SDS-PAGE (data not shown). These results indicated that an important domain for assembly and packaging of the RNA was affected by the truncation of the C protein to 71 aa. NLPs failed to assemble when tRNA was omitted from the reaction. As a control, C 1–170 was purified under denaturing conditions (5 M urea) and dialysed against the assembly buffer; this protein did not result in the formation of NLPs (Fig. 3H), appearing instead to aggregate into elongated, unorganized structures. The formation of filaments was probably due to a non-specific interaction of the protein with the tRNA.

HCV C 1–75 was the smallest protein domain of the core that generated NLPs in vitro. Assembly with C 1–75 resulted in the formation of heterogeneous NLPs ranging in size from 15 to more than 150 nm in diameter with a mean size of 25 nm (Fig. 4E). A large proportion (70%) of these structures observed by EM appeared as irregular protein aggregates of less than 25 nm in diameter. However, with longer recombinant proteins, such as C 1–79, we observed that most NLPs were well formed, with a defined structure similar to that observed with the wt capsid preparation. We estimated that 90% of the NLPs composed of C 1–79 were between 15 and 40 nm in diameter (Fig. 4D). We also observed a small proportion of larger particles of up to 80 nm. The NLPs of C 1–170, C 1–112, C 1–92 and C 1–82 were stable and uniform in shape and diameter (Fig. 3A–D). The mean diameter of C 1–82 NLPs was 26 nm. More than 70% of the NLPs generated with the recombinant C 1–170 were between 25 and 40 nm in diameter (Fig. 4A) with a mean of 33 nm, which is consistent with the size of the wt virus capsid (Andre et al., 2002; Ishida et al., 2001) and suggested that the protein was properly folded. The density of the C 1–170 particles was estimated at 1.24 g ml⁻¹ and that of C 1–82 particles at 1.27 g ml⁻¹ after continuous sucrose density gradients (20–63%) (data not shown). We also tested all the recombinant proteins with the first 600 nt of HCV genomic RNA, which resulted in the formation of similar NLPs (data not shown).

Particles derived from the HCV C protein were confirmed using an immunogold labelling technique with a rabbit antibody raised against denatured C 1–170. NLPs formed from C 1–170 (Fig. 3I) and C 1–82 (Fig. 3J) were labelled by the gold particles; however, C 1–82 particles were less susceptible to recognition by the anti-C protein IgG, probably because of the deletion of the immunoreactive sequence. In the absence of primary antibodies or using pre-immune serum as the primary antibody, we did not observe gold particles at the surface of the NLPs.
Purification of NLPs formed from C 1–82

NLPs assembled from C 1–82 were further purified on a CsCl gradient (Fig. 5A). The presence of HCV C protein in the collected fractions was assayed by ELISA and confirmed by Western blot analysis. Particles collected from the major peak at a density of 1.32 g cm$^{-3}$ were denatured and visualized by SDS-PAGE (Fig. 5B, lane 2). Only one protein band was revealed by Coomassie blue staining and the apparent molecular mass corresponded to the C 1–82 protein. The major C protein fraction observed by EM contained uniform NLP structures, similar to those obtained previously (Fig. 5C). The second peak of protein with a high density corresponded to broken particles with darker centres (data not shown). HCV C protein was also detected in the low-density fractions collected from the CsCl experiment. These protein fractions possibly represented the non-encapsidated portion of the C protein or small oligomer complexes, precursor of the NLPs, since no capsids were detected by EM (data not shown).

In vivo assembly of the recombinant HCV C proteins in *P. pastoris*

The HCV C protein contains a hydrophobic sequence at the C terminus that anchors the protein to the ER membranes. It has been demonstrated that *P. pastoris* contains all the cellular machinery necessary to generate NLPs from the...
construct of the HCV full-length C protein sequence and part of the E1 protein (Acosta-Rivero et al., 2002). To investigate the formation of HCV-like particles produced in vivo, we transformed yeast cells with constructs containing full-length C protein under the control of a methanol-inducible promoter. Total yeast extract expressing C 1–191 contained a protein that cross-reacted with anti-C protein IgG (Fig. 6A, lane 2). We also generated yeast cells expressing the first 82 aa of the C protein alone or with the C-terminal ER-anchoring domain (Fig. 1C). The fusion of the ER protein signal (aa 170–191) to the first 82 aa allowed the production of the truncated protein in P. pastoris (Fig. 6A, lane 4). The apparent molecular mass of the protein after SDS-PAGE was similar to the C 1–82 protein produced in E. coli (Fig. 6A, lane 5). In the absence of the ER-targeting signal, no C 1–82 protein could be detected in total yeast extracts (Fig. 6A, lane 3) or that of different yeast transformants (data not shown). The presence and the integrity of the gene inside the cell were confirmed by producing a PCR product from chromosomal DNA (data not shown). We believed that the protein without the ER domain was more susceptible to degradation. However, the addition of proteosome inhibitor MG132 in the medium during induction did not increase C protein content (data not shown).

