New frontiers in the investigation of structural functional RNA domains in viral genomes. Understanding the hepatitis C virus (HCV).

Alfredo Berzal-Herranz* and Cristina Romero-López*

Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC. PTS Granada, Armilla 18016. Granada. Spain.

*To whom correspondence should be addressed.
E-mail: aberzalh@ipb.csic.es; cristina_romero@ipb.csic.es

Abstract. RNA structure is used as a dynamic and robust supracoding system, beyond the nucleotidic sequence, to store critical information in a minimum size. The building blocks of this system are highly conserved folded genomic regions, the so-called functional RNA domains. They play, on their own, biological roles, such as interactions with proteins or nucleic acids. Therefore, deciphering the RNA structural code and understanding the structure-function relationship is critical for a deep knowledge of RNA function. This takes special relevance in the case of RNA viruses, which exhaustively make use of functional RNA domains to get replication efficient and compact genomes. The RNA genome of the hepatitis C virus (HCV) can be considered a prototype in the investigations about functional genomic regions. It contains multiple structurally conserved domains mostly located in the 5'- and 3'-untranslated regions (5'UTR, 3'UTR). These elements participate in HCV translation and replication by recruiting viral and host protein factors and also by establishing a complex and active all-RNA interaction network. Long-range RNA-RNA contacts are related to conformational rearrangements at the 5' end of the HCV genome and also to the structural transition in the 3'UTR from the replication-competent conformer to the dimerizable form. Together, the essentiality of functional RNA domains and their high conservation rate makes them potential therapeutic targets. This review provides an overview of those methodologies that have been mainly used for analyzing RNA folding and their application to the study of HCV molecular biology.

Keywords. Functional RNA domains; hepatitis C virus; RNA-RNA interactions; RNA structure; untranslated regions.

1 Introduction

In recent years, the genome of many organisms has been sequenced. These studies have revealed that protein coding genes are a low percentage of the total genetic information. Furthermore, the complexity and variety of biological functions cannot be explained only by protein-mediated catalysis. Therefore, additional elements must be coded into genomes to fill these gaps. The knowledge that RNA may act as a regulatory and catalytic element prompted numerous studies that finally demonstrated that RNA itself is a key component of the metabolic machinery in all the living organisms [1]. As for protein catalysis, RNA function depends on its three-dimensional conformation. RNA structure is determined by nucleotidic interactions that define loops and stems (the secondary structure). These are the building blocks to further generate helices duplexes, triple-stranded structures and other long-distant connections that finally yield the three-dimensional folding of an RNA molecule (tertiary structure) [2]. Structure enables RNA to recruit proteins, small ligands and metabolites; and even to interact with other nucleic acids or with itself. Understanding the biophysical relationship between structure and biological function is now the main goal in RNA investigations to know in depth the mechanisms underlying the regulation exerted by RNA molecules. For that purpose, several biochemical techniques have been developed. Together with the use of novel bioinformatic approaches, these methodologies provide an accurate way for deciphering RNA folding. These investigations can also help to expand our knowledge about important biological process controlled by RNA molecules, such as transcription elongation, splicing or translation [1].

The versatility and dynamism of RNA acquires special importance in RNA viruses. Their genome is a dynamic and sophisticated element that contains all the information required for viral propagation, but is packaged in a minimum size. This is achieved by overlapping different coding levels. Thus, besides the protein coding sequence, nucleotides also code for structural information that is translated into an intricate regulatory network governed by an all-RNA based mechanism. This system grants important advantages to RNA viruses. Thus, the existence of functional structural regions, the so-called functional RNA domains, provides a robust genetic background [3-5] to resist the inclusion of mutations during viral replication. Interestingly, the same features that achieve proficient viral pools also provide us a potential tool for destroying them. Thus, novel nucleic acids-based drugs targeting conserved functional genomic domains are now fundamental components in the antiviral toolbox [6].

This review will provide a brief summary of those biochemical techniques used for the elucidation of RNA folding and their application to a practical case, the hepatitis C virus genome. It will also argue the importance of RNA structure as a regulatory element that is critical for the consecution of the viral cycle.

2 Studying RNA folding

A major point to take into account for the understanding of RNA folding is that the acquisition of the three-dimensional conformation is a dynamic process. RNA has the ability to fold back in itself to adopt complex structures based on other simpler ele-

ments; it can also adopt different geometries, all of them with similar thermodynamic parameters and, therefore, equally probable, thus yielding complex conformational pools. The study of the final structure is, therefore, complicated and must follow a reliable route.

In the initial step of the analysis, comparative sequence phylogenetic studies are used to predict potential base-pairs in a wide range of RNA molecules [7-10]. These results are further confirmed by the application of biochemical procedures. Treating RNA with chemical or enzymatic reagents able to specifically modify or cleave nucleotides in a structure-dependent way is a common strategy for inferring the secondary and, in some cases, tertiary structure. Modified or cleaved positions are detected as stops in a reverse transcriptase-mediated primer extension reaction, and further resolved by electrophoresis (Fig. 1A). This methodology has been implemented by the incorporation of high-resolution capillary gel electrophoresis [11]. Reactions are compared to a non-reagent experiment and to a dideoxy sequencing marker for the identification of each nucleotide position [12, 13].

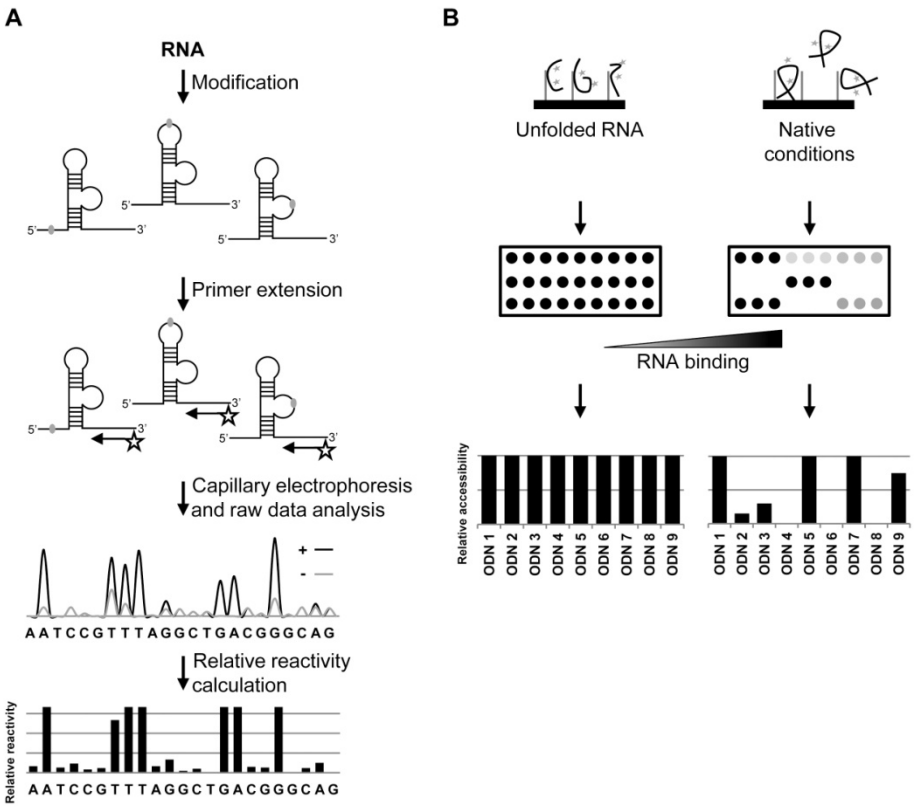

Fig. 1. RNA structure mapping.

