2618–2633 *Nucleic Acids Research, 2006, Vol. 34, No. 9*
doi:10.1093/nar/gkl240

Analysis of hepatitis C virus RNA dimerization and core–RNA interactions

Roland Ivanyi-Nagy, Igor Kanevsky[1], Caroline Gabus, Jean-Pierre Lavergne[2], Damien Ficheux[2], François Penin[2], Philippe Fossé[1] and Jean-Luc Darlix*

LaboRetro, Unité INSERM de Virologie Humaine, Ecole Normale Supérieure de Lyon, IFR 128 Biosciences Lyon-Gerland, 69364 Lyon Cedex 07, France, [1]CNRS-UMR 8113, LBPA-Alembert, Ecole Normale Supérieure de Cachan, 94235 Cachan Cedex, France and [2]Institut de Biologie et Chimie des Protéines, CNRS-UMR 5086, Université Claude Bernard Lyon I, IFR 128 Biosciences Lyon-Gerland, 69367 Lyon Cedex 07, France

Received January 16, 2006; Revised February 8, 2006; Accepted March 29, 2006

ABSTRACT

The core protein of hepatitis C virus (HCV) has been shown previously to act as a potent nucleic acid chaperone *in vitro*, promoting the dimerization of the 3′-untranslated region (3′-UTR) of the HCV genomic RNA, a process probably mediated by a small, highly conserved palindromic RNA motif, named DLS (dimer linkage sequence) [G. Cristofari, R. Ivanyi-Nagy, C. Gabus, S. Boulant, J. P. Lavergne, F. Penin and J. L. Darlix (2004) *Nucleic Acids Res.*, 32, 2623–2631]. To investigate in depth HCV RNA dimerization, we generated a series of point mutations in the DLS region. We find that both the plus-strand 3′-UTR and the complementary minus-strand RNA can dimerize in the presence of core protein, while mutations in the DLS (among them a single point mutation that abolished RNA replication in a HCV subgenomic replicon system) completely abrogate dimerization. Structural probing of plus- and minus-strand RNAs, in their monomeric and dimeric forms, indicate that the DLS is the major if not the sole determinant of UTR RNA dimerization. Furthermore, the N-terminal basic amino acid clusters of core protein were found to be sufficient to induce dimerization, suggesting that they retain full RNA chaperone activity. These findings may have important consequences for understanding the HCV replicative cycle and the genetic variability of the virus.

INTRODUCTION

Hepatitis C virus (HCV) infection is a major health and economic problem with >170 million people chronically infected worldwide (1). In the majority (~80%) of cases, HCV infection leads to viral persistence, eventually entailing serious sequalae, including chronic hepatitis, steatosis, cirrhosis and hepatocellular carcinoma (2). HCV is a small, enveloped virus with a single-stranded RNA genome of positive polarity. Based on its distinct features, it has been classified as the sole member of the genus Hepacivirus in the *Flaviviridae* family (3) which include two other genera with important human and animal pathogens: flaviviruses (e.g. West Nile virus, dengue virus, yellow fever virus) and pestiviruses (e.g. bovine viral diarrhoea virus and classical swine fever virus). In addition, the *Flaviviridae* family include yet unassigned viruses named GBV-A, GBV-B and GBV-C. GBV-B, the genetically closest relative of HCV is hepatotropic and causes acute and chronic hepatitis in New World monkeys (4,5).

The 9.6 kb genome of HCV contains a single long open reading frame (ORF), flanked by highly conserved 5′- and 3′-untranslated regions (5′- and 3′-UTRs). These regions are highly structured and constitute *cis*-acting elements that play essential roles in the viral replicative cycle [reviewed in (6)]. The 5′-UTR contains an IRES (internal ribosome entry segment) that directs cap-independent translation of the viral polyprotein (7,8), and an overlapping RNA signal that is necessary for viral replication (9,10). The 3′-UTR has a tripartite organization, being composed of a ~40 nt long variable region, a poly(U/UC) tract, and a highly conserved 98 nt segment, named X RNA (11–14). Except for the variable region, 3′-UTR structures were shown to be necessary for viral replication and productive infection (15–19). Translation of the ORF yields a large polyprotein which is co-translationally and post-translationally cleaved into at least 10 proteins: the structural proteins (core, E1 and E2 glycoproteins), the hydrophobic peptide p7 and the non-structural proteins NS2 to NS5B (20).

The core protein of HCV is a small (~21 kDa), highly basic RNA-binding protein generated in a mature form via successive cleavages of the viral polyprotein (21–23). The

*To whom correspondence should be addressed. Tel: +33 4 72 72 81 69; Fax: +33 4 72 72 87 77; Email: Jean-Luc.Darlix@ens-lyon.fr

mature form of core is a dimeric alpha-helical protein exhibiting membrane binding properties via its C-terminal domain D2 (24). The N-terminal part of core (domain D1) was shown to bind to different regions of the HCV genomic RNA, as well as to cellular RNA molecules (25–27). In the presence of highly structured RNAs, such as tRNA or the 5′-UTR of HCV, recombinant core protein was sufficient for particle formation *in vitro* (28–31). Thus, core protein is believed to bind the HCV genomic RNA, leading to genome packaging and formation of the viral nucleocapsid. Besides its structural role in the viral particle, core has also been implicated in a variety of cellular processes, including regulation of gene expression, cell apoptosis, signalling and lipid metabolism (32–34).

Previously, we have shown that the core protein possesses potent nucleic acid chaperone activities, similarly to retroviral NC proteins (35). RNA chaperone proteins facilitate the correct folding of RNA molecules by preventing their misfolding or by resolving misfolded RNA species (36–38). They bind nucleic acids with broad sequence specificity and mediate RNA conformational changes in an ATP-independent manner. In cells, a variety of proteins (e.g. hnRNP A1, YB-1/p50, FMRP) carry out this essential function, thereby regulating transcription, RNA processing, transport and translation (37). RNA viruses also take advantage of RNA chaperones for structural rearrangements necessary for the translation and replication of their genomes. These proteins are either encoded by the viruses themselves or 'hijacked' from the host cells, as in the case of the La protein, which is utilized by numerous viruses—including HCV—to regulate IRES-mediated translation (39,40). Viral-encoded RNA chaperones include also the small nucleocapsid (NC) proteins of retroviruses that mediate critical structural rearrangements of the retroviral genome required for replication and virus assembly (41,42).

We have reported earlier that RNA chaperoning by core results in important structural rearrangements of the 3′-UTR of the HCV genome, leading to RNA dimerization *in vitro* (35). The RNA signal mediating dimerization involves a 16 nt long palindrome (named dimer linkage sequence or DLS), located in the highly conserved 3′ region, termed X RNA. In this report, we further characterized the core protein-mediated dimerization of the HCV UTR regions. We show that both the plus-strand 3′-UTR and the complementary negative-strand RNA dimerize in the presence of the core protein, and that the DLS is probably the sole determinant of dimerization. Furthermore, based on enzymatic and chemical probing of wild-type and mutant RNAs, we provide a structural model for monomeric and dimeric X RNAs, and propose a mechanism for the dimerization process.

MATERIALS AND METHODS

Plasmid DNA construction

Plasmid pSP64-3′UTR was constructed by inserting a PCR product [generated using pVan37 (35) and oligonucleotides GAEL66 and SP6], containing a T7 promoter followed by the 3′-UTR region of HCV isolate Con1 (GenBank accession no. AJ238799), between the SalI and HindIII sites of pSP64 (Promega). The resulting plasmid thus contains the HCV 3′-UTR flanked by T7 and SP6 promoters, making possible the *in vitro* transcription of plus- and minus-strand UTR RNAs, respectively. Point mutations were introduced in pSP64-3′UTR by using a PCR-based mutagenesis protocol (43), with three common oligonucleotides (mut, SP6 and T7) and one mutation-specific oligonucleotide (pal-1–pal-11) for each mutant. To obtain plasmids containing the 3′ X region, plasmid pX(+) and pX(+)-Pal-4 were generated by PCR using oligonucleotides PF1 and PF2, and plasmids pSP64-3′UTR and pMutPal-4, respectively. Plasmid pX(−) and pX(−)-Pal-4 were generated by PCR using oligonucleotides PF3 and PF4, and plasmids pSP64-3′UTR and pMutPal-4, respectively. The resulting PCR products were digested with EcoRI and HindIII and ligated into pSP64-3′UTR digested with the same enzymes. For all PCRs, the Expand High Fidelity PCR System (Roche Molecular Biochemicals) was used, and DNA constructs were verified by sequencing. Oligonucleotides used in this study are listed in Table 1.

In vitro RNA synthesis and purification

DNA templates for *in vitro* transciption of plus- and minus-strand HCV RNAs were generated by digestion of the appropriate plasmids with HindIII or FokI, respectively. RNAs were prepared by *in vitro* transcription with T7 or SP6 RNA polymerase, according to the manufacturer's instructions (RiboMax kit, Promega), labelled by incorporation of [^{32}P]UMP during *in vitro* transcription, and purified by PAGE in 7 M urea, 50 mM Tris–borate, pH 8.3 and 1 mM EDTA (0.5× TBE), followed by elution and ethanol precipitation. Plasmids pX(+) and pX(+)-Pal-4 were linearized with ScaI to generate templates for the *in vitro* synthesis of X(+) and X(+)-Pal-4 RNAs, respectively. Plasmids pX(−) and pX(−)-Pal-4 were linearized with SmaI to generate templates for the *in vitro* synthesis of X(−) and X(−)-Pal-4 RNAs, respectively. X RNAs were transcribed *in vitro* using the T7-MEGAshortscript™ high yield transcription kit (Ambion) and purified by electrophoresis on a denaturing 12% PAGE as described previously (44). RNAs were treated with calf intestinal alkaline phosphatase (Roche Molecular Biochemicals) and 5′ end-labelled using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP. The 5′ end-labelled RNAs were purified by electrophoresis on a denaturing 12% PAGE and recovered by elution, followed by ethanol precipitation.