NLPs C 1–191 and C 1–82 ERa were enriched from the yeast lysate by sucrose cushion centrifugation after treatment with Sarkosyl and analysed by EM. We observed NLPs with both constructs (Fig. 6B and C). The NLPs formed from the truncated protein were smaller in size (mean diameter of 25 nm) than the full-length protein (32 nm) but exhibited a similar diameter to the C 1–82 NLPs generated after in vitro assembly (Fig. 5C). These structures were not observed in yeast cells expressing the vector only (data not shown). Immunogold labelling with anti-C protein antibodies was closely associated with both NLPs (Fig. 6B and C, inserts). These NLPs were totally membrane free, probably because detergent was used during the enrichment process. Non-enveloped NLPs have been observed in yeast cells expressing the polyprotein C–E1 (Falcon et al., 1999) and also in liver cells (Falcon et al., 2003) and the sera of infected patients (Maillard et al., 2001). It is likely that membrane-free NLPs were also produced in our yeast expression system.

**DISCUSSION**

*In vitro* assembly of the HCV C protein into NLPs using purified recombinant proteins from E. coli (Kunkel et al., 2001; Kunkel & Watowich, 2002) was recently reported. The particles were formed only in the presence of structured RNA stem–loops, such as the HCV IRES sequence or with small tRNA molecules. The recent findings that a tRNA-like domain is a general feature of IRES RNA, including the IRES of HCV (Lyons & Robertson, 2003), may explain the affinity of the C protein for tRNA. The NLPs presented in these reports were heterogeneous in size and larger than the expected wt HCV nucleocapsids, which have been reported to have a mean diameter of 35 nm (Andre et al., 2002; Falcon et al., 2003; Ishida et al., 2001). It is likely

---

**Fig. 4.** Histograms showing the distribution of the diameter of particles observed by EM. The diameter of 300 particles was precisely measured in each case from electron micrographs at a magnification of 120 000 ×. Diameters of NLPs derived from C 1–170 (A), C 1–112 (B), C 1–82 (C), 1–79 (D) and C 1–75 (E) are shown.
that the denaturation–renaturation process described in this study partially affected the folding of the protein, which led to the formation of larger agglomerates. In this paper, the removal of the C-terminal hydrophobic anchoring domain of the HCV C protein and the addition of a His tag increased the solubility of the protein and facilitated purification under native conditions. The recombinant protein expressing the 170 N-terminal residues of the C protein, which covers almost the entire sequence of the mature protein (179 aa) (McLauchlan et al., 2002), generated NLPs similar in size and appearance to the wt HCV nucleocapsid (Falcon et al., 2003; Ishida et al., 2001). Sucrose density gradients suggested that the C 1–170 protein is organized within the particles in a similar way to the wt HCV nucleocapsid, since the density of the C 1–170 NLPs was evaluated at 1·24 g ml⁻¹; the density of non-enveloped particles from infected sera have been estimated at 1·22–1·25 g ml⁻¹ (Ishida et al., 2001).

**Fig. 5.** Purification of NLPs of C 1–82 using CsCl density-gradient centrifugation. (A) Density of the fractions collected from the top of the tube together with quantification of the HCV C protein by ELISA using polyclonal rabbit anti-C protein IgG. (B) SDS-PAGE stained with Coomassie blue showing the protein fraction associated with the highest value obtained in ELISA before (lane 1) and after (lane 2) CsCl density-gradient centrifugation. (C) Electron micrograph of NLPs generated from the recombinant protein shown in (B) (lane 2). Bar, 50 nm.
We produced several versions of the HCV C protein with deletions to determine the smallest portion of the protein capable of assembly. We showed that the N-terminal 75 aa (C 1–75) were sufficient to trigger the formation of NLPs in vitro, while the construct C 1–71 failed. The C 1–75 protein contains only four additional amino acids (PEGR), which delimit the border of the second basic cluster of the C protein, which can play an important role in virus assembly. More than 70% of the structures observed with the C 1–75 protein were, however, aggregates with a diameter exceeding the expected range for the HCV nucleocapsid. The addition of a further four amino acids in the construct C 1–79 stabilized the particles and gave rise to uniform and structured particles resembling HCV nucleocapsids. Also, the NLPs produced from C 1–82 formed one band with a density of 1.32 g ml⁻¹ in a CsCl gradient.
suggesting that all the NLPs were uniform in diameter, appearance and density (Fig. 5). Based on these results, we suggest that the minimal assembly domain of the HCV C protein resides in the first 79 aa of the protein. Previous studies have described regions of the C protein involved in protein–protein interactions located in the C-terminal half (Nolandt et al., 1997; Yan et al., 1998). Our results do not exclude the fact that these regions may contribute to the stability of the particle inside the enveloped virion.

