

# An unexpected RNA distal interaction mode found in an essential region of the hepatitis C virus genome

Ángel Cantero-Camacho and José Gallego\*

Facultad de Medicina, Universidad Católica de Valencia, C/Quevedo 2, 46001 Valencia, Spain

Received December 01, 2017; Revised January 12, 2018; Editorial Decision January 23, 2018; Accepted January 24, 2018

## ABSTRACT

**The 3'X tail is a functionally essential 98-nt sequence located at the 3'-end of the hepatitis C virus (HCV) RNA genome. The domain contains two absolutely conserved dimer linkage sequence (DLS) and *k* nucleotide segments involved in viral RNA dimerization and in a distal base-pairing interaction with stem-loop 5BSL3.2, respectively. We have previously shown that domain 3'X forms an elongated structure comprising two coaxially stacked SL1' and SL2' stem-loops. This conformation favors RNA dimerization by exposing a palindromic DLS segment in an apical loop, but buries in the upper stem of hairpin SL2' the *k* nucleotides involved in the distal contact with 5BSL3.2. Using nuclear magnetic resonance spectroscopy and gel electrophoresis experiments, here we show that the establishment of the complex between domain 3'X and stem-loop 5BSL3.2 only requires a rearrangement of the nucleotides forming the upper region of subdomain SL2'. The results indicate that the interaction does not occur through a canonical kissing loop mechanism involving the unpaired nucleotides of two terminal loops, but rather involves a base-paired stem and an apical loop and may result in a kissing three-way junction. On the basis of this information we suggest how the 3'X tail switches between monomer, homodimer and heterodimer states to regulate the HCV viral cycle.**

## INTRODUCTION

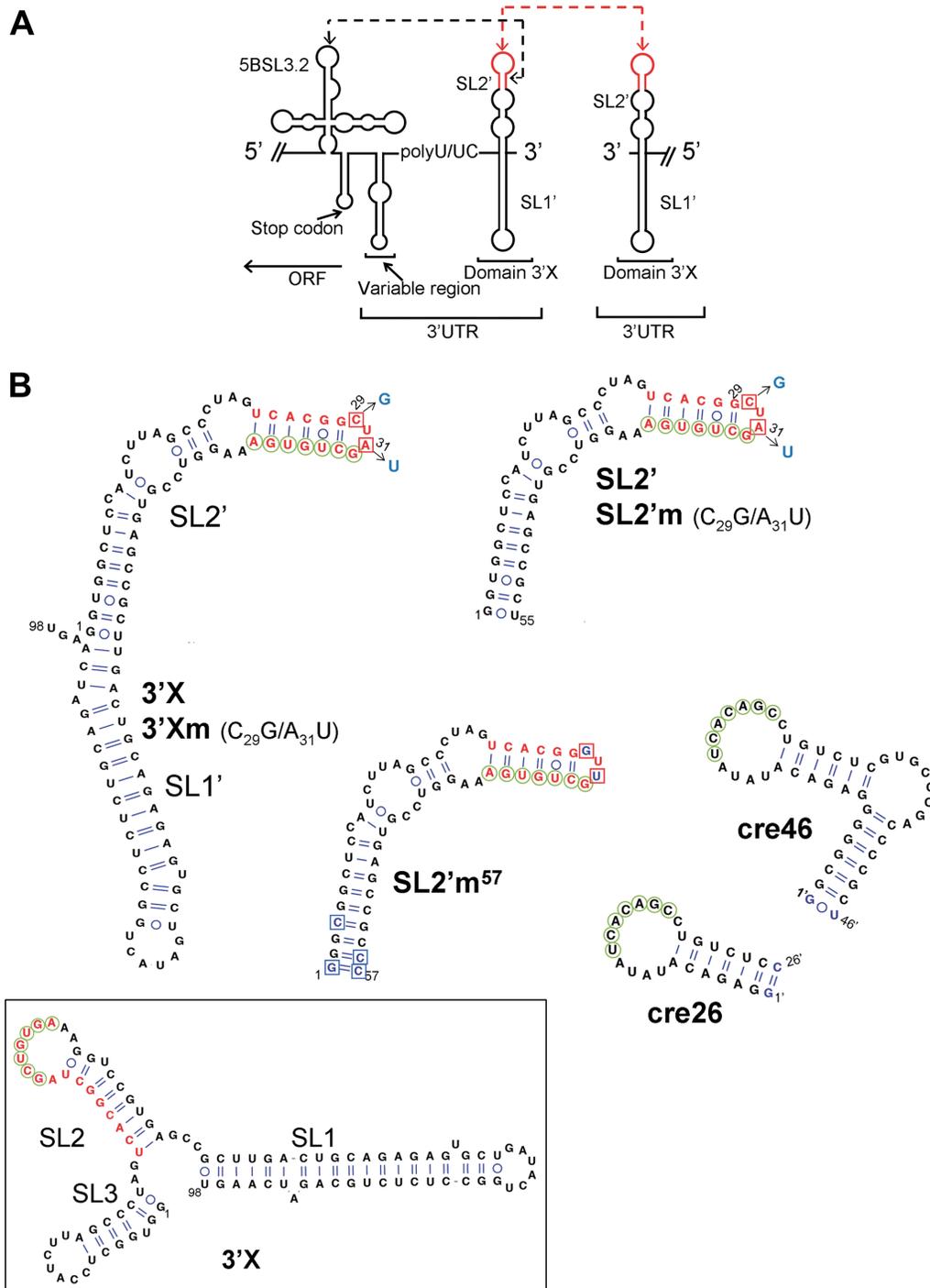
RNA occupies a central role in the functioning of living organisms (1–3), and many of the biological functions performed by this biopolymer are linked to the presence of three-dimensional (3D) folds. The overall shape of RNA structures depends on the packaging of double helices in three dimensions, which is modulated by a variety of tertiary interactions. These contacts encompass coaxial stacking, tetraloop-receptor motifs, pseudoknots and loop–loop interactions among many others (4,5).

Loop–loop interactions, also called kissing loop contacts, occur when the unpaired nucleotides in one hairpin loop form Watson–Crick base pairs with the unpaired nucleotides of another hairpin loop. They are formed by RNAs of many living organisms, and have been observed to play a wide variety of roles (6). Loop–loop interactions are frequently involved in intermolecular RNA complexes. They were first found to mediate interactions between target sequences and antisense RNAs modulating gene expression in bacteria (7,8). In retroviruses, kissing contacts between loops with palindromic sequences are thought to be responsible for the dimerization and packaging of two copies of retroviral genomes into virion particles (9). In an intramolecular context, loop–loop interactions are particularly prominent in mediating contacts between distantly positioned sequences (10), but have also been found to participate in the architecture of local RNA folds (11) and viroids (12).

Two properties of loop–loop interactions set them apart from other classes of tertiary contacts and help to explain their biological functions: they involve Watson–Crick pairings and are therefore specific, and can also be unusually stable (13,14). If the loop–loop contact is not constrained within a local fold, the helices of each intervening hairpin and the new helix created by the pairing of the complementary loops are usually coaxially stacked (15–18). This effectively results in the formation of a single composite helix that likely enhances the stability of the interaction. These characteristics contribute to explain why loop–loop interactions are frequently involved in specific contacts between sequences belonging to different molecules, or located many nucleotides apart within the same molecule.

The genomes of positive-sense RNA viruses frequently form loop–loop interactions between remote sequences that mediate important viral activities (10). The genomic RNA of hepatitis C virus (HCV), an important human pathogen (19), provides several examples of this class of interactions. One of these contacts involves the 3'X domain, a strongly conserved 98-nt structure located at the 3' terminus of the viral genome (20,21). This domain is recognized by the viral replication complex to start synthesis of the opposite polarity strand and has been found to be essential for replication (22–26). A 7-nt *k* sequence within the 3'X domain

\*To whom correspondence should be addressed. Tel: +34 963 637 412; Fax: +34 963 944 590; Email: jose.gallego@ucv.es



**Figure 1.** RNA sequences and distal contacts involving the 3'X terminal domain of HCV. **(A)** Schematic representation of functional *cis*-acting structures identified in the 3' region of the HCV genome. Distal RNA–RNA interactions proposed to have functional relevance are identified with double-headed arrows. **(B)** Predicted secondary structure of RNA sequences analyzed in this study: 98-nt full-length domain 3'X, 55-nt subdomain SL2', 46-nt subdomain 5BSL3.2 (cre46) and 26-nt apical 5BSL3.2 hairpin cre26. 3'Xm and SL2'm are dimerization-defective variants of the wild-type sequences 3'X and SL2', respectively, containing a double C<sub>29</sub>G/A<sub>31</sub>U mutation in the apical palindromic tetraloop of SL2'. SL2'm<sup>57</sup> is a dimerization-defective subdomain SL2' variant with two G:U pairs replaced with G:C and a G:C pair added in the lower stem (blue squares). The blue-colored nucleotides in cre26 and cre46 indicate changes relative to the wild-type sequence, introduced to increase transcription yield. 3'X sequence *k* (indicated with green circles) has been shown to establish a distal interaction involving Watson–Crick pairs with a complementary *k'* sequence (green circles) in the apical loop of 5BSL3.2. In the control SL2'm-*k'* sequence (Supplementary Figure S1), the *k'* segment was replaced with the *k'* tract of 5BSL3.2. The 16-nt palindromic DLS sequence (depicted with red nt) overlaps with the *k* tract and is absolutely conserved among all HCV isolates. The alternative three-stem conformation of domain 3'X (shown in an inset) exposes differently the *k* and DLS sequences.

has been shown by genetic experiments to establish a distal interaction with a complementary *k'* sequence located in stem-loop 5BSL3.2 of the open reading frame (ORF) (Figure 1A), and this interaction has been observed to influence the replication and translation processes of the virus (27–31). This contact was proposed to be a kissing interaction between one apical loop of domain 3'X and the apical loop of hairpin 5BSL3.2 (also termed SL9266).