Protein expression and peptide synthesis

Full-length HCV core protein (amino acids 2–169) was expressed in *Escherichia coli* and purified as described previously (24,35). Peptides 2BD and 3BD (Figure 3) were synthesized on a Milligen 9050 apparatus with Fmoc-OH/DIC/Hobt chemistry. The peptides were cleaved by a TFA solution with classical scavengers and precipitated in diethyl ether. The precipitate was centrifuged and the pellet was solubilized in water and lyophilised. The crude peptides were dissolved in water and purified on Vydac column (C18, 5 μm, 250 × 10 mm) in TFA 0.1% (A) with an appropriate gradient of B (70% acetonitrile, 0.09% TFA). Purified peptides were characterized by electrospray mass spectrum (SCIEX API 165) at 6696 and 9385 UMA for peptides 2BD and 3BD, respectively, and by RP-HPLC on analytical column Vydac

Table 1. Oligonucleotides used in this study

Name	Sense	Sequence (5′ to 3′)
GAEL66	(+)	GGGCTCGAGTAATACGACTCACTATAGGGACGGGGAGCTAAACACTCCAGGCC
mut	(+)	CAGAGGAGCAGGCCAATAGGCCATC
SP6	(−)	TATTTAGGTGACACTATAG
T7	(+)	TAATACGACTCACTATAGGG
pal-1	(−)	GGACCTTT**CACtGgTAGCCGTGA**CT
pal-2	(−)	TT**TCACAGCTAGgCGTcA**CTAGGGCTAAG
pal-3	(−)	TCACGGACCTT**aCACAGCTAGCC**
pal-4	(−)	ACCTT**TCACAGgTAGCCGTGA**CT
pal-5	(−)	ACCTT**TCACAGaTAGCCGTGA**CT
pal-6	(−)	CTT**TCACAGCTAtCCGTGA**CTAGG
pal-7	(−)	CTT**TCACAGCTAcCCGTGA**CTAGG
pal-8	(−)	ACCTT**TCACAGCaAGCCGTGA**CTA
pal-9	(−)	ACGGACCTT**TCtCtGCTAGCaGaGACT**AGGGCTA
pal-10	(−)	GGC**TCACGGAggTTTCACAGCTAGCCGTGA**CTAGacCTAAGATGGAGC
pal-11	(−)	GGCTCACGGAggTT**TCACAGgTAGCCGTGA**CTAGacCTAAGATGGAGC
PF1	(+)	GAAGAATTCTGGAGTAATACGACTCACTATAGGTGGCTCCATCTTAG
PF2	(−)	CCGACGCCAAGCTTAGTACTTGATCTGCAGAGAG
PF3	(−)	GAAGAATTCTGGAGTAATACGACTCACTATAGACTTGATCTGCAGAGAG
PF4	(+)	CATGACAAGCTTCCCGGGTGGCTCCATCTTAG
PF5	(−)	CAGTCAAGCGGCTCACG
PF6	(+)	GGTGGCTCCATCTTAGC

Mutations introduced in the X RNA region are in lower case. The T7 promoter sequence is underlined and the DLS is in boldface.

(C18, 5 μm, 250 × 4.6 mm) using a 30 min gradient from 10 to 90% of B.

Dimerization assay

FL and/or X RNAs were incubated with proteins (at protein to nucleotide molar ratios as indicated in the figure legends) in 10 μl Core buffer (20 mM Tris–HCl, pH 7.0, 30 mM NaCl, 0.1 mM $MgCl_2$ and 10 μM $ZnCl_2$) containing 8 U RNasin (Promega) at 37°C for 10 min. Reactions were stopped by adding SDS–EDTA solution (at 0.5% SDS and 5 mM EDTA final concentrations). Proteins were removed by proteinase K digestion (2 μg) at room temperature for 15 min. RNAs were phenol-chloroform extracted and analysed by PAGE under native conditions, followed by autoradiography.

Analysis of the thermal stability of dimeric RNAs

X(+) RNA or X(−) RNA (8 pmol) were incubated in the presence of 3BD (40 pmol) and 64 U RNasin (Promega) for 10 min at 37°C in a final volume of 80 μl Core buffer. RNAs were recovered by phenol-chloroform extraction. Aliquots (10 μl) of each sample were incubated for 7 min at temperatures ranging from 20 to 65°C, followed by addition of 2.5 μl of loading buffer (50 w/v% glycerol, 0.05 w/v% bromophenol blue, 0.05 w/v% xylene cyanol). The aliquots were loaded on 1.5% agarose (Seakem®LE agarose) gel and electrophoresis was carried out at 20°C in 0.5× TBE buffer. The percentage of dimerization was determined as described previously (45). The labelled RNA was detected by autoradiography as described previously (45) and quantified with a Storm 840 (Molecular Dynamics).

Enzymatic RNA probing

Assays were carried out in a final volume of 10 μl. ^{32}P-labelled RNA (1 pmol) was dissolved in 6 μl of water, heated at 90°C for 1 min and chilled on ice. Then, 8 U RNasin and Core buffer were added (final concentrations: 20 mM Tris–HCl, pH 7.0, 30 mM NaCl, 0.1 mM $MgCl_2$ and 10 μM $ZnCl_2$) and the sample was incubated in the presence or absence of 3BD (5 pmol) for 10 min at 37°C. RNA was recovered by phenol-chloroform extraction when 3BD was removed before enzymatic probing. To equalize the quantity of RNA in each assay, 1 μg of a tRNA mixture from *E.coli* MRE 600 (Roche Molecular Biochemicals) was added to the sample before the addition of a ribonuclease. RNA was incubated with RNases T1 (Life Technologies), T2 (Life Technologies) or V1 (Ambion) for 7 min at 37°C. The cleavage reactions were stopped by extraction with phenol-chloroform followed by precipitation in ethanol. The dry pellets were dissolved in 7 M urea, 0.03 w/v% bromophenol blue and 0.03 w/v% xylene cyanol and loaded on denaturing 10 or 12% polyacrylamide gels. The cleavage sites were identified by running in parallel the products of alkaline hydrolysis and limited RNase T1 digestion of the same RNA. Alkaline hydrolysis was performed at 90°C for 5 min in 45 mM $NaHCO_3/Na_2CO_3$ (pH 8.9). Partial RNase T1 digestion was performed with 0.1 U of the enzyme at 50°C for 5 min in 13 mM sodium citrate (pH 5), 3 M urea, 0.6 mM EDTA, 0.03 w/v% bromophenol blue and 0.02 w/v% xylene cyanol.

Chemical RNA probing with CMCT

1-Cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT) was purchased from Sigma–Aldrich. Assays were carried out in a final volume of 10 μl. RNA (1 pmol) was dissolved in 6 μl of water, heated at 90°C for 1 min and chilled on ice. Then, 8 U RNasin and probing buffer were added (final concentrations: 20 mM HEPES–KOH, pH 7.8, 30 mM NaCl, 0.1 mM $MgCl_2$ and 10 μM $ZnCl_2$) and the sample was incubated in the presence or absence of 3BD (5 pmol) for 10 min at 37°C. RNA was recovered by phenol-chloroform extraction when 3BD was removed before CMCT probing. To equalize the quantity of RNA in each assay, 1 μg of a tRNA mixture from *E.coli* MRE 600 (Roche Molecular Biochemicals) was

added to the sample before the addition of CMCT. RNA was incubated with 17 mg/ml of CMCT for 5 and 10 min at 37°C. The CMCT treatment was stopped by adding 30 µl ethanol and 1 µl of 3 M sodium acetate. After ethanol precipitation, the dry pellet was dissolved in water and the modified bases were identified by primer extension with avian myeloblastosis RT (Invitrogen) as described previously (46). ODN PF5 complementary to X(+) and X(+)-Pal-4 RNAs, and PF6 complementary to X(−) and X(−)-Pal-4 RNAs were used as primers and were 5′ end-labelled using T4 polynucleotide kinase and [γ-^{32}P]ATP. The samples were analysed by electrophoresis on a denaturing 8% polyacrylamide gel.

RESULTS

Core protein mediates dimerization of plus- and minus-strand HCV RNAs *in vitro*

Previously, we have shown that a 16 nt long palindrome (5′-UCACGGCUAGCUGUGA-3′; Figure 1B and C) is involved in the dimerization of the HCV plus-strand 3′-UTR (35). By analogy with retroviruses, we named this palindromic region DLS. This RNA motif is universally present in the X RNA region (Figure 1A and B) of all HCV sequences reported so far, suggesting that it plays an important role in the viral life cycle.

Since HCV replicates via a negative-strand RNA intermediate, which in turn serves as a template for plus-strand RNA synthesis, we examined whether the minus-strand genomic RNA would be also able to dimerize. RNAs corresponding to the full-length plus-strand 3′-UTR or to the complementary minus-strand RNA [FL(+) and FL(−), respectively], and to the X regions [X(+) and X(−) RNAs] were synthesized by run-off transcription and incubated for 10 min at 37°C in the presence of increasing concentrations of core protein. After dissociation of the core from the RNA by extensive proteinase K digestion and SDS treatment prior to phenol-chloroform extraction, RNAs were resolved on a native polyacrylamide gel (Figure 1D). Incubation with core protein induced the appearance of a new RNA species both for FL(+) and FL(−) (Figure 1D, lanes 1–5), as well as for X(+) and X(−) RNAs (Figure 1D, lanes 6–10). The resulting bands likely correspond to a new structural form of the RNA molecules induced by the interaction with the core protein. Accordingly, when FL and X RNAs were incubated together with core protein (Figure 1D, lanes 11–15), a new intermediate-sized band appeared, corresponding to FL-X heterodimers. These data clearly demonstrate that the slowly migrating RNA forms result from dimerization of the UTR RNAs, induced by the core protein.

Analysis of the thermal stability of dimeric RNAs showed that the X(+) RNA dimers are more stable than the X(−) RNA dimers. Indeed, the X(+) dimers melt at 55–60°C while the X(−) dimers melt at ∼50°C (Supplementary Figure 1). This may be due to mismatches in the minus-strand DLS dimer (Figure 1C).

Sequence determinants for HCV RNA dimerization

In order to map the major determinants of HCV UTR RNA dimerization, we introduced point mutations at various positions in the DLS and monitored the dimerization of the resulting UTR (FL) RNAs at increasing core protein/RNA molar ratios. The mutations and their influence are summarized in Figure 2A. Representative results are shown for FL(+) RNAs in Figure 2B and for FL(−) RNAs in Figure 2C.

Based on the most stable FL(+) RNA structure predicted by the Mfold program version 3.2 (47), the DLS(+) forms an apical stem–loop (Figure 1B, right-hand structure and Figure 2A) where the central [$_{29}$CUAG$_{32}$] is unpaired, while the two ends of the palindrome constitute the stem. Importantly, even single point mutations in the loop completely abolished core-mediated dimerization [Figure 2B, Pal-4(+)–Pal-7(+)], except for Pal-8(+) where the RNA with a [$_{29}$CUUG$_{32}$] loop was still able to dimerize, albeit with reduced efficiency. In Pal-8(+), formation of non-canonical U–U base pairs may account for the observed dimerization. Indeed, in internal loops closed by G–C base pairs, U–U pairs may be stabilizing by forming hydrogen bonds (48). Moreover, complementarity in the DLS(+) stem is also important for dimerization, as evidenced by Pal-2(+) which did not dimerize upon disruption of the stem. In contrast, when complementarity is preserved by introducing symmetrical mutations [Pal-9(+)], or by *trans*-complementation (Supplementary Figure 2), dimerization still occurred. Interestingly, we observed a slight, but nevertheless reproducible decrease in the dimerization efficiency of Pal-9(+) RNA compared with the wild type. This may be due to the stabilization of the monomeric structure [−7.4 kcal/mol calculated for the region 23–38 in Pal-9(+) compared with −6.5 kcal/mol for the wild type] by changing a G–U base pair to A–U, resulting in a higher energy barrier for the disruption of the original base pairing.