There is little sequence conservation among C proteins of the Flaviviridae family. However, when we compared the hydrophobicity plots for the C protein of several flaviviridae, we highlighted structural features that are common between all members of this family (Fig. 7), as also proposed by McLauchlan et al. (2002). The C proteins of the flaviviridae family are similar in size (mean size of 114 aa) and harbour two regions with a net positive charge separated by a conserved internal hydrophobic segment often rich in G residues (region 1a) (Mandl et al., 1988; Markoff et al., 1997). The N terminus of the HCV C protein presents the same arrangement for the first N-terminal 75 aa. An ER-anchoring region present at the C terminus is also common to all types. However, we noticed that the HCV C protein harbours additional regions (1b and 2), located between the second charged segment of region 1a and the ER-anchoring domain (region 3), that are not found in the C proteins of other flaviviruses (Fig. 7). This additional region is composed of two hydrophobic regions and one charged region. If we deleted most of the additional region of HCV C (C 1–82 ERa), we produced a protein of 103 aa with a similar arrangement to the other flaviviridae. We chose the N-terminal 82 residues because the in vitro NLPs produced with this construct appeared to be more stable than the C 1–79 NLPs. This protein generated uniform NLPs of 26 nm in vivo, confirming our in vitro data that the N terminus of C is sufficient for NLP formation. They appeared similar in size and appearance to the NLPs produced in vitro from the recombinant protein C 1–82 purified from E. coli. We could not detect the protein C 1–82 lacking the ER-anchoring domain in P. pastoris, probably because of its susceptibility to degradation by proteasomes (McLauchlan et al., 2002).

According to our results, we have assigned specific tasks for the elaboration of the capsid to the different domains of the C protein (Fig. 8). The C-terminal domain (region 3) involved in ER targeting (Santolini et al., 1994) and

**Fig. 7.** Schematic comparison of the C protein of several flaviviridae. +, Positively charged region; grey bars, hydrophobic region; black bars, ER-anchoring domain. YFV, yellow fever virus; JEV, Japanese encephalitis virus; DV, dengue virus; WNV, West Nile virus; TBEV, tick-borne encephalitis virus; BVDV, bovine viral diarrhoea virus; SFV, swine fever virus; HCV, hepatitis C virus; HCV C 1–82 ERa, truncated construct of the HCV C protein with the additional regions 1b and 2, found only in the HCV C protein, deleted.
maturation of the protein (Lemberg & Martoglio, 2002) appears to be essential to accumulate and concentrate the protein at this specific site to initiate the encapsidation process. The N-terminal domain, which includes the two charged regions (region 1a), is important for interaction with the nucleic acid and for protein interactions between the subunits to trigger the assembly process, as proposed by Li et al. (2003). The hydrophobic segment rich in G in region 1a that separates the two charged regions resembles the G-rich spacers that often separate functional protein domains. These are known to improve protein flexibility and rotation, which maximizes the interaction between the protein subunits and the packaged nucleic acid (Acharya & Varshney, 2002; Ilari et al., 2002). The next two regions of the C protein (1b and 2) are unique to the HCV C protein amongst the flaviviridae, even if the sequence homology between members of this group is not particularly important. Additionally, the HCV C protein harbours additional regions (1b and 2) that are dispensable for the formation of NLPs. These studies may be useful in the development of new antiviral compounds that interfere with capsid formation, as recently discovered for human immunodeficiency virus type 1 (Tang et al., 2003) and hepatitis B virus (Deres et al., 2003).

In summary, we have shown that a native soluble form of the HCV C protein expressed in E. coli can produce HCV NLPs in vitro that are similar in size, shape and density to native HCV capsids. We have presented evidence that the N-terminal half of the HCV C protein contains all the information necessary for assembly of the viral nucleocapsid in vitro and in vivo. Our results confirm that the N terminus of the HCV C protein is similar to the C protein of other flaviviridae, even if the sequence homology between members of this group is not particularly important. Additionally, the HCV C protein harbours additional regions (1b and 2) that are dispensable for the formation of NLPs. These studies may be useful in the development of new antiviral compounds that interfere with capsid formation, as recently discovered for human immunodeficiency virus type 1 (Tang et al., 2003) and hepatitis B virus (Deres et al., 2003).

ACKNOWLEDGEMENTS

We thank the Canadian Institute of Health Research of Canada (CIHR) for funding our research programme on hepatitis C virus.

REFERENCES