One of the biophysical techniques that has gained relevance during last years is the SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) methodology [14, 15], a robust method for analyzing local nucleotide flexibility in RNA molecules [16]. This chemistry informs about the conformation of the ribose-phosphate backbone in all those dynamic residues, even if they remain solvent inaccessible. Thus, it may be used to ultimately define the overall geometry of an RNA molecule. SHAPE has been used to model the secondary structure of numerous RNA molecules [12], to detect conformational rearrangements [13, 17-19] and to monitor interactions with proteins [20] and small ligands [21, 22].

Either by chemical probing or by SHAPE analysis, relative reactivity quantification at each position is a requisite for inferring the secondary and tertiary RNA structure. This is a laborious work that can be accomplished by different softwares, such as CAFA [23], FAST [24], SHAPE-CE [25], HiTRACE [26] or the most widely used ShapeFinder software [27]. All these computational tools require many manual selections, which turns data analysis into a challenging and subjective work [28]. To overcome this, Karabiber et al. developed the QuShape package [29], which is a user-friendly implementation of previous algorithms.

The resulting experimental reactivity values can be further used to constrain the most probable secondary structure by using different software tools. Calculating pseudo-free energy change terms at each independent nucleotide from SHAPE experimental data yields robust structural models [30]. The algorithm is included in the RNAstructure tool [31] (http://rna.urmc.rochester.edu/RNAstructure.html). The computational package also integrates several modules that allow for calculating pseudoknots [32], predicting the master structure for a set of related sequences [33] or inferring the binding affinity of oligonucleotides to RNA target regions [34, 35], a useful tool for designing small interfering RNAs.

It is noteworthy that most of the *in silico* folding prediction strategies only consider canonical base-pairs (Watson-Crick and wobble interactions). This drawback has been overcome by the MC-Fold and MC-Sym pipeline [36], which considers the energetic contribution of every nucleotidic interaction by defining sets of nucleotide cyclic motifs. Thus, MC-Fold predicts secondary structures that can be further modeled to three-dimensional conformations by MC-Sym. The pipeline exhaustively explores the structural space of an RNA molecule and includes modifications to accommodate relative reactivity values derived from different chemical probing methods.

As well as modeling the secondary and tertiary structure of an RNA molecule, in some cases, it may also be useful to get some information about the ability of certain regions in the target RNA to interact with oligonucleotides. These data can be used for designing specific and efficient antisense oligos or siRNAs directed against long RNA molecules, which are difficult to model *in silico*. Antisense oligonucleotides microarray methodology has been designed to inform about the accessibility of consecutive, overlapping stretches of nucleotides [37]. It relies in the strong correlation existing between the native folding of a target RNA and its capacity to differentially interact with a set of complementary DNA oligonucleotides (Fig. 1B). Thus, this technique provides data about the overall three-dimensional folding of the molecule and defines the most favorable regions to be targeted. DNA microarrays have been successfully used for the analysis of RNA fragments derived from the hepatitis C

virus (HCV) [13, 19, 37], the foot-and-mouth disease virus (FMDV) [38, 39] and the human immunodeficiency virus (HIV) [40].
Together, the above mentioned biochemical and computational techniques provide complementary and not overlapping information about the two- and three-dimensional structure of a target RNA. The next sections will describe the practical use of these methodologies for investigating the complex interaction network in the genomic RNA of HCV and its potential as antiviral target.

3 Functional structural RNA domains in the HCV genome

The HCV genome is a ~ 9.6 kb long, single-stranded positive RNA molecule, encoding a single open reading frame (ORF) flanked by highly conserved untranslated regions (UTRs) [41-43]. Throughout the infective cycle, the genomic RNA actively participates in the execution of different steps by using functional domains mainly located at both the 5' and the 3' ends.
During early infection, viral protein synthesis is initiated by a highly structured element that functions as an internal ribosome entry site (IRES), mostly placed at the 5'UTR (Fig. 2) [44, 45]. The translation initiation mechanism used by HCV greatly differs to that used by most cellular mRNAs [46, 47] and is primarily accomplished by functional RNA domains [48] that replace the requirements of initiation protein factors. Importantly, the presence of structural elements at the 3' end of the viral genome may also modulate the initiation and elongation steps involved in HCV translation [49-54].
The secondary structure of the HCV IRES was firstly modeled by comparative sequence analysis and thermodynamic-based predictions [55]. Further application of biochemical probing techniques, SHAPE methodology, nuclear magnetic resonance (NMR) and X-ray crystallography have refined and completed the initial theoretical model. Under physiological magnesium conditions, the IRES folds as a dynamic and extended element with tightly compact regions [56]. The minimum IRES element is defined by two major domains, II and III, plus the short stem-loop IV containing the start translation codon (Fig. 2) [56]. By using SHAPE chemistry and X-ray crystallography, it has been shown that domains II and III are aligned at both sides of a double-pseudoknot structure (PK1 and PK2; Fig. 2) that organizes the overall folding of the IRES and directs the positioning of the start codon at the ribosomal P site [57]. Many of the essential structural motifs required for efficient initiation of translation reside in the highly branched domain III (Fig. 2). It is composed by six hairpins (designated from IIIa-IIIf) organized around three- and four-way junctions that define the binding platforms for the eIF3 (junction IIIabc) [58] and the 40S ribosomal subunit (junction IIIdef) [59]. The essential subdomain IIId is a highly conserved G-rich element that functions as the core binding center for the 40S subunit [60-63]. Chemical and enzymatic probing assays, SHAPE analyses, NMR and molecular dynamics studies have reported that subdomain IIId is a dynamic stem-loop with a rigid, asymmetric internal E-loop that resembles the sarcin-loop of the ribosome [13, 24, 64-66]. The capping apical loop contains a phylogenetically conserved GGG triplet and adopts a typical U-turn geometry that provides an improved interacting ability with proteins and nucleic

acids [60, 65]. Thus, the IRES could be considered as an all-RNA translation initiation factor.