The profound effect of mutations observed both in the stem and in the loop region indicate formation of an extended duplex and suggest that a 'kissing loop' interaction alone is not sufficient for stable dimerization. In agreement with this, we were unable to observe dimerization of wild-type(+) RNA in the presence of 5 mM Mg^{2+} and migration at 4°C, conditions described to facilitate kissing loop formation and stabilization (49).

The most stable predicted structure for DLS(−) consists of an 8 nt loop and a shorter stem than for the complementary sequence. Nevertheless, dimerization of mutant FL(−) RNAs (Figure 2C) mirrored the pattern observed for the FL(+) RNAs, indicating that the same determinants play a role in DLS(+) and DLS(−) RNA dimerization.

Core domains promoting HCV RNA dimerization

The HCV core protein consists of a basic N-terminal RNA binding domain with three highly charged amino acid clusters (Figure 3A), and a hydrophobic C-terminal domain that is probably involved in the intracellular targeting of the protein to lipid droplets (24,50–52). The N-terminal region (amino acids 2–117) alone proved to be active in RNA chaperone assays and promoted UTR dimerization efficiently [(35) and data not shown]. To gain further insights into the determinants of core chaperoning, we examined the nucleic acid chaperone activity of the N-terminal basic clusters of HCV core protein. Peptides 3BD and 2BD, corresponding to all three or the second and third highly basic clusters,

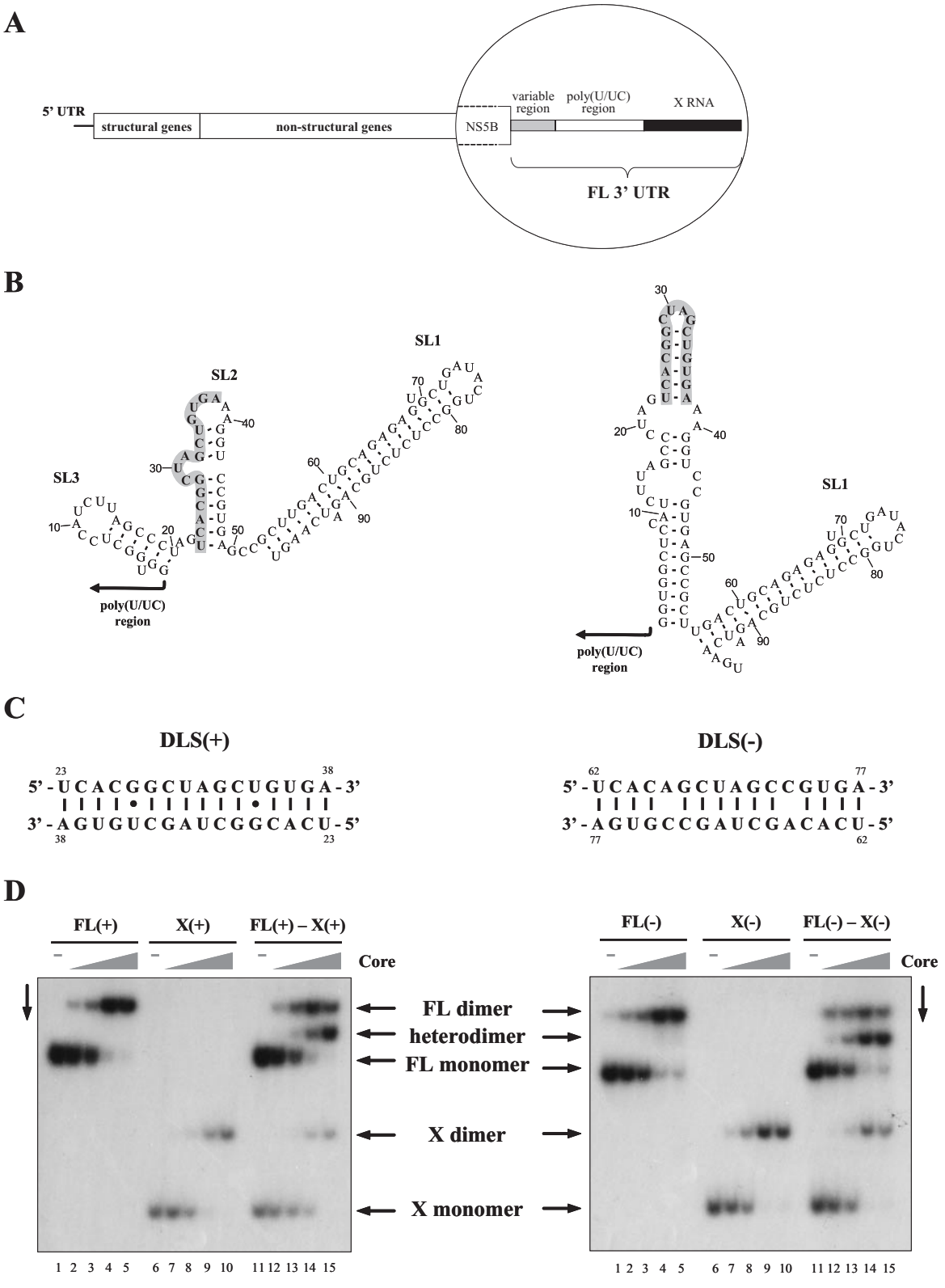
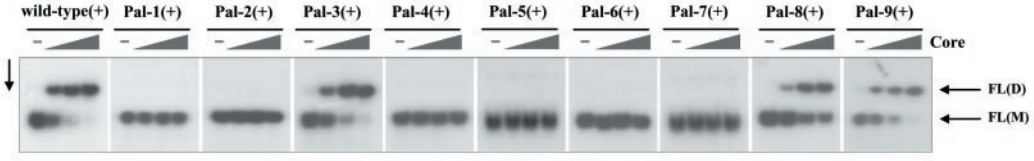


Figure 1. (**A**) Organization of the HCV genome. The three regions of the 3′-UTR [variable region, poly(U/UC) tract and X RNA] are illustrated in the inset. (**B**) Alternative stable conformations of X RNA, predicted by the Mfold program (47). The palindromic region is illustrated in boldface on a grey background. Numbering begins at the 5′ first nucleotide of the X RNA. (**C**) DLS(+) and DLS(−) sequences, illustrating the potential base pairing interactions in the dimeric form. Numbering begins at the 5′ end of X(+) and X(−) RNA, respectively. (**D**) Heterodimerization of plus-strand full-length 3′-UTR [FL(+)] and X(+) RNAs, and of the complementary minus-strand RNA regions [FL(−) and X(−)] in the presence of core protein. The full-length UTR (lanes 1–5), X RNA (lanes 6–10), or both (lanes 11–15) were incubated with the core protein at protein to nucleotide molar ratios of 0, 1:80, 1:40, 1:20 and 1:10. After 10 min incubation at 37°C, core protein was removed by SDS treatment and proteinase K digestion and RNA were analysed by native gel electrophoresis and autoradiography.

A

	Plus-strand		Minus-strand	
	Mutations	Dimerization	Mutations	Dimerization
Wild-type	-	+++	-	+++
Pal-1	^{32}G→C ^{34}U→A	-	^{66}A→U ^{68}C→G	-
Pal-2	^{24}C→G ^{28}G→C	-	^{72}C→G ^{76}G→C	-
Pal-3	^{38}A→U	+++	^{62}U→A	+
Pal-4	^{32}G→C	-	^{68}C→G	-
Pal-5	^{32}G→U	-	^{68}C→A	-
Pal-6	^{29}C→A	-	^{71}G→U	-
Pal-7	^{29}C→G	-	^{71}G→C	-
Pal-8	^{31}A→U	++	^{69}U→A	-
Pal-9	^{25}A→U ^{27}G→U ^{34}U→A ^{36}U→A	+++	^{64}A→U ^{66}A→U ^{73}C→A ^{75}U→A	++

B



C

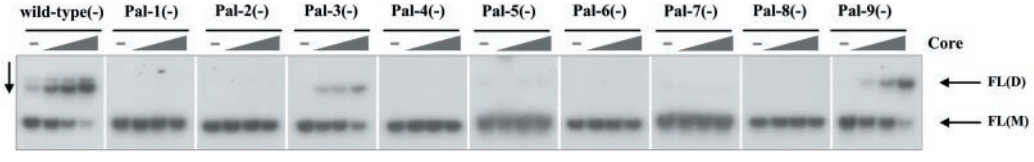


Figure 2. Influence of DLS mutations on HCV RNA dimerization. (**A**) Mutations introduced in the X RNA region, and their effect on plus-strand 3′-UTR and complementary minus-strand RNA dimerization. Mutations are highlighted with black background circles. For the sake of clarity, potential conformational changes in the DLS resulting from the mutations are not illustrated here (for details see Supplementary Figure 3). Numbering begins at the 5′ end of X(+) and X(−) RNA, respectively. (**B**) Dimerization of wild-type and mutant FL(+) RNAs. ^{32}P-labelled RNA molecules were incubated with increasing amounts of core protein (at protein to nucleotide molar ratios of 0, 1:80, 1:40 and 1:20). Following incubation, core protein was removed and RNA were analysed by native gel electrophoresis and autoradiography. FL(D) corresponds to dimeric RNA, while FL(M) to the monomeric form. (**C**) Dimerization of wild-type and mutant FL(−) RNAs. Experiments were done as in (B).