In addition to the 7-nt *k* sequence interacting with the ORF, the 3'X region contains a 16-nt palindromic sequence that promotes dimerization of viral genomes *in vitro* (32–34) through the formation of an intermolecular kissing loop contact between two palindromic loops (34) (Figure 1A). This tract is termed the dimer linkage sequence (DLS) by analogy to the dimerization sequences of retroviruses. The DLS overlaps with the *k* sequence (Figure 1B and Supplementary Figure S1) and is absolutely conserved among the different HCV genotypes and isolates (35). Given the high genetic variability of HCV (19), the conservation of the DLS/*k* sequence is a strong indication of a key role in the viral life cycle. Interestingly, the establishment of kissing interactions by the *k* and DLS segments would be mutually exclusive, since the exposure of either palindromic DLS or *k* nucleotides in a loop requires different secondary structures. This suggests a switch function for the highest conserved sequence in HCV.

Using nuclear magnetic resonance (NMR) and small angle X-ray scattering (SAXS) experiments, we have recently shown that in isolation, the 3'X domain adopts an elongated structure comprising two SL1' and SL2' double-helical stems stabilized by coaxial stacking (36,37) (Figure 1B). This conclusion has been recently corroborated by an independent NMR analysis of the 3'-end region of the HCV RNA (38). The two-stem structure exposes a palindromic DLS tetranucleotide in the apical loop of subdomain SL2', and promotes dimerization of both full-length 3'X and subdomain SL2' sequences at higher ionic strength and/or RNA concentration (36,37).

The two-stem conformation of domain 3'X is seemingly incompatible with the formation of a kissing loop complex with 5BSL3.2, since the *k* nucleotides are almost completely buried in the upper double-helical stem of hairpin SL2' (Figure 1B). Formation of the *k*–*k'* interaction apparently requires the adoption by the 3'X domain of an alternative, three-stem structure where the *k* sequence is exposed in the apical loop of hairpin SL2 and available for pairing with the *k'* nucleotides of the 5BSL3.2 apical loop (Figure 1B, inset). This conformation was proposed on the basis of chemical modification experiments (33,39–42). It has also been suggested that the two-stem and three-stem folds were part of an RNA-based switch signaling the transition between the replication, translation and possibly genome packaging processes of the virus (31,34,41–44). It was therefore important to elucidate the structural details of the 3'X–5BSL3.2 interaction, to find out whether the establishment of the interaction implied a conformational change in domain 3'X and to shed light on the nature of the proposed riboswitch.

Here we have used NMR spectroscopy and gel electrophoresis experiments to study the structure of the complex formed between domain 3'X and hairpin 5BSL3.2. The results indicate that the domain retains the SL1' and SL2'

conformation upon binding to 5BSL3.2. Rather than inducing an SL1, SL2 and SL3 conformation where the *k* nucleotides are exposed in the apical loop of hairpin SL2, the establishment of the distal contact implies a conformational change of the SL2' upper region where the *k* nucleotides are base-paired, and the formation of new Watson–Crick pairs with the *k'* nucleotides of the 5BSL3.2 terminal loop. Thus instead of implying the unpaired nucleotides of two hairpin loops, as in canonical loop–loop interactions, the 3'X–5BSL3.2 distal interaction involves a base-paired stem and a terminal loop. This unexpected tertiary contact may result in the formation of a three-way 'kissing junction' rather than a single composite helix.

## MATERIALS AND METHODS

### Sequences

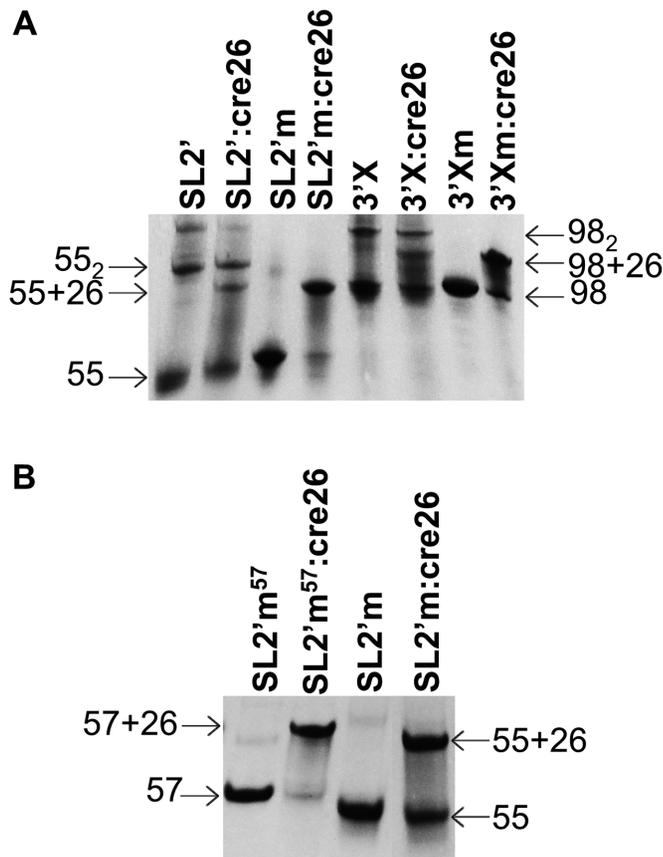
The sequences analyzed in this study correspond to HCV genotype 1b and were obtained from vector pFK-I<sub>389</sub>FLUCNS3–3'ET (45), kindly provided by Dr R. Barten-schlager (University of Heidelberg, Germany).

### Secondary structure predictions

Secondary structure predictions were carried out with the RNAfold (<http://rna.tbi.univie.ac.at/>) (46) and/or mfold (<http://mfold.rna.albany.edu>) (47) web servers. RNA structures were drawn using VARNA (<http://varna.lri.fr/>) (48).

### Preparation of RNA samples for NMR spectroscopy and gel electrophoresis experiments

Dimerization-defective 3'X subdomain constructs SL2'm (55 nt), SL2'm-*k'* (55 nt) and SL2'm<sup>57</sup> (57 nt), as well as 5BSL3.2 sequences cre26 (26 nt) and cre46 (46 nt) (Figure 1B and Supplementary Figure S1), were prepared by T7-polymerase *in vitro* transcription using synthetic oligonucleotide DNA templates. The dimerization-defective 3'X domain sequence 3'Xm (98 nt) was transcribed from a pUC19 plasmid containing a *Sca*I restriction site for linearization and run-off transcription. This vector was generated by cloning a C<sub>29</sub>G/A<sub>31</sub>U-mutant 3'X sequence between the *Sma*I and *Eco*RI sites of pUC19. Wild-type domain 3'X and subdomain SL2' constructs were prepared as described previously (36). For sequences 3'Xm, SL2'm and cre26, we also generated uniformly <sup>13</sup>C/<sup>15</sup>N-labeled transcripts using nucleoside triphosphates obtained from CortecNet. An SL2'm sample with specific <sup>13</sup>C/<sup>15</sup>N-labeling of G and C nucleotides was also prepared to facilitate the assignment process. All constructs were purified on denaturing gels containing 20% acrylamide and 8 M urea. After electroelution from the gel, the RNAs were ethanol-precipitated two times and desalted with Sephadex G-25 cartridges. Prior to NMR or gel electrophoresis experiments, all samples were transferred by diafiltration into aqueous solutions containing 10 mM sodium phosphate (pH 6.0) and 0.1 mM ethylenediaminetetraacetic acid with no added salts, or additionally containing 2 mM MgCl<sub>2</sub>. Before each NMR and gel electrophoresis experiment, the RNA samples were heated at 95°C for 5 min and immediately placed on ice for 30 min. For experiments analyzing



**Figure 2.** Gel electrophoresis analysis of 3'X domain sequences and their complexes with the 5BSL3.2 hairpin cre26. (A) Native gel comparing the electrophoretic mobility of wild-type 3'X and SL2' and dimerization-defective 3'Xm and SL2'm constructs in the absence and presence of one molar equivalent of cre26. (B) Native gel comparing the electrophoretic mobility of constructs SL2'm and SL2'm<sup>57</sup> in the absence and presence of one molar equivalent of cre26. All experiments used a running buffer containing Mg<sup>2+</sup> ions, which have been shown to promote homodimerization and 5BSL3.2 complex formation.

intermolecular complex formation the samples were snap-cooled in the absence of MgCl<sub>2</sub>, and then incubated with 2 mM MgCl<sub>2</sub> for 150 min at 25°C, in the absence or presence of the interaction partner.

### NMR spectroscopy

The RNA concentration of the NMR samples ranged between 45 and 350 μM. NMR spectra were acquired on a cryoprobe-equipped, 600 MHz Bruker Avance III spectrometer, and analyzed using Topspin 3.5 (Bruker Biospin) and Sparky 3.110 (49). The unlabeled systems were studied using 2D watergate-NOESY (with 150 ms mixing time) and watergate-TOCSY experiments (60 ms mixing time) recorded in 90% H<sub>2</sub>O/10% D<sub>2</sub>O, typically at two temperatures (16 and 27°C). TOCSY and NOESY (120 and 250 ms mixing time) experiments were also acquired in D<sub>2</sub>O, typically at 27 and 37°C. The SL2'm sequence was additionally examined with watergate-NOESY and watergate-ROESY experiments at low mixing time (20 ms) to identify cross-peaks due to conformational exchange. The recycle delays

were 1.6 and 2 s for all homonuclear TOCSY and NOESY experiments, respectively.

