

# The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA *in vitro*

Gaël Cristofari, Roland Ivanyi-Nagy, Caroline Gabus, Steeve Boulant<sup>1</sup>, Jean-Pierre Lavergne<sup>1</sup>, François Penin<sup>1</sup> and Jean-Luc Darlix\*

LaboRetro, INSERM #412, ENS, 46, allée d'Italie, 69364 Lyon Cedex 07, France and <sup>1</sup>IBCP, CNRS, 7 passage du Vercors, 69367 Lyon Cedex 07, France

Received March 2, 2004; Revised April 1, 2004; Accepted April 12, 2004

## ABSTRACT

The hepatitis C virus (HCV) is an important human pathogen causing chronic hepatitis, liver cirrhosis and hepatocellular carcinoma. HCV is an enveloped virus with a positive-sense, single-stranded RNA genome encoding a single polyprotein that is processed to generate viral proteins. Several hundred molecules of the structural Core protein are thought to coat the genome in the viral particle, as do nucleocapsid (NC) protein molecules in Retroviruses, another class of enveloped viruses containing a positive-sense RNA genome. Retroviral NC proteins also possess nucleic acid chaperone properties that play critical roles in the structural remodelling of the genome during retrovirus replication. This analogy between HCV Core and retroviral NC proteins prompted us to investigate the putative nucleic acid chaperoning properties of the HCV Core protein. Here we report that Core protein chaperones the annealing of complementary DNA and RNA sequences and the formation of the most stable duplex by strand exchange. These results show that the HCV Core is a nucleic acid chaperone similar to retroviral NC proteins. We also find that the Core protein directs dimerization of HCV (+) RNA 3' untranslated region which is promoted by a conserved palindromic sequence possibly involved at several stages of virus replication.

## INTRODUCTION

The hepatitis C virus (HCV) is an important human pathogen associated with non-A, non-B hepatitis, and thus is considered to be the major cause of chronic hepatitis and liver cirrhosis, leading to hepatocellular carcinoma (1). HCV, an enveloped virus member of the *Flaviviridae* family, has a positive-sense, single stranded RNA genome of ~9600 nt. The genome contains a single open-reading frame (ORF) encoding a

polyprotein of 3010 amino acids that is flanked by 5' and 3' untranslated regions (UTR) of ~340 and ~230 nt, respectively. The 5' and 3' UTRs bear highly conserved RNA structures that are essential for protein synthesis and viral RNA replication. The viral polyprotein precursor is cleaved by cellular and viral proteases to generate at least 10 viral proteins identified as the structural proteins (Core, E1, E2 and p7) and the non-structural proteins and enzymes (NS2, NS3, NS4a, NS4b, NS5a and NS5b) (2). The most N-terminal viral protein is the basic Core protein, several hundreds of molecules of which are thought to coat the genome in the viral particle, and which is a major HCV antigen (3,4). Core binds RNA and DNA, forms virus-like particles in the presence of related or unrelated RNA (5) and is probably involved in genomic RNA packaging (6,7).

Retroviruses such as human immunodeficiency virus (HIV) and murine leukaemia virus (MuLV) are also enveloped viruses containing a positive-sense RNA genome that is extensively coated by several hundred molecules of the basic nucleocapsid (NC) protein in the viral particle. In addition, retroviral NC proteins were found to possess nucleic acid chaperone properties *in vitro* that play critical roles in the structural remodelling of the genome during retrovirus replication, notably chaperoning genomic RNA packaging and compaction within the NC structure (8). Additionally, NC protein has been implicated in promoting the conversion of the plus-strand genomic RNA into a double-stranded DNA by reverse transcriptase during the early phases of retrovirus replication (9). Such a parallel between these two families of enveloped viruses containing a positive-sense genomic RNA and the analogy between the HCV Core protein and retroviral NC proteins prompted us to investigate the nucleic acid chaperoning properties of the HCV Core protein *in vitro*.

Nucleic acid chaperone proteins, like retroviral NC proteins, can bind nucleic acids with broad sequence specificities and in a cooperative manner. They facilitate the formation of the most stable nucleic acid structure and circumvent folding traps. Importantly, they do not require ATP to function and once the most stable nucleic acid structure has been reached, their binding is no longer required to maintain the new conformation (10,11). Apart from their basic nature, no motif or signature specific for nucleic acid

\*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

Present address:

Gaël Cristofari, ISREC, Chemin des Boveresses 155, 1066 Epalinges, Switzerland

**Table 1.** DNA oligonucleotides used in the present study

Name	Sense	Sequence (from 5' to 3')	Description	Reference
Tar(+)	(+)	GGTCTCTCTTGTAGACCAGG-TCGAGCCCGGAGCTCTCTGGCTAGCAAGGAACCC	Nucleotides 1–56 of HIV-1 MAL strain genomic RNA	Lapadat-Tapolsky <i>et al.</i> (21)
Tar(-)	(-)	GGGTTCCCTTGCTAGCCAGAGA-GCTCCCGGGCTCGACCTGGTCTAACAAAGAGAGACC	Nucleotides 1–56 of HIV-1 MAL strain genomic RNA	Lapadat-Tapolsky <i>et al.</i> (21)
R(-) <sup>wt</sup>	(-)	GGGAGGCACTCAAGGCAAGC-TTTATTGAGGCTTAAGCAGTG-GGTTCCCTTGCTAGCCAGAGAG-CTCCCGGGCTCGACCTGGTCT-AACAAGAGAGACC	Nucleotides 1–96 of HIV-1 MAL strain genomic RNA	Lapadat-Tapolsky <i>et al.</i> (21)
R(-) <sup>mut</sup>	(-)	GGGAGGCACTCAAGGCAAGC-TTTATTGAGGCTTAAGCAGTG-GGTTCCCTTGCTAGCCAGAGAG-CTCCCGGGCTCGACCTGGTCT-AACATCAGTCTCTA	Complementary to R(+), but with 3' terminal mismatches	Lapadat-Tapolsky <i>et al.</i> (21)
R(+)	(+)	GGTCTCTCTTGTAGACCAGG-TCGAGCCCGGAGCTCTCTGGCTAGCAAGGAACCCACTGCTT-AAGCCTCAATAAAGCTTGCCT-TGAGTGCCTCCC	Nucleotides 1–96 of HIV-1 MAL strain genomic RNA	Lapadat-Tapolsky <i>et al.</i> (21)
5'Core	(+)	TATCATATGAGCACAAATCCT-AAACCTCAAAGA	NdeI site, 5' end HCV 1b Core protein (ORF)	This study
3'Core117	(-)	TATCTGCAGACGCGACTACG-CCGGGGGTCTGT	PstI site, 3' end HCV 1b Core protein (ORF)	This study
3'Core169	(-)	TATCTGCAGCAGATTCCTGTT-GCATAGTTCAC	PstI site, 3' end HCV 1b Core protein (ORF)	This study
GAEL58	(+)	GGAATTCTAATACGACTCACTA-TAGGGTGGCTCCATCTAGCCC-TAGTCACGG	EcoRI site, T7 promoter and nucleotides 9508–9534 of HCV 1b RNA	This study
GAEL59	(-)	ACTTGATCTGCAGAGAGGCCAG-TATC	Nucleotides 9605–9580 HCV 1b RNA	This study
GAEL65	(+)	GGAATTCACGGGGAGCTAAACA-CTCCAGGCC	EcoRI site and nucleotides 9375–9398 of HCV 1b RNA	This study
GAEL66	(+)	GGGCTCGAGTAATACGACTCAC-TATAGGGACGGGGAGCTAAACA-CTCCAGGCC	XhoI site, T7 promoter and nucleotides 9375–9398 of HCV 1b RNA	This study
GAEL67	(-)	GGGGAAGCTTACTTGATCTGCA-GAGAGGCCAGTATC	HindIII site, nucleotides 9605–9580 HCV 1b RNA	This study
as-SL3	(-)	CTAGGGCTAAGATGGAGCCAC		This study
as-SL2/1	(-)	TTTCACAGCTAGCCGTGACTA		This study
as-SL2/2	(-)	TCAAGCGGCTCACGGACCTT		This study

chaperones has been identified. Cellular nucleic acid chaperones take part in a wide variety of processes, and notably play fundamental roles in RNA metabolism including mRNA processing, transport, maintenance and translation (10,11). Viral nucleic acid chaperones, of which retroviral NC proteins are prototypes, are a key component of the viral particle and are involved in stages of virus replication as diverse as genome replication, translation-to-replication switch or viral particle formation (9). Here we show that the HCV Core protein has potent nucleic acid chaperone properties analogous to retroviral NC proteins. Strikingly, this activity directs an important structural rearrangement of HCV 3'UTR, resulting in RNA dimerization *in vitro*. By providing a biochemical activity to a major HCV structural protein, these findings shed new light on the dynamics of HCV components.

## MATERIALS AND METHODS

### Plasmid DNA construction

Plasmid DNAs were constructed by cloning a high-fidelity PCR generated fragment containing two restriction sites at its

ends (Roche) into pSP64 (Promega) for *in vitro* transcription or pT7-7(6xHis) for recombinant protein expression in *Escherichia coli*. Plasmid DNAs were purified using Qiagen Maxiprep kit. All DNA constructs have been sequenced. Oligodeoxynucleotides (ODNs) used as PCR primers are reported in Table 1. Plasmid containing the HCV 3' RNA region was generated by PCR using the Hepatitis C virus type 1b clone DNA as template (isolate Con1, a kind gift from R. Bartenschlager, GenBank accession number AJ238799) (12), and the ODN primers described in Table 1. The amplified DNA fragment was cloned into pSP64 (Promega) to generate pVan37. Similarly, a 507 bp fragment corresponding to amino acids 1–169 of the HCV Core protein (WT, see Fig. 1A) was amplified by PCR from sequence D89872 [a kind gift from K. Shimotohno (13)] and with two specific primers containing additional NdeI site in 5' and PstI site in 3'. The NdeI–PstI fragment was cloned into the bacterial expression vector pT7-7(6xHis) (14). Plasmid DNAs pHIV-CG4 containing part of HIV-1 cDNA (15), pTL9 containing the sequence of tRNALys,3 (16) and pTy1-CG73, containing part of Ty1 cDNA (17) were previously described.

## RNAs

Recombinant RNAs were prepared by *in vitro* transcription with T7 RNA polymerase according to the manufacturer's instructions (RiboMax kit; Promega). RNAs were further purified by polyacrylamide gel electrophoresis (PAGE) containing 7 M urea and 50 mM Tris-borate pH 8.3, 1 mM EDTA (0.5× TBE), ethanol precipitation and spin-column exclusion chromatography (S-300 HR; Amersham Biosciences). Where indicated, RNA was labelled by incorporation of [<sup>32</sup>P]UMP during transcription. To generate RNAs comprising part of or the full-length HCV 3'UTR, templates were obtained by PCR amplification of plasmid pVan37 using primers GAEL66 and GAEL59 (HCV 3' FL RNA) or GAEL58 and GAEL59 (HCV 3' X RNA). For 3' truncations of HCV RNA, plasmid pVan37 was digested with HindIII, Fnu4HI or NheI. An HIV-derived RNA including nucleotides 1–415 (HIV 5' RNA) was synthesized using plasmid pHIV-CG4 digested with HaeIII as template. Template for tRNA<sup>Lys,3</sup> transcription was obtained by digesting plasmid pTL9 with BanI.

## Proteins

The Core protein C[2–169] (denoted WT or Core protein) and its truncated version C[2–117] (NT) were expressed in *E. coli* with a C-terminal (His)<sub>6</sub>-tag and purified by Ni-chelate chromatography and RP-HPLC to homogeneity as previously reported (see Fig. 1A and B) (18) and stored in 20 mM Tris-HCl pH 7.4, 1 mM DTT and 0.1% N-dodecylmaltoside (Buffer TDD). For fragment generation, C[2–169] protein was incubated with endoproteinase Glu-C [enzyme/substrate ratio of 5/100 (w/w)] for 3 h at 37°C in a buffer containing 20 mM Tris-HCl pH 7.4, 1 mM DTT and 10% glycerol. The fragments were purified by HPLC on a Nucleosil C8 column (300 Å, 4.6 × 250 mm) using a 10–100% linear gradient of acetonitrile in 0.1% trifluoroacetic acid at 1 ml/min flow rate. Chromatography was monitored by measuring ODs at 220 and 280 nm. Proteins corresponding to the main peaks were lyophilized, identified by mass spectrometry and N-terminal sequencing and stored in Buffer TDD.