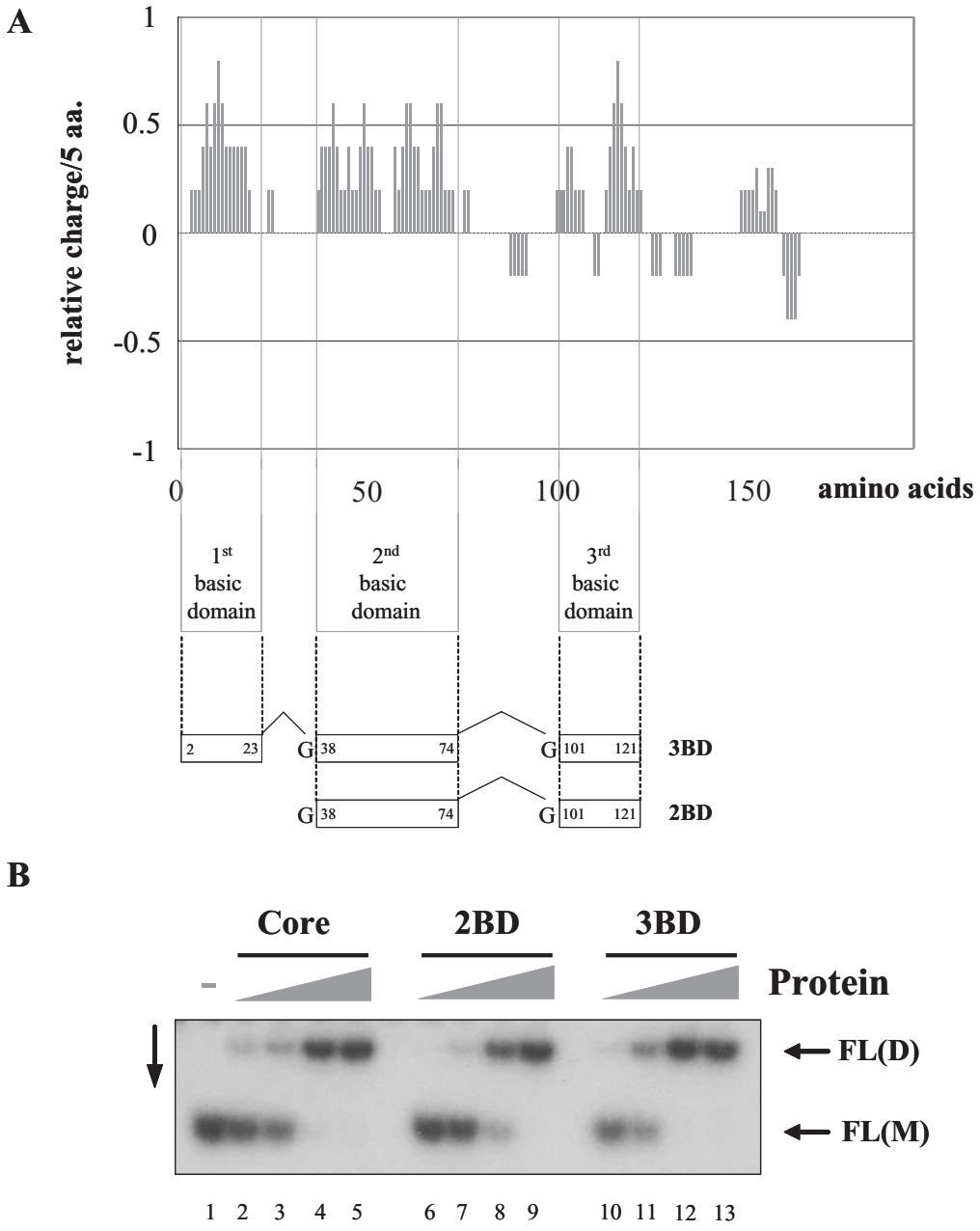


Figure 3. Activity of core peptides in RNA dimerization. (**A**) The bar chart illustrates the charge distribution of amino acids in the core protein, as calculated by the charge function of the EMBOSS package, using default parameters and a sliding window of five amino acids. Synthetic peptides 2BD and 3BD correspond to two or three basic amino acid clusters of the core protein, respectively. (**B**) Dimerization of FL(+) RNAs induced by core peptides 2BD and 3BD. ^{32}P-labelled RNA molecules were incubated with increasing amounts of the corresponding polypeptides (at polypeptide to nucleotide molar ratios 1:80, 1:40, 1:20 and 1:10). Following incubation, polypeptides were removed and RNA were analysed by native gel electrophoresis and autoradiography. FL(D) corresponds to dimeric RNA, while FL(M) to the monomeric form.

respectively (Figure 3A), were chemically synthesized and purified as described in Materials and Methods.

Dimerization of wild-type FL(+) RNA was monitored at increasing peptide to nucleotide molar ratios (Figure 3B). Intriguingly, the ability of 2BD and 3BD to dimerize HCV UTR RNAs was close to or equal, respectively, to that of the full-length core protein. Importantly, this dimerization reflects the RNA chaperone activity of the peptides, and cannot be accounted for by simple charge neutralization, since a highly basic peptide (pI = 9.9) lacking chaperone activity was not able to direct FL(+) dimerization (35).

Experimental strategy for the structural probing of the HCV X RNAs

In order to better understand the mechanism of core-mediated dimerization, we analysed the structure of wild-type plus- and minus-strand X RNAs in their monomeric and dimeric forms.

In parallel experiments, we also analysed the structure of a dimerization-deficient mutant RNA (Pal-4) which was selected because this point mutation completely abolishes HCV replication in a subgenomic replicon system (18). All four RNAs, in the presence or absence of the 3BD peptide, were probed with RNases T1, T2 and V1 at 37°C under conditions described in Materials and Methods. We used the 3BD peptide for the probing experiments for several reasons. Firstly, it was available in quantities large enough to perform several times the experiments. Secondly, the core basic clusters behave similarly to the full-length core with respect to RNA binding (data not shown), 3′-UTR dimerization (Figure 3B) and capsid assembly (30). Since the N-terminal region of core protein is essentially unfolded in the absence of RNA (24,53), deletion of the linker regions between the basic clusters probably does not change the structure of the core protein–RNA complex.

RNase T1 is specific for unpaired guanines, while RNase T2 cleaves single-stranded regions with a preference for adenines and RNase V1 is specific for double-stranded and stacked regions. The cleavage positions were identified using 5′ end-labelled RNAs. To investigate the DLS, we also used CMCT that modifies N3 position of uridines and, at lower rate, N1 position of guanines. The modified bases were identified by primer extension (Materials and Methods). Representative examples of probing experiments are shown in Figures 4 and 6, and the results of a series of independent experiments are summarized in Figures 5 and 7. To generate the secondary structural models, we selected the foldings predicted by the Mfold program (47) that were the most consistent with the experimental data.

To identify possible conformational changes and protections induced by 3BD in the X region, we compared the enzymatic and CMCT probing patterns of the four RNAs in the presence or absence of 3BD. To discriminate between reactivity changes induced by RNA dimerization and those resulting from the binding of 3BD, enzymatic and CMCT probings of wild-type RNAs were also performed after removal of 3BD by extraction with phenol-chloroform. We verified that after phenol-chloroform extraction and incubation at 37°C, the temperature used for RNA probing, the dimers did not dissociate (Supplementary Figure 1).

Structural probing of wild-type X(+) and mutant X(+)-Pal-4 RNAs

The probing data are consistent with the formation of stem–loop SL1 and a structural heterogeneity for the 1–20 sequence in the X(+) and X(+)-Pal-4 RNAs under both the monomeric and dimeric forms (Figure 5). Furthermore, as observed previously (54,55), our probing data support a structural heterogeneity for the 21–55 region of the monomeric X(+) RNA (Figure 5A). Interestingly, we found that 3BD induced a significant increase in RNase T1 sensitivity for G_{46} and G_{48}, and a strong decrease in RNase T1 sensitivity for guanine residues in the DLS of the X(+) RNA (Figure 4A, lanes 3BD). In addition, new RNAse V1 cleavages sites (G_{28}, U_{30}, C_{33} and U_{34}) appeared in the DLS upon incubation with 3BD. The RNase cleavage patterns of the dimeric X(+) RNA, following removal of 3BD, were identical to those obtained in the presence of 3BD (Figure 4A, compare lanes 3BD and 3BD rm), in agreement with the proposed chaperoning role of core, where the protein is only required to establish the dimer but not to maintain it. Taken together, our results support a secondary structural model in which the reactivity changes are caused only by RNA dimerization through the DLS (Figure 5B).

In contrast with the wild-type sequence, the probing data suggest that the 21–55 region in the monomeric X(+)-Pal-4 RNA forms a unique stem–loop structure (Figure 5C), which is not changed in the presence of the 3BD peptide. Thus, a possible explanation for the inability of Pal-4 RNAs to dimerize could be that the DLS is 'trapped' in a conformation favouring the monomeric structure. In order to discriminate between the effects of sequence versus structural changes on dimerization, we introduced additional point mutations in Pal-4 outside the DLS region that would strongly favour formation of the alternative SL2 structure, where the DLS(+) constitutes an apical stem–loop (Pal-11 RNA in Supplementary Figure 4). However, the second-site mutations did not restore dimerization of the 3′-UTR, indicating that correct folding and DLS sequence both play a role in dimerization.

Structural probing of wild-type X(−) and mutant X(−)-Pal-4 RNAs

The probing data are consistent with the proposed secondary structural models for the X(−) and X(−)-Pal-4 RNAs (Figure 7A and C). Interestingly, we found that 3BD induced a strong decrease in RNase T1 and T2 sensitivity for G_{67}, G_{71}, G_{76}, U_{69} and A_{70} that lie within the DLS of the X(−) RNA (Figure 6A, lanes 3BD, and Figure 7A). In addition, new RNAse V1 cleavage sites (U_{69}, C_{73} and A_{80}) appeared upon incubation with 3BD or in the dimeric X(−) RNA after removal of 3BD (Figure 6A, lanes 3BD, and 3BD rm). V1 cleavage at U_{69} and C_{73} is compatible with the formation of an extended duplex that links the two subunits in the dimeric RNA (Figure 7B). However, our probing data do not exclude the possibility that the linkage structure is restricted to a loop–loop complex involving the DLS sequence. It was surprising that moderate T1 and T2 cleavages occurred within the DLS when 3BD was extracted from the dimer (Figure 6A). A possible explanation is that 3BD stabilizes the extended duplex that contains two mismatches. The helix between the mismatches may become less stable and more sensitive to single-strand specific RNases after removal of the core peptide.

In the structure of the X(−)-Pal-4 RNA, G_{91} and G_{92} are predicted to lie within a short stem, but they were a moderate T1 target (Figure 7C). This observation suggests that the base pairing interaction between nucleotides $_{78}CUA_{80}$ and $_{90}UGG_{91}$ is transient. The RNase cleavage patterns of the X(−)-Pal-4 RNA were not changed in the presence of 3BD (Figure 6B).