The <sup>13</sup>C/<sup>15</sup>N-labeled samples were analyzed with 2D <sup>1</sup>H-<sup>15</sup>N HSQC and HNN-COSY (50) experiments recorded in 90% H<sub>2</sub>O/10% D<sub>2</sub>O at 27°C, as well as 2D <sup>1</sup>H-<sup>13</sup>C HSQC, <sup>1</sup>H-<sup>1</sup>H HCCH-TOCSY (51) and filtered/edited <sup>1</sup>H-<sup>1</sup>H NOESY (52) (usually with 250 ms mixing time) experiments recorded in D<sub>2</sub>O, typically at 27 and/or 37°C. The SL2'm samples with uniform and G,C-specific <sup>13</sup>C/<sup>15</sup>N-labeling were additionally analyzed with 3D <sup>13</sup>C-edited NOESY-HSQC experiments (200 ms mixing time). For <sup>1</sup>H-<sup>15</sup>N HSQC experiments we acquired 128 indirect experiments with 96 scans per experiment. For HNN-COSY experiments, the delay for evolution of the <sup>2</sup>J<sub>NN</sub> coupling was set to 15 ms and we collected 128 complex points in the t<sub>1</sub> dimension with 320 scans for each t<sub>1</sub> increment. <sup>1</sup>H-<sup>13</sup>C HSQC experiments involved 512 points in the indirect dimension and 40 scans per point, and NOESY-HSQC experiments were executed with 64 and 400 increments in the indirect carbon and hydrogen dimensions, respectively, and eight repetitions per increment. The recycle delays were between 1.0 and 1.3 s for all experiments with labeled samples.

*Assignments and secondary structure determination.* The assignment of 5BSL3.2 construct cre26 and dimerization-defective subdomain sequence SL2'm involved nitrogen, carbon and exchangeable and non-exchangeable protons, and was based on standard analyses (53) of 2D NOESY, TOCSY, HSQC, HCCH-TOCSY and HNN-COSY data, as well as 3D NOESY-HSQC spectra for SL2'm samples, most of them acquired in the absence and presence of 2 mM MgCl<sub>2</sub> and at two different temperatures. Following this procedure, we assigned the signals (typically aromatic nuclei as well as H1' and H2' protons) of all cre26 residues. For SL2'm 35 of 55 residues were assigned, including those forming the lower stem (G<sub>1</sub>-C<sub>8</sub> and G<sub>48</sub>-U<sub>55</sub>), middle segment (U<sub>14</sub>-C<sub>17</sub> and G<sub>42</sub>) and upper stem-loop (C<sub>24</sub>-G<sub>37</sub>), and we also generated tentative assignments for U<sub>13</sub>, C<sub>45</sub> and eight purine nucleotides based on NOEs, chemical shift statistics (54,55) and comparisons between uniformly and G,C-specifically labeled SL2'm spectra. The remaining 10 pyrimidine nucleotides could not be identified due to extensive overlap. Determination of the secondary structure of the dimerization-defective 3'Xm full-length domain sequence was based on comparisons of HNN-COSY and watergate-NOESY data of 3'Xm with those of previously studied (36) wild-type domain 3'X and subdomains SL1 and SL2', as well as dimerization-defective subdomain SL2'm, all acquired under identical conditions. The secondary structure of wild-type domain 3'X and subdomain SL2' and SL1 sequences was previously determined by NMR spectroscopy in our laboratory based on exchangeable proton spectra (36).

*Analysis of the distal interaction.* These studies were performed by comparing <sup>1</sup>H-<sup>15</sup>N HSQC and HNN-COSY spectra obtained with <sup>13</sup>C/<sup>15</sup>N-labeled domain 3'Xm or subdomain SL2'm sequences in the absence and presence of unlabeled cre26. For uniformly and G,C-specifically labeled SL2'm sequences, we also compared the <sup>1</sup>H-<sup>13</sup>C HSQC spectra in the absence and presence of unlabeled cre26. Con-

versely, we also analyzed the  $^1\text{H}$ - $^{13}\text{C}$  HSQC,  $^1\text{H}$ - $^{15}\text{N}$  HSQC and HNN-COSY spectra of a  $^{13}\text{C}/^{15}\text{N}$ -labeled cre26 construct in the absence and presence of an unlabeled SL2'm sequence. By labeling with  $^{15}\text{N}/^{13}\text{C}$  isotopes one interaction partner at a time, the NMR spectra of the complexes were simplified (56,57) since all of the HN, HNN and HC crosspeaks corresponded to 3'Xm or SL2'm sequences and not unlabeled cre26, or *vice versa*. Likewise, the HN resonances detected in the mixture without a coupled N must correspond to intermolecular N-H...N hydrogen bonds.

### Gel electrophoresis experiments

Native gels were run at 4°C for 12 h under constant voltage (80 V). We used 20% 19:1 acrylamide:bisacrylamide gels and a running buffer containing 89 mM tris-borate and 2 mM  $\text{MgCl}_2$  (TBM). These experiments involved 10–20  $\mu\text{M}$  RNA samples, prepared as specified above. All gels were stained with methylene blue and destained with water.

## RESULTS

### A mutation in the palindromic DLS loop of subdomain SL2' inhibits RNA dimerization without affecting the capacity to establish a tertiary interaction with stem-loop 5BSL3.2

We have recently reported that the isolated 3'X domain adopts an extended structure comprising two coaxially stacked SL1' and SL2' double-helical stems (36,37) (Figure 1B). This structure exposes a palindromic DLS  $\text{C}_{29}\text{UAG}_{32}$  tetranucleotide in the apical loop of subdomain SL2', and promotes dimerization of both full-length 3'X and subdomain SL2' sequences at higher ionic strength (36,37). This was not surprising, as the homodimerization process of these sequences has been shown to be triggered by a kissing complex interaction between two palindromic SL2' loops (34).

Formation of 3'X–5BSL3.2 complexes requires the presence of  $\text{Mg}^{2+}$  ions, and in these conditions 3'X and SL2' homodimers are also formed (36). To facilitate a detailed study of the 3'X–5BSL3.2 complexes, we generated full-length 3'X and subdomain SL2' constructs where the palindromic  $\text{C}_{29}\text{UAG}_{32}$  sequence of the SL2' tetraloop was changed to  $\text{G}_{29}\text{UUG}_{32}$ . This mutation abolished the DLS palindrome without affecting the *k* nucleotides proposed to establish a distal base-pairing interaction with stem-loop 5BSL3.2 (Figure 1B). Gel electrophoresis experiments indicated that domain 3'X and subdomain SL2' constructs containing the double  $\text{C}_{29}\text{G}/\text{A}_{31}\text{U}$  mutation (hereafter identified as 3'Xm and SL2'm) did not dimerize in the presence of  $\text{Mg}^{2+}$ , and retained the capacity to form complexes with the 5BSL3.2 sequence cre26 observed in the wild-type sequences (Figure 2A). Likewise, NMR spectroscopy experiments showed that the conformation of these mutant 3'Xm and SL2'm constructs did not vary significantly relative to the wild-type sequences (Supplementary Figures S2–4). This supported our earlier conclusions regarding the solution structure of the 3'X and SL2' systems (36,37), since mutations of nucleotides located in a terminal loop were not expected to affect the conformation of the constructs. The NMR spectra also confirmed the absence of homodimerization of the mutant SL2'm and 3'Xm sequences in the presence of  $\text{MgCl}_2$

(Supplementary Figures S3B and 4A). On the other hand, gel electrophoresis experiments indicated that the cre26 sequence representing the apical stem-loop of subdomain 5BSL3.2 had the same capacity to bind 3'X sequences as the full-length subdomain (Figure 1B and Supplementary Figure S5). Likewise, control experiments showed that a SL2'm-*k*' construct where the *k* sequence was replaced with the *k*' tract present in the 5BSL3.2 loop (Supplementary Figure S1) was incapable of interacting with cre26, as expected (Supplementary Figure S6).