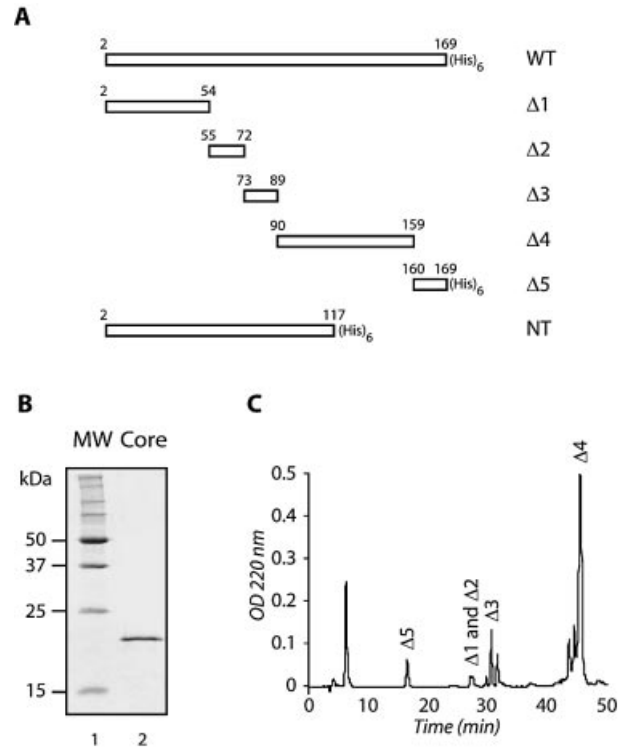
HIV-1 wild-type NC protein NCp7 and mutant NC[12–53] (denoted ΔNCp7) were chemically synthesized and purified by RP-HPLC, as described before (19). Proteins were dissolved at 1 mg/ml in buffer containing 30 mM HEPES pH 6.5, 30 mM NaCl and 0.1 mM ZnCl<sub>2</sub>. T4 gp32 was from Roche, bovine serum albumin (BSA) from NEB.

## ODN labelling

ODNs were labelled with 50 μCi of [<sup>γ</sup>-<sup>32</sup>P]ATP using T4 polynucleotide kinase (Invitrogen) and subsequently purified by PAGE.

## DNA annealing assay

Tar(+) and <sup>32</sup>P-Tar(−) ODNs were incubated (50 fmol each) in 5 μl of buffer A (20 mM Tris-HCl pH 7.0, 30 mM NaCl, 0.1 mM MgCl<sub>2</sub>, 5 mM DTT and 10 μM ZnCl<sub>2</sub>) with protein to nucleotide molar ratios as indicated in the figure legends. Reactions were performed at 37°C for 10 min except for the positive control that was incubated at 62°C. To stop the reaction and denature the protein in order to release it from the <sup>32</sup>P-ODN, we added 2.5 μl of a solution containing 20%



**Figure 1.** Purification of HCV Core protein. (A) Core protein and fragments used in this study. (B) SDS-PAGE and Coomassie blue staining of the purified full-length Core protein. Lane 1, molecular weight marker (MW). Lane 2, purified Core protein. (C) [C<sub>HCV1-169</sub>(6xHis)] was digested with endoproteinase Glu-C and the fragments were submitted to RP-HPLC and then identified by mass spectrometry as indicated above the peaks. The peak containing fragments Δ1 and Δ2 was further separated using a second RP-HPLC chromatography step. The chromatogram was monitored at 220 nm for more sensitivity and small peaks flanking Δ3 and Δ4 were found to be devoid of any contaminant protein.

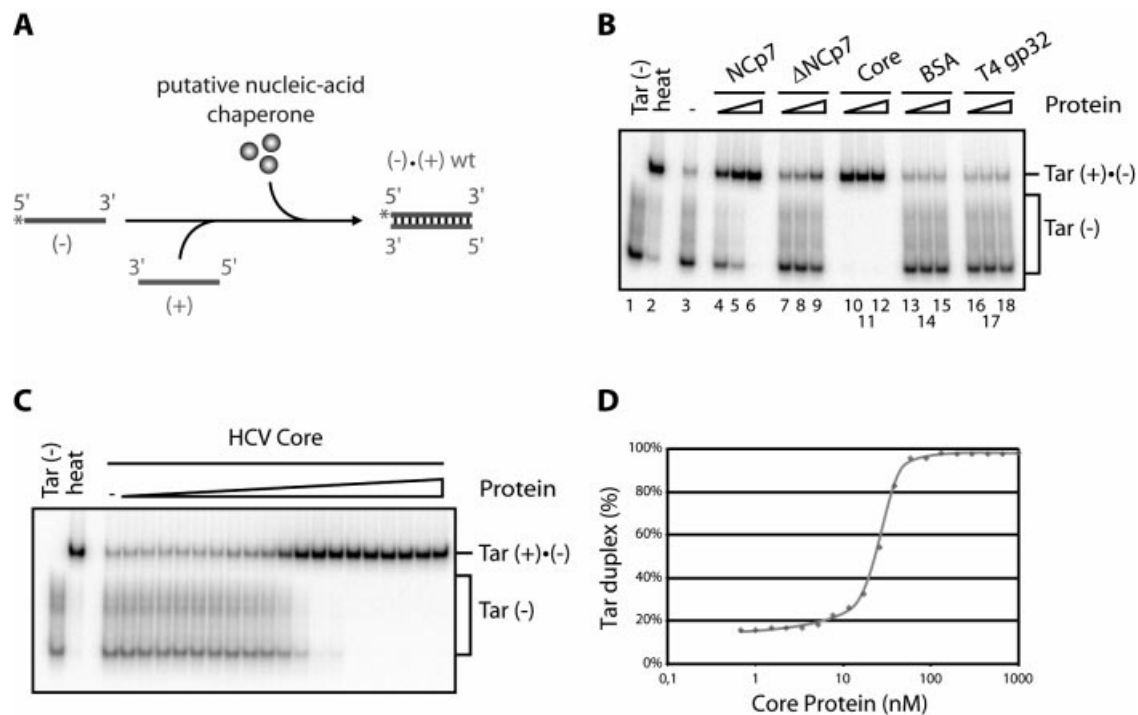
glycerol, 20 mM EDTA pH 8.0, 2% SDS, 0.25% bromophenol blue and 0.4 mg/ml calf liver tRNA, as described before (20). Samples were resolved by 6% native PAGE in 50 mM Tris-Borate pH 8.3, 1 mM EDTA (0.5× TBE) at 4°C. The amount of single-stranded versus double-stranded DNA was assessed by PhosphorImaging (see legends).