DISCUSSION

RNA chaperone activity of HCV core protein and X RNA dimerization

The HCV core protein was found to bind nucleic acids in a relatively sequence-independent manner and to chaperone

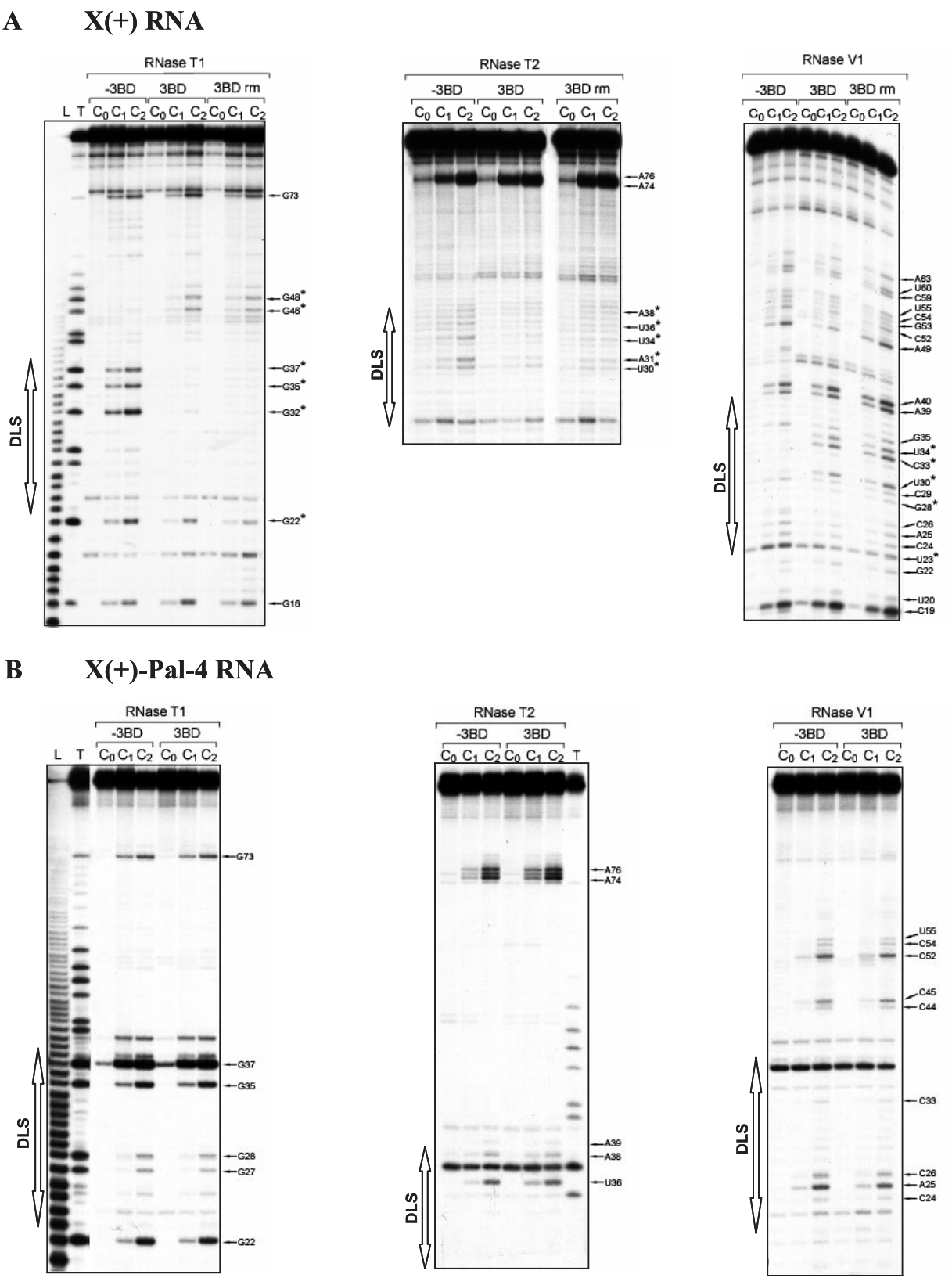
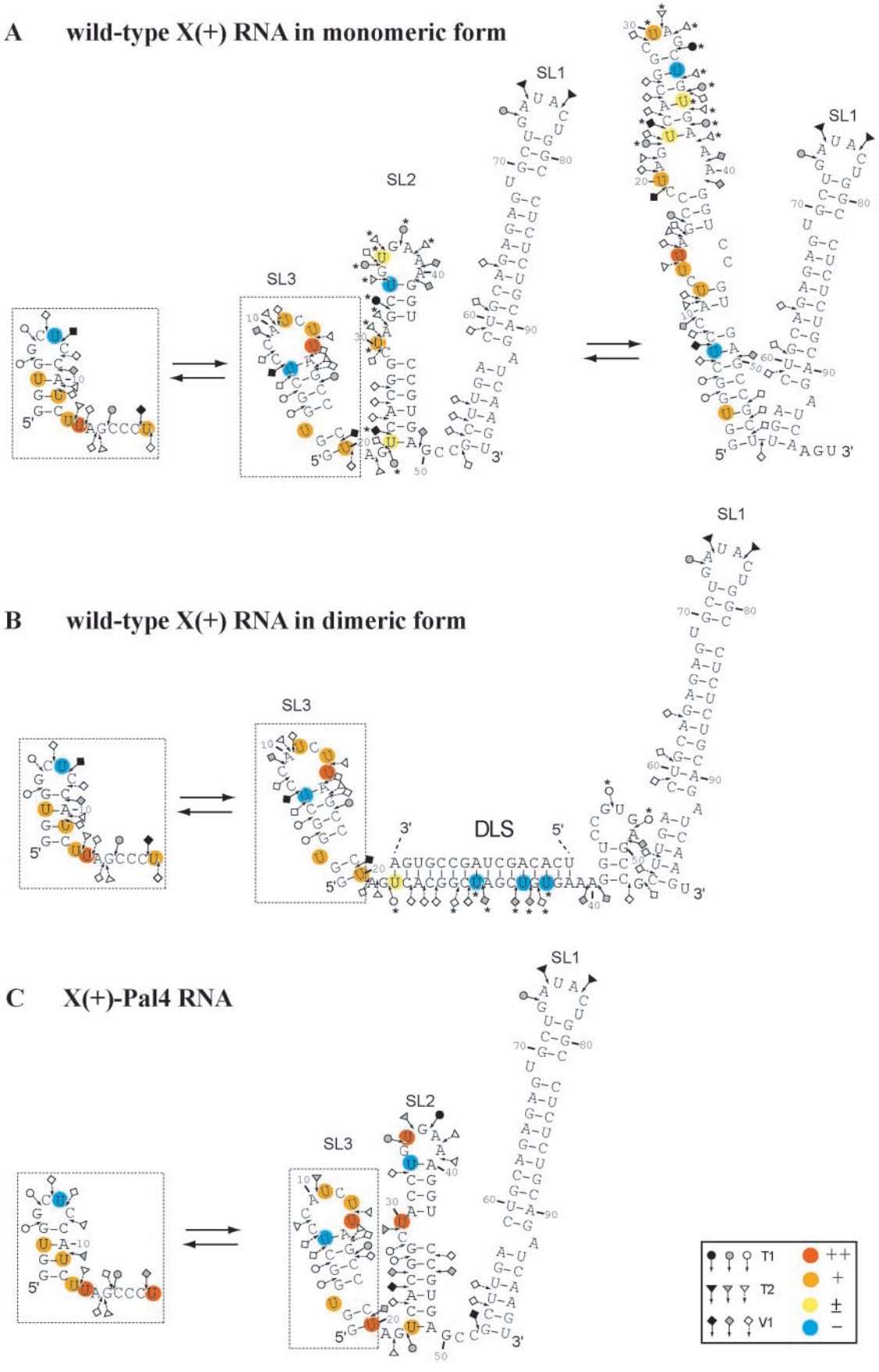


Figure 4. Enzymatic probing of wild-type X(+) RNA (**A**) and X(+)-Pal-4 RNA (**B**). 5′ End-labelled RNAs in the absence (lanes −3BD and 3BD rm) or in the presence of 3BD (lanes 3BD) were incubated with RNases T1 [0.06 (lanes C_1) and 0.12 (lanes C_2) unit], T2 [0.06 (lanes C_1) and 0.12 (lanes C_2) unit] or V1 [0.006 (lanes C_1) and 0.012 (lanes C_2) unit] as described in Materials and Methods. Lanes C_0 are controls without RNase treatment. The 3BD to nucleotide molar ratio was 1:20. For lanes 3BD rm, 3BD was removed before enzymatic probing. T and L refer to RNAse T1 sequencing and to the alkaline ladder, respectively. Arrows point out the cleavage sites. The reactivity changes induced by RNA dimerization are indicated by asterisks.


A wild-type X(+) RNA in monomeric form
SL1
SL2
SL3
5'
3'
B wild-type X(+) RNA in dimeric form
SL1
SL3
DLS
5'
3'
AGUGCCGAUCGACACU
C X(+)-Pal4 RNA
SL1
SL2
SL3
T1
T2
V1
++
+
±
–

the formation of the most stable nucleic acid structure *in vitro*, as evidenced by strand annealing (hybridization) and strand exchange assays (35). Upon binding to its partner, the genomic RNA of HCV, the core protein triggered important structural rearrangements, inducing the dimerization of the plus-strand 3′-UTR, and also that of the complementary minus-strand RNA region. Importantly, once the dimeric RNA conformation has been reached, core protein was no longer necessary to maintain the newly formed structure, confirming a bona fide chaperone function. Furthermore, the N-terminal basic amino acid clusters of core protein were found to retain nucleic acid chaperone activity and induce dimerization of the UTR RNAs (Figure 3B). Indeed, the N-terminal part of the protein was shown previously to be sufficient for RNA-binding (25) and particle formation (28,29,31), and uncharged regions connecting the basic clusters were found to be dispensable for assembly (30), indicating that RNA chaperoning and capsid assembly require the same core protein regions.

Mutational analysis and structural probing of plus- and minus-strand X RNAs (Figures 2, 4 and 6 and Supplementary Figure 4) provided strong evidence that dimerization is mediated by a 16 nt long palindromic sequence, the DLS. Mutations disrupting the palindromic nature of the DLS completely abolished UTR dimerization, while a mutant RNA preserving self-complementarity dimerized efficiently (Figure 2). Besides changing the DLS sequence, the mutations could also influence dimerization efficiency by modifying the conformation of the DLS region, or by altered core protein binding properties. Despite our attempts, we could not detect significant differences in core protein binding affinity for wild-type versus mutant UTR RNAs (data not shown). However, Mfold analyses (Supplementary Figures 3 and 4) and structural probing of a dimerization-incompetent mutant RNA (Pal-4) (Figures 4 and 6) show that the mutations may change the conformation of the DLS region or skew the equilibrium between co-existing alternative conformations. Overall, our results support a mechanism whereby both RNA sequence and structure are important determinants of core-mediated viral RNA dimerization.

Conformational changes in X RNA

The 98 nt long X RNA constituting the 3′ end of the HCV genome is extremely conserved between HCV genotypes (56) and is required for virus replication in cell culture (17–19) and in the chimpanzee (15,16). Besides replication, roles in translational regulation and RNA stability have also been proposed for the X RNA (57–59) but these issues still remain controversial (17,60).

Based on chemical and enzymatic probing of plus- and minus-strand X RNAs, we propose structural models for their monomeric and dimeric forms (Figures 5 and 7). A plausible mechanism for dimerization would be that core protein binds to (transiently) single-stranded RNA regions and prevents their immediate re-annealing. This metastable structure could then be stabilized by either regaining the monomeric conformation, or by the thermodynamically favoured dimerization. At the moment, we cannot distinguish between potential alternative (loop–loop, stem–loop or stem–stem) dimerization mechanisms.

Due to its importance in HCV replication, the conformation of X RNA has been extensively studied (54–55,61–63) and it was suggested to adopt several alternative conformations, including two different pseudoknot interactions involving the DLS region (63,64). Friebe and co-workers suggested that nucleotides [$_{32}$GCUGUGA$_{38}$] in SL2 would form a pseudoknot with a complementary sequence in a conserved stem–loop structure (5BSL3.2) in the NS5B region of the HCV genome (64). Mutations in either SL2 or 5BSL3.2 abolished replication of a subgenomic replicon, while compensatory mutations restoring the proposed base pairing rescued replication, albeit only at low level [i.e. 50-fold less than for the wild-type replicon construct, (64)]. Importantly, all structural (and subgenomic replicon) studies so far have been conducted in the absence of the core protein. Our results showing profound structural rearrangements in the X RNA region upon core protein binding argue for a functional relevance for the structures proposed in Figures 5 and 7. Indeed, RNA molecules encompassing both the NS5B and 3′-UTR regions of HCV dimerize efficiently upon core protein binding (data not shown), indicating that the dimeric form, at least *in vitro*, is favoured over the pseudoknot structures in the presence of the core protein. Thus, pseudoknot formation would probably occur only at the beginning of the replicative cycle, when core protein levels are still low. Then, concomitant with the accumulation of core, the more stable dimeric conformation would become predominant.