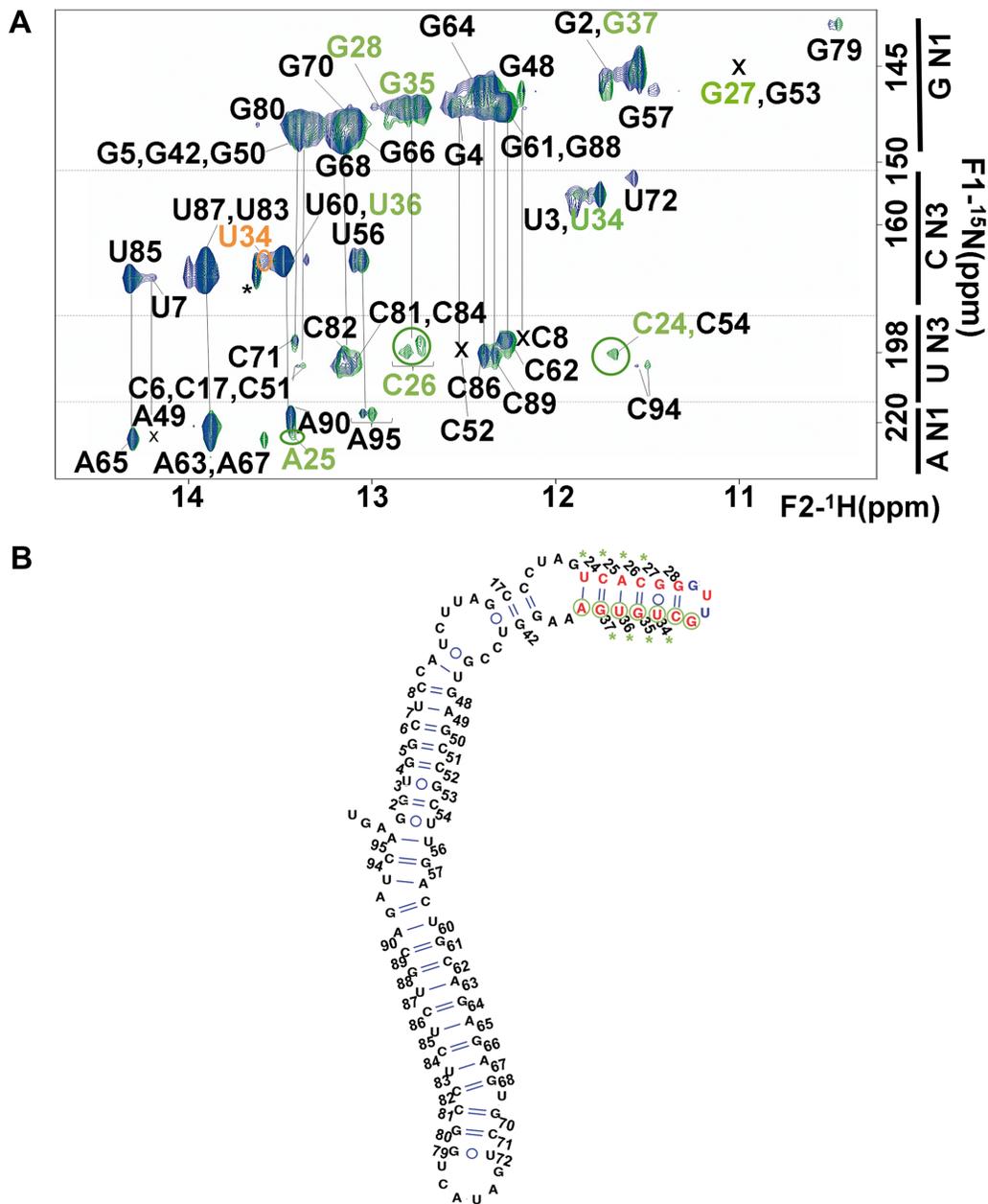
### Domain 3'X retains the SL1' and SL2' conformation when interacting with stem-loop 5BSL3.2

Genetic experiments indicated that the 7-nt  $\text{G}_{32}\text{CUGUGA}_{38}$  *k* sequence of domain 3'X establishes a distal base-pairing interaction with a complementary UCACAGC *k*' sequence located in the apical loop of stem-loop 5BSL3.2 (27). It was suggested that this distal interaction involved the three-stem fold of domain 3'X and particularly hairpin SL2, since in this context the *k* bases are exposed in the apical loop of SL2 and available to form kissing loop–loop pairs with the *k*' nucleotides (27–29,43,44) (Figure 1B, inset).

To assess whether 5BSL3.2 complex formation implied a conformational change relative to the two-stem structure adopted by the isolated domain, we used NMR spectroscopy to monitor the interaction between a  $^{13}\text{C}/^{15}\text{N}$ -labeled 3'Xm domain sequence and an unlabeled cre26 construct in the presence of  $\text{Mg}^{2+}$  (Figure 3). After addition of cre26, the HN imino crosspeaks detected in the 3'Xm:cre26 mixture matched those corresponding to the isolated 3'Xm domain (see superposed spectra in Figure 3). Furthermore, the presence of cre26 did not promote detection of HN imino crosspeaks corresponding to the terminal G53:U98, C54:G97 and U55:A96 pairs of hairpin SL1 present in the tree-stem domain conformation (Figures 1B, 3 and Supplementary Figure S4). These base pairs are absent in the shortened SL1' subdomain present in the two-stem domain fold detected by NMR spectroscopy (36) and SAXS (37), where G53, C54 and U55 are paired at the base of the SL2' stem, and the terminal A96, G97 and U98 nt form an unpaired tail (Figure 1B). Altogether, these results indicated that the 3'X domain retained the SL1' and SL2' conformation upon binding to cre26.

### The interaction of domain 3'X with 5BSL3.2 implies an exchange of base-pairing partners of the *k* nucleotides forming the upper stem of subdomain SL2'

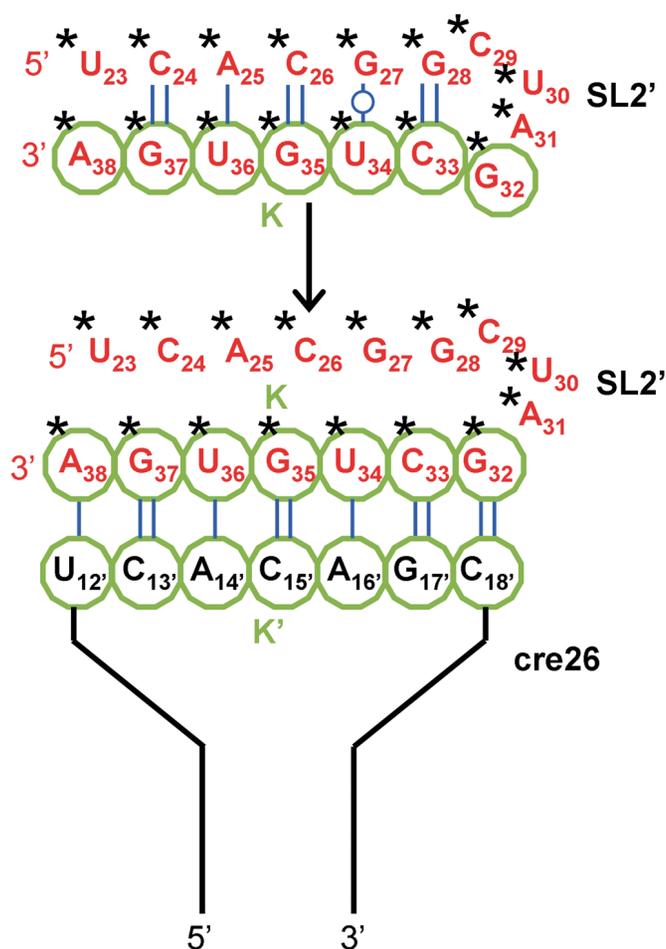
Upon  $^{13}\text{C}/^{15}\text{N}$ -3'Xm complex formation with unlabeled cre26, a new U HN3 imino crosspeak appeared. This crosspeak had a chemical shift typical of Watson–Crick A:U pairs, but did not have an A N1 partner in the HNN-COSY spectrum (Figure 3, orange circle). Likewise, the HN1 imino crosspeaks of G35 and G37 lost their respective HNN-COSY C26 and C24 N3 connectivities, and the same was observed the U36 HN3 imino, whose HNN-COSY A25 N1 partner became undetectable (Figure 3, green circles). Despite these changes, the chemical shift of these imino resonances did not vary upon cre26 addition.



**Figure 3.** NMR spectroscopy analysis of the distal interaction formed between full-length domain sequence 3'Xm and 5BSL3.2 hairpin cre26. (A) Superposition of  $^1\text{H}$ - $^{15}\text{N}$  HNN-COSY spectra of 3'Xm, acquired in the absence (green) and presence (blue) of one molar equivalent of unlabeled cre26 sequence. Since the experiment used unlabeled cre26, all of the HN and HNN crosspeaks correspond to 3'Xm. The crosspeaks of upper SL2' stem nucleotides are identified with green labels, and the HNN crosspeaks that disappear with the addition of cre26 are marked with green circles. The HN crosspeak that appears upon cre26 binding is marked in orange and was assigned as U34. Crosspeaks marked with crosses are visible at a lower contour level. The crosspeak corresponding to a minor conformation of U36 is marked with an asterisk. (B) Secondary structure of isolated domain 3'Xm supported by the NMR data. DLS nucleotides are depicted in red, *k* nucleotides are indicated with green circles, and the C<sub>29</sub>G and A<sub>31</sub>U mutations blocking dimerization are colored blue. Residues with assigned HNN resonances are numbered, and those whose HNN-COSY crosspeak patterns undergo changes upon addition of cre26 are marked with green asterisks. Conditions: 45  $\mu\text{M}$  3'Xm, 2 mM MgCl<sub>2</sub>, 27°C.

All G35, U36 and G37 nt belong to the seven-nt G<sub>32</sub>CUGUGA<sub>38</sub> *k* sequence of the domain. Taking into account that a  $^{13}\text{C}/^{15}\text{N}$ -labeled 3'Xm sample was titrated with unlabeled cre26, the observed changes in the HNN-COSY crosspeak patterns can be attributed to the formation of an intermolecular 3'X:cre26 duplex where the 3'X G<sub>35</sub>U<sub>G37</sub> *k* nucleotides pair in antiparallel with comple-

mentary C<sub>13</sub>'AC<sub>15</sub>' *k*' nucleotides present in the cre26 apical loop (Figure 4). This would explain the disappearance of the C N3 and A N1 HNN-COSY crosspeaks connected to the G35, U36 and G37 imino protons of 3'Xm, since they would be hydrogen-bonded to magnetically inactive N3 or N1 atoms upon complexation with the unlabeled cre26 construct (see Figure 4). Moreover, the chemical shift of these



**Figure 4.** Schematic representation of the base-pairs proposed to be formed by the  $k$  nucleotides of the SL2' subdomain in the absence (top) and presence (bottom) of 5BSL3.2 hairpin cre26. SL2' DLS nucleotides are shown in red and marked with asterisks, and  $k$  and  $k'$  residues are identified with green circles. The NH $\cdots$ N hydrogen bonds present in the base pairs are only detected by NMR HNN-COSY experiments when the two bases are  $^{15}\text{N}$ -labeled.

iminos would not vary with complex formation because the chemical environment of the intermolecular  $k$ - $k'$  duplex would be very similar relative to that of the upper duplex of SL2', where the  $k$  nucleotides are base-paired with complementary nucleotides in the opposite strand (Figure 4). Furthermore, the appearance of a new Watson-Crick U HN3 imino crosspeak without an A N1 HNN-COSY partner can be attributed to the formation of an intermolecular base pair between 3'Xm U34 and cre26 A16', which would replace the intramolecular U34:G27 wobble pair present in the upper stem of SL2' (Figure 4).