## DNA strand exchange assay

<sup>32</sup>P-R(+) and R(−)<sup>mut</sup> ODNs (50 fmol of each) in water were heat denatured for 2 min at 90°C and chilled on ice. The reaction mixture was adjusted to buffer A composition in 10 μl and incubated for 30 min at 62°C to generate an imperfect duplex between the R(+) and R(−)<sup>mut</sup> ODNs. Then, 50 fmol of R(−)<sup>wt</sup> and the nucleic acid chaperone protein were sequentially added with protein to nucleotide molar ratios as indicated in the figure legends. Reactions were incubated for 5 min at 37°C and processed as for the annealing assay.

## HCV RNA dimerization

HCV 3'UTR or X RNA and protein (at protein to nucleotide molar ratios as indicated in the figure legends) were incubated in 10 μl of buffer A containing 8 U RNasin at 37°C for 10 min. Reactions were stopped by adding 0.5% SDS and 5 mM EDTA final. Proteins were removed by proteinase K digestion (2 μg) at room temperature for 10 min and phenol-chloroform



**Figure 2.** The HCV Core protein has DNA annealing activity. (A) Scheme of the annealing assay. Note that protein is removed before native gel analysis. (B) Annealing of Tar(-) and Tar(+) ODNs is activated by Core.  $^{32}\text{P}$ -Tar(-) ODN was incubated with Tar(+) ODN in the presence of increasing amounts of the protein as indicated at the top of the figure. Protein to nucleotide molar ratios were 1:9, 1:6 and 1:4. Lane 1, labelled  $^{32}\text{P}$ -Tar(-) ODN alone (5%); lane 2, Tar(-):Tar(+) upon heating at 62°C, without protein (70%); lane 3, Tar(-):Tar(+) duplex formed spontaneously at 37°C (25%); lanes 4–6, NCp7 (55, 70 and 95%); lanes 7–9,  $\Delta\text{NCp7}$  (30, 35 and 45%); lanes 10–12, Core protein (97% in all lanes); lanes 13–15, BSA (25% in all lanes); lanes 16–18, T4 gp32 (30% in all lanes). (C) Dose dependence of the HCV Core protein annealing activity. Conditions were as in (B). Protein to nucleotide molar ratios were 1:500, 1:330, 1:220, 1:150, 1:98, 1:65, 1:43, 1:29, 1:19, 1:13, 1:9, 1:6, 1:4, 1:2.6, 1:1.7, 1:1.1, 1.3:1, 2:1 and 3:1. (D) Quantification of the data shown in (C).

extracted. RNA was analysed by 1.3% agarose gel electrophoresis in 50 mM Tris-Borate pH 8.3. The gel was fixed in 10% trichloroacetic acid, dried and autoradiographed.

## RESULTS

### HCV Core protein has DNA annealing activity

The Core protein C[2–169] (denoted Core protein or WT) and its truncated version C[2–117] (NT) were expressed in *E. coli* with a C-terminal (His)<sub>6</sub>-tag and purified as previously reported (Fig. 1A and B) (18). Additional Core fragments were obtained by proteolytic digestion of the C[2–169] protein and purified by RP-HPLC (see Fig. 1C).

To assess the putative nucleic acid chaperone activity of the HCV Core protein, the ability of the protein to enhance the annealing of complementary DNA oligonucleotides (ODNs) was assessed (see Fig. 2A). To this end, complementary  $^{32}\text{P}$ -Tar(-) and Tar(+) ODNs (56 nt) were incubated with the Core protein and duplex formation was analysed by native gel electrophoresis after protein removal, as described in Materials and Methods. The  $^{32}\text{P}$ -Tar(-) ODN can form a stable hairpin, which minimizes spontaneous annealing with Tar(+) ODN at 37°C, and therefore it does not migrate as a single band by PAGE (Fig. 2B, lane 1). In the absence of protein, significant annealing was only achieved by incubating

the two complementary ODNs at 62°C (Fig. 2B, compare lanes 2 and 3). The addition of the HCV Core protein strongly enhanced duplex formation at 37°C (Fig. 2B, compare lane 3 and lanes 10–12), as did the canonical chaperone NCp7 (Fig. 2B, lanes 4–6) (20,21). This effect is not simply due to molecular crowding since at the same protein concentration, BSA was inactive (Fig. 2B, lanes 13–15). Consistent with previous reports (22), this enhancement is also not a general property of DNA binding proteins nor of positively charged peptides since the single-stranded DNA binding protein T4gp32 or the  $\Delta\text{NCp7}$  protein ( $pI = 9.9$ ) were unable to promote duplex formation at similar concentrations (Fig. 2B, lanes 16–18 and 7–9, respectively).

The Core protein annealing activity was dose-dependent within a narrow range of concentration and appeared to be cooperative (see Fig. 2C and D). At the minimal active Core protein concentration, duplex formation occurred at least 200-fold faster than in the absence of protein (data not shown). This is an underestimate since complete annealing with Core protein was achieved in <30 s and the activity at shorter time points was difficult to precisely assess. For comparison, only 20% duplex formation was obtained in the absence of protein after 10 min incubation. Similar qualitative results were obtained with longer ODNs (R, 96 nt), although the minimal concentration needed to obtain complete annealing was different (data not shown and Supplementary Fig. S1).

### HCV Core protein has DNA strand exchange activity

Another property of nucleic acid chaperones is to promote the formation of the most stable nucleic acid conformation. This can be assayed by examining the strand exchange activity of the protein (see Fig. 3A). First, an imperfect DNA duplex with two partially complementary ODNs was formed by heating. The Core protein was then added together with another ODN with the potential to form a perfect duplex with one of the two initial ODNs. After incubation, the protein was removed and the ratio of the two DNA duplexes, the perfect to the imperfect, was evaluated by native gel electrophoresis. The migration of the different duplexes are shown in Figure 3B, lanes 1–3, using labelled R(+) ODN. Strikingly, Core protein exchanged the R(-)<sup>mut</sup> mismatched strand in the duplex for the matched strand R(-)<sup>wt</sup>, as has previously been observed with HIV-1 NCp7 (Fig. 3B, lanes 4–6 and 7–9). In contrast, the level of imperfect duplex remained unaffected by BSA, the single-strand DNA binding protein T4 gp32 or the basic deletion mutant of HIV-1 NCp7 (*pI* = 9.9, called ΔNCp7, lanes 4–9 and 10–18), underlining the fact that molecular crowding, DNA binding or basic residues are not sufficient to promote DNA strand exchange.