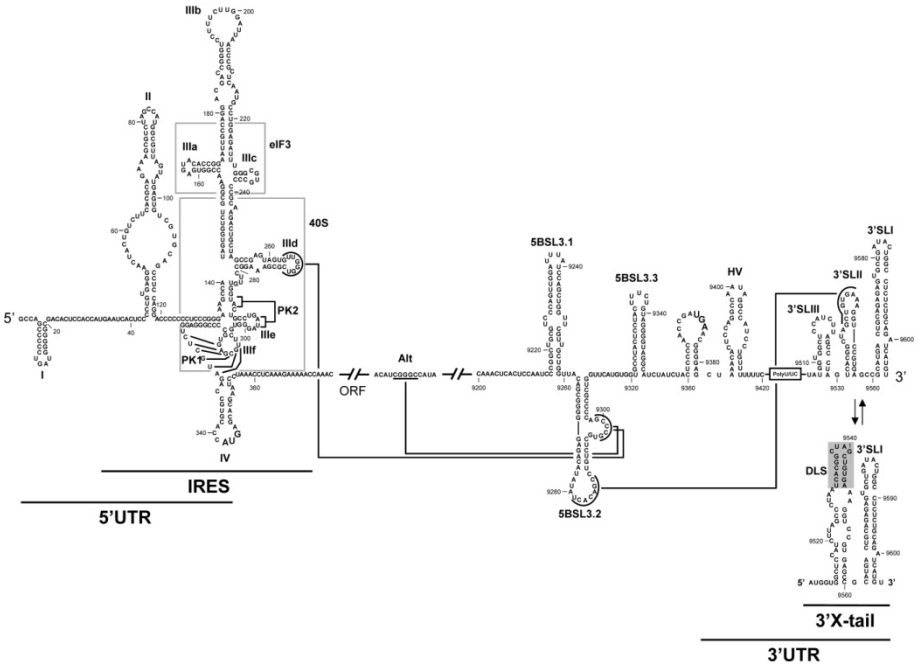

Fig. 2. Functional domains in the HCV genome.

The switch from translation to replication accounts at optimum viral proteins concentration levels. At this time, the genomic RNA moves to the endoplasmic reticulum to constitute the replication complex. The 3'UTR of the HCV genome contains phylogenetically conserved structural elements that recruit a number of viral and cellular factors required for viral replication and translation (Fig. 2) [67-77]. It also encompasses a palindromic sequence motif (dimer linkage sequence, DLS) that is related to viral genomic dimerization [78-80] (Fig. 2). Interestingly, this process is absent in the rest of the members of the *Flaviviridae* family. This suggests that it should play an additional role in the HCV cycle different to just promoting the encapsidation and release of the infective virions.

The essential 3'X-tail region occupies the very 3' terminus of the HCV RNA genome (Fig. 2). By using enzymatic probing assays, SHAPE analyses and *in silico* thermodynamic-based predictions, it has been shown that it folds into two mutually exclusive conformations [19, 79]. Both models preserve the essential, highly conserved 3'SLI domain at the very 3' end. The upstream segment swaps from the two stem-loops (3'SLII and 3'SLIII) conformation, which occludes the DLS motif, to the dimerizable form with one stem-loop exposing the DLS [80] (Fig. 2). Dimerization initiates by the establishment of an apical loop-apical loop interaction to yield a thermodynamically favored kissing complex, which could then progress to a stable extended duplex in the presence of the viral core chaperone protein *in vitro* [78-80].

The use of sequence phylogenetic studies and mutational analysis, as well as further application of thermodynamic folding computational tools, has discovered a set of conserved stem-loop structures placed at the 3' end of the coding sequence with potential roles in the progression of the viral cycle [81-85] (Fig. 2). Among them, the so-called 5BSL3.2 domain is indispensable for efficient replication [86, 87] and also regulates the IRES function, even in the presence of the translational enhancer 3'UTR [54]. The 5BSL3.2 domain is composed by two G-rich helixes, connected by an eight-base bulge and capped by an apical loop [86, 88] (Fig. 2). It is embedded into a higher-order cruciform structure and delimited by the two adjacent stem-loops 5BSL3.1 and 5BSL3.3. Importantly, both sequence and conformation are requisites for the efficient functioning of the 5BSL3.2 domain, which points to the establishment of interactions with viral and host protein factors [89, 90], as well as with other RNA domains of the HCV genome [84, 86, 91, 92].

4 The RNA-RNA interaction network tunes the folding of essential functional RNA domains in the HCV genome

Throughout the course of the infective cycle, transitions between different stages are finely regulated to achieve the adaptive fitness. In this context, the dynamism of RNA is a great advantage for the virus, which uses functional genomic domains for further regulatory activities.

The 5BSL3.2 domain is the prototype of a multi-function RNA element. Besides its role as recruiting agent for the viral RNA-dependent RNA polymerase [89], it has emerged as the central organizing partner for the establishment of a complex long-distance RNA-RNA interaction network that operates in the HCV genome (Fig. 2). The apical loop interacts with a complementary sequence that appears selectively exposed in the apical loop of domain 3'SLII, in the non-dimerizable 3'X-tail conformer. The resulting kissing loop contact is essential for HCV replication *in vitro* [86, 92]. SHAPE analyses performed by Tuplin *et al*. [92] have demonstrated that the structural consequences of this interaction are related to the viral genotype. Thus, in genotype 1 viruses, the 5BSL3.2 and 3'SLII domains fold as discrete stem-loops and the interaction seems to be absent; while in genotype 2 genomes, the contact formation is well-supported and induces a clear increase in NMIA reactivity for the residues composing the 3'SLII stem. This increase could be associated to some degree of disorder around the interacting sequences [92] or could be the result of profound conformational rearrangements related to the acquisition of the flexible dimerizable form in the 3'X-tail [19].

A second interaction involves the 8-nt bulge of the 5BSL3.2 domain and a complementary sequence placed around position 9110 (Alt sequence) (Fig. 2). Interestingly, the folding of this region is also affected by the viral genotype [92].

Numerous efforts have been aimed at elucidating whether these two interactions are mutually exclusive, or on the contrary, they occur simultaneously to form an extended pseudoknot. Interestingly, SHAPE assays performed by Tuplin *et al*. [92] support the notion that, in genotype 2 viral variants, the 5BSL3.2 domain forms the core of an extended pseudoknot in which both distant contacts account simultaneously but in an

independent fashion, i.e., disruption of one contact does not significantly affects to the other. By the opposite, in genotype 1 viruses, only the 5BSL3.2-Alt interaction seems to be stable enough to be detected by SHAPE chemistry.