#### The distal interaction implies a conformational change in subdomain SL2' that does not affect the nucleotides of the lower stem

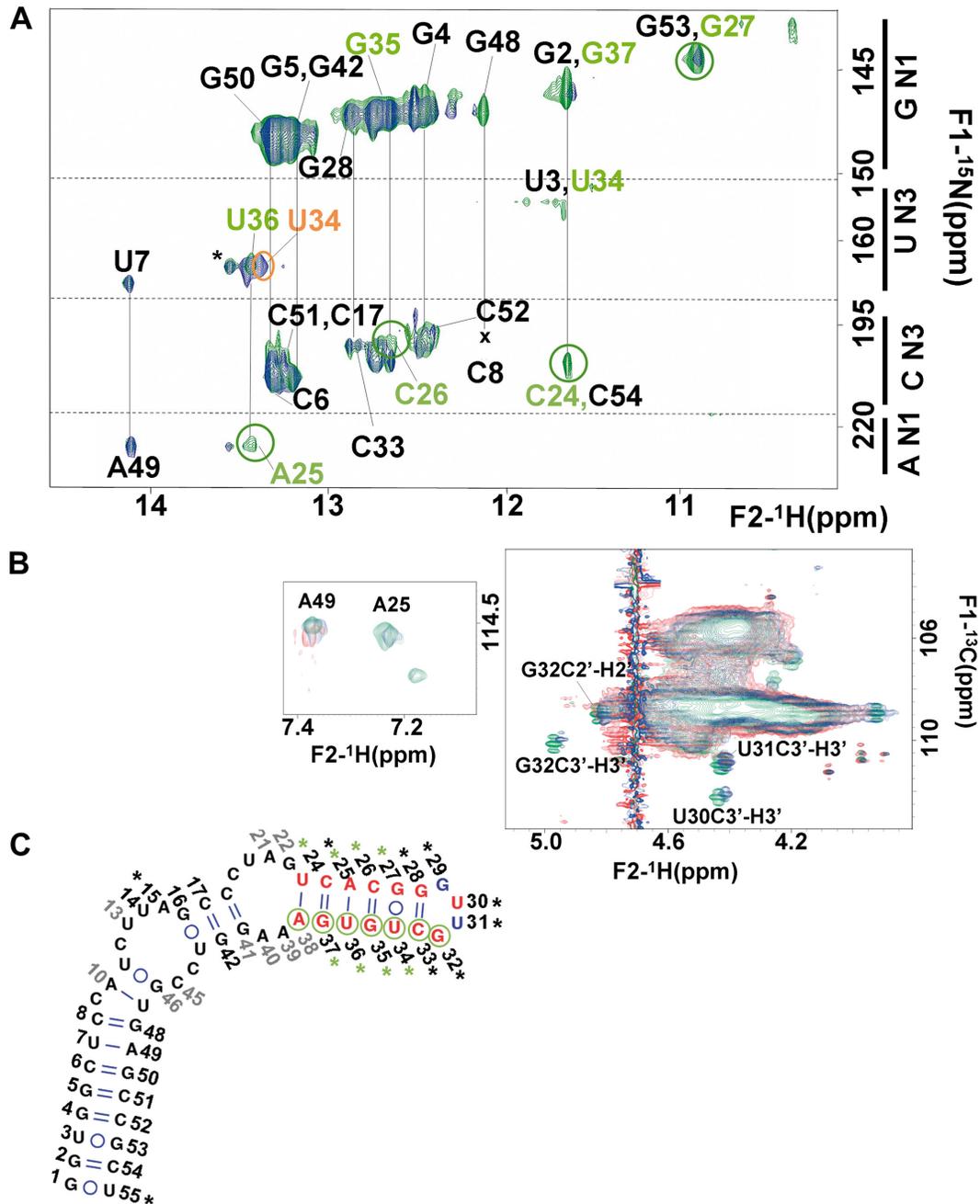
Gel electrophoresis experiments indicated that, relative to the full-length domain, subdomain SL2' had the same capacity to form complexes with 5BSL3.2 hairpin cre26 (Figure 2A) and NMR analyses indicated that the full-length

domain retained the structure composed of stems SL1' and SL2' upon binding to cre26 (Figure 3). Based on these observations, we next focused on the study of the complex formed by subdomain SL2' and cre26 to simplify the NMR spectra and obtain a better view of the molecular details of the interaction.

As with the 3'Xm system, we monitored the interaction between a  $^{13}\text{C}/^{15}\text{N}$ -labeled SL2'm sequence and an unlabeled cre26 construct in the presence of  $\text{Mg}^{2+}$ . Upon addition of cre26 a new SL2'm U HN3 imino crosspeak appeared without an A N1 partner in the HNN-COSY spectrum and with a chemical shift typical of Watson-Crick A:U pairs (Figure 5A). At the same time, the G HN1 crosspeak corresponding to the G27:U34 wobble pair of the upper SL2' stem weakened significantly. The new crosspeak was thus assigned to the intermolecular U34:A16' pair of the SL2'-cre26 duplex, which would replace the intramolecular G27:U34 pair in the upper SL2' duplex (Figure 4). Likewise, the HN1 imino crosspeaks of G35 and G37 lost their respective HNN-COSY C26 and C24 N3 connectivities, and the same was observed for the U36 HN3 imino, whose HNN-COSY A25 N1 partner became undetectable (Figure 5A). These same changes were detected when the 3'Xm:cre26 complex was formed. In contrast to the upper stem, the HNN-COSY crosspeak patterns of the base pairs forming the lower stem of subdomain SL2', specifically G4:C52, G5:C51, C6:G50 and U7:A49, remained unchanged in the presence of cre26 (Figure 5A). This indicated that the lower stem of subdomain SL2' was undisturbed upon cre26 complex formation, as observed with the full-length sequence (Figure 3).

Analyses of the non-exchangeable proton and carbon nuclei of SL2'm in the absence and presence of unlabeled cre26 confirmed the observations based on imino signals, and indicated that the upper region of the SL2'm hairpin underwent a significant conformational rearrangement with complex formation. In the isolated subdomain, the first nucleotide of the  $k$  segment, G32, is part of the G<sub>29</sub>UUG<sub>32</sub> apical tetraloop (Figure 1B), where it adopts an unusual *syn* conformation. In addition, the sugars of the tetraloop nt U30 and U31 have a strong C2'-*endo* character. Upon addition of unlabeled cre26, most of the resolved sugar signals of U30, U31 and G32 weakened or disappeared (Figure 5B), and significant chemical shift changes were detected in the resonances of G29 and C33. This indicated a conformational change in the apical stem-loop upon complex formation. The resonances of other nucleotides of the upper region of SL2'm with resolved resonances, including A15 and A25, also underwent significant perturbations (Figure 5B), suggesting that the rearrangement of the SL2' hairpin was not restricted to the apical stem-loop containing the  $k$  segment. In contrast, with the exception of the terminal U55 residue, we did not detect significant changes in the nucleotides of the lower stem with resolved resonances, specifically G1, G2, U3, G4, G5, C6, A49, G50, C51 and G53 (Figure 5C). These results supported the imino observations pointing to the preservation of the lower stem.

This latter conclusion was independently supported by analyzing with electrophoretic experiments the cre26-binding ability of a new subdomain SL2'm<sup>57</sup> sequence. In this construct, an additional G:C pair was added at the stem



**Figure 5.** NMR spectroscopy analysis of subdomain SL2'm in the absence and presence of 5BSL3.2 hairpin cre26. (A) Comparison of  $^1\text{H}$ - $^{15}\text{N}$  HNN-COSY spectra of construct SL2'm acquired in the absence (green) and presence (blue) of one molar equivalent of unlabeled cre26 sequence. Since the experiment used unlabeled cre26, all of the HN and HNN crosspeaks correspond to SL2'm. The crosspeaks of upper SL2' stem nucleotides are identified with green labels, and the HNN crosspeaks that disappear with the addition of cre26 are marked with green circles. The HN diagonal crosspeak that appears upon cre26 binding is marked in orange and was assigned as U34. The C8 HNN crosspeak marked with a cross is visible at a lower contour level in the absence of cre26. The crosspeak corresponding to a minor conformation of U36 (see Supplementary Figure S2) is marked with an asterisk. (B) Superposition of selected regions of the  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra of construct SL2'm acquired in the absence (green) and presence of one (blue) and two (red) molar equivalents of unlabeled cre26 sequence. In the C2'-H2 region (left), A25 H2 in the upper stem disappears when adding cre26, whereas A49 H2 (lower stem) remains relatively undisturbed; in the C2'/3'-H2'/3' region (right), the resolved sugar signals of the apical tetraloop residues U30, U31 and G32 weaken or disappear in the presence of cre26. (C) Secondary structure of isolated subdomain SL2'm supported by the NMR results. DLS nucleotides are depicted in red, *k* nucleotides are indicated with green circles, and the C<sub>29</sub>G and A<sub>31</sub>U mutations blocking dimerization are colored blue. Nucleotides with unambiguous and tentative assignments are identified with black and gray numbers, respectively, whereas unassigned nucleotides are not numbered. Unambiguously assigned nucleotides that undergo spectral changes in the presence of cre26 are marked with green asterisks (representing variations in HNN-COSY crosspeak patterns) and/or black asterisks (changes in carbon and/or non-exchangeable hydrogen resonances). Conditions: 81  $\mu\text{M}$  SL2'm, 2 mM MgCl<sub>2</sub>, 27°C (A); 101  $\mu\text{M}$  SL2'm, 2 mM MgCl<sub>2</sub>, 27°C (B).