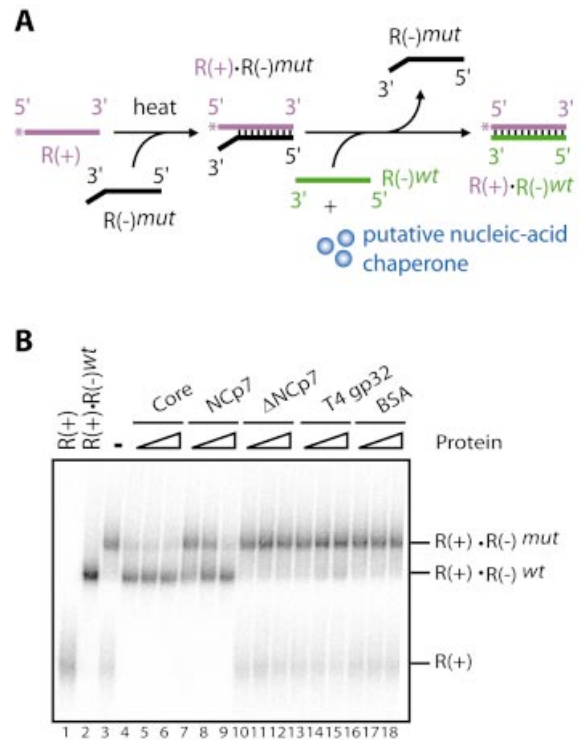
Altogether these results demonstrate that the HCV Core protein has both DNA annealing and strand exchange activities, indicating that it is a DNA chaperone.

### HCV Core protein induces viral (+) RNA dimerization

Since HCV is an RNA virus with no known DNA step in its life cycle, we asked whether the Core protein might also have an RNA chaperone activity. Preliminary experiments using well characterized *in vitro* HIV-1 and Ty1 model systems to examine RNA annealing activities (17,23,24) showed that HCV Core protein fully substituted for HIV-1 NCp7 or Ty1 Gag RNA chaperone activities (Supplementary Fig. S2).

These results prompted us to test whether HCV Core might influence the folding of its own genomic RNA. To this end, RNA representing the conserved 3'UTR of HCV (+) RNA, or its very 3' domain called the X RNA, was incubated with Core protein. RNA was then purified by SDS/proteinase K treatment and phenol-chloroform extraction and analysed by gel electrophoresis in native conditions. To our surprise the Core protein mediated a dramatic structural remodelling of the HCV X RNA. This was deduced from the lower mobility of HCV X RNA in a native gel after Core protein treatment and extraction, whereas the migration of the X RNA extracted in the same way but without the addition of Core protein was unaffected (Fig. 4A, compare lanes 1 and 2–4). Under similar concentrations, the canonical nucleic acid chaperone NCp7 was less effective in inducing this conformational change, since only half of the X RNA was converted to the new lower mobility form (Fig. 4A, compare lanes 2–4 and 10–12). The N-terminal fragment of HCV Core Δ1 was also found to be less active than the full-length protein (Fig. 4A, compare lanes 2–4 and 5–7). In contrast, HIV-1 NC mutant ΔNCp7 and Core fragment Δ2, which were inactive in stimulating ODN duplex formation (Fig. 2B and data not shown), were also unable to induce X RNA conformational change.

In separate experiments we observed that the migration of the full-length 3'UTR, which includes the X RNA region, by native gel electrophoresis, was also altered after incubation



**Figure 3.** The HCV Core protein has DNA strand exchange activity. (A) Schematic of the strand exchange assay. R(+):R(-)<sup>mut</sup> duplex was first formed and then R(-)<sup>wt</sup> ODN and protein were added. After the reaction, protein was removed and duplexes were analysed under native conditions. (B) Core protein promotes formation of the most stable duplex. Protein to nucleotide molar ratios were 1:40, 1:20 and 1:10. Lane 1, <sup>32</sup>P-R(+) ODN alone; lane 2, R(+):R(-)<sup>mut</sup> duplex formed by heating at 62°C, without protein or R(-)<sup>mut</sup>; lane 3, no protein added.