In addition to organize the folding of the genomic 3' end, the bulge of the 5BSL3.2 domain also establishes a long-distance interaction with the apical loop of the subdomain IIId of the IRES region [91]. As occurred with the Alt-5BSL3.2 contact, in genotype 1 replication competent RNA transcripts the interaction IIId-5BSL3.2 induced a fine-tuning effect over the involved regions and surrounding residues [13, 19], as detected by chemical probing, SHAPE assays and antisense oligonucleotides microarrays. RNA structure modeling of subdomain IIId was performed using the MC-Fold/MC-Sym pipeline [36] for the replicative RNA (Rep) and compared to that obtained for a transcript containing the isolated IRES region (I) (Fig. 3) (Romero-López and Berzal-Herranz, unpublished results). The experimental constraints derived from chemical probing and SHAPE analyses were applied to achieve accurate predictions [13]. The resulting models predict major structural changes in the Rep construct with respect to variant I at the residues in the apical loop (local root-mean square deviation RMSD of 6.84 Å), with minor differences in the E loop. These structural reorganization events provide important clues regarding the implications for IRES function of conformational rearrangements mediated by the 3' end of the RNA genome and could be likely associated to the regulation of viral translation [13, 54].

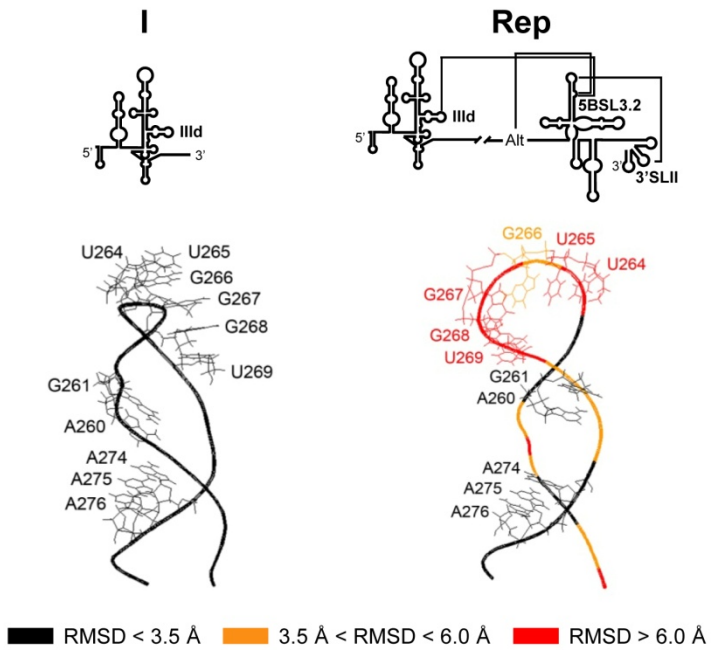

Fig. 3. RNA structural modeling of subdomain IIId.

It has been recently reported that all these three interactions are equally probable [93]. Therefore, choosing between different contacts might depend on the presence of addi-

tional host and/or viral proteins. An important consequence derived from the establishment of this interaction network is the induction of profound conformational rearrangements, not only in the directly involved residues and surrounding areas, but also in the rest of partners that compose such a network [19]. Thus, it has been shown that the IIId-5BSL3.2 contact regulates the structural swapping at the 3'X-tail to promote the acquisition of the dimerizable form [19].

All this information has been used to draw a working model that tries to explain the transitions between different steps of the viral cycle [19] (Fig. 4). In the early infection, the IRES would be mostly occupied by the translational machinery, thus favoring the contacts Alt-5BSL3.2 and 3'SLII-5BSL3.2. After protein synthesis, the binding of the NS5B and other protein factors to 3'X-tail and the 5BSL3.2 domain would promote a molecular context in which the 5BSL3.2-IIId and 5BSL3.2-Alt interactions could be equally feasible. Exchange between them could contribute to the creation of an enhanced replicative process [84] by repressing translation [54]. Further increase in the copy number of the viral RNA would alter this equilibrium to favor the acquisition of the dimerizable form exposing the DLS motif in the presence of the core chaperone protein [78].

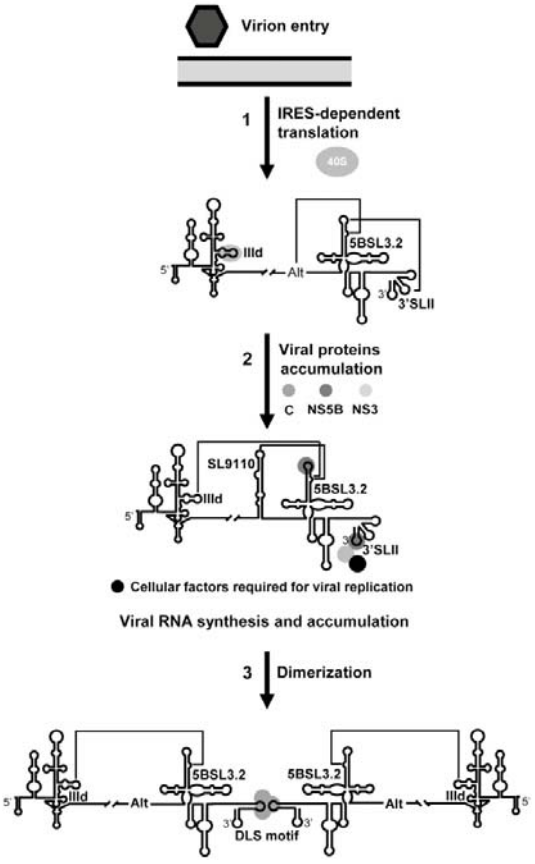

Fig. 4. Long-range RNA-RNA interactions in the HCV infective cycle.

Hence, the 5BSL3.2 domain would act as a nucleation partner to bring into close proximity both ends of the viral genome. The acquisition of such circular topology would favor the establishment of a cross-talk between both ends, with important benefits for the virus: promote an increase in the local concentration of essential proteins and cofactors and additional protection against the action of exonucleases. In addition, reducing the spatial distance between different functional domains would improve the regulation mediated by RNA elements.

5 Targeting HCV genomic RNA with RNA ligands

RNA viruses, such as HCV, present complex evolutionary replication dynamics that produce a wide spectrum of mutants. This feature greatly complicates the development of efficient antiviral drugs. In the case of HCV, the combined use of generic compounds, such as α-interferon, with modified nucleotides and/or direct-acting agents achieves viral sustained responses for short periods of time, finally leading to the appearance of resistant variants [94]. Therefore, designing novel therapeutic strategies and antiviral drugs is a major goal in HCV investigations.