Although the exact role of these alternative conformations is still poorly understood, it is tempting to speculate that interconversions between the various forms—including the monomeric and dimeric RNAs, as well as the proposed pseudoknot structures—constitute an elaborate system of riboswitches that regulate major transitions between translation/replication and replication/packaging of the viral RNA (54–55,63). Since RNA chaperones, including core, have been shown to exert their effect on RNA transconformations in a highly concentration-dependent manner (35), gradual core protein accumulation following infection could be a major driving force in orchestrating the various switches. Moreover, as numerous other cellular and viral proteins have been described to bind to the X RNA region of the HCV genomic RNA, including polypyrimidine-tract binding protein, several

Figure 5. Secondary structure models for the HCV X(+) RNAs. Closed, grey and open symbols indicate strong, medium and weak cleavage sites, respectively, for the various RNases (circle for T1, triangle for T2 and diamond for V1). The colour codes used for the reactivity of uridines is indicated in the inset. The absence of colour indicates that the reactivity of the given nucleotide could not be determined because of the presence of a band in the control lane. The reactivity of nt 45–98 towards CMCT has not been determined. The reactivity changes between the monomeric and dimeric conformations are indicated by asterisks. In the monomeric and dimeric conformations of the X(+) RNA and X(+)-Pal-4 RNA, the 53–98 sequence forms stem–loop SL1, and the 1–20 sequence (dotted inset) forms stem–loop SL3 and an alternative secondary structure that are in a dynamic equilibrium. (**A**) Monomeric conformations of X(+) RNA. The reactivities of nucleotides towards CMCT and RNases were determined in the absence of 3BD. The 1–55 sequence forms stem–loops SL2 and SL3, and alternative secondary structures that are in a dynamic equilibrium. (**B**) Dimeric conformations of X(+) RNA. The reactivities of nucleotides towards CMCT and RNases were determined after removal of 3BD. The extended duplex is formed by the 23–38 sequence of two X(+) RNA molecules. Only the extended duplex and one strand of the dimer are drawn. (**C**) Monomeric conformations of X(+)-Pal-4 mutant RNA in the absence or presence of 3BD.

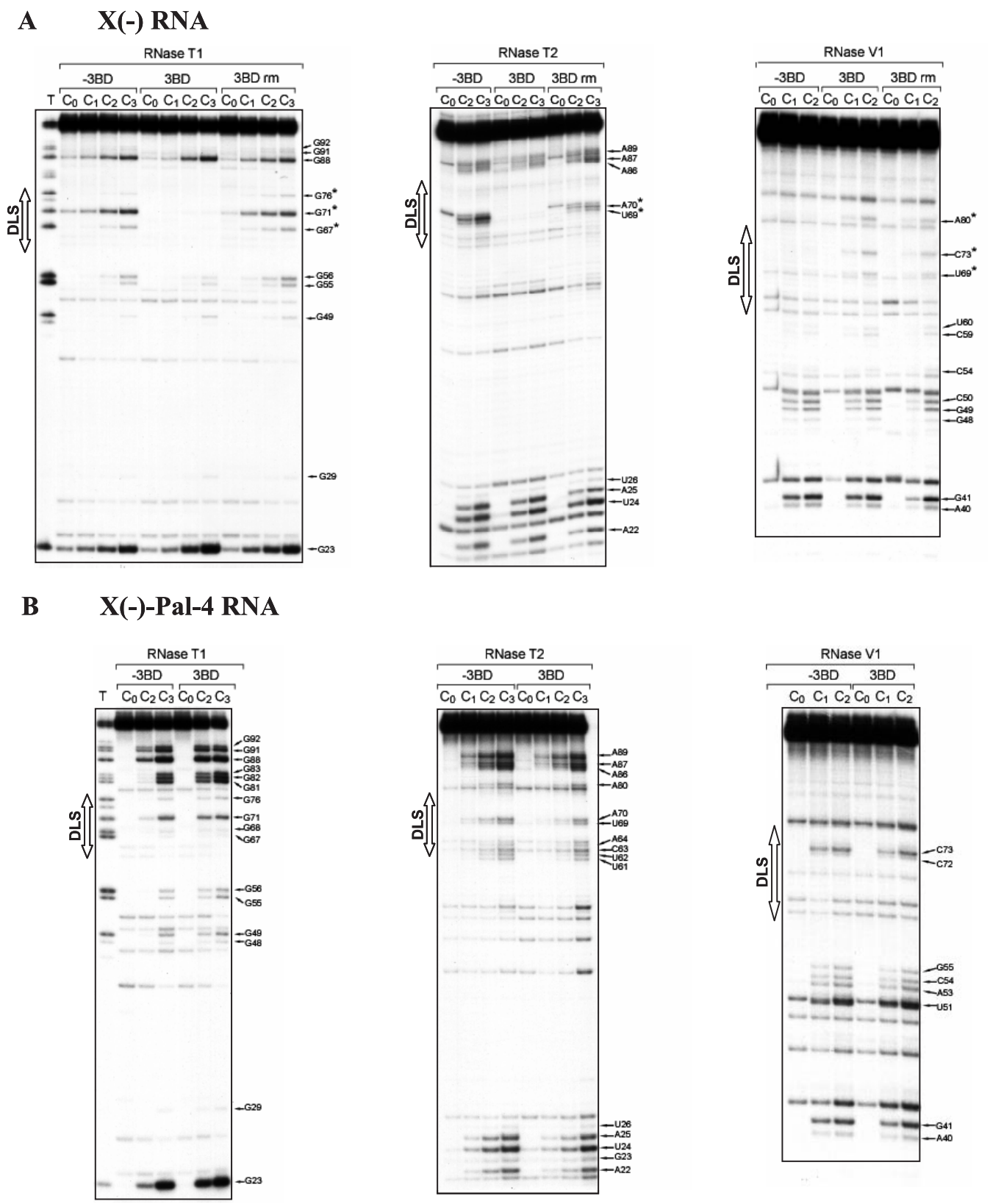


Figure 6. Enzymatic probing of wild-type X(−) RNA (**A**) and X(−)-Pal-4 RNA (**B**). 5′ End-labelled RNAs in the absence (lanes −3BD and 3BD rm) or in the presence of 3BD (lanes 3BD) were incubated with RNases T1 [0.03 (lanes C_1), 0.06 (lanes C_2) and 0.12 (lanes C_3) unit], T2 [0.03 (lanes C_1), 0.06 (lanes C_2) and 0.12 (lanes C_3) unit] or V1 [0.006 (lanes C_1) and 0.012 (lanes C_2) unit] as described in Materials and Methods. Lanes C_0 are controls without RNase treatment. The 3BD to nucleotide molar ratio was 1:20. For lanes 3BD rm, 3BD was removed before enzymatic probing. T refers to RNAse T1 sequencing. Arrows point out the cleavage sites. The reactivity changes induced by 3BD are indicated by asterisks.

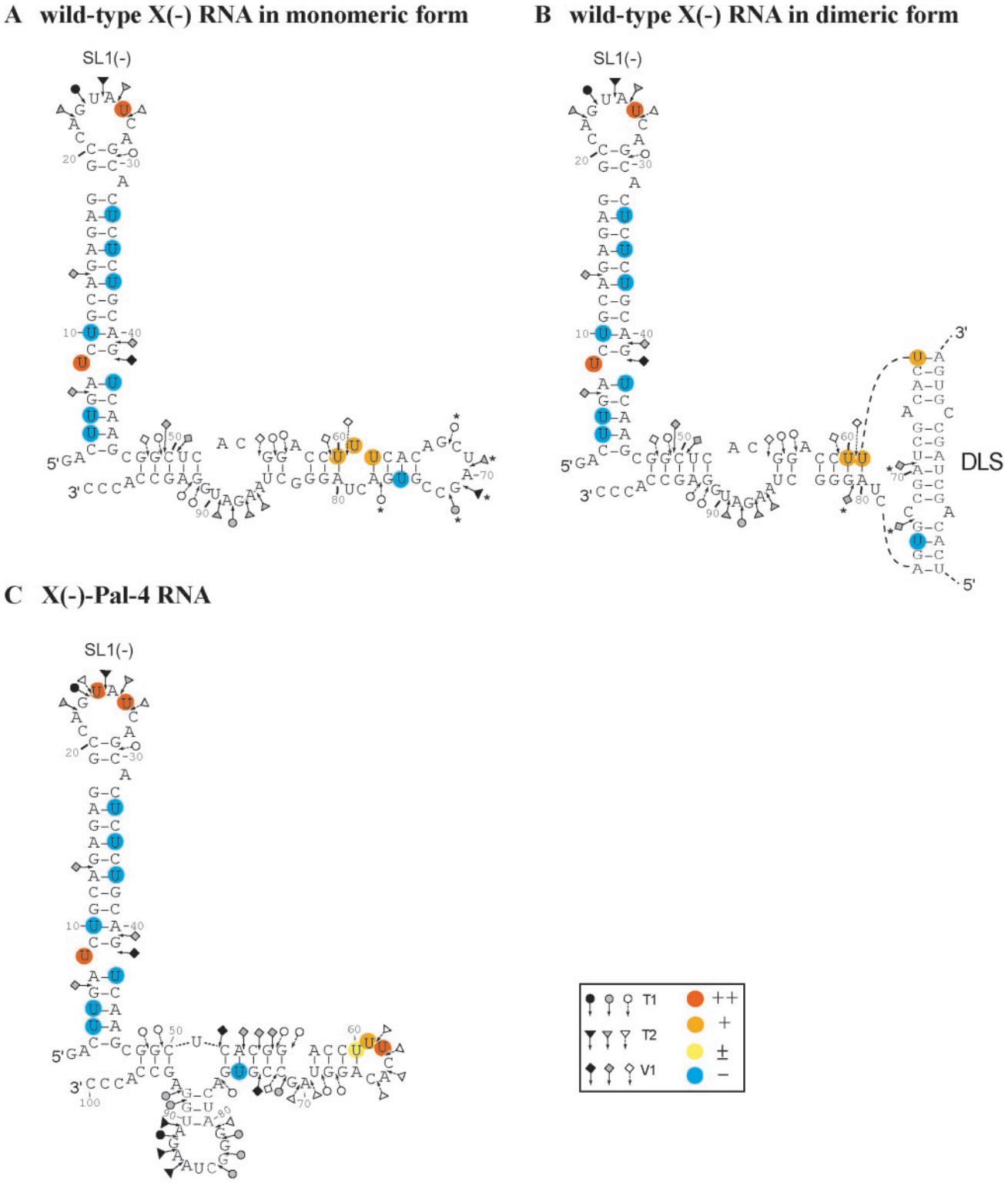


Figure 7. Secondary structure model for the X(−) RNAs. Closed, grey and open symbols indicate strong, medium and weak cleavage sites, respectively, for the various RNases (circle for T1, triangle for T2 and diamond for V1). The colour codes used for the reactivity of uridines are indicated in the inset. The absence of colour indicates that the reactivity of the given nucleotide could not be determined because of the presence of a band in the control lane. The reactivity of nt 83–100 towards CMCT has not been determined. In the monomeric and dimeric conformations of the X(−) and X(−)-Pal-4 mutant RNAs, the 3–46 sequence forms the SL1(−) stem–loop. (**A**) Monomeric conformation of X(−) RNA. The reactivities of nucleotides towards CMCT and RNases were determined in the absence of 3BD. The reactivity decreases induced by 3BD are indicated by asterisks. (**B**) Dimeric conformation of X(−) RNA. The reactivities of nucleotides towards CMCT and RNases were determined in the presence of 3BD. The reactivity increases induced by 3BD are indicated by asterisks. The extended duplex is formed by the 62–77 sequence of two X(−) RNA molecules. Only the extended duplex and one strand of the dimer are drawn. (**C**) Monomeric conformation of X(−)-Pal-4 mutant RNA in the absence or presence of the 3BD peptide.