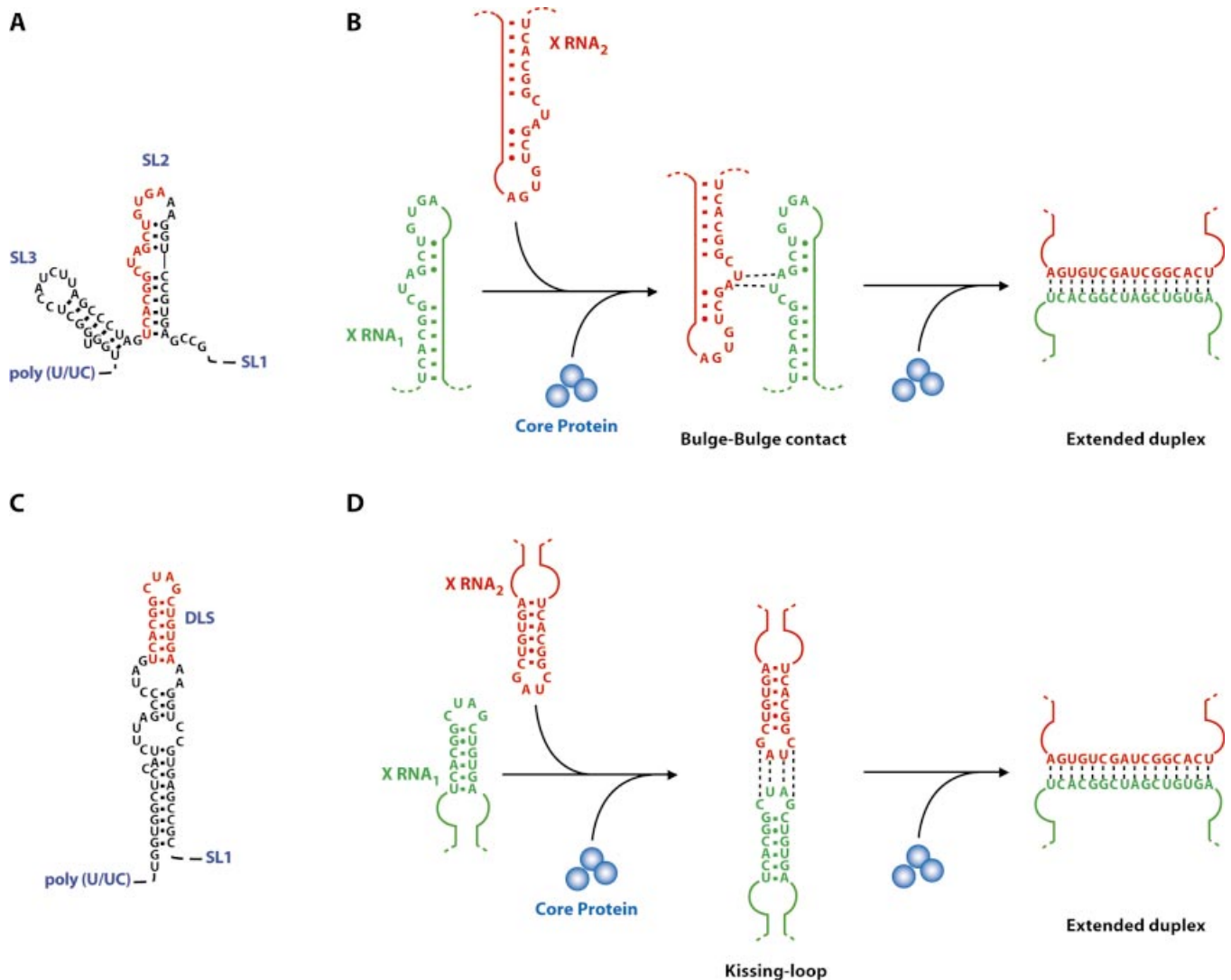
with and removal of the Core protein (data not shown and Fig. 4B, lanes 5–8). In both cases, the apparent size of the new conformer was about twice the size expected for the RNA. The most obvious explanation for these observations is that Core protein drives the dimerization of HCV RNA. This hypothesis was confirmed by the finding that, when incubated together, the full-length (FL) and X RNAs gave rise not only to their respective homodimers, X(D) and FL(D), but also to the predicted single, intermediate-sized heterodimer, X/FL(D) (Fig. 4B, compare lanes 1–4 and 5–8 to lanes 9–12).

### A conserved palindrome in SL2 promotes (+) RNA dimerization

The X RNA secondary structure is composed of three highly conserved stem-loops (called SL1–SL3, see Fig. 5A). To define which of these SLs are important for HCV 3'UTR dimerization, we examined the dimerization of 3' truncated RNAs. The N, F and H RNAs start at the beginning of the HCV 3'UTR and stops at the NheI, Fnu4HI and HindIII restriction sites, respectively (Fig. 5A). F and H RNAs were able to fully dimerize (Fig. 5B, lanes 1–5 and 6–10), whereas the N RNA did not (lanes 11–15). This allowed us to delineate a region in SL2 required for HCV 3'UTR dimerization. Interestingly, the 5' strand of the SL2 stem-loop contains a palindromic sequence that could be involved in this process.







**Figure 6.** Models for HCV (+) RNA dimerization. **(A)** Structure of the 3' X RNA as proposed by Kolykhalov *et al.* (43), and Ito and Lai (44). The palindromic sequence is shown in red. **(B)** Dimerization of the X RNA starting from the structure shown in (A). This model proposes that two SL2 stems initiate interactions by a limited bulge-bulge contact, followed by the formation of a stable extended duplex of 16 bp. **(C)** An alternative structure of 3' X RNA as proposed by Yi and Lemon (27). The proposed DLS is shown in red. **(D)** A kissing-loop model. This model proposes that two DLS palindromic sequences [denoted X RNA1 and X RNA2, shown in (C)] initiate interactions by a kissing-loop mechanism, followed by the formation of a stable extended duplex of 16 bp.

and 4, and Supplementary Figs S1 and S2) and allows the formation of the most stable structure by strand exchange (Fig. 3). The HCV Core protein shows a broad range of sequence specificity, and once nucleic acid molecules have been refolded, it is no longer required to maintain the new conformation. In addition, we found that remodelled nucleic acids, while in a nucleoprotein complex with Core protein, are still accessible to other processes and are not hidden in a protected inactive state. Indeed, Core protein-coated RNA can be copied into cDNA by reverse transcriptase (G. Cristofari, C. Gabus and J.-L. Darlix, unpublished observations).

The molecular mechanism underlying the Core nucleic acid chaperone activity is not known. Any proposed mechanism should account for the ability of the Core protein to enable nucleic acids to escape metastable folding traps and to reach

their most thermodynamically stable state. As already shown for retroviral NC proteins, hnRNPA1 or the StpA proteins (9,11,25,26), we can propose several complementary modes of action: shielding of negatively charged phosphate backbones from each other, secondary or tertiary structure destabilization promoted by chaperone-base interactions and matchmaker properties through homotypic interactions while interacting with the nucleic acid molecules.

Deletion analyses show that amino acids 2–117 (NT mutant) are sufficient to obtain the wild-type level of nucleic acid chaperone activity (see Supplementary Figs S1 and S2). The activities of other Core fragments were as follows: WT = NT > Δ1 >> Δ4 > Δ2 = Δ3 = Δ5 = 0 (Supplementary Figs S1 and S2). Core fragment Δ4 exhibited some remaining activity but we could not discriminate between a redundant nucleic

acid chaperone domain in the C-terminal fragment or an activity present in the fragment overlapping NT and  $\Delta 4$  since we were unable to generate a peptide encompassing residues 118–169. Fragments  $\Delta 2$ ,  $\Delta 3$  and  $\Delta 5$  are very short and therefore may be unable to fold properly. However, circular dichroism and NMR analyses indicate that the Core and Core fragments are devoid of any stable structure in the absence of a nucleic acid molecule (J.-P. Lavergne and F. Penin, unpublished results).