From a wide point of view, targeting conserved structural and functional genomic domains with RNA molecules is an excellent approach. As it has been mentioned, these regions exhibit high genetic robustness and the use of different RNA-based compounds directed against multiple viral genomic elements might contribute greatly to reduce the appearance of resistant variants. Among the multiple antiviral strategies with nucleic acid-based inhibitors, the use of antisense oligonucleotides [6], small interfering RNAs (siRNAS) [95] and aptamers [96, 97] has rendered promising results [6]. These studies have also shown that several challenges must be overcome for the efficient use of RNA-based inhibitors, such as specific cell targeting, delivery and stabilization [98]. In this context, the advances in chemical synthesis have allowed for the incorporation of modified nucleotides that prevent nuclease-mediated degradation of the antiviral molecule while improve pharmacokinetic properties and diminish immunogenicity [99]. These modifications include chemical substitutions at the ribose 2' hydroxyl group, such as the inclusion of 2'-O-methyl, 2'-O-fluoro and 2'-O-methoxyethyl groups. Thus, new generation of modified nucleic acids-based drugs are currently being extensively developed and tested in clinical trials [6].

6 Conclusions

Investigations in the field of RNA folding have been mainly focused on deciphering the relationship between structure and function. Classical and novel methodologies are now actively contributing to understand the rules that govern RNA folding, conformational transitions and their regulatory roles. This is of key importance in RNA viruses, which use RNA structure to encode critical genetic information for adaptive viral fitness. Thus, complex and sophisticated control systems governed by RNA elements are being found in many viral genomes. These regions present high genetic

robustness and may be considered excellent candidates for novel antiviral nucleic acids-based strategies. Improvements in DNA and RNA synthesis will likely help to develop innovative compounds that target functional structured domains to achieve sustained therapeutic responses with minimal toxicity and secondary effects.

Abbreviations

DLS, dimer linkage sequence
HCV, hepatitis C virus
IRES, internal ribosome entry site
SHAPE, selective 2'-hydroxyl acylation analyzed by primer extension
siRNAs, small interfering RNAs
UTR, untranslated region

Competing interests

The authors declare that they have no competing interests

Author's contributions

Cristina Romero-López and Alfredo Berzal-Herranz wrote the paper.

Acknowledgements

Work in our laboratory is supported by grants from the Spanish Ministry of Economy and Competitiveness, BFU2012-31213 and *Junta de Andalucía*, CVI-7430, to A.B.-H.; and is partially supported by FEDER funds from the EU.

References

1. Mauger DM, Siegfried NA, Weeks KM: **The genetic code as expressed through relationships between mRNA structure and protein function.** *FEBS Lett* 2013, **587:**1180-1188.
2. Reiter NJ, Chan CW, Mondragon A: **Emerging structural themes in large RNA molecules.** *Curr Opin Struct Biol* 2011, **21:**319-326.
3. Wagner A, Stadler PF: **Viral RNA and evolved mutational robustness.** *J Exp Zool* 1999, **285:**119-127.
4. Contreras AM, Hiasa Y, He W, Terella A, Schmidt EV, Chung RT: **Viral RNA mutations are region specific and increased by ribavirin in a full-length hepatitis C virus replication system.** *J Virol* 2002, **76:**8505-8517.
5. Churkin A, Cohen M, Shemer-Avni Y, Barash D: **Bioinformatic analysis of the neutrality of RNA secondary structure elements across genotypes reveals evidence for direct evolution of genetic robustness in HCV.** *J Bioinform Comput Biol* 2010, **8:**1013-1026.
6. Haasnoot J, Berkhout B: **Nucleic acids-based therapeutics in the battle against pathogenic viruses.** *Handb Exp Pharmacol* 2009**:**243-263.

7. Cech TR: **Conserved sequences and structures of group I introns: building an active site for RNA catalysis--a review.** *Gene* 1988, **73:**259-271.
8. Michel F, Westhof E: **Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis.** *J Mol Biol* 1990, **216:**585-610.
9. Brown JW, Haas ES, James BD, Hunt DA, Liu JS, Pace NR: **Phylogenetic analysis and evolution of RNase P RNA in proteobacteria.** *J Bacteriol* 1991, **173:**3855-3863.
10. Romero DP, Blackburn EH: **A conserved secondary structure for telomerase RNA.** *Cell* 1991, **67:**343-353.
11. Deigan KE, Li TW, Mathews DH, Weeks KM: **Accurate SHAPE-directed RNA structure determination.** *Proc Natl Acad Sci USA* 2009, **106:**97-102.
12. Weeks KM, Mauger DM: **Exploring RNA Structural Codes with SHAPE Chemistry.** *Acc Chem Res* 2011, **44:**1280-1291.
13. Romero-López C, Barroso-delJesus A, García-Sacristán A, Briones C, Berzal-Herranz A: **The folding of the hepatitis C virus internal ribosome entry site depends on the 3'-end of the viral genome.** *Nucleic Acids Res* 2012, **40:**11697-11713.
14. Chamberlin SI, Weeks KM: **Differential helix stabilities and sites pre-organized for tertiary interactions revealed by monitoring local nucleotide flexibility in the bI5 group I intron RNA.** *Biochemistry* 2003, **42:**901-909.
15. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM: **RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE).** *J Am Chem Soc* 2005, **127:**4223-4231.
16. McGinnis JL, Dunkle JA, Cate JH, Weeks KM: **The mechanisms of RNA SHAPE chemistry.** *J Am Chem Soc* 2012, **134:**6617-6624.
17. Julien KR, Sumita M, Chen PH, Laird-Offringa IA, Hoogstraten CG: **Conformationally restricted nucleotides as a probe of structure-function relationships in RNA.** *RNA* 2008, **14:**1632-1643.
18. Mortimer SA, Weeks KM: **Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution.** *Nat Protoc* 2009, **4:**1413-1421.
19. Romero-López C, Barroso-delJesus A, García-Sacristán A, Briones C, Berzal-Herranz A: **End-to-end crosstalk within the hepatitis C virus genome mediates the conformational switch of the 3'X-tail region.** *Nucleic Acids Res* 2014, **42:**567-582.
20. McGinnis JL, Duncan CD, Weeks KM: **High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly.** *Methods Enzymol* 2009, **468:**67-89.
21. Wang B, Wilkinson KA, Weeks KM: **Complex ligand-induced conformational changes in tRNA(Asp) revealed by single-nucleotide resolution SHAPE chemistry.** *Biochemistry* 2008, **47:**3454-3461.
22. Wrzesinski J, Blaszczyk L, Wronska M, Kasprowicz A, Stokowa-Soltys K, Nagaj J, Szafraniec M, Kulinski T, Jezowska-Bojczuk M, Ciesiolka J: **Mapping the interactions of selected antibiotics and their Cu2+**

complexes with the antigenomic delta ribozyme. *FEBS J* 2013, **280:**2652-2664.