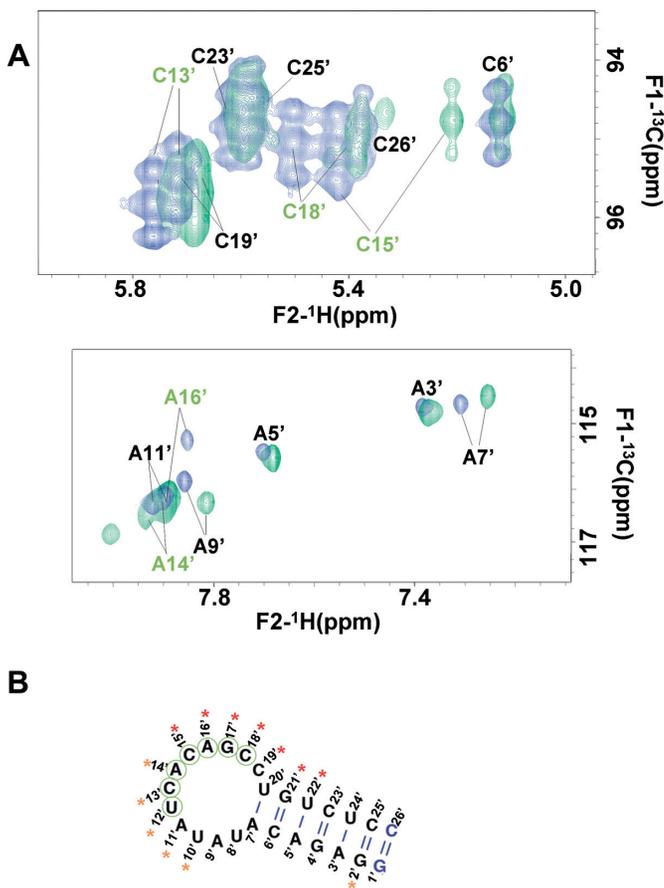
terminus and two G:U pairs were replaced with G:C, so that the lower stem of the SL2' hairpin was significantly stabilized (Figure 1B). This mutant had similar or greater cre26-binding capacity relative to dimerization-defective SL2'm or wild-type SL2' constructs, respectively (Figure 2). This result corroborated that the interaction with 5BSL3.2 did not imply a disruption of the lower double-helical stem of subdomain SL2'.

### The distal interaction primarily involves nucleotides located in the apical loop of 5BSL3.2

To observe the consequences of complex formation in subdomain 5BSL3.2, we examined the interaction between a  $^{13}\text{C}/^{15}\text{N}$ -labeled cre26 hairpin and an unlabeled SL2'm construct in the presence of  $\text{Mg}^{2+}$ . Analyses of the exchangeable and non-exchangeable proton resonances of the isolated cre26 construct indicated that this sequence formed a hairpin closed by a 12-nt  $\text{U}_8\text{AUAUCACAGCC}_{19}$  apical loop containing the 7-nt *k'* segment (underlined; Figure 6 and Supplementary Figure S7). This conclusion agreed with earlier NMR analyses of the 5BSL3.2 subdomain (27). Sequential as well as A H2 cross-strand NOEs indicated that this loop was structured and contained two segments of stacked nucleotides,  $\text{U}_8\text{-A}_{11}$  and  $\text{A}_{14}\text{-C}_{19}$ . Upon addition of unlabeled SL2'm, the imino hydrogens of the *k'* residues  $\text{U}_{12}$  and  $\text{G}_{17}$ , located in the apical loop of the isolated hairpin, were expected to form hydrogen bonds with their  $\text{A}_{38}$  and  $\text{C}_{33}$  *k* acceptors in the SL2'm sequence (Figure 4) and become visible. These imino resonances were not detected (data not shown), likely because of their proximity to the ends of the intermolecular duplex. However, significant chemical shift perturbations were detected in the carbon and non-exchangeable hydrogen resonances of almost all nucleotides of the cre26 apical loop in the presence of unlabeled SL2'm, and these perturbations were particularly strong for the  $\text{C}_{15}\text{-CAGC}_{18}$  residues of the *k'* segment as well as  $\text{C}_{19}$  (Figure 6). This result indicated that the interaction between cre26 and subdomain SL2' primarily involved the apical loop of hairpin 5BSL3.2, and supported the conclusions obtained with the analyses of the  $^{13}\text{C}/^{15}\text{N}$ -labeled 3'Xm and SL2'm constructs regarding the formation of the *k-k'* duplex.

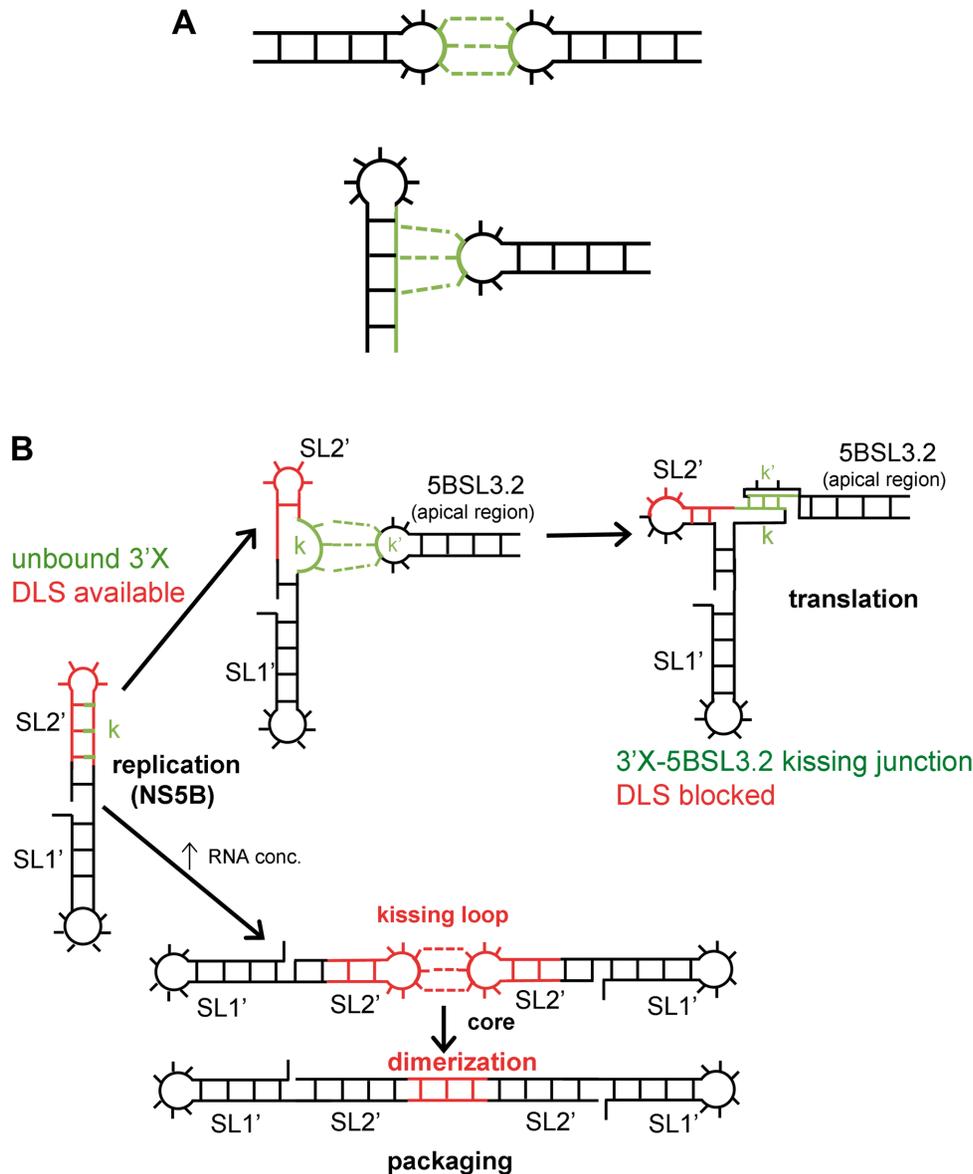
## DISCUSSION

The distal interaction between domain 3'X and hairpin 5BSL3.2 in the 3' region of the HCV genomic RNA has been found to be essential for the replication and translation processes of the virus (27–31). In this manuscript we show that the establishment of this contact implies the formation of a Watson–Crick duplex between *k* nucleotides base-paired in the upper stem of 3'X subdomain SL2' and unpaired *k'* nucleotides located in the terminal loop of 5BSL3.2. HNN-COSY analyses indicated that the interaction only affected nucleotides located in the upper region of SL2'; the domain retained the lower double-helical stem of this subdomain as well as the SL1' stem upon binding to 5BSL3.2 (Figure 3). This conclusion was confirmed by the study of the complex of subdomain construct SL2'm with 5BSL3.2 (Figure 5). As observed with the full-length