### HCV RNA dimerization and virus replication

Our results show that the HCV (+) 3'UTR RNA dimerizes *in vitro*, a process which is chaperoned by the HCV Core protein (Fig. 4). In addition we also find that the primary RNA–RNA interaction is mediated through a highly conserved palindromic sequence within the SL2 stem–loop in the X region of the plus strand RNA (Fig. 5). A model of the dimerization process induced by the Core protein is shown in Figure 6B. Interestingly, Yi and Lemon have recently reported that the X RNA could adopt an alternative conformation with similar stability, where SL3 and SL2 are fused to generate a single long stem–loop with two internal loops (Fig. 6C) (27). Based on this alternate structure and on experiments shown in Figures 4 and 5 and by analogy with HIV RNA dimerization (28–33), we propose a kissing-loop to extended duplex transition as a mechanism for X RNA dimerization (see Fig. 6D). Note that the two models are not mutually exclusive since the RNA chaperone activity of the Core protein might change the conformation of the monomeric RNA (and could thus shift the equilibrium between the forms shown in Fig. 6A and C). In this alternative structure, the palindromic sequence required for X RNA dimerization folds into an independent apical stem–loop. We propose calling this stem–loop structure the dimer linkage sequence (DLS) by analogy with retroviruses.

We can only speculate on the physiological relevance of the viral RNA dimerization, since studies on HCV production and replication have been, and are still, hampered by the lack of a cell culture system permitting the efficient production of infectious HCV virions. In infected cells, the (+) strand RNA is both the genomic RNA used as a template for (–) strand RNA synthesis by the viral RNA polymerase NS5B and the mRNA coding for the viral polyprotein. The X region of the (+) RNA presumably contains the promoter for (–) strand RNA synthesis which is recognized by the viral RNA polymerase NS5B (34,35). In addition, the X region has been reported to modulate HCV (+) RNA translation (36,37). Thus, HCV RNA dimerization might be a means to regulate (–) strand RNA synthesis by promoter conformational change, either positively by providing assistance to the recruitment of the viral RNA polymerase or negatively by masking the promoter. The latter possibility could correspond to a molecular switch where (+) strand RNA dimerization would facilitate the transition from (–) strand RNA synthesis to viral RNA translation. Alternatively, HCV RNA dimerization might cause the release of (+) strand RNA from the translation apparatus thus facilitating (+) strand RNA packaging into the viral particle. This situation would be reminiscent of the 3'UTR-mediated dimerization of *bicoid* mRNA which permits its incorporation into Staufen ribonucleoparticles and subsequent correct localization in *Drosophila* egg cells (38), or of

the situation that prevails in retroviruses where the genomic RNA is converted into a dimeric form before encapsidation (8,9).

The status of HCV genomic RNA in the viral particle, either monomeric or dimeric, is not yet known mainly for the reasons stated above. We can only speculate that the presence of a genomic RNA dimer during RNA replication in the infected cell and/or in the HCV viral particle would impact the genetic variability of HCV and thus its ability to escape the immune system or antiviral therapies by promoting RNA recombination as in retroviruses (7). The recent findings of natural recombinants between HCV subtypes 2k and 1b or 1a and 1b supports the existence of ongoing recombination processes in HCV (39,40).

### Possible roles of the Core as a nucleic acid chaperone

When expressed on its own in cells or transgenic mice, the HCV Core protein was found to impact several basic processes including transcriptional regulation, and apoptosis or cell growth promotion and immortalization. Additionally, Core protein has been implicated in the induction of hepatocellular carcinoma and a possible immunoregulatory role (3). Since the HCV Core protein is a potent nucleic acid chaperone, it is tempting to speculate that the observed pleiotropic phenotype might result from the Core chaperoning activity interfering with that of cellular nucleic acid chaperones such as p53, hnRNP A1 and YB-1/p50 (9).

By analogy with the retroviral NC proteins (7), the RNA binding and chaperoning properties of the Core protein are probably a driving force in genomic RNA packaging and compaction in the viral particle processes that require important structural remodelling (7–9).

Another intriguing observation is that the Core chaperone activity should antagonize that of the viral RNA helicase NS3. While the first one promotes duplex formation (this study), the second one causes duplex unwinding (41). It has been proposed that the function of some cellular helicases could be to remove RNA-binding proteins from RNA, an activity called RNase (42). Thus, we can envisage that NS3 may release the Core from the genomic RNA at some stages during virus replication. This is presently under investigation. These two antagonistic activities encoded by the same viral genome offer an unprecedented range of nucleic acid structural rearrangement possibilities.

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

### ACKNOWLEDGEMENTS

We are grateful to R. Bartenschlager and K. Shimotohno for HCV plasmids and to C. Kelleher and J. Lingner for critical reading of the manuscript. This work was supported by INSERM and EC (TRIOH) grants (J.L.D.) and by a Fondation pour la Recherche Médicale (FRM) fellowship (G.C.).

### REFERENCES

1. Lauer, G.M. and Walker, B.D. (2001) Hepatitis C virus infection. *N. Engl. J. Med.*, **345**, 41–52.



2. Penin, F. (2003) Structural biology of hepatitis C virus. *Clin. Liver Dis.*, **7**, 1–21, vii.
3. Ray, R.B. and Ray, R. (2001) Hepatitis C virus core protein: intriguing properties and functional relevance. *FEMS Microbiol. Lett.*, **202**, 149–156.
4. McLauchlan, J. (2000) Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes. *J. Viral Hepat.*, **7**, 2–14.
5. 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.
6. 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.
7. Baumert, T.F., Ito, S., Wong, D.T. and Liang, T.J. (1998) Hepatitis C virus structural proteins assemble into viruslike particles in insect cells. *J. Virol.*, **72**, 3827–3836.
8. Rein, A., Henderson, L.E. and Levin, J.G. (1998) Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.*, **23**, 297–301.
9. 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.
10. Herschlag, D. (1995) RNA chaperones and the RNA folding problem. *J. Biol. Chem.*, **270**, 20871–20874.
11. Cristofari, G. and Darlix, J.L. (2002) The ubiquitous nature of RNA chaperone proteins. *Prog. Nucleic Acid Res. Mol. Biol.*, **72**, 223–268.
12. Lohmann, V., Korner, F., Koch, J., Herian, U., Theilmann, L. and Bartenschlager, R. (1999) Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science*, **285**, 110–113.
13. Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M. and Shimotohno, K. (1991) Gene mapping of the putative structural region of the hepatitis C virus genome by *in vitro* processing analysis. *Proc. Natl Acad. Sci. USA*, **88**, 5547–5551.
14. Cortay, J.C., Negre, D., Scarabel, M., Ramseier, T.M., Vartak, N.B., Reizer, J., Saier, M.H., Jr and Cozzone, A.J. (1994) *In vitro* asymmetric binding of the pleiotropic regulatory protein, FruR, to the ace operator controlling glyoxylate shunt enzyme synthesis. *J. Biol. Chem.*, **269**, 14885–14891.
15. Bieth, E., Gabus, C. and Darlix, J.L. (1990) A study of the dimer formation of Rous sarcoma virus RNA and of its effect on viral protein synthesis *in vitro*. *Nucleic Acids Res.*, **18**, 119–127.
16. Barat, C., Le Grice, S.F. and Darlix, J.L. (1991) Interaction of HIV-1 reverse transcriptase with a synthetic form of its replication primer, tRNA(Lys<sub>3</sub>). *Nucleic Acids Res.*, **19**, 751–757.
17. Cristofari, G., Ficheux, D. and Darlix, J.L. (2000) The GAG-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. *J. Biol. Chem.*, **275**, 19210–19217.
18. Boulant, S., Becchi, M., Penin, F. and Lavergne, J.P. (2003) Unusual multiple recoding events leading to alternative forms of hepatitis C virus core protein from genotype 1b. *J. Biol. Chem.*, **278**, 45785–45792.
19. de Rocquigny, H., Ficheux, D., Gabus, C., Fournie-Zaluski, M.C., Darlix, J.L. and Roques, B.P. (1991) First large scale chemical synthesis of the 72 amino acid HIV-1 nucleocapsid protein NCp7 in an active form. *Biochem. Biophys. Res. Commun.*, **180**, 1010–1018.
20. Tsuchihashi, Z. and Brown, P.O. (1994) DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.*, **68**, 5863–5870.
21. Lapadat-Tapolsky, M., Pernelle, C., Borie, C. and Darlix, J.L. (1995) Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1. *Nucleic Acids Res.*, **23**, 2434–2441.
22. Martin, S.L. and Bushman, F.D. (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.*, **21**, 467–475.
23. Darlix, J.L., Gabus, C., Nugeyre, M.T., Clavel, F. and Barre-Sinoussi, F. (1990) Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.*, **216**, 689–699.
24. Cristofari, G., Bampi, C., Wilhelm, M., Wilhelm, F.X. and Darlix, J.L. (2002) A 5'-3' long-range interaction in Ty1 RNA controls its reverse transcription and retrotransposition. *EMBO J.*, **21**, 4368–4379.
25. Waldsich, C., Grossberger, R. and Schroeder, R. (2002) RNA chaperone StpA loosens interactions of the tertiary structure in the td group I intron *in vivo*. *Genes Dev.*, **16**, 2300–2312.
26. Portman, D.S. and Dreyfuss, G. (1994) RNA annealing activities in HeLa nuclei. *EMBO J.*, **13**, 213–221.
27. Yi, M. and Lemon, S.M. (2003) 3' nontranslated RNA signals required for replication of hepatitis C virus RNA. *J. Virol.*, **77**, 3557–3568.
28. Laughrea, M. and Jette, L. (1996) Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms and all sequences upstream or downstream of hairpin 248–271 are dispensable for dimer formation. *Biochemistry*, **35**, 1589–1598.
29. Muriaux, D., Fosse, P. and Paoletti, J. (1996) A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process *in vitro*. *Biochemistry*, **35**, 5075–5082.
30. Muriaux, D., De Rocquigny, H., Roques, B.P. and Paoletti, J. (1996) NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J. Biol. Chem.*, **271**, 33686–33692.
31. Skripkin, E., Paillart, J.C., Marquet, R., Ehresmann, B. and Ehresmann, C. (1994) Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization *in vitro*. *Proc. Natl Acad. Sci. USA*, **91**, 4945–4949.
32. Paillart, J.C., Marquet, R., Skripkin, E., Ehresmann, B. and Ehresmann, C. (1994) Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J. Biol. Chem.*, **269**, 27486–27493.
33. Laughrea, M. and Jette, L. (1994) A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry*, **33**, 13464–13474.
34. Oh, J.W., Ito, T. and Lai, M.M. (1999) A recombinant hepatitis C virus RNA-dependent RNA polymerase capable of copying the full-length viral RNA. *J. Virol.*, **73**, 7694–7702.
35. Oh, J.W., Sheu, G.T. and Lai, M.M. (2000) Template requirement and initiation site selection by hepatitis C virus polymerase on a minimal viral RNA template. *J. Biol. Chem.*, **275**, 17710–17717.
36. Michel, Y.M., Borman, A.M., Paulous, S. and Kean, K.M. (2001) Eukaryotic initiation factor 4G-poly(A) binding protein interaction is required for poly(A) tail-mediated stimulation of picornavirus internal ribosome entry segment-driven translation but not for X-mediated stimulation of hepatitis C virus translation. *Mol. Cell. Biol.*, **21**, 4097–4109.
37. Ito, T., Tahara, S.M. and Lai, M.M. (1998) The 3'-untranslated region of hepatitis C virus RNA enhances translation from an internal ribosomal entry site. *J. Virol.*, **72**, 8789–8796.
38. Ferrandon, D., Koch, I., Westhof, E. and Nusslein-Volhard, C. (1997) RNA-RNA interaction is required for the formation of specific bicoid mRNA 3' UTR-STAUFIN ribonucleoprotein particles. *EMBO J.*, **16**, 1751–1758.
39. Kalinina, O., Norder, H., Mukomolov, S. and Magnus, L.O. (2002) A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J. Virol.*, **76**, 4034–4043.
40. 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.
41. Pang, P.S., Jankowsky, E., Planet, P.J. and Pyle, A.M. (2002) The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. *EMBO J.*, **21**, 1168–1176.
42. Jankowsky, E., Gross, C.H., Shuman, S. and Pyle, A.M. (2001) Active disruption of an RNA-protein interaction by a DEXH/D RNA helicase. *Science*, **291**, 121–125.
43. 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.
44. 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.