23. Mitra S, Shcherbakova IV, Altman RB, Brenowitz M, Laederach A: **High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis.** *Nucleic Acids Res* 2008, **36:**e63.
24. Pang PS, Elazar M, Pham EA, Glenn JS: **Simplified RNA secondary structure mapping by automation of SHAPE data analysis.** *Nucleic Acids Res* 2011, **39:**e151.
25. Aviran S, Trapnell C, Lucks JB, Mortimer SA, Luo S, Schroth GP, Doudna JA, Arkin AP, Pachter L: **Modeling and automation of sequencing-based characterization of RNA structure.** *Proc Natl Acad Sci USA* 2011, **108:**11069-11074.
26. Yoon S, Kim J, Hum J, Kim H, Park S, Kladwang W, Das R: **HiTRACE: high-throughput robust analysis for capillary electrophoresis.** *Bioinformatics* 2011, **27:**1798-1805.
27. Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC: **ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis.** *RNA* 2008, **14:**1979-1990.
28. Leonard CW, Hajdin CE, Karabiber F, Mathews DH, Favorov OV, Dokholyan NV, Weeks KM: **Principles for understanding the accuracy of SHAPE-directed RNA structure modeling.** *Biochemistry* 2013, **52:**588-595.
29. Karabiber F, McGinnis JL, Favorov OV, Weeks KM: **QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis.** *RNA* 2013, **19:**63-73.
30. Gherghe CM, Shajani Z, Wilkinson KA, Varani G, Weeks KM: **Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA.** *J Am Chem Soc* 2008, **130:**12244-12245.
31. Reuter JS, Mathews DH: **RNAstructure: software for RNA secondary structure prediction and analysis.** *BMC Bioinformatics* 2010, **11:**129.
32. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM: **Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots.** *Proc Natl Acad Sci USA* 2013, **110:**5498-5503.
33. Xu Z, Mathews DH: **Multilign: an algorithm to predict secondary structures conserved in multiple RNA sequences.** *Bioinformatics* 2011, **27:**626-632.
34. Mathews DH, Burkard ME, Freier SM, Wyatt JR, Turner DH: **Predicting oligonucleotide affinity to nucleic acid targets.** *RNA* 1999, **5:**1458-1469.
35. Matveeva OV, Mathews DH, Tsodikov AD, Shabalina SA, Gesteland RF, Atkins JF, Freier SM: **Thermodynamic criteria for high hit rate antisense oligonucleotide design.** *Nucleic Acids Res* 2003, **31:**4989-4994.
36. Parisien M, Major F: **The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data.** *Nature* 2008, **452:**51-55.
37. Martell M, Briones C, de Vicente A, Piron M, Esteban JI, Esteban R, Guardia J, Gomez J: **Structural analysis of hepatitis C RNA genome using DNA microarrays.** *Nucleic Acids Res* 2004, **32:**e90.

38. Fernández N, Fernandez-Miragall O, Ramajo J, García-Sacristán A, Bellora N, Eyras E, Briones C, Martínez-Salas E: **Structural basis for the biological relevance of the invariant apical stem in IRES-mediated translation.** *Nucleic Acids Res* 2011, **39:**8572-8585.
39. Fernández N, García-Sacristán A, Ramajo J, Briones C, Martínez-Salas E: **Structural analysis provides insights into the modular organization of picornavirus IRES.** *Virology* 2011, **409:**251-261.
40. Ooms M, Verhoef K, Southern E, Huthoff H, Berkhout B: **Probing alternative foldings of the HIV-1 leader RNA by antisense oligonucleotide scanning arrays.** *Nucleic Acids Res* 2004, **32:**819-827.
41. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M: **Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome.** *Science* 1989, **244:**359-362.
42. Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, Shimotohno K: **Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis.** *Proc Natl Acad Sci USA* 1990, **87:**9524-9528.
43. Takamizawa A, Mori C, Fuke I, Manabe S, Murakami S, Fujita J, Onishi E, Andoh T, Yoshida I, Okayama H: **Structure and organization of the hepatitis C virus genome isolated from human carriers.** *J Virol* 1991, **65:**1105-1113.
44. Reynolds JE, Kaminski A, Kettinen HJ, Grace K, Clarke BE, Carroll AR, Rowlands DJ, Jackson RJ: **Unique features of internal initiation of hepatitis C virus RNA translation.** *EMBO J* 1995, **14:**6010-6020.
45. Wang TH, Rijnbrand RC, Lemon SM: **Core protein-coding sequence, but not core protein, modulates the efficiency of cap-independent translation directed by the internal ribosome entry site of hepatitis C virus.** *J Virol* 2000, **74:**11347-11358.
46. Tsukiyama-Kohara K, Iizuka N, Kohara M, Nomoto A: **Internal ribosome entry site within hepatitis C virus RNA.** *J Virol* 1992, **66:**1476-1483.
47. Wang C, Sarnow P, Siddiqui A: **Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism.** *J Virol* 1993, **67:**3338-3344.
48. Lukavsky PJ: **Structure and function of HCV IRES domains.** *Virus Res* 2009, **139:**166-171.
49. Ito T, Tahara SM, Lai MM: **The 3'-untranslated region of hepatitis C virus RNA enhances translation from an internal ribosomal entry site.** *J Virol* 1998, **72:**8789-8796.
50. McCaffrey AP, Ohashi K, Meuse L, Shen S, Lancaster AM, Lukavsky PJ, Sarnow P, Kay MA: **Determinants of hepatitis C translational initiation *in vitro*, in cultured cells and mice.** *Mol Ther* 2002, **5:**676-684.
51. Bradrick SS, Walters RW, Gromeier M: **The hepatitis C virus 3'-untranslated region or a poly(A) tract promote efficient translation subsequent to the initiation phase.** *Nucleic Acids Res* 2006, **34:**1293-1303.
52. Song Y, Friebe P, Tzima E, Junemann C, Bartenschlager R, Niepmann M: **The hepatitis C virus RNA 3'-untranslated region strongly enhances**