**Figure 6.** NMR spectroscopy analysis of 5BSL3.2 hairpin cre26 in the absence and presence of subdomain SL2'm. (A) Superposition of selected regions of the  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra of construct cre26 acquired in the absence (green) and presence (blue) of one molar equivalent of unlabeled SL2'm sequence. *k'* nucleotides are marked with green font. Note the strong shifts of *k'* nucleotides  $\text{C}_{15}$  and  $\text{C}_{18}$  in the C5-H5 region (top), as well as  $\text{A}_{14}$  and  $\text{A}_{16}$  in the C2-H2 section (bottom). (B) NMR-supported secondary structure of isolated hairpin cre26, with *k'* nucleotides indicated with green circles. Residues whose aromatic carbon and hydrogen resonances (C/H2, C/H5, C/H6 and/or C/H8) undergo significant chemical shift changes upon addition of one molar equivalent of unlabeled SL2'm are marked with orange (>1 standard deviation) and red (>2 standard deviation) asterisks. All cre26 nucleotides were unambiguously assigned. Conditions: 171  $\mu\text{M}$  cre26, 2 mM  $\text{MgCl}_2$ , 27°C.

sequence, HNN-COSY data showed that the interaction implied an exchange of base-pairing partners of the *k* nucleotides of the upper stem of SL2'm, whereas the lower stem remained relatively undisturbed (Figure 5A). Changes in carbon and non-exchangeable proton resonances indicated a significant conformational change in the upper region of subdomain SL2'm upon complex formation that did not include most nucleotides of the lower double-helical stem (Figure 5B). On the other hand, experiments focused on 5BSL3.2 revealed that the interaction primarily affected nucleotides of the apical loop of the hairpin where the *k'* segment is located (Figure 6). Genetic experiments have indicated that the 3'X–5BSL3.2 contact involved two complementary tracts of 7 nt:  $\text{G}_{32}\text{CUGUGA}_{38}$  of 3'X (*k*) and  $\text{U}_{12}\text{-CACAGC}_{18}$  of 5BSL3.2 (*k'*) (27). With our  $^{13}\text{C}/^{15}\text{N}$ -labeling strategy we have been able to detect via HNN-



**Figure 7.** Scheme summarizing the new RNA distal interaction mode identified in this study and its possible impact on the HCV life cycle. (A) In a classical kissing loop interaction (top), the bases of two apical loops form Watson–Crick pairs, whereas in the tertiary interaction mode identified in this study, the distal contact involves an apical loop and a base-paired stem. (B) Secondary structure model of the 3'X–5BSL3.2 interaction, together with the possible impact of the different 3'X structures on the HCV life cycle. The 3'X *k* bases directly involved in the distal contact are base-paired in the upper SL2' stem, and their complements are exchanged by intermolecular ones upon 5BSL3.2 complex formation. The lower SL2' stem and subdomain SL1' remain unaffected, and the rest of the nucleotides forming the upper half of subdomain SL2' may fold into a hairpin. The switching function of the 3'X domain is probably based on the fact that the DLS is exposed in the isolated domain but blocked upon 5BSL3.2 complex formation. In addition, the unusual structure of the 3'X–5BSL3.2 'kissing junction' proposed in this model may be a platform for protein binding.

COSY experiments the 4 bp internal segment of this intermolecular duplex, U<sub>34</sub>GUG<sub>37</sub>:C<sub>13</sub>ACA<sub>16</sub>' (Figure 4), probably because the outer pairs were subject to terminal effects that increased solvent exchange rates and hampered NMR detection of the corresponding imino protons.

On the basis of the abovementioned genetic experiments, it was proposed that the distal contact between domain 3'X and hairpin 5BSL3.2 was established through a canonical kissing-loop interaction involving the unpaired nucleotides of two terminal loops: the *k'* nucleotides located in the apical loop of 5BSL3.2, and the *k* nucleotides exposed in the

terminal loop of hairpin SL2, present in the three-stem conformation of domain 3'X (27–29,43,44) (Figure 1B, inset). However, we have not obtained NMR evidence supporting the formation of the tree-stem SL1, SL2 and SL3 subdomains in the presence of 5BSL3.2. Such rearrangement would imply disruption of the lower stem of SL2' as well as formation of additional base pairs at the base of subdomain SL1'. For the shorter subdomain constructs SL2' or SL2'm, it would generate an unpaired tail of 6 nt at their 3' termini (36). None of these changes was observed. Furthermore, the ability of SL2' subdomain construct SL2'm<sup>57</sup>, contain-

ing a significantly stabilized lower stem (Figure 1B), to bind 5BSL3.2 with the same affinity as the wild-type subdomain (Figure 2) confirmed that the interaction with 5BSL3.2 did not involve a disruption of the lower stem of subdomain SL2' (to form SL2 and SL3), but rather affected the upper region of the hairpin only. Thus, the 3'X–5BSL3.2 distal interaction takes place in an unexpected way: the 3'X tail retains the two stem conformation composed of subdomains SL1' and SL2' upon binding to 5BSL3.2, and complex formation implies disruption of the SL2' upper region where the *k* nucleotides are base-paired and the establishment of new Watson–Crick pairs with the *k*' nucleotides of the 5BSL3.2 terminal loop.

In order to bind to the 5BSL3.2 apical loop, the upper stem of subdomain SL2' must transiently adopt one or several open conformations. These activated states are probably facilitated by the presence of the SL2' internal loops (Figure 1B) and are short-lived, since they were not detected by NMR in the isolated 3'X or SL2' systems (36). In this respect, SAXS analyses indicated increased flexibility of the isolated SL2' subdomain relative to the full-length domain or subdomain SL1 (37). 3'X–5BSL3.2 complex formation likely occurs through the stabilization of these transient SL2' activated states by the 5BSL3.2 loop.

Taking the above results together, the 3'X–5BSL3.2 distal contact of HCV represents a novel RNA tertiary interaction mode that involves formation of Watson–Crick pairs between a base-paired stem and a terminal loop rather than between the unpaired nucleotides of two hairpin loops, as in canonical loop–loop interactions (Figure 7A). This mechanism can explain why replication was not rescued to wild-type levels with the double mutations introduced in the genetic study of Friebe *et al.* (27), since mutations involving *k* nucleotides (like C33G, A38U or G37C; see reference (27)) directly affected the stability of the upper SL2' stem involved in the distal contact (Figure 4), even when they were compensated by complementary mutations in the 5BSL3.2 loop.

Several authors have suggested that the two-stem and three-stem folds of the 3'X tail are part of an RNA-based switch signaling the transition between the replication, translation and possibly genome packaging processes of the virus (31,34,41–44,58). Since the presence of unpaired nucleotides after a stable secondary structure promotes primer-independent initiation of RNA synthesis by the viral NS5B polymerase (59), the two-stem fold would facilitate replication by exposing the three terminal nucleotides of the HCV genome at the end of the SL1' stem (36,37). This conformation would also promote viral RNA dimerization by exposing a palindromic DLS tract in the apical tetraloop of subdomain SL2' (36,37) that allows the establishment of the initial kissing contact between two palindromic loops (34). In contrast, the three-stem fold would conceal the palindromic and terminal nucleotides, and display the *k* sequence in the apical loop of the SL2 hairpin for 5BSL3.2 binding.

If formation of the 3'X–5BSL3.2 contact does not imply a reorganization of the 3'X tail secondary structure to form the three-stem conformation, how is this switching function accomplished? We propose that the switch function may reside in the fact that the DLS nucleotides driving

viral RNA dimerization are available in the unbound 3'X conformation, but blocked in the 3'X–5BSL3.2 complex as depicted in the model shown in Figure 7B. At the same time, the 3'X–5BSL3.2 contact likely affects the interaction of 5BSL3.2 with the internal ribosome entry site (60), modulating in this way the translation process (30). This model has some similarities with the recently proposed switching mechanism of the 5'-leader RNA region of HIV-1, which can also adopt two possible conformations. In the monomer conformer, the DLS palindromic loop of HIV-1 (termed in this case dimer initiation sequence, or DIS) is sequestered by base-pairing with the U5 segment of the leader, whereas in the dimer-promoting conformer, the U5 segment pairs with the start codon, blocking translation and the DIS hairpin becomes available for dimerization (61). Retroviruses pack two RNA genomes, and in these viruses RNA dimerization and packaging are closely linked (62,63). The role of RNA dimerization is not as well understood in HCV (32–34,58): it may be involved in packaging as in retroviruses, or contribute to the switching mechanism controlling the replication and translation processes, as explained above. In this respect, it has been recently reported that RNA dimerization is essential for HCV replication (26).

The unusual RNA tertiary interaction mechanism found for the 3'X–5BSL3.2 complex may allow an expansion of bioinformatics searches aimed to identify new functional RNA–RNA contacts. Moreover, the interaction between 3'X and 5BSL3.2 may result in the formation of a three-way kissing junction of the type depicted in Figure 7B, rather than the single composite helix usually generated by canonical loop–loop interactions. This is potentially important, as the unusual structure of the kissing junction might provide a binding platform for a *trans*-acting factor. Interestingly, formation of this intermolecular junction involves more SL2' nt than a simple loop–loop contact: since RNA residues involved in tertiary interactions usually exhibit a higher degree of sequence conservation (4), this would also be consistent with the high conservation of the 55-nt SL2' segment, which is absolutely invariable among HCV genotypes (35). Further work is currently in progress in our laboratory to determine the structural details of the 3'X–5BSL3.2 distal complex of HCV.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Martina Palomino for assistance with the NMR experiments, as well as Tamara Ovejero and Elisa Oltra for providing a p-GLICU plasmid containing the C<sub>29</sub>G/A<sub>31</sub>U mutant 3'X sequence.

## FUNDING

Ministerio de Economía y Competitividad of Spain [BFU2015–65103-R to J.G.]; Universidad Católica de Valencia of Spain [PRUCV/2015/629, 2017–114-001 to J.G.]. Funding for open access charge: Ministerio de Economía y Competitividad of Spain [BFU2015–65103-R].

*Conflict of interest statement.* None declared.