ribosomal proteins, and the viral non-structural NS3 and NS5B proteins (58,65–68), the actual conformational state of X RNA—and the resulting function—would probably reflect the complex interplay of these proteins.

Interestingly, RNA chaperone-mediated remodelling of regulatory regions in genomic RNA seems to be a widespread strategy employed by distant virus families to regulate diverse steps of the viral life cycle. The best known example is the nucleocapsid (NC) protein of retroviruses and retroelements, which is instrumental in virus structure, genomic RNA replication, dimerization/packaging, and recombination (42). Moreover, the nucleocapsid (N) protein of hantaviruses (segmented negative-sense RNA viruses of the Bunyaviridae family) has also been suggested to act as an RNA chaperone, binding with high affinity the terminal panhandle structures (69).

Possible consequences of HCV RNA dimerization on viral variability

The status of HCV genomic RNA, i.e. if it is monomeric or dimeric, either in infected cells or in virions, is currently unknown. Lack of efficient cell culture systems producing infectious viruses had for long time precluded studies aimed at understanding viral assembly and structure. With the recent development of cell culture models supporting complete replication of HCV (70–72), these studies became possible, and we are currently investigating the role of RNA dimerization in HCV replication and packaging. Nevertheless, several independent pieces of evidence support an important role for core protein chaperoning and RNA dimerization in the viral life cycle: (i) the DLS is completely conserved in all reported HCV sequences, and mutations in it abolish replication (18,64); (ii) the core proteins of related viruses (such as the flavivirus WNV, the pestivirus BVDV, and GBV-B, an unclassified virus most closely related to HCV) all show potent RNA chaperone activities *in vitro* (data to be published elsewhere); and (iii) the GBV-B 3′-UTR adopts a dimeric conformation in the presence of GBV-B core, similarly to HCV (data to be published elsewhere).

A possible consequence of dimerization may be the facilitation of recombination between various HCV genotypes and subtypes, leading to increased viral variability, as it is well documented in retroviruses (73,74). To our knowledge, this is the first description of an RNA virus whose genome may dimerize—and thus might undergo recombination by a copy-choice mechanism—both in the positive-strand and the replicative negative-strand forms. Reports of circulating recombinant HCV strains (75–77) demonstrate that superinfection and recombination do occur in natural populations, although their frequency and their determinants are still poorly understood.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Gaël Cristofari for critical reading of the manuscript and Françoise Chaminade for skilful technical assistance. This work was supported by grants from ANRS (to J.-L.D., J.-P. L. and P.F.). R. I-N. is the recipient of an ANRS PhD fellowship. Funding to pay the Open Access publication charges for this article was provided by ANRS.

Conflict of interest statement. None declared.

REFERENCES

1. World Health Organization (1999) Global surveillance and control of hepatitis C. Report of a WHO Consultation organized in collaboration with the Viral Hepatitis Prevention Board, Antwerp, Belgium. *J. Viral Hepat.*, **6**, 35–47.
2. Poynard,T., Ratziu,V., Benhamou,Y., Opolon,P., Cacoub,P. and Bedossa,P. (2000) Natural history of HCV infection. *Baillieres Best Pract. Res. Clin. Gastroenterol.*, **14**, 211–228.
3. Lindenbach,B.D. and Rice,C.M. (2001) *Flaviviridae*: the viruses and their replication. In Knipe,D.M. and Howley,P.M. (eds), *Fields Virology*.. Lippincott Williams & Wilkins, Philadelphia, PAVol. I, pp. 991–1041.
4. Bukh,J., Apgar,C.L., Govindarajan,S. and Purcell,R.H. (2001) Host range studies of GB virus-B hepatitis agent, the closest relative of hepatitis C virus, in New World monkeys and chimpanzees. *J. Med. Virol.*, **65**, 694–697.
5. Martin,A., Bodola,F., Sangar,D.V., Goettge,K., Popov,V., Rijnbrand,R., Lanford,R.E. and Lemon,S.M. (2003) Chronic hepatitis associated with GB virus B persistence in a tamarin after intrahepatic inoculation of synthetic viral RNA. *Proc. Natl Acad. Sci. USA*, **100**, 9962–9967.
6. Bartenschlager,R., Frese,M. and Pietschmann,T. (2004) Novel insights into hepatitis C virus replication and persistence. *Adv. Virus Res.*, **63**, 71–180.
7. Tsukiyama-Kohara,K., Iizuka,N., Kohara,M. and Nomoto,A. (1992) Internal ribosome entry site within hepatitis C virus RNA. *J. Virol.*, **66**, 1476–1483.
8. Wang,C., Sarnow,P. and Siddiqui,A. (1993) Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *J. Virol.*, **67**, 3338–3344.
9. Friebe,P., Lohmann,V., Krieger,N. and Bartenschlager,R. (2001) Sequences in the 5′ nontranslated region of hepatitis C virus required for RNA replication. *J. Virol.*, **75**, 12047–12057.
10. Kim,Y.K., Kim,C.S., Lee,S.H. and Jang,S.K. (2002) Domains I and II in the 5′ nontranslated region of the HCV genome are required for RNA replication. *Biochem. Biophys. Res. Commun.*, **290**, 105–112.
11. Kolykhalov,A.A., Feinstone,S.M. and Rice,C.M. (1996) Identification of a highly conserved sequence element at the 3′ terminus of hepatitis C virus genome RNA. *J. Virol.*, **70**, 3363–3371.
12. Tanaka,T., Kato,N., Cho,M.J. and Shimotohno,K. (1995) A novel sequence found at the 3′ terminus of hepatitis C virus genome. *Biochem. Biophys. Res. Commun.*, **215**, 744–749.
13. Tanaka,T., Kato,N., Cho,M.J., Sugiyama,K. and Shimotohno,K. (1996) Structure of the 3′ terminus of the hepatitis C virus genome. *J. Virol.*, **70**, 3307–3312.
14. Yamada,N., Tanihara,K., Takada,A., Yorihuzi,T., Tsutsumi,M., Shimomura,H., Tsuji,T. and Date,T. (1996) Genetic organization and diversity of the 3′ noncoding region of the hepatitis C virus genome. *Virology*, **223**, 255–261.
15. Yanagi,M., St. Claire,M., Emerson,S.U., Purcell,R.H. and Bukh,J. (1999) *In vivo* analysis of the 3′ untranslated region of the hepatitis C virus after *in vitro* mutagenesis of an infectious cDNA clone. *Proc. Natl Acad. Sci. USA*, **96**, 2291–2295.
16. Kolykhalov,A.A., Mihalik,K., Feinstone,S.M. and Rice,C.M. (2000) Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3′ nontranslated region are essential for virus replication *in vivo*. *J. Virol.*, **74**, 2046–2051.
17. Friebe,P. and Bartenschlager,R. (2002) Genetic analysis of sequences in the 3′ nontranslated region of hepatitis C virus that are important for RNA replication. *J. Virol.*, **76**, 5326–5338.
18. Yi,M. and Lemon,S.M. (2003) 3′ Nontranslated RNA signals required for replication of hepatitis C virus RNA. *J. Virol.*, **77**, 3557–3568.
19. Yi,M. and Lemon,S.M. (2003) Structure-function analysis of the 3′ stem–loop of hepatitis C virus genomic RNA and its role in viral RNA replication. *RNA*, **9**, 331–345.