translation directed by the internal ribosome entry site. *J Virol* 2006, **80:**11579-11588.
53. Bung C, Bochkaeva Z, Terenin I, Zinovkin R, Shatsky IN, Niepmann M: **Influence of the hepatitis C virus 3'-untranslated region on IRES-dependent and cap-dependent translation initiation.** *FEBS Lett* 2010, **584:**837-842.
54. Romero-López C, Berzal-Herranz A: **The functional RNA domain 5BSL3.2 within the NS5B coding sequence influences hepatitis C virus IRES-mediated translation.** *Cell Mol Life Sci* 2012, **69:**103-113.
55. Brown EA, Zhang H, Ping LH, Lemon SM: **Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs.** *Nucleic Acids Res* 1992, **20:**5041-5045.
56. Kieft JS, Zhou K, Jubin R, Murray MG, Lau JY, Doudna JA: **The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold.** *J Mol Biol* 1999, **292:**513-529.
57. Berry KE, Waghray S, Mortimer SA, Bai Y, Doudna JA: **Crystal structure of the HCV IRES central domain reveals strategy for start-codon positioning.** *Structure* 2011, **19:**1456-1466.
58. Kieft JS, Zhou K, Grech A, Jubin R, Doudna JA: **Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation.** *Nat Struct Biol* 2002, **9:**370-374.
59. Lukavsky PJ, Kim I, Otto GA, Puglisi JD: **Structure of HCV IRES domain II determined by NMR.** *Nat Struct Biol* 2003, **10:**1033-1038.
60. Jubin R, Vantuno NE, Kieft JS, Murray MG, Doudna JA, Lau JY, Baroudy BM: **Hepatitis C virus internal ribosome entry site (IRES) stem loop IIId contains a phylogenetically conserved GGG triplet essential for translation and IRES folding.** *J Virol* 2000, **74:**10430-10437.
61. Kolupaeva VG, Pestova TV, Hellen CU: **An enzymatic footprinting analysis of the interaction of 40S ribosomal subunits with the internal ribosomal entry site of hepatitis C virus.** *J Virol* 2000, **74:**6242-6250.
62. Odreman-Macchioli FE, Tisminetzky SG, Zotti M, Baralle FE, Buratti E: **Influence of correct secondary and tertiary RNA folding on the binding of cellular factors to the HCV IRES.** *Nucleic Acids Res* 2000, **28:**875-885.
63. Babaylova E, Graifer D, Malygin A, Stahl J, Shatsky I, Karpova G: **Positioning of subdomain IIId and apical loop of domain II of the hepatitis C IRES on the human 40S ribosome.** *Nucleic Acids Res* 2009, **37:**1141-1151.
64. Klinck R, Westhof E, Walker S, Afshar M, Collier A, Aboul-Ela F: **A potential RNA drug target in the hepatitis C virus internal ribosomal entry site.** *RNA* 2000, **6:**1423-1431.
65. Lukavsky PJ, Otto GA, Lancaster AM, Sarnow P, Puglisi JD: **Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function.** *Nat Struct Biol* 2000, **7:**1105-1110.
66. Golebiowski J, Antonczak S, Di-Giorgio A, Condom R, Cabrol-Bass D: **Molecular dynamics simulation of hepatitis C virus IRES IIId domain: structural behavior, electrostatic and energetic analysis.** *J Mol Model* 2004, **10:**60-68.

67. Yi M, Lemon SM: **3' nontranslated RNA signals required for replication of hepatitis C virus RNA.** *J Virol* 2003, **77:**3557-3568.
68. Kolykhalov AA, Mihalik K, Feinstone SM, Rice CM: **Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication *in vivo*.** *J Virol* 2000, **74:**2046-2051.
69. Friebe P, Bartenschlager R: **Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication.** *J Virol* 2002, **76:**5326-5338.
70. Huang L, Hwang J, Sharma SD, Hargittai MR, Chen Y, Arnold JJ, Raney KD, Cameron CE: **Hepatitis C virus non-structural protein 5A (NS5A) is a RNA-binding protein.** *J Biol Chem* 2005, **280:**36417-36428.
71. Harris D, Zhang Z, Chaubey B, Pandey VN: **Identification of cellular factors associated with the 3'-nontranslated region of the hepatitis C virus genome.** *Mol Cell Proteomics* 2006, **5:**1006-1018.
72. Tingting P, Caiyun F, Zhigang Y, Pengyuan Y, Zhenghong Y: **Subproteomic analysis of the cellular proteins associated with the 3' untranslated region of the hepatitis C virus genome in human liver cells.** *Biochem Biophys Res Commun* 2006, **347:**683-691.
73. Kim CS, Seol SK, Song OK, Park JH, Jang SK: **An RNA-binding protein, hnRNP A1, and a scaffold protein, septin 6, facilitate hepatitis C virus replication.** *J Virol* 2007, **81:**3852-3865.
74. Scheller N, Mina LB, Galao RP, Chari A, Gimenez-Barcons M, Noueiry A, Fischer U, Meyerhans A, Diez J: **Translation and replication of hepatitis C virus genomic RNA depends on ancient cellular proteins that control mRNA fates.** *Proc Natl Acad Sci USA* 2009, **106:**13517-13522.
75. Weinlich S, Huttelmaier S, Schierhorn A, Behrens SE, Ostareck-Lederer A, Ostareck DH: **IGF2BP1 enhances HCV IRES-mediated translation initiation via the 3'UTR.** *RNA* 2009, **15:**1528-1542.
76. Yu KL, Jang SI, You JC: **Identification of *in vivo* interaction between Hepatitis C Virus core protein and 5' and 3' UTR RNA.** *Virus Res* 2009, **145:**285-292.
77. Bai Y, Zhou K, Doudna JA: **Hepatitis C virus 3'UTR regulates viral translation through direct interactions with the host translation machinery.** *Nucleic Acids Res* 2013, **41:**7861-7874.
78. Cristofari G, Ivanyi-Nagy R, Gabus C, Boulant S, Lavergne JP, Penin F, Darlix JL: **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* 2004, **32:**2623-2631.
79. Ivanyi-Nagy R, Kanevsky I, Gabus C, Lavergne JP, Ficheux D, Penin F, Fosse P, Darlix JL: **Analysis of hepatitis C virus RNA dimerization and core-RNA interactions.** *Nucleic Acids Res* 2006, **34:**2618-2633.
80. Shetty S, Kim S, Shimakami T, Lemon SM, Mihailescu MR: **Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate.** *RNA* 2010, **16:**913-925.