## REFERENCES

- Eddy, S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
- Sharp, P.A. (2009) The centrality of RNA. *Cell*, **136**, 577–580.
- Cech, T.R. and Steitz, J.A. (2014) The noncoding RNA revolution—trashing old rules to forge new ones. *Cell*, **157**, 77–94.
- Batey, R.T., Rambo, R.P. and Doudna, J.A. (1999) Tertiary motifs in RNA structure and folding. *Angew. Chem. Int. Ed. Engl.*, **38**, 2326–2343.
- Butcher, S.E. and Pyle, A.M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.*, **44**, 1302–1311.
- Brunel, C., Marquet, R., Romy, P. and Ehresmann, C. (2002) RNA loop-loop interactions as dynamic functional motifs. *Biochimie*, **84**, 925–944.
- Eguchi, Y. and Tomizawa, J. (1991) Complexes formed by complementary RNA stem-loops. Their formations, structures and interaction with Cole1 Rom protein. *J. Mol. Biol.*, **220**, 831–842.
- Argaman, L. and Altuvia, S. (2000) fhlA repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. *J. Mol. Biol.*, **300**, 1101–1112.
- Paillart, J.C., Shehu-Xhilaga, M., Marquet, R. and Mak, J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.*, **2**, 461–472.
- Nicholson, B.L. and White, K.A. (2014) Functional long-range RNA-RNA interactions in positive-strand RNA viruses. *Nat. Rev. Microbiol.*, **12**, 493–504.
- de la Pena, M., Dufour, D. and Gallego, J. (2009) Three-way RNA junctions with remote tertiary contacts: a recurrent and highly versatile fold. *RNA*, **15**, 1949–1964.
- Flores, R., Serra, P., Minoia, S., Di Serio, F. and Navarro, B. (2012) Viroids: from genotype to phenotype just relying on RNA sequence and structural motifs. *Front. Microbiol.*, **3**, 217.
- Li, P.T., Bustamante, C. and Tinoco, I. (2006) Unusual mechanical stability of a minimal RNA kissing complex. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 15847–15852.
- Bouchard, P. and Legault, P. (2014) A remarkably stable kissing-loop interaction defines substrate recognition by the Neurospora Varkud Satellite ribozyme. *RNA*, **20**, 1451–1464.
- Chang, K.Y. and Tinoco, I. (1997) The structure of an RNA “kissing” hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.*, **269**, 52–66.
- Lee, A.J. and Crothers, D.M. (1998) The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence. *Structure*, **6**, 993–1005.
- Mujeeb, A., Clever, J.L., Billici, T.M., James, T.L. and Parslow, T.G. (1998) Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat. Struct. Biol.*, **5**, 432–436.
- Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C. and Dumas, P. (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.*, **8**, 1064–1068.
- Bartenschlager, R. (ed). (2013) *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*. Springer-Verlag, Berlin Heidelberg.
- 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.
- 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.
- 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.
- 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.
- Yi, M. and Lemon, S.M. (2003) 3′ nontranslated RNA signals required for replication of hepatitis C virus RNA. *J. Virol.*, **77**, 3557–3568.
- 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.
- Masante, C., Jaubert, C., Palau, W., Plissonneau, J., Besnard, L., Ventura, M. and Di Primo, C. (2015) Mutations of the SL2 dimerization sequence of the hepatitis C genome abrogate viral replication. *Cell Mol. Life Sci.*, **72**, 3375–3385.
- 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.
- Murayama, A., Weng, L., Date, T., Akazawa, D., Tian, X., Suzuki, T., Kato, T., Tanaka, Y., Mizokami, M., Wakita, T. et al. (2010) RNA polymerase activity and specific RNA structure are required for efficient HCV replication in cultured cells. *PLoS Pathog.*, **6**, e1000885.
- You, S. and Rice, C.M. (2008) 3′ RNA elements in hepatitis C virus replication: kissing partners and long poly(U). *J. Virol.*, **82**, 184–195.
- Tuplin, A., Struthers, M., Cook, J., Bentley, K. and Evans, D.J. (2015) Inhibition of HCV translation by disrupting the structure and interactions of the viral CRE and 3′ X-tail. *Nucleic Acids Res.*, **43**, 2914–2926.
- Tuplin, A., Struthers, M., Simmonds, P. and Evans, D.J. (2012) A twist in the tail: SHAPE mapping of long-range interactions and structural rearrangements of RNA elements involved in HCV replication. *Nucleic Acids Res.*, **40**, 6908–6921.
- 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.
- Ivanyi-Nagy, R., Kanevsky, I., Gabus, C., Lavergne, J.P., Ficheux, D., Penin, F., Fossé, P. and Darlix, J.L. (2006) Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. *Nucleic Acids Res.*, **34**, 2618–2633.
- Shetty, S., Kim, S., Shimakami, T., Lemon, S.M. and Mihailescu, M.R. (2010) Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate. *RNA*, **16**, 913–925.
- 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.
- Cantero-Camacho, Á. and Gallego, J. (2015) The conserved 3′X terminal domain of hepatitis C virus genomic RNA forms a two-stem structure that promotes viral RNA dimerization. *Nucleic Acids Res.*, **43**, 8529–8539.
- Cantero-Camacho, A., Fan, L., Wang, Y.-X. and Gallego, J. (2017) Three-dimensional structure of the full-length 3′X-tail of hepatitis C virus RNA in monomeric and dimeric states. *RNA*, **23**, 1465–1476.
- Kranawetter, C., Brady, S., Sun, L., Schroeder, M., Chen, S.J. and Heng, X. (2017) Nuclear magnetic resonance study of RNA Structures at the 3′-End of the Hepatitis C Virus Genome. *Biochemistry*, **56**, 4972–4984.
- 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.
- 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.
- 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.
- Romero-López, C., Barroso-Deljesus, A., García-Sacristán, A., Briones, C. and Berzal-Herranz, A. (2014) End-to-end crosstalk within the hepatitis C virus genome mediates the conformational switch of the 3′X-tail region. *Nucleic Acids Res.*, **42**, 567–582.
- Shetty, S., Stefanovic, S. and Mihailescu, M.R. (2013) Hepatitis C virus RNA: molecular switches mediated by long-range RNA-RNA interactions? *Nucleic Acids Res.*, **41**, 2526–2540.
- Palau, W., Masante, C., Ventura, M. and Di Primo, C. (2013) Direct evidence for RNA-RNA interactions at the 3′ end of the Hepatitis C virus genome using surface plasmon resonance. *RNA*, **19**, 982–991.
- Lohmann, V., Körner, 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.

46. Lorenz, R., Bernhart, S.H., Höner Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
47. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
48. Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
49. Goddard, T.D. and Kneller, D.G. (2004) *Sparky 3.110*. University of California, San Francisco.
50. Dingley, A.J. and Grzesiek, S. (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (2)J(NN) couplings. *J. Am. Chem. Soc.*, **120**, 8293–8297.
51. Marino, J.P., Prestegard, J.H. and Crothers, D.M. (1994) Correlation of adenine H2/H8 resonances in uniformly <sup>13</sup>C labeled RNAs by 2D HCCH-TOCSY: a new tool for <sup>1</sup>H assignment. *J. Am. Chem. Soc.*, **116**, 2205–2206.
52. Peterson, R.D., Theimer, C.A., Wu, H. and Feigon, J. (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J. Biomol. NMR*, **28**, 59–67.
53. Varani, G., Aboul-Ela, F. and Allain, F.H.-T. (1996) NMR investigation of RNA structure. *Prog. Nucl. Magn. Reson. Spectrosc.*, **29**, 51–127.
54. Aeschbacher, T., Schmidt, E., Blatter, M., Maris, C., Duss, O., Allain, F.H., Güntert, P. and Schubert, M. (2013) Automated and assisted RNA resonance assignment using NMR chemical shift statistics. *Nucleic Acids Res.*, **41**, e172.
55. Clay, M.C., Ganser, L.R., Merriman, D.K. and Al-Hashimi, H.M. (2017) Resolving sugar puckers in RNA excited states exposes slow modes of repuckering dynamics. *Nucleic Acids Res.*, **45**, e134.
56. Kitamura, A., Jardine, P.J., Anderson, D.L., Grimes, S. and Matsuo, H. (2008) Analysis of intermolecular base pair formation of prohead RNA of the phage phi29 DNA packaging motor using NMR spectroscopy. *Nucleic Acids Res.*, **36**, 839–848.
57. Puglisi, E.V. and Puglisi, J.D. (2011) Secondary structure of the HIV reverse transcription initiation complex by NMR. *J. Mol. Biol.*, **410**, 863–874.
58. Holmstrom, E.D., Nettels, D. and Schuler, B. (2017) Conformational plasticity of hepatitis C virus core protein enables RNA-induced formation of nucleocapsid-like particles. *J. Mol. Biol.*, doi:10.1016/j.jmb.2017.10.010.
59. Kao, C.C., Yang, X., Kline, A., Wang, Q.M., Barket, D. and Heinz, B.A. (2000) Template requirements for RNA synthesis by a recombinant hepatitis C virus RNA-dependent RNA polymerase. *J. Virol.*, **74**, 11121–11128.
60. Romero-López, C. and Berzal-Herranz, A. (2009) A long-range RNA-RNA interaction between the 5' and 3' ends of the HCV genome. *RNA*, **15**, 1740–1752.
61. Keane, S.C. and Summers, M.F. (2016) NMR studies of the structure and function of the HIV-1 5'-leader. *Viruses*, **8**, 38.
62. Moore, M.D. and Hu, W.S. (2009) HIV-1 RNA dimerization: It takes two to tango. *AIDS Rev.*, **11**, 91–102.
63. Johnson, S.F. and Telesnitsky, A. (2010) Retroviral RNA dimerization and packaging: the what, how, when, where, and why. *PLoS Pathog.*, **6**, e1001007.