20. Penin,F., Dubuisson,J., Rey,F.A., Moradpour,D. and Pawlotsky,J.M. (2004) Structural biology of hepatitis C virus. *Hepatology*, **39**, 5–19.
21. McLauchlan,J., Lemberg,M.K., Hope,G. and Martoglio,B. (2002) Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. *EMBO J.*, **21**, 3980–3988.
22. Kato,T., Miyamoto,M., Furusaka,A., Date,T., Yasui,K., Kato,J., Matsushima,S., Komatsu,T. and Wakita,T. (2003) Processing of hepatitis C virus core protein is regulated by its C-terminal sequence. *J. Med. Virol.*, **69**, 357–366.
23. Okamoto,K., Moriishi,K., Miyamura,T. and Matsuura,Y. (2004) Intramembrane proteolysis and endoplasmic reticulum retention of hepatitis C virus core protein. *J. Virol.*, **78**, 6370–6380.
24. Boulant,S., Vanbelle,C., Ebel,C., Penin,F. and Lavergne,J.P. (2005) Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. *J. Virol.*, **79**, 11353–11365.
25. Santolini,E., Migliaccio,G. and La Monica,N. (1994) Biosynthesis and biochemical properties of the hepatitis C virus core protein. *J. Virol.*, **68**, 3631–3641.
26. Fan,Z., Yang,Q.R., Twu,J.S. and Sherker,A.H. (1999) Specific *in vitro* association between the hepatitis C viral genome and core protein. *J. Med. Virol.*, **59**, 131–134.
27. Shimoike,T., Mimori,S., Tani,H., Matsuura,Y. and Miyamura,T. (1999) Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. *J. Virol.*, **73**, 9718–9725.
28. Kunkel,M., Lorinczi,M., Rijnbrand,R., Lemon,S.M. and Watowich,S.J. (2001) Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein. *J. Virol.*, **75**, 2119–2129.
29. Klein,K.C., Polyak,S.J. and Lingappa,J.R. (2004) Unique features of hepatitis C virus capsid formation revealed by *de novo* cell-free assembly. *J. Virol.*, **78**, 9257–9269.
30. Klein,K.C., Dellos,S.R. and Lingappa,J.R. (2005) Identification of residues in the hepatitis C virus core protein that are critical for capsid assembly in a cell-free system. *J. Virol.*, **79**, 6814–6826.
31. Majeau,N., Gagne,V., Boivin,A., Bolduc,M., Majeau,J.A., Ouellet,D. and Leclerc,D. (2004) The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation. *J. Gen. Virol.*, **85**, 971–981.
32. McLauchlan,J. (2000) Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes. *J. Viral Hepat.*, **7**, 2–14.
33. Ray,R.B. and Ray,R. (2001) Hepatitis C virus core protein: Intriguing properties and functional relevance. *FEMS Microbiol. Lett.*, **202**, 149–156.
34. Giannini,C. and Brechot,C. (2003) Hepatitis C virus biology. *Cell Death Differ.*, **10**, S27–S38.
35. Cristofari,G., Ivanyi-Nagy,R., Gabus,C., Boulant,S., Lavergne,J.P., Penin,F. and Darlix,J.L. (2004) The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA *in vitro*. *Nucleic Acids Res.*, **32**, 2623–2631.
36. Herschlag,D. (1995) RNA chaperones and the RNA folding problem. *J. Biol. Chem.*, **270**, 20871–20874.
37. Cristofari,G. and Darlix,J.L. (2002) The ubiquitous nature of RNA chaperone proteins. *Prog. Nucleic Acid Res. Mol. Biol.*, **72**, 223–268.
38. Schroeder,R., Barta,A. and Semrad,K. (2004) Strategies for RNA folding and assembly. *Nature Rev. Mol. Cell Biol.*, **5**, 908–919.
39. Ali,N. and Siddiqui,A. (1997) The La antigen binds 5′ noncoding region of the hepatitis C virus RNA in the context of the initiator AUG codon and stimulates internal ribosome entry site-mediated translation. *Proc. Natl Acad. Sci. USA*, **94**, 2249–2254.
40. Wolin,S.L. and Cedervall,T. (2002) The La protein. *Annu. Rev. Biochem.*, **71**, 375–403.
41. Darlix,J.L., Lapadat-Tapolsky,M., de Rocquigny,H. and Roques,B.P. (1995) First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.*, **254**, 523–537.
42. Darlix,J.L., Cristofari,G., Rau,M., Pechoux,C., Berthoux,L. and Roques,B. (2000) Nucleocapsid protein of human immunodeficiency virus as a model protein with chaperoning functions and as a target for antiviral drugs. *Adv. Pharmacol.*, **48**, 345–372.
43. Mikaelian,I. and Sergeant,A. (1992) A general and fast method to generate multiple site directed mutations. *Nucleic Acids Res.*, **20**, 376.
44. Polge,E., Darlix,J.L., Paoletti,J. and Fosse,P. (2000) Characterization of loose and tight dimer forms of avian leukosis virus RNA. *J. Mol. Biol.*, **300**, 41–56.
45. Fosse,P., Motte,N., Roumier,A., Gabus,C., Muriaux,D., Darlix,J.L. and Paoletti,J. (1996) A short autocomplementary sequence plays an essential role in avian sarcoma-leukosis virus RNA dimerization. *Biochemistry*, **35**, 16601–16609.
46. Kanevsky,I., Chaminade,F., Ficheux,D., Moumen,A., Gorelick,R., Negroni,M., Darlix,J.L. and Fosse,P. (2005) Specific interactions between HIV-1 nucleocapsid protein and the TAR element. *J. Mol. Biol.*, **348**, 1059–1077.
47. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
48. Schroeder,S.J., Burkard,M.E. and Turner,D.H. (1999) The energetics of small internal loops in RNA. *Biopolymers*, **52**, 157–167.
49. Marquet,R., Paillart,J.C., Skripkin,E., Ehresmann,C. and Ehresmann,B. (1994) Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucleic Acids Res.*, **22**, 145–151.
50. Moradpour,D., Englert,C., Wakita,T. and Wands,J.R. (1996) Characterization of cell lines allowing tightly regulated expression of hepatitis C virus core protein. *Virology*, **222**, 51–63.
51. Barba,G., Harper,F., Harada,T., Kohara,M., Goulinet,S., Matsuura,Y., Eder,G., Schaff,Z., Chapman,M.J., Miyamura,T. and Brechot,C. (1997) Hepatitis C virus core protein shows a cytoplasmic localization and associates to cellular lipid storage droplets. *Proc. Natl Acad. Sci. USA*, **94**, 1200–1205.
52. Hope,R.G. and McLauchlan,J. (2000) Sequence motifs required for lipid droplet association and protein stability are unique to the hepatitis C virus core protein. *J. Gen. Virol.*, **81**, 1913–1925.
53. Kunkel,M. and Watowich,S.J. (2004) Biophysical characterization of hepatitis C virus core protein: implications for interactions within the virus and host. *FEBS Lett.*, **557**, 174–180.
54. Blight,K.J. and Rice,C.M. (1997) Secondary structure determination of the conserved 98-base sequence at the 3′ terminus of hepatitis C virus genome RNA. *J. Virol.*, **71**, 7345–7352.
55. Ito,T. and Lai,M.M. (1997) Determination of the secondary structure of and cellular protein binding to the 3′-untranslated region of the hepatitis C virus RNA genome. *J. Virol.*, **71**, 8698–8706.
56. Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F., Miyakawa,Y. and Mayumi,M. (1998) The entire nucleotide sequences of three hepatitis C virus isolates in genetic groups 7–9 and comparison with those in the other eight genetic groups. *J. Gen. Virol.*, **79**, 1847–1857.
57. Ito,T., Tahara,S.M. and Lai,M.M.C. (1998) The 3′-untranslated region of hepatitis C virus RNA enhances translation from an internal ribosomal entry site. *J. Virol.*, **72**, 8789–8796.
58. Ito,T. and Lai,M.M. (1999) An internal polypyrimidine-tract-binding protein-binding site in the hepatitis C virus RNA attenuates translation, which is relieved by the 3′-untranslated sequence. *Virology*, **254**, 288–296.
59. Spangberg,K., Wiklund,L. and Schwartz,S. (2001) Binding of the La autoantigen to the hepatitis C virus 3′ untranslated region protects the RNA from rapid degradation *in vitro*. *J. Gen. Virol.*, **82**, 113–120.
60. Kong,L.K. and Sarnow,P. (2002) Cytoplasmic expression of mRNAs containing the internal ribosome entry site and 3′ noncoding region of hepatitis C virus: effects of the 3′ leader on mRNA translation and mRNA stability. *J. Virol.*, **76**, 12457–12462.
61. Smith,R.M., Walton,C.M., Wu,C.H. and Wu,G.Y. (2002) Secondary structure and hybridization accessibility of hepatitis C virus 3′-terminal sequences. *J. Virol.*, **76**, 9563–9574.
62. Smith,R.M. and Wu,G.Y. (2004) Secondary structure and hybridization accessibility of the hepatitis C virus negative strand RNA 5′-terminus. *J. Viral Hepat.*, **11**, 115–123.
63. Dutkiewicz,M. and Ciesiolka,J. (2005) Structural characterization of the highly conserved 98-base sequence at the 3′ end of HCV RNA genome and the complementary sequence located at the 5′ end of the replicative viral strand. *Nucleic Acids Res.*, **33**, 693–703.
64. Friebe,P., Boudet,J., Simorre,J.P. and Bartenschlager,R. (2005) Kissing-loop interaction in the 3′ end of the hepatitis C virus genome essential for RNA replication. *J. Virol.*, **79**, 380–392.
65. Tsuchihara,K., Tanaka,T., Hijikata,M., Kuge,S., Toyoda,H., Nomoto,A., Yamamoto,N. and Shimotohno,K. (1997) Specific interaction of polypyrimidine tract-binding protein with the extreme 3′-terminal structure of the hepatitis C virus genome, the 3′X. *J. Virol.*, **71**, 6720–6726.

66. Wood,J., Frederickson,R.M., Fields,S. and Patel,A.H. (2001) Hepatitis C virus 3′X region interacts with human ribosomal proteins. *J. Virol.*, **75**, 1348–1358.
67. Banerjee,R. and Dasgupta,A. (2001) Specific interaction of hepatitis C virus protease/helicase NS3 with the 3′-terminal sequences of viral positive- and negative-strand RNA. *J. Virol.*, **75**, 1708–1721.
68. Oh,J.W., Sheu,G.T. and Lai,M.M.C. (2000) Template requirement and initiation site selection by hepatitis C virus polymerase on a minimal viral RNA template. *J. Biol. Chem.*, **275**, 17710–17717.
69. Mir,M.A. and Panganiban,A.T. (2005) The hantavirus nucleocapsid protein recognizes specific features of the viral RNA panhandle and is altered in conformation upon RNA binding. *J. Virol.*, **79**, 1824–1835.
70. Wakita,T., Pietschmann,T., Kato,T., Date,T., Miyamoto,M., Zhao,Z., Murthy,K., Habermann,A., Krausslich,H.G., Mizokami,M. *et al.* (2005) Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Med.*, **11**, 791–796.
71. Lindenbach,B.D., Evans,M.J., Syder,A.J., Wolk,B., Tellinghuisen,T.L., Liu,C.C., Maruyama,T., Hynes,R.O., Burton,D.R., McKeating,J.A. *et al.* (2005) Complete replication of hepatitis C virus in cell culture. *Science*, **309**, 623–626.
72. Zhong,J., Gastaminza,P., Cheng,G., Kapadia,S., Kato,T., Burton,D.R., Wieland,S.F., Uprichard,S.L., Wakita,T. and Chisari,F.V. (2005) Robust hepatitis C virus infection *in vitro*. *Proc. Natl Acad. Sci. USA*, **102**, 9294–9299.
73. Mikkelsen,J.G., Rasmussen,S.V. and Pedersen,F.S. (2004) Complementarity-directed RNA dimer-linkage promotes retroviral recombination *in vivo*. *Nucleic Acids Res.*, **32**, 102–114.
74. Chin,M.P.S., Rhodes,T.D., Chen,J., Fu,W. and Hu,W.S. (2005) Identification of a major restriction in HIV-1 intersubtype recombination. *Proc. Natl Acad. Sci. USA*, **102**, 9002–9007.
75. Kalinina,O., Norder,H., Mukomolov,S. and Magnius,L.O. (2002) A natural intergenotypic recombinant of hepatitis C virus identified in St Petersburg. *J. Virol.*, **76**, 4034–4043.
76. Kalinina,O., Norder,H. and Magnius,L.O. (2004) Full-length open reading frame of a recombinant hepatitis C virus strain from St Petersburg: proposed mechanism for its formation. *J. Gen. Virol.*, **85**, 1853–1857.
77. Colina,R., Casane,D., Vasquez,S., Garcia-Aguirre,L., Chunga,A., Romero,H., Khan,B. and Cristina,J. (2004) Evidence of intratypic recombination in natural populations of hepatitis C virus. *J. Gen. Virol.*, **85**, 31–37.