81. Smith DB, Simmonds P: **Characteristics of nucleotide substitution in the hepatitis C virus genome: constraints on sequence change in coding regions at both ends of the genome.** *J Mol Evol* 1997, **45:**238-246.
82. Hofacker IL, Fekete M, Flamm C, Huynen MA, Rauscher S, Stolorz PE, Stadler PF: **Automatic detection of conserved RNA structure elements in complete RNA virus genomes.** *Nucleic Acids Res* 1998, **26:**3825-3836.
83. Tuplin A, Wood J, Evans DJ, Patel AH, Simmonds P: **Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus.** *RNA* 2002, **8:**824-841.
84. Diviney S, Tuplin A, Struthers M, Armstrong V, Elliott RM, Simmonds P, Evans DJ: **A hepatitis C virus *cis*-acting replication element forms a long-range RNA-RNA interaction with upstream RNA sequences in NS5B.** *J Virol* 2008, **82:**9008-9022.
85. Chu D, Ren S, Hu S, Wang WG, Subramanian A, Contreras D, Kanagavel V, Chung E, Ko J, Amirtham Jacob Appadorai RS, et al: **Systematic analysis of enhancer and critical *cis*-acting RNA elements in the protein-encoding region of the hepatitis C virus genome.** *J Virol* 2013, **87:**5678-5696.
86. Friebe P, Boudet J, Simorre JP, Bartenschlager R: **Kissing-loop interaction in the 3' end of the hepatitis C virus genome essential for RNA replication.** *J Virol* 2005, **79:**380-392.
87. You S, Rice CM: **3' RNA elements in hepatitis C virus replication: kissing partners and long poly(U).** *J Virol* 2008, **82:**184-195.
88. You S, Stump DD, Branch AD, Rice CM: **A *cis*-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication.** *J Virol* 2004, **78:**1352-1366.
89. Zhang J, Yamada O, Sakamoto T, Yoshida H, Araki H, Murata T, Shimotohno K: **Inhibition of hepatitis C virus replication by pol III-directed overexpression of RNA decoys corresponding to stem-loop structures in the NS5B coding region.** *Virology* 2005, **342:**276-285.
90. Oakland TE, Haselton KJ, Randall G: **EWSR1 binds the hepatitis C virus *cis*-acting replication element and is required for efficient viral replication.** *J Virol* 2013, **87:**6625-6634.
91. Romero-López C, Berzal-Herranz A: **A long-range RNA-RNA interaction between the 5' and 3' ends of the HCV genome.** *RNA* 2009, **15:**1740-1752.
92. Tuplin A, Struthers M, Simmonds P, Evans DJ: **A twist in the tail: SHAPE mapping of long-range interactions and structural rearrangements of RNA elements involved in HCV replication.** *Nucleic Acids Res* 2012, **40:**6908-6921.
93. Shetty S, Stefanovic S, Mihailescu MR: **Hepatitis C virus RNA: molecular switches mediated by long-range RNA-RNA interactions?** *Nucleic Acids Res* 2013, **41:**2526-2540.
94. Zhu Y, Chen S: **Antiviral treatment of hepatitis C virus infection and factors affecting efficacy.** *World J Gastroenterol* 2013, **19:**8963-8973.
95. Blake SJ, Bokhari FF, McMillan NA: **RNA interference for viral infections.** *Curr Drug Targets* 2012, **13:**1411-1420.

96. James W: **Aptamers in the virologists' toolkit.** *J Gen Virol* 2007, **88:**351-364.
97. Sundaram P, Kurniawan H, Byrne ME, Wower J: **Therapeutic RNA aptamers in clinical trials.** *Eur J Pharm Sci* 2013, **48:**259-271.
98. Rossi JJ: **Controlled, targeted, intracellular expression of ribozymes: progress and problems.** *Trends Biotechnol* 1995, **13:**301-306.
99. Wang RE, Wu H, Niu Y, Cai J: **Improving the stability of aptamers by chemical modification.** *Curr Med Chem* 2011, **18:**4126-4138.

Figure legends

Fig. 1. RNA structure mapping. A) RNA folding analysis by chemical probing or SHAPE analysis. The RNA is treated with chemical reagents that modify nucleotides at specific positions. RNA modifications act as stop signals in a reverse-transcription reaction. Fluorescently labeled color-coded primers are used to further map each modified residue. The resulting cDNA products are resolved by automated capillary electrophoresis. Raw data are scaled and normalized to finally yield the relative reactivity values at each nucleotide. B) Antisense oligonucleotides microarray assays. Under native folding conditions, the RNA under study is fluorescently labeled and hybridized with a customized panel of antisense, overlapping DNA oligonucleotides. Differential hybridization ability for each oligonucleotide is related to different solvent exposure at the target region. Fluorescent signal is quantified and normalized to render the relative accessibility pattern.

Fig. 2. Functional domains in the HCV genome. Figure show the secondary structure proposed for the 5' and the 3' ends of the HCV genomic RNA and the long-range RNA-RNA interactions established between distant regions. At the 5' terminus, the minimum region for IRES activity is depicted. The entire 3' end contains the 3'UTR plus the stem-loops 5BSL3.1-5BSL3.3 and the Alt sequence motif at the NS5B coding sequence. The 3'X-tail folds into two different conformers with distinct functional roles. Dimer linkage sequence (DLS) is shown in grey. Pseudoknot elements are indicated as PK1 and PK2. The translations start and stop codons are shown in bold. Nucleotide numbering corresponds to HCV Con1 isolate.

Fig. 3. RNA structural modeling of subdomain IIId. PDB RNA structure prediction of subdomain IIId using the MC-Fold/MC-Sym pipeline. Experimental relative reactivity values were used to construct three dimensional models for subdomain IIId in the transcript I and the replicative RNA. Root mean-square deviation (RMSD) value was calculated from the comparison of the subdomain IIId between both molecules in order to infer important differences in the stem-loop conformation. Color code: black, residues with a RMSD <3.5 Å with respect to molecule I; orange, nucleotides with a RMSD ranging from 3.5 to 6.0 Å with respect to I; red, residues with a RMSD >6.0 Å.

Fig. 4. Long-range RNA-RNA interactions in the HCV infective cycle. 1) During early infection, the naked genomic RNA initiates viral translation by an IRES-dependent mechanism by directly recruiting the 40S ribosomal subunit at the subdomain IIId. This avoids the interaction IIId-5BSL3.2 and enhances the conformational reorganization of the 3' end mediated by the 5BSL3.2 domain. 2) Viral protein accumulation unleashes the transition toward the replication step. The ribosome is released from the IRES while protein factors bind to the 3'SLII. This favors a translational-repressed state by the establishment of the IIId-5BSL3.2 contact and an enhanced replication process dependent on the interaction Alt-5BSL3.2. 3) The newly synthesized RNA molecules with the preferred interaction IIId-5BSL3.2 would expose the dimer linkage sequence (DLS) in an apical loop. In the presence of the core chaperone protein, dimerization is favored in these conformers. Figure adapted from